

Recovery and microbial host assignment of mobile genetic elements in complex microbiomes: insights from a spiked gut sample

Bram Bloemen,^{1,2} Maud Delvoye,¹ Stefan Hoffman,¹ Kathleen Marchal,^{2,3} Kevin Vanneste,¹ Marie-Alice Fraiture,¹ Nancy H. C. Roosens,¹ Sigrid C. J. De Keersmaecker¹

AUTHOR AFFILIATIONS See affiliation list on p. 13.

ABSTRACT Mobile genetic elements (MGEs) are major drivers of horizontal gene transfer, including the spread of antimicrobial resistance (AMR) genes. However, determining the microbial host of an MGE in complex microbiomes remains challenging. Here, we spike a niche-specific *Bacillus velezensis* strain carrying a plasmid and linear phage-plasmid into a batch bioreactor simulating the human gut, and use it as a spike-in control to assess the performance of Hi-C sequencing and Oxford Nanopore Technologies (ONT)-enabled DNA methylation detection to identify MGE-host pairs. To improve recovery of low-abundance genomes, we used a novel ONT adaptive sampling (AS) strategy that depletes *de novo* assembled, sample-specific high-abundance contigs, rather than relying on reference genomes. This approach led to an approximately twofold enrichment of low-abundance replicons, including the spike-in strain. Methylation-based host assignment failed for the *B. velezensis* MGEs, likely due to the absence of DNA methylation. In contrast, Hi-C successfully linked the phage-plasmid to its host, but not the plasmid, likely due to non-intact cells, and only after removing artefactual signals through bioinformatic processing. For a native *Escherichia coli* strain, Hi-C and methylation data linked it to two plasmids. Selective isolation and whole-genome sequencing of both the native *E. coli* and spike-in *B. velezensis* then confirmed the metagenomic observations. Our results highlight that Hi-C and methylation data can provide powerful insights into MGE-host associations, but their interpretation requires careful computational analysis and biological validation. Moreover, our AS strategy offers a cost-efficient method to boost coverage of low-abundance genomes, improving metagenomic investigation of MGEs in complex microbiomes.

IMPORTANCE Mobile genetic elements are important contributors to horizontal gene transfer, including of antimicrobial resistance genes. Understanding which microbes carry these mobile elements is vital to assess the spread of resistance. Here, we use a nanopore adaptive sampling approach to increase detection of low-abundance bacteria and mobile elements and use DNA methylation detection and Hi-C sequencing to determine mobile element hosts. By introducing a known bacterium and isolating a native strain, we could evaluate the performance of these methods, indicating that although powerful, they require careful experimental design, interpretation, and validation. However, when combined, these approaches enable a comprehensive investigation of mobile elements and gene transfer dynamics in complex environments.

KEYWORDS antimicrobial resistance, microbiome, nanopore sequencing, mobile genetic elements, MGE-host prediction, AMR

Editor Yogesh Hooda, Child Health Research Foundation, Dhaka, Bangladesh

Address correspondence to Sigrid C. J. De Keersmaecker, sigrid.dekeersmaecker@sciensano.be.

The authors declare no conflict of interest.

Received 1 September 2025

Accepted 6 January 2026

Published 26 January 2026

Copyright © 2026 Bloemen et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Mobile genetic elements (MGEs), including phages and plasmids, play a crucial role in prokaryotic evolution by enabling the rapid spread of genes through horizontal gene transfer (HGT) (1). MGEs frequently carry antimicrobial resistance (AMR) genes and can transmit these to new species in microbial communities (2–6). Therefore, determining their hosts is vital to understanding AMR dynamics and assessing future AMR risks. Traditionally, this is done by cultivation-dependent approaches, such as isolation, to identify the original host, and culture-based mating assays to determine the host range (7). However, as these methods rely on cultivation and the presence of selectable traits such as AMR, their applicability in natural microbial environments is limited, e.g. for uncultivable strains or environments with high background AMR rates, such as the human gut (8). In recent years, several culture-independent tools have been developed that allow studying MGEs in their natural environments (7, 8). Among these, shotgun metagenomic sequencing has emerged as a powerful technique to comprehensively investigate entire microbial communities. However, accurately reconstructing MGEs, their respective hosts, and the broader genetic composition of these communities remains challenging. While long-read metagenomics has enabled significant advances in terms of assembly contiguity and accuracy (9), recovery of low-abundance genomes or MGEs remains difficult at commonly achieved sequencing depths (10, 11). Recently, Oxford Nanopore Technologies (ONT) adaptive sampling (AS) has been applied to alleviate this problem by enriching sequences of interest through the ejection of non-target reads from the nanopore (12–17). In complex natural microbiomes, AS has been applied to increase coverage of low-abundance genomes and MGEs, but these approaches depended on prior availability of external reference sequences (12–14). However, the use of *a priori* available reference genomes results in suboptimal AS performance due to mismatches with the actually present genomes, as demonstrated by Martin et al. (12). Therefore, the target sequences used for AS should be as similar as possible to the actual genomes (12). Once an MGE has been detected, another challenge in metagenomics is defining its host, especially for extrachromosomal MGEs (7, 8). Although computational tools exist to do this based on sequence composition, their performance decreases with increasing taxonomic resolution, especially for broad host range MGEs shared between strains with varying sequence composition (18–21). Recent improvements have enabled long-read sequencing technologies to detect bacterial methylation motifs, caused by methyltransferase (MTase) activity, commonly as part of restriction-modification systems. This information has been applied to attribute MGEs to their hosts by detecting shared methylation motifs (22–25). However, discriminating between different microbiome strains requires MTase activity and sufficient variation in detectable methylation motifs. Hi-C provides another way to predict MGE origins by chemically crosslinking different parts of the genome, capturing their *in vivo* spatial proximity. This has been applied in various environments, including gut microbiomes and wastewater, to comprehensively study MGE host range and transmission (5, 18, 26). However, Hi-C data can be noisy, and many studies do not use spike-in controls or appropriate bioinformatic methods to deal with this noise (27).

In this study, we assessed the performance of the above-described methods in terms of detecting MGEs and determining their host by spiking a well-characterized, niche-atypic *Bacillus velezensis* strain into a batch bioreactor simulating a human gut environment (28, 29). Importantly, the *B. velezensis* strain contains both a linear phage-plasmid and an AMR-carrying plasmid (29). Furthermore, we developed a novel AS approach where *de novo* assembled, sample-specific high-abundance genomes are depleted, thereby favoring the sequencing of low-abundance genomes, including the spike-in. Next, methylation motifs and Hi-C were evaluated for MGE-host linking, using the *B. velezensis* MGEs as an internal control. Finally, we extended the validation of metagenomic observations by selectively isolating both the spiked *B. velezensis* and a native *Escherichia coli* strain, based on the obtained metagenomic information. Isolated data for these strains were then compared to their MAG counterparts. By investigating

these methods and providing insights, we aimed to advance the use of metagenomics to study MGEs and their hosts in complex microbiomes.

RESULTS

Enhanced recovery of low-abundance genomes and MGEs from metagenomes with adaptive sampling based on sample-specific reference genomes

To evaluate metagenomic sequencing for plasmid origin determination in complex microbiomes, we spiked a *B. velezensis* strain, which is not common to the human microbiome environment, carrying high copy numbers of a known plasmid and phage-plasmid into the CoaP batch bioreactor. qPCR detected spike-in plasmid presence after incubation, and re-isolation of the spike-in confirmed the presence of both the plasmid and its host (Table S2). In contrast, and as expected, the control bioreactor was qPCR-negative. Next, we sequenced spiked and control bioreactor DNA extracts on nanopore MinION flow cells with standard conditions, but did not retrieve the *B. velezensis* strain nor its plasmid or phage-plasmid in the spiked sample (Table 1Table 1). To increase sensitivity, we generated a preliminary metagenomic assembly in order to deplete high-abundance contigs (>30× depth) with AS in a next MinION sequencing run (Fig. 1a). While the *B. velezensis* plasmid was detected, horizontal coverage of the chromosome only mildly increased (Table 1). By resequencing on PromethION to increase depth and using AS in ~50% of channels to deplete *de novo* assembled, and hence sample-specific high-abundance contigs, we detected the plasmid and chromosome on both the AS and non-AS part (Table 1). Despite initially containing fewer active channels (1,065 instead of 1,463) and generating a lower yield (40.99 Gb versus 90.69 Gb), the AS part of the flow cell obtained a ~1.3-fold higher coverage depth for *B. velezensis* than the non-AS part (Table 1, Fig. 1b). Normalizing by the initial ratio of active channels increased the enrichment to ~1.8, reflecting the expected enrichment if the AS and non-AS parts had equal active channel counts at the start (Fig. 1c). Interestingly, AS enrichment was higher initially but converged to no enrichment toward the end (Fig. 1d).

