



# Identifying and tracking mobile elements in evolving compost communities yields insights into the nanobiome

Bram van Dijk<sup>1,2</sup>, Pauline Buffard<sup>1</sup>, Andrew D. Farr<sup>1</sup>, Franz Giersdorf<sup>1</sup>, Jeroen Meijer<sup>2</sup>, Bas E. Dutilh<sup>2,3</sup> and Paul B. Rainey<sup>1,4</sup>

© The Author(s) 2023

Microbial evolution is driven by rapid changes in gene content mediated by horizontal gene transfer (HGT). While mobile genetic elements (MGEs) are important drivers of gene flux, the nanobiome—the zoo of Darwinian replicators that depend on microbial hosts—remains poorly characterised. New approaches are necessary to increase our understanding beyond MGEs shaping individual populations, towards their impacts on complex microbial communities. A bioinformatic pipeline (xenoseq) was developed to cross-compare metagenomic samples from microbial consortia evolving in parallel, aimed at identifying MGE dissemination, which was applied to compost communities which underwent periodic mixing of MGEs. We show that xenoseq can distinguish movement of MGEs from demographic changes in community composition that otherwise confounds identification, and furthermore demonstrate the discovery of various unexpected entities. Of particular interest was a nanobacterium of the candidate phylum radiation (CPR) which is closely related to a species identified in groundwater ecosystems (*Candidatus Saccharibacterium*), and appears to have a parasitic lifestyle. We also highlight another prolific mobile element, a 313 kb plasmid hosted by a *Cellvibrio* lineage. The host was predicted to be capable of nitrogen fixation, and acquisition of the plasmid coincides with increased ammonia production. Taken together, our data show that new experimental strategies combined with bioinformatic analyses of metagenomic data stand to provide insight into the nanobiome as a driver of microbial community evolution.

ISME Communications; <https://doi.org/10.1038/s43705-023-00294-w>

## INTRODUCTION

Horizontal gene transfer (HGT) can markedly affect the evolutionary fate of microbes [1–3]. Besides transformation—where bacteria directly take up environmental DNA—all horizontal movement of genetic material is catalysed by mobile genetic elements (MGEs), Darwinian entities with dynamics of their own [4]. Last century, a multiplicity of MGEs has been observed, ranging from the distinctly parasitic bacteriophages [5–9] and transposons [10–12], to plasmids [13–15], and integrative and conjugative elements (ICEs) [16–19]. More recently, new mobile elements are being discovered all across the microbial world, such as REPINs [20, 21], Starships [22], and Borgs [23], and even entire fungal chromosomes appear to be on the move [24–26].

The relationship between MGEs and hosts are complex, ever changing, and highly context-dependent. For example, while conjugative elements are typically benign, they may also promote self-survival at the expense of hosts [19, 27]. Similarly, while bacteriophages are typically predatory or parasitic, they can be co-opted to benefit hosts [5, 28–30]. Moreover, MGEs may recombine with one another or parasitise other mobile elements [31–35]. Taken together, these processes may be fundamental to our understanding of microbial communities as collections of locally adaptive genes, rather than locally adapted species [36, 37]. While the scope and scale of DNA flux through microbial communities

via MGEs is currently poorly understood, recent work suggests that the flux may be highly significant even to the extent that it defines and drives a community-level process with effects similar to sex within populations [38, 39].

In addition to moving genes necessary for self-replication and transmission, MGEs often mediate transfer of host genes to which they become linked. Whether by design, or accident, MGEs that acquire genes that enhance host fitness stand to be rapidly amplified by selection with captured genes being widely disseminated. On occasion the effects can be highly consequential, for example, movement of ICEs carrying genes for nodulation and nitrogen fixation convert non-symbiotic rhizobia into plant symbionts in a single step [40, 41]. ICEs have also been identified by observing the movement of antimicrobial resistance [42, 43], and heavy metal resistance genes [18]. However, such routes to discovery depend on both abilities to culture focal microbes and carriage of selectable phenotypic traits.

With increasing ability to sequence complex communities, discovery of MGEs has been fuelled by metagenomics through culture-independent assembly of DNA replicons, without prior assumptions about biological relevance. Moreover, bioinformatic tools have been developed that can separate a wide range of candidate MGEs from microbial hosts [44–51]. Although such tools allow differentiation of chromosomal DNA from phages, plasmids,

<sup>1</sup>Department of Microbial Population Biology, Max Planck Institute for Evolutionary Biology, Plön, Germany. <sup>2</sup>Theoretical Biology and Bioinformatics, Department of Biology, Science for Life, Utrecht University, Utrecht, the Netherlands. <sup>3</sup>Institute of Biodiversity, Faculty of Biological Sciences, Cluster of Excellence Balance of the Microverse, Friedrich Schiller University, Jena, Germany. <sup>4</sup>Laboratory of Biophysics and Evolution, CBI, ESPCI Paris, Université PSL CNRS, Paris, France. ✉email: b.vandijk@uu.nl; rainey@evolbio.mpg.de

Received: 17 February 2023 Revised: 2 August 2023 Accepted: 8 August 2023

Published online: 28 August 2023

and other MGEs, discovery of MGEs is constrained by existing databases and training sets that are based on known MGEs. Moreover, metagenomic detection of MGEs *per se* rarely yields insight into the ecological relevance of the element. Unbiased detection and characterisation of unknown MGEs remains a major challenge.

Recently, Quistad et al. reported a generally applicable experimental strategy that connects movement of genes via MGEs to the ecology and functioning of complex microbial communities [52]. The strategy exploits the fact that long-term persistence of MGEs—and especially those that are costly to hosts—depends on horizontal transmission [53–55]. In the absence of opportunity to encounter new hosts, MGEs lose replicative capacity, resulting in either extinction, or co-option as permanent components of the host genome, at least in cases where elements encode host-beneficial fitness effects. Frequent exposure to new hosts, may instead breath ‘evolutionary life’ into MGEs, fuelling the continued co-evolution of MGEs and their hosts [13, 14, 56].

In the study by Quistad et al., garden compost communities were established and maintained for 48 weeks in glass mesocosms. These communities were further split into two pairs: horizontal- and vertical communities. For vertical communities (hereafter: V communities), a sample of the community (including MGEs) is periodically transferred to a fresh mesocosm. Horizontal communities (hereafter: H communities) are treated similarly, but with one important difference: at the time of transfer, a sample from all independent H communities is passed through a 0.2 µm filter and the filtrate collected. Filtration removes bacteria and larger entities, but allows collection of material smaller than 0.2 µm that stands to include recognised MGEs such as phages, but also yet-to-be-discovered elements, and additionally naked DNA, minerals, nutrients and so forth. This filtrate is referred to as an ‘MGE-cocktail’. The combined MGE cocktail of all independent communities is then redistributed across all H communities (see Fig. 1a). Evolved communities were then subjected to metagenomic sequencing, which—by virtue of the experimental design—allowed the identification of sequences in the H communities that were not detected in the respective ancestral communities (hereafter referred to as ‘unique sequences’). While their initial detection was not dependent on matches to existing viral databases, many of these unique sequences were shown to indeed encode phage-related proteins. In principle, however, any element in the MGE-cocktail with the ability to amplify after introduction into a new horizontal community, could be detected with this experimental design.

Here, we describe a broadly applicable ‘xenoseq’ pipeline (github.com/bramvandijk88/xenoseq) that takes as input, raw sequencing reads from time series experiments, and identifies sequences that have been transferred from allopatric communities and amplified by replication in the current community. When applied to metagenomic data from the experiment performed by Quistad et al., we show that xenoseq readily detects candidate MGEs whose dynamics (including community of origin) can be followed through time. One important attribute of xenoseq is ability to distinguish between sequences that are selected due to demographic changes in patterns of species abundance, and those that are horizontally disseminated from an allopatric community. The latter are hereafter referred to as ‘xenotypic sequences’. We show that xenotypic sequences are enriched in recognisable components of phages and IS-elements, that is, canonical selfish genetic elements (SGEs). Less expected was the observation of horizontally transmitted nanobacteria and large plasmids. To explore the dynamics of MGEs through the course of the year-long study, Metagenome-Assembled Genomes (MAGs) were constructed by cross assembly, providing insights into the potential hosts of these unexpected elements, and allowing the dissemination of MGEs to be linked to the community-level changes in ammonia production rates.

## MATERIALS AND METHODS

### Bioinformatic pipeline to detect transfer events

The xenoseq pipeline (<https://github.com/bramvandijk88/xenoseq>, Fig. 1c–e) is a wrapper that combines read trimming, assembly, read mapping, read filtering, and local alignment to seek evidence of horizontal gene transfer or the transfer of other nanoscale entities between evolving communities. This pipeline takes as input raw (untrimmed) fastq files containing shotgun metagenomic data of derived samples (query, in which to search for newly introduced sequences) and datasets from ancestral samples (reference). Query reads are trimmed using fastp [57] (v0.23.2), mapped against the corresponding reference contigs using BWA [58, 59] (v0.7.17) using default options, and samtools [60] (v1.15.1) is used with the flag ‘-4’ to extract unmapped reads. These reads are assembled into ‘unique contigs’ using Megahit [61] (v1.2.9) (*xenoseq\_find*, Fig. 1c). Next, these unique contigs are blasted against a local database using NCBI Blast [62] (v2.13.0) of all other reference communities to link the emergence of unique contigs to allopatric ‘donor’ communities. Linked contigs are extracted from unique contigs using Seqkit [63] (v2.3.0) (*xenoseq\_link*, Fig. 1d). By default, contigs are linked to a donor when blast has at least one high-scoring segment with a minimum length of 300 with at least 99% nucleotide identity. While the algorithm and cut-offs limit the sensitivity over larger evolutionary distances, they are designed to detect sequences that have recently diverged and are still largely identical. The contigs that match target sequences in allopatric communities are referred to as ‘*xenotypic contigs*’. Finally, to detect shifts in abundance of the detected contigs, reads from all query and reference samples are mapped (*xenoseq\_trace*), and coverage/breadth statistics are saved in a tab-separated file. This allows for visualisation of transfer across communities (Fig. 1e). In this manuscript, we further analyse sequences >10 kb in length.

By default, xenoseq runs all subroutines (*xenoseq\_find*, *xenoseq\_link*, *xenoseq\_trace*), but each subroutine can be run independently by modifying the command-line flags. Finally, xenoseq uses GNU parallel<sup>59</sup> to run multiple jobs simultaneously.

