Examining horizontal gene transfer in microbial communities

Ilana Lauren Brito

Abstract | Bacteria acquire novel DNA through horizontal gene transfer (HGT), a process that enables an organism to rapidly adapt to changing environmental conditions, provides a competitive edge and potentially alters its relationship with its host. Although the HGT process is routinely exploited in laboratories, there is a surprising disconnect between what we know from laboratory experiments and what we know from natural environments, such as the human gut microbiome. Owing to a suite of newly available computational algorithms and experimental approaches, we have a broader understanding of the genes that are being transferred and are starting to understand the ecology of HGT in natural microbial communities. This Review focuses on these new technologies, the questions they can address and their limitations. As these methods are applied more broadly, we are beginning to recognize the full extent of HGT possible within a microbiome and the punctuated dynamics of HGT, specifically in response to external stimuli. Furthermore, we are better characterizing the complex selective pressures on mobile genetic elements and the mechanisms by which they interact with the bacterial host genome.

Mobile genetic elements
(MGEs). Units of DNA that can be transferred within a genome or between genomes.

Speciation
The evolution of a clade of organisms that are genetically and often ecologically distinct from neighbouring clades.

Genotoxins
Chemicals that can elicit changes in the DNA, including but not limited to strand breakage, deamination and cross-linking, which can lead to mutation.

Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY, USA.
e-mail: ibrito@cornell.edu

https://doi.org/10.1038/s41579-021-00534-7

Bacteria acquire novel DNA through the process of horizontal gene transfer (HGT), which enables them to adapt to changing environmental conditions. This foreign DNA (that is, DNA that can be horizontally transferred between bacteria) can contain elements that expand the niche of an organism, change its relationships with its host or provide a competitive edge against other organisms within its environment. Most notably, antibiotic resistance genes carried on transferrable cassettes may render infections recalcitrant to first-line antibiotic treatments. The functions conferred by mobile genetic elements (MGEs) extend beyond antibiotic resistance and are quite diverse, including digestion of most classes of carbohydrates[1], mercury resistance[2,3], virulence[4] and catabolism used in bioremediation[5]. Despite the importance for defining large phenotypic differences between strains, surprisingly little is known about the what, when and how of HGT within natural microbial communities, in large part due to technical difficulties in examining the mobile gene pool in situ.

Collectively, this foreign DNA (also referred to as mobile DNA) has been dubbed the 'mobilome', and provides ecological insight into the processes of adaptation and speciation. Although profiling communities via 16S ribosomal RNA marker sequencing has become commonplace in inferring putative functional differences between bacterial communities[6,7], this ignores the mobile gene pool as a source for phenotypic variation. In reality, upwards of half of the genome of an organism may comprise mobile genes[8,9], and numerous case examples prove that strain-level heterogeneity based on horizontally acquired traits can drastically alter natural ecosystems, such as the impact of the cholera toxin-encoding phage CTXφ on the emergence of toxigenic strains of *Vibrio cholerae*[10] or the introduction of an integrative conjugative element that encodes genotoxins into enterobacteria linked to colorectal cancer[11]. Answering the question of how extensively HGT shapes the function of natural microbial communities is starting to be within reach given a wider suite of tools to probe HGT with unprecedented resolution and breadth.

DNA can be mobilized through several means: conjugation, transposition or transformation being the most characterized, although other means are coming to light, including outer membrane vesicle uptake[12] and transfer via virus-like particles[13] (FIG. 1). DNA involved in these processes is heterogeneous and dynamic. Viral genomes may be packaged in phage virions or, in the case of lysogenic phages, integrated in the genome; plasmids can remain as extrachromosomal circularized DNA or as linearized DNA or may integrate into the genome; and transposons can be found within the genome or within plasmids or phages. Although foreign DNA is a helpful way of delineating contributions to the genome that are horizontally versus vertically acquired, a complete analysis of this eclectic DNA requires an assortment of genomic techniques.

The process of HGT has been studied in laboratories starting with seminal work performed by Joshua Lederberg and Oswald Avery in the 1940s, and methods to