As our AS method depletes high-abundance contigs, it can enrich other low-abundance genomes alongside the spike-in strain, and potentially improve metagenome assembly. To assess this, we compared the assemblies from the AS and non-AS parts of the PromethION flow cell, only correcting for the initial number of active channels. Despite the lower yield, the AS method recovered 19 additional high- and medium-quality MAGs (≥50% completeness and <10% contamination) compared to the non-AS assembly (Fig. 2a). Next, we assessed the impact of the 30× threshold, beyond which contigs from the preliminary assembly are depleted (Fig. 1a). Between 1× and 30×, completeness strongly increased for the majority of MAGs (Fig. 2b), indicating that 30× is an appropriate threshold in most cases. Still, some MAGs remained incomplete at even higher depths, indicating constraints for assembly and binning beyond depth. We then used non-AS reads to estimate true relative abundances in the higher-quality AS assembly, since AS distorts coverage, and compared AS depth to non-AS depth to calculate enrichment (normalized for initial active pore counts). This revealed that

TABLE 1 Metagenomic detection of spike-in *B. velezensis* and its plasmid and phage-plasmid in both the control and spiked samples using various sequencing methods

Sample	Flow cell	Method	Yield (Gb)	Plasmid coverage breadth (depth)	Phage-plasmid coverage breadth (depth)	Chromosome coverage breadth (depth)
Control	MinION	Rapid	16.6	n.d. ^a	n.d.	1.25% (1.84)
Control	MinION	Rapid + AS	3.43	n.d.	n.d.	0.97% (0.07)
Spiked	MinION	Rapid	6.67	35.91% (0.51)	11.77% (0.12)	4.22% (0.22)
Spiked	MinION	Rapid + AS	3.5	67.9% (5.09)	28.87% (0.41)	6.63% (0.15)
Spiked	PromethION: channels 0–1,500	Ligation	90.69	100.00% (1,569.84)	100.00% (319.45)	99.98% (34.44)
Spiked	PromethION: channels 1,500–3,000	Ligation + AS	40.99	100.00% (2,005.15)	100.00% (388.31)	99.98% (43.89)

^an.d., not detected.

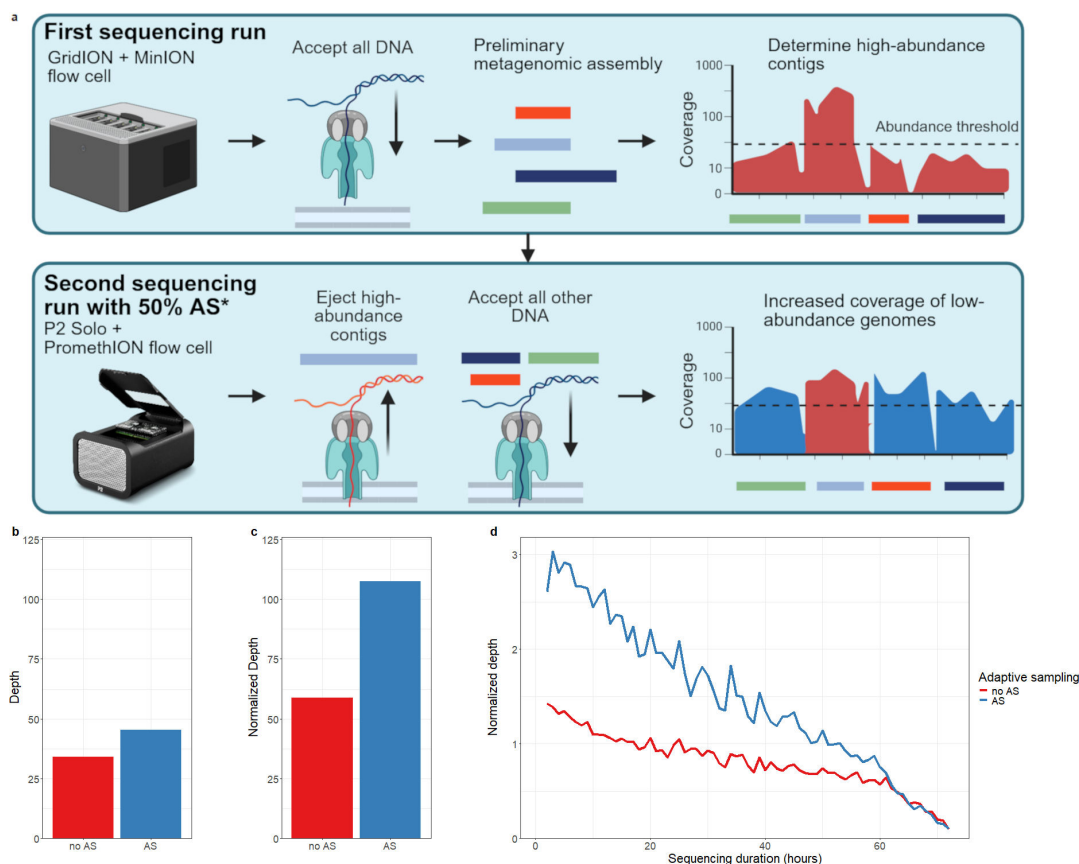


FIG 1 (a) Overview of the AS method. A first sequencing run was performed on MinION and followed by metagenomic assembly. Contigs exceeding a certain coverage threshold were retained and depleted with AS in a second sequencing run on a PromethION flow cell. *: Only ~50% of channels were configured to use AS, with the other half running regular sequencing as control. Made on biorender.com, P2 solo image courtesy of Oxford Nanopore Technologies. (b) Comparison of spike-in sequencing depth of the *B. velezensis* chromosome with and without the AS method. (c) Same as panel b, but normalized for the number of active channels at the start of the experiment. (d) Normalized sequencing depth of the spike-in throughout the 72 h experiment, indicating faster decay of throughput in the AS part of the flow cell.

the majority of MAGs were present at low relative abundance (below 1%, Fig. 2c). These low-abundance genomes were a median 2.03-fold enriched (95% bootstrap confidence interval 1.95–2.21). In contrast, high abundance MAGs, such as *E. coli*, were depleted (median 5.95-fold). Also, MAG completeness was significantly higher (one-sided Wilcoxon signed-rank, $P = 3.303 \times 10^{-12}$) in the AS assembly than in the non-AS assembly, especially for MAGs below 0.1% (Fig. 2d). Finally, we observed a median rejected read length of 675 bases with AS on PromethION, compared to only 445 bases with AS on MinION.

Methylation-based MGE-origin prediction depends on detectable methylation motifs

We mapped both AS and non-AS methylation-called reads to the AS assembly and investigated whether the MGEs carried by the spike-in could be linked to their host through shared methylation motifs. However, in the *B. velezensis* MAG, no methylation motifs were identified, and neither its plasmid nor phage-plasmid could therefore be linked. To ensure sufficient coverage for methylation motif identification, we then performed an ONT sequencing run on the *B. velezensis* isolate. While again no motifs were detected despite the high coverage (4,500×), MicrobeMod (30) detected several potential MTases and type IV restriction enzymes (RE) with high similarity to known REBASE entries (Table 2) (31). In contrast to *B. velezensis*, manual inspection of the *E. coli*

