



UvA-DARE (Digital Academic Repository)

Assembly-based analysis of the infant gut microbiome reveals novel ubiquitous plasmids

Noordzij, Hanna Theodora; Wortel, Meike T.; Heintz-Buschart, Anna; Petrikonyte, Patricija; de Muinck, Eric J.; Trosvik, Pål

DOI

[10.1016/j.plasmid.2025.102761](https://doi.org/10.1016/j.plasmid.2025.102761)

Publication date

2025

Document Version

Final published version

Published in

Plasmid

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Noordzij, H. T., Wortel, M. T., Heintz-Buschart, A., Petrikonyte, P., de Muinck, E. J., & Trosvik, P. (2025). Assembly-based analysis of the infant gut microbiome reveals novel ubiquitous plasmids. *Plasmid*, 134, Article 102761. <https://doi.org/10.1016/j.plasmid.2025.102761>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.
UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)



Original Research Article

Assembly-based analysis of the infant gut microbiome reveals novel ubiquitous plasmids

Hanna Theodora Noordzij^{a,*}, Meike T. Wortel^b, Anna Heintz-Buschart^b, Patricija Petrikonyte^a, Eric J. de Muinck^{a,c}, Pål Trosvik^a

^a Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway

^b Swammerdam Institute for Life Sciences (SILS), University of Amsterdam, Amsterdam, Netherlands

^c Department of Pharmacy, University of Oslo, Oslo, Norway



ARTICLE INFO

Keywords:

Infant
Gut
Microbiome
Plasmid
Mobile genetic element
De novo assembly
Metagenome

ABSTRACT

Little is known about the role of mobile genetic elements in natural ecosystems such as the infant gut microbiome. Here, we conduct the most comprehensive longitudinal study of the infant plasmidome to date by analyzing monthly fecal samples from 12 infants from birth to one year of age. We employ an assembly-based bioinformatic pipeline for the reconstruction and identification of full-length plasmids, including a novel approach for assigning putative plasmid hosts. We then investigated plasmid content and dynamics in the infant gut microbiome. After assembly and identification, we identified 620 unique circular plasmids in the infant cohort, including a number of novel sequences. Independent assembly of the same plasmids in several samples and infants helped corroborate the authenticity of the plasmids. Among the observed plasmids was the recently described ubiquitous and abundant *Bacteroides* plasmid pBI143. Overall, the genus *Bacteroides* had the highest plasmid carriage, while the highest plasmid diversity was observed in *Clostridium*, including 5 previously unknown widespread plasmids. Lastly, we leveraged the longitudinal nature of our dataset to investigate contemporaneous correlations between temporal variations in plasmid abundances and species dynamics. This enabled us to link co-residing plasmids and tightly linked plasmid-taxon pairs within each infant. These insights into plasmid ecology help us understand determinants driving plasmid distribution in complex microbial communities.

1. Introduction

In contrast to the relatively stable microbiome of the adult human gastrointestinal (GI) tract, the infant GI microbiome is highly dynamic, individual, and passes through a series of developmental stages (de Muinck and Trosvik, 2018; Ferretti et al., 2018). Colonization of the GI tract begins during birth and is typically dominated by facultatively aerobic species, such as members of the *Enterobacteriaceae* family, in the first weeks of life. As the gut environment becomes more anaerobic, microbes that can utilize available oxygen are outcompeted by strict anaerobes and the community gradually progresses towards a more adult-like state that is dominated by *Bacteroidota* and *Firmicutes* (Penders et al., 2014).

Throughout the colonization period of the first year of life, the GI microbiome becomes more diverse (Yatsunenko et al., 2012; de Muinck and Trosvik, 2018). Interestingly, although taxonomic diversity is lower

in infants than adults, the diversity of the plasmid composition within the microbiome has been observed to be higher in infants than their mothers (He et al., 2024). The reasons for this are not well understood, but it is likely that plasmids play an important role in adaptive processes of the GI microbiome during early life, by dissemination of genes harbored on plasmids or other processes facilitated by plasmids such as the evolution of new alleles (He et al., 2024).

In environments of high bacterial density, like the GI tract, horizontal gene transfer (HGT) is common (McInnes et al., 2020; Smillie et al., 2011), and exchange of genetic material is one of the fundamental mechanisms of evolution in bacteria (McInnes et al., 2020; Perron et al., 2015; Haudiquet et al., 2022). Furthermore, HGT can happen at particularly high frequencies in newly colonized environments (Bravo et al., 2023) like the infant GI tract, suggesting an enhanced rate of adaptation during ecosystem establishment. Focusing on processes that are affected by plasmids, bacteria can evolve through acquisition of

* Corresponding author.

E-mail address: h.t.noordzij@ibv.uio.no (H.T. Noordzij).

<https://doi.org/10.1016/j.plasmid.2025.102761>

Received 7 November 2024; Received in revised form 4 September 2025; Accepted 7 October 2025

Available online 13 October 2025

0147-619X/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

novel genes, increased gene expression through regulation of plasmid copy numbers, or obtaining novel gene functionality by de novo acquisition of a chromosomally encoded gene that is decoupled from selective restraints (Rodríguez-Beltrán et al., 2021). Importantly, the complement of plasmids within a host-associated microbiome has implications for health and disease, and are linked to increased carriage of antimicrobial resistance genes (ARGs) (Stockdale et al., 2022).

Different lineages of bacteria have varying affinities for plasmid carriage (Rodríguez-Beltrán et al., 2021). For example, *Bacteroides* species, a nearly ubiquitous genus colonizing the GI of both infants (He et al., 2024) and adults (Nguyen and Vedantam, 2011), often carry one or more plasmids. On the other hand, the predominant gut bacterium *Faecalibacterium prausnitzii* is not typically observed to harbor plasmids (Fitzgerald et al., 2018; Machado et al., 2022). In addition, plasmids can also be shared by multiple lineages, such as the pB10, which was transferred to 15 genera within both alpha- and gammaproteobacteria (De Gelder et al., 2005). Another example is the small cryptic plasmid pBI143 (without known functionality), which is highly abundant in the human GI tract of industrialized populations. This plasmid has been found in 11 species across 3 genera; *Bacteroides*, *Phocaeicola*, and *Parabacteroides* (Fogarty et al., 2024). In general, broad host-range plasmids have been found to be relatively wide-spread in the human gut microbiome, and HGT through plasmid dissemination is thought to occur even across phyla (Yang et al., 2023).

Most research on plasmids has been motivated either by clinical relevance, e.g. HGT of plasmid borne carbapenem resistance in hospitals (León-Sampedro et al., 2021), or by applications in biotechnology and molecular biology, e.g. tuning plasmid copy number for optimized protein production (Rouches et al., 2022). Furthermore, studies are typically limited to plasmids that are retained under laboratory conditions by cultivable host bacteria. Recently, however, bioinformatic tools have enabled extensive identification of putative plasmid DNA sequences (Camargo et al., 2023), including a large, cross-sectional meta-analysis of the plasmid content of the infant gut microbiome (He et al., 2024). These tools have been instrumental in beginning to uncover the natural ecology and distribution of plasmids in complex microbial assemblages.

Here, we apply a pipeline of established bioinformatics tools that we optimized for taxonomic assignment to identify plasmids and their putative hosts in a longitudinal data set of deeply sequenced metagenomes of a cohort of infants born to term in Oslo, Norway (BabyBiome cohort), from birth and throughout the first year of life (Trosvik et al., 2024). We found that plasmids associated with *Bacteroides* were the most abundant, including the plasmid pBI143. The highest diversity of plasmids was found within *Clostridium*, where we uncover several novel plasmids and identify widespread types, including one potentially clinically relevant plasmid carrying a glycopeptide ARG. Although overall plasmid abundance fluctuated a great deal over the first year of life, plasmid diversity displayed significant increases over the same period. Among some lineages we saw opposing trends, e.g. with *Bacteroides* and *Clostridium*-associated plasmids increasing over time, while plasmids associated with *Enterococcus*, *Klebsiella* and *Streptococcus* decreased over the same period. Lastly, we leveraged the longitudinal nature of our data to investigate contemporaneous correlations between temporal variations in plasmid abundances and species dynamics. This enabled us to link co-residing plasmids and tightly linked plasmid-taxon pairs within each infant. Further analysis even uncovered the potential for detection of an HGT event in natural ecosystems. Together, these insights into plasmid ecology help us to understand determinants driving plasmid distribution and sequestration in complex microbial communities.

2. Methods

2.1. Sample set

A total of 149 samples collected from 12 infants born to term in Oslo,

Norway, were analyzed in this study (Trosvik et al., 2024). Most fecal samples were frozen at -20°C immediately upon collection, pending transport, on dry ice, to a -80°C storage facility at the University of Oslo. A subset of fecal samples was collected while families were traveling (de Muinck and Trosvik, 2018), and stored in 95 % alcohol before delivery to the University of Oslo.

DNA extraction from all samples was done with the PowerSoil 96 well DNA isolation kit (MO BIO Laboratories Inc., Carlsbad, CA, USA), per instructions provided by the manufacturer. Shotgun sequencing libraries were prepared with the Kapa Hyper library kit (Roche) and sequenced on an Illumina NovaSeq 6000 apparatus using a full S4 flow cell and 150 PE mode. Raw sequence reads were trimmed and filtered using Trimmomatic v0.39 with default settings apart from a MINLEN of 50 (Bolger et al., 2014). Only read pairs where both reads survived trimming and filtering were used in subsequent analysis. Reads with a host origin were identified by alignment to the human genome sequence CRCh38.p13 using Bowtie2 v2.4.223 (Langmead and Salzberg, 2012) and removed using SAMTools v1.1124 (Danecek et al., 2021). Quality control of trimmed and filtered sequencing libraries with FastQC v0.12.1 and MultiQC v1.22.3 showed no evident quality issues. After trimming and filtering the mean per library read depth was 121,121,684 reads ($\pm 32,477,513$ s.d.; median 120,785,932; range 8,799,966-291,140,538). Sequence data can be found in the NCBI SRA database under BioProject PRJNA1085219.

2.2. Plasmid assembly

The clean and trimmed sequences were here assembled for circular plasmids using metaplasmidSPAdes.py from SPAdes v3.15.3 (Antipov et al., 2020) (Fig. 1). Assembled circular contigs from metaplasmidSPAdes were identified as plasmids using end-to-end geNomad v1.7.4 with default settings (Camargo et al., 2023). Using the default settings of MobMess v1 (Yu et al., 2024), plasmid contigs were clustered to a non-redundant subset of plasmid representatives per infant. In addition, all infant plasmid representatives were compared and placed in evolutionary plasmid systems (containing the same plasmid backbone) using MobMess. Plasmid representatives that clustered with plasmids from other infants were giving the same plasmid cluster number.

2.3. Protein annotation

Plasmid representatives per infant were annotated for ARGs using Resistance Gene Identifier (RGI) software v6.0.1 and the Comprehensive Antibiotic Resistance Database (CARD) v3.3.0 (Alcock et al., 2020). The option `-low_quality` was used for potential partial genes at the start and end of the sequence, and `-keep` to keep found ORFs. ORFs were further annotated with the non-redundant UniRef90 database from Universal Protein Resource (UniProt) (The UniProt Consortium, 2023) using DIAMOND v11.3.0 (Buchfink et al., 2021). CARD annotations with at least 80 % fragment coverage and identity were considered hits. UniRef90 annotations with at least 80 % fragment coverage and an e-value <0.001 were considered hits.

