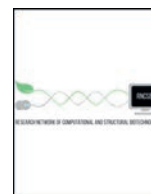




COMPUTATIONAL
AND STRUCTURAL
BIOTECHNOLOGY
JOURNAL

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



Mini Review

Reconsidering plasmid maintenance factors for computational plasmid design

Hirokazu Yano ^{a,*}, 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, Imizu, Toyama 939-0398, Japan

ARTICLE INFO

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

ABSTRACT

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).

| | |
|----------------------------------|----|
| Competing Interests | 78 |
| Authors' Contributions | 78 |
| Acknowledgements | 78 |
| References | 78 |

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 *Escherichia coli* and other gram-negative bacteria. Reviews of the mechanisms of action of each element of a plasmid's basic function can be found elsewhere ([4,5] for partition, [6,7] for transfer, [8–10] for replication, and [11] for toxin-antitoxin mechanisms).

In this review, incompatibility (Inc) group classification is used to refer to plasmid groups [12,13]. Inc. groups and representative plasmid vectors relevant to gram-negative bacteria are listed in Table 1. Different plasmids belonging to the same Inc. group are incompatible and unable to be inherited in a single bacterial cell line. We note that there are also conditions, however, in which very similar or identical replicons can co-exist in the same cell [14]. Some Inc. groups defined in *Pseudomonas* are equivalent to those defined in *Escherichia coli*; for example, IncP-1, IncP-3, IncP-4, and IncP-6 are equivalent to IncP, A/C, IncQ, and IncG/U, respectively [15,16].

2. Key Factors in Plasmid Design

Based on recent progress in plasmid biology and bioinformatics, we consider three factors that should be taken into account to design a synthetic plasmid (Table 2): 1) plasmid gene content; 2) interaction with host (host factors and fitness cost imposed by plasmids); and 3) constraints in genome (size, sequence composition [e.g., G + C content, oligonucleotide composition, and codon usage], and gene direction). These factors are described in detail in the subsequent sections.

2.1. Plasmid Gene Content

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 $<1e-5$) comparison, replication initiation (Rep) proteins for the six Inc. groups (RepB and RepE for IncFI, RepA for IncFII, 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].

2.1.2.1. Replication Module. Plasmids can carry two types of replication origins; one is a vegetative origin (*oriV*), and the other is a transfer origin (*oriT*). In this section, we describe the replication module that uses *oriV*. In ColE1-type plasmids, the replication module consists of *oriV* and genes for two noncoding RNAs (RNA I and RNA II) and Rop protein,

Table 1
Lists of plasmids in different incompatibility groups.

| Incompatibility ^a | Representative plasmid (original host) ^b | Accession number ^c | RIP ^d | MOB ^e | MPF ^e | Host range ^f | Reference |
|--------------------------------|---|--|---|------------------|------------------|--|----------------------------|
| Inc groups | | | | | | | |
| A/C ₁ (=IncA) | RA1 (<i>Aeromonas hydrophila</i>) | NC_012885 | RepA | MOB _H | MPF _F | <i>Gammaproteobacteria</i> | [151] |
| A/C ₂ (=IncC) | pRMH760 (<i>Klebsiella pneumoniae</i>) | KF976462 | RepA | MOB _H | MPF _F | <i>Gammaproteobacteria</i> | [152] |
| IncB/O | R3521 (<i>Escherichia coli</i>) | GU256641 | RepA | MOB _p | MPF _I | <i>Escherichia</i> | [153] |
| IncD | R711b (<i>Providencia</i>) | NA | NA | NA | NA | <i>Escherichia</i> , <i>Salmonella</i> , <i>Proteus</i> | [154] |
| IncFI | F (<i>Escherichia coli</i>) | AP001918 | RepE (for RepFIA replicon) | MOB _F | MPF _F | <i>Enterobacteriaceae</i> , <i>Yersiniaceae</i> | [155] |
| IncFII | R1 (<i>Salmonella enterica</i>) | KY749247 | RepA | MOB _F | MPF _F | <i>Escherichia</i> , <i>Salmonella</i> | [156] |
| IncG/U (=IncP-6) | Rms149 (<i>Pseudomonas aeruginosa</i>) | AJ877225 | RepA | MOB _p | – | <i>Proteobacteria</i> | [157,158] |
| | RA3 (<i>Aeromonas hydrophila</i>) | DQ401103 | | | | | |
| IncH | R27 (<i>Salmonella typhi</i>) | AF250878 | RepH11A, RepH11B | MOB _H | MPF _F | <i>Enterobacteriaceae</i> , <i>Yersiniaceae</i> , <i>Erwiniaceae</i> | [159] |
| Incl | R64 (<i>Salmonella enterica</i>) | AP005147 | RepZ | MOB _p | MPF _I | <i>Escherichia</i> , <i>Salmonella</i> , <i>Shigella</i> | [160] |
| IncJ | R391 (<i>Providencia rettgeri</i>) | AY090559 | – | MOB _H | MPF _F | NA | [161] |
| IncK | R387 (<i>Shigella flexneri</i>) | NCTC50022 | – | MOB _p | MPF _I | NA | [162] |
| IncL | R471 (<i>Serratia marcescens</i>) | KM406489 | RepA | MOB _p | MPF _I | <i>Proteobacteria</i> | [163,164] |
| | pK01-34 (<i>Klebsiella oxytoca</i>) | AB715422 | | | | | |
| IncM | R69 (<i>Salmonella enterica</i>) | KM406488 | RepA | MOB _p | MPF _I | <i>Proteobacteria</i> | [165] |
| | pEL60 (<i>Erwinia amylovora</i>) | NC_005246 | | | | | |
| IncN | N3 (<i>Escherichia coli</i>) | NC_015599 | RepA | MOB _F | MPF _T | <i>Escherichia</i> , <i>Klebsiella</i> , <i>Salmonella</i> | [166] |
| IncP (=IncP-1) | RK2 (<i>Pseudomonas aeruginosa</i>) | BN000925 | TrfA | MOB _p | MPF _T | <i>Proteobacteria</i> | [167] |
| IncQ | RSF1010 (<i>Escherichia coli</i>) | M28829 | RepABC | MOB _Q | – | <i>Proteobacteria</i> | [168] |
| IncR | pKP1780 (<i>Klebsiella pneumoniae</i>) | JX424614 | RepB | – | – | <i>Klebsiella</i> | [169] |
| IncS (=IncH12) | R478 (<i>Serratia marcescens</i>) | BX664015 | RepH1A | MOB _H | MPF _F | <i>Serratia</i> | [170] |
| IncT | Rts1 (<i>Proteus vulgaris</i>) | AP004237 | – | MOB _H | MPF _F | <i>Proteus</i> , <i>Citrobacter</i> | [171] |
| IncV | R753 (<i>Escherichia coli</i>) | NCTC50521 (planned) | NA | NA | NA | <i>Proteus</i> | [172] |
| IncW | R388 (<i>Escherichia coli</i>) | NC_028464 | RepA | MOB _F | MPF _T | <i>Salmonella</i> , <i>Escherichia</i> , <i>Providencia</i> | [173] |
| IncX | R6K (<i>Escherichia coli</i>) | NCTC50005 | π | MOB _p | MPF _T | <i>Enterobacteriaceae</i> | [174] |
| IncY | P1 (<i>Escherichia virus</i>) | AF234172 (phage P1 mod749::IS5 c1–100) | RepA | – | – | <i>Enterobacteriaceae</i> | [175,176] |
| | pMCR-1-P3 (<i>Escherichia coli</i>) | KX880944 | | | | | |
| IncZ | pEI545 (<i>Klebsiella pneumoniae</i>) | M93064 (partial) | RepA | NA | NA | <i>Klebsiella</i> | [162] |
| PromA | pMRAD02 (<i>Methylobacterium radiotolerans</i>) | NC_010509 | RepA | MOB _p | MPF _T | <i>Proteobacteria</i> | [177–179] |
| | pIPO2 ([§] unknown) | AJ297913 | | | | | |
| | pSN0729-62 ([§] unknown) | AP018705 | | | | | |
| IncP-9 | pM3 (<i>Pseudomonas putida</i>) | AF078924 (partial) | Rep | MOB _F | MPF _T | <i>Pseudomonas</i> , <i>Escherichia</i> | [180,181] |
| | NAH7 (<i>Pseudomonas putida</i>) | NC_007926 | | | | | |
| Representative cloning vectors | | | | | | | |
| Not assigned | pUC18/19, pET, pBluescript | L09136 | pMB1 copy-up type, pMB1 type, pMB1 copy-up type | – | – | <i>Escherichia</i> | [175] Merck Millipore[182] |
| Not assigned | pACYC148 | X06403 | p15A type | – | – | <i>Escherichia</i> | [183] |
| Not assigned | pBBR1MCS (<i>Bordetella bronchiseptica</i>) | NC_025015 | pBBR1 Rep | MOB _V | – | <i>Proteobacteria</i> | [184] |
| Not assigned | pSC101 (<i>Salmonella enterica</i>) | NC_002056 | RepA | MOB _Q | – | <i>Salmonella</i> , <i>Escherichia</i> | [185] |
| Not assigned | pABC1 (<i>Rhizobium etli</i>) | KY031728 | p42b RepC | – | – | <i>Rhizobiales</i> | [44] |
| Not assigned | pME6041 (<i>Pseudomonas</i> sp.) | AF118812 | pVS1 RepA | – | – | <i>Proteobacteria</i> | [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) GenBank and RefSeq (<https://www.ncbi.nlm.nih.gov/Seqin/acc.html>) and Wellcome Sanger Institute (prefix "NCTC" <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 MPF 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.