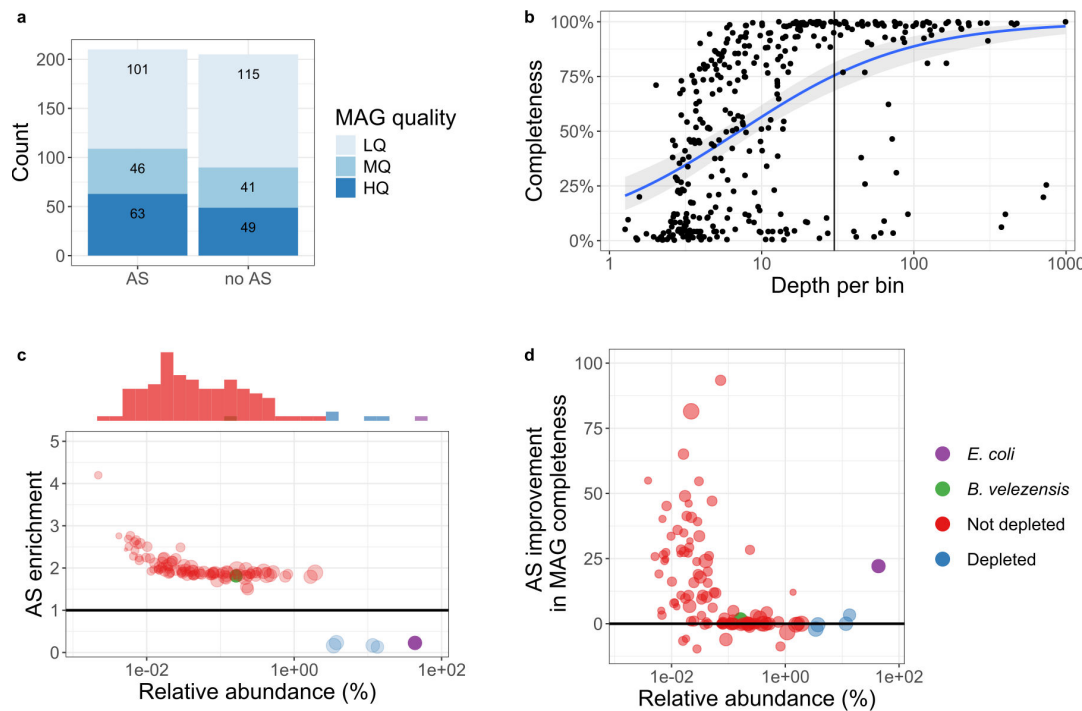


FIG 2 (a) Number of MAGs and their quality for the AS and non-AS data. LQ: low quality, <50% complete and <10% contamination; MQ: medium quality, ≥50% complete and <10% contamination; HQ: high quality, >90% complete and <10% contamination. (b) MAG completeness by depth for both AS and non-AS assemblies, with a vertical line indicating 30× coverage. The blue line represents a logistic regression, with standard error in gray. (c) Normalized enrichment of MAGs (≥50% completeness and <10% contamination) by relative abundance. The histogram on top indicates the distribution of MAGs by relative abundance. (d) Comparison of MAG completeness between MAGs from the AS and non-AS metagenomic assemblies.

MAG and other contigs with similar methylation motifs (Fig. 3) highlighted a potential association between several MGEs (contigs 5688, 17164, 11219, and several others) and a single-contig *E. coli* MAG (contig 1551, Fig. 4). Finally, while methylation sometimes clearly indicated a single host MAG for MGEs, e.g., for *Bifidobacterium adolescentis* and contigs 5454 and 13033 (Fig. 4), the potential origin was more ambiguous for many other MGEs (Fig. 3).

Noisy Hi-C data complicates linking spike-in to its MGEs, highlighting the importance of appropriate normalization and filtering

With DNA methylation detection unable to establish the origin of the plasmid and phage-plasmid carried by the spike-in control, we next sought whether proximity ligation (Hi-C) sequencing could determine this link, using metaCC to correct for various

TABLE 2 MTases and restriction enzymes detected in the *B. velezensis* spike-in isolate by MicrobeMod^a

Type	(Predicted) motif or activity	HMM e-value	REBASE homolog	REBASE AAI
Type IV RE	cuts m5C	0.0	Ppo12321ORFAP	100.0
Type IV RE	cuts m5C and m6A	1.3e-55	BveP12MrrP	100.0
Type IV RE	cuts m5C	1.5e-12	n.d.	n.d.
Type II MTase	m5C	7.2e-34	M.BveP22ORF2000P	100.0
Type II MTase	n.d.	9.4e-52	M.Bve116265ORF11045P	98.0
Type II MTase	n.d.	5.5e-13	n.d.	n.d.
Type II MTase	n.d.	9.5e-06	n.d.	n.d.
Type II MTase	n.d.	9.1e-19	n.d.	n.d.

^aRE, restriction enzyme; MTase, methyltransferase; HMM, hidden Markov model; REBASE, restriction enzyme database; AAI, amino-acid identity; n.d., not determined.

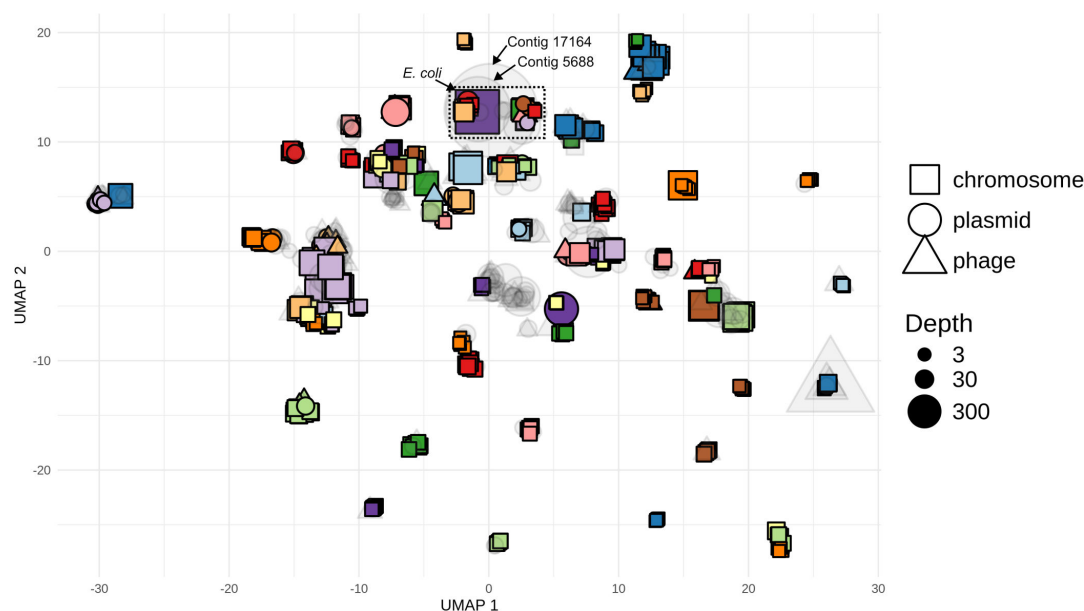


FIG 3 UMAP of methylation patterns in the metagenomic assembly of the PromethION adaptive sampling data set. Mobile elements such as plasmids or phages are indicated by circles and triangles, respectively. Squares represent chromosomal contigs and are colored by MAG. Only MAGs with $\geq 50\%$ completeness and $< 10\%$ contamination were considered. Colored symbols indicate binned contigs, while transparent, gray triangles and circles indicate unbinned mobile elements. The *E. coli* MAG (contig 1551, purple square) and two high-depth plasmids (transparent gray circles; contigs 5688 and 17164) are indicated. The dotted square indicates the contigs analyzed further in Fig. 4. *B. velezensis* is absent from this figure, since it contained no identified methylation motifs.

Hi-C biases (32). While no Hi-C links were observed between the *B. velezensis* chromosome and its plasmid, the phage-plasmid could be linked (Fig. 5a). However, this signal could not be discriminated from links to, e.g., *E. coli* in the non-normalized Hi-C data, as only two Hi-C read pairs supported this connection (Fig. S1a). Only after metaCC normalization (32), the correct signal to *B. velezensis* could be distinguished from noise (Fig. S1b). In general, *B. velezensis* had a low number of Hi-C reads compared to other MAGs at comparable relative abundances (0.15%, Fig. 5c), and a majority of these Hi-C contacts mapped to other MAGs (Fig. 5d). Additionally, the low median ONT read length of 874 bases (Fig. 5b) indicates possible DNA degradation (which might be linked to the used growth conditions), as the median length was 2,968 bases in isolate data, and other high and medium-quality MAGs had a median read length of 3,136 bases. In contrast to the *B. velezensis* spike-in, the *E. coli* MAG contained a large number of Hi-C contacts ($> 10^8$) (Fig. 5b and c), most of which were intra-MAG contacts (Fig. 5d). After normalization and filtering, the *E. coli* MAG was linked to two plasmids: contig 5688 and contig 17164 (Fig. 5a), which were previously seen to carry a similar methylation pattern (Fig. 4). Contig 5688 represents a 118 kbp plasmid carrying several AMR genes, including *aadA1* and *blaSHV-1* (Table 3). However, many other contigs with similar methylation patterns (e.g., contig 11219) were not connected to the *E. coli* MAG by Hi-C links. To verify the Hi-C-based links, we selectively cultured the sample on medium supplemented with streptomycin (*aadA1*) and ampicillin (*blaSHV-1*), in order to specifically isolate the strain carrying this plasmid. qPCR was then used to confirm *E. coli* as the species, and the presence of the *aadA1* and *blaSHV-1* genes. The isolate was sequenced separately, which confirmed the presence of the contig 5688 and contig 17164 plasmids, but not the other unbinned MGEs on Fig. 4.