2.4. Classification of mobility

Nucleotide sequences were screened for origins of replication (DoriC) and origins of transmission (OriT) using blastn in BLAST+ v2.13.0 (Camacho et al., 2009) and the DoriC v10.0 and OriT databases respectively (Luo and Gao, 2019; Ares-Arroyo et al., 2023). Hits with at least 95 % identity and 90 % coverage were kept. In addition, CONJScan from MacSyFinder v2 (Néron et al., 2023) was used to classify plasmids that can conjugate (CONJ), including their type 4 secretion system (T4SS). The remaining plasmids were typed either as mobile with the help of another residing CONJ plasmid (MOB) or not mobile (MOBless). Plasmids assigned MOBless by MacSyFinder but that have an OriT were

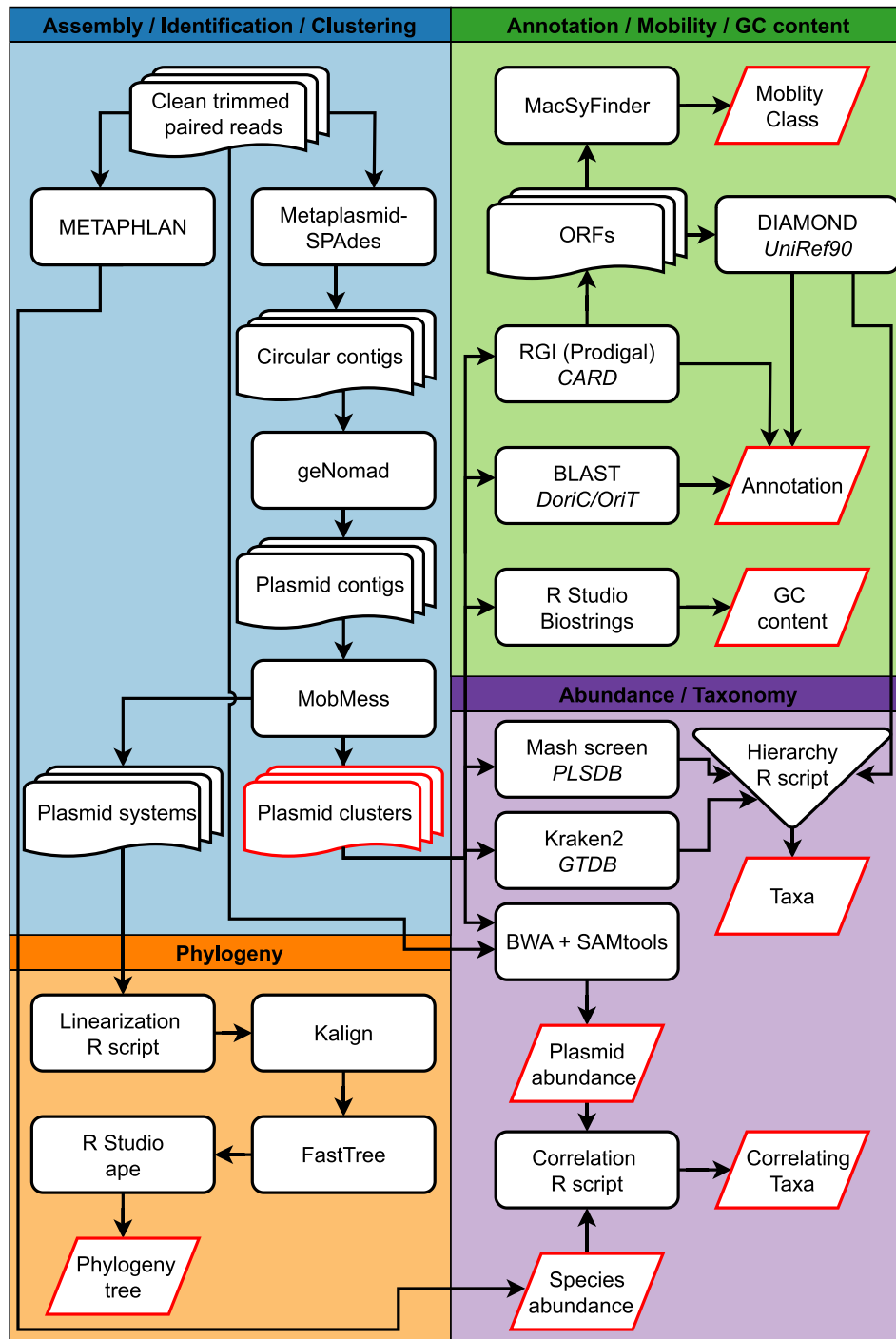


Fig. 1. Flowchart of the methods. Clean and trimmed reads from monthly samples of 12 infants were used (Trosvik et al., 2024). Reads were assembled with metaplasmidSPAdes and resulting contigs were identified as plasmids using geNomad. Plasmid contigs were clustered with MobMess. The same tool was used to identify evolutionary plasmid systems between all found clusters. Resulting plasmid representatives were annotated for ARGs and origins of replication/transfer. Unannotated ORFs were further annotated with UniRef90. Plasmid classification was done using MacSyFinder. Containment of plasmids within the PLSDB was done using mash screen. Taxonomic assignment of all plasmids was done with Kraken2 and the GTDB genome and plasmid databases. With a custom R script, putative plasmid hosts were assigned for each plasmid. Reads were mapped back to plasmid representatives per infant and used for calculating abundance. Plasmid families appearing in >5 infants were linearized to start on the same protein. These sequences were aligned with kalign, trees calculated with FastTree and drawn with the R library ape. Lastly, species abundance per sample from METAPHLAN was correlated with plasmid abundance to find potential plasmid hosts within the infants. The flowchart was made using draw.io and RColorBrewer colors.

reassigned to MOB. Gene maps were made with geneviewer (van der Velden, 2025).

2.5. Plasmid abundance

Reads were mapped to the infant's own plasmid representatives using BWA mem v0.7.17 (H. Li and Durbin, 2009). Mapped reads were filtered, sorted, indexed, and coverage was calculated using SAMtools

v1.17 (Danecek et al., 2021). Proper paired reads with at least 120 bp alignment ($-f 2 -e \text{'rlen}>120'$) were kept. Plasmid presence was decided for plasmids with at least 80 % coverage and a minimum readdepth of 10. Plasmid abundance was calculated as reads per kilobase per million reads mapped (RPKM), i.e. (mapped reads * 10^9) / (plasmid length * total reads) (Kaminski et al., 2015).

2.6. Taxonomy

Plasmid representatives per infant contained within the plasmid database v2024_05_313_v2 (PLSDB) (Schmartz et al., 2022), a database of plasmids derived from cultivating isolates, were found using mash screen v2.3 (Ondov et al., 2019), with the options $-v 0.1 -i 0.99$. Some plasmids from PLSDB were found in more than one taxonomic genus and were assigned as putative Broad Host Range (pBHR) plasmids, followed by their shared taxonomic order of all found genera. If these plasmids were found in more than one order, they were assigned as pBHR other. In addition, plasmid sequences were taxonomically classified with Kraken v2 (D. E. Wood and Salzberg, 2014), using the Genome Taxonomy Database (GTDB) (Parks et al., 2022), specifically, the previously separated databases for chromosomal and plasmid sequences (GTDBp and GTDBc) (Barrantes-Jiménez et al., 2025). Lastly, the taxonomic origin of the annotated UniRef90 proteins was used to assign plasmids to a putative host to the genus level. The highest number of ORFs belonging to a certain genus was used, while no host was assigned in the case of an equal count.

Potential bacterial host genus of all plasmids were assigned using the following hierarchy (Fig. S4A). First, plasmids classified using the GTDBp with the same genus of (one of the) contained plasmid(s) within the PLSDB were assigned the genera provided from PLSDB. Second, host genera of the remainder of plasmids that were contained within PLSDB. Third, classified genera using the GTDBc and GTDBp were equal. Fourth, classified genus using the GTDBp was equal to the result of UniRef90 annotation. Fifth, classified genus using the GTDBc was equal to the result of UniRef90 annotation. In addition, there was no classification of the GTDBp. Sixth, classification using the GTDBc. Seventh, classification using the annotation from UniRef90.

The pipeline was validated using the 9 genera hosting the highest plasmid diversity in our dataset (see results 4.2.). The same search terms as used by PLSDB (Galata et al., 2019) were used, together with our genus of interest, at <https://www.ncbi.nlm.nih.gov/nucleotide>. In addition, hits were filtered on a release date after the latest PLSDB update (1st of June 2024 to 30th of June 2025). The number of hits were uneven, with ~10.000 *Escherichia* and *Klebsiella*, 2–3.000 *Staphylococcus* and *Enterococcus*, and 160 or less found plasmids for the remaining genera. No *Collinsella* plasmids were found using the release date constraints, and only 20 plasmids when this constraint was lifted. As such, the 20 most recent plasmids using the date release constraints were selected, together with the only uploaded *Collinsella* plasmids.

A second method for assignment of taxa was done by computing the contemporaneous correlation between species and plasmid abundance time series (Trosvik et al., 2024), and was stored in addition to the previous database based taxonomic assignment. In short, for this analysis only plasmids and species observed at least 5 time points within a series were included, and candidate assignments were considered at a Pearson's R^2 of >0.95 . Read based species abundance was computed using MetaPhlan v3 (Beghini et al., 2021; Trosvik et al., 2024). In addition, we compared the observed number of $R^2 > 0.95$ correlations with a false positive rate, represented by random correlations to time series from other infants. If most of the observed correlations are artefactual then those values should be similar. The previously mentioned plasmid-taxon correlations within infants were taken as the observed value. The mean count of species that correlated with at least one plasmid per infant compared to all other infants was taken as the expected measure, i.e. the rate of false positives produced by chance correlation between unrelated time series. For comparisons between

infants, samples were sorted on days after birth and the last sample was removed in case of unequal number of samples (12/13 per infant). Resulting observed and expected values for all 12 infants were then compared using the Chi squared test using 11 degrees of freedom. Null hypothesis: the same number of species correlate with plasmid abundance within and across infants.

2.7. GC content

For the GC content of the species genomes, previously assembled high quality MAGs were used (unpublished, for quality information see (Trosvik et al., 2024)). Briefly, reads were assembled with MEGAHIT v1.2.9 (D. Li et al., 2015). Contigs were binned using CONCOCT v1.1.0 (Alneberg et al., 2014). Bins were inspected for completeness, contamination and strain heterogeneity using CheckM v1.1.3 (Parks et al., 2015). Taxonomic classification of MAGs was done with GTDB-Tk v1.7 (Chaumeil et al., 2019). GC content of the CheckM results were used for the MAGs of relevant taxa. GC content of the plasmid representatives per infant was calculated using 'letterFrequency' of the R library Biostrings (Pagès et al., 2024).

2.8. Phylogeny

Nucleotide sequences of plasmid representatives of plasmid clusters and systems across infants were aligned to start on the same protein sequence with a custom R script. Multiple sequence alignment of the linearized plasmids was done with kalign v3.3.5 (Lassmann, 2019). A phylogenetic tree was calculated with FastTree v2.1.11 and the options $-nt -gtr -gamma$ (Price et al., 2010). The trees were visualized with the R library ape (Paradis and Schliep, 2019).

2.9. Bacterial isolates

Fecal samples were stored at -80°C and thawed on ice before inoculation. A plastic inoculating loop was used to collect a small amount of fecal material to inoculate into liquid media. Used media were Bifidobacterium specific medium (BSM) Millipore 90,273, Lurial Broth plus ampicillin (LB + amp), Lurial Broth Miller (LB) Sigma-Aldrich L3522, and Anaerobic basal Broth (ABB), Thermo Fisher Scientific (CM0957). Under aerobic conditions, the cultures were incubated at 37°C with 200 rpm shaking. Anaerobic cultures were incubated at 37°C without shaking. All cultures were grown to an OD of 600 nm. To obtain pure cultures, 50 μl culture was spread onto selective media and incubated at 37°C . Selective media used; McConkey agar (Cm0507B), anaerobic basal agar (CM0972), Lurial Broth agar Sigma-Aldrich (L3027), and Lurial Broth agar plus ampicillin. One colony was sub-cultured into fresh liquid media using a sterile plastic inoculation loop and incubated as the previous liquid media. In the case of mixed colonies on the same plate, the ones that showed different colony morphology were inoculated in their individual flasks. After the liquid bacterial cultures reached an OD of 600 nm they were used for gDNA isolation.

DNA extraction was performed using the QIAamp PowerFecal Pro DNA Kit (Qiagen) following the manufacturer's instructions. Samples (65) were submitted to the Norwegian Sequencing Centre for library preparation using the Rapid Sequencing Kit V14 and Rapid Sequencing Adapters Kit 96 V14. Whole genome sequencing was performed on an Oxford Nanopore MinION Mk1C instrument using a Flowcell version R10.4.1 for 72 h. Dorado 0.5.0 was used for base calling with model dna_r10.4.1_e8.2_400bps_hac@v5.0.0. Sequences were assembled with Flye v2.9.2 and the $-\text{nano-raw}$ option (Kolmogorov et al., 2019). Quality of the assemblies were evaluated with Quast v5.2.0 (Mikheenko et al., 2018). Taxonomic assignment of genomes was done using GTDB-Tk v2.1.0 (Chaumeil et al., 2022). geNomad was used as previously described to identify plasmid and phage sequences. Found plasmids were clustered with MobMess with the infant's plasmid representatives to find matches.