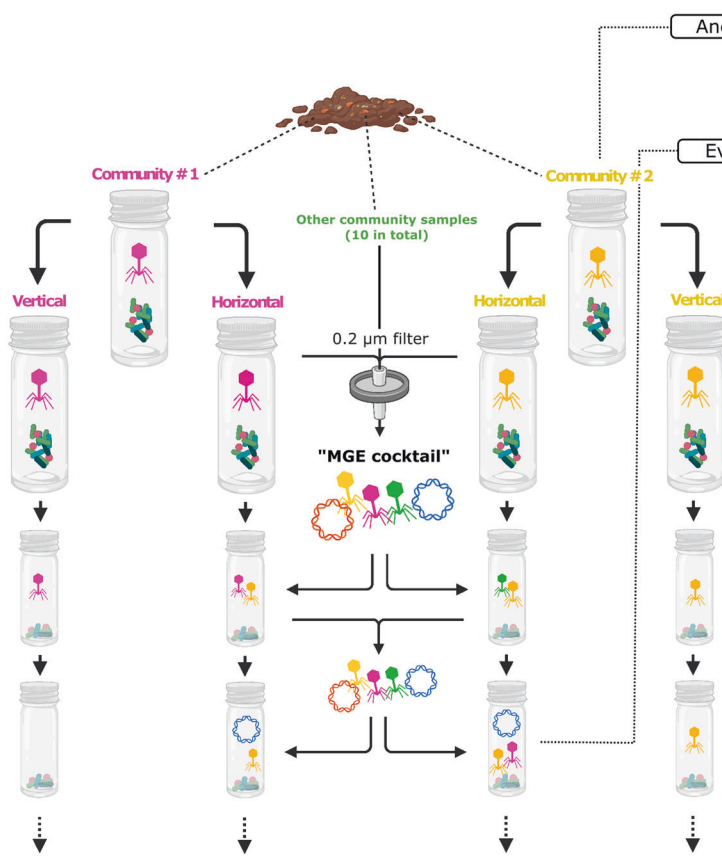
### Xenoseq benchmark

To benchmark the pipeline, introduction of MGEs into mock-communities was simulated (see Supplementary Material I full details). Six bacterial genome sequences were downloaded from RefSeq. Two mock-communities were then generated, one with an even taxon distribution (easy dataset), and one with a highly skewed taxon distribution (hard dataset). Simulated MGE sequences were either i) randomly inserted into genome sequences, representing integrative elements, or ii) included as separate replicons linked to a single host genome by addition to the genome fasta file as a separate contig. Illumina sequencing was simulated from the resulting ‘communities’ with ART<sup>62</sup>, using both default and ten-fold elevated error rates, which were used as input to benchmark the pipeline. Simulation of mock-communities and horizontal gene transfer events was done in R v4.1.3 using the packages biostrings<sup>60</sup> v2.62.0 and seqinr<sup>61</sup> v4.2.16. Another (small) mock-community dataset is available on the repository, and can be used to rapidly test whether the pipeline and its dependencies are configured correctly.

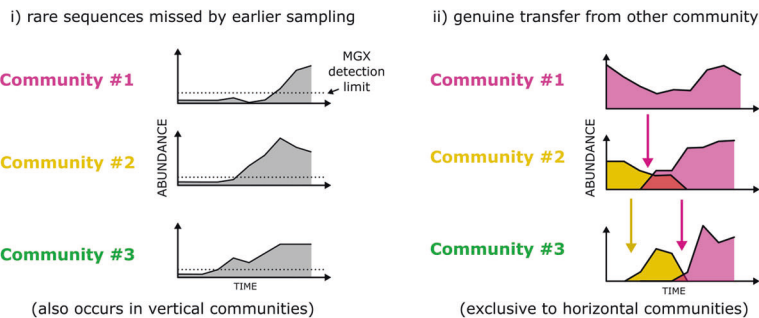
### Metagenome-assembled genomes

We generated MAGs from all twenty compost communities. By combining samples across multiple time points, we improved the potential detection of rare types whose coverage in a single sample is insufficient for assembly. Quality control of sequencing reads was done with Prinseq version 0.20.4 [64] using ‘-derep 14 -lc method -c threshold 20’. We trimmed adapters with Flexbar v.3.5.0 [65] using the ‘-adapter-preset Nextera -ap ON’ flag. Combined reads from multiple time points for each community were cross-assembled *de novo* using metaSPAdes v.3.14.0 [66]. Reads from each sample were mapped back to the assembled contigs using bwa-mem v.0.7.17-r1188 [59], and coverage calculations were performed with SAMtools v1.7 [60]. Contigs were binned using metaBAT2 v2.12.1 [67] and MAG quality was assessed with CheckM v1.1.3 [68]. All MAGs and contigs were annotated using BAT and CAT (Bin/Contig Annotation Tool, respectively, v.5.1.2) [69], which uses prodigal [70] to predict open reading frames and diamond [71] to query open reading frames against the NR database [72]. Where BAT or CAT were unable to reliably assign a taxon due to conflicting annotations between ORFs, a higher order taxonomic rank was assigned to the MAG: that is, to genus level if species identity could not be assigned, family level if genus could not be assigned, etc. CAT was called using parameters ‘-index\_chunk 1 --block\_size 5 --top 11’ using the CAT database and taxonomy constructed on 2021-04-30.

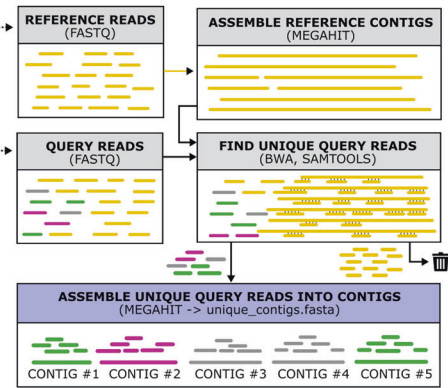
## a. Example experimental setup



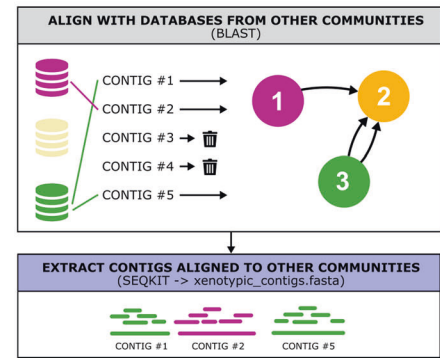
## b. Evolution in horizontal communities yields two types of "unique sequences"



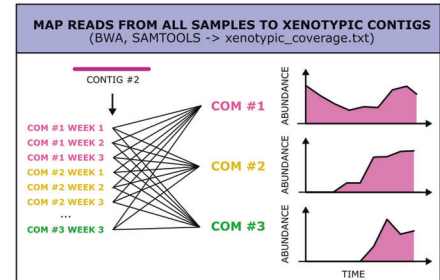
## c. Identify unique sequences (xenoseq\_find)



## d. Find evidence for parallel origin (xenoseq\_link)



## e. Reconstruct MGE dissemination (xenoseq\_trace)



**Fig. 1 Experimental design and overview of bioinformatic pipeline.** Overview of (a) experimental setup, (b) sources of unique sequences, and (c–e) the bioinformatic pipeline. **a** Experimental evolution protocol to identify novel MGEs and other mobile sequences [52]. Multiple parallel communities were established from garden compost, which were then propagated on cellulose as the sole carbon source and serially transferred every two weeks. Horizontal treatments were provided with an 'MGE cocktail' which was derived by pooling material collected from samples passed through 0.2 µm filters. This manipulation allows MGEs and linked genes to transfer horizontally from one community to the next. **b** Sequencing samples obtained from mesocosms yields two types of 'unique sequences'. Shown are cartoons to visualise how some unique sequences (left) are the result of rare sequences that were undetected by earlier sampling, while other unique sequences (right) are the result of genuine transfer from another community. **c** The first subroutine of xenoseq (xenoseq\_find) compares raw DNA sequences reads in fastq format from one sample (query, hereafter called *evolved*) with another sample (subject, hereafter *ancestral*). Ancestral samples are assembled into 'unique contigs', stretches of DNA that have newly appeared in the community. **d** To distinguish between two sources of unique sequence shown in panel b, the second subroutine (xenoseq\_link) identifies candidate mobile elements by aligning them to all other ancestral communities. The subset of contigs that align to allopatric communities are referred to as *xenotypic contigs*. **e** Finally, the dissemination patterns of xenotypic contigs is reconstructed by mapping DNA sequences reads from all communities against assembled contigs. Sequencing depth and breadth are stored in a tabular text file.

Read mapping of samples from individual time points (above) was used to study the relative abundance of individual MAGs. The nanobacterium MAG was placed into a 16S phylogeny with MAGs from groundwater ecosystems from He et al. [73] by extracting 16S sequences with Barnapp 0.9, and a neighbour-joining tree was created with Geneious 2023.1.2.

## DNA degradation assay

To investigate whether the identified sequences are derived from living cells or DNA liberated by lysed cells, we tested the ability of microbial compost communities to degrade extracellular DNA. To do this, four experimental microbial communities were established from samples of four independent compost heaps. Compost was washed in M9 salt

solution and stored with glycerol saline at  $-80^{\circ}\text{C}$ . Frozen stocks were thawed on ice and 4.25 mL of the stock was then washed of glycerol by two successive cycles of centrifugation (4000 *g*, 10 min) and resuspension (in 5 mL sterile M9 salt solution), followed by a final resuspension in 1 mL M9 salt solution. The washed cells were then added to 100 mL bottles with 19 mL M9 media and a piece of cellulose paper as a carbon source. Mesocosms were incubated without shaking at  $28^{\circ}\text{C}$  for 14 days to allow community growth. The entire volume including any remaining paper was then transferred to 50 mL centrifuge tubes, vortexed for five minutes to produce a slurry, and 1 mL of slurry was transferred to fresh M9 media for another 14 days of incubation. During this time a stock of genomic DNA (gDNA) from *Escherichia coli* (REL606) was made using the 'DNeasy Ultraclean Microbial kit' (Qiagen), resulting in a stock of  $\sim 30$  mL of gDNA at 20  $\mu\text{g}/\text{mL}$  (stored at  $-20^{\circ}\text{C}$ ).

Communities were then propagated a second time to four replicate cellulose-mesocosms spiked with gDNA. Community-free mesocosms were also established which acted as controls. All mesocosms were spiked with ten  $\mu\text{g}$  of *E. coli* gDNA, and were destructively harvested after 0, 1, 2 and 14 days of incubation. For sampling, the contents of each mesocosm was transferred to a 20 mL centrifuge tube, reduced to a slurry and then centrifuged at 4000 *g* for 5 min. The supernatant was transferred to 15 mL sterile tubes and frozen at  $-80^{\circ}\text{C}$  for later DNA extraction. In total, four replicate communities and three community-free controls were harvested at each time-point.

To measure relative DNA degradation by communities over the time-course, supernatant samples were thawed on ice, and further spiked with an internal standard of one  $\mu\text{g}/\text{mL}$  of gDNA of *Pseudomonas fluorescens* SBW25. Samples were then vortexed and samples were filtered using a 0.2  $\mu\text{m}$  syringe-filter. DNA of 2 mL of spiked filtrate was isolated using a Phage DNA Isolation Kit (Norgen Biotek Corp) following the manufacturer's instructions, except for loading two (rather than one) mL sample through a single purification column to improve DNA yield. Extracted DNA was then sequenced using a NextSeq MidOutput 300 cycle run resulting in  $2 \times 150$  bp paired-end reads (see Supplementary Table 1). Reads were aligned to the non-redundant (NR) protein database [72] with Diamond<sup>54</sup> using the blastx subroutine, where the top hit was used to identify the read as being either *P. fluorescens*, *E. coli*, or 'other' (the latter representing either species present in the communities or erroneously assigned reads).

### MGE classification

Viral sequences were predicted using Vibrant (v1.2.0) [44], Virsorter2 (v2.2.3) [45], Seeker (v1.0.3) [46], and CheckV (v0.8.1) [74]. While CheckV is primarily designed to estimate completeness of candidate viral genomes, we found that sequences with either low, medium or high scores were typically identified as viral by other tools as well (see Fig. 3 in main text). Plasmid sequences were predicted using PlasFlow (v1.1.0)w [47] and PlasClass (v0.1) [48] and, if relevant, the origin of replication with OriFinder 2022 [75]. ICEs were identified with ICEfinder [49]. IS elements were predicted with ISEscan (v1.7.2.3) [51]. Integrons were predicted using Integronfinder2 (v2.0rc6) [50]. All open reading frames on MGEs were annotated with prokka [76] with both the default database and the PHROG database [77] for identification of viral genes. All tools were run with default options, results are shown in Supplementary Table 1, Sheet 3.