Validation of metagenomic assembly and metagenomic MGE-host links by isolate whole-genome sequencing

To assess metagenomic detection and assembly of mobile elements and their hosts, we compared the *E. coli* and *B. velezensis* isolate assemblies to their metagenomic

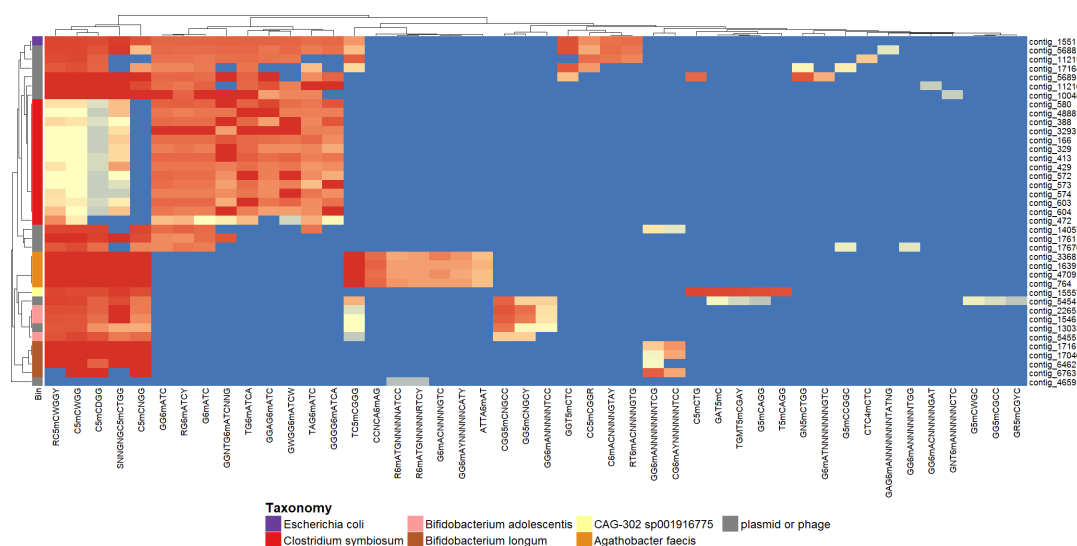


FIG 4 Heatmap of methylation patterns on *E. coli* and contigs from MAGs with similar methylation motifs (dotted square on Fig. 3), clustered by binary distance. Apart from unbinned mobile genetic elements, only MAGs with $\geq 50\%$ completeness and $< 10\%$ contamination were plotted.

counterparts (Table 3). For the spike-in *B. velezensis*, the phage-plasmid was detected in the PromethION data by geNomad and consisted of a single 13.9 kbp contig (5667) covering 99.17% of the reference with 100% identity. For the *B. velezensis* plasmid, seven metagenomic contigs were assembled, covering 71.04% of the reference with 100% identity. Some of these contigs overlapped (Fig. S2a), and only one was reported as a plasmid. The chromosome itself was assembled into four contigs with 99.71% coverage and 99.99% identity to the isolate chromosome. For *E. coli*, two plasmid contigs were assembled in the isolate, and for each, a single metagenomic contig (i.e., 5688 and 17,164) could be aligned with near-perfect coverage and identity to the reference. However, the second plasmid formed a 13.3 kbp contig in the metagenomic assembly, which is a concatenated version of the 6.6 kbp plasmid in the isolate Autocycler assembly (Fig. S2b). We also compared the *E. coli* methylation motifs from the metagenomic data set to those observed in the isolate (see below, Table 4). For motifs with a high methylation degree in the isolate ($> 80\%$), Nanomotif found a corresponding MTase, and the same or similar motifs were retrieved from the metagenomic data. In particular, the motif GGT5mCTC discriminated some *E. coli*-MGE pairs from other MAGs in methylation-based clustering (Fig. 4). Its reverse complement, GAG6mACC, was also methylated, but this motif was not detected as such in the MAG. For motifs with $< 80\%$ methylation in the isolate (e.g., CC5mCGGA), Nanomotif could not identify a MTase.

DISCUSSION

Understanding HGT dynamics in natural microbial communities is important to assess the genetic exchange of critical traits, such as antimicrobial resistance (7). MGEs play a central role in HGT, but studying them in their natural environments poses several challenges. Many current methods rely on cultivation or can only track a limited number of MGEs or microbes at once (7, 8). In contrast, metagenomics enables a cultivation-independent and comprehensive view of microbiomes and the MGEs within them. However, it depends on sufficient sequencing coverage, which is challenging for low-abundance strains and MGEs (10, 11). Additionally, determining the microbial hosts of MGE is difficult, as DNA extraction disrupts cells, creating a mixed pool of DNA fragments of unknown cellular origins. In this study, we addressed these limitations using three approaches: nanopore AS, detection of shared methylation motifs in MGEs and their host replicons, and Hi-C proximity ligation. Furthermore, we used both a spike-in control and an isolated native strain to validate metagenomic observations.

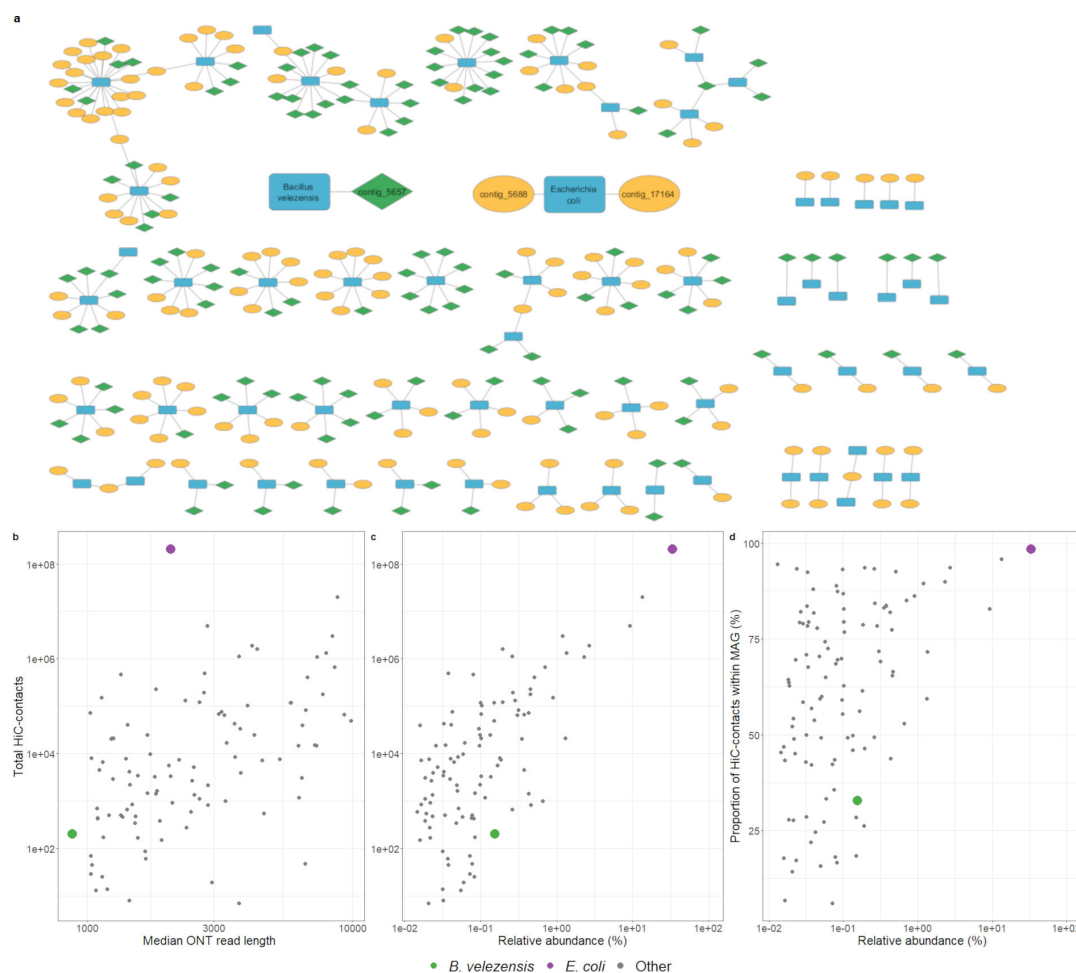


FIG 5 (a) Network graph of Hi-C links between plasmids (yellow), phages (green), and their hosts (blue). The *B. velezensis* and *E. coli* MAGs and their associated phages or plasmids are labeled. Mobile elements or MAGs without links (e.g., *B. velezensis* plasmid) are not plotted. (b) Number of total raw Hi-C contacts by median ONT read length. (c) Number of total raw Hi-C contacts by relative abundance. (d) Proportion of intra-MAG Hi-C contacts by relative abundance. In all plots, only high and medium-quality MAGs are included (contamination <10%, completeness >50%).