3. Statistics

All statistical analysis were carried out with R (R Core Team, 2023). The Shannon index was calculated with the function ‘diversity’ of the vegan package (Oksanen et al., 2025). A binary matrix of absence/presence data of plasmid representatives per infant was used as input for the ‘estimateR’ function of the vegan package to calculate plasmid Chao richness. Comparing GC content between plasmids and chromosomes was done with the two sided ‘wilcox.test’. Benjamini–Hochberg false discovery rate correction was done with the ‘p.adjust’ function with method set to ‘BH’. In testing for correlation, we used Spearman’s Rho for examining general time trends, due to the robustness of this approach, while Pearson’s correlation was used for matching specific time series for reasons of sensitivity. GAM modelling was done with the ‘gam’ function in the mgcv package (S. N. Wood, 2011) using five degrees of freedom. The Pearson’s chi-squared test was calculated with the “observed” and “expected” correlations with 11 degrees of freedom. Presence/absence of all plasmids, i.e. plasmid community, were compared between infants using Jaccard similarity test ‘jaccard.test’ with the option “mca” to get centered Jaccard similarity coefficient from the R package Jaccard (Chung et al., 2019). Heatmaps were made using pheatmap (Kolde, 2018). Colors used in all figures are from base R or RColorBrewer (Neuwirth, 2022).

4. Results

4.1. Plasmid reconstruction from infant metagenomes

Monthly samples collected from 12 infants born to term in Oslo, Norway, were analyzed in this study (Trosvik et al., 2024). Assembly with metaplasmidSPAdes resulted in 3825 circular contigs of which 651 and 1862 were identified as viruses and plasmids by geNomad (Camargo et al., 2023). The plasmid contigs were clustered with MobMess (Yu et al., 2024), resulting in 880 plasmid representatives for the 12 infants, with an average of 73.3 per infant (s.d. 26.6) (Fig. S1). Similar plasmid representatives across infants were clustered in 620 plasmid clusters. Presence or absence was measured by mapping sample reads to the infants’ own plasmid representatives. To limit false positives, plasmids with at least 10× coverage over 80 % of the fragment’s length were considered present. 594 of the 620 unique plasmids were present in at least one sample, suggesting we lost a few low abundant plasmids due to our stringent criteria. Most plasmids were observed at a single time point (221 plasmids, Fig. S2, Table S1) and in only one infant (491 plasmids, Table S2), however one plasmid was found in 11 infants and 10 plasmids were found in at least 6 infants (Table 1, 4.5.). The siblings, and

Table 1
Widespread plasmids.

System	Plasmid(s)	# Infants	Taxonomy	Alternative name
PS26	604	11	Clostridium	
	450;451;452	9	Clostridium	
	608	9	pBHR_Bacteroidales; Bacteroides	
PS37	619	8	Haemophilus;	
	552;553	7	Cuspidothrix	PTU-E3
	426	7	pBHR_Enterobacterales	
	517	7	Clostridium	
	572	7	Eggerthella;Collinsella	
PS22	584	7	Clostridium	
	423;424;507;508	6	Bifidobacterium	PTU-Bif1
PS29	473;611	6	pBHR_other;	
PS33	528;529;530	6	pBHR_Enterobacterales	PTU-E9
		6	pBHR_Enterobacterales	pPIGDM1
		6	pBHR_Lactobacillales;	
	512	6	Enterococcus	
	609	6	pBHR_other	pBI143

especially the twins, had significantly more similar plasmid carriage to one another relative to the rest of the infants (Fig. S3). Between birth and one year of age we saw a significant overall increase in plasmid richness (Fig. 2), although we did not see a significant net time trend in total plasmid abundance (Table 2). Protein, origin of replication, and origin of transfer annotation of all plasmids can be found in Table S3.

4.2. Taxonomic assignment

Plasmids are not necessarily bound to a specific host species, for example, the cryptic plasmid pBI143 was recently found in multiple genera of the Bacteroidales order (Fogarty et al., 2024). In addition, assembled plasmids from metagenomic data are hard to assign to a specific host. Here we assigned putative plasmid hosts using the plasmid database (PLSDB) (Schmartz et al., 2022), the genome taxonomic database (GTDB) (Parks et al., 2022) (separated into plasmid and chromosomal sequences, GTDBp and GTDBc respectively (Barrantes-Jiménez et al., 2025)) and the UniProt Reference Clusters 90 database (UniRef90) (The UniProt Consortium, 2023) (see methods 2.6.), using a custom hierarchy (Fig. S4A). We classified plasmids contained within the PLSDB hosted by more than one genus as putative Broad Host Range (pBHR) plasmids. The pBHR plasmids were further grouped on their shared taxonomic Order or alternatively assigned as “pBHR other”.

Of the 880 plasmids, 253 were contained within the PLSDB whereof 122 were assigned to (one of) the same putative genus host using the GTDBp (Fig. S4B, Table S4). When a plasmid was found in more than one genus from the PLSDB, it was assigned as a putative Broad Host Range (pBHR) plasmid. In total 122 pBHR plasmids were found. pBHR Enterobacterales was the most diverse with 83 plasmids, whereof 43 were previously described as plasmid taxonomic units (PTUs) of the Enterobacterales order (Table S4) (Redondo-Salvo et al., 2020). Furthermore, 99 plasmids were assigned to a genus based on the GTDBp, 279 based on the GTDBc, and 137 based on UniRef90 protein annotation, leaving 112 plasmids unassigned. Plasmid diversity was decidedly highest for the putative plasmid host *Clostridium* with 180 plasmids (Table S5). The pBHR Enterobacterales was second highest with 83 plasmids, whereof 72 and 62 plasmids included the genera *Escherichia* and *Klebsiella*, respectively. Plasmids assigned to *Escherichia* and *Klebsiella* were broadly categorized as pBHR plasmids, whereas plasmids associated with *Clostridium* and *Streptococcus* were generally linked exclusively to those genera.

We validated our method by assigning putative plasmid hosts for plasmids originating from isolated bacteria. Plasmid DNA sequences from the 9 genera hosting the highest plasmid diversity in our dataset were downloaded from NCBI (see methods 2.6.). All 180 plasmids (20 per genus) were assigned to a putative host by our pipeline. Since most of the *Collinsella* plasmids were contained within the PLSDB, this part of the pipeline was not used for the *Collinsella* plasmids. 92 of the plasmids were contained within the PLSDB, while 22 remained within an unknown host (Fig. S4C, Table S6). Interestingly, the accuracy of assigned putative plasmid hosts depended on the genus and the database used (Fig. S5). For *Escherichia* and *Klebsiella* plasmids, the PLSDB was the most accurate with 17 and 20 correct assignments, respectively. While the two GTDB databases resulted in between 9 and 11 incorrect assignments for these two genera, they gave the best accuracy for *Bacteroides* plasmids with 1 incorrect assignment (*Parabacteroides*). *Bifidobacterium* plasmids were correctly assigned by GTDBc for 18 out of 20 plasmids, while there were no correct assignments using GTDBp. *Clostridium* was most accurately assigned using UniRef90 with 15 out of 20. Overall, our method using input from all databases resulted in the most accurate assignment of plasmid hosts. However, incorrect and no assignment occurred, including almost half of the *Clostridium* plasmids.

Previously, it was found that the GC content of bacterial genomes and plasmids are highly correlated, but the GC content of plasmids was often lower than their respective host chromosome (Nishida, 2012; Almpanis et al., 2018). In this infant cohort we observed the same trends

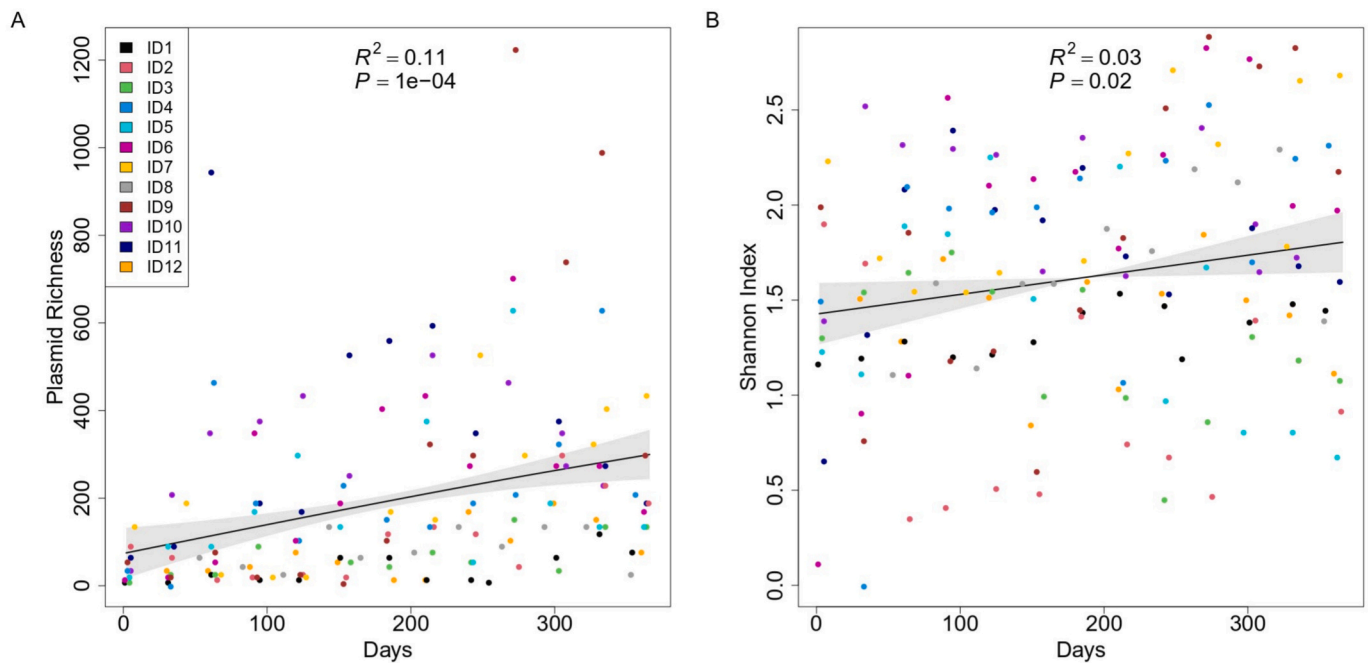


Fig. 2. Plasmid richness and diversity from birth to one year of age. A. Plasmid richness calculated per sample using the Chao estimate. B. Diversity of the plasmids using Shannon index. A/B. Calculations per sample were colored by infant ID and plotted against the days since birth. Richness and diversity had a significant linear ($\text{edf} = 1$) upward trend that was found with a generalized additive model (black line), with a 95 % confidence interval in grey.

with GC content of the plasmids lower than the GC content of the metagenome-assembled genomes (MAGs) except in the cases of *Bifidobacterium* and *Collinsella*, (Fig. 3A). Chromosomal and plasmid DNA sequences assigned to *Clostridium* had a relatively low GC content of around 30 %, whereas *Bifidobacterium* had the highest at around 60 %. The variation in plasmid GC content for *Bacteroides*, *Escherichia*, and *Klebsiella* was wider than observed for the other genera. Although mechanistically poorly understood (Mahajan and Agashe, 2022), this may suggest a higher diversity of donor and recipient plasmids in these taxa.

4.3. Mobility potential

We then assessed the conjugation potential of the plasmids. Plasmids with a full conjugation machinery were typed as CONJ. Plasmids that can conjugate with the help of a second plasmid, including plasmids only containing an origin of transfer, were classified as MOB. The remaining plasmids were then typed as MOBless. Similar to taxonomic assignment, some plasmid references (per infant) within the same plasmid cluster had dissimilar conjugation potential due to small sequence differences (Table S4).

Most plasmids were classified as MOB (662), followed by MOBless (136), with CONJ being the least common category (82). In addition, the MOB plasmids were the most abundant in the infants (Fig. S6). Interestingly, of the commonly found putative plasmid hosts, *Bacteroides* and pBHR Bacteroidales had the highest proportions of CONJ plasmids (Fig. 3B). All CONJ plasmids were > 30 kbp long, whereas MOB plasmids range from 1 to 70 kbp. As expected, the type 4 secretion system (T4SS) of the CONJ plasmids were type B in *Bacteroides* and type FATA in the Firmicutes *Clostridium* and *Enterococcus*, whereas the type T was found in *Clostridium* and *Escherichia* (Table S4) (Guglielmini et al., 2013).