[§] Original hosts are unknown because exogenous plasmid capturing was used.

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 point mutation in RNA II result in a dramatic increase in copy numbers

Table 2
Key factors in the construction of a plasmid vector.

| 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 <i>cis</i> -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 unwound 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 β -clamp, leading to unidirectional replication [31]. Dimeric Rep proteins prevent *oriV* melting by pairing iterons in a phenomenon called handcuffing (Fig. 2C) [32–34]. An increased monomer to dimer ratio dissociates the paired iterons [32,33].

A Rep protein mutant of R6K π 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 *parS*), 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 *par* 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 *parA-parB* operon and its downstream *parS* 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.

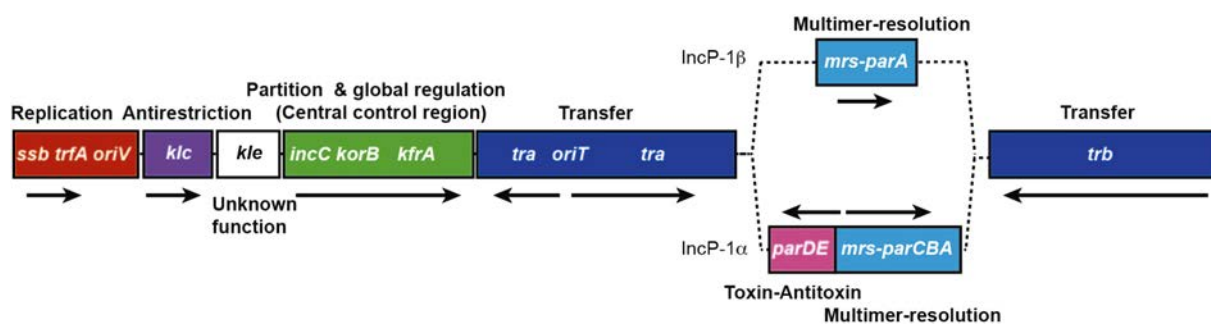


Fig. 1. Modular structure of the plasmid genome. An example of the IncP-1 plasmid genome is shown. The replication module is the only element essential for plasmid maintenance. The other modules assist in plasmid maintenance. IncP-1 β and IncP-1 α are lineages of IncP-1 plasmids. Horizontal arrows indicate 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].

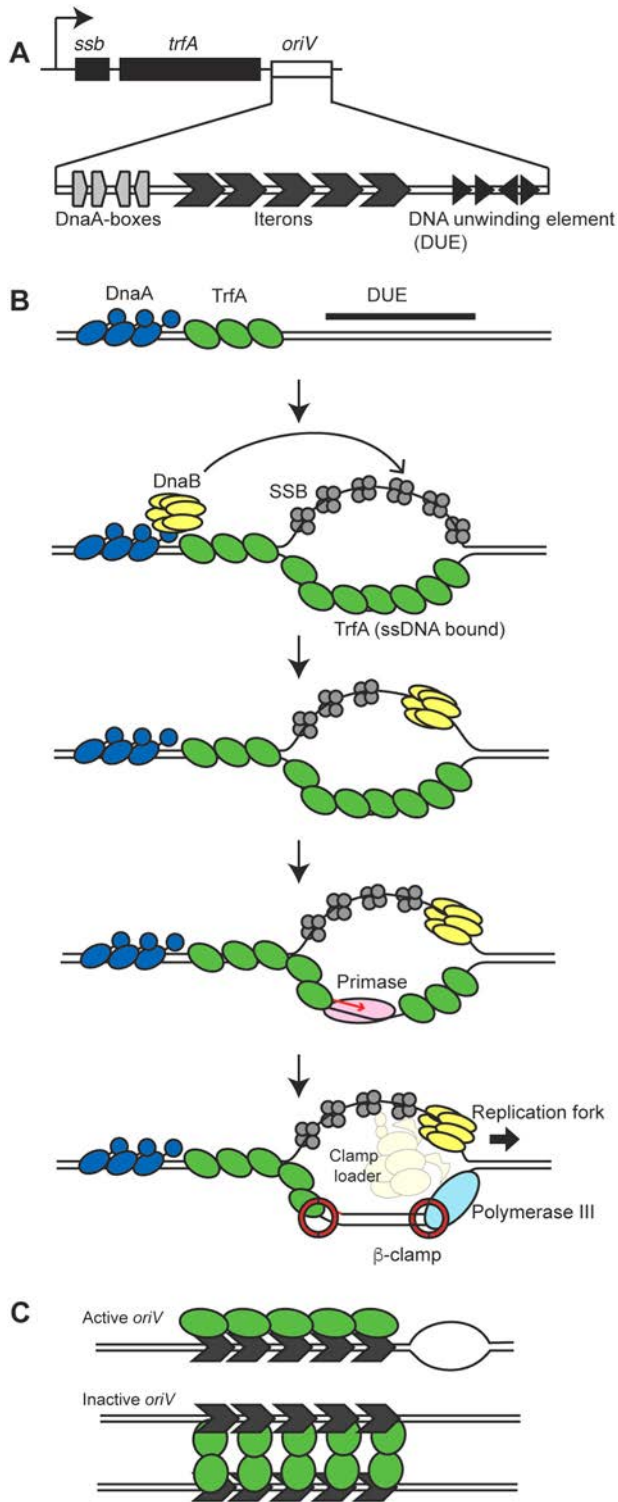


Fig. 2. Replication module. (A) Replication module of IncP-1 plasmids. *Ssb* encodes a single-strand DNA binding protein. *trfA* encodes the Rep protein. (B) Replication initiation of IncP-1 plasmid. Monomeric TrfA bound to iterons opens base pairs in the DUE, and then host DnaA or TrfA itself recruits DnaB onto the unwound DUE. TrfA molecules bind to single-stranded DNA to assist replisome 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].

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

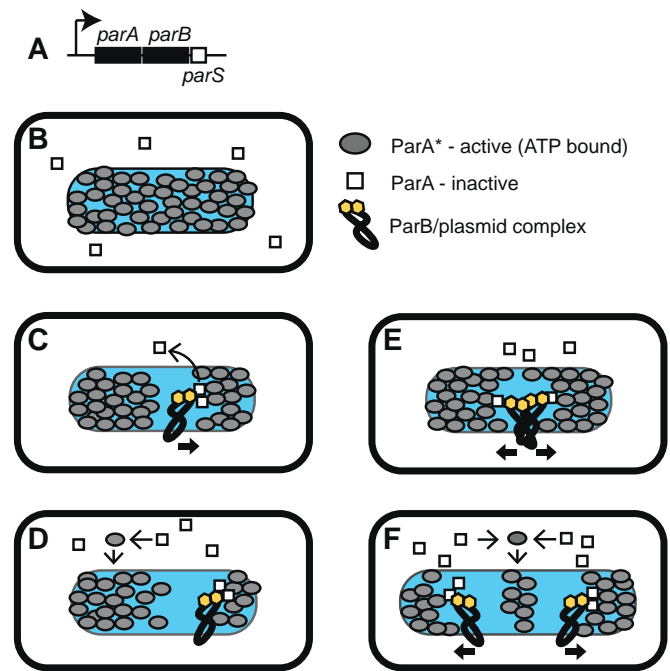


Fig. 3. Partition module. (A) *par* locus of P1 prophage (type I partition system). *parA* encodes Walker-type ATPase. *parB* encodes centromere-binding protein. *parS* is centromere. (B–F) Diffusion-ratchet model of plasmid motion. (B) ParA has two states: ParA*, the ATP-bound form, active in binding 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.

