Plasmidomics: studying plasmids as ecological entities beyond their hosts

Bradd Mendoza-Guido[1*] and Keilor Rojas-Jiménez[2]

1. Instituto de Investigaciones en Salud, Universidad de Costa Rica

2. Escuela de Biología, Universidad de Costa Rica

*Corresponding author: bradd.mendoza@ucr.ac.cr

1. Abstract

Plasmids are autonomous mobile genetic elements whose ecology extends beyond individual bacterial hosts. As molecular symbionts, they traverse strains, species, and environments, disseminating adaptive genes and shaping microbial community structure through dynamics that are often decoupled from host taxonomy. Plasmidomics—the omics discipline dedicated to the study of plasmids—has revealed that plasmid diversity and distribution respond to environmental gradients independently of their hosts, underscoring their roles as ecological entities. However, the field faces critical methodological constraints: short-read assemblies fragment plasmid sequences, culture-dependent approaches underrepresent environmental diversity, and most metagenomic methods fail to capture plasmid–host associations. In addition, a universally accepted classification framework is still lacking. Advancing plasmidomics will require the integration of long-read sequencing, Hi-C proximity ligation, and ecology-informed classification frameworks grounded in genomic and functional criteria. Together, these approaches are essential to uncover the true diversity, evolutionary significance, and ecological dynamics of plasmids across complex microbial ecosystems.

Keywords: Microbiology, Mobile Genetic Elements, Horizontal Gene Transfer, Molecular Symbionts, Metagenomics, Molecular Evolution

2. Plasmids as Ecological Entities

Plasmids are autonomous, mobile genetic elements (MGEs) whose ecology extends well beyond the boundaries of their bacterial hosts. Defined by a dedicated replication system often supported by maintenance functions that coordinate replication with the host cell cycle [1], they traverse strains, species, and environments, accumulating and disseminating adaptive genes across microbial communities. Beyond their canonical role in horizontal gene transfer (HGT), plasmids exert a direct influence on bacterial ecology and evolution [2,3]. For instance, plasmid communities can function as dynamic genomic reservoirs in which plasmid-encoded genes evolve under selective regimes distinct from those acting on chromosomes, partly driven by elevated copy numbers and the coexistence of multiple plasmid variants within a single host [2].

Within this framework, plasmids and other MGEs are best understood as molecular symbionts, engaging in ecological host interactions that span a parasitism–mutualism continuum shaped by evolutionary trajectories, community context, and environmental selective pressures [3–5]. Importantly, the macro and micro-evolutionary trajectory and persistence of plasmids in bacterial communities appear to depend more on ecological dynamics than on the biology of any individual host [6], indicating that plasmids diversify in ways only partially coupled to host taxonomy.

This perspective provides a framework for interpreting key aspects of plasmid biology beyond the constraints of individual hosts, encompassing technological limitations, challenges in current classification systems and plasmid ecological dynamics. Advancing the field requires integrative plasmidomics approaches that combine genomic, ecological, and evolutionary perspectives. In particular, metagenomic strategies incorporating long-read sequencing and proximity-ligation methods are essential to recover complete plasmid architectures and resolve their interactions with host organisms and other molecular symbionts in natural communities, alongside the development of robust and biologically meaningful classification systems. Together, these advances are critical to uncover the true diversity, ecological dynamics, and evolutionary significance of plasmids across complex and largely uncultured ecosystems.

3. Evolution and challenges of Plasmidomics

Plasmidomics is the omics discipline focused on studying plasmids, including their interactions with host organisms, other molecular entities, and biotic and abiotic components of their ecosystems. If genomics studies genomes, plasmidomics studies the collection of plasmids (the “plasmidome”) present in a cell, organism, or environment.

The sequencing of plasmids and bacterial genomes has been widely applied for decades, substantially advancing our understanding of plasmid biology and their roles in bacterial adaptation. In recent years, the rapid expansion in the number of available bacterial genomes (both draft and complete) has driven the development of numerous bioinformatic tools and databases aimed at plasmid identification and classification. Initial methods relied on the identification of hallmark plasmid features, such as replication initiation and relaxase proteins, and the assignment of incompatibility groups through similarity searches against curated databases, as implemented in tools like PlasmidFinder [7].

However, plasmid reconstruction from draft genomes remains inherently difficult, largely due to the abundance of repetitive elements and mobile genetic components that disrupt assembly continuity and lead to fragmented or ambiguous contigs [8]. To overcome these limitations, subsequent approaches incorporated structural information by clustering contigs based on shared genomic distances and incompatibility constraints to infer contigs belonging to the same plasmid, as exemplified by tools such as MOB-suite. [9]. In parallel, machine-learning-based (MLB) methods emerged, classifying contigs as plasmid or chromosomal by integrating sequence composition, gene content, and architectural features, thereby reducing dependence on complete assemblies and reference bias [10,11]. These strategies have improved plasmid identification in draft genomes, supported by the biological assumption that bacteria typically harbor a limited number of extrachromosomal replicons and that plasmids of the same incompatibility group rarely coexist.