4.4. Plasmids associated with *Bacteroides* are highly abundant and increase over time

Both in terms of plasmid abundance and bacterial host composition,

there was substantial variation among the infants. There was also considerable variation over time within each infant (Figs. 4, S7). Even though our assignment of putative plasmid hosts was based on several methods, the abundance profiles were similar for *Bacteroides* plasmids based on the separate databases (Figs. S8–11). A noteworthy exception was plasmid cluster 608, assigned to *Enterococcus* by UniRef90 and highly abundant in ID2. Of the other most common genera, *Clostridium* and *Collinsella* assignment was mostly based on the GTDBp/c and UniRef90 databases. Even though the highest plasmid diversity was found in *Clostridium* when considering the full taxonomic assignment, the plasmids with the highest overall numerical abundance were associated with *Bacteroides* and pBHR Bacteroidales (and pBHR Other which was mostly driven by pBI143 see 4.5; Fig. 4, Tables S4 and S8). However, ID6 and ID8 had no *Bacteroides*-associated plasmids at all. ID6 was previously found to have extremely low levels of *Bacteroidota* (de Muinck and Trosvik, 2018), and in this infant we found a highly abundant plasmid on day 2, with the putative host *Sphingobium*, and otherwise relatively low plasmid abundances. *Bacteroides*-associated plasmids were also rare in ID7, an infant with appreciable *Bacteroides* abundances, demonstrating the high degree of inter-individual variability. Plasmids assigned to *Bifidobacterium* in ID7 were highly abundant during the first two months, and later on substantial amounts of plasmids associated with *Phocaeicola*, a genus with species previously classified as *Bacteroides*, were found.

ID8 had several plasmids of unknown origin throughout the first year of life. There was also a general trend for early samples to be dominated by plasmids associated with taxa other than *Bacteroides*, e.g. *Escherichia* and pBHR Enterobacterales plasmids were highly abundant during the first 6 months in ID3. Abundances of plasmids associated with *Bacteroides*, *Clostridium*, *Collinsella*, and pBHR Bacteroidales increased significantly over time, whereas *Enterococcus*, *Staphylococcus*, and *Streptococcus*-associated plasmids decreased (Table 1). *Clostridium* plasmids showed sporadically high abundances in the twins, ID10 and ID11, and was also observed in several samples from ID2, 6, 8, 9 and 12.

Table 2

Time correlations of the most found plasmid genera (>16 plasmids). Spearman correlations were made between the abundance (RPKM) of plasmid hosts and sampling day (day after birth). Total plasmid abundance did not change significantly throughout the study ($p > 0.05$ for Total). *Bacteroides*, *Clostridium*, *Collinsella*, and pBHR Bacteroidales had significant increases in abundance from birth to one year, whereas *Enterococcus*, *Staphylococcus*, and *Streptococcus* had a significant decrease. The plasmid count (#) was per infant, potentially counting the same plasmid multiple times. In the last two columns, spearman correlations of host taxa from MetaPhlAn and sampling day were added. Taxa with significant time trends, *Bacteroides*, Bacteroidales, *Collinsella*, Enterobacterales, *Klebsiella*, *Staphylococcus* and *Streptococcus*, had similar Rhos with the abundance of their plasmids. P-values were adjusted with the Benjamini-Hochberg method to correct for multiple hypothesis testing.

Putative genus	Plasmids			Metaphlan	
	#	Rho	P	Rho	P
Total	880	7.44E-02	3.98E-01	–	–
Bacteroides	17	3.69E-01	8.99E-04	3.14E-01	4.71E-04
Bifidobacterium	17	1.63E-01	1.15E-01	–1.66E-01	5.69E-02
Clostridium	188	2.37E-01	9.58E-03	8.57E-02	3.26E-01
Collinsella	19	2.19E-01	3.80E-02	2.19E-01	3.07E-02
Enterococcus	18	–3.29E-01	8.99E-04	–1.56E-01	6.97E-02
Escherichia	25	–2.97E-02	7.69E-01	–7.05E-02	3.93E-01
Klebsiella	33	–1.74E-01	1.25E-01	–2.11E-01	1.66E-02
Staphylococcus	24	–5.64E-01	1.63E-07	–7.09E-01	5.12E–23
Streptococcus	60	–2.97E-01	2.61E-03	–2.42E-01	5.91E-03
pBHR Bacteroidales	17	2.67E-01	1.55E-02	3.90E-01	3.90E-01
pBHR Enterobacterales	83	–9.34E-02	3.30E-01	–4.42E-01	1.02E-07
pBHR Lactobacillales	8	–1.36E-01	3.24E-01	–3.75E-01	8.05E-06

4.5. Identity of plasmids found in multiple infants

Plasmids are known to acquire and lose genes but generally conserve a backbone of plasmid maintenance genes. MobMess assigns plasmids that are similar along their full length into clusters, while plasmids with the same backbone but different additional genes are put in plasmid systems (PS), see methods 2.2. In the full catalog of plasmid clusters and systems in our dataset, there were 14 plasmid clusters and systems that we designated as widespread (present in at least 50 % of the infants; Table 1). One of these was the previously characterized cryptic plasmid, pBI143 that is widespread in the human gut microbiome in industrialized populations (Fogarty et al., 2024), and was found in 6 of the 12 infants, including the twins and the sibling pair (Tables 1, S4, Fig. S12). Fogarty et al. described three variants of pBI143 of 2751, 2784 and 2782 bp carrying a MOB and REP protein. We found the plasmid cluster 609 with a variant of 2750–2754 bp and a variant of 2783 bp, both with >99.8 % sequence identity to pBI143. Interestingly, when investigating all plasmid contigs before clustering, the 2783 bp version was found in both the siblings (ID4 and ID5), whereas the shortest version was only found in the youngest (ID4). All found contigs of plasmid 609 were aligned with the three versions of pBI143, and phylogenetic analysis confirmed the close relation of the 2750–2754 bp version with the first version of pBI143, and the 2783 bp version with the second (Fig. S10). Like Fogarty et al. (2024), we found that pBI143 was highly abundant, representing 18.8 % of sequence reads mapped to plasmids in the entire dataset. Furthermore, we found a highly significant overall increase in pBI143 abundance over time for infants carrying this plasmid

(Spearman's rho = 0.40, $p < 0.001$).

The pBHR Enterobacterales PS33 was found in 6 of the infants, with a 2.5 kbp backbone identical to that of plasmid pPIGDM1, originally isolated from *Enterobacter agglomerans* (Mikiewicz et al., 1997). The PS33 variants had many similar plasmids found within the PLSDB, most originating from *Enterobacter hormaechei*, followed by *Enterobacter cloacae* (Table S4). Other genera found through the PLSDB included *Klebsiella* and *Salmonella*, suggesting that PS33 can transfer between genera. ID7 carried not only the 2.5 kbp backbone, but an additional variant of 8.8 kbp carrying additional plasmid maintenance genes (Table S3, Fig. S13). PS33 was only found as a relatively large 16.9 kbp plasmid in ID8, though the additional genes are without a clear function beside plasmid mobility and transposition. The five 2.5 kbp variants clustered in two groups, with the twins carrying the same variant and clustering with pPIGDM1 (Fig. S9).

Interestingly, 5 of the 14 widespread plasmids were novel and predicted to be hosted by the genus *Clostridium* based on their nucleotide and protein sequences. These included PS26 found in 9 infants and connected to PS6 and PS24 in ID6 and 7 (Figs. S14–15). Surprisingly, no known functionality beside plasmid transfer and maintenance was found for this ubiquitous plasmid. Another cryptic plasmid of 4.6 kbp linked to *Clostridium* was found in 11 of the 12 infants, this ubiquitous plasmid was >98.9 % identical over the full length in all infants. The plasmid 517, linked to *Clostridium*, was >97 % identical, carried in 7 infants and encoded a glycopeptide antibiotic resistance protein.

4.6. Contemporaneous plasmid-taxon correlations

We have previously shown that antibiotic and virulence genes can be linked with specific taxa by correlating temporal gene abundance patterns with taxon dynamics (Trosvik et al., 2024). This approach leverages the strengths of a time series sampling design. In the case of plasmids, this premise is not a given due to increased chances of HGT, multiple species harboring the same plasmid, or fluctuating plasmid copy numbers. Although we know very little about how plasmids behave in natural environments, laboratory studies have shown that they can have large variations in copy number within a population (Carapuca et al., 2007), which can be tuned for protein production (Rouches et al., 2022). Although poorly understood, the temporal dynamics of a particular plasmid may depend on a number of factors like population density, environmental cues, or the dynamics of other plasmids in the population (Wein and Dagan, 2020). In addition, HGT events may produce compound plasmid dynamics resulting from the superimposition of fluctuation patterns in two or more populations. Fig. 5A,B illustrates some theoretical scenarios of plasmid-taxon contemporaneous correlation.

To get a sense of the probability that species and plasmid abundances may correlate by chance, we tested if the correlation was higher between species and plasmids within infants than between infants (Chi squared test, $p < 0.001$, Table 3). However, since our expected values were higher than zero, some of our results will include false positives. Looking at all plasmid-taxon pairs within each infant, a total of 82 plasmids correlated with a species (Table S4). 75 of these plasmids had been assigned to a putative host. Of the already assigned plasmids, 35 corresponded with the previously assigned taxa on the genus level (we considered the genera *Bacteroides* and *Phocaeicola* as one genus). An example of a taxon that did not match was cluster 240, assigned to *Enterobacter* in ID5, but correlated with *Klebsiella* abundances (Fig. 5C). In another example, three plasmids were highly correlated with *Collinsella aerofaciens* (Fig. 5D), but at different relative abundances, possibly indicating different plasmid copy numbers. Three plasmids (599, 608, and 609) in ID1 correlated with *Bacteroides thetaiotaomicron*, while plasmid 608 was correlated with *Bacteroides dorei* (*B. dorei*) in ID2. These three clusters were linked with several *Bacteroides* species in PLSDB, suggesting we found realistic hosts for these genetic elements.

The twins were very similar in their plasmid composition (Fig. S3)