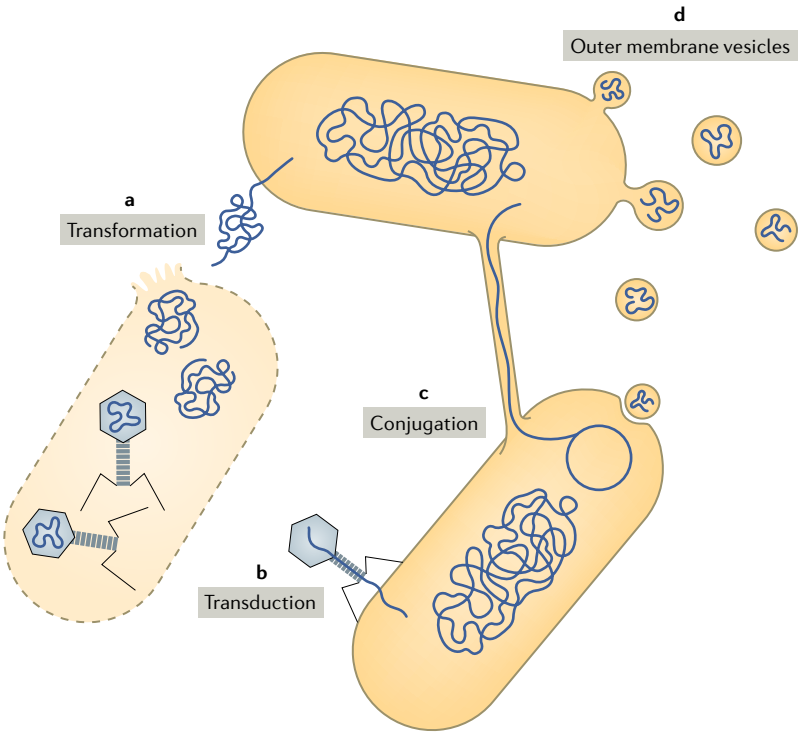

Fig. 1 | **General routes of horizontal gene transfer within natural communities.** The schematic shows the different mechanisms of horizontal gene transfer. **a** | Transformation involves the uptake of naked DNA from lysed cells in the environment. **b** | During transduction, genetic material is introduced from a phage into bacterial genomes. **c** | Conjugation involves the transfer of DNA through conjugative pili and is the predominant mechanisms by which DNA is transferred between bacteria. **d** | Additional mechanisms involve outer membrane vesicles and DNA packaged into virus-like particles (not shown), although their contribution to overall horizontal gene transfer is unknown.

coax bacteria to take up DNA have since been used to manipulate various organisms[14]. These types of test-tube experiments have been particularly useful in measuring the rates of HGT and testing general triggers for HGT, and although we can artificially induce HGT through electroporation or chemically induced competence, much less is known about HGT in natural communities. Pioneering work in this field using reporters on plasmids (reviewed in REF.[15]), for example to study HGT in biofilms[16,17], has set the stage for measuring the rates of gene transfer more widely, and this method is being revisited now with next-generation sequencing tools within the context of a microbiome. Similarly, methods to detect genomic signatures of HGT[18,19] have been further advanced by the exponential increase in the number of available genomes. This Review focuses primarily on methods to assess recent HGT (TABLE 1), as methods to study ancient gene transfer may differ. Nevertheless, many of the methods to study ancient HGT events (reviewed in REF.[20]) are mature and may be robust for identifying modern events. As this field is rapidly emerging with novel technologies that may not yet be fully vetted, cross-verifying results using multiple methods and confirming results using databases or culture-based approaches will be essential in determining which of these methods prove accurate and reliable.

Biofilms
The organization of unicellular organisms, often multiple species, into an adherent, cohesive mat, often involving extracellular polymeric substances and distinct changes in function compared with planktonic cells.

k-mer composition
A tally of the unique DNA fragments *k* base pairs long in a genome or a dataset of sequencing reads or contigs.

A major goal of this Review is to synthesize advances made across disciplines. Many of the basic questions about HGT are common across different ecologies: what genes are most often transferred? Between which bacteria? By what means? Under what conditions? With what barriers? And at what rate? Many of the methods described here are universally applicable, requiring only minor adjustments suited to different microbiota characteristics; that is, whether they are dilute (for example, marine or skin microbiomes), whether there is heavy particulate matter (for example, human stool); whether there are many hard-to-lyse Gram-positive organisms (for example, soil) or whether PCR inhibitors are present (for example, avian cloacal microbiomes). MGEs differ across environments in terms of G+C proportion[21], *k*-mer composition[22] and gene content. For example, plasmids carrying biocide resistance or metal resistance genes tend to be much larger and more likely to harbour toxin–antitoxin systems than plasmids carrying antibiotic resistance genes[23], which may reflect underlying selective pressures in different environments. With better methods to assess gene transfer, we will soon be able to answer second-order questions such as the following: what is the role of HGT in the resiliency and adaptability of communities? Is crosstalk between transferred genes and bacterial host genes[24,25] the exception or the rule? Does gene transfer confer generally beneficial or deleterious outcomes for microbial recipients? And what is the relative contribution of HGT mechanisms to speciation and to the evolution of microbial communities at large?

Mobile genetic elements in metagenomes

Shotgun metagenomic sequencing captures some MGEs, but short-read lengths (100–300 bp with the Illumina platform) pose major obstacles in assembling and identifying horizontally transferred regions (FIG. 2a). MGEs often contain components that are overrepresented and present in multiple genomes[26]; within a sample, they may have recombined within and between MGEs[27]; and the presence of direct or inverted repeats flanking MGEs[28] complicates de novo assembly (FIG. 2b,c). Furthermore, assemblers also struggle with variable sequencing depths of MGEs compared with their host genomes that result from free-floating phages, high-copy plasmids[29] or the presence of common mobile genes across MGEs (FIG. 2d). Metagenomic assemblers, especially those that use de Bruijn graph assembly, generate fragmented contigs in the face of complicated graphs[30].