More recently, these methodological frameworks have been adapted to metagenomic data, where plasmid detection is further complex by extreme contig fragmentation, higher plasmid diversity, and the absence of direct host linkage information unless supported by technologies such as Hi-C. In this context, MLB tools specifically trained on metagenomic datasets have become central to plasmid discovery [10,12,13]. Notably, Yu et al. demonstrated that plasmid sequences predicted from human gut metagenomes using PlasX could be recovered from cultured isolates and experimentally transferred between *Bacteroides fragilis* strains [14].

A major limitation in plasmid research remains the poor characterization of uncultured bacterial communities and, consequently, their plasmids. Currently available metagenome-assembled genome (MAG) reconstruction methods tend to miss the recruitment of plasmid sequences due to differences in GC content, copy number, and coverage relative to chromosomal DNA [15]. This limitation results in a substantial underrepresentation of plasmids from uncultured bacteria and limits the diversity captured in reference datasets. Therefore, MLB models trained on restricted and biased collections of well-characterized plasmids may fail to detect novel or divergent plasmid sequences. Additionally, many proteins are shared between plasmids and other MGE, such as integrative and conjugative elements (ICEs), which can lead to the misclassification of ICE-derived contigs as plasmids and vice versa [16]. In addition, the ability of ICES to transition into plasmids, and of plasmids to integrate into chromosomes as ICEs, further complicates their discrimination [17].

Long-read sequencing technologies offer clear advantages for plasmid assembly and detection, particularly in metagenomic datasets. Both PacBio and Oxford Nanopore platforms routinely generate reads exceeding 10 kb, which can span most plasmids, as these are often smaller than this size [1]. For larger plasmids, long reads are frequently sufficient to bridge repetitive regions that otherwise prevent correct overlap and contig reconstruction when using short reads alone. Despite these advantages, applications of long-read sequencing for plasmid identification in metagenomes remain scarce [18–20], likely due to the higher cost and technical complexity of long-read metagenomic sequencing. Moreover, systematic benchmarking studies comparing plasmid recovery and classification using long-read approaches versus conventional short-read–based and hybrid methods are still lacking.

An alternative strategy involves enriching plasmids from microbial communities prior to sequencing to reduce chromosomal DNA contamination. Most enrichment protocols rely on alkaline lysis, which selectively denatures chromosomal DNA while allowing small, supercoiled plasmids to renature and remain soluble [21]. However, this approach preferentially recovers small plasmids and often leads to the loss of large plasmids during extraction [22], resulting in their underrepresentation despite their potential functional importance.

Other methods employ PlasmidSafe DNase to selectively digest linear double-stranded DNA, preserving large circular plasmids [23,24]. However, this strategy excludes linear plasmids from downstream analyses. Using this approach, Calderón-Osorio et al. [23] reported that only 8.2% of metagenomic reads from river samples were classified as plasmid-derived based on the plasmid RefSeq database. This low proportion suggests that a substantial fraction of plasmid diversity remains uncharacterized, either due to incomplete reference databases or residual chromosomal contamination in plasmid-enriched fractions. Collectively, these findings underscore major limitations in plasmid ecology studies and highlight the urgent need to expand and curate plasmid reference databases [25,26]. In this regard, relevant tools and databases for plasmidomic analyses are summarized in Table 1.

Table 1. Relevant tools and databases for plasmidomic analyses.

Tool / Database	Input Data	Approach	Typical Application	Main Limitations
PlasmidFinder [7]	Assembled contigs	Replicon-based identification using	Plasmid typing and incompatibility group assignment	- Limited sensitivity; detects only known replicons