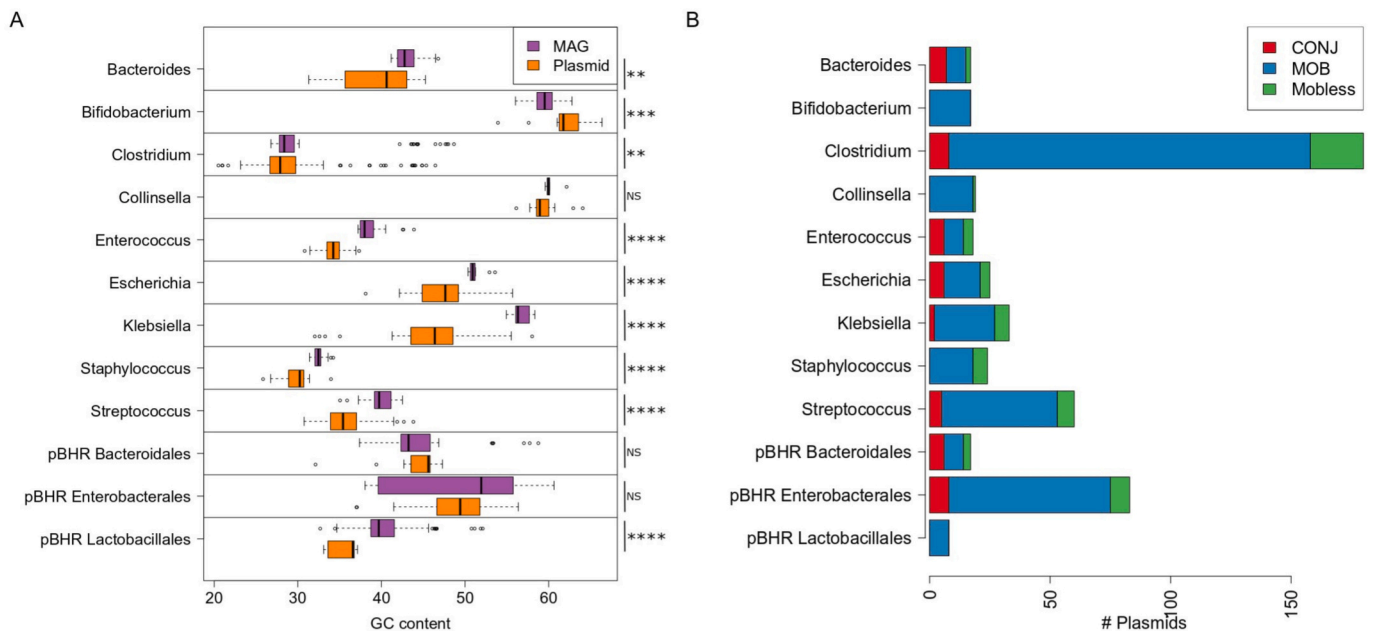


Fig. 3. GC content and mobility of plasmids from the most prevalent genera. A. GC content of plasmids and MAGs of the nine most prevalent genera and the three pBHR categories with a specified order. Apart from *Collinsella*, pBHR Bacteroidales and pBHR Enterobacterales, GC content of the plasmids is significantly different from MAG GC content. General GC content per genera is varying, with *Clostridium* on the lower end, and *Bifidobacterium* at the highest end. *P*-values have been adjusted with the Benjamini-Hochberg method to correct for multiple hypothesis testing. Symbols for *P*-values; ns > 0.05, * ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001, **** ≤ 0.0001. B. Prediction of the conjugation ability of plasmids from the same categories as 3 A. Plasmids with full conjugation machinery in red (CONJ), decayed or incomplete conjugation machinery in blue (MOB) and plasmids that can't conjugate in green (MOBless). Plasmids were counted per infant, counting ubiquitous plasmids several times. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and illustrate an interesting case of plasmid dynamics. *B. dorei* was observed in ID11 on day 6 (Fig. 5F), but not until day 158 in ID10 (Fig. 5E). In both twins, plasmid 536 correlated strongly with *B. dorei*, while the plasmids 560 and 561 only correlated strongly in ID10. In ID11, plasmids 560 and 561 followed the abundance of *B. dorei* very closely until a peak at day 186, after which the pattern started to break down, and on day 304 these plasmids' abundances showed an opposite trend from *B. dorei*. For ID10, all three plasmids tracked *B. dorei* closely throughout the time series. Together, it seems likely that the twins were colonized at different timepoints by the same *B. dorei* carrying the same plasmids.

4.7. Plasmid assembly and host confirmation using culturing isolates

To corroborate our results, we cultured, sequenced, and assembled the complete genomes of ten isolates, two *Enterobacter hormaechei*, seven *Enterococcus faecalis*, and one *Klebsiella michiganensis* (Tables S8–10). These were selected from a total of 45 isolates based on a contamination of less than 10 %, assembly of maximal 20 contigs, carrying at least one plasmid, and finally, carrying at least one plasmid in the same plasmid cluster or system as one of our previous assembled plasmids (21, 2, 4, 8 removed isolates, respectively) (Table 4). The 4 isolates without any plasmids were from the species *Bifidobacterium bifidum*, *Citrobacter freundii*, *Enterococcus B durans*, and *Escherichia coli*. Isolated species carrying no overlapping plasmids were *Citrobacter freundii* (1×), *Escherichia coli* (4×), *Enterococcus faecalis* (1×), *Staphylococcus aureus* (1×), and *Staphylococcus warneri* (1×).

The 2.6 kbp pBHR Enterobacterales plasmid 528 (of PS33) was found in both *Enterobacter hormaechei* isolates, albeit with different lengths (Table 4). In addition, the plasmid correlated with *Enterobacter cloacae complex* in ID10, a species group including *Enterobacter hormaechei*. Other potential plasmids that we did not find within this species isolated from ID10 were, three pBHR Enterobacterales plasmids and one *Enterobacter* plasmid.

All *Enterococcus faecalis* isolated (from IDs 4, 9, 10 and 11) carried

the ~71 kbp *Enterococcus* plasmid 538 identified in the twins. Interestingly, plasmid 528 found in both *Enterobacter hormaechei* isolates (ID10) was also found in one *Enterococcus faecalis*, isolated from a sample from ID11 taken on day 36. The plasmid was present as a high copy number plasmid in *Enterobacter hormaechei* (24 and 32 times), though only 4 copies were present in *Enterococcus faecalis*. While the plasmid 528 carried by both *Enterobacter hormaechei* and one *Enterococcus faecalis* isolates were very similar to the assembled versions in ID10 and ID11 (Fig. S12B), when we compared the abundance profiles of these species and plasmids in both twins, we could not discern a HGT event in vivo (Fig. S16).

Lastly, *Klebsiella michiganensis* isolated from ID11 was found to have 3 large plasmids of 36, 100 and 161 kbp respectively. The large 161 kbp plasmid, potentially carrying three ARGs according to geNomad (Table S10), was previously assembled in ID10 as plasmid 564 (161 kbp, pBHR Enterobacterales).

5. Discussion

5.1. Plasmid reconstruction

A major motivation for studying MGEs is their potential for facilitating the spread of ARGs to pathogens. However, plasmid dynamics and ecology in natural ecosystems remains understudied and poorly understood. Recent research, based on mapping of metagenomic sequencing reads to databases, has provided important insights into the content MGEs and ARGs in the infant microbiome (Bargheet et al., 2025; Trosvik et al., 2024; Pärnänen et al., 2018), supported by assembly-based approaches (Bargheet et al., 2025; Trosvik et al., 2024; Vatanen et al., 2022). Assembly-based techniques are needed for analyzing the potential mobility of ARGs, as physical linkage with an MGE is generally required for HGT, but these methods have their limitations when using short-read sequencing, as ARGs are often surrounded by a conserved repetitive genetic background from different species, resulting in fragmentation of contigs during assembly (Abramova et al., 2024). This

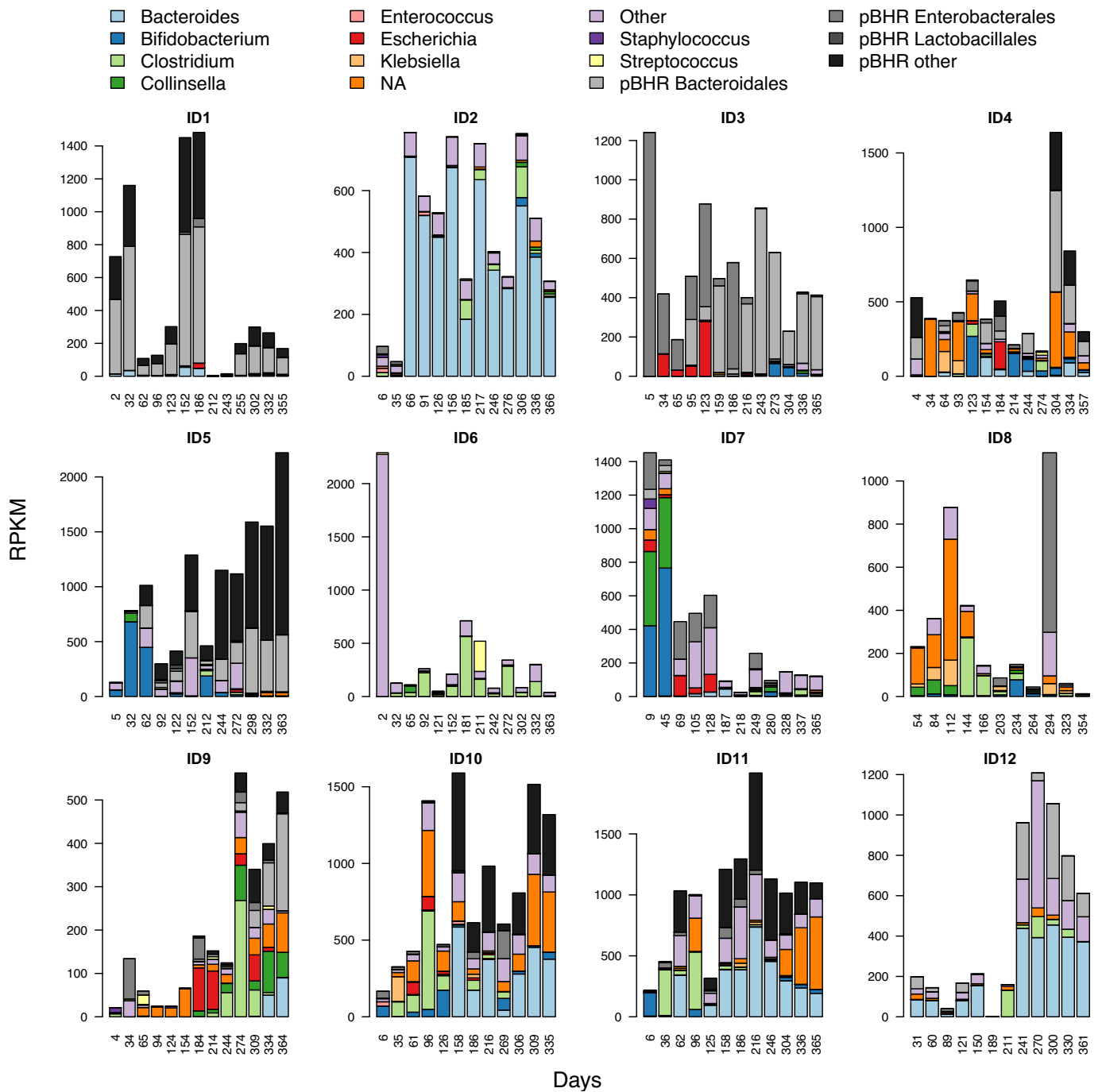


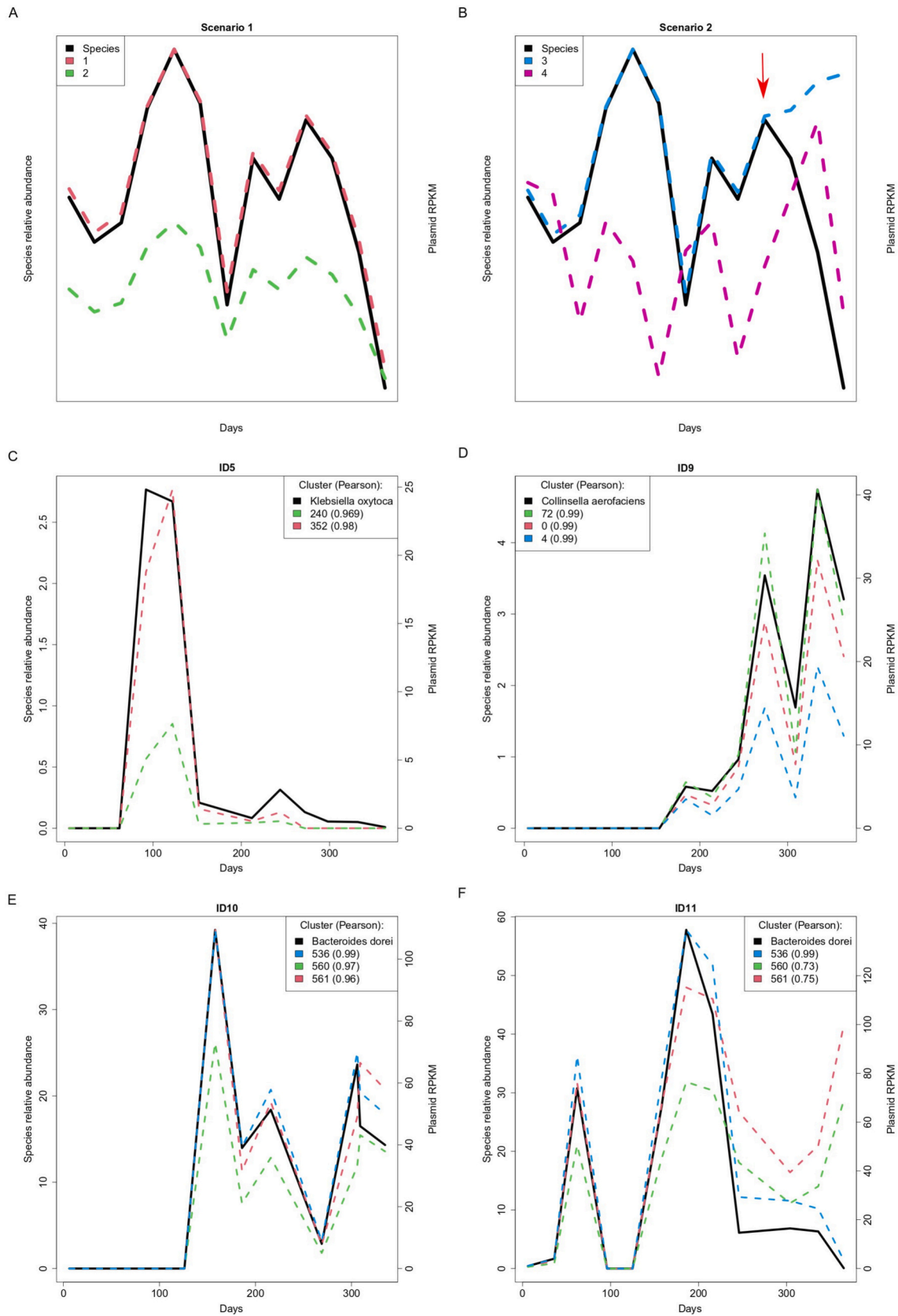
Fig. 4. Plasmid abundance with different genera hosts per sample. One bar plot per infant, with bars for all samples ordered from birth (left) to one year (right). The top nine most abundant genera are colour coded, the putative Broad Host Range (pBHR) plasmids are shown in greys, other genera are gathered under “Other” (light purple), with “NA” (orange) for plasmids without known hosts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

potentially prevents successful assembly of ARG-carrying plasmids (Paganini et al., 2021) or collapse of MGE or ARG variants into one (Kerkvliet et al., 2024).