There are defining features or markers consistent across MGEs that can be used for their identification, such as machinery proteins (for example, resolvases), enzymes and structural proteins (for example, phage capsids) involved in the process of HGT. Specifically, these include proteins that assemble into phages (for example, phage capsid and tail proteins), proteins involved in conjugation (for example, plasmid relaxases, which nick DNA at the origin of transfer to induce mobilization, or the Tra genes, which include conjugative pili proteins) and transposases and insertion sequences. Raw reads or assembled contigs can be aligned directly to MGE databases[31–35], but these databases may not be congruent

with all samples. Hybrid identification approaches combine annotation with gene-agnostic methods, such as differences in G+C content, methylation patterns, *k*-mer content or the ability to identify circular plasmid genomes[36–39]. Because much of the sample-unique genetic diversity may arise from MGEs[40], a large portion of the mobile gene pool may go unnoticed when reference-based alignments are used. In a study of the gut microbiota of Fiji Islanders, the mobile genes observed were substantially less complete for alignment to MGEs identified from existing reference genomes than to a dataset that also included MGEs identified in bacterial genomes from the Fijian population being surveyed[26]. Overall, reference databases of MGEs are notoriously incomplete and biased towards well-studied pathogenic organisms. The diversity of phages is extensive[41], although computational approaches are being applied to understand their movement between hosts[42]. Ultimately, there is no comprehensive reference database of MGEs to draw comparisons with, nor could there easily be as MGEs have rapidly evolving gene content. There is a need for computationally cheap methods to identify MGEs in metagenomic shotgun sequences while preserving their content and genomic context.

Short-read mapping. Horizontally transferred regions can be detected by comparing reads or assemblies with reference genomes. Large gaps or extreme peaks in coverage within a reference genome may reflect structural variants indicative of HGT[29] within the sequenced sample. Similarly, paired-end reads individually mapping to different references or assemblies or a single read partially mapping to different genomes may also suggest recombination of transferred DNA, gene gain or gene loss. MetaCHIP (metagenomics community-level HGT identification pipeline) is one such method that examines the best-match homology for open reading frames within an assembled contig[43]. Putatively transferred DNA is identified as those parts of assemblies showing high-homology alignments to alternative genomes and little homology to the flanking DNA. An alternative approach is to look for regions in a reference genome that do not recruit paired metagenomic reads, despite decent coverage of the rest of the reference genome[44,45]. One such method, split-read insertion detection (SRID), implemented in packages such as MGEFinder[138] and DaisyGPS[46], has been used for genomes[137] and metagenomes[32] alike. Whereas MetaCHIP is useful for finding transferred regions in

Table 1 | **Approaches to study horizontal gene transfer in microbial communities**

Goal	Method	Advantages	Limitations
Identifying mobile genetic element	Phage, transposon or plasmid identification using gene markers in metagenomic assembled genomes[34,35,37]	High-confidence hits	Largely dependent on comprehensiveness of reference databases
	Sequence comparison (*k*-mer-based partitioning or binning) of metagenomic reads or contigs[22,36]	De novo method that is straightforward to implement	Low sensitivity, potential false associations
	Split-read insertion detection[32,46,137,138]	Easy to implement	May be complicated by intragenomic recombination, low sensitivity, prone to false positives
	Sequencing phages[53–57]	Higher rate of capture than standard sequencing approaches	Laborious, captures only lytic phage
	Sequencing plasmids[66,67,139]	Higher rate of capture than standard sequencing approaches	Potentially biased capture
	Comparison of identical regions in distantly related organisms' genomes[26,43,87,89]	High-confidence identification of recent HGT	Low sensitivity for full MGE sequences
Obtaining genomic context	Long-read sequencing[77–79]	High-confidence hits, can obtain full-length mobile elements, may retrieve host associations for integrated elements	Does not capture plasmid–host associations
	Inverse PCR[75]	Gains contextual information	May provide local context that may not contain taxonomic information
	Methylation sequencing[83,140]	High accuracy	Limited resolution
Linking mobile genetic elements to hosts	Whole-genome sequencing	Accurate, standard approach	Culturing is laborious
	Single-cell sequencing[26,100]	Obtain host-relevant information on MGEs without the need for culturing	Genomes can be incomplete, certain methods are prone to contamination
	Viral tagging, reporter constructs and other FACS-based sorting with tagged phage or plasmids[107–109]	Straightforward implementation	Cell sorting is technically difficult, may require genetic engineering
	Proximity ligation (Hi-C or XRM–seq)[112,115–118,141,142]	Comprehensive	Low sensitivity, expensive, laborious
	Single-cell fusion PCR (epicPCR, OIL-PCR)[119,121]	High sensitivity	Low throughput (few genes can be identified at once)

epicPCR, emulsion, paired isolation and concatenation PCR; FACS, fluorescence-activated cell sorting; HGT, horizontal gene transfer; MGE, mobile genetic element; OIL-PCR, one-step isolation and lysis PCR; XRM–seq, ribosome crosslinking and sequencing.