To address limited sequencing coverage, we used ONT AS to redistribute sequencing capacity from *de novo* sample-specific assembled high-abundant contigs to low-abundant replicons, instead of relying on external, *a priori* known reference genomes for AS. Depleting our *de novo* assembled contigs resulted in an average of approximately twofold enrichment by yield of low-abundance MAGs, and in particular 1.83-fold for our spike-in control. Other studies report similar enrichments of around twofold by depth but did not normalize by the initial number of active pores (14–17). By enriching all low-abundance genomes, our method led to improved MAG recovery despite reducing overall sequencing yield, similar to results by Sun et al. (14). However, comparing small plasmids in the *B. velezensis* and *E. coli* isolates to their metagenomic counterparts revealed known issues in metagenomic assemblers (33). Also of note, rejected reads were markedly longer for PromethION (675 bases) compared to MinION (445 bases), the latter being in line with other studies (13, 15, 16). These longer rejected reads could indicate bottlenecks in the ability of the GridION to keep up with data generation by the PromethION (run via a P2 Solo device coupled to a GridION). Alleviating these bottlenecks through faster base calling and AS algorithms could reduce decision times, enabling higher enrichment and increasing the benefits of AS (34). Additionally, regular nuclease cleansing and reloading of the flow cell to preserve pore activity (which we did not do) and longer DNA fragments could enable more enrichment (12, 35). Also,

TABLE 3 Comparison of *B. velezensis* and *E. coli* isolate assemblies to metagenomic assemblies^a

Isolate contig	Metagenomic detection with geNomad	AMR genes	Reference length (bp)	Metagenomic assembly length (bp)	Number of metagenomic contigs	Reference coverage (%)	Identity (%)
<i>B. velezensis</i> chromosome	n.a. ^c	<i>satA</i>	4,402,877	4,282,440	4	99.71	99.99
<i>B. velezensis</i> plasmid	1/7 contigs	<i>aadA1</i> , <i>bleO</i>	6,775	7,747	7	71.04	100.00
<i>B. velezensis</i> phage-plasmid ^b	Yes	<i>f</i> ^d	13,899	13,877	1	99.17	100.00
<i>E. coli</i> chromosome	n.a.	/	5,112,102	5,112,101	1 (contig 1,551)	100.00	99.99
<i>E. coli</i> plasmid 1	Yes	<i>aadA1</i> , <i>sul1</i> , <i>qacEdelta1</i> , <i>blaSHV-1</i>	118,065	118,065	1 (contig 5,688)	100.00	100.00
<i>E. coli</i> plasmid 2	Yes	/	6,647	13,292	1 (contig 17,164)	100.00	99.99

^aAMR, antimicrobial resistance.^bAutocycler did not generate a separate phage-plasmid contig, so we used the contig from the previous reference (accession [NZ_OU015426.1](#)).^cn.a., not applicable.^d/, none detected.

our current method requires two flow cells, but could be performed on a single flow cell by pausing the run once sufficient coverage has been reached for microbes at a predetermined abundance threshold. This would be followed by *de novo* assembly and then continuing the run with AS depletion of the high-abundance contigs. Furthermore, dynamic implementations of this method using real-time assembly and real-time updating of the AS decision strategy have also been proposed (12, 13). Nevertheless, the AS method improved assembly of low-abundance genomes and MGEs without requiring higher sequencing depth, thereby increasing cost-efficiency.

While AS can improve recovery of low-abundance MGEs, it does not necessarily resolve the issue of linking extrachromosomal MGEs to their host microbe. To address this, we leveraged ONT-enabled methylation motif detection. This revealed several MGE-*E. coli* pairs sharing common methylated motifs, some exclusive to these pairs, with two MGE-*E. coli* pairs later confirmed by Hi-C and isolation. Most of these shared motifs were further validated by identifying a corresponding MTase and consistency between metagenomic and isolate motifs. Interestingly, CC5mCGGA was one of the *E. coli* motifs for which no MTase was detected and could also be explained by potential erroneous base calling of a C5mCWGG pattern, with a T (W) being base called as 5mC, and the second 5mC as C. In contrast to *E. coli*, the plasmid and linear phage-plasmid of the spike-in control *B. velezensis* could not be linked, as no methylation motifs were reported by the state-of-the-art methylation motif calling tool (36), even in isolate whole-genome sequencing data with ~4,500-fold coverage. Despite the absence of detected motifs, several candidate type II MTases were identified by MicrobeMod. On the other hand, two type IV REs were also detected with high confidence. Since type IV REs are known to cleave at methylated positions with poor sequence specificity (37),

TABLE 4 Comparison of methylation motifs detected in *E. coli* isolate and in its metagenomic-assembled genome

Motif in isolate ^a	Degree of methylation (%)	Candidate MTase	Motif in metagenomic assembly ^a	Reverse complement
G6mATC	97.9	M.SspPIBDamP	G6mATC (and variations)	
C5mCWGG	99.2	M.SfiC32DcmP	C5mCWGG (and variations)	
GAG6mACC	84.3	M2.Ecos9ORF22615P	n.d. ^b	Reverse complement
GGT5mCTC	99.0	M2.EcosS3ORF17125P	GGT5mCTC (and variations)	
RT6mACNNNNGTG	98.0	M.EcosS3ORF17480P	RT6mACNNNNGTG	Reverse complement
C6mACNNNNGTAY	96.9		C6mACNNNNGTAY	
CC5mCGGA	55.1	n.d.	CC5mCGGR	Reverse complement
YC5mCGGG	64.1	n.d.	TC5mCGGG	
CTC5mCGGAKRA	73.0	n.d.	n.d.	

^aBold indicates the methylated nucleotide.^bn.d., not determined.

we suspect that these would be incompatible with active DNA methylation, potentially explaining why our *B. velezensis* strain seems to lack any type II methylation motifs. In this case, methylation-based MGE host linking would indeed not be feasible. For many other MGEs, methylation-based host attribution was ambiguous. These observations illustrate that methylation-based MGE-host attribution requires the presence and activity of methyltransferases, along with sufficient variation in detectable methylation motifs, which is not always the case (23, 38).

Whereas the resolution of DNA methylation binning is inherently limited by the number and types of methylation motifs, Hi-C sequencing enables a higher resolution by directly preserving DNA proximity information before lysis and subsequent DNA extraction. Nonetheless, Hi-C data needs to be interpreted with caution, as McCallum et al. demonstrated frequent artefactual Hi-C links and highlighted the importance of spike-in controls (27, 39). In our case, we used a thoroughly characterized *B. velezensis* spike-in control and could link it to its phage-plasmid, which was not possible with methylation data. However, no links were found to its plasmid, and the link to the phage-plasmid was supported by only two Hi-C contacts. Furthermore, before data processing, most raw spike-in Hi-C read pairs originating from the chromosome or the phage-plasmid connected them to other MAGs, indicating likely artefactual links. One possible explanation is that most *B. velezensis* DNA was present extracellularly, resulting in fewer and more random Hi-C contacts, as extracellular DNA is more likely to come into physical proximity with DNA from other strains (27, 40). Extracellular DNA degradation could also explain the shorter *B. velezensis* ONT read lengths compared to other species (41, 42). More stringent washing steps or DNase treatment could improve the Hi-C signal-to-noise ratio by removing extracellular DNA before the crosslinking step. After normalization and filtering of the raw Hi-C data, the artefactual Hi-C links to the spike-in were removed, highlighting the power of noise reduction and normalization strategies (27, 32). For *E. coli*, the processed Hi-C results were clearer and indicated that it hosted two plasmids, which were also seen in methylation-based linking, with one of these carrying several AMR genes. Using this information, we selectively isolated streptomycin- and ampicillin-resistant *E. coli* and then confirmed plasmid presence through isolate sequencing. This serves as proof of concept of how methylation motif detection and Hi-C can be applied in complex communities to obtain information on MGE hosts and to subsequently isolate MGEs of interest. Both methods have advantages and disadvantages, with Hi-C having higher resolution but depending on cellular integrity and accurate noise removal, while methylation detection is less dependent on cellular integrity, but on active MTases and sufficient variation in detectable methylation motifs.

While our study was performed on a single bioreactor sample, it reflected natural community structures, as it originated from a human donor and had at least 60 different genomes and a highly uneven distribution. As uneven distributions are common in microbial ecology (43), our AS approach could improve recovery of rare genomes and MGEs across different environments. Nevertheless, methylation binning efficiency likely depends on community MTase diversity, and Hi-C sequencing might require additional sample collection and processing steps in more complex sample matrices, e.g., to remove noise originating from extracellular debris.

In conclusion, our results demonstrate how combining nanopore AS, methylation profiling, and Hi-C sequencing could enable recovery of MGE-host pairs at low abundances. By including a spike-in control and isolating a native microbiome member, we showed that these approaches can generate biological insights but require careful experimental design and validation. Future work using ONT metagenomic sequencing could benefit from increased read lengths, faster AS algorithms, and improved DNA modification calling. Hi-C sequencing results could be improved by separating cellular debris from truly intact cells and using appropriate bioinformatic methods for interpretation. Taken together, these advancements will facilitate accurate, comprehensive, and cost-efficient tracking of MGEs in natural microbiomes, which is crucial to monitor the spread of AMR.