Here, we used metaplasmidSPAdes to assemble circular contigs due to its relative accuracy in assembling the genetic context around ARGs and MGEs compared to other tools (Abramova et al., 2024). However, the use of short-read metagenomic sequencing has likely resulted in a bias towards smaller plasmids in our dataset (Yorki et al., 2023), limiting the finding of conjugative plasmids. Another limitation of our method’s design is the potential presence of (pro)phages. The tool geNomad is designed to identify both plasmid and phage sequences (Camargo et al.,

2023). We have decided not to remove contigs with viral content that were identified as plasmid, since phage-plasmids were found to play a potential role in HGT of ARGs (Pfeifer et al., 2022). Longitudinal sampling can help mitigate some of the difficulties with short-read assembly for plasmid detection by increasing the chance for reappearing plasmids to be detected in samples where assembly was not successful. Furthermore, assembly of the same plasmid in several samples helps corroborate the authenticity of the plasmid sequence.

We recently conducted a study on ARG dynamics in the infant gut microbiome, using the same data set as the one analyzed here (Trosvik et al., 2024). That study also included an analysis of linkage between



(caption on next page)

Fig. 5. Correlation of plasmid abundance with species abundance. Relative abundance of the species from MetaPhlan v3 and RPKM of the plasmids for each infant were correlated using the Pearson method. In general, both plasmid and species were present in at least 5 timepoints and only correlations with a > 0.95 score were kept, however, lower score examples are illustrated here. A/B Theoretical correlation scenarios. A. Plasmid 1 and 2 correlate strongly with a species abundance, where plasmid 2 has half the abundance of plasmid 1. This suggests that plasmids 1 and 2 are carried by this species with different copy numbers. B. Plasmid 3 correlates strongly to the species until a certain timepoint, which suggests 1) accidental correlation, 2) plasmid copy number is changed due to (environmental) circumstances, or 3) the plasmid has spread to a new host and now follows more than one species abundance. Plasmid 4 does not correlate with the species abundance, which could indicate that the plasmid is not hosted by this species, the copy number is not related to the species abundance, or the plasmid is carried by several species. C. Correlation of plasmids 240 and 352 with *Klebsiella oxytoca* in ID5. These plasmids were assigned the putative hosts of *Enterobacter* and pBHR *Enterobacteriales* respectively. D. Three plasmids correlating with *Collinsella aerofaciens* in ID9. The plasmids are present in the same host with different, but stable, copy numbers. E/F. Plasmids (not) correlating with *Bacteroides dorei* in the twins (ID10, ID11). E. The abundance of plasmids 536, 560, and 561 correlate with the abundance of *Bacteroides dorei*. F. plasmid 536 also correlated to *Bacteroides dorei* in ID11, however, the other two plasmids did not.

Table 3

Pearson's chi-squared test on correlating species with at least one plasmid within and across infants. Contemporaneous correlations between species and plasmid abundance were made when both were present in at least 5 timepoints. "Observed" values are the number of species correlating with at least one plasmid within the same infant, reporting the results per infant (rows). "Expected" values were calculated by making the same correlations but now with plasmids from other infants (false positive correlations) and averaging over all other infants. Null hypothesis: the same number of species correlate with plasmid abundance within and across infants. Pearson's chi-squared test of 299 and 11 degrees of freedom resulted in $p < 0.001$.

ID	# Correlations	# Not correlated	Observed	Expected	(o-e) ²	(o-e) ² / e
1	2	273	2	0.82	1.40	1.71
2	11	413	11	0.73	105.53	145.10
3	2	475	2	0.64	1.86	2.92
4	7	805	7	1.36	31.77	23.30
5	6	534	6	3.18	7.94	2.50
6	3	1173	3	1.27	2.98	2.34
7	7	533	7	1.45	30.75	21.14
8	6	253	6	3.45	6.48	1.88
9	8	1108	8	1.64	40.50	24.75
10	10	746	10	5.18	23.21	4.48
11	15	953	15	3.00	144.00	48.00
12	7	521	7	1.45	30.75	21.14
Chi squared						299.26

ARGs and MGEs, reporting 557 instances of relatively close physical linkage between ARGs and putative plasmid sequence markers, in contrast to only 13 plasmids carrying ARGs identified in the present study. The focus of each of these studies are widely different. The previous study attempted to limit false negatives by solely identifying ARG and MGE markers co-localized on linear contigs. The present study is focused on limiting false positives through de-novo assembly of complete circular plasmids. Furthermore, while the previous study only required a certain percentage of mapping coverage in order for an MGE marker to be accepted, the present study used more stringent filtering criteria.

5.2. Taxonomic assignment of putative plasmid hosts

Often, it is of pivotal importance to know which species carries a specific plasmid, especially plasmids encoding one or more ARGs. However, there are no methods for assigning a plasmid to a certain host in short-read shotgun metagenome studies. To this end data may be supplemented by culturing and sequencing of isolates or by DNA linking techniques like Hi-C (Yaffe and Relman, 2020). Here, we have expanded current methodology for predicting the most likely plasmid host by novel use of common tools. The PLSDB is a reasonable source for plasmid host assignment, since the plasmids were identified in cultured isolates (Schmartz et al., 2022). In addition, plasmids that were found in several genera within the PLSDB were likely mobile and were assigned as putative Broad Host Range (pBHR) plasmids. Approximately half of our pBHR *Enterobacteriales* plasmids were previously described as plasmid clusters carried by species within the *Enterobacteriales* order (Redondo-Salvo et al., 2020). However, the PLSDB will result in a strong bias towards species that are relatively easy to culture and are often cultured for diagnostic or research reasons. We confirmed that plasmids carried by *Bifidobacterium*, *Clostridium*, and *Streptococcus* were difficult to assign using the PLSDB. To circumvent this problem, we have added taxonomic assignment of the plasmids based on their nucleotide

sequence and proteins. Since these assignments were not based on plasmids from isolated bacteria, we interpreted conflicting genera as an uncertainty instead of assigning the plasmid as pBHR. Together, these tools gave strong evidence for assigning putative plasmids host genera in many of our plasmids. In case of putative plasmid host assignment with limited evidence, for example many of the *Clostridium* plasmids, instead of viewing these solely as unreliable, we think these can be used as leads for further investigation by research groups that are interested in this genus or other understudied species. In addition, we found that our method identified around half of the NCBI *Clostridium* plasmids, suggesting that this genus might have an even higher diversity of plasmids in infants than found in this study. By combining the high-throughput method of short-read metagenomic sequencing and assembly with the best available host assignment methods, we can uncover new plasmids and putative plasmid hosts. This paves the road for expanding our knowledge of plasmid ecology and dynamics, although such findings should always be corroborated with additional (experimental) methods.

5.3. Plasmid composition of the human infant gut microbiome

The first thorough investigation of the infant gut plasmidome was published recently (He et al., 2024), identifying the phylum *Bacteroidota* as the main plasmid hosts. Our investigation corroborates this result in terms of total plasmid abundance, while the greatest diversity was seen in the *Clostridium* genus (*Firmicutes* phylum). Plasmid abundance associated with the genera *Bacteroides* and *Clostridium* generally increased over time from birth to one year of age. Conversely, plasmids associated with *Enterococcus*, *Klebsiella*, and *Streptococcus* decreased in abundance. These results are in general agreement with population trends over the observed period, with strictly anaerobic species increasing over time, while high levels of facultative anaerobes are more closely associated with the earliest life stages. About a third of *Bacteroides* plasmid abundance was due to the cryptic plasmid pBI143, previously found to be abundant in the adult gut microbiome (Fogarty et al., 2024),

Table 4

Shared plasmids between cultured isolates and assembled plasmid through metagenomic sequencing. All found plasmids carried by cultured isolates from infant samples. Most of these plasmids clustered together with the metaplasmidSPAdes assembled plasmids from short-read shotgun metagenomic sequences. Average nucleotide identity (ANI) was taken from the full-length metagenome assembled plasmid compared to the cultured plasmid given by MobMess (I_global). Marked in RED when the plasmids did not originate from the same infant.

Barcode	Cultured				Metagenome				
	ID	Taxonomy	contig	length	Cluster	System	ID	length	ANI
18	10	Enterobacter hormaechei	1	4058	528	PS33	10	2622	0.987
25	4	Enterococcus faecalis	2	70855	538		10	70985	0.988
28	11	Enterococcus faecalis	2	2479	528	PS33	11	2622	0.977
28	11	Enterococcus faecalis	3	70856	538		11	70985	0.988
36	10	Enterobacter hormaechei	1	3924	528	PS33	10	2622	0.987
39	11	Enterococcus faecalis	2	70857	538		11	70985	0.988
42	10	Enterococcus faecalis	2	70854	538		10	70985	0.988
49	11	Enterococcus faecalis	2	70858	538		11	70985	0.988
50	11	Klebsiella michiganensis	2	160674	564		10	160805	0.988
50	11	Klebsiella michiganensis	3	99532					
50	11	Klebsiella michiganensis	20	35956					
55	9	Enterococcus faecalis	2	70853	538		10	70985	0.988
59	11	Enterococcus faecalis	3	70857	538		11	70985	0.988

demonstrating a wider dissemination of this particular plasmid. In general, we found that plasmid diversity increased towards the first birthday, which is in accordance with previous reports on infant MGE and plasmid diversity (Pärnänen et al., 2018; He et al., 2024; Bargheet et al., 2025). Together, the available data suggest that at some point in the child's development, after the first birthday, MGE and plasmid diversity reaches a high point, after which it decreases to an adult level.

The twins' plasmidomes were very similar to one another compared to the other infants of this study, while the siblings' plasmidomes resembled one another more than those of unrelated infants, but to a lesser extent than the twins. The same was found on a taxonomic level for the twins, but not the siblings (de Muinck and Trosvik, 2018). Furthermore, one of the twins underwent extensive antibiotic treatment while the other did not, which might have been expected to promote some degree of divergence of their microbiomes. The twins were dizygotic and thus no more closely related than the siblings, indicating that a closely shared environment might be important in shaping the plasmidome in the infant gut. While this is only anecdotal evidence, the extent to which interactions with siblings, and other close relatives, influence the development of the plasmidome is a venue of future research.

5.4. Widespread plasmids

Two highly abundant and widespread plasmids were plasmid 609 (the cryptic plasmid pBI143) found in 6 of the 12 infants, and plasmid 608, found in 9 of the 12 infants. We found several other widespread plasmids in our infants, like the pBHR Enterobacteriales PS33 found in 7 of the infants. PS33 was generally assembled as a short backbone

plasmid of 2.6 kbp but carried additional genes in ID7 (8.8 kbp) and ID8 (16.9 kbp). The same plasmid backbone sequence was carried by three isolates of the species *Enterobacter hormaechei* and *Enterococcus faecalis*. The PS33 backbone sequence was also described in the literature as pPIGDM1 originally isolated from *Enterobacter agglomerans*, and found to carry the transposon Tn1331 in other isolates (Alavi et al., 2011). The fact that these plasmids were found with similar or identical backbone sequences in several infants and/or timepoints, lends credence to the authenticity of these plasmid sequences.