		incompatibility group markers		- Restricted to certain taxonomic groups (mainly Enterobacteriaceae and Gram-positives) - Poor performance on novel or highly fragmented plasmids - Limited for genome assemblies
plasmidSPAdes [27]	Raw sequencing reads	Assembly graph–based reconstruction using coverage differences	De novo plasmid assembly from WGS data	- Requires high and differential coverage - Poor resolution for low-copy plasmids with coverage similar to chromosome - Limited for genome assemblies
MOB-suite [9,28]	Assembled contigs	Marker gene sequences + contig clustering using mash genomic distance + MOB typing	Plasmid reconstruction, relaxase/MOB typing, conjugation potential prediction and host range prediction	- Limited for genome assemblies - Dependent on assembly quality - Database-dependent for novel plasmids
PlasX [12]	Assembled metagenomic contigs	Machine learning (logistic regression) using de novo gene families (1,090,132 families). Trained on 16,827 plasmids + 14,367 chromosomes.	Identification of plasmid-derived contigs in metagenomes	- Model-dependent - User should choose score threshold (arbitrary) - Cannot recognize plasmids with signatures not in training data - 81.5% recall on PLSDB plasmids
PlasFlow [13]	Assembled metagenomic contigs (≥1 kb recommended)	Deep neural network using k-mer frequencies and genome signatures. Ensemble voting classifier with probability threshold filtering Trained on 7,604 plasmids + 1,961 chromosomes	Identification of plasmid-derived contigs with phylum-level taxonomic classification (96% accuracy)	- No longer actively maintained (highly important) - Model-dependent - Cannot recognize plasmids with signatures not in training data - 85.98% recall - May confuse integrated elements - Not recommended for short sequences
Platon [10]	Assembled contigs	Replicon Distribution Score (RDS) + plasmidic marker protein sequences + heuristics	Identification of plasmid-derived contigs in isolate genomes and metagenomes	- Model-dependent - Cannot recognize plasmids with signatures not in training data - Moderate sensitivity (~70%) - Automatically classifies contigs >500 kbp as chromosomal - 37.4% recall on PLSDB plasmids
PlasmidScope [26]	Plasmid sequences	Comprehensive reference database (852,600 plasmids	Plasmid annotation, classification, comparative analysis,	- Database tool (not de novo identification), plasmid identification must be done separately

		from 10 repositories) + annotation pipeline	and online analytical tools	
IMG/PR [25]	Plasmid sequences	Curated reference database with functional, taxonomic, and ecological metadata (699,973 plasmids from genomes, metagenomes and metatranscriptomes)	Comparative plasmid genomics, plasmid taxonomic unit (PTU) definition, host range and functional profiling	- Database resource only (requires external tools for analysis), plasmid identification must be done separately
PLSDB [29,30]	Plasmid sequences	Curated reference database derived from NCBI prokaryotic genomes (72 360 plasmids)	Plasmid comparison and annotation of complete plasmids. Plasmid detection using genomic distances against the database.	- Includes only complete plasmids from NCBI; plasmids derived from metagenomic datasets are underrepresented. - Database-dependent for novel plasmids

These methodological constraints are not merely technical limitations; they directly determine the resolution at which plasmid ecology can be studied. Gaps in assembly continuity, database coverage, and host assignment each introduce systematic blind spots in our understanding of how plasmid communities are structured across environments.

4. The Need for a Consolidated Plasmid Classification System

Another major limitation for understanding plasmid ecological dynamics is the absence of a universal classification system. Existing frameworks based on host range, or by marker typing systematically fail to capture ecologically meaningful units; particularly given that a substantial proportion of environmental plasmids lack these canonical markers or cannot be identified, and that host-range inferences are often constrained by the limited diversity of cultivar-derived reference databases. Although several classification approaches have been proposed (reviewed in [31] and summarized in Table 2), none has achieved the broad acceptance or level of standardization comparable to taxonomic systems developed for cellular organisms or, more recently, for viruses [32].

The precedent of virus taxonomy constitutes a valuable guideline to consider. Despite not being considered living organisms, viruses have long benefited from a formal taxonomic framework (with the exception of most phages); and as the availability of viral genomes expanded exponentially (including from uncultured sources) the International Committee on Taxonomy of Viruses (ICTV) adopted the genome-based similarity thresholds, defining species at ~95% ANI and genera at ~70%. This transition, however, highlights the broader challenge of how taxonomic systems should be updated and on what criteria they should be based.

In plasmid biology, we are now facing an analogous inflection point. The rapid expansion of plasmid sequences has substantially broadened our understanding of their diversity and ecology, yet also exposes the limitations of existing classification frameworks. This growing body of data demands a classification system grounded in genomic and ecological criteria, rather than relying on phenotypic traits or host-associated proxies. Accordingly, developing a consolidated, ecology-informed plasmid taxonomy is not an isolated methodological challenge, but a direct consequence of recognizing plasmids as autonomous ecological entities—the central argument advanced throughout this perspective.

Table 2. Advantages and limitations of the main plasmid classification systems reviewed in [31].

Classification system	**Key features**	**Advantages**	**Limitations**
Fertility inhibition (fi)	- Ability to inhibit F plasmid conjugative transfer; *fi+* plasmids inhibit F transfer, *fi–* plasmids do not	- Early systematic approach; helped identify relationships between plasmids	- Applied to two different phenomena - Very limited scope; gradually abandoned
Incompatibility	- Stable coexistence in the same cell; shared replication and partition systems	- Broad consensus effort with several Inc groups established - widely used in epidemiological studies -Constituted the basis for the replicon typing ssytem	- Not testable in plasmids derived from unculturable MAGs. - Plasmids may break incompatibility rules - Requires mating experiments to test plasmid stability - Requires different antibiotic markers - Exponentially increasing test combinations
Replicon typing	- Replication origin sequences - Initially based on DNA hybridization and PCR, later moved to genome sequence–based detection (PlasmidFinder) - Includes plasmid Multi Locus Sequence Type for some Inc groups - 471 replicon sequences	- Universal plasmid characteristics - Well established databases	- Essentially a genetic analogue of the incompatibility system - Restricted to certain taxonomic groups (mainly Enterobacteriaceae and Gram-positives) - Multi-replicon plasmids cannot be classified into a single group. - Same replicon in different PTUs - Cut-off values do not guarantee incompatibility - Dependent of known replicon sequences
MOB class	- Relaxase gene identity and phylogeny - 9 MOB classes	- Robust universal marker - Only gene shared by most transmissible elements	- Cannot be applied to non-mobilizable plasmids which represent a great portion of plasmids