### MAG metabolic functions

The limiting resources in the M9 medium of the experiment by Quistad et al. are carbon (provided by the cellulose paper) and nitrogen (1 mM ammonium chloride from M9 salts), both of which are added every 14 days during serial passaging. To obtain insight into community function, MAGs were interrogated for the presence of genes predicted to be involved in cellulose degradation and nitrogen metabolism. Although many relevant proteins are involved, we focus on endoglucanases and cellobiosidases (cellulose degradation), nitrogenases (nitrogen fixation), and nitrate/nitrite reductase (nitrogen reduction). More 'private' functions, such as the ability to take up glucose, glycolysis, and downstream pathways, are not taken into consideration for this study. Protein sequences were extracted from the Uniprot database (keywords: 'endoglucanase', 'cellobiosidase', 'nitrate reductase', 'nitrite reductase', and 'nitrogenase', all with the additional query 'reviewed:true') and aligned to MAGs using diamond blastx [71] with default settings. A score was assigned to each MAG by calculating the fraction of queries matched. For example, when one out of two cellobiosidases from Uniprot gave a significant match, a score of 0.5 was assigned. All scores are given in Supplementary Table 1, Sheet 5. For Fig. 5b, a metabolic function

was assigned when the score was greater than 0.1. When two MAGs with the same CAT annotation occurred in multiple communities, but the metabolic scores were not identical due to differences in genomic coverage, the highest score was used.

### Measuring the DNA concentration of MGE cocktails

To measure the DNA concentration present in DNA cocktails, compost mesocosms were established using the protocol described in Quistad et al. Briefly, samples were obtained from a compost heap (Plön, Germany) in September 2021. Five grams from each compost sample was transferred into a 100 mL glass flask containing a four  $\text{cm}^2$  piece of cellulose paper as a complex carbon source (Whatman cellulose filter paper) in 20 mL minimal M9 medium which contains 0.935 mM ammonium chloride. Subsequently, the founder mesocosms were incubated without shaking at  $28^{\circ}\text{C}$  for 2 weeks. The lids remained slightly opened, allowing gas exchange.

After a 2-week incubation period, the cellulose paper was transferred into a falcon tube with 20 mL of minimal M9 medium and vortexed into a slurry. The slurries were vortexed, and 12 mL of cellulose slurry was centrifuged, for 10 min, at 4000 *g*. Subsequently, 10 mL supernatant was filtered through a 0.2  $\mu\text{m}$  filter to produce the MGE cocktail. For DNA extraction, MGE cocktails were concentrated using an ultracentrifuge at  $\sim 26,500$  *g* for 45 min. The supernatant was discarded and the pellet resuspended in 2 mL medium. DNA was extracted using the Phage DNA Isolation kit (Norgen Biotek Corp) or QIAprep Spin Miniprep Kit, following the manufacturer's protocol. The amount of extracted DNA was assessed by Qubit HS assay. The measurements are found in Supplementary Table 1, Sheet 4.

### Data analysis and visualisation

Data analysis and visualisation was done in R [78] v4.1.3, using the packages ggplot2 [79, p. 2], dplyr [80] v1.0.9, ggplotly [81] v3.4.1, patchwork [82] v1.1.1, gggenomes [83] v0.9.5.9000, ggraph (github.com/thomasps85/ggraph) v2.0.5, and tracklayer [84] v1.60.0.

## RESULTS

A bioinformatic pipeline was developed that identifies novel MGEs and other biologically relevant entities from metagenomic data. The pipeline does not rely on existing databases, but instead detects transfer of candidate MGEs based on sequences newly introduced into evolving communities. As such sequences are of 'foreign' nature, we refer to these as 'xenotypic sequences': the pipeline is termed xenoseq (<https://github.com/bramvandijk88/xenoseq>). For a full description of the bioinformatic pipeline see methods and Supplementary Material I and II.

Xenoseq was first benchmarked using simulated mock communities (Supplementary Materials II), and then applied to datasets from Quistad et al. The data by Quistad et al. is of particular interest, because the experimental design allows for dissemination of MGEs within *horizontal communities*, without allowing the migration of microbial cells (Fig. 1a). As short-read metagenome samples were prepared from all communities at various timepoints, it is possible to identify sequences that are not native to the evolved communities, that is, candidates for transfer of phages, plasmids, or other MGEs. However, not all sequences that newly appear are due to horizontal transfer of the nanobiome between allopatric communities. False positives can emerge when a sequence increases in abundance within sympatric communities, which can happen with rare species that are initially below the metagenomic detection limit (Fig. 1b). This false positive signal due to demographic change ought to apply equally to *horizontal* and *vertical* communities. Hence, after identifying newly emerged 'unique sequences' (xenoseq\_find, Fig. 1c), further evidence for transfer is sought by aligning these contigs against sequences from allopatric communities (xenoseq\_link, Fig. 1d). Note that false positive can still occur when parallel communities are very similar in composition, however in the Quistad experiment communities are diverse in composition (see ref. [52]). Then, as a final step, read mapping provides further insight into the origin and dissemination of xenotypic contigs (see Fig. 1e).

Note that xenoseq can in principle be applied to datasets that deviate from this particular experimental design. While Quistad et al. use glass mesocosms, mice would equally suffice when one is interested in the evolution of gut microbiomes. The experimental design could reveal selection pressures experienced by such natural communities, where the movement of MGEs (and the genes they carry as cargo) reflects what traits are relevant for community function under those conditions. Additionally, one could expose evolving communities to particular selection pressures, e.g. for antimicrobial resistance, to discover new MGEs that encode traits relevant under that condition. The only requirement is that longitudinal samples are taken from parallel communities which undergo exchange of MGEs or other biologically relevant entities.

### Horizontal compost mesocosms are enriched in xenotypic sequences

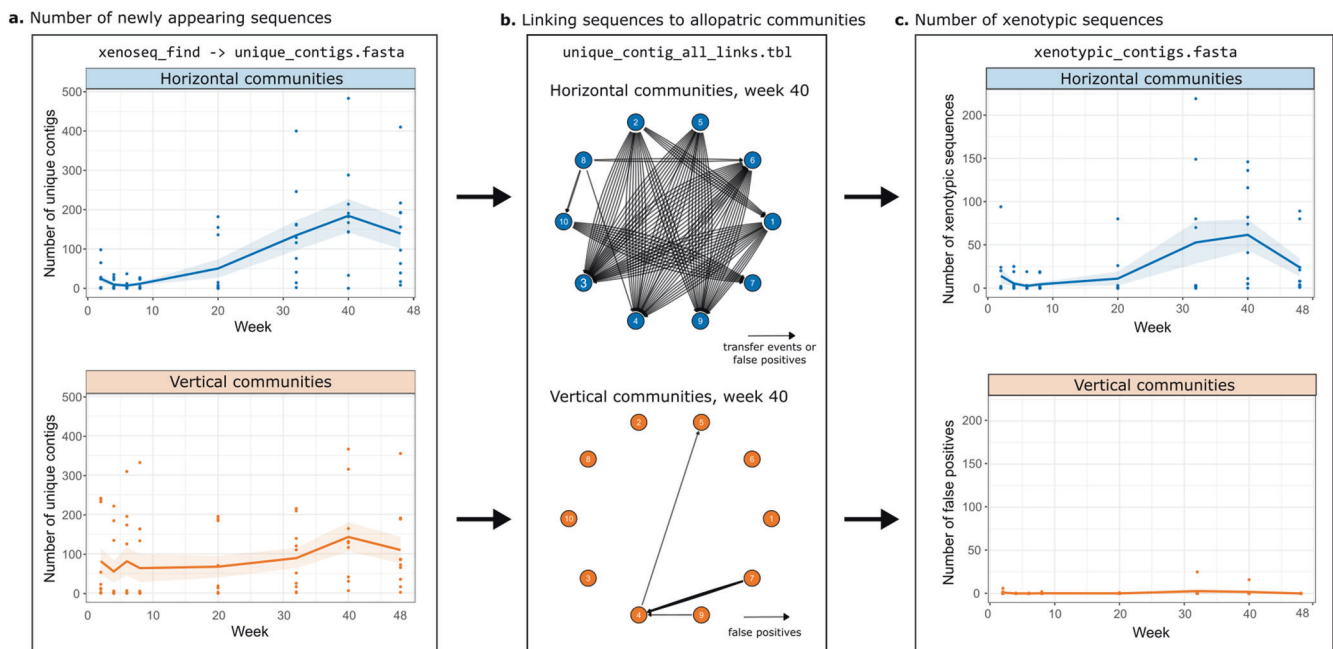
The mesocosms from the study by Quistad et al. represent a 'challenging case' for the xenoseq pipeline, because of the unprecedented diversity and abundance of rare types. As mentioned above, shifts in abundance of rare types (as observed by Quistad et al.) can generate false positives (Fig. 1b). As a control, we therefore ran xenoseq on vertical communities, which are 'closed' and unaffected by movement of MGEs from allopatric communities. Indeed, we found that unique sequences are present in both horizontal and vertical communities (Fig. 2a, a total of 5617 and 6883 sequences respectively). However, only horizontal communities contained unique contigs that could be linked to allopatric communities (Fig. 2b). In total, 1756 (31.2%) of contigs in horizontal communities could be linked to an allopatric 'donor' community, versus only 58 (0.8%) of contigs from vertical communities. The remaining 58 contigs may be spuriously linked to allopatric communities due to overlap between ancestral

compost communities. While all derived compost communities showed evidence of DNA sequences that were amplified over time, evidence for allopatric origins of these sequences was found exclusively in horizontal communities.

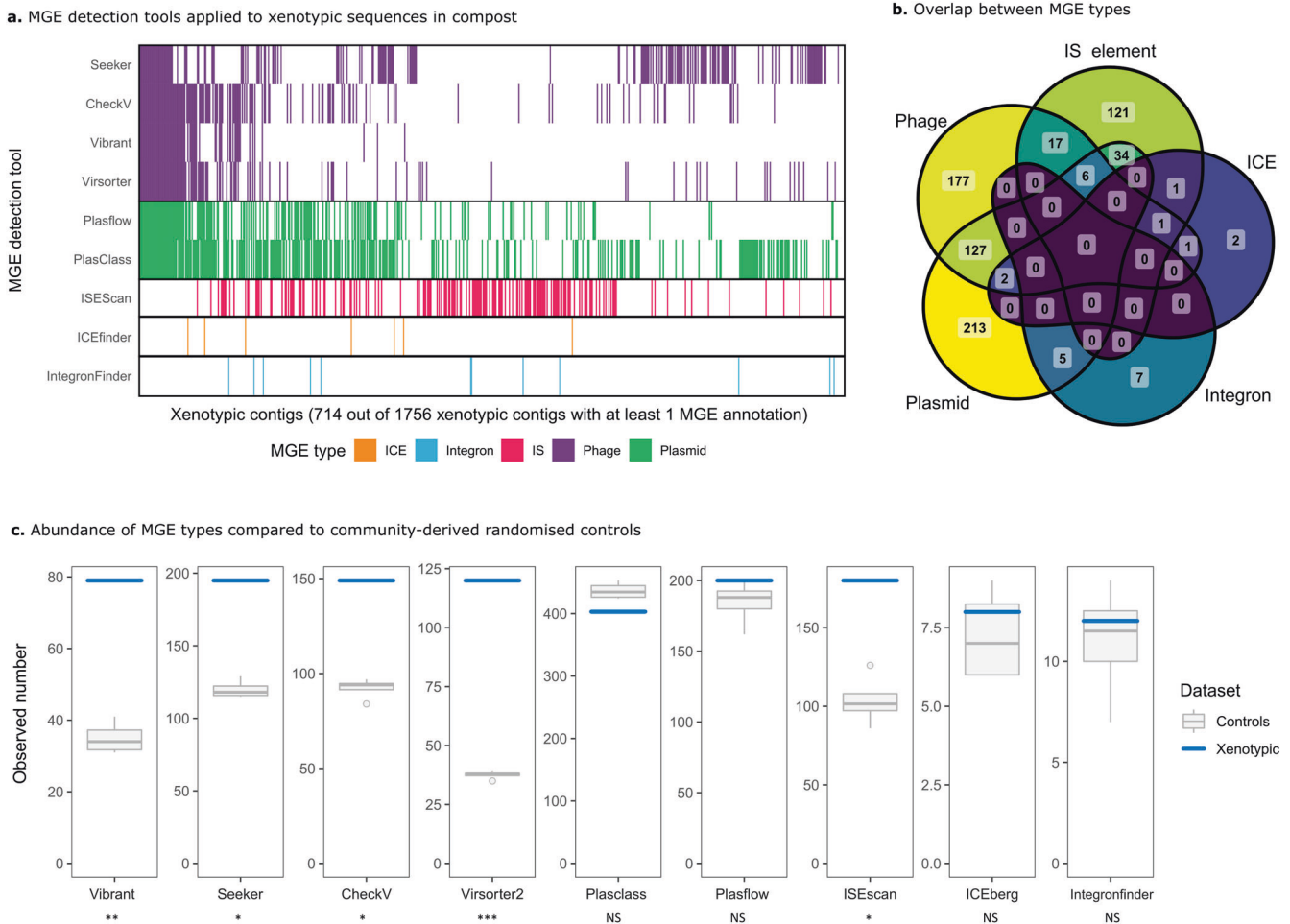
A possible confounding factor in the detection of xenotypic sequences arises if naked DNA included in the MGE cocktail were to remain intact during the 14-day incubation period thus ending up in the metagenomic data. To test this, we added exogenous gDNA from *Escherichia coli* to the communities and monitored its persistence over time (see Methods). In the presence of the microbial communities, *E. coli* gDNA was no longer detectable within 48 h after addition (Supplementary Fig. 3), while it remained detectable throughout the 14-day experiment in community-free controls. Furthermore, with DNA concentrations of 0.67 ng/ml in the initial MGE cocktails, we consider it highly unlikely that these sequences would significantly contribute to the community metagenome samples after 14 days. Taken together, these data demonstrate that in order for xenoseq to predict horizontal transmission, DNA sequences must first persist degradation, and then be sufficiently amplified within the recipient community. Such amplification indicates some form of selection for the candidate MGE, e.g. because it is a selfish element or conveys a substantial fitness benefit to a host.