Of the 14 widespread plasmid clusters and systems, 5 novel sequences were found to have *Clostridium* as the putative host based on the presence of specific gene variants. The widespread plasmid 517 showed strong contemporaneous correlation with *Clostridium paraputrificum* in ID10 and ID11, strengthening the case for assignment to the genus. The fact that these widespread plasmids were not represented in PLSDDB might reflect the limited research done on *Firmicutes* plasmids in general (Lanza et al., 2015), even though these plasmids might be involved in HGT of ARGs in *Clostridium*, like the transfer of an Erm-carrying plasmid from *Clostridium ljungdahlii* to a wild-type *Clostridium acetobutylicum* (Charubin et al., 2024). Here, we also found a small *Clostridium* plasmid carrying plasmid maintenance genes and a predicted glycopeptide ARG (VanZ family protein). A different ortholog of the VanZ has shown light resistance to teicoplanin in *Clostridium difficile* (Woods et al., 2018). However, the functionality of our found gene should be experimentally confirmed.

5.5. Contemporaneous correlations between plasmid and species abundances

Using contemporaneous correlations, we were able to make putative taxonomic assignments of 75 plasmids, with nearly half of the assignments matching our putative plasmid host assignment on the genus level. These results enabled us to identify strongly linked plasmid-taxon pairs. While we used cutoffs to exclude plasmids and species that were observed transiently within an infant, some correlations were still based on one large abundance peak and a few low abundance observations. In such cases there is a heightened probability that the correlation is spurious, or that plasmid dynamics will correlate with the dynamics of several species just by chance, leading to false positives or ambiguous assignments. In addition, even though we showed significantly higher plasmid-taxon correlations within infants compared to between infants, the correlations can be the result of similar read depth ratios of the samples, correlating species dynamics, or other reasons. Furthermore, random correlations were also found across infants, i.e. false positives. As such, we do not suggest using this method as a definitive method of taxonomic assignment, but rather as a tool to form hypotheses and to strengthen previously assigned putative plasmid hosts. It is important to keep in mind that bioinformatically linking a plasmid to their host is a very difficult task. We are of the opinion that any means that we may have to generate workable hypotheses can be of potential value. We recognize the weaknesses and limitations to this approach but given that certain criteria are met (e.g. relatively stable copy numbers, enough non-zero measurement points in time series, absence of HGT events during observation period) this approach could still be useful. In addition, the method is database independent and allows for de novo hypotheses.

We would again point out that the correlation approach can only be expected to be effective if average plasmid copy number per cell remain relatively constant over time in a given host population. If copy numbers are highly variable due to some more complex regulation mechanism (Rouches et al., 2022; Rodríguez-Beltrán et al., 2021) the signal will be obscured, and this scenario is likely for many plasmids in relatively dynamic natural environments. In the case of multiple host populations similar difficulties may arise, although correlation with higher order taxa may be preserved in such cases. In the hypothetical case of a cross-species HGT event during the observation period we might see a

departure from correlation following transfer. We identify a candidate for such an event in ID11. While this is highly speculative, and based on little data, the result highlights the potential usefulness of a longitudinal sampling regime and represents a novel method of HGT detection in metagenomic data. Inferring HGT can be extremely challenging in complex bacterial communities, and especially in longer time series of communities with relatively stable membership, ecological co-dynamics could reveal potential transfer events that can be investigated further, e.g. by single cell techniques.

5.6. Conclusion

Here, we have presented what is, to our knowledge, the most comprehensive longitudinal study of the infant plasmidome to date. While our analytical pipeline is relatively conservative, we present an extensive catalog of circular assemblies bearing the hallmarks of plasmid sequences, including several novel and widespread plasmids. We present time trends and putative taxonomic assignments of host bacteria, with results highlighting the need for more research into underrepresented host taxa like *Clostridium*. Finally, we outline how a longitudinal sampling scheme can be used to identify putative plasmid hosts, as well as a novel theory for detection of recent HGT events.

Author contribution

All authors conceptualized the project. H.T.N analyzed the data and wrote the paper with input from all other authors. P.P. cultured the isolates. All authors have read and agreed to the published version of the manuscript.

CRedit authorship contribution statement

Hanna Theodora Noordzij: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Meike T. Wortel:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Anna Heintz-Buschart:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Patricija Petrikonyte:** Methodology, Investigation. **Eric J. de Muinck:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Pål Trosvik:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Funding

This study was funded by the Research Council of Norway grant 300948.

Declaration of competing interest

The authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.plasmid.2025.102761>.

Data availability

The DNA sequence data of the isolates has been uploaded to the NCBI SRA database as BioProject PRJNA1283097.

Fasta file of all plasmids and the supplementary tables are made available at Doi: <https://doi.org/10.5281/zenodo.14045300>.

R and shell scripts are made available at https://github.com/HThNoordzij/Plasmid2025_scripts.

References

- Abramova, Anna, Karkman, Antti, Bengtsson-Palme, Johan, 2024. Metagenomic assemblies tend to break around antibiotic resistance genes. *BMC Genomics* 25 (1), 959. <https://doi.org/10.1186/s12864-024-10876-0>.
- Alavi, Mohammad R., Antonic, Vlado, Ravizee, Adrien, Weina, Peter J., Izadjoo, Mina, Stojadinovic, Alexander, 2011. An enterobacter plasmid as a new genetic background for the transposon Tn1331. *Infect. Drug Resist.* 4 (November), 209–213. <https://doi.org/10.2147/IDR.S25408>.
- Alcock, Brian P., Raphenya, Amogelang R., Lau, Tammy T.Y., et al., 2020. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48 (D1), D517–D525. <https://doi.org/10.1093/nar/gkz935>.
- Almpanis, Apostolos, Swain, Martin, Gatherer, Derek, McEwan, Neil, 2018. Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. *Microb. Genom.* 4 (4), e000168. <https://doi.org/10.1099/mgen.0.000168>.
- Alneberg, Johannes, Bjarnason, Brynjar Smári, de Bruijn, Ino, et al., 2014. Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11 (11), 1144–1146. <https://doi.org/10.1038/nmeth.3103>.
- Antipov, Dmitry, Raiko, Mikhail, Lapidus, Alla, Pevzner, Pavel A., 2020. Metaviral SPAdes: assembly of viruses from metagenomic data. *Bioinf. (Oxf)* 36 (14), 4126–4129. <https://doi.org/10.1093/bioinformatics/btaa490>.
- Ares-Arroyo, Manuel, Coluzzi, Charles, Rocha, Eduardo P.C., 2023. Origins of transfer establish networks of functional dependencies for plasmid transfer by conjugation. *Nucleic Acids Res.* 51 (7), 3001–3016. <https://doi.org/10.1093/nar/gkac1079>. April 24.
- Bargheet, Ahmed, Noordzij, Hanna Theodora, Ponsoero, Alise J., et al., 2025. Dynamics of gut resistome and mobilome in early life: a meta-analysis. *EBioMedicine* 114 (March), 105630. <https://doi.org/10.1016/j.ebiom.2025.105630>.
- Barrantes-Jiménez, Kenia, Lejzerowicz, Franck, Tran, Tam, et al., 2025. Anthropogenic imprint on riverine plasmidome diversity and proliferation of antibiotic resistance genes following pollution and urbanization. *Water Res.* 281 (August), 123553. <https://doi.org/10.1016/j.watres.2025.123553>.
- Beghini, Francesco, McIver, Lauren J., Blanco-Míguez, Aitor, et al., 2021. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* 10 (May), e65088. <https://doi.org/10.7554/eLife.65088>.
- Bolger, Anthony M., Lohse, Marc, Usadel, Bjoern, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bravo, Alicia, Moreno-Blanco, Ana, Espinosa, Manuel, 2023. One earth: the equilibrium between the human and the bacterial worlds. *Int. J. Mol. Sci.* 24 (20), 20. <https://doi.org/10.3390/ijms242015047>.
- Buchfink, Benjamin, Reuter, Klaus, Drost, Hajk-Georg, 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 18 (4), 366–368. <https://doi.org/10.1038/s41592-021-01101-x>.
- Camacho, Christiam, Coulouris, George, Avagyan, Vahram, et al., 2009. BLAST+: architecture and applications. *BMC Bioinform.* 10 (December), 421. <https://doi.org/10.1186/1471-2105-10-421>.
- Camargo, Antonio Pedro, Roux, Simon, Schulz, Frederik, et al., 2023. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* 1–10. <https://doi.org/10.1038/s41587-023-01953-y>. September 21.
- Carapuça, Elisabete, Azzoni, Adriano R., Prazeres, Duarte M.F., Monteiro, Gabriel A., Mergulhão, Filipe J.M., 2007. Time-course determination of plasmid content in eukaryotic and prokaryotic cells using real-time PCR. *Mol. Biotechnol.* 37 (2), 120–126. <https://doi.org/10.1007/s12033-007-0007-3>.
- Charubin, Kamil, Hill, John D., Papoutsakis, Eleftherios Terry, 2024. DNA transfer between two different species mediated by heterologous cell fusion in *Clostridium* coculture. *mBio* 15 (2), e0313323. <https://doi.org/10.1128/mbio.03133-23>.
- Chaumeil, Pierre-Alain, Müssig, Aaron J., Hugenholtz, Philip, Parks, Donovan H., 2019. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinf. (Oxf)* 36 (6), 1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
- Chaumeil, Pierre-Alain, Müssig, Aaron J., Hugenholtz, Philip, Parks, Donovan H., 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinf. (Oxf)* 38 (23), 5315–5316. <https://doi.org/10.1093/bioinformatics/btac672>.
- Chung, Neo Christopher, Miasojedow, Blażej, Startek, Michał, Gambin, Anna, 2019. Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data. *BMC Bioinform.* 20 (15), 644. <https://doi.org/10.1186/s12859-019-3118-5>.
- Danecek, Petr, Bonfield, James K., Liddle, Jennifer, et al., 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10 (2), giab008. <https://doi.org/10.1093/gigascience/giab008>.
- De Gelder, Leen, Vandecasteele, Frederik P.J., Brown, Celeste J., Forney, Larry J., Top, Eva M., 2005. Plasmid donor affects host range of promiscuous IncP-1β plasmid pB10 in an activated-sludge microbial community. *Appl. Environ. Microbiol.* 71 (9), 5309–5317. <https://doi.org/10.1128/AEM.71.9.5309-5317.2005>.
- de Muinck, Eric J., Trosvik, Pål, 2018. Individuality and convergence of the infant gut microbiota during the first year of life. *Nat. Commun.* 9 (1), 2233. <https://doi.org/10.1038/s41467-018-04641-7>.
- Ferretti, Pamela, Pasolli, Edoardo, Tett, Adrian, et al., 2018. Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. *Cell Host Microbe* 24 (1), 133–145 e5. <https://doi.org/10.1016/j.chom.2018.06.005>.
- Fitzgerald, Cormac Brian, Shkoporov, Andrey N., Sutton, Thomas D.S., et al., 2018. Comparative analysis of *Faecalibacterium Prausnitzii* genomes shows a high level of