	- PCR typing available (for those in Gammaproteobacteria) - MOBscan web server [33]	- More consistent than replicon typing - Smaller number of classes (9 vs. 471 replicons)	- MOB and Rep phylogenies not always congruent - Single-gene limitations
pATLAS [34]	- Alignment-free (Mash, *k* = 21) - visual network tool including 13,924 plasmids Jaccard index threshold = 0.9	- User-friendly exploration tool - can identify plasmid s similar to queries. - includes metadata	- No further community definition - No true classification - Not updated
Plasmid cliques [35]	- Alignment-free (BinDash, *k* = 21) - 566 cliques (plasmid groups) with 561 containing 3 or more plasmids - Jaccard index threshold = 0.3	Highly correlated with gene content, host, GC content, replicon type, and MOB class - Can infer HGT events	- Some plasmid assigned to multiple clique s - Stringent threshold (low diversity) - No automatic allocation tool for new plasmid sequences
MOB-suite clustering [9]	- Alignment-free (Mash, *k* = 21) - 7,298 clusters - Complete-linkage hierarchical clustering - Mash distance threshold = 0.06	- Includes MOB-typer and MOB-recon tools - Plasmid reconstruction from draft assemblies - Automated classification tool available	- Replicon prediction does not totally agree with PlasmidFinder database - Stringent threshold leading to 53% singletons - limited to Proteobacteria MOB/MPF types
PTU (Plasmid Taxonomic Units) [36]	- Alignment-based (ANIL50) - 380 PTUs - Hierarchical Bayesian stochastic block modeling–based algorithm using ≥70% sequence identity across ≥50% alignment coverage of the smaller genome.	- No bias toward specific characteristics - clear genetic discontinuity - includes cliques and paracliques - COPLA: tool for automatic assignment (web based) host-range classification	- 253/380 PTUs lack PlasmidFinder replicons - Members of the same PTU may be compatible (do not match Inc groups) - Plasmids within the same PTU can harbor distinct relaxase (MOB) classes - Purely threshold-based clustering with limited biological relevance - Studies linking PTU structure to biological or ecological traits remain scarce

A more recent genome-based strategy was explored by Redondo-Salvo et al. (2020), [36] who clustered more than 10,000 plasmids into plasmid taxonomic units (PTUs) based on high sequence similarity. PTU reconstruction relied on an alignment fraction (AF) ≥50% and an average nucleotide identity (ANI) threshold of 70%, a combination designed to avoid spurious clustering of short plasmid sequences sharing only transposable elements. This approach was later adopted by the Integrated Microbial Genome (IMG) team, which applied similar approach for constructing their plasmid database (Table 1) [25].

However, clustering-based approaches can vary substantially depending on the software and algorithms used to group plasmid sequences, even when applying the same parameters. This issue is particularly pronounced when clustering large graphs, where small changes in network structure can lead to markedly different cluster definitions. For example, the IMG team employed vclust with the Leiden algorithm to define PTUs using the parameters described above [25]; however, vclust implements multiple clustering algorithms, each of which can produce distinct plasmid groupings [37]. Similar limitations apply to other tools for plasmid classification which are based on clustering algorithms. Importantly, the choice of clustering approach should be guided by the specific scientific objective. Although most of these methods do not provide a formal taxonomic framework, they can still be valuable for addressing ecological questions related to plasmid diversity and structure.

The main advantage of the PTU-based approaches is its reliance on whole-genome sequences for clustering. By using complete sequence information, it captures a broader spectrum of plasmid features and allows the inclusion of plasmids that lack canonical mobilization genes commonly used for classification, or even the absence of replication proteins (for plasmid fragments or with unknown replicases). And despite the ANI-based tools do not infer phylogenetic relationships by modeling the evolutionary trajectory of sequences, in bacterial genomes such distances may strongly correlate with phylogenetic analyses, often reflecting evolutionary classifications [38].

Nevertheless, this strategy remains constrained by important limitations. For example, in the IMG/PR analysis, 699,973 plasmid sequences were clustered into 214,950 PTUs, with approximately 63% represented by a single plasmid. This high proportion of singletons indicates that even whole-genome–based approaches fail to fully reconstruct plasmid evolutionary relationships [25].