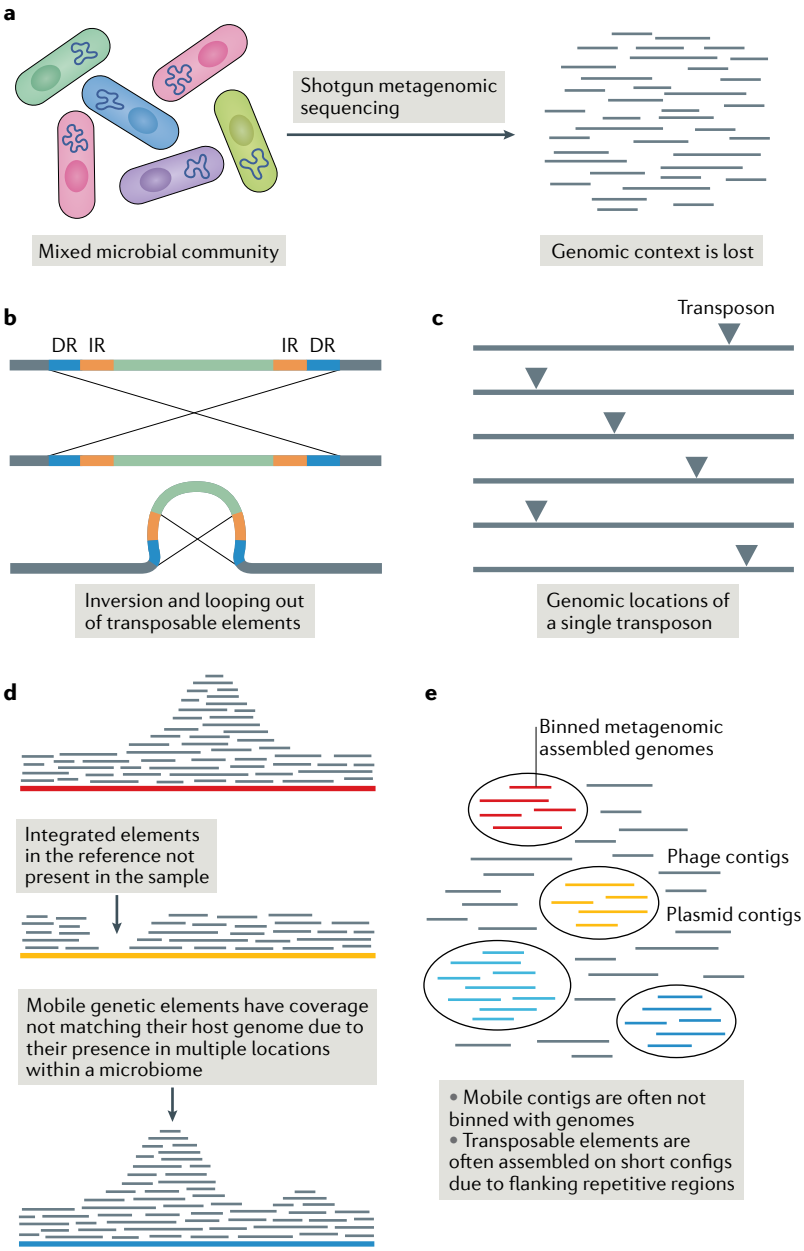

Fig. 2 | **Metagenomic assessment of the mobilome.** **a** | Metagenomic shotgun reads of a mixed microbial community lose their genomic context when they are sequenced on short-read sequencing platforms. Plasmids lose their genomic context, even with long-read sequencing. **b** | Mobile genetic elements (MGEs) are especially recombinogenic, due to repeated elements (direct repeats (DR) or inverted repeats (IR)) present in phages, transposable elements and common machinery genes that can result in translocation, rearrangement or looping out of DNA. The resulting heterogeneity within a population complicates metagenomic assembly. **c** | The mobility of transposons within a single genome complicates metagenomic assembly. **d** | MGEs present in multiple genomes or sharing similar regions may improperly recruit sequencing reads derived from a different MGE or genome within a microbiome to reference genomes or metagenomic assembled genomes (coloured segments). Similarly, a reference genome may fail to recruit reads to regions in the genome representing MGEs. Coverage of mobile DNA therefore may poorly reflect the abundance of a particular genome. **e** | Contigs containing mobile DNA are often left unbinned and/or are incorporated into only a subset of their host genomes. Many cannot be annotated at all due to the relatively poor annotations of mobile genes compared with core genes.

assemblies, SRID is useful for finding flexible regions in isolates or reference genomes. Resulting regions must be interpreted with caution as these methods are prone to false positives[137], due to the recruitment of reads from closely related organisms and recombination within individual genomes.

Binning methods. Binning raw metagenomic reads or contigs to assemble genomes with higher completeness also results in better assembly of MGEs, and in some cases, can be used to link MGEs with their bacterial hosts (FIG. 2e). Binning methods generally use *k*-mer or G+C composition and/or the abundances of *k*-mers or contigs[47–49]. Contigs from highly promiscuous MGEs will fail to be binned with their hosts if there are many variants. However, this method has proven to be useful with MGEs that are more closely linked with one host, namely host-specific phages. Emergent self-organizing map clustering of time-series contigs assembled from metagenomic data, according to their abundance profiles, was used to find *Staphylococcus epidermidis*-specific phages and plasmids[50], as well as plasmids for *Enterococcus faecalis* in infant gut microbiota[51]. Similarly, latent strain analysis[52], a binning method based on the *k*-mer profiles of individual reads across many related samples, was able to find phages along with their host genomes. These methods provide an incomplete picture of the MGEs in samples, but simultaneously suggest that there may be opportunities to improve metagenomic assembly further to account for this fast-evolving portion of the microbiota.