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].

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.

2.1.3. Testing the Functionality of Functional Modules

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 kondratievae* plasmid have been confirmed using these vectors [70].

2.1.4. 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 ColE1 have been used in an antibiotic-free host-vector system [73].

2.2. Interactions with the Host

Early biochemical studies and recent experimental evolution studies have suggested the importance of host factors and fitness cost for plasmid carriage. These factors are discussed below.

2.2.1. Host Factors

Most plasmids require the host's replication initiator *DnaA* and DNA helicase encoded by the host chromosome or plasmid itself upon replication initiation from *oriV* [9]. Whether plasmids can load DNA helicase at the *oriV* using *DnaA* or plasmid's *Rep* protein determines the capability of plasmid replication in the host cells and their replication host range [74,75]. Nucleoid-associated proteins (NAPs), such as histone-like nucleoid-structuring protein (H-NS) are known to make the DNA structure more compact [76]. Moreover, chromosomally encoded NAPs have been shown to affect gene expression from the IncP-7 plasmid pCAR1 [77,78].

2.2.2. Fitness Cost Imposed by Plasmids

When plasmids are introduced to novel hosts, plasmids initially impose a fitness cost on the hosts and are thus not necessarily stably maintained, particularly in laboratory systems [72,79,80]. It should be noted that in nature, plasmids can persist without positive selection, despite their detectable costs in laboratory systems [72]. Resequencing of experimentally evolved plasmid-host pairs in several independent studies suggests that initial interactions between the host gene and plasmid gene are unfavorable for the host's growth. Although the cause of the cost can be different among plasmid-host pairs, reduced interaction appears to improve host growth and plasmid maintenance [81–83]. These observations are consistent with the complexity hypothesis, which states that the number of interaction partners predicts the horizontal transfer ability of a gene [84,85]. Using a series of antibiotic resistance

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

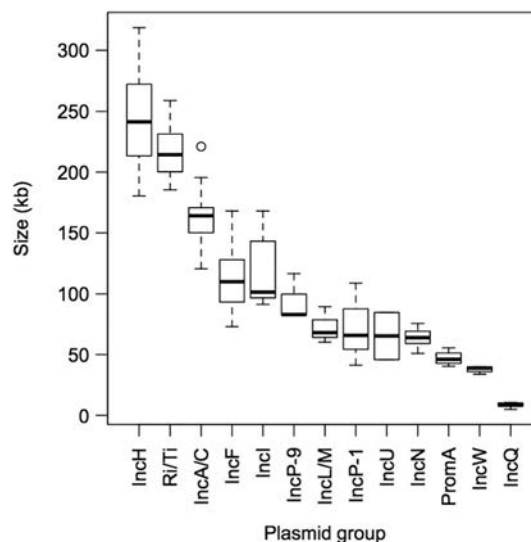


Fig. 4. Box-and-whisker plots displaying the distributions of plasmid sizes (kb) for various plasmid groups based on a five-number summary (minimum, 25th percentile, median, 75th percentile, and maximum). Outliers are plotted as open circles. Data are from Suzuki et al. [17].

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 group 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 IncHI2, 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 [97,99]. 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 genic 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 tetranucleotides (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 suggests 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 tumefaciens* replicons (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

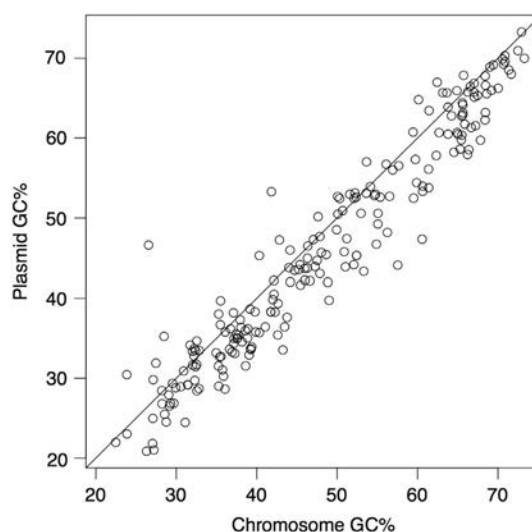


Fig. 5. Plot of G + C contents of 209 plasmids and their host chromosomes. Each point represents a plasmid–chromosome pair from 209 prokaryotes. To minimize the bias in the numbers of sequenced organisms and replicons available in public databases (e.g., thousands of genome projects for *Escherichia coli*, and multireplicons for *Borrelia* species), RefSeq data for completely sequenced prokaryotes that consist of one chromosome and plasmid were retrieved on April 17, 2017 from a list of all selected representative prokaryotic genomes (ftp://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/prok_representative_genomes.txt). The G + C contents of plasmids tend to be lower than (and are correlated with) those of the host chromosomes.

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-throughput DNA 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.insdc.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]. Orlek et al. [135] presented a curated dataset of complete Enterobacteriaceae plasmids compiled from the NCBI database (<https://figshare.com/s/18de8bdcbba47dbaba41>). The web servers PLSDB (<https://ccb-microbe.cs.uni-saarland.de/plsdb/>) [136] and pATLAS (<http://www.patlas.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 from 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. plasmidSPAdes (for monocultured bacterial genomes) and Recycler (for metagenomes) may be complementary methods for plasmid reconstruction [141]. Two studies have compared the performances of the four plasmid analysis tools (cBar, PlasmidFinder, plasmidSPAdes, and Recycler) [148,149].

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

Table 3

List of bioinformatics tools for plasmid sequence analysis and vector design.

| Usage | Name | URL |
|---|------------------------|---|
| <i>Viewing/editing plasmid sequences</i> | | |
| | ApE (A plasmid Editor) | http://biologylabs.utah.edu/jorgensen/wayned/ape/ |
| | SnapGene | http://www.snapgene.com |
| | Benchling | https://benchling.com/molecular-biology |
| <i>Designing vectors by assembling modules</i> | | |
| | SEVA-DB | http://seva.cnb.csic.es |
| <i>Reconstructing plasmids from sequencing reads or assembly graphs</i> | | |
| | PLACNET | https://sourceforge.net/projects/placnet/ |
| | PLACNETw | https://castillo.dicom.unican.es/upload/ |
| | plasmidSPAdes | http://cab.spbu.ru/software/spades/ |
| | Recycler | https://github.com/Shamir-Lab/Recycler |
| | PlasmidTron | https://github.com/sanger-pathogens/plasmidtron |
| <i>Detecting plasmids in assembled contigs</i> | | |
| | PlasmidFinder | https://cge.cbs.dtu.dk/services/PlasmidFinder/ |
| | cBar | http://csbl.bmb.uga.edu/~ffzhou/cBar |
| | PlasFlow | https://github.com/smaegol/PlasFlow |
| | MOB-suite | https://github.com/phac-nml/mob-suite |
| <i>Detecting plasmids in unassembled reads</i> | | |
| | PlasmidSeeker | https://github.com/bioinfo-ut/PlasmidSeeker |

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. and M.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