Plasmids evolve in a highly modular manner, undergo frequent recombination, and can change rapidly over short evolutionary timescales, thereby readily disrupting fixed identity and coverage thresholds used for clustering. More fundamentally, this limitation echoes an assumption previously raised by Garcillán-Barcia et al. [1]: when a plasmid has gradually replaced its entire original gene repertoire, does it retain its identity, or has it effectively become a different entity? This “plasmid Theseus’ paradox” highlights the conceptual difficulty of reconstructing plasmid macro-evolutionary lineages and suggests intrinsic limits to any classification method for plasmids. Collectively, these observations underscore the need for future studies that move beyond static bioinformatic schemes to explicitly address the ecological and evolutionary processes driving plasmid modularity to these limits.

An alternative approach that can capture the gradual functional changes of plasmids is the use of best bidirectional hits (BBHs) between gene sequences, particularly focusing on proteins that tend to be more conserved. This strategy has been widely applied to cluster many different molecular symbionts such as plasmids, ICEs and phage–plasmids [17,39]. By relying on shared protein content rather than strict ANI-based thresholds, BBH-based approaches can recover deeper and more divergent relationships among molecular symbionts that retain at least a minimal set of homologous proteins. This enables the detection of large-scale evolutionary patterns that are often obscured by ANI-based methods.

Moreover, gene-content–based frameworks could provide a more biologically meaningful basis for plasmid classification by applying a hierarchical strategy in which plasmids are first grouped according to key functional modules involved in replication (Inc groups), mobility (relaxases) and conjugation (secretion systems), and subsequently refined using conserved gene content across the plasmid backbone.

Despite these advantages, BBH-based clustering has not yet been systematically applied to develop a plasmid classification system using a comprehensive plasmid dataset, representing a critical gap in current plasmid research. A key limitation is the requirement for accurate annotation of coding sequences, a computational and tool-dependent step that may introduce variability across analyses. Nevertheless, integrating gene-content– and ANI-based approaches together may provide a more robust framework for resolving plasmid evolutionary relationships and advancing toward a biologically grounded plasmid classification system.

The limitations outlined above extend beyond methodological limits and point to a more fundamental issue in how plasmids are studied and classified. Recent plasmidomic studies have begun to address this gap, revealing ecological patterns that challenge traditional host-centric views of plasmid biology. Importantly, the findings discussed in the following section should be interpreted as conservative estimates of the true complexity of plasmid ecology, still constrained by the methodological and taxonomic limitations described above, yet already compelling enough to underscore the need for a more systematic and ecology-informed exploration of plasmid dynamics.

5. Plasmid Ecology goes Beyond Host Ecology

One major limitation of classifying or studying plasmids based on host taxonomy is that plasmids can exhibit ecological patterns that are decoupled from their hosts. For example, plasmids derived from the human gut microbiomes of individuals from industrialized countries have been reported to form distinct plasmid systems compared with those from non-industrialized countries [12], with variable gene repertoires, including genes related to antibiotic resistance and efflux pumps. Notably, most predicted plasmids in that study did not display distributions correlated with their putative host taxonomy, highlighting plasmids as highly dynamic components of microbiomes with ecological patterns differ from those of their bacterial hosts (Figure 1A). Although the approach used by the authors to estimate host-taxonomy represents a more rigorous mathematical framework—leveraging Pearson correlations and Jaccard indices to estimate ecological co-occurrence between plasmids and potential hosts across metagenomes—it remains dependent on short-read data and read mapping-based taxonomic assignment, inheriting the assembly fragmentation and databases-detection biases outlined above.

Further illustrating this disconnect, the cryptic plasmid pBI143 appears to be one of the most abundant genetic elements in the human gut, predominantly detected in individuals from industrialized countries compared with those from non-industrialized [40]. Despite no measurable impact on host fitness *in vivo*, its presence has been associated with inflammatory states in the human colon, underscoring the importance of plasmid dynamics in microbiome function. In that study, the authors raise the question of whether current and future research should expand the concept of a “core microbiome” beyond microbial taxa to include plasmids and other molecular symbionts that are more prevalent than their hosts across populations, potentially advancing our understanding of microbial ecology (Figure 1B). This is a compelling reframing with broad implications for microbiome research. Resolving it, however, requires methods capable of directly linking plasmids (and other molecular symbionts) to their hosts within complex communities such as Hi-C metagenomics.

Similarly, studies have shown that in a urban river, increasing levels of pollution are associated with higher plasmid diversity, while chromosomal diversity is reduced or remains stable [23,41]. This pattern suggests that plasmid communities may respond more rapidly or sensitively to environmental stressors, consistent with the well-established role of plasmids as vectors of adaptive traits shaped by strong selective pressures (Figure 1C). However, these analyses rely on short-read data mapped to plasmid databases derived from known hosts and use bacterial host taxonomy to interpret plasmid diversity, which may not accurately reflect the true diversity of plasmids or their associated hosts.