### Xenotypic sequences are enriched for phages and IS-elements

The MGE cocktail that was distributed among mesocosms could contain bacteriophages, plasmids, naked DNA, membrane vesicles, and potential other (unknown) vehicles of transfer. The 1756 xenotypic sequences were interrogated using a variety of MGE detection tools, which provided predictions as to whether the sequences are viral [44–46, 74], plasmid<sup>45,46</sup>, IS-carrying<sup>49</sup>, ICE-carrying<sup>47</sup>, or integron-carrying<sup>48</sup>. We found that 714 (40.6%) of xenotypic sequences were identified by at least one MGE prediction



**Fig. 2 Horizontal communities are enriched in xenotypic sequences.** **a** Unique contigs within horizontal (blue) and vertical (orange) communities identified by *xenoseq\_find*. The solid lines represent the mean across 10 communities, shaded areas represents the standard error. **b** Aligning contigs identified by *xenoseq\_find* to ancestral samples allows linking of donor-acceptor pairs (example network shown for horizontal and vertical communities at week 40, also see Supplementary Figs. 1+2). For horizontal communities, links in the network represent candidate elements that are horizontally transferred. For vertical communities, the links instead represent false positives, as transfer from allopatric communities is excluded by the experimental design. **c** The number of xenotypic contigs (sequences linked to at least one donor community) identified in horizontal communities is shown in blue, and false positives observed in vertical communities is shown in orange. Dots represent 10 different replicate communities sampled at each time point, the solid lines represent the mean across communities, and the shaded area represents the standard error.



**Fig. 3 MGE detection tools reveal a diverse set of mobile elements in compost communities.** **a** Various tools were used to predict MGEs including phages, plasmids, IS-elements, ICEs, and integrons in xenotypic sequences. On the x-axis are 714/1756 candidate MGE contigs with at least one MGE prediction. MGE tools are depicted on the y-axis. Coloured bars indicate that the contig was predicted by the corresponding tool. **b** A Venn diagram shows substantial overlap between MGEs, especially phages and plasmids. This may indicate interactions or hybridisation of phages and plasmids, but could also be the result of erroneous assignment by one of the tools. **c** The number of MGE annotations derived from xenotypic sequences is compared with four controls. Controls consist of contigs with similar lengths to the xenotypic contigs, but are sampled from the whole compost community metagenome. A modified T-test (Crawford-Howell [110]) was used to infer whether phages, plasmids, IS-elements, ICEs, or Integrons were over- or under-represented among xenotypic sequences (\* =  $p$  value < 0.05, \*\* =  $p$  value < 0.01, \*\*\* =  $p$  value < 0.001). See precise p-values, including those corrected for multiple testing using a Bonferroni adjustment, in Supplementary Table 1, sheet 3.

tool (see Fig. 3a). Considerable overlap was found between phages and plasmids (Fig. 3b), but also between other elements. These results indicate the existence of recombined or hybrid MGEs like phage-plasmids<sup>29</sup>. They may also highlight challenges of unambiguously predicting the MGE type from sequence using the selected tools.

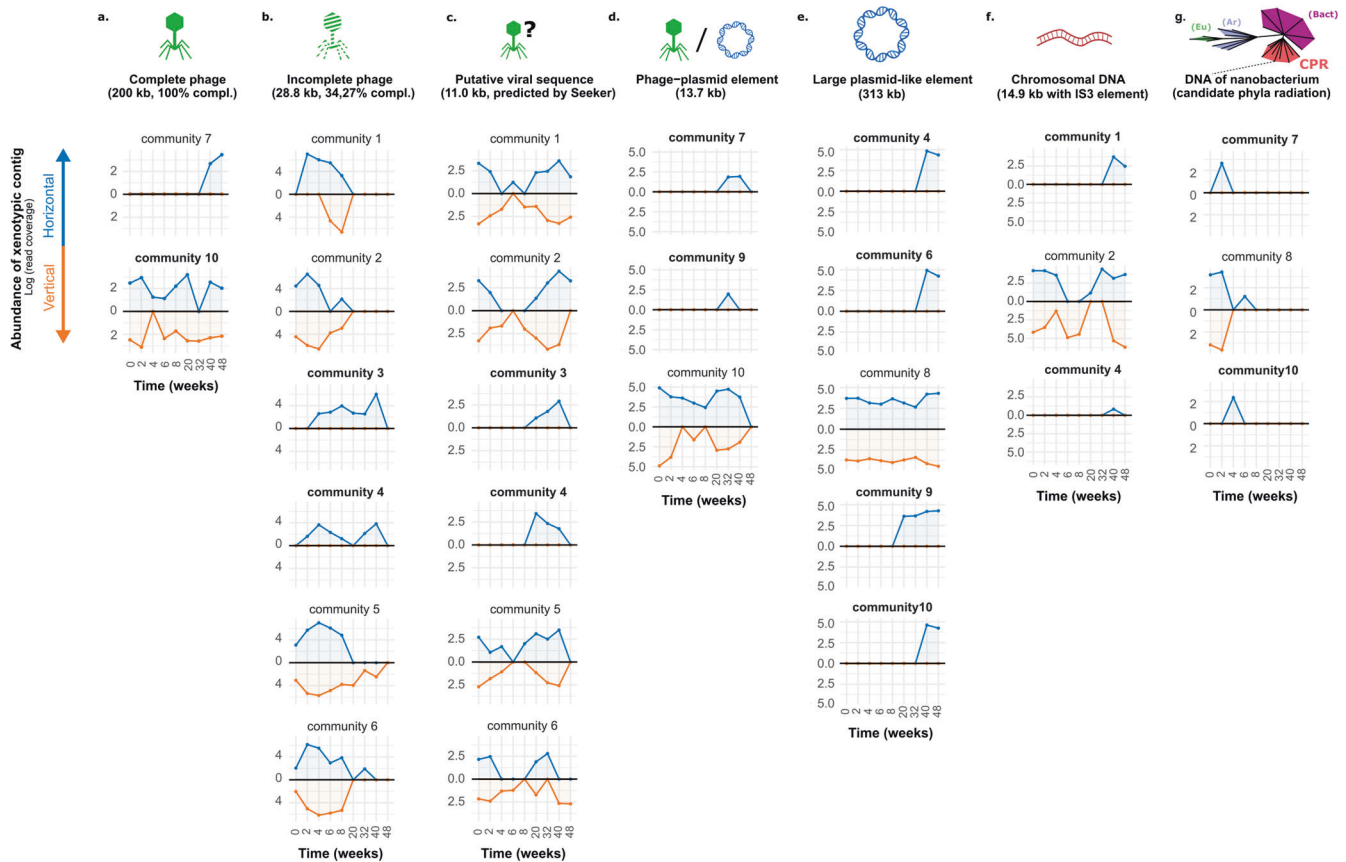
To test whether xenotypic sequences are enriched in MGEs, these data were compared to an arbitrary set of sequences with a similar length distribution, sampled from the metagenomic contigs. Note that these community samples are still expected to contain many MGEs, but perhaps less than the xenotypic sequences. Indeed, phages and IS-elements appear to be over-represented among xenotypic sequences, whereas the numbers of plasmids, ICEs, and integrons are not significantly different (Fig. 3c). When corrected for multiple testing, phages remained significantly overrepresented but this did not hold for IS elements (see Supplementary Table 1, Sheet 3 for all  $p$  values). In summary, because canonically selfish elements appear enriched, the results may further corroborate the importance of sequence amplification after introduction into a new community.

Interestingly, after applying nine state-of-the-art tools for predicting five categories of MGEs, 1042 xenotypic contigs (59.4%) remained unidentified.

#### Movement of xenotypic sequences between communities

To investigate the dissemination and dynamics of xenotypic contigs, xenoseq was used to map reads from all communities to the xenotypic contigs representing candidate MGEs. Figure 4 shows the results for seven selected examples: a complete 200 kb phage (Fig. 4a), an incomplete but abundant phage sequence (Fig. 4b), a putative viral sequence exclusively predicted by Seeker (Fig. 4c), an element predicted as both phage and plasmid unanimously (Fig. 4d), a large 313 kb plasmid (Fig. 4e), and a stretch of apparent chromosomal DNA adjacent to an IS3 element (Fig. 4f).

Xenoseq also revealed transfer of chromosomal bacterial DNA fragments lacking MGE predictions, many of which were annotated as *Candidatus Saccharibacterium* (Fig. 4g), a member of the candidate phylum radiation (CPR) bacteria first observed in human oral cavities [85]. Although this nanobacterium could



**Fig. 4** Tracing selected xenotypic sequences across communities. Results of *xenoseq\_trace* for seven xenotypic sequences; **a** a phage with 100% read coverage as reported by CheckV (abbreviated as compl), **b** an incomplete yet highly abundant phage sequence (at certain time points covered by  $>10^6$  reads), **c** a putative viral sequence exclusively predicted by seeker, **d** a sequence reported to be both phage and plasmid by all relevant MGE tools, **e** a large plasmid that successfully establishes in four communities, **f** a putative chromosomal region flanked by a transposable element, and **g** a sequence not predicted to be an MGE, but annotated by CAT as being *Candidatus Saccharibacterium*. For all panels, abundances (y axes, average read coverage) are shown across communities over all time points (x axes). Abundance in horizontal communities is shown in blue, whereas abundance in vertical communities is shown on the opposing axes in orange. Communities in which sequence are unique to the horizontal regime are shown in bold. Communities in which the contig was not observed are omitted for clarity. The full interactive data set is available as Supplementary Material.

indeed be sufficiently small to pass through the  $0.2\ \mu\text{m}$  filter, members of this species appear in both horizontal communities (community 1, 3, 4, 5, 7, 8, 9, and 10) as well as vertical communities (community 1, 3, 4, 7, 8, 9, 10), with varying estimates of genome completion (see Supplementary Table 1). The best representative MAG of this species was observed in HC10, which has a small genome (818 kb) that was nevertheless estimated to be 97% complete, and indeed encoded many of the relevant housekeeping genes (e.g. *gyrA*, *recA*, *polA*, *topA*, *rpoB*, and a single 16S ribosomal RNA gene). Comparing the 16S sequences with those from recently published CPR bacteria and DPANN archaea from groundwater samples [73], shows it clusters together with other CPR bacteria annotated as *Candidatus Saccharibacterium* (see Supplementary Fig. 4). Interestingly, alignment-free estimates of average nucleotide identity (fastANI v1.33) revealed that the (near-complete) nanobacterial MAGs from the communities are highly similar (up to 99.4% ANI). However, fastANI with respect to the groundwater CPR bacteria yielded no results, which occurs when the genomes compared are too dissimilar. Taken together, these analyses suggests that the nanobacterium in our compost mesocosms are closely related to, but distinct from, these previously published bacterial MAGs.