Direct sequencing of plasmids or phages. Phage sequencing has opened doors into understanding the ecology of microbial systems, as they seem to have a key role in modifying community structure. For example, phage populations may have a role in shaping the developing microbiota in infants[53], and are altered in health conditions such as inflammatory bowel disease[54,55] and malnutrition[56]. Phage concentrations may differ greatly, sometimes requiring large sample sizes (for example, upwards of 500 g for human stool[57]) depending on the protocol. The diversity of phage genomes — DNA based or RNA based, single stranded or double stranded — presents a challenge in isolating all types of viruses simultaneously. Rather than identifying prophages in metagenomic data, one can directly isolate and sequence lytic phages to examine their gene content, while disregarding their respective hosts[58]. Commonly used methods involve using a CsCl density gradient to isolate phages, tangential-flow filtration and polyethylene glycol precipitation[59–61]. These methods favour smaller viral or virus-like particles, and therefore several extremely large viruses identified in wastewater[62] or the human gut (for example, the 97-kb crAssphage[63] and the 540-kb

Prophages
Phage genomes that are integrated and replicated along within the genome of their host. These are either phages in their lysogenic phase or they are inactive (mutated so they no longer can enter a lytic phase).

Lytic phages
Phages that reproduce within a cell and subsequently lyse the cell to release the virions.

megaphage of *Prevotella* species[64]) may be missed. Furthermore, these methods are prone to sample loss, and verification using staining, functional screening or testing for the presence of genomic DNA contamination using 16S or 18S ribosomal RNA primers is advisable.

Sequencing plasmids, albeit in isolation from their hosts, can reveal differences in the recombination frequencies of plasmid-associated transposable elements[65] and has been used to identify functional differences in a wide range of environments, including cow rumen[66] and wastewater treatment plant sludge[67,139]. Plasmids can be isolated using alkaline lysis, although the efficiency may vary by microbiome type, or filtration methods similar to those mentioned above. Plasmid DNA can be enriched in metagenomic DNA preparations by selective digestion of linear DNA with a plasmid-safe exonuclease and by use of multiple displacement amplification, which amplifies plasmids particularly efficiently due to the processivity of the Φ29 polymerase. An alternative approach for amplifying plasmids called 'transposon-aided capture' involves tagging isolated plasmids with a transposon carrying antibiotic resistance, amplifying them in an exogenous host, such as *Escherichia coli*, and then recovering plasmids after antibiotic selection[68,69]. Quantifying the abundance or copy number of plasmids in natural communities has remained elusive.

By combining virome sequencing with sequencing of the microbiota, it is sometimes possible to identify putative hosts by analysing the content of CRISPR arrays. As part of a bacterial adaptive immune response, CRISPR–Cas systems protect against deleterious phages or plasmids by incorporating DNA complementary to the transferred DNA into their CRISPR arrays, functioning as a registry of the cell's previous exposures. CRISPR arrays have been mined to understand HGT within host-associated environments[70] and marine environments[71]. In some cases, this enables the linking of single CRISPR elements to their MGE targets[72,73] and, if contigs can be taxonomically annotated, their host bacteria[74].

Additional genomic context of MGEs

Beyond identification of MGEs, several technologies provide a better localized context in which to situate MGEs. In some cases, these methods may be sufficient to assign MGEs to specific hosts.

Inverse PCR. The genomic contexts of integrated MGEs are often lost during assembly, especially for integrated MGEs that are especially promiscuous. Although the synteny of reads has been used to try to establish the immediate genomic context of an MGE, this method is limited to the quality of mapping of individual short reads[26]. Rather, inverse PCR is a low-throughput method, gene or MGE specific, and can be used to obtain the local genomic contexts of MGEs by sequencing the area surrounding a gene or MGE of interest, which may additionally contain taxonomic information. Rather than designing primers that amplify the DNA segment interior to the two primers, the inverse PCR approach enables sequencing of the regions exterior to a primer pair. In short, DNA is digested into large fragments (tens of kilobases or longer) and circularized. Outward-facing primers are then used to amplify distal regions. This method has been applied to mobile antimicrobial resistance genes harboured by wastewater communities[75]. Its utility depends on the genomic diversity of the contexts in which an MGE is situated. For example, it could be used to identify host bacterial species if species-level markers are situated in the MGE-adjacent regions.

Long-read and synthetic long-read sequencing. MGEs in natural microbial communities range in size from 1 kb to upwards of 1 Mb (REF.[76]). Within the gut microbiome, two phage families, including crAssphage[63] and Lak megaphages (more than 500 kb in size)[64] are widespread. Illumina shotgun sequencing results in paired-reads with, at most, 600 sequenced bp, so there is a utility in applying long-read sequencing technologies to MGEs and other difficult-to-assemble genomic regions, despite their higher error rates. Newer methods, such as barcoding DNA fragments arising from one larger fragment and sequencing them on the 10X Genomics platform, have been used to examine intraspecies differences[77,78], but these methods are still being developed. Long-read sequencing technologies such as PacBio's single-molecule, real-time (SMRT) platform and Oxford Nanopore's sequencing platform typically sequence DNA fragments well in excess of 10,000 bp, and in the case of the latter upwards of 100,000 bp, which is sufficiently long to span a complex transposon insertion, a prophage and even some full plasmids[79]. These technologies both have higher error rates and higher costs than short-read platforms, such as the Illumina platform, although these metrics are rapidly improving[80]. The choice of one platform, or both[81], will depend on the goal of the sequencing effort (identifying integrated MGEs or examining sequence variants). These methods for obtaining large contiguous genomic sequences are especially useful for integrative elements and in assembling plasmids that frequently recombine; however, they are inherently limited in their ability to link non-integrative plasmids with their host microorganisms.