MATERIALS AND METHODS

In vitro microbiome cultivation and spiking

The human fecal sample was collected by ProDigest (Ghent, Belgium) from an anonymized individual according to the ethical approval of the University Hospital Ghent. It was then stored anaerobically at 4°C until processing (within 48 h after sample collection). The fecal suspension was sparged with nitrogen gas before flash-freezing at –80°C, using a modified version of the cryoprotectant used by Hoefman et al. (44). One milliliter of a 7.5% suspension of the fecal suspension was inoculated into the Colon-on-a-Plate (CoaP, Prodigest) reactor, in 10 mL total volume. The *B. velezensis* strain, described in reference 29, was grown overnight in Brain-Heart infusion broth (BHI, ThermoFisher Scientific) at 37°C, with 5 µg/mL of kanamycin (Sigma-Aldrich). One hundred microliters were then transferred to a 250 mL flask containing 25 mL BHI (ThermoFisher Scientific) and grown to OD₆₀₀ 0.2, corresponding to 10⁸ CFU/mL. One milliliter was centrifuged at 5,000 × *g* for 1 min, washed with 1 mL of PBS (ThermoFisher Scientific), and pelleted at 5,000 × *g*. The pellet was resuspended in 100 µL PBS and added to the CoaP reactor. Another reactor was not spiked (negative control). Both reactors were then incubated for 48 h at 37°C under anaerobic conditions and mild shaking (90 rpm).

Sampling and DNA extraction

For Hi-C sequencing and isolation, 500 µL of CoaP fluid was combined with 500 µL of the cryoprotectant described above, and flash-frozen at –80°C. For DNA extraction, 1 mL of CoaP fluid was pelleted (at 5,000 × *g*) and resuspended in 250 µL of DNA/RNA shield (Zymo Research). DNA was extracted using a modified version of the three-peaks extraction method (45), as described in Bloemen et al. (25). For DNA extraction from *B. velezensis* and *E. coli* isolates, colonies were inoculated in 10 mL Luria-Bertani (LB) broth (Sigma-Aldrich) with appropriate antibiotics (5 µg/mL kanamycin and 10 µg/mL streptomycin, respectively), and incubated overnight at 37°C with shaking. Next, 1 mL was pelleted at 5,000 × *g*, resuspended in 250 µL of DNA/RNA shield (Zymo Research), and DNA was extracted according to the same method as for the CoaP sample.

qPCR and isolation of spiked *B. velezensis* strain and native *E. coli* strain

qPCR to detect *B. velezensis* was done by adding 5 µL of extracted DNA to 20 µL qPCR mastermix containing 1× Ssoadvanced universal probes mastermix (Biorad) and the Bvel_plasmid primer/probe set (28) (Table S1). A C_q >35 was interpreted as negative (46). *B. velezensis* was re-isolated by inoculating 10 µL of cryopreserved CoaP fluid in 10 mL LB with 5 µg/mL kanamycin, and growing overnight at 37°C while shaking, followed by three overnight passages in 5 µg/mL kanamycin LB. Dilution series were then plated on 5 µg/mL kanamycin LB agar. Colonies were picked and incubated in 100 µL nuclease-free water (Invitrogen) for 10 min at 95°C. Five microliters of this served as a template for qPCR, containing 20 µL of 1× SYBR Green Real Time PCR Mastermix (Biosense) and the Bvel plasmid and chromosome primers (Table S1). For *E. coli* isolation, 10 µL of cryopreserved CoaP sample was inoculated in 10 mL LB with 10 µg/mL streptomycin, grown overnight at 37°C with shaking. Dilutions were plated on MacConkey agar (Oxoid) with streptomycin (10 µg/mL) and ampicillin (35 µg/mL), and colonies were screened for *blaSHV*, *sul1*, and *uidA* using primers as outlined in Table S1, using the same method as for *B. velezensis* colony qPCR. For all assays, the PCR protocol was 10 min at 95°C, followed by 45 cycles of 15 s at 95°C and 1 min extension at 60°C.

Nanopore sequencing, adaptive sampling, and base calling

Metagenomic libraries were generated according to the manufacturer's instructions using ONT rapid sequencing (SQK-RAD114) for MinION R10.4.1 flow cells (FLO-MIN114), and ONT ligation sequencing (SQK-LSK114) for PromethION (FLO-PRO114M). Samples were sequenced on a GridION or P2Solo connected to GridION (ONT) for 72 h. Isolates

were sequenced with SQK-LSK114 on MinION R10.4.1 flow cells. Dorado 0.9.1 and the dna_r10.4.1_e8.2_400bps_sup@v5.0.0 model were used for base calling, with option “--modified-bases 4mC_5mC 6mA,” and methylation models v5.0.0_4mC_5mC@v3 and v5.0.0_6mA@v3. Reads were filtered for Q >10 and length >500 bp using seqkit v2.3.1 (47). The AS runs on metagenomic samples were carried out as follows (Fig. 1): reads were assembled from MinION data sets without AS using Flye v2.9.4 (48), with options “--nano-hq --meta -i 1 --read-error 0.03.” Contigs with coverage >30× were used for the adaptive sampling deplete feature in MinKNOW (35). In the MinION AS runs on metagenomic libraries, all channels were used for AS, while on PromethION, channels 1–1,500 were selected for AS, with the remaining channels serving as control. To normalize coverage depths, we used the number of active channels at the start, determined from “pore_activity.csv.”

Detection of *B. velezensis* spike-in

Reads were mapped to the *B. velezensis* reference genome (ncbi accession [GCF_907164735.1](https://.ncbi.nlm.nih.gov/assembly/GCF_907164735.1)) with Minimap2 v2.28 (49), and filtered with SAMtools v1.21 (50), with flags “-F0x904.” Finally, a filtering step using pysam v0.22.1 removed reads with <90% identity or <80% mapped bases.

Proximity ligation (Hi-C) sequencing and read filtering

Proximity ligation was carried out using Proximeta (Phase Genomics), following the Proximeta v4 protocol. The resulting libraries were sequenced on a P2 flow cell on NextSeq 1000, generating 424 M 150 bp paired-end reads. For Hi-C reads, adaptor sequences, low-quality reads, and PCR duplicates were removed using the bbdut, clumpify, and repair scripts from BBmap v38.44 (51), according to the methods by Du and Sun (32). Hi-C reads were then mapped to the ISFinder database, downloaded from <https://github.com/thanhleviet/ISfinder-sequences> (52) using BWA-MEM v0.7.17 (53) with options “E50 -L0,” and reads with ≥60% coverage and ≥99% identity were removed to remove artefactual Hi-C links (27).

Metagenomic analysis

The PromethION data were split according to the AS and non-AS parts of the flow cell. Next, Flye (48) was used with options “--nano-hq --meta -i 2 --read-error 0.03.” Contigs were polished using Medaka v2.0.0 with the dna_r10.4.1_e8.2_400bps_sup@v5.0.0 model. Hi-C reads were mapped to the assembly using BWA-MEM v0.7.17, with preset “--5SP.” Minimap2 v2.28 (49) was used with parameters “-ax map-ont -y” to map ONT reads back to the assembly. Binning was done with MetaBAT2 v2.15 (54), GraphMB v0.1.5 (55) with parameters “--mincontig 3000 --assembly_type flye --vamb --minbin 250000” and SemiBin v2.1.0 (56) with parameters “--environment global --sequencing-type long_read --compression=none.” MetaCC v1.2.0 was used for Hi-C binning, with options “-e MluCI -e Sau3AI” for the normalization step (32). DASTool v1.1.6 (57) with parameter “--score_threshold 0” was used to select the best binning results. Methylation pileups were generated using Modkit v0.3.0, followed by Nanomotif v0.6.2 (24) “motif_discovery” with parameters “--read-level-methylation --threshold-valid-coverage 1,” and “include_contig” with “--run_detect_contamination.” MAGs were classified using “gtdbtk classify_wf” from GTDB-Tk v2.3.2 with GTDB database release 214 (58). MAG quality was assessed with “checkm lineage_wf” from CheckM v1.2.2 (59). MAGs were matched across different metagenomic assemblies by taking the closest FastANI hit with size >500 kbp (60), and compared to isolate assemblies using Mummer v4.0.0rc1 (61). R v4.4.3 and ggplot2 were used to visualize assembly statistics.

Plasmid and phage elements were identified with geNomad v1.8.0 (62). Prophage contigs with <50% of prophage content were excluded from MGE-host contact maps, as these are more host than prophage. To generate MGE-host Hi-C contact graphs,

MetaCC-normalized data were used, and the number of normalized contacts was aggregated by MAG. Contigs from bins with <50% completeness or >10% contamination were considered “unbinned.” Spurious Hi-C contacts were removed by filtering out MGE-bin links with <300,000 normalized contacts, and only retaining links representing $\geq 30\%$ of all contacts to a given MGE. The final graph was visualized in Cytoscape v3.10.2 (63). For methylation analysis, the same thresholds were used (<50% completeness or >10% contamination) for regarding contigs as unbinned. Only motifs detected at the bin level were retained, and motifs with <25% methylation in a bin (mean of median methylation degree per contig, weighted by contig length) or contig were removed. R v4.4.3 and the R packages ComplexHeatmap v2.10.0 (64) and UMAP v0.2.10.0 (65) were then used to visualize methylation data. For UMAP, a cosine distance matrix was used, while the Jaccard binary distance was used for the heatmap clustering.