Co-occurrence analysis (see Supplementary Fig. 5a) further reveals a potential connection between the identified nanobacterium and the genus *Cellvibrio*. Tracking the abundance of both

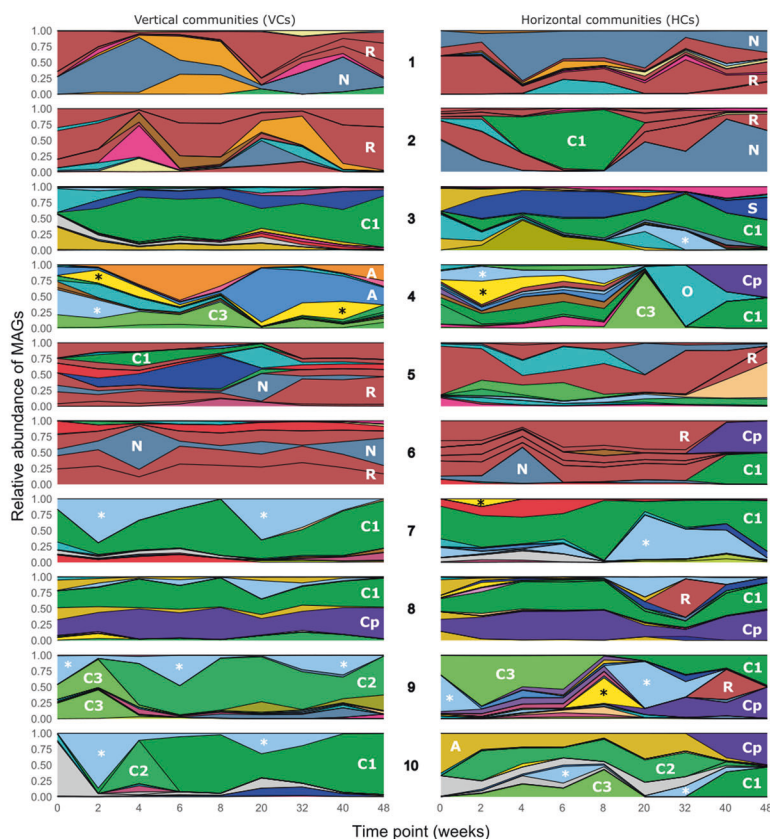
players over times suggests a potential boom-and-bust dynamic, punctuated by and long periods of stasis (Supplementary Fig. 5b). Such dynamics may be indicative of a pathogenic lifestyle.

### Compost mesocosms form distinct community types

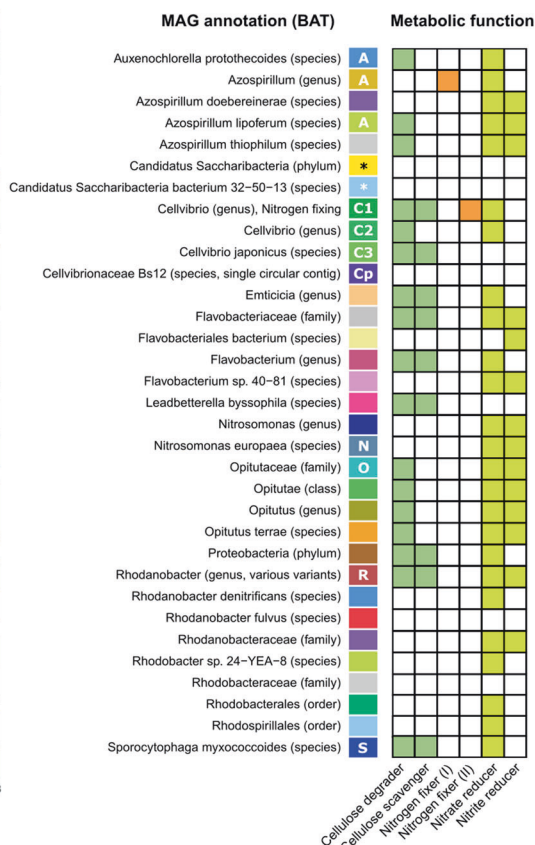
Thus far, we have illustrated that treating evolving microbial communities with an ‘MGE cocktail’ derived from all communities promotes the movement of various MGEs, and even nanobacteria between communities. We have also shown that xenotypic contigs are especially enriched in SGEs, suggesting the importance of independent sequence amplification for the survival of MGEs after introduction into allopatric communities. In the following sections, we examine the ecological and evolutionary consequences of this treatment. To investigate whether and how the horizontal communities are distinct from the vertical communities, we studied the abundance of MAGs. These MAGs were assigned a taxonomic classification using the Bin Annotation Tool (BAT). When taxonomic rank could not be determined reliably due to conflicting ORFs, a higher-order taxonomic rank was assigned instead, ensuring all MAGs had a robust classification.

Each MAG was screened for various genes related to two metabolically relevant functions: cellulose degradation and nitrogen metabolism (see Methods and Supplementary Table 1, Sheet 5). The relative abundance of dominant MAGs is shown

a. Temporal trajectories of dominant MAGs



b. Metabolic functions of MAGs



c. Cp is a large plasmid-like mobile element with viral genes and a putative conjugative region

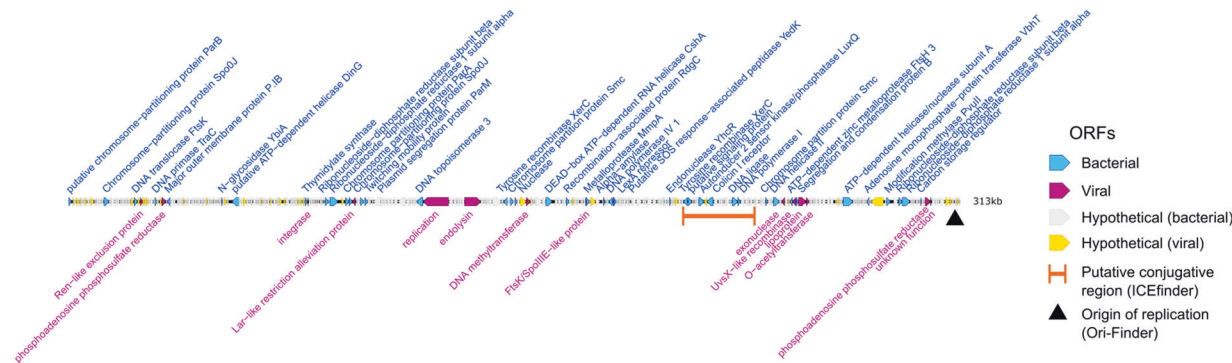


Fig. 5 Horizontal communities show repeated shifts in dominant MAGs. a For all 20 replicate experiments (10 horizontal, 20 vertical), abundance (relative read coverage) is shown for MAGs over time. For clarity, only MAGs that are highly abundant across multiple time points are shown (representing on average 59.6% of the total community), and abundance is given as the fraction of reads mapping only to these particular MAGs. A subset of MAGs are labelled with a letter (C1, R, N, Cp, etc.) corresponding to the legend in (b). b Metabolic functions assigned to each MAG. MAGs encoding putative endoglucanases, cellobiosidases, nitrogenases (I and II), and nitrate/nitrite reductase were assigned roles to cellulose degraders, cellulose scavengers, nitrogen fixers, nitrate reducers, and nitrite reducers, respectively (see Methods). A score was calculated based on the number of significant hits to these protein sequences, and MAGs with a score greater than 0.1 are coloured in the heatmap. c A circular, 313 kb mobile element (Cp) that emerges in four independent horizontal communities. The element contains both plasmid-like and phage-like traits.

for each community in Fig. 5a (for an interactive graph of all MAGs, see Supplementary Files). The MAGs are dominated by either *Rhodanobacter* (shown in brown in Fig. 5a) or *Cellvibrio* lineages (shown in green). Members of both these genera are able to degrade cellulose, the sole carbon source in the mesocosms (Fig. 5b). As many other MAGs do not have this ability, *Rhodanobacter* and *Cellvibrio* appear to occupy a similar niche and are the primary degraders of cellulose, which could explain

why they appear to be mutually exclusive. The two community types furthermore have distinct lineages that cohabit the mesocosms. For example, the communities dominated by *Rhodanobacter* lineages often coexist with *Nitrosomonas europaea* (steel blue), which can reduce nitrate and nitrite (Fig. 5b). Communities dominated by *Cellvibrio* instead frequently host species of above-mentioned *C. Saccharibacteria* (denoted by an asterisk in Fig. 5a, b).

Downloaded from https://academic.oup.com/ismecommun/article/3/1/901/7584896 by MPI Evolutionary Biology user on 22 February 2024



### Horizontal communities show more changes in primary degraders

As can be seen from Fig. 5a, vertical communities (VCs, left) are relatively stable in the composition of dominant MAGs, establishing either with *Cellvibrio* or *Rhodanobacter* as primary cellulose degraders. Horizontal communities (HCs) however show rapid shifts in these primary degraders. For example, HC2 initially establishes with *Rhodanobacter* (brown) as the primary degrader. However, during week 4 to 8, a *Cellvibrio* lineage (C1, green) becomes more dominant relative to other MAGs. Finally, *Rhodanobacter* re-emerges as the primary degrader. *Vice versa*, HC8 and HC9 are initially established as *Cellvibrio*-communities, and show a transient appearance of a *Rhodanobacter* lineage after 32/40 weeks.

While the above-mentioned disruptions caused by the MGE cocktails are transient, *Cellvibrio* C1 in HC6 appears to completely overturn the ecosystem structure that was established during the preceding 32 weeks, although it is unclear whether this disruption is transient since we do not have measurements after 48 weeks. *Cellvibrio* C1 also emerges in three other mesocosms independently (in HC9 at week 8, and HC4 and HC10 at week 32), and its emergence is always accompanied by a MAG containing only a single 313 kb contig (shown in purple in Fig. 5a, henceforth called Cp). This contig is of circular nature, and is in fact the same plasmid-like element earlier identified by xenoseq (Fig. 4e). The large mobile element carries plasmid partitioning proteins (e.g. ParB and ParM) typically associated with low copy-number plasmids [86, 87] (see Fig. 5c), which is consistent with it co-occurring with *Cellvibrio* C1 in an approximate 1:1 ratio. In addition to features that are akin to large plasmids, Cp also carries many ORFs associated with phages as evident in matches to the viral PHROGs-database<sup>66</sup> (for example, integrase and endolysin). Finally, Cp carries a putative conjugative region identified by ICEfinder. As this plasmid is present in earlier time points only in VC/HC8, this suggests this community as the donor. We hypothesise that Cp transfers horizontally, enabling *Cellvibrio* C1 (the putative host) to displace the previously established ecosystem structure.