Methylation signatures. MGEs are exposed to host methyltransferases and acquire methylation patterns matching their host's DNA, and these organism-specific methylation signatures can therefore be used to bin MGEs with their host genomes. DNA methylation can be detected by bisulfite sequencing, PacBio's SMRT sequencing platform or IonTorrent's long-read sequencing platform, as chemical modifications of DNA affect the kinetics of sequencing[82,83]. By exploitation of this feature, MGEs have been linked with their host genomes in a mouse gut microbiome[140] and a simple cheese microbiome[84]. Among the challenges in scaling up this approach is the necessity to use a platform that provides information about the methylation status of DNA sequences, either through bisulfite sequencing or SMRT sequencing, which is more expensive and of lower throughput than its short-read relatives. It is also limited to larger MGEs, which contain adequate coverage of unique or identifiable methylation motifs to be able to assign host-MGE associations. Additionally, methylation patterns can be altered in response to environmental stimuli[85], which

limits the utility of such a method for examining changes in plasmid–host associations over time or in comparing the same species present in different hosts[86].

Associating MGEs with their hosts

A major goal of many of the approaches outlined here is to obtain accurate and complete information about host–MGE associations within a community so we can dissect the dynamics of HGT in natural communities.

Comparative genomics. Comparative genomics using whole-genome sequences has been the gold standard method for examining HGT, providing high-quality data for associating integrated MGEs and plasmids with their host. Although genome sequencing is of low throughput and covers a fraction of the community, it has still been useful in constructing HGT networks in various environments[87], including the gut microbiome[26,88] and within cheese communities[89]. Improvements in culturomics have resulted in capturing greater diversity than ever before within a microbiome[90,91], enabling a more comprehensive snapshot of the HGT network. Still, care must be taken to avoid contamination, which may falsely appear as HGT, and artifacts introduced during culturing, such as conjugation that occurs during culturing and the gain and loss of plasmids during plate-based culturing[92,93]. In closely related strains, detecting horizontally transferred regions requires observing DNA segments that differ across isolates[88] that are not likely to be explained by gene loss. Others have applied a simple, conservative heuristic to identify recently transferred DNA by identifying nearly identical DNA in distantly related organisms[87], by examining the length of stretches of nearly identical DNA in closely related organisms[94] or by examining insertion sites within sequenced genomes[95]. Identifying regions of recent HGT in isolates of intermediate relatedness can be more challenging, relying on subtler techniques (for example, differences in codon usage[96]).

Newer technologies such as single-cell sequencing promise to increase the number of available genomes by several magnitudes. Single-cell sequencing, through cell-sorting[97], microfluidic[98] or hydrogel-based[99] isolation, is attractive for several reasons: it does not require a priori knowledge of culturing conditions; it is taxonomically unbiased; and it may be less labour-intensive and, theoretically, higher throughput than culture-based assays, which require colony picking. Single-cell sequencing is not without caveats, as assemblies are often incomplete and more fragmented than those of sequenced isolates[98], and are more prone to contamination. Combining single-cell sequencing with metagenomics offers the possibility of identifying mobile genes in a population and then surveying the mobile gene pool in a large number of samples as well as examining differences in the architecture of MGEs across samples[26,100].

Mobile reporter constructs in microbial communities. The traditional way of studying HGT in vitro is to use reporter constructs built into plasmids or phages that enable the detection through fluorescence or antibiotic selection, and these methods have been illustrative in examining conjugation in natural communities[101,102] and within biofilms[103] (FIG. 3). Rates of HGT can be inferred from increases in the number of transconjugants, transformants or transductants over time. This method is inherently limited to those MGEs that can be genetically altered or are tractable, but has been used, with some success, to track HGT in microbial communities in vivo. In a recent study, a genetically modified phage, SopEΦ, was introduced into two trackable *Salmonella enterica* subsp. *enterica* serovar Typhimurium strains[104]. Newly formed lysogens were identified by the incorporation of antibiotic resistance traits carried by the phage, resulting in strains of donor and recipient bacteria that had different resistance profiles. Barcoding a subset of phage virions enabled the study authors to examine the rates of independent transfer events and conclude that the frequency of transfer events must be high.

Extending this framework to examine multiple MGEs or various organisms may be challenging. Regarding the use of antibiotic resistance reporters, organisms differ in their intrinsic or acquired resistance to antibiotics, which may obscure the detection of organisms with newly acquired resistance traits. Alternatively, fluorescent reporters may be used to screen HGT events between diverse species, but expression may differ across species. In one such attempt, plasmids isolated from soil communities were engineered to express green fluorescent protein (GFP) and introduced into donor cells[105]. Donor cells carried red fluorescent protein (RFP) and a repressor for the promotor driving GFP expression so that GFP would be expressed only in cells that received the plasmid. Laboratory conjugation experiments between these bacterial donors and soil communities were performed and then subjected to fluorescence-activated cell sorting followed by 16S sequencing, revealing subpopulations with transconjugants that were able to acquire the plasmid. This study had provocative results, showing extensive transfer across the bacterial phylogeny. Because this method may be subject to false positives due to challenges of sorting heterogeneous populations of bacteria, they necessitate genetic confirmation. Viral tagging has also been effective in obtaining phage–host interactions; isolated phages are fluorescently labelled and adsorbed to the cellular envelopes of specific hosts[106] and sequenced in bulk or are reintroduced into a microbial community and sorted by fluorescence-activated cell sorting into individual wells[107–109]. This method can provide comprehensive information about phage–host associations and does not require genetic engineering.