Isolate assembly and methylation analysis

For both the *B. velezensis* spike-in and the native *E. coli*, we generated long-read assemblies using ONT R10.4.1 data (obtained as described above) with Autocycler v0.3.0 (66), using the wrapper script https://github.com/rrwick/Autocycler/blob/main/pipelines/autocycler_wrapper_by_iskold/autocycler_wrapper.sh with default settings for *E. coli*. A modified version was used to obtain an assembly for *B. velezensis*. Flye v2.9.4 was used as described above with all reads. For the other Autocycler assemblers, reads were downsampled to 10% (~450× coverage) for efficiency, and a single assembly was made with each assembler. For both *E. coli* and *B. velezensis*, all reads were used for methylation calling and motif detection with Nanomotif as described above. For *B. velezensis*, MicrobeMod v1.1.0 was additionally used to detect MTase genes, since no methylation motifs were identified.

ACKNOWLEDGMENTS

Funding for this study was provided through internal Sciensano funding.

AUTHOR AFFILIATIONS

¹Transversal Activities in Applied Genomics, Sciensano, Brussels, Belgium

²Department of Plant Biotechnology and Bioinformatics, Ghent University, Zwijnaarde, Belgium

³Department of Information Technology, Ghent University, Zwijnaarde, Belgium

AUTHOR ORCID*s*

Bram Bloemen  <http://orcid.org/0000-0003-1895-2221>

Kevin Vanneste  <http://orcid.org/0009-0000-4051-0628>

Nancy H. C. Roosens  <http://orcid.org/0000-0001-9218-078X>

Sigrid C. J. De Keersmaecker  <http://orcid.org/0000-0003-4198-4133>

AUTHOR CONTRIBUTIONS

Bram Bloemen, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft | Maud Delvoye, Methodology, Resources | Stefan Hoffman, Methodology, Resources | Kathleen Marchal, Supervision, Writing – review and editing | Kevin Vanneste, Data curation, Resources, Software, Writing – review and editing | Marie-Alice Fraiture, Resources, Writing – review and editing | Nancy H. C. Roosens, Funding acquisition, Resources, Writing – review and editing | Sigrid C. J. De Keersmaecker, Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The sequencing data generated in this study have been deposited in ENA under project [PRJEB95107](#). The bioinformatic pipeline for metagenomic assembly and Hi-C data processing can be found at https://github.com/brambloemen/ONT_MAG_pipeline/tree/MGE-host_linking_study.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental tables and figures (mSystems01282-25-s0001.docx). Table S1 and S2; Figures S1 and S2.

REFERENCES

- Arnold BJ, Huang I-T, Hanage WP. 2022. Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol* 20:206–218. <https://doi.org/10.1038/s41579-021-00650-4>
- Khedkar S, Smyshlyaev G, Letunic I, Maistrenko OM, Coelho LP, Orakov A, Forslund SK, Hildebrand F, Luetge M, Schmidt TSB, Barabas O, Bork P. 2022. Landscape of mobile genetic elements and their antibiotic resistance cargo in prokaryotic genomes. *Nucleic Acids Res* 50:3155–3168. <https://doi.org/10.1093/nar/gkac163>
- Souque C, González Ojeda I, Baym M. 2024. From petri dishes to patients to populations: scales and evolutionary mechanisms driving antibiotic resistance. *Annu Rev Microbiol* 78:361–382. <https://doi.org/10.1146/annurev-micro-041522-102707>
- Forster SC, Liu J, Kumar N, Gulliver EL, Gould JA, Escobar-Zepeda A, Mkandawire T, Pike LJ, Shao Y, Stares MD, Browne HP, Neville BA, Lawley TD. 2022. Strain-level characterization of broad host range mobile genetic elements transferring antibiotic resistance from the human microbiome. *Nat Commun* 13:1445. <https://doi.org/10.1038/s41467-022-29096-9>
- Risely A, Newbury A, Stalder T, Simmons BI, Top EM, Buckling A, Sanders D. 2024. Host- plasmid network structure in wastewater is linked to antimicrobial resistance genes. *Nat Commun* 15:555. <https://doi.org/10.1038/s41467-024-44827-w>
- Pfeifer E, Bonnin RA, Rocha EPC. 2022. Phage-plasmids spread antibiotic resistance genes through infection and lysogenic conversion. *mBio* 13:e0185122. <https://doi.org/10.1128/mbio.01851-22>
- Brito IL. 2021. Examining horizontal gene transfer in microbial communities. *Nat Rev Microbiol* 19:442–453. <https://doi.org/10.1038/s41579-021-00534-7>
- Yang QE, Gao JT, Zhou SG, Walsh TR. 2025. Cutting-edge tools for unveiling the dynamics of plasmid-host interactions. *Trends Microbiol* 33:496–509. <https://doi.org/10.1016/j.tim.2024.12.013>
- Sanderson ND, Hopkins KMV, Colpus M, Parker M, Lipworth S, Crook D, Stoesser N. 2024. Evaluation of the accuracy of bacterial genome reconstruction with Oxford Nanopore R10.4.1 long-read-only sequencing. *Microb Genom* 10:001246. <https://doi.org/10.1099/mgen.0.001246>
- Yorki S, Shea T, Cuomo CA, Walker BJ, LaRocque RC, Manson AL, Earl AM, Worby CJ. 2023. Comparison of long- and short-read metagenomic assembly for low-abundance species and resistance genes. *Brief Bioinformatics* 24. <https://doi.org/10.1093/bib/bbad050>
- Bickhart DM, Kolmogorov M, Tseng E, Portik DM, Korobeynikov A, Tolstoganov I, Uritskiy G, Liachko I, Sullivan ST, Shin SB, Zorea A, Andreu VP, Panke-Buisse K, Medema MH, Mizrahi I, Pevzner PA, Smith TPL. 2022. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat Biotechnol* 40:711–719. <https://doi.org/10.1038/s41587-021-01130-z>
- Martin S, Heavens D, Lan Y, Horsfield S, Clark MD, Leggett RM. 2022. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol* 23:11. <https://doi.org/10.1186/s13059-021-02582-x>
- Weilguny L, De Maio N, Munro R, Manser C, Birney E, Loose M, Goldman N. 2023. Dynamic, adaptive sampling during nanopore sequencing using Bayesian experimental design. *Nat Biotechnol* 41:1018–1025. <https://doi.org/10.1038/s41587-022-01580-z>
- Sun Y, Cheng Z, Li X, Yang Q, Zhao B, Wu Z, Xia Y. 2023. Genome enrichment of rare and unknown species from complicated microbiomes by nanopore selective sequencing. *Genome Res* 33:612–621. <https://doi.org/10.1101/gr.277266.122>
- Ulrich J-U, Epping L, Pilz T, Walther B, Stingl K, Semmler T, Renard BY. 2024. Nanopore adaptive sampling effectively enriches bacterial plasmids. *mSystems* 9:e0094523. <https://doi.org/10.1128/msystems.00945-23>
- De Meulenaere K, Cuypers WL, Gauglitz JM, Guetens P, Rosanas-Urgell A, Laukens K, Cuypers B. 2024. Selective whole-genome sequencing of *Plasmodium* parasites directly from blood samples by nanopore adaptive sampling. *mBio* 15:e0196723. <https://doi.org/10.1128/mbio.01967-23>
- Herbert J, Thompson S, Beckett AH, Robson SC. 2025. Impact of microbiological molecular methodologies on adaptive sampling using nanopore sequencing in metagenomic studies. *Environ Microbiome* 20:47. <https://doi.org/10.1186/s40793-025-00704-7>
- Shamash M, Sinha A, Maurice CF. 2025. Improving gut virome comparisons using predicted phage host information. *mSystems* 10:e0136424. <https://doi.org/10.1128/msystems.01364-24>
- Suzuki H, Yano H, Brown CJ, Top EM. 2010. Predicting plasmid promiscuity based on genomic signature. *J Bacteriol* 192:6045–6055. <https://doi.org/10.1128/JB.00277-10>
- Mahony J. 2024. Biological and bioinformatic tools for the discovery of unknown phage-host combinations. *Curr Opin Microbiol* 77:102426. <https://doi.org/10.1016/j.mib.2024.102426>
- Feng T, Chen X, Wu S, Tang W, Zhou H, Fang Z. 2025. Predicting the bacterial host range of plasmid genomes using the language model-based one-class support vector machine algorithm. *Microb Genom* 11:001355. <https://doi.org/10.1099/mgen.0.001355>
- Beaulaurier J, Zhu S, Deikus G, Mogno I, Zhang X-S, Davis-Richardson A, Canepa R, Triplett EW, Faith JJ, Sebra R, Schadt EE, Fang G. 2018. Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. *Nat Biotechnol* 36:61–69. <https://doi.org/10.1038/nbt.4037>
- Tourancheau A, Mead EA, Zhang X-S, Fang G. 2021. Discovering multiple types of DNA methylation from bacteria and microbiome using nanopore sequencing. *Nat Methods* 18:491–498. <https://doi.org/10.1038/s41592-021-01109-3>
- Heidelberg S, Dall SM, Bøjer JS, Nissen J, Maas LNL, Sereika M, Kirkegaard RH, Jensen SI, Kousgaard SJ, Thorlacius-Ussing O, Hose K, Nielsen TD, Albertsen M. 2024. Nanomotif: identification and exploitation of DNA methylation motifs in metagenomes using Oxford Nanopore sequencing. *bioRxiv*. <https://doi.org/10.1101/2024.04.29.591623>
- Bloemen B, Gand M, Ringenier M, Bogaerts B, Vanneste K, Marchal K, Roosens NHC, Dewulf J, Boyen F, De Keersmaecker SCJ. 2025. Overcoming challenges in metagenomic AMR surveillance with nanopore sequencing: a case study on fluoroquinolone resistance. *Front Microbiol* 16:1614301. <https://doi.org/10.3389/fmicb.2025.1614301>