### Ammonia production appears elevated in Cp+ communities

Because cellulose is the only exogenously provided carbon source in the mesocosms, the ability to degrade cellulose is not surprisingly found in many MAGs (see Fig. 5b). While all the *Cellvibrio* lineages have the ability to degrade cellulose, one particular *Cellvibrio* (C1) is unique in also carrying a nitrogen fixing enzyme (flavoprotein, EC 1.19.6.1). The ability to fix nitrogen is likely important, because 1 mM ammonium chloride from the M9 medium added every 14 days is the only source of exogenously supplied nitrogen. We hypothesised that the Cp plasmid may have regulatory impacts on C1 host metabolism, perhaps favouring fixation of nitrogen.

In the study by Quistad et al., ammonia production was measured at the end of the experiment ( $T = 48$  weeks). By plotting the abundance of *Cellvibrio* C1 against data on ammonia production, a strong positive correlation was observed (Fig. 6a) in horizontal communities, but much less so in vertical communities. Note that C1 in most horizontal communities carries the Cp plasmid, and C1 in vertical communities does not. We suggest that this indicates that C1 may be responsible for the accumulation of ammonia, but only in the presence of the Cp plasmid. Indeed, by taking the slope of the linear regression, we found that horizontal communities show a significantly steeper correlation between C1 and ammonia production rates, especially early in the growth cycle (Fig. 6b). Finally, when focussing only on horizontal communities where Cp was absent (HC3 and HC7), no positive correlation was observed (Supplementary Fig. 6). These data suggest that Cp promotes the fixation of nitrogen by *Cellvibrio* C1.

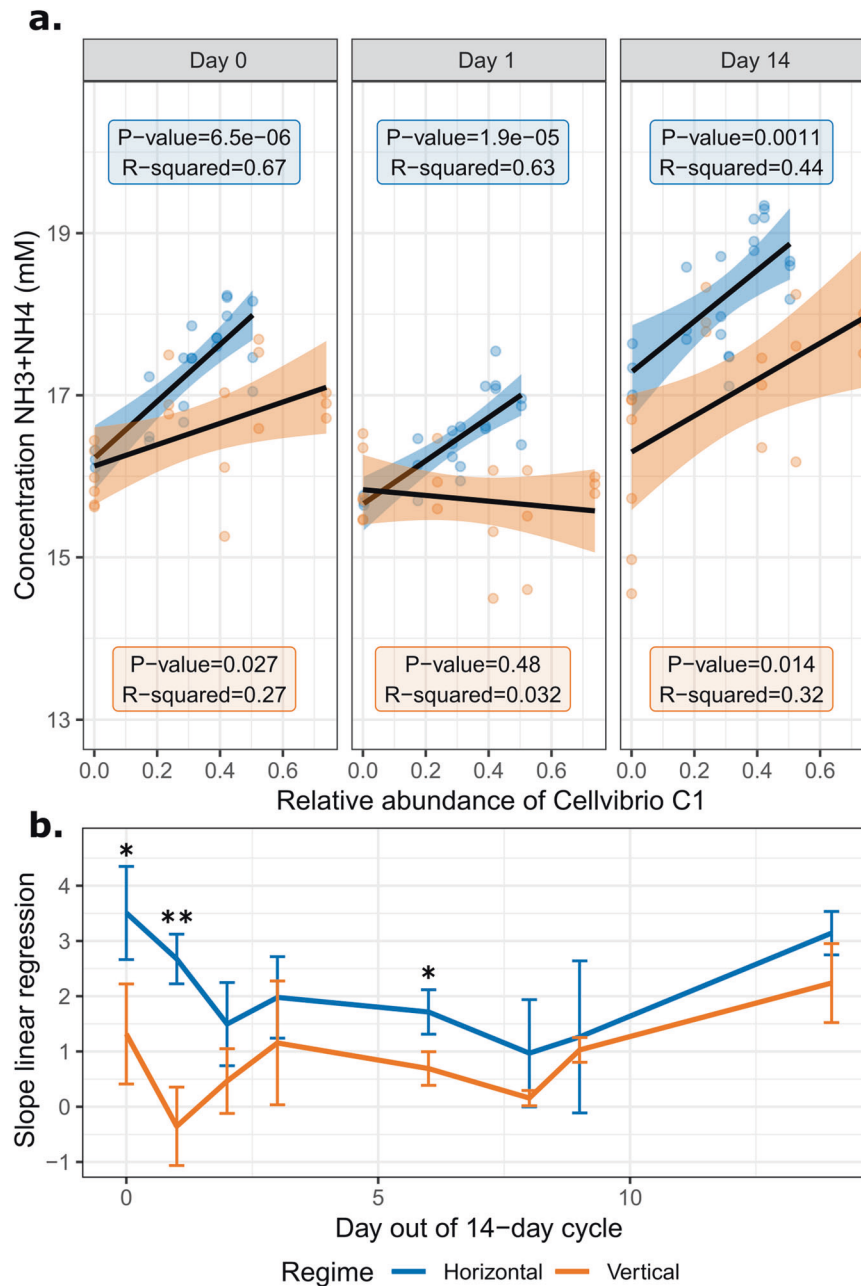
### DISCUSSION

MGEs are important determinants of microbial evolution with far-reaching effects within the context of communities. As diverse as microbiomes are, the nanobiome—the set of Darwinian entities dependent on microbial hosts—is likely even more diverse. We developed a bioinformatic pipeline called xenoseq to shine light on the nanobiome, inspired by experiments from Quistad et al. Using time-resolved metagenomic data, xenoseq can distinguish between two sources of novel sequences, (i) those that arise due to local demographic changes, and (ii) those that are the consequence of horizontal transmission of nanoscale entities between parallel communities. We show that the latter category of *xenotypic sequences* requires amplification after introduction into allopatric communities. We found that xenotypic sequences were especially enriched in selfish genetic elements (phages and IS elements). While other MGEs such as plasmids and ICEs were not enriched, we did find some unanticipated players to transfer horizontally via the filtrate, including a 313 kb plasmid and a CPR nanobacterium. Taken together, our data shows that the pipeline can successfully identify a broad range of interesting MGEs, without any prior knowledge as to their identity.

The experimental strategy of Quistad et al. and our bioinformatic pipeline can be applied to any microbial community and can accommodate specific selection pressures, such as antimicrobial resistance [88, 89], heavy metal resistance [90], or bioremediation capacity [91]. Furthermore, even without any evident selection pressures, the method can provide direct evidence of selection pressures experienced by communities, through identification of traits transferred horizontally. Quistad et al. observed the enrichment of nitrogen metabolism genes in horizontal communities, which in this study, we have been able to link to the proliferation of a large 313 kb plasmid-like sequence. Given its apparent indirect role in nitrogen fixation, we hypothesise that amplification of this element confers a fitness advantage (Fig. 6). The mechanism of horizontal transfer is unknown, but is unlikely to be via naked DNA, leading to the possibility that the element maybe packaged inside membrane vesicles (MVs) [92–95], as recently shown in *Klebsiella pneumoniae* [96].

Besides the movement of MGEs, our study also suggests competition between ecologically relevant species is enhanced in horizontal communities. This enhanced competition may be due to ‘kill the winner’ dynamics [97], whereby phages preferentially reduce the abundance of established microbial species, while at the same time improving their own evolutionary fate, making niche space available for competitors. A similar dynamic was observed in a recent study that illustrates how the induction of prophages can significantly impact the assembly of chitin-degrading communities [98]. Notably, these shifts in abundances could exacerbate the problem with false positives as we observed in this study. Taken together, we argue that it is important to improve understanding of how MGEs influence the competition between microbial species—for example in societally relevant systems such as the human gut [99] or plant rhizospheres [100, 101]—which may furthermore enable us to make better evolutionary predictions in the future [102].

An unexpected finding of our study was the apparent transmission of a CPR nanobacterium of the phylum *Saccharibacteria*. Indeed, close relatives of this organism are not much larger than a phage<sup>75</sup>, and the reduced genome (~818 kb) is indicative of a symbiotic relationship with another species. The nanobacterium was shown to cluster closely with CPR bacteria observed in groundwater ecosystems [73], where many ultrasmall cells are shown to stick to the cell surface of larger bacteria. Similar patterns have been observed for *Vampirococcus lugossi*, which are also thought to be epibionts that attach to photosynthetic bacteria [103]. Rather than epibionts, however, we suggest that the nanobacterium in our study may be an intracellular parasite, as the boom-and-bust dynamics shown punctuated by



**Fig. 6** *Cellvibrio C1* appears to fix more nitrogen in horizontal (Cp+) communities. **a** The relative abundance of *Cellvibrio C1* was determined by the proportion of reads from the sample at week 48 mapping against this MAG, which was then plotted against ammonia production measurements from Quistad et al. [52]. *Cellvibrio C1* was present in five vertical communities and seven horizontal communities, yielding a total of 12 abundance values for *Cellvibrio C1*, plotted on the x-axis. For each of these values, three technical replicate measurements of ammonia were performed at several time points in the 14-day cycle. The y-axis shows the measured ammonia concentrations (NH<sub>3</sub> + NH<sub>4</sub>)<sup>2</sup>. A linear regression model was fitted to the data points. For day 0, 1 and 14, these fits are shown with their corresponding *p* values and *R*<sup>2</sup> values (see Supplementary Fig. 6 for all days). **b** The slope of the linear regression is plotted for all days in the bottom panel, revealing that horizontal communities consistently show a steeper slope, suggesting that the large Cp plasmid promotes nitrogen-fixation by the *Cellvibrio C1* MAG. Error bars denote the standard deviation of the slope across three technical replicates of ammonia production measurements. The asterisks indicate significant differences (Student's T test) between the horizontal and vertical communities at that time point (\* = *p* value < 0.05, \*\* = *p* value < 0.01).

and long periods of stasis shown in Supplementary Fig. 5b are reminiscent of predatory dynamics as observed for *Bdellovibrio bacteriovorus*. Our co-occurrence analysis (see Supplementary Fig. 5a) revealed a potential connection with a bacterial lineage belonging to the genus *Cellvibrio*. Identification of this nanobacterium, and the newly generated hypothesis of a host-parasite relationship, highlight an additional power to our approach. Excitingly, many species of *Cellvibrio* can be isolated and cultured,

paving the way for future studies on the nanobacterium and its interactions with the host.

A number of annotation-free methods to track HGT in microbial ecosystems have recently been developed, relying for example on differential read coverage [104], assembly graphs [105, 106], discordant read pair mapping [107], or Hi-C metagenomics [88, 108, 109]. However, compared to our approach, these methods do not provide insights into the dynamics and functional

consequences of MGE transmission, and how this flux of DNA shapes microbial communities. Our strategy combines experimental intervention, metagenomics and bioinformatic analysis to discover elements capable of transmission as nanoparticles and subsequent amplification, without prior knowledge of the elements themselves. In other words, using our method, it becomes possible to interrogate the nanobiome free from pre-conceived notions of the identity of its members.

While we illustrate the discovery of novel and interesting biological entities, there is also considerable potential for false positives and false negatives. As partially addressed in this study, metagenomic detection limits make it extremely difficult to say for certain whether a sequence was ancestrally present in a community, or whether it was acquired from an allopatric community. Moreover, it is possible that treating communities with cocktails of MGEs has indirect impacts on the community composition, which if resulting in rare species becoming more abundant, would further exacerbate this problem. While our pipeline tries to reduce these problems by linking sequences to their community of origin, this strategy trades off with false negatives, as genuine horizontal gene transfer events may be erroneously discarded. While we argue that advancements like Hi-C metagenomics may help to alleviate some of these difficulties, we also suggest that future studies apply the experimental protocol to simpler (synthetic) communities, making it easier to distinguish foreign DNA from rare sequences. Note however, that such simplification comes with its own trade off, as it hampers our ability to study the processes that are inherent to complexity itself [38]. We therefore argue that in order to further our understanding of complex microbiomes, it is also important to embrace complexity—with all the potential sources of noise that entails—and identify the simple rules that drive the eco-evolutionary dynamics of microbial communities.