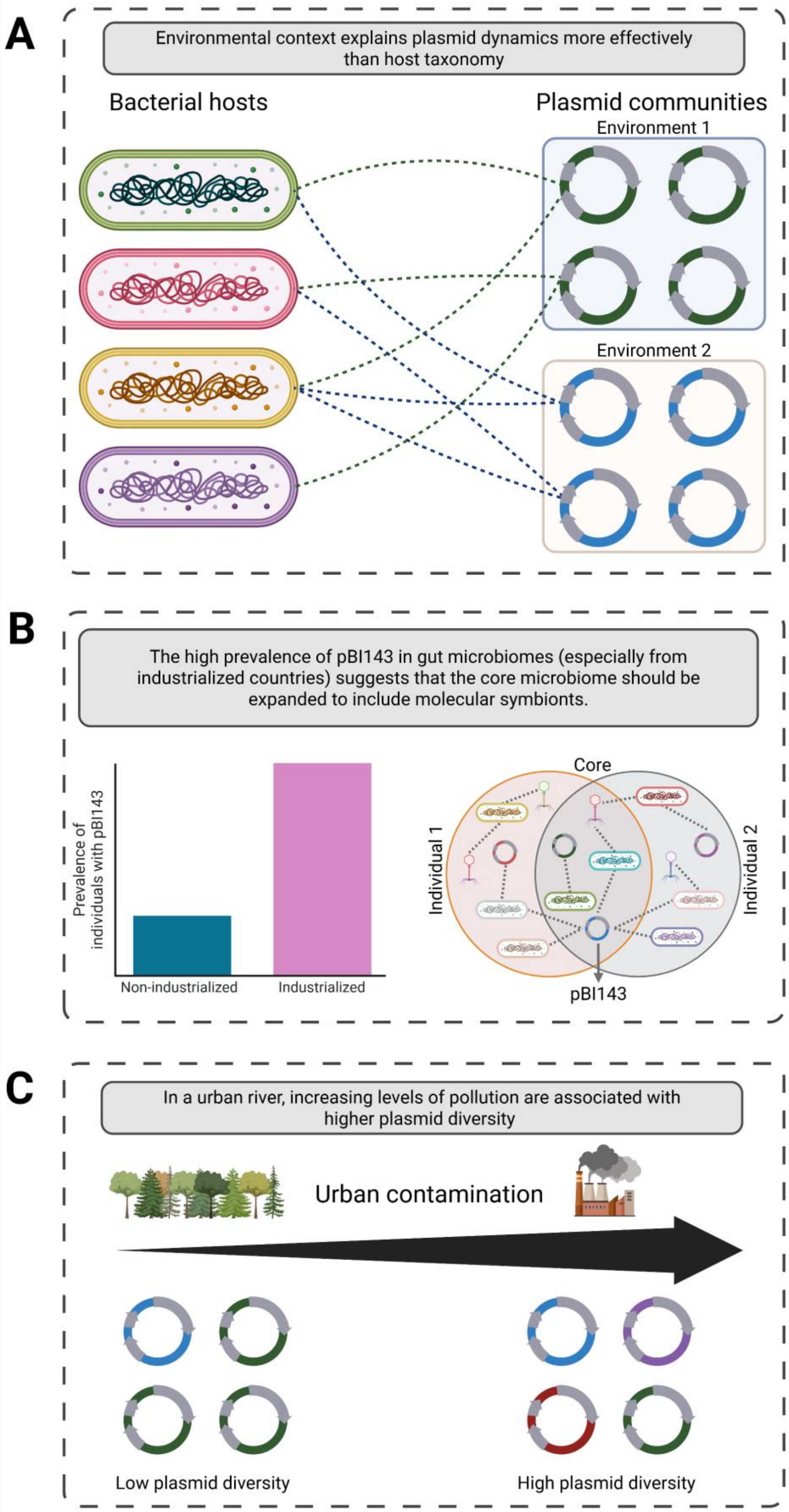
A
Environmental context explains plasmid dynamics more effectively than host taxonomy
Bacterial hosts
Plasmid communities
Environment 1
Environment 2
B
The high prevalence of pBI143 in gut microbiomes (especially from industrialized countries) suggests that the core microbiome should be expanded to include molecular symbionts.
Prevalence of individuals with pBI143
Non-industrialized
Industrialized
Core
Individual 1
Individual 2
pBI143
C
In a urban river, increasing levels of pollution are associated with higher plasmid diversity
Urban contamination
Low plasmid diversity
High plasmid diversity

Figure 1. Some examples of how plasmid ecology goes beyond host taxonomy. **A)** Plasmid communities are structured more strongly by environment than by host taxonomy. **B)** The prevalence of pBI143, one of the most abundant genetic elements in the human gut, is higher in industrialized countries, underscoring the need to expand the “core microbiome” concept to include molecular symbionts. The gray dashed lines indicate molecular symbiont–host interactions, illustrating that some elements, such as pBI143, can interact with multiple hosts; while the hosts themselves may not belong to the core microbiome, the plasmid does. **C)** Plasmid diversity increases with pollution levels.