- genome plasticity and warrants separation into new species-level taxa. *BMC Genomics* 19 (1), 931. <https://doi.org/10.1186/s12864-018-5313-6>.
- Fogarty, Emily C., Schechter, Matthew S., Lolans, Karen, et al., 2024. A cryptic plasmid is among the most numerous genetic elements in the human gut. *Cell* 187 (5), 1206–1222 e16. <https://doi.org/10.1016/j.cell.2024.01.039>.
- Galata, Valentina, Fehlmann, Tobias, Backes, Christina, Keller, Andreas, 2019. PLSDb: a resource of complete bacterial plasmids. *Nucleic Acids Res.* 47 (D1), D195–D202. <https://doi.org/10.1093/nar/gky1050>.
- Guglielmini, Julien, de la Cruz, Fernando, Rocha, Eduardo P.C., 2013. Evolution of conjugation and type IV secretion systems. *Mol. Biol. Evol.* 30 (2), 315–331. <https://doi.org/10.1093/molbev/mss221>.
- Haudiquet, Matthieu, de Sousa, Jorge Moura, Touchon, Marie, Rocha, Eduardo P.C., 2022. Selfish, promiscuous and sometimes useful: how mobile genetic elements drive horizontal gene transfer in microbial populations. *Philos. Trans. R. Soc. B* 377 (1861), 20210234. <https://doi.org/10.1098/rstb.2021.0234>.
- He, Wanli, Russel, Jakob, Klinck, Franziska, Nesme, Joseph, Sørensen, Søren Johannes, 2024. Insights into the ecology of the infant gut plasmidome. *Nat. Commun.* 15 (1), 6924. <https://doi.org/10.1038/s41467-024-51398-3>.
- Kaminski, James, Gibson, Molly K., Franzosa, Eric A., Segata, Nicola, Dantas, Gautam, Huttenhower, Curtis, 2015. High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLoS Comput. Biol.* 11 (12), e1004557. <https://doi.org/10.1371/journal.pcbi.1004557>.
- Kerkvliet, Jesse J., Bossers, Alex, Kers, Jannigje G., Meneses, Rodrigo, Willems, Rob, Schürch, Anita C., 2024. Metagenomic assembly is the main bottleneck in the identification of mobile genetic elements. *PeerJ* 12 (January), e16695. <https://doi.org/10.7717/peerj.16695>.
- Kolde, Raivo, 2018. Pheatmap: Pretty Heatmaps. Released. <https://github.com/raivokolde/pheatmap>.
- Kolmogorov, Mikhail, Yuan, Jeffrey, Lin, Yu, Pevzner, Pavel A., 2019. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37 (5), 540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
- Langmead, Ben, Salzberg, Steven L., 2012. Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9 (4), 357–359. <https://doi.org/10.1038/nmeth.1923>.
- Lanza, Val Fernández, Tedim, Ana P., Martínez, José Luis, Baquero, Fernando, Coque, Teresa M., 2015. The plasmidome of Firmicutes: impact on the emergence and the spread of resistance to antimicrobials. *Microbiol. Spectr.* 3 (2). <https://doi.org/10.1128/microbiolspec.PLAS-0039-2014>. PLAS-0039-2014.
- Lassmann, Timo, 2019. Kalign 3: multiple sequence alignment of large data sets. *Bioinf (Oxf)* 36 (6), 1928–1929. <https://doi.org/10.1093/bioinformatics/btz795>.
- León-Sampedro, Ricardo, DelaFuente, Javier, Díaz-Agero, Cristina, et al., 2021. Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients. *Nat. Microbiol.* 6 (5), 606–616. <https://doi.org/10.1038/s41564-021-00879-y>.
- Li, Heng, Durbin, Richard, 2009. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinf (Oxf)* 25 (14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li, Dinghua, Liu, Chi-Man, Luo, Ruihang, Sadakane, Kunihiko, Lam, Tak-Wah, 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinf (Oxf)* 31 (10), 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
- Luo, Hao, Gao, Feng, 2019. DoriC 10.0: an updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Res.* 47 (D1), D74–D77. <https://doi.org/10.1093/nar/gky1014>.
- Machado, Daniela, Barbosa, Joana Cristina, Domingos, Melany, et al., 2022. Revealing antimicrobial resistance profile of the novel probiotic candidate *Faecalibacterium Prausnitzii* DSM 17677. *Int. J. Food Microbiol.* 363 (February), 109501. <https://doi.org/10.1016/j.ijfoodmicro.2021.109501>.
- Mahajan, Saurabh, Agashe, Deepa, 2022. Evolutionary jumps in bacterial GC content. *G3 (Bethesda, Md.)* 12 (8), jkacl08. <https://doi.org/10.1093/g3journal/jkacl08>.
- McInnes, Ross S., McCallum, Gregory E., Lamberte, Lisa E., van Schaik, Willem, 2020. Horizontal transfer of antibiotic resistance genes in the human gut microbiome. *Curr. Opin. Microbiol.* 53 (February), 35–43. <https://doi.org/10.1016/j.mib.2020.02.002>.
- Mikheenko, Alla, Prjibelski, Andrey, Saveliev, Vladislav, Antipov, Dmitry, Gurevich, Alexey, 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinf (Oxf)* 34 (13), i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>.
- Mikiewicz, Diana, Wróbel, Borys, Węgrzyn, Grzegorz, Płucienniczak, Andrzej, 1997. Isolation and characterization of a ColE1-like plasmid from *Enterobacter agglomerans* with a novel variant of *Rom* gene. *Plasmid* 38 (3), 210–219. <https://doi.org/10.1006/plas.1997.1312>.
- Néron, Bertrand, Denise, Rémi, Coluzzi, Charles, Touchon, Marie, Rocha, Eduardo P.C., Abby, Sophie S., 2023. MacSyFinder v2: improved modelling and search engine to identify molecular systems in genomes. *Peer Community J.* 3. <https://doi.org/10.24072/pcjournal.250>.
- Neuwirth, Erich, 2022. RColorBrewer: ColorBrewer Palettes. R Package Version 1.1-3. Released April 3. <https://CRAN.R-project.org/package=RColorBrewer>.
- Nguyen, Mai, Vedantam, Gayatri, 2011. Mobile genetic elements in the genus bacteroides, and their mechanism(s) of dissemination. *Mob. Genet. Elem.* 1 (3), 187–196. <https://doi.org/10.4161/mge.1.3.18448>.
- Nishida, Hiromi, 2012. Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids. *Int. J. Evol. Biol.* 2012, 342482. <https://doi.org/10.1155/2012/342482>.
- Oksanen, Jari, Simpson, Gavin L., Blanchet, Guillaume F., et al., 2025. Vegan: Community Ecology Package. Version R Package Version 2.8-0. Released. <https://vegan.github.io/vegan/>.
- Ondov, Brian D., Starrett, Gabriel J., Sappington, Anna, et al., 2019. Mash screen: high-throughput sequence containment estimation for genome discovery. *Genome Biol.* 20 (1), 232. <https://doi.org/10.1186/s13059-019-1841-x>.
- Paganini, Julian A., Plantinga, Nienke L., Arredondo-Alonso, Sergio, Willems, Rob J.L., Schürch, Anita C., 2021. Recovering *Escherichia Coli* plasmids in the absence of long-read sequencing data. *Microorganisms* 9 (8), 1613. <https://doi.org/10.3390/microorganisms9081613>.
- Pages, Hervé, Aboyoun, Patrick, Gentleman, Robert, DebRoy, Saikat, 2024. Biostrings: Efficient Manipulation of Biological Strings. Released. <https://bioconductor.org/packages/Biostrings>.
- Paradis, Emmanuel, Schliep, Klaus, 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35 (3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>.
- Parks, Donovan H., Imelfort, Michael, Skennerton, Connor T., Hugenholtz, Philip, Tyson, Gene W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25 (7), 1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Parks, Donovan H., Chuvochina, Maria, Rinke, Christian, Mussig, Aaron J., Chaumeil, Pierre-Alain, Hugenholtz, Philip, 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50 (D1), D785–D794. <https://doi.org/10.1093/nar/gkab776>.
- Pärnänen, Katriina, Karkman, Antti, Hultman, Jenni, et al., 2018. Maternal gut and breast milk microbiota affect infant gut antibiotic resistance and mobile genetic elements. *Nat. Commun.* 9 (1), 1. <https://doi.org/10.1038/s41467-018-06393-w>.
- Penders, John, Gerhold, Kerstin, Thijs, Carel, et al., 2014. New insights into the hygiene hypothesis in allergic diseases: mediation of sibling and birth mode effects by the gut microbiota. *Gut Microbes* 5 (2), 239. <https://doi.org/10.4161/gmic.27905>.
- Perron, Gabriel G., Fredrik Inglis, R., Pennings, Pleuni S., Cobey, Sarah, 2015. Fighting microbial drug resistance: a primer on the role of evolutionary biology in public health. *Evol. Appl.* 8 (3), 211–222. <https://doi.org/10.1111/eva.12254>.
- Pfeifer, Eugen, Bonnin, Rémy A., Rocha, Eduardo P.C., 2022. Phage-plasmids spread antibiotic resistance genes through infection and lysogenic conversion. *mBio*. <https://doi.org/10.1128/mbio.01851-22> ahead of print, September 26. 1752 N St., N.W., Washington, DC.
- Price, Morgan N., Dehal, Paramvir S., Arkin, Adam P., 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5 (3), e9490. <https://doi.org/10.1371/journal.pone.0009490>.
- R Core Team, 2023. R: A Language and Environment for Statistical Computing. Version (R Foundation for Statistical Computing, 2023). Released. <https://www.r-project.org/>.
- Redondo-Salvo, Santiago, Fernández-López, Raúl, Ruiz, Raúl, et al., 2020. Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat. Commun.* 11 (1), 3602. <https://doi.org/10.1038/s41467-020-17278-2>.
- Rodríguez-Beltrán, Jerónimo, DelaFuente, Javier, León-Sampedro, Ricardo, MacLean, R. Craig, Millán, Álvaro San, 2021. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nat. Rev. Microbiol.* 19 (6), 347–359. <https://doi.org/10.1038/s41579-020-00497-1>.
- Rouches, Miles V., Xu, Yasu, Cortes, Louis Brian Georges, Lambert, Guillaume, 2022. A plasmid system with tunable copy number. *Nat. Commun.* 13 (1), 3908. <https://doi.org/10.1038/s41467-022-31422-0>.
- Schmartz, Georges P., Hartung, Anna, Hirsch, Pascal, et al., 2022. PLSDb: advancing a comprehensive database of bacterial plasmids. *Nucleic Acids Res.* 50 (D1), D273–D278. <https://doi.org/10.1093/nar/gkab1111>.
- Smillie, Chris S., Smith, Mark B., Friedman, Jonathan, Cordero, Otto X., David, Lawrence A., Alm, Eric J., 2011. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480 (7376), 7376. <https://doi.org/10.1038/nature10571>.
- Stockdale, S.R., Harrington, R.S., Shkoporov, A.N., et al., 2022. Metagenomic assembled plasmids of the human microbiome vary across disease cohorts. *Sci. Rep.* 12 (1), 9212. <https://doi.org/10.1038/s41598-022-13313-y>.
- The UniProt Consortium, 2023. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 51 (D1), D523–D531. <https://doi.org/10.1093/nar/gkac1052>.
- Trosvik, Pål, Noordzij, Hanna Theodora, de Muinck, Eric J., 2024. Antibiotic resistance gene dynamics in the commensal infant gut microbiome over the first year of life. *Sci. Rep.* 14 (1), 18701. <https://doi.org/10.1038/s41598-024-69275-w>.
- van der Velden, Niels, 2025. Geneviewer. Released. <https://github.com/nvelden/geneviewer>.
- Vatanen, Tommi, Jabbar, Karolina S., Ruotuhola, Terhi, et al., 2022. Mobile genetic elements from the maternal microbiome shape infant gut microbial assembly and metabolism. *Cell* 185 (26), 4921–4936 e15. <https://doi.org/10.1016/j.cell.2022.11.023>.
- Wein, Tanita, Dagan, Tal, 2020. Plasmid evolution. *Curr. Biol.* 30 (19), R1158–R1163. <https://doi.org/10.1016/j.cub.2020.07.003>.
- Wood, Simon N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 73 (1), 3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>.
- Wood, Derrick E., Salzberg, Steven L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15 (3), R46. <https://doi.org/10.1186/gb-2014-15-3-r46>.
- Woods, Emily C., Wetzel, Daniela, Mukerjee, Monjori, McBride, Shonna M., 2018. Examination of the *Clostridioides (Clostridium) Difficile* VanZ ortholog, CD1240. *Anaerobe* 53 (October), 108–115. <https://doi.org/10.1016/j.anaerobe.2018.06.013>.
- Clostridia 2017: 10th International Conference on the Molecular Biology and Pathogenesis of the Clostridia.

- Yaffe, Eitan, Relman, David A., 2020. Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. *Nat. Microbiol.* 5 (2), 343–353. <https://doi.org/10.1038/s41564-019-0625-0>.
- Yang, Lili, Mai, Guoqin, Hu, Zheng, et al., 2023. Global transmission of broad-host-range plasmids derived from the human gut microbiome. *Nucleic Acids Res.* 51 (15), 8005–8019. <https://doi.org/10.1093/nar/gkad498>.
- Yatsunenکو, Tanya, Rey, Federico E., Manary, Mark J., et al., 2012. Human gut microbiome viewed across age and geography. *Nature* 486 (7402), 222–227. <https://doi.org/10.1038/nature11053>.
- Yorki, Sosie, Shea, Terrance, Cuomo, Christina A., et al., 2023. Comparison of long- and short-read metagenomic assembly for low-abundance species and resistance genes. *Brief. Bioinform.* 24 (2), bbad050. <https://doi.org/10.1093/bib/bbad050>.
- Yu, Michael K., Fogarty, Emily C., Murat Eren, A., 2024. Diverse plasmid systems and their ecology across human gut metagenomes revealed by PlasX and MobMess. *Nat. Microbiol.* 9 (3), 830–847. <https://doi.org/10.1038/s41564-024-01610-3>.