### Concluding remarks

Our work confirms earlier conclusions that microbial ecosystems are greatly influenced by the flux of DNA created by MGEs, even when these fluxes are initially subtle. Hence, we argue that understanding the nanobiome—the zoo of Darwinian entities much smaller than bacteria—is crucial to understanding microbial ecology, and how these systems eventually scale up to impact entire ecosystems.

### DATA AVAILABILITY

Raw sequencing data are taken from Quistad et al., and are publicly available online (<https://www.mg-rast.org/mgmain.html?mgpage=project&project=mgp18485>). MAGs, interactive datasets, and R scripts for designing the mock communities are published on Zenodo (<https://doi.org/10.5281/zenodo.7589193>). Any further requests for data can be sent to [rainey@evolbio.mpg.de](mailto:rainey@evolbio.mpg.de).

### REFERENCES

- Kolstø AB. Dynamic bacterial genome organization. *Mol Microbiol.* 1997;24:241–8. <https://doi.org/10.1046/j.1365-2958.1997.3501715.x>.
- Snel B, Bork P, Huynen MA. Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome Res.* 2002;12:17–25. <https://doi.org/10.1101/gr.176501>.
- Puigbò P, Lobkovsky AE, Kristensen DM, Wolf YI, Koonin EV. Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes. *BMC Biol.* 2014;12:66. <https://doi.org/10.1186/s12915-014-0066-4>.
- Garoña A, Dagan T. Darwinian individuality of extrachromosomal genetic elements calls for population genetics tinkering. *Environ Microbiol Rep.* 2021;13:22–6. <https://doi.org/10.1111/1758-2229.12894>.
- Chen J, et al. Genome hypermobility by lateral transduction. *Science.* 2018;362:207–12. <https://doi.org/10.1126/science.aat5867>.
- Humphrey S, et al. Bacterial chromosomal mobility via lateral transduction exceeds that of classical mobile genetic elements. *Nat Commun.* 2021;12:1. <https://doi.org/10.1038/s41467-021-26004-5>.

- Dutilh BE, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun.* 2014;5:1. <https://doi.org/10.1038/ncomms5498>.
- Garmaeva S, et al. Stability of the human gut virome and effect of gluten-free diet. *Cell Rep.* 2021;35:109132.
- de Jonge PA, et al. Gut virome profiling identifies a widespread bacteriophage family associated with metabolic syndrome. *Nat Commun.* 2022;13:1. <https://doi.org/10.1038/s41467-022-31390-5>.
- Siguier P, Filée J, Chandler M. Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol.* 2006;9:526–31. <https://doi.org/10.1016/j.mib.2006.08.005>.
- Touchon M, Rocha EPC. Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol.* 2007;24:969–81. <https://doi.org/10.1093/molbev/msm014>.
- Wu Y, Aandahl RZ, Tanaka MM. Dynamics of bacterial insertion sequences: can transposition bursts help the elements persist? *BMC Evol Biol.* 2015;15:288. <https://doi.org/10.1186/s12862-015-0560-5>.
- Hall JP, Wood AJ, Harrison E, Brockhurst MA. Source–sink plasmid transfer dynamics maintain gene mobility in soil bacterial communities. *Proc Natl Acad Sci.* 2016;113:8260–5. <https://doi.org/10.1073/pnas.1600974113>.
- Hülter NF, Wein T, Effe J, Garoña A, & Dagan T. Intracellular competitions reveal determinants of plasmid evolutionary success. *Front Microbiol.* 2020;11. <https://www.frontiersin.org/articles/10.3389/fmicb.2020.02062>. Accessed 18 Jul 2022.
- Garoña A, Hülter NF, Romero Picazo D, Dagan T. Segregational drift constrains the evolutionary rate of prokaryotic plasmids. *Mol Biol Evol.* 2021;38:5610–24. <https://doi.org/10.1093/molbev/msab283>.
- Burrus V. Mechanisms of stabilization of integrative and conjugative elements. *Curr Opin Microbiol.* 2017;38:44–50. <https://doi.org/10.1016/j.mib.2017.03.014>.
- Carraro N, Burrus V. The dualistic nature of integrative and conjugative elements. *Mob Genet Elem.* 2015;5:98–102. <https://doi.org/10.1080/2159256X.2015.1102796>.
- Colombi E, Straub C, Künzel S, Templeton MD, McCann HC, Rainey PB. Evolution of copper resistance in the kiwifruit pathogen *Pseudomonas syringae* pv. actinidiae through acquisition of integrative conjugative elements and plasmids. *Environ Microbiol.* 2017;19:819–32. <https://doi.org/10.1111/1462-2920.13662>.
- Delavat F, Miyazaki R, Carraro N, Pradervand N, van der Meer JR. The hidden life of integrative and conjugative elements. *FEMS Microbiol Rev.* 2017;41:512–37. <https://doi.org/10.1093/femsre/fux008>.
- Bertels F, Rainey PB. Within-genome evolution of REPINs: a new family of miniature mobile DNA in bacteria. *PLOS Genet.* 2011;7:e1002132. <https://doi.org/10.1371/journal.pgen.1002132>.
- Bertels F, Rainey PB. Ancient Darwinian replicators nested within eubacterial genomes. *BioEssays.* 2023;45:2200085. <https://doi.org/10.1002/bies.202200085>.
- Gluck-Thaler E, et al. Giant starship elements mobilize accessory genes in fungal genomes. *Mol. Biol. Evol.* 2022;39:msac109. <https://doi.org/10.1093/molbev/msac109>.
- Al-Shayeb B, et al. Borgs are giant genetic elements with potential to expand metabolic capacity. *Nature.* 2022;610. <https://doi.org/10.1038/s41586-022-05256-1>.
- Ma L-J, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature.* 2010;464. <https://doi.org/10.1038/nature08850>.
- Li J, Fokkens L, van Dam P, Rep M. Related mobile pathogenicity chromosomes in *Fusarium oxysporum* determine host range on cucurbits. *Mol. Plant Pathol.* 2020;21:761–76. <https://doi.org/10.1111/mpp.12927>.
- Li J, Fokkens L, Conneely LJ, Rep M. Partial pathogenicity chromosomes in *Fusarium oxysporum* are sufficient to cause disease and can be horizontally transferred. *Environ Microbiol.* 2020;22:4985–5004. <https://doi.org/10.1111/1462-2920.15095>.
- Jurénas D, Fraikin N, Goormaghtigh F, & Van Melderen L. Biology and evolution of bacterial toxin–antitoxin systems. *Nat Rev Microbiol.* 2022;1–16. <https://doi.org/10.1038/s41579-021-00661-1>.
- Lang AS, Zhaxybayeva O, Beatty JT. Gene transfer agents: phage-like elements of genetic exchange. *Nat Rev Microbiol.* 2012;10:472–82. <https://doi.org/10.1038/nrmicro2802>.
- Stanton TB. Prophage-like gene transfer agents—novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. *Anaerobe.* 2007;13:43–49. <https://doi.org/10.1016/j.anaerobe.2007.03.004>.
- Rakonjac J, Bennett NJ, Spagnuolo J, Gagic D, & Russel M. Filamentous bacteriophage: biology, phage display and nanotechnology applications. *Curr Issues Mol Biol.* 2011;13. <https://doi.org/10.21775/cimb.013.051>.
- Haskett TL, Ramsay JP, Bekuma AA, Sullivan JT, O'Hara GW, Terpililli JJ. Evolutionary persistence of tripartite integrative and conjugative elements. *Plasmid.* 2017;92:30–36. <https://doi.org/10.1016/j.plasmid.2017.06.001>.
- Mark Osborn A, Böltner D. When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid.* 2002;48:202–12. [https://doi.org/10.1016/S0147-619X\(02\)00117-8](https://doi.org/10.1016/S0147-619X(02)00117-8).