- Skippington E, Ragan MA. Lateral genetic transfer and the construction of genetic exchange communities. *FEMS Microbiol Rev* 2011;35:707–35. <https://doi.org/10.1111/j.1574-6976.2010.00261.x>.
- Wiedenbeck J, Cohan FM. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* 2011;35:957–76. <https://doi.org/10.1111/j.1574-6976.2011.00292.x>.
- 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>.
- 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>.
- 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>.
- Fernández-López C, Bravo A, Ruiz-Cruz S, Solano-Collado V, Garsin DA, Lorenzo-Díaz F, et al. Mobilizable rolling-circle replicating plasmids from gram-positive bacteria: a low-cost conjugative transfer. *Plasmids Biol Impact Biotechnol Discov* 2014;2:257–76. <https://doi.org/10.1128/microbiolspec.PLAS-0008-2013> American Society of Microbiology.
- Cabezón E, Ripoll-Rozada J, Peña A, de la Cruz F, Arechaga I. Towards an integrated model of bacterial conjugation. *FEMS Microbiol Rev* 2015;39:81–95. <https://doi.org/10.1111/1574-6976.12085>.
- Wegrzyn KE, Gross M, Uciechowska U, Konieczny I. Replisome assembly at bacterial chromosomes and iteron plasmids. *Front Mol Biosci* 2016;3:39. <https://doi.org/10.3389/fmolb.2016.00039>.
- del Solar G, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Díaz-Orejeras R. Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* 1998;62:434–64.
- Lilly J, Camps M. Mechanisms of theta plasmid replication. *Microbiol Spectr* 2015. <https://doi.org/10.1128/microbiolspec.PLAS-0029-2014> 3:PLAS-0029-2014.
- Harms A, Brodersen DE, Mitarai N, Gerdes K. Toxins, targets, and triggers: an overview of toxin-antitoxin biology. *Mol Cell* 2018;70:768–84. <https://doi.org/10.1016/j.molcel.2018.01.003>.
- Bouet J-Y, Nordström K, Lane D. Plasmid partition and incompatibility—the focus shifts. *Mol Microbiol* 2007;65:1405–14. <https://doi.org/10.1111/j.1365-2958.2007.05882.x>.
- Novick RP. Plasmid incompatibility. *Microbiol Rev* 1987;51:381–95.
- Velappan N, Sblattero D, Chasteen L, Pavlik P, Bradbury ARM. Plasmid incompatibility: more compatible than previously thought? *Protein Eng Des Sel* 2007;20:309–13. <https://doi.org/10.1093/protein/gzm005>.
- Thomas CM, Haines AS. Plasmids of the genus *Pseudomonas*. *Pseudomonas*, Boston, MA: Springer US; 2004; 197–231. https://doi.org/10.1007/978-1-4419-9086-0_7.
- Haines AS, Cheung M, Thomas CM. Evidence that IncG (IncP-6) and IncU plasmids form a single incompatibility group. *Plasmid* 2006;55:210–5. <https://doi.org/10.1016/j.plasmid.2005.11.003>.
- Suzuki H, Yano H, Brown CJ, Top EM. Predicting plasmid promiscuity based on genomic signature. *J Bacteriol* 2010;192:6045–55. <https://doi.org/10.1128/JB.00277-10>.
- Galperin MY, Koonin EV. Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol* 2000;18:609–13. <https://doi.org/10.1038/76443>.
- Treangen TJ, Rocha EPC. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* 2011;7:e1001284. <https://doi.org/10.1371/journal.pgen.1001284>.
- Thomas CM. Transcription regulatory circuits in bacterial plasmids. *Biochem Soc Trans* 2006;34:1072–4. <https://doi.org/10.1042/BST0341072>.
- Fernández-López R, de la Cruz F. Rebooting the genome: the role of negative feedback in horizontal gene transfer. *Mob Genet Elem* 2014;4:1–6. <https://doi.org/10.4161/2159256X.2014.988069>.
- Lorenzo-Díaz F, Fernández-López C, Lurz R, Bravo A, Espinosa M. Crosstalk between vertical and horizontal gene transfer: plasmid replication control by a conjugative relaxase. *Nucleic Acids Res* 2017;45:7774–85. <https://doi.org/10.1093/nar/gkx450>.
- Norberg P, Bergström M, Jethava V, Dubhashi D, Hermansson M. The IncP-1 plasmid backbone adapts to different host bacterial species and evolves through homologous recombination. *Nat Commun* 2011;2:268. <https://doi.org/10.1038/ncomms1267>.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402.
- Adamczyk M, Jagura-Burdzy G. Spread and survival of promiscuous IncP-1 plasmids. *Acta Biochim Pol* 2003;50:425–53 doi: 1035002425.
- Shetty RP, Endy D, Knight TF. Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* 2008;2:5. <https://doi.org/10.1186/1754-1611-2-5>.
- Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A, de Las Heras A, Et Al. The standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res* 2013;41:D666–75. <https://doi.org/10.1093/nar/gks1119>.
- Brantl S. Antisense RNAs in plasmids: control of replication and maintenance. *Plasmid* 2002;48:165–73.
- Sambrook JRD. *Molecular Cloning: A Laboratory Manual*. 3rd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001.
- Rakowski SA, Filutowicz M. Plasmid R6K replication control. *Plasmid* 2013;69:231–42. <https://doi.org/10.1016/j.plasmid.2013.02.003>.
- Wawrzycka A, Gross M, Wasaznik A, Konieczny I. Plasmid replication initiator interactions with origin 13-mers and polymerase subunits contribute to strand-specific replisome assembly. *Proc Natl Acad Sci U S A* 2015;112:E4188–96. <https://doi.org/10.1073/pnas.1504926112>.
- Bury K, Wegrzyn K, Konieczny I. Handcuffing reversal is facilitated by proteases and replication initiator monomers. *Nucleic Acids Res* 2017;45:3953–66. <https://doi.org/10.1093/nar/gkx166>.
- Das N, Chattoraj DK. Origin pairing ('handcuffing') and unpairing in the control of P1 plasmid replication. *Mol Microbiol* 2004;54:836–49. <https://doi.org/10.1111/j.1365-2958.2004.04322.x>.
- Gasset-Rosa F, Díaz-López T, Lurz R, Prieto A, Fernández-Tresguerres ME, Giraldo R. Negative regulation of pPS10 plasmid replication: origin pairing by zipping-up DNA-bound RepA monomers. *Mol Microbiol* 2008;68:560–72. <https://doi.org/10.1111/j.1365-2958.2008.06166.x>.
- Chen D, Feng J, Krüger R, Urh M, Inman RB, Filutowicz M. Replication of R6K γ origin *in vitro*: discrete start sites for DNA synthesis dependent on π and its copy-up variants. *J Mol Biol* 1998;282:775–87. <https://doi.org/10.1006/jmbi.1998.2055>.
- Metcalfe WW, Jiang W, Wanner BL. Use of the *rep* technique for allele replacement to construct new *Escherichia coli* hosts for maintenance of R6K gamma origin plasmids at different copy numbers. *Gene* 1994;138:1–7.
- Pogliano J, Ho TQ, Zhong Z, Helinski DR. Multicopy plasmids are clustered and localized in *Escherichia coli*. *Proc Natl Acad Sci* 2001;98:4486–91. <https://doi.org/10.1073/pnas.081075798>.
- Chang PC, Cohen SN. Bidirectional replication from an internal origin in a linear streptomyces plasmid. *Science* 1994;265:952–4.
- Horiuchi T, Hidaka M. Core sequence of two separable terminus sites of the R6K plasmid that exhibit polar inhibition of replication is a 20 bp inverted repeat. *Cell* 1988;54:515–23.
- Salje J, Gayathri P, Löwe J. The ParMRC system: molecular mechanisms of plasmid segregation by actin-like filaments. *Nat Rev Microbiol* 2010;8:683–92. <https://doi.org/10.1038/nrmicro2425>.
- Bouet J-Y, Funnell BE. P1 ParA interacts with the P1 partition complex at *parS* and an ATP-ADP switch controls ParA activities. *EMBO J* 1999;18:1415–24. <https://doi.org/10.1093/emboj/18.5.1415>.
- Williams DR, Macartney DP, Thomas CM. The partitioning activity of the RK2 central control region requires only *incC*, *korB* and *KorB*-binding site O_B3 but other *KorB*-binding sites form destabilizing complexes in the absence of O_B3 . *Microbiology* 1998;144:3369–78. <https://doi.org/10.1099/00221287-144-12-3369>.
- Cevallos MA, Cervantes-Rivera R, Gutiérrez-Ríos RM. The repABC plasmid family. *Plasmid* 2008;60:19–37. <https://doi.org/10.1016/j.plasmid.2008.03.001>.
- Döhlemann J, Wagner M, Happel C, Carrillo M, Sobetzko P, Erb TJ, et al. A family of single copy *repABC*-type shuttle vectors stably maintained in the alpha-proteobacterium *Sinorhizobium meliloti*. *ACS Synth Biol* 2017;6:968–84. <https://doi.org/10.1021/acssynbio.6b00320>.
- Gerdes K, Rasmussen PB, Molin S. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proc Natl Acad Sci U S A* 1986;83:3116–20.
- Mruk I, Kobayashi I. To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res* 2014;42:70–86. <https://doi.org/10.1093/nar/gkt711>.