More broadly, inferring plasmid host taxonomy based solely on previously reported associations is inherently limited. Plasmids often exhibit broad host ranges spanning from species to phyla [36] many of which remain poorly characterized. Moreover, plasmids are often mosaic elements containing sequences derived from multiple replicons, which further reduces the accuracy of read-mapping-based host assignment.

Although accurately identifying plasmid hosts remains challenging with current methods, emerging approaches such as Hi-C metagenomic sequencing offer promising alternatives and have been successfully applied to resolve molecular symbionts–host interactions [42,43]. Other tools attempt to infer plasmid host taxonomy using MLB or genetic distances approaches [28,44], but these methods share similar limitations because they rely on datasets of known plasmid–host associations, which likely underestimate the true breadth of plasmid host ranges.

6. Concluding Remarks and Future Perspectives

Plasmidomics is consolidating plasmids as autonomous ecological entities whose diversity and dynamics are shaped as much (if not more) by environmental context than by their bacterial hosts. The evidence reviewed here consistently challenges host-centric frameworks: plasmid communities can respond independently to environmental gradients, with patterns that are not reflected in host taxonomy, highlighting the pervasive and distinct ecological roles of plasmids across ecosystems, including the human microbiome. Consequently, studying plasmids in isolation from their ecological context is no longer tenable, and future plasmidomic research must adopt integrative, ecology-informed frameworks.

Yet, the field faces compounding limitations that constrain this perception. Short-read assemblies and culture-dependent approaches systematically underrepresent environmental plasmid diversity, meaning that current ecological patterns likely capture only a fraction of the true plasmid landscape. At the same time, the lack of a universal, ecology-informed classification system continues to hinder meaningful comparisons across studies and ecosystems. The integration of long-read sequencing, particularly when combined with Hi-C proximity ligation, will be critical to resolve plasmid structure and host associations at scale. Finally, recognizing plasmids as ecological units calls for a reassessment of foundational concepts such as the “core microbiome”, expanding it to include persistent mobile genetic elements alongside their microbial hosts.

7. *Acknowledgments*

BMG thanks Eduardo P. C. Rocha for insightful discussions on plasmid and bacterial biology, as well as for reading the manuscript and providing valuable comments.

8. References

1. Garcillán-Barcia, M.P. *et al.* (2025) The extended mobility of plasmids. *Nucleic Acids Res* 53, gkaf652
2. Rodríguez-Beltrán, J. *et al.* (2021) Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nat Rev Microbiol* 19, 347–359
3. Mendoza-Guido, B. and Rojas-Jimenez, K. (2025) Beyond plasmid addiction: the role of toxin–antitoxin systems in the selfish behavior of mobile genetic elements. *Journal of Bacteriology* 207, e00232-25
4. Jalasvuori, M. and Koonin, E.V. (2015) Classification of prokaryotic genetic replicators: between selfishness and altruism. *Annals of the New York Academy of Sciences* 1341, 96–105
5. Filée, J. (2018) Giant viruses and their mobile genetic elements: the molecular symbiosis hypothesis. *Current Opinion in Virology* 33, 81–88
6. Bottery, M.J. (2022) Ecological dynamics of plasmid transfer and persistence in microbial communities. *Current Opinion in Microbiology* 68, 102152
7. Carattoli, A. and Hasman, H. (2020) PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS). In *Horizontal Gene Transfer: Methods and Protocols* (de la Cruz, F., ed), pp. 285–294, Springer US
8. Arredondo-Alonso, S. *et al.* (2017) On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microbial Genomics* 3, e000128
9. Robertson, J. and Nash, J.H.E. (2018) MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microbial Genomics* 4
10. Schwengers, O. *et al.* (2020) Platon: identification and characterization of bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores. *Microbial Genomics* 6, e000398
11. Pradier, L. *et al.* (2021) PlasForest: a homology-based random forest classifier for plasmid detection in genomic datasets. *BMC Bioinformatics* 22, 349
12. Yu, M.K. *et al.* (2024) Diverse plasmid systems and their ecology across human gut metagenomes revealed by PlasX and MobMess. *Nat Microbiol* 9, 830–847
13. Krawczyk, P.S. *et al.* (2018) PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res* 46, e35
14. Zou, X. *et al.* (2022) Classification of bacterial plasmid and chromosome derived sequences using machine learning. *PLOS ONE* 17, e0279280