33. Rocha E, Haudiquet M, de Sousa JM, and Touchon M. Selfish, promiscuous, and sometimes useful: how mobile genetic elements drive horizontal gene transfer in microbial populations. *EcoEvoRxiv*. 2021. <https://doi.org/10.32942/osf.io/7t2jh>.
34. Pfeifer E, Moura de Sousa JA, Touchon M, Rocha EPC. Bacteria have numerous distinctive groups of phage-plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Res*. 2021;49:2655–73. <https://doi.org/10.1093/nar/gkab064>.
35. Pfeifer E, Bonnin RA, Rocha EP. Phage-plasmids spread antibiotic resistance genes through infection and lysogenic conversion. *MBio*. 2022;13:e01851-22. <https://doi.org/10.1128/mbio.01851-22>.
36. Hall JPJ. Is the bacterial chromosome a mobile genetic element? *Nat Commun*. 2021;12:6400. <https://doi.org/10.1038/s41467-021-26758-y>.
37. Boucher Y, et al. Local mobile gene pools rapidly cross species boundaries to create endemicity within global vibrio cholerae populations. *mBio*. 2011;2:e00335–10. <https://doi.org/10.1128/mBio.00335-10>.
38. Rainey PB, Quistad SD. Toward a dynamical understanding of microbial communities. *Philos Trans R Soc B Biol Sci*. 2020;375:20190248 <https://doi.org/10.1098/rstb.2019.0248>.
39. Inkpen SAD. Adaptive regeneration across scales: replicators and interactors from limbs to forests. *Philos Theory Pract Biol*. 2021;13. <https://doi.org/10.3998/ptpbio.16039257.0013.001>.
40. Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc Natl Acad Sci*. 1995;92:8985–9. <https://doi.org/10.1073/pnas.92.19.8985>.
41. Sullivan JT, Ronson CW. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci*. 1998;95:5145–9. <https://doi.org/10.1073/pnas.95.9.5145>.
42. Burrus V, Pavlovic G, Decaris B, Guédon G. Conjugative transposons: the tip of the iceberg. *Mol. Microbiol*. 2002;46:601–10. <https://doi.org/10.1046/j.1365-2958.2002.03191.x>.
43. Gawron-Burke C & Clewell DB. A transposon in *Streptococcus faecalis* with fertility properties. *Nature*. 1982;300. <https://doi.org/10.1038/300281a0>.
44. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*. 2020;8:90 <https://doi.org/10.1186/s40168-020-00867-0>.
45. Guo J, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*. 2021;9:37 <https://doi.org/10.1186/s40168-020-00990-y>.
46. Auslander N, Gussow AB, Benler S, Wolf YI, Koonin EV. Seeker: alignment-free identification of bacteriophage genomes by deep learning. *Nucleic Acids Res*. 2020;48:e121 <https://doi.org/10.1093/nar/gkaa856>.
47. Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res*. 2018;46:e35 <https://doi.org/10.1093/nar/gkx1321>.
48. Pellow D, Mizrahi I, Shamir R. PlasClass improves plasmid sequence classification. *PLOS Comput. Biol*. 2020;16:e1007781 <https://doi.org/10.1371/journal.pcbi.1007781>.
49. Liu M, et al. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res*. 2019;47:D660–D665. <https://doi.org/10.1093/nar/gky1123>.
50. Néron B, Littner E, Haudiquet M, Perrin A, Cury J, & EPC Rocha EPC. IntegronFinder 2.0: identification and analysis of integrons across bacteria, with a focus on antibiotic resistance in *Klebsiella*. *Microorganisms*. 2022;10. <https://doi.org/10.3390/microorganisms10040700>.
51. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics*. 2017;33:3340–7. <https://doi.org/10.1093/bioinformatics/btx433>.
52. Quistad SD, Doulcier G, Rainey PB. Experimental manipulation of selfish genetic elements links genes to microbial community function. *Philos Trans R Soc B Biol Sci*. 2020;375:20190681 <https://doi.org/10.1098/rstb.2019.0681>.
53. Doolittle WF & Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature*. 1980;284. <https://doi.org/10.1038/284601a0>.
54. Baltrus DA. Exploring the costs of horizontal gene transfer. *Trends Ecol Evol*. 2013;28:489–95. <https://doi.org/10.1016/j.tree.2013.04.002>.
55. van Dijk B, Hogeweg P, Doekes HM, Takeuchi N. Slightly beneficial genes are retained by bacteria evolving DNA uptake despite selfish elements. *eLife*. 2020;9:e56801 <https://doi.org/10.7554/eLife.56801>.
56. van Dijk B, Bertels F, Stolk L, Takeuchi N, Rainey PB. Transposable elements promote the evolution of genome streamlining. *Philos Trans R Soc B Biol Sci*. 2022;377:20200477 <https://doi.org/10.1098/rstb.2020.0477>.
57. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
58. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589–95. <https://doi.org/10.1093/bioinformatics/btp698>.
59. Vasimuddin M, Misra S, Li H, & Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems, In: *Proceedings of IEEE international parallel and distributed processing symposium (IPDPS)*, 2019, p. 314–24. <https://doi.org/10.1109/IPDPS.2019.00041>.
60. Li H, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
61. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31:1674–6. <https://doi.org/10.1093/bioinformatics/btv033>.
62. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
63. Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLOS One*. 2016;11:e0163962 <https://doi.org/10.1371/journal.pone.0163962>.
64. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27:863–4. <https://doi.org/10.1093/bioinformatics/btr026>.
65. Dodt M, Roehr JT, Ahmed R, Dieterich C. FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. *Biology*. 2012;1. <https://doi.org/10.3390/biology1030895>.
66. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res*. 2017;27:824–34. <https://doi.org/10.1101/gr.213959.116>.
67. Kang DD, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*. 2019;7:e7359 <https://doi.org/10.7717/peerj.7359>.
68. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25:1043–55. <https://doi.org/10.1101/gr.186072.114>.
69. von Meijenfeldt FAB, Arkhipova K, Cambuy DD, Coutinho FH, Dutilh BE. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol*. 2019;20:217 <https://doi.org/10.1186/s13059-019-1817-x>.
70. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform*. 2010;11:119 <https://doi.org/10.1186/1471-2105-11-119>.
71. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*. 2015;12:59–60. <https://doi.org/10.1038/nmeth.3176>.
72. Sayers EW, et al. GenBank. *Nucleic Acids Res*. 2022;50:D161–D164. <https://doi.org/10.1093/nar/gkab1135>.
73. He C, et al. Genome-resolved metagenomics reveals site-specific diversity of epibiotic CPR bacteria and DPANN archaea in groundwater ecosystems. *Nat Microbiol*. 2021;6. <https://doi.org/10.1038/s41564-020-00840-5>.
74. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, & Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol*. 2021;39. <https://doi.org/10.1038/s41587-020-00774-7>.
75. Dong MJ, Luo H, Gao F. Ori-Finder 2022: a comprehensive web server for prediction and analysis of bacterial replication origins. *Genomics, Proteomics & Bioinformatics*. 2022;20:1207–13. <https://doi.org/10.1016/j.gpb.2022.10.002>.
76. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9. <https://doi.org/10.1093/bioinformatics/btu153>.
77. Terzian P, et al. PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genom Bioinform*. 2021;3:lqab067 <https://doi.org/10.1093/nargab/lqab067>.
78. R Core Team, R: a language and environment for statistical computing. <https://www.R-project.org/>.
79. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. [Online]. <https://ggplot2.tidyverse.org>.
80. Wickham H, François R, Henry L, Müller K, & Studio R. dplyr: a grammar of data manipulation. 2015. Available: <https://CRAN.R-project.org/package=dplyr>. Accessed 16 Jun 2020.
81. Interactive web-based data visualization with R, plotly, and shiny (Carson Sievert) - Li - 2021 - Biometrics - Wiley Online Library. <https://onlinelibrary.wiley.com/doi/10.1111/biom.13474>. Accessed 11 Aug 2022.
82. The Composer of Plots. <https://patchwork.data-imaginist.com/>. Accessed 11 Aug 2022.
83. Hackl T. gggenomes. Aug. 08, 2022. Available: <https://github.com/thackl/gggenomes>. Accessed 11 Aug 2022.
84. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinform*. 2009;25:1841–2. <https://doi.org/10.1093/bioinformatics/btp328>.

85. He X, et al. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc Natl Acad Sci USA*. 2015;112:244–9. <https://doi.org/10.1073/pnas.1419038112>.
86. Pinto UM, Pappas KM, & Winans SC. The ABCs of plasmid replication and segregation. *Nat Rev Microbiol*. 2012;10. <https://doi.org/10.1038/nrmicro2882>.
87. Møller-Jensen J, Borch J, Dam M, Jensen RB, Roepstorff P, Gerdes K. Bacterial mitosis: ParM of plasmid R1 moves plasmid DNA by an actin-like insertional polymerization mechanism. *Mol Cell*. 2003;12:1477–87. [https://doi.org/10.1016/S1097-2765\(03\)00451-9](https://doi.org/10.1016/S1097-2765(03)00451-9).
88. Beyi AF, Hassall A, Phillips GJ, & Plummer PJ. Tracking reservoirs of antimicrobial resistance genes in a complex microbial community using metagenomic Hi-C: the case of bovine digital dermatitis. *Antibiotics*. 2021;10. <https://doi.org/10.3390/antibiotics10020221>.
89. Geyrhofer L, Ruelens P, Farr AD, Pesce D, de Visser JAGM and Brenner N. Race to survival during antibiotic breakdown determines the minimal surviving population size. *bioRxiv*. 2022. <https://doi.org/10.1101/2022.08.04.502802>.
90. Hesse E, et al. Ecological selection of siderophore-producing microbial taxa in response to heavy metal contamination. *Ecol Lett*. 2018;21:117–27. <https://doi.org/10.1111/ele.12878>.
91. Shibasaki S, Mitri S. Controlling evolutionary dynamics to optimize microbial bioremediation. *Evol Appl*. 2020;13:2460–71. <https://doi.org/10.1111/eva.13050>.
92. Soler N, Krupovic M, Marguet E, Forterre P. Membrane vesicles in natural environments: a major challenge in viral ecology. *ISME J*. 2015;9:793–6. <https://doi.org/10.1038/ismej.2014.184>.
93. Tashiro Y, et al. Interaction of bacterial membrane vesicles with specific species and their potential for delivery to target cells. *Front Microbiol*. 2017;8:571 <https://doi.org/10.3389/fmicb.2017.00571>.
94. Biller SJ, Schubotz F, Roggensack SE, Thompson AW, Summons RE, Chisholm SW. Bacterial vesicles in marine ecosystems. *Science*. 2014;343:183–6. <https://doi.org/10.1126/science.1243457>.
95. Marcilla A & Sánchez-López CM. Extracellular vesicles as a horizontal gene transfer mechanism in *Leishmania*. *Trends Parasitol*. 2022;51471492222001817. <https://doi.org/10.1016/j.pt.2022.08.004>.
96. Altamia MA, et al. *Teredinibacter haidensis* sp. nov., *Teredinibacter purpureus* sp. nov. and *Teredinibacter franksiae* sp. nov., marine, cellulolytic endosymbiotic bacteria isolated from the gills of the wood-boring mollusc *Bankia setacea* (Bivalvia: Teredinidae) and emended description of the genus *Teredinibacter*. *Int J Syst Evol. Microbiol*. 2021;71:004627 <https://doi.org/10.1099/ijsem.0.004627>.
97. Koskella B, Brockhurst MA. Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev*. 2014;38:916–31. <https://doi.org/10.1111/1574-6976.12072>.
98. Szabo RE, et al. Historical contingencies and phage induction diversify bacterioplankton communities at the microscale. *Proc Natl Acad Sci*. 2022;119:e2117748119 <https://doi.org/10.1073/pnas.2117748119>.
99. Reyes A, Semenkovich NP, Whiteson K, Rohwer F, and Gordon JL. Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat Rev Microbiol*. 2012;10. <https://doi.org/10.1038/nrmicro2853>.
100. Wang X, et al. Phage combination therapies for bacterial wilt disease in tomato. *Nat. Biotechnol*. 2019;37. <https://doi.org/10.1038/s41587-019-0328-3>.
101. Wang X, et al. Parasites and competitors suppress bacterial pathogen synergistically due to evolutionary trade-offs. *Evolution*. 2017;71:733–46. <https://doi.org/10.1111/evo.13143>.
102. Wortel MT, et al. Towards evolutionary predictions: current promises and challenges. *Evol Appl*. 2023;16:3–21. <https://doi.org/10.1111/eva.13513>.
103. Moreira D, Zivanovic Y, López-Archilla AL, Iniesto M, & López-García P. Reductive evolution and unique predatory mode in the CPR bacterium *Vampirococcus lugosii*. *Nat Commun*. 2021;12. <https://doi.org/10.1038/s41467-021-22762-4>.
104. Durrant MG, Li MM, Siranosian BA, Montgomery SB, Bhatt AS. A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. *Cell Host Microbe*. 2020;27:140–153.e9. <https://doi.org/10.1016/j.chom.2019.10.022>.
105. Antipov D, Raiko M, Lapidus A, Pevzner PA. MetaviralSPAdes: assembly of viruses from metagenomic data. *Bioinformatics*. 2020;36:4126–9. <https://doi.org/10.1093/bioinformatics/btaa490>.
106. Brown CT, Moritz D, O'Brien MP, Reidl F, Reiter T, Sullivan BD. Exploring neighborhoods in large metagenome assembly graphs using spacegraphs reveals hidden sequence diversity. *Genome Biol*. 2020;21:164 <https://doi.org/10.1186/s13059-020-02066-4>.
107. Bae J, Lee KW, Islam MN, Yim H-S, Park H, Rho M. iMGEins: detecting novel mobile genetic elements inserted in individual genomes. *BMC Genom*. 2018;19:944 <https://doi.org/10.1186/s12864-018-5290-9>.
108. Yaffe E, Relman DA. Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. *Nat Microbiol*. 2020;5:343–53. <https://doi.org/10.1038/s41564-019-0625-0>.
109. Kent AG, Vill AC, Shi Q, Satlin MJ, & Brito IL. Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. *Nat Commun*. 2020;11. <https://doi.org/10.1038/s41467-020-18164-7>.
110. Crawford JR, Howell DC. Comparing an individual's test score against norms derived from small samples. *Clin. Neuropsychol*. 1998;12:482–6. <https://doi.org/10.1076/clin.12.4.482.7241>.

## ACKNOWLEDGEMENTS

We thank Eitan Yaffe for fruitful discussions during the development of the pipeline. We also thank David Rogers for help with statistical methods, elucidating potential sources of sequencing artefacts that could disrupt our analysis, and generally highly insightful discussions.

## AUTHOR CONTRIBUTIONS

BvD designed and tested the pipeline, curated, analysed and visualised the metagenomic data, wrote the original draft and, the was responsible for the review and editing during revisions. PB tested and benchmarked the pipeline throughout development. ADF and FG performed additional experiments to test the stability of DNA in the compost mesocosms. JM and BED performed cross-assembly of MAGs. PB contributed ideas and concepts and was involved in crafting of the final manuscript. All authors have read the final manuscript and gave approval for publication, and agree to be held accountable for the work.

## FUNDING

BvD, AF, PB and PBR acknowledge support from the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Center 1182 'Origin and Function of Metaorganisms' (grant no. SFB1182, Project C4 to PBR). PBR acknowledges generous core funding from the Max-Planck Society. BED was supported by the European Research Council (ERC) Consolidator grant 865694: DiversiPHI, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2051—Project-ID 390713860, and the Alexander von Humboldt Foundation in the context of an Alexander von Humboldt Professorship funded by the German Federal Ministry of Education and Research. Open Access funding enabled and organized by Projekt DEAL.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43705-023-00294-w>.

**Correspondence** and requests for materials should be addressed to Bram van Dijk or Paul B. Rainey.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023