- [47] Hall AM, Gollan B, Helaine S. Toxin–antitoxin systems: reversible toxicity. *Curr Opin Microbiol* 2017;36:102–10. <https://doi.org/10.1016/j.mib.2017.02.003>.
- [48] Weggerer A, Sun T, Altenbuchner J. Optimization of an *E coli* L-rhamnose-inducible expression vector: test of various genetic module combinations. *BMC Biotechnol* 2008;8:2. <https://doi.org/10.1186/1472-6750-8-2>.
- [49] Bernard P, Couturier M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol* 1992;226:735–45.
- [50] Stieber D, Gabant P, Szpirer C. The art of selective killing: plasmid toxin/antitoxin systems and their technological applications. *Biotechniques* 2008;45:344–6. <https://doi.org/10.2144/000112955>.
- [51] Le Roux F, Binesse J, Saulnier D, Mazel D. Construction of a *Vibrio splendidus* mutant lacking the metalloprotease gene *vsm* by use of a novel counterselectable suicide vector. *Appl Environ Microbiol* 2007;73:777–84. <https://doi.org/10.1128/AEM.02147-06>.
- [52] Val M-E, Skovgaard O, Ducos-Galand M, Bland MJ, Mazel D. Genome engineering in *Vibrio cholerae*: a feasible approach to address biological issues. *PLoS Genet* 2012;8:e1002472. <https://doi.org/10.1371/journal.pgen.1002472>.
- [53] Paul S, Summers D, ArgR and PcpA, accessory proteins for XerCD-mediated resolution of ColE1 dimers, are also required for stable maintenance of the P1 prophage. *Plasmid* 2004;52:63–8. <https://doi.org/10.1016/j.plasmid.2004.04.003>.
- [54] Guhathakurta A, Viney I, Summers D. Accessory proteins impose site selectivity during ColE1 dimer resolution. *Mol Microbiol* 1996;20:613–20.
- [55] Cornet F, Mortier I, Patte J, Louarn JM. Plasmid pSC101 harbors a recombination site, *psi*, which is able to resolve plasmid multimers and to substitute for the analogous chromosomal *Escherichia coli* site *dif*. *J Bacteriol* 1994;176:3188–95.
- [56] Hiraga S, Sugiyama T, Itoh T. Comparative analysis of the replicon regions of eleven ColE2-related plasmids. *J Bacteriol* 1994;176:7233–43.
- [57] Eberl L, Kristensen CS, Givskov M, Grohmann E, Gerlitz M, Schwab H. Analysis of the multimer resolution system encoded by the *parCBA* operon of broad-host-range plasmid RP4. *Mol Microbiol* 1994;12:131–41.
- [58] Loftie-Eaton W, Yano H, Burleigh S, Simmons RS, Hughes JM, Rogers LM, et al. Evolutionary paths that expand plasmid host-range: implications for spread of antibiotic resistance. *Mol Biol Evol* 2016;33:885–97. <https://doi.org/10.1093/molbev/msv339>.
- [59] Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Front Microbiol* 2015;6:242. <https://doi.org/10.3389/fmicb.2015.00242>.
- [60] Christie PJ, Gordon JE. The *Agrobacterium* Ti plasmids. *Plasmids Biol Impact Biotechnol Discov* 2014;2:295–313. <https://doi.org/10.1128/microbiolspec.PLAS-0010-2013> American Society of Microbiology.
- [61] Ummels R, Abdallah AM, Kuiper V, Aajoud A, Sparrus M, Naem R, et al. Identification of a novel conjugative plasmid in mycobacteria that requires both type IV and type VII secretion. *MBio* 2014;5:e01744–14. doi:<https://doi.org/10.1128/mBio.01744-14>.
- [62] Thoma L, Muth G. Conjugative DNA-transfer in *Streptomyces*, a mycelial organism. *Plasmid* 2016;87–88:1–9. <https://doi.org/10.1016/j.plasmid.2016.09.004>.
- [63] Meyer R. Replication and conjugative mobilization of broad host-range IncQ plasmids. *Plasmid* 2009;62:57–70. <https://doi.org/10.1016/j.plasmid.2009.05.001>.
- [64] Pappas CJ, Benaroudj N, Picardeau M. A replicative plasmid vector allows efficient complementation of pathogenic *Leptospira* strains. *Appl Environ Microbiol* 2015;81:3176–81. <https://doi.org/10.1128/AEM.00173-15>.
- [65] Heeb S, Itoh Y, Nishijyo T, Schneider U, Keel C, Wade J, et al. Small, stable shuttle vectors based on the minimal pVSI replicon for use in gram-negative, plant-associated bacteria. *Mol Plant Microbe Interact* 2000;13:232–7. <https://doi.org/10.1094/MPMI.2000.13.2.232>.
- [66] Obranić S, Babić F, Maravić-Vlahoviček G. Improvement of pBRR1MCS plasmids, a very useful series of broad-host-range cloning vectors. *Plasmid* 2013;70:263–7. <https://doi.org/10.1016/j.plasmid.2013.04.001>.
- [67] Serfiotis-Mitsa D, Roberts GA, Cooper LP, White JH, Nutley M, Cooper A, et al. The *orf18* gene product from conjugative transposon Tn916 is an ArdA antirestriction protein that inhibits type I DNA restriction-modification systems. *J Mol Biol* 2008;383:970–81. <https://doi.org/10.1016/j.jmb.2008.06.005>.
- [68] Belogurov AA, Delver EP, Agafonova OV, Belogurova NG, Lee L-Y, Kado CI. Antirestriction protein Ard (Type C) encoded by IncW plasmid pSa has a high similarity to the “protein transport” domain of TraC1 primase of promiscuous plasmid RP4. *J Mol Biol* 2000;296:969–77. <https://doi.org/10.1006/jmbi.1999.3493>.
- [69] Serfiotis-Mitsa D, Herbert AP, Roberts GA, Soares DC, White JH, Blakely GW, et al. The structure of the KlcA and ArdB proteins reveals a novel fold and antirestriction activity against Type I DNA restriction systems *in vivo* but not *in vitro*. *Nucleic Acids Res* 2010;38:1723–37. <https://doi.org/10.1093/nar/gkp1144>.
- [70] Bartosik AA, Glabinski K, Kulinska A, Lewicka E, Godziszewska J, Markowska A, et al. Convenient broad-host-range unstable vectors for studying stabilization cassettes in diverse bacteria. *BMC Microbiol* 2016;16:59. <https://doi.org/10.1186/s12866-016-0674-y>.
- [71] Mignon C, Sodoyer R, Werle B. Antibiotic-free selection in biotherapeutics: now and forever. *Pathog (Basel, Switzerland)* 2015;4:157–81. <https://doi.org/10.3390/pathogens4020157>.
- [72] Vandermeulen G, Marie C, Scherman D, Pr at V. New generation of plasmid backbones devoid of antibiotic resistance marker for gene therapy trials. *Mol Ther* 2011;19:1942–9. <https://doi.org/10.1038/mt.2011.182>.
- [73] Mairhofer J, Pfaffenzeller I, Merz D, Grabherr R. A novel antibiotic free plasmid selection system: advances in safe and efficient DNA therapy. *Biotechnol J* 2008;3:83–9. <https://doi.org/10.1002/biot.200700141>.