Approaches that use genetic reporters, like the barcoded SopEΦ phage, may leverage next-generation sequencing to increase their throughput. Cas machinery has also been used to inducibly record HGT events[110]. Cas1 and Cas2 proteins, which incorporate mobile DNA into protospacers, were introduced into an *E. coli* reporter strain. This reporter strain was introduced into communities, after which the CRISPR array was sequenced to determine the cellular exposures.

Proximity ligation. Hi-C, a type of proximity ligation method, is a method that generates long-range DNA–DNA interactions mediated by DNA-associated proteins

Culturomics
The study of bacterial cell culture using high-throughput methods, usually with the goal of isolating diverse organisms from complex microbial communities.

Protospacers
Small DNA fragments found within CRISPR arrays that are derived from invading mobile genetic DNA (plasmids or phages).

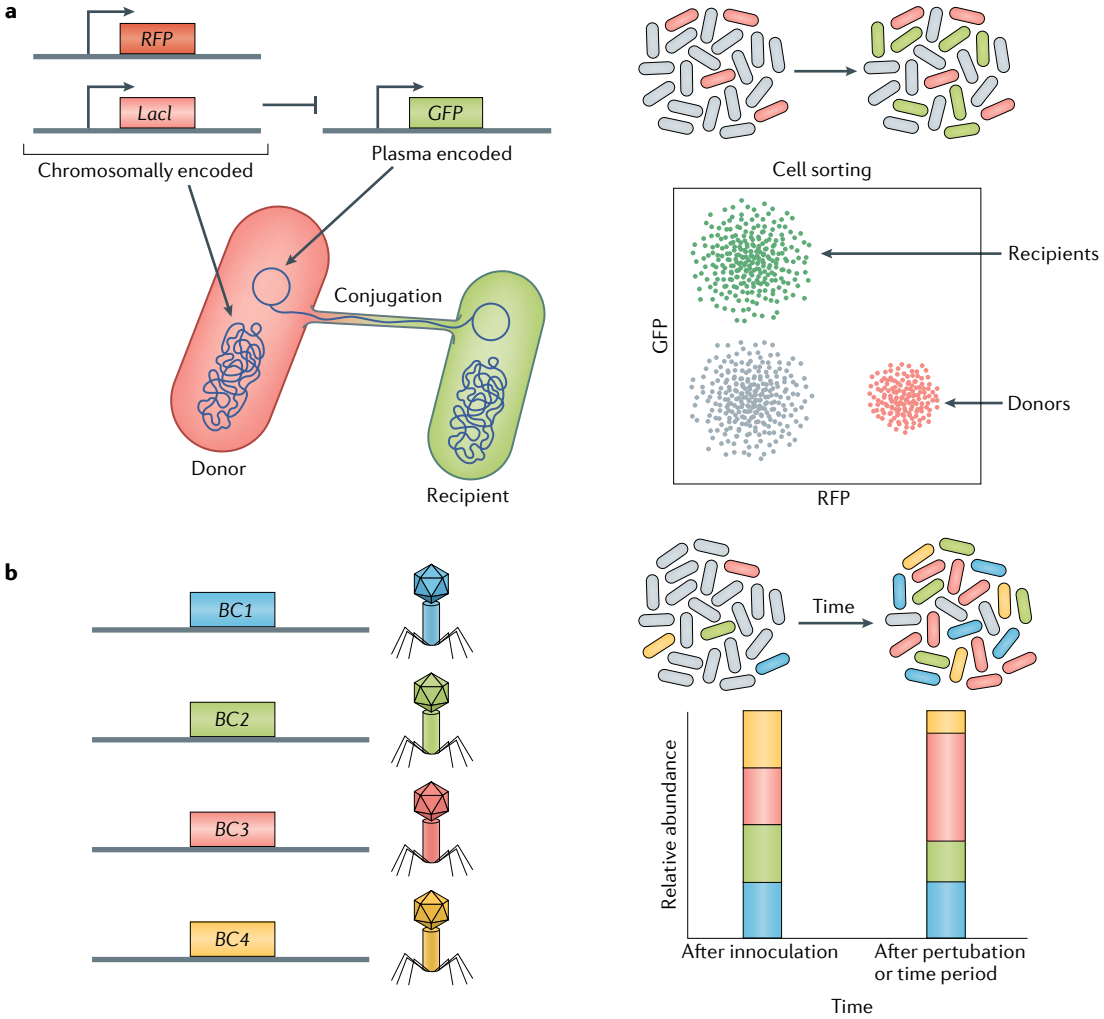

Fig. 3 | **Reporter constructs for examining recipients of horizontal gene transfer and movement of mobile genetic elements. a** | Fluorescent reporters have been a mainstay for examining horizontal gene transfer events. In this example, a donor strain encodes red fluorescent protein (RFP) and LacI on its chromosome, and green fluorescent protein (GFP) is encoded on a plasmid. LacI represses GFP fluorescence in the donor, and therefore the donor appears red. In recipients, GFP is expressed and therefore transconjugants fluoresce green. Donors and recipient strains can be distinguished using fluorescence-activated cell sorting. **b** | Genomic reporters are now being used to monitor the movement of specific mobile genetic elements, as in the case of phage and plasmids. A library of phages carrying different barcodes (BC1–BC4) can be introduced into a bacterial population, and the frequencies of phage transfer events can be detected over time, with or without a perturbation such as infection or antibiotic treatment. The largest libraries thus far have comprised seven barcoded phages[104].