15. Maguire, F. *et al.* (2020) Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic Islands. *Microbial Genomics* 6, e000436
16. Mendoza-Guido, B. *et al.* (2025) Comprehensive plasmidomic analysis of Helicobacter pylori reveals the potential role of plasmids in pathogenic adaptation and a novel putative toxin-antitoxin system. *FEMS Microbiol Lett* 372, fnaf092
17. Cury, J. *et al.* (2018) Host Range and Genetic Plasticity Explain the Coexistence of Integrative and Extrachromosomal Mobile Genetic Elements. *Mol Biol Evol* 35, 2230–2239
18. Peng, K. *et al.* (2023) Long-read metagenomic sequencing reveals that high-copy small plasmids shape the highly prevalent antibiotic resistance genes in animal fecal microbiome. *Science of The Total Environment* 893, 164585
19. Masuda, S. *et al.* (2024) Uncovering microbiomes of the rice phyllosphere using long-read metagenomic sequencing. *Commun Biol* 7, 357
20. Dai, D. *et al.* (2022) Long-read metagenomic sequencing reveals shifts in associations of antibiotic resistance genes with mobile genetic elements from sewage to activated sludge. *Microbiome* 10, 20
21. Singh, V.K. *et al.* (2022) Protocols for Isolation of Plasmid DNA. In *Protocols for the Diagnosis of Pig Viral Diseases* (Deb, R. et al., eds), pp. 91–107, Springer US
22. Clemson, M. and Kelly, W.J. (2003) Optimizing alkaline lysis for DNA plasmid recovery. *Biotechnology and Applied Biochemistry* 37, 235–244
23. Calderón-Osorno, M. *et al.* (2026) Insights into the riverine metaplasmidome: The role of anthropogenic factors in shaping microbial communities. *The Microbe* 10, 100640
24. Brown Kav, A. *et al.* (2013) A method for purifying high quality and high yield plasmid DNA for metagenomic and deep sequencing approaches. *Journal of Microbiological Methods* 95, 272–279
25. Camargo, A.P. *et al.* (2024) IMG/PR: a database of plasmids from genomes and metagenomes with rich annotations and metadata. *Nucleic Acids Res* 52, D164–D173
26. Li, Y. *et al.* (2024) PlasmidScope: a comprehensive plasmid database with rich annotations and online analytical tools. *Nucleic Acids Research* DOI: 10.1093/nar/gkae930
27. Antipov, D. *et al.* (2016) plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 32, 3380–3387
28. Robertson, J. *et al.* (2020) Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance. *Microbial Genomics* 6, e000435
29. Galata, V. *et al.* (2019) PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Research* 47, D195–D202
30. Molano, L.-A.G. *et al.* (2025) The PLSDB 2025 update: enhanced annotations and improved functionality for comprehensive plasmid research. *Nucleic Acids Res* 53, D189–D196
31. Garcillán-Barcia, M.P. *et al.* (2023) Plasmid classifications. *Plasmid* 126, 102684
32. Moraru, C. *et al.* (2020) VIRIDIC—A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. *Viruses* 12

33. Garcillán-Barcia, M.P. *et al.* (2020) MOBscan: Automated Annotation of MOB Relaxases. In *Horizontal Gene Transfer: Methods and Protocols* (de la Cruz, F., ed), pp. 295–308, Springer US
34. Jesus, T.F. *et al.* (2019) Plasmid ATLAS: plasmid visual analytics and identification in high-throughput sequencing data. *Nucleic Acids Res* 47, D188–D194
35. Acman, M. *et al.* (2020) Large-scale network analysis captures biological features of bacterial plasmids. *Nat Commun* 11, 2452
36. Redondo-Salvo, S. *et al.* (2020) Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat Commun* 11, 3602
37. Zielezinski, A. *et al.* (2025) Ultrafast and accurate sequence alignment and clustering of viral genomes. *Nat Methods* 22, 1191–1194
38. Abram, K. *et al.* (2021) Mash-based analyses of Escherichia coli genomes reveal 14 distinct phylogroups. *Commun Biol* 4, 1–12
39. Pfeifer, E. and Rocha, E.P.C. (2024) Phage-plasmids promote recombination and emergence of phages and plasmids. *Nat Commun* 15, 1545
40. Fogarty, E.C. *et al.* (2024) A cryptic plasmid is among the most numerous genetic elements in the human gut. *Cell* 187, 1206-1222.e16
41. Barrantes-Jiménez, K. *et al.* (2025) Influence of Seasonality and Pollution on the Presence of Antibiotic Resistance Genes and Potentially Pathogenic Bacteria in a Tropical Urban River. *Antibiotics (Basel)* 14, 798
42. Wu, R. *et al.* (2023) Hi-C metagenome sequencing reveals soil phage–host interactions. *Nat Commun* 14, 7666
43. Robinson, C.R.P. *et al.* (2025) Hi-C-resolved metagenomics reveals host range variation among mobile genetic elements within the European honey bee. *mBio* 16, e02243-25
44. Ji, Y. *et al.* (2023) HOTSPOT: hierarchical host prediction for assembled plasmid contigs with transformer. *Bioinformatics* 39, btad283