- [74] Caspi R, Helinski DR, Pacek M, Konieczny I. Interactions of DnaA proteins from distantly related bacteria with the replication origin of the broad host range plasmid RK2. *J Biol Chem* 2000;275:18454–61. <https://doi.org/10.1074/jbc.M000552200>.
- [75] Jiang Y, Pacek M, Helinski DR, Konieczny I, Toukdarian A. A multifunctional plasmid-encoded replication initiation protein both recruits and positions an active helicase at the replication origin. *Proc Natl Acad Sci U S A* 2003;100:8692–7. <https://doi.org/10.1073/pnas.1532393100>.
- [76] Dillon SC, Dorman CJ. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* 2010;8:185–95. <https://doi.org/10.1038/nrmicro2261>.
- [77] Yun C-S, Takahashi Y, Shintani M, Takeda T, Suzuki-Minakuchi C, Okada K, et al. MvaT family proteins encoded on IncP-7 plasmid pCAR1 and the host chromosome regulate the host transcriptome cooperatively but differently. *Appl Environ Microbiol* 2015;82:832–42. <https://doi.org/10.1128/AEM.03071-15>.
- [78] Yun C-S, Suzuki C, Naito K, Takeda T, Takahashi Y, Sai F, et al. Pmr, a histone-like protein H1 (H-NS) family protein encoded by the IncP-7 plasmid pCAR1, is a key global regulator that alters host function. *J Bacteriol* 2010;192:4720–31. <https://doi.org/10.1128/JB.00591-10>.
- [79] Sota M, Yano H, M Hughes J, Daughdrill Gw, Abdo Z, Forney Lj, et al.. Shifts in the host range of a promiscuous plasmid through parallel evolution of its replication initiation protein. *ISME J* 2010;4:1568–80. <https://doi.org/10.1038/ismej.2010.72>.
- [80] De Gelder L, Williams JJ, Ponciano JM, Sota M, Top EM. Adaptive plasmid evolution results in host-range expansion of a broad-host-range plasmid. *Genetics* 2008;178:2179–90. <https://doi.org/10.1534/genetics.107.084475>.
- [81] Millan AS, Toll-Riera M, Qi Q, Maclean RC. Interactions between horizontally acquired genes create a fitness cost in *Pseudomonas aeruginosa*. *Nat Commun* 2015;6:6845. <https://doi.org/10.1038/ncomms7845>.
- [82] Harrison E, Guymier D, Spiers AJ, Paterson S, Brockhurst MA. Parallel compensatory evolution stabilizes plasmids across the parasitism–mutualism continuum. *Curr Biol* 2015;25:2034–9. <https://doi.org/10.1016/j.cub.2015.06.024>.
- [83] Yano H, Wegryzn K, Loftie-Eaton W, Johnson J, Deckert GE, Rogers LM, et al. Evolved plasmid–host interactions reduce plasmid interference cost. *Mol Microbiol* 2016;101:743–56. <https://doi.org/10.1111/mmi.13407>.
- [84] Cohen O, Gophna U, Pupko T. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol Biol Evol* 2011;28:1481–9. <https://doi.org/10.1093/molbev/msq333>.
- [85] Jain R, Rivera MC, Lake JA. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A* 1999;96:3801–6.
- [86] Porse A, Schou TS, Munck C, Ellabaan MMH, Sommer MOA. Biochemical mechanisms determine the functional compatibility of heterologous genes. *Nat Commun* 2018;9:522. <https://doi.org/10.1038/s41467-018-02944-3>.
- [87] Baltrus DA. Exploring the costs of horizontal gene transfer. *Trends Ecol Evol* 2013;28:489–95. <https://doi.org/10.1016/j.tree.2013.04.002>.
- [88] Doyle M, Fookes M, Ivens A, Mangan MW, Wain J, Dorman CJ. An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science* 2007;315:251–2. <https://doi.org/10.1126/science.1137550>.
- [89] Lang KS, Johnson TJ. Characterization of Acr2, an H-NS-like protein encoded on A/C2-type plasmids. *Plasmid* 2016;87–88:17–27. <https://doi.org/10.1016/j.plasmid.2016.07.004>.
- [90] Shintani M, Suzuki-Minakuchi C, Nojiri H. Nucleoid-associated proteins encoded on plasmids: occurrence and mode of function. *Plasmid* 2015;80:32–44. <https://doi.org/10.1016/j.plasmid.2015.04.008>.
- [91] Suzuki-Minakuchi C, Hirotani R, Shintani M, Takeda T, Takahashi Y, Matsui K, et al. Effects of three different nucleoid-associated proteins encoded on IncP-7 plasmid pCAR1 on host *Pseudomonas putida* KT2440. *Appl Environ Microbiol* 2015;81:2869–80. <https://doi.org/10.1128/AEM.00023-15>.
- [92] van Passel MWJ, Bart A, Luyf ACM, van Kampen AHC, van der Ende A. Compositional discordance between prokaryotic plasmids and host chromosomes. *BMC Genomics* 2006;7:26. <https://doi.org/10.1186/1471-2164-7-26>.
- [93] Suzuki H, Sota M, Brown CJ, Top EM. Using Mahalanobis distance to compare genomic signatures between bacterial plasmids and chromosomes. *Nucleic Acids Res* 2008;36:e147. <https://doi.org/10.1093/nar/gkn753>.
- [94] Zhong C, Peng D, Ye W, Chai L, Qi J, Yu Z, et al. Determination of plasmid copy number reveals the total plasmid DNA amount is greater than the chromosomal DNA amount in *Bacillus thuringiensis* YBT-1520. *PLoS One* 2011;6:e16025. <https://doi.org/10.1371/journal.pone.0016025>.
- [95] Yano H, Deckert GE, Rogers LM, Top EM. Roles of long and short replication initiation proteins in the fate of IncP-1 plasmids. *J Bacteriol* 2012;194:1533–43. <https://doi.org/10.1128/JB.06395-11>.
- [96] Rawlings DE, Tietze E. Comparative biology of IncQ and IncQ-like plasmids. *Microbiol Mol Biol Rev* 2001;65:481–96. <https://doi.org/10.1128/MMBR.65.4.481-496.2001>.
- [97] San Millan A, Santos-Lopez A, Ortega-Huedo R, Bernabe-Balas C, Kennedy SP, Gonzalez-Zorn B. Small-plasmid-mediated antibiotic resistance is enhanced by increases in plasmid copy number and bacterial fitness. *Antimicrob Agents Chemother* 2015;59:3335–41. <https://doi.org/10.1128/AAC.00235-15>.
- [98] Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, et al. Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. *Sci Transl Med* 2014;6. <https://doi.org/10.1126/scitranslmed.3009845> 254ra126.
- [99] Plotka M, Wozniak M, Kaczorowski T. Quantification of plasmid copy number with single colour droplet digital PCR. *PLoS One* 2017;12:e0169846. <https://doi.org/10.1371/journal.pone.0169846>.
- [100] Raghavan R, Kelkar YD, Ochman H. A selective force favoring increased G+C content in bacterial genomes. *Proc Natl Acad Sci U S A* 2012;109:14504–7. <https://doi.org/10.1073/pnas.1205683109>.