within the genome before sequencing (FIG. 4), can be used to link mobile genes with their host genomes. Hi-C was originally developed to determine chromosomal structure in eukaryotes. In short, DNA from bacterial cells is crosslinked by formaldehyde links between DNA-bound proteins while cells are intact, then while cells are intact, then subsequently digested, biotin labelled and ligated in dilute solution to associate DNA fragments bound in the same crosslinked complex. Biotinylated fragments are concentrated and sequenced to reveal interactions between distal regions of DNA existing in the original cells[111]. Although the main use of Hi-C has been to improve the quality of genomic assemblies from microbiome samples[112], Hi-C sequencing has also proved its ability to associate distant portions of bacterial chromosomes or a bacterium's chromosome with its plasmid or phage[141], even in environmental[113,114] or human[115,116] microbiomes. One study focusing on MGEs used Hi-C to show that HGT has resulted in the widespread dispersal of MGEs within individuals' gut microbiota[117]. This has led to the development of MGE-specific Hi-C protocols that implement computational pipelines specifically geared towards associating MGEs with their hosts, the large portion of which has recapitulated known host–MGE associations, lending credence to these approaches. Built on the same premise, XRM–seq (ribosome crosslinking and sequencing)[118] instead crosslinks ribosomal RNA with total mRNA, including viral mRNA that is being transcribed. Chimeric reads can associate active viral genomes with bacterial hosts.

Proximity ligation methods favour comprehensiveness, potentially capturing plasmids, prophages and

transposable elements concurrently in a wide range of cells. As Hi-C reads align to contigs with few base pairs (on average, half of a read pair), Hi-C read pairs may therefore ambiguously map to multiple places. The data must be interpreted in consideration of the redundancy of MGEs. In addition, most of the paired sequence reads are proximal to one another within the genome; that is, they are uninformative at providing the long-range information required to link MGEs to their nascent genomes. This method also may not be sensitive to low-abundance interactions. Nevertheless, it is notable in its ability to link MGEs with their hosts in a high-throughput, unbiased fashion.

Single-cell fusion PCR. Fusion PCR approaches have been used to study the host associations of rare genes, and the method is slowly being applied to genes that may be mobile. For example, emulsion, paired isolation and concatenation PCR (epicPCR)[119] has been used to detect the bacterial hosts of a specific targeted gene with high sensitivity. Functional genes are fused with phylogenetic markers originating from the same cell during PCR (FIG. 5). Although this method has been used to establish the host associations of mobile antibiotic resistance genes in wastewater microbial communities[120], the biomaterials that encapsulate beads may enable extrachromosomal elements, such as plasmids, to escape and contaminate adjacent beads, increasing the false-positive rates by erroneously linking plasmid and host. Alternative platforms are being developed, such as one-step isolation and lysis PCR (OIL-PCR)[121], to enable scalable high-fidelity capture of mobile genes and their hosts.

Insights into HGT within communities

Due to the aforementioned technologies, we are acquiring a broader understanding of the types of mobile genes, their genomes and their genomic contexts. As these technologies enable the study of what gets transferred, how it is transferred and between whom it is transferred, it will be necessary to place our observations of HGT in the context of the dual processes of HGT and natural selection (FIG. 6) to understand the role of HGT in shaping microbial communities.

Extensive within-microbiome transfer is possible. After analysis of several thousand reference genomes from different environments, it became clear that, in addition to genetic relatedness, shared ecology governs HGT[87]. However, the extent to which organisms in the same shared environment engaged in HGT is not well understood. Several studies have revealed the potential for extremely widespread transfer in situ by use of fluorescent reporters, as mentioned earlier. Broad host range plasmids encoding GFP that were introduced into a mouse microbiome via a donor *E. coli* strain were able to disperse widely across phyla[122]. Importantly, this occurred more dramatically within the gut of the mouse than when reporter strains were introduced into equivalent microbiome samples in vitro. Similar experiments in soil communities in vitro showed that three different plasmids harboured by three different bacterial species were able to disperse across phyla into a large number of recipient clades[105]. These experiments highlight the role of selection rather than dispersal in shaping mobile gene pools. Similarly, ongoing work suggests that mobile genes are widespread within a single individual's gut microbiota, confirming that contact rates and shared environmental conditions may favour HGT and selection for specific MGEs. More work is needed to determine the contributions of HGT between commensal and transient organisms and to better evaluate differences in the rates of HGT between environments.

Punctuated bursts of HGT rather than gradual gene flow. Several in vivo experiments suggest that the rate