26. Kent AG, Vill AC, Shi Q, Satlin MJ, Brito IL. 2020. Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. *Nat Commun* 11:4379. <https://doi.org/10.1038/s41467-020-18164-7>
27. McCallum GE, Rossiter AE, Quraishi MN, Iqbal TH, Kuehne SA, van Schaik W. 2023. Noise reduction strategies in metagenomic chromosome confirmation capture to link antibiotic resistance genes to microbial hosts. *Microb Genom* 9:001030. <https://doi.org/10.1099/mgen.0.001030>
28. Fraiture M-A, Bogaerts B, Winand R, Deckers M, Papazova N, Vanneste K, De Keersmaecker SCJ, Roosens NHC. 2020. Identification of an unauthorized genetically modified bacteria in food enzyme through whole-genome sequencing. *Sci Rep* 10:7094. <https://doi.org/10.1038/s41598-020-63987-5>
29. D'aes J, Fraiture M-A, Bogaerts B, De Keersmaecker SCJ, Roosens NHC, Vanneste K. 2021. Characterization of genetically modified microorganisms using short- and long-read whole-genome sequencing reveals contaminations of related origin in multiple commercial food enzyme products. *Foods* 10:2637. <https://doi.org/10.3390/foods10112637>
30. Crits-Christoph A, Kang SC, Lee HH, Ostrov N. 2023. MicrobeMod: a computational toolkit for identifying prokaryotic methylation and restriction-modification with nanopore sequencing. *Microbiology*. <https://doi.org/10.1101/2023.11.13.566931>
31. Roberts RJ, Vincze T, Posfai J, Macelis D. 2023. REBASE: a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* 51:D629–D630. <https://doi.org/10.1093/nar/gkac975>
32. Du Y, Sun F. 2023. MetaCC allows scalable and integrative analyses of both long-read and short-read metagenomic Hi-C data. *Nat Commun* 14:6231. <https://doi.org/10.1038/s41467-023-41209-6>
33. Trigodet F, Sachdeva R, Banfield JF, Eren AM. 2025. Assemblies of long-read metagenomes suffer from diverse errors. *bioRxiv*. <https://doi.org/10.1101/2025.04.22.649783>
34. Munro R, Payne A, Holmes N, Moore C, Cahyani I, Loose M. 2025. Enhancing nanopore adaptive sampling for PromethION using readfish at scale. *Genome Res* 35:877–885. <https://doi.org/10.1101/gr.279329.124>
35. Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M. 2021. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat Biotechnol* 39:442–450. <https://doi.org/10.1038/s41587-020-00746-x>
36. Soto-Serrano A, Li W, Panah FM, Hui Y, Atienza P, Fomenkov A, Roberts RJ, Deptula P, Krych L. 2024. Matching excellence: Oxford Nanopore technologies' rise to parity with pacific biosciences in genome reconstruction of non-model bacterium with high G+C content. *Microb Genom* 10:001316. <https://doi.org/10.1099/mgen.0.001316>
37. Loenen WAM, Raleigh EA. 2014. The other face of restriction: modification-dependent enzymes. *Nucleic Acids Res* 42:56–69. <https://doi.org/10.1093/nar/gkt747>
38. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, Froula J, Kang DD, Malmstrom RR, Morgan RD, Posfai J, Singh K, Visel A, Wetmore K, Zhao Z, Rubin EM, Korlach J, Pennacchio LA, Roberts RJ. 2016. The epigenomic landscape of prokaryotes. *PLoS Genet* 12:e1005854. <https://doi.org/10.1371/journal.pgen.1005854>
39. Castañeda-Barba S, Ridenhour BJ, Top EM, Stalder T. 2025. Detection of rare plasmid hosts using a targeted Hi-C approach. *ISME Commun* 5:ycae161. <https://doi.org/10.1093/ismeco/ycae161>
40. Nagano T, Várnai C, Schoenfelder S, Javierre B-M, Wingett SW, Fraser P. 2015. Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol* 16:175. <https://doi.org/10.1186/s13059-015-0753-7>
41. Calderón-Franco D, van Loosdrecht MCM, Abeel T, Weissbrodt DG. 2021. Free-floating extracellular DNA: systematic profiling of mobile genetic elements and antibiotic resistance from wastewater. *Water Res* 189:116592. <https://doi.org/10.1016/j.watres.2020.116592>
42. Beebe TJ. 1993. Identification and analysis of nucleic acids in natural freshwaters. *Sci Total Environ* 135:123–129. [https://doi.org/10.1016/0048-9697\(93\)90282-B](https://doi.org/10.1016/0048-9697(93)90282-B)
43. Shoemaker WR, Locey KJ, Lennon JT. 2017. A macroecological theory of microbial biodiversity. *Nat Ecol Evol* 1:107. <https://doi.org/10.1038/s41559-017-0107>
44. Hoefman S, Pommerening-Röser A, Samyn E, De Vos P, Heylen K. 2013. Efficient cryopreservation protocol enables accessibility of a broad range of ammonia-oxidizing bacteria for the scientific community. *Res Microbiol* 164:288–292. <https://doi.org/10.1016/j.resmic.2013.01.007>
45. Quick J. 2019. The “Three Peaks” faecal DNA extraction method for long-read sequencing v2. <https://doi.org/10.17504/protocols.io.rshmf6e>
46. D'aes J, Fraiture M-A, Bogaerts B, Van Laere Y, De Keersmaecker SCJ, Roosens NHC, Vanneste K. 2025. Metagenomics-based tracing of genetically modified microorganism contaminations in commercial fermentation products. *Food Chem (Oxf)* 10:100236. <https://doi.org/10.1016/j.fochms.2024.100236>
47. Shen W, Sipos B, Zhao L. 2024. SeqKit2: a Swiss army knife for sequence and alignment processing. *iMeta* 3:e191. <https://doi.org/10.1002/imt2.191>
48. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Pevzner PA. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* 17:1103–1110. <https://doi.org/10.1038/s41592-020-00971-x>
49. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
50. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
51. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. tech rep, lawrence berkeley national Lab(LBNL), Berkeley, CA (United States)
52. Siguer P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 34:D32–6. <https://doi.org/10.1093/nar/gkj014>
53. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. <https://doi.org/10.48550/arXiv.1303.3997>
54. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>
55. Lamurias A, Sereika M, Albertsen M, Hose K, Nielsen TD. 2022. Metagenomic binning with assembly graph embeddings. *Bioinformatics* 38:4481–4487. <https://doi.org/10.1093/bioinformatics/btac557>
56. Pan S, Zhao X-M, Coelho LP. 2023. SemiBin2: self-supervised contrastive learning leads to better MAGs for short- and long-read sequencing. *Bioinformatics* 39:i21–i29. <https://doi.org/10.1093/bioinformatics/btad209>
57. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 3:836–843. <https://doi.org/10.1038/s41564-018-0171-1>
58. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38:5315–5316. <https://doi.org/10.1093/bioinformatics/btad672>
59. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
60. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>
61. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. *PLOS Comput Biol* 14:e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>
62. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, Chain PSG, Nayfach S, Kyrpides NC. 2024. Identification of mobile genetic elements with geNomad. *Nat Biotechnol* 42:1303–1312. <https://doi.org/10.1038/s41587-023-01953-y>
63. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>
64. Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32:2847–2849. <https://doi.org/10.1093/bioinformatics/btw313>
65. McInnes L, Healy J, Melville J. 2020. UMAP: uniform manifold approximation and projection for dimension reduction. *ArXiv*
66. Wick RR, Hovden BP, Stinear TP. 2025. Autocycler: long-read consensus assembly for bacterial genomes. *Bioinformatics*. <https://doi.org/10.1101/2025.05.12.653612>