- [101] Muto A, Osawa S. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci U S A* 1987;84:166–9.
- [102] Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res* 2005;33:1141–53. <https://doi.org/10.1093/nar/gki242>.
- [103] Kelkar YD, Phillips DS, Ochman H. Effects of Genic Base Composition on Growth Rate in G+C-rich Genomes. *G3 (Bethesda)* 2015;5:1247–52. doi:<https://doi.org/10.1534/g3.115.016824>.
- [104] McCutcheon JP, McDonald BR, Moran NA. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet* 2009;5:e1000565. <https://doi.org/10.1371/journal.pgen.1000565>.
- [105] Rocha EPC, Danchin A. Base composition bias might result from competition for metabolic resources. *Trends Genet* 2002;18:291–4. [https://doi.org/10.1016/S0168-9525\(02\)02690-2](https://doi.org/10.1016/S0168-9525(02)02690-2).
- [106] Nishida H. Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids. *Int J Evol Biol* 2012;2012:342482. <https://doi.org/10.1155/2012/342482>.
- [107] Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, Libby SJ, et al. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* 2006;313:236–8. <https://doi.org/10.1126/science.1128794> (80-).
- [108] Karlin S. Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. *Trends Microbiol* 2001;9:335–43. [https://doi.org/10.1016/S0966-842X\(01\)02079-0](https://doi.org/10.1016/S0966-842X(01)02079-0).
- [109] Campbell A, Mrázek J, Karlin S. Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. *Proc Natl Acad Sci U S A* 1999;96:9184–9.
- [110] Lawrence JG, Ochman H. Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* 1997;44:383–97.
- [111] Wilkins BM, Chilley PM, Thomas AT, Pocklington MJ. Distribution of restriction enzyme recognition sequences on broad host range plasmid RP4: molecular and evolutionary implications. *J Mol Biol* 1996;258:447–56. <https://doi.org/10.1006/jmbi.1996.0261>.
- [112] Thorsted PB, Macartney DP, Akhtar P, Haines AS, Ali N, Davidson P, et al. Complete sequence of the IncP β plasmid R751: implications for evolution and organisation of the IncP backbone. *J Mol Biol* 1998;282:969–90. <https://doi.org/10.1006/jmbi.1998.2060>.
- [113] Das MK, Dai H-K. A survey of DNA motif finding algorithms. *BMC Bioinformatics* 2007;8:S21. <https://doi.org/10.1186/1471-2105-8-S7-S21>.
- [114] Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 2011;12:32–42. <https://doi.org/10.1038/nrg2899>.
- [115] Novoa EM, Ribas De Pouplana L. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet* 2012;28:574–81. <https://doi.org/10.1016/j.tig.2012.07.006>.
- [116] Quax TEF, Claessens NJ, Söll D, van der Oost J. Codon bias as a means to fine-tune gene expression. *Mol Cell* 2015;59:149–61. <https://doi.org/10.1016/j.molcel.2015.05.035>.
- [117] Hanson G, Collier J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol* 2017. <https://doi.org/10.1038/nrm2017.91>.
- [118] Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 1985;2:13–34.
- [119] Kanaya S, Yamada Y, Kudo Y, Ikemura T. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* 1999;238:143–55.
- [120] Ermolaeva MD. Synonymous codon usage in bacteria. *Curr Issues Mol Biol* 2001;3:91–7.
- [121] Peixoto L, Zavala A, Romero H, Musto H. The strength of translational selection for codon usage varies in the three replicons of *Sinorhizobium meliloti*. *Gene* 2003;320:109–16.
- [122] Davis JJ, Olsen GJ. Modal codon usage: Assessing the typical codon usage of a genome. *Mol Biol Evol* 2010;27:800–10. <https://doi.org/10.1093/molbev/msp281>.
- [123] Karberg KA, Olsen GJ, Davis JJ. Similarity of genes horizontally acquired by *Escherichia coli* and *Salmonella enterica* is evidence of a supraspecies pangenome. *Proc Natl Acad Sci* 2011;108:20154–9. <https://doi.org/10.1073/pnas.1109451108>.
- [124] Sernova NV, Gelfand MS. Identification of replication origins in prokaryotic genomes. *Brief Bioinform* 2008;9:376–91. <https://doi.org/10.1093/bib/bbn031>.
- [125] Guo F-B, Yuan J-B. Codon usages of genes on chromosome, and surprisingly, genes in plasmid are primarily affected by strand-specific mutational biases in *Lawsonia intracellularis*. *DNA Res* 2009;16:91–104. <https://doi.org/10.1093/dnares/dsp001>.
- [126] Arakawa K, Tomita M. Measures of compositional strand bias related to replication machinery and its applications. *Curr Genomics* 2012;13:4–15. <https://doi.org/10.2174/138920212799034749>.
- [127] Arakawa K, Suzuki H, Tomita M. Quantitative analysis of replication-related mutation and selection pressures in bacterial chromosomes and plasmids using generalised GC skew index. *BMC Genomics* 2009;10:640. <https://doi.org/10.1186/1471-2164-10-640>.
- [128] Rocha EPC, Danchin A. Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat Genet* 2003;34:377–8. <https://doi.org/10.1038/ng1209>.
- [129] French S. Consequences of replication fork movement through transcription units in vivo. *Science* 1992;258:1362–5.
- [130] Srivatsan A, Tehrani DM, MacAlpine DM, Wang JD. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet* 2010;6:e1000810. <https://doi.org/10.1371/journal.pgen.1000810>.
- [131] Zheng W-X, Luo C-S, Deng Y-Y, Guo F-B. Essentiality drives the orientation bias of bacterial genes in a continuous manner. *Sci Rep* 2015;5:16431. <https://doi.org/10.1038/srep16431>.
- [132] Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility of plasmids. *Microbiol Mol Biol Rev* 2010;74:434–52. <https://doi.org/10.1128/MMBR.00020-10>.
- [133] Leplae R, Hebrant A, Wodak SJ, Toussaint A. ACLAME: a CLAssification of Mobile genetic elements. *Nucleic Acids Res* 2004;32:D45–9. <https://doi.org/10.1093/nar/gkh084>.
- [134] Leplae R, Lima-Mendez G, Toussaint A. A first global analysis of plasmid encoded proteins in the ACLAME database. *FEMS Microbiol Rev* 2006;30:980–94. <https://doi.org/10.1111/j.1574-6976.2006.00044.x>.
- [135] Orlek A, Phan H, Sheppard AE, Doumith M, Ellington M, Peto T, et al. A curated dataset of complete *Enterobacteriaceae* plasmids compiled from the NCBI nucleotide database. *Data Br* 2017;12:423–6. <https://doi.org/10.1016/j.dib.2017.04.024>.
- [136] Galata V, Fehlmann T, Backes C, Keller A. PLSDb: a resource of complete bacterial plasmids. *Nucleic Acids Res* 2018. <https://doi.org/10.1093/nar/gky1050>.
- [137] Jesus TF, Ribeiro-Gonçalves B, Silva DN, Bortolola V, Ramirez M, Carrizo JA. Plasmid ATLAS: plasmid visual analytics and identification in high-throughput sequencing data. *Nucleic Acids Res* 2018. <https://doi.org/10.1093/nar/gky1073>.
- [138] Zhou F, Xu Y. cBar: a computer program to distinguish plasmid-derived from chromosome-derived sequence fragments in metagenomics data. *Bioinformatics* 2010;26:2051–2. <https://doi.org/10.1093/bioinformatics/btq299>.
- [139] Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;58:3895–903. <https://doi.org/10.1128/AAC.02412-14>.
- [140] Lanza VF, de Toro M, Garcillán-Barcia MP, Mora A, Blanco J, Coque TM, et al. Plasmid flux in *Escherichia coli* ST131 sublineages, analyzed by plasmid constellation network (PLACNET), a new method for plasmid reconstruction from whole genome sequences. *PLoS Genet* 2014;10:e1004766. <https://doi.org/10.1371/journal.pgen.1004766>.
- [141] Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 2016. <https://doi.org/10.1093/bioinformatics/btw493>.
- [142] Rozov R, Brown Kav A, Bogumil D, Shterzer N, Halperin E, Mizrahi I, et al. Recycler: an algorithm for detecting plasmids from *de novo* assembly graphs. *Bioinformatics* 2016;33. <https://doi.org/10.1093/bioinformatics/btw651>.
- [143] Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res* 2018;46:e35. <https://doi.org/10.1093/nar/gkx1321>.
- [144] Page AJ, Wailan A, Shao Y, Judge K, Dougan G, Klemm EJ, et al. PlasmidTron: assembling the cause of phenotypes and genotypes from NGS data. *Microb Genomics* 2018;4. <https://doi.org/10.1099/mgen.0.000164>.
- [145] Roosaare M, Puustusmaa M, Möls M, Vaheer M, Remm M. PlasmidSeeker: identification of known plasmids from bacterial whole genome sequencing reads. *Peer J* 2018;6:e4588. <https://doi.org/10.7717/peerj.4588>.
- [146] Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genomics* 2018;4. <https://doi.org/10.1099/mgen.0.000206>.
- [147] Vielva L, de Toro M, Lanza VF, de la Cruz F. PLACNETw: a web-based tool for plasmid reconstruction from bacterial genomes. *Bioinformatics* 2017;33:3796–8. <https://doi.org/10.1093/bioinformatics/btx462>.
- [148] Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb Genomics* 2017;3:e000128. <https://doi.org/10.1099/mgen.0.000128>.
- [149] Laczny CC, Galata V, Plum A, Posch AE, Keller A. Assessing the heterogeneity of *in silico* plasmid predictions based on whole-genome-sequenced clinical isolates. *Brief Bioinform* 2017. <https://doi.org/10.1093/bib/bbx162>.
- [150] Heap JT, Pennington OJ, Cartman ST, Minton NP. A modular system for *Clostridium* shuttle plasmids. *J Microbiol Methods* 2009;78:79–85. <https://doi.org/10.1016/j.mimet.2009.05.004>.
- [151] Fricke WF, Welch TJ, McDermott PF, Mammel MK, Leclerc JE, White DG, et al. Comparative genomics of the IncA/C multidrug resistance plasmid family. *J Bacteriol* 2009;191:4750–7. <https://doi.org/10.1128/JB.00189-09>.
- [152] Harmer CJ, Hall RM. pRMH760, a precursor of A/C₂ plasmids carrying *bla*_{CMY} and *bla*_{NDM} genes. *Microb Drug Resist* 2014;20:416–23. <https://doi.org/10.1089/mdr.2014.0012>.
- [153] Papagiannitsis CC, Tzouveleki LS, Kotsakis SD, Tzelepi E, Miriagou V. Sequence of pR3521, an IncB plasmid from *Escherichia coli* encoding ACC-4, SCO-1, and TEM-1 β -lactamases. *Antimicrob Agents Chemother* 2011;55:376–81. <https://doi.org/10.1128/AAC.00875-10>.
- [154] Bradley DE. Characterization of pili determined by drug resistance plasmids R711b and R778b. *J Gen Microbiol* 1977;102:349–63. <https://doi.org/10.1099/00221287-102-2-349>.
- [155] Manwaring NP, Skurray RA, Firth N. Nucleotide sequence of the F plasmid leading region. *Plasmid* 1999;41:219–25. <https://doi.org/10.1006/plas.1999.1390>.
- [156] Cox KEL, Schildbach JF. Sequence of the R1 plasmid and comparison to F and R100. *Plasmid* 2017;91:53–60. <https://doi.org/10.1016/j.plasmid.2017.03.007>.
- [157] Haines AS, Jones K, Cheung M, Thomas CM. The IncP-6 plasmid Rms149 consists of a small mobilizable backbone with multiple large insertions. *J Bacteriol* 2005;187:4728–38. <https://doi.org/10.1128/JB.187.14.4728-4738.2005>.
- [158] Kulinska A, Czeredy M, Hayes F, Jagura-Burdzy G. Genomic and functional characterization of the modular broad-host-range RA3 plasmid, the archetype of the IncU Group. *Appl Environ Microbiol* 2008;74:4119–32. <https://doi.org/10.1128/AEM.00229-08>.
- [159] Sherburne CK, Lawley TD, Gilmour MW, Blattner FR, Burland V, Grotbeck E, et al. The complete DNA sequence and analysis of R27, a large IncHI plasmid from

- Salmonella typhi* that is temperature sensitive for transfer. *Nucleic Acids Res* 2000; 28:2177–86.
- [160] Sampei G, Furuya N, Tachibana K, Saitou Y, Suzuki T, Mizobuchi K, et al. Complete genome sequence of the incompatibility group I1 plasmid R64. *Plasmid* 2010;64: 92–103. <https://doi.org/10.1016/j.plasmid.2010.05.005>.
- [161] Carraro N, Poulin D, Burrus 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] Praszkiar J, Wei T, Siemering K, Pittard J. Comparative analysis of the replication regions of IncB, IncK, and IncZ plasmids. *J Bacteriol* 1991;173:2393–7.
- [163] Borrell L, Yang J, Pittard AJ, Praszkiar J. Interaction of initiator proteins with the origin of replication of an IncL/M 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 pK01-34, an IncL/M plasmid carrying *bla_{IMP-3a}* in *Klebsiella oxytoca* isolated in Japan. *Antimicrob Agents Chemother* 2016;60:3156–62. <https://doi.org/10.1128/AAC.02507-15>.
- [165] Carattoli A, Seiffert SN, Schwendener S, Perreten V, Endimiani A. Differentiation of IncL 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. Fitness of *Escherichia coli* strains carrying expressed and partially silent IncN and IncP1 plasmids. *BMC Microbiol* 2012;12:53. <https://doi.org/10.1186/1471-2180-12-53>.
- [167] Pansegrau W, Lanka E, Barth PT, Figurski DH, Guiney DG, Haas D, et al. Complete nucleotide sequence of Birmingham IncPα plasmids. *J Mol Biol* 1994;239:623–63. <https://doi.org/10.1006/jmbi.1994.1404>.
- [168] Scholz P, Haring V, Wittmann-Liebold B, Ashman K, Bagdasarian M, Scherzinger E. Complete nucleotide sequence and gene organization of the broad-host-range plasmid RSF1010. *Gene* 1989;75:271–88.
- [169] Papagiannitsis CC, Miriagou V, Giakkoupi P, Tzouvelelis LS, Vatopoulos AC. Characterization of pKP1780, a novel IncR plasmid from the emerging *Klebsiella pneumoniae* ST147, encoding the VIM-1 metallo-β-lactamase. *J Antimicrob Chemother* 2013;68:2259–62. <https://doi.org/10.1093/jac/dkt196>.
- [170] Gilmour MW, Thomson NR, Sanders M, Parkhill J, Taylor DE. The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. *Plasmid* 2004;52:182–202. <https://doi.org/10.1016/j.plasmid.2004.06.006>.
- [171] Murata T, Ohnishi M, Ara T, Kaneko J, Han C-G, Li YF, et al. Complete nucleotide sequence of plasmid Rts1: implications for evolution of large plasmid genomes. *J Bacteriol* 2002;184:3194–202.
- [172] Hedges RW. R Factors from *Proteus mirabilis* and *P. vulgaris*. *J Gen Microbiol* 1975; 87:301–11. <https://doi.org/10.1099/00221287-87-2-301>.
- [173] Revilla C, Garcillan-Barcia MP, Fernandez-Lopez R, Thomson NR, Sanders M, Cheung M, et al. Different pathways to acquiring resistance genes illustrated by the recent evolution of IncW plasmids. *Antimicrob Agents Chemother* 2008;52: 1472–80. <https://doi.org/10.1128/AAC.00982-07>.
- [174] Kelley WL, Bastia D. Conformational changes induced by integration host factor at origin gamma of R6K and copy number control. *J Biol Chem* 1991;266:15924–37.
- [175] 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.
- [176] Zhang C, Feng Y, Liu F, Jiang H, Qu Z, Lei M, et al. A phage-like IncY plasmid carrying the *mcr-1* gene in *Escherichia coli* from a pig farm in China. *Antimicrob Agents Chemother* 2017;61:e02035–16. doi:<https://doi.org/10.1128/AAC.02035-16>.
- [177] Ito H, Iizuka H. Taxonomic studies on a radio-resistant *Pseudomonas*. *Agric Biol Chem* 1971;35:1566–71. <https://doi.org/10.1080/00021369.1971.10860119>.
- [178] Schröder G, Bailey MJ, Pembroke T, Smalla K, Lanka E, Pühler A, et al. The complete nucleotide sequence and environmental distribution of the cryptic, conjugative, broad-host-range plasmid pIPO2 isolated from bacteria of the wheat rhizosphere. *Microbiology* 2002;148:1637–53. <https://doi.org/10.1099/00221287-148-6-1637>.
- [179] Yanagiya K, Maejima Y, Nakata H, Tokuda M, Moriuchi 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>.
- [180] Sevastyanovich YR, Titok MA, Krasowiak R, Bingle LEH, Thomas CM. Ability of IncP-9 plasmid pM3 to replicate in *Escherichia coli* is dependent on both *rep* and *par* functions. *Mol Microbiol* 2005;57:819–33. <https://doi.org/10.1111/j.1365-2958.2005.04732.x>.
- [181] Kishida K, Inoue K, Ohtsubo Y, Nagata Y, Tsuda M. Conjugative transfer system of IncP-9 naphthalene-catabolic plasmid NAH7: characterization of its *oriT* region and relaxase and host range of conjugative system. *Appl Environ Microbiol* 2016; 83:e02359–16. doi:<https://doi.org/10.1128/AEM.02359-16>.
- [182] Norrander J, Kempe T, Messing J. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 1983;26:101–6.
- [183] Chang AC, Cohen SN. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* 1978;134:1141–56.
- [184] Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM, et al. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 1995;166:175–6.
- [185] Bernardi A, Bernardi F. Complete sequence of pSC101. *Nucleic Acids Res* 1984;12: 9415–26.
- [186] Lawley Trevor, Frost BMW Laura S. Bacterial conjugation in gram-negative bacteria. In: Phillips GJ, Funnell BE, editors. *Plasmid Biol*. American Society of Microbiology; 2004. p. 473–91. <https://doi.org/10.1128/9781555817732>.
- [187] Guglielmini J, Néron B, Abby SS, Garcillán-Barcia MP, la Cruz F de, Rocha EPC. Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res* 2014;42:5715–27. doi:<https://doi.org/10.1093/nar/gku194>.
- [188] Zatyka M, Jagura-Burdzy G, Thomas CM. Regulation of transfer genes of promiscuous IncP plasmid RK2: repression of Tra1 region transcription both by relaxosome proteins and by the Tra2 regulator TrbA. *Microbiology* 1994;140:2981–90. <https://doi.org/10.1099/13500872-140-11-2981>.
- [189] Herman D, Thomas CM, Stekel DJ. Global transcription regulation of RK2 plasmids: a case study in the combined use of dynamical mathematical models and statistical inference for integration of experimental data and hypothesis exploration. *BMC Syst Biol* 2011;5:119. <https://doi.org/10.1186/1752-0509-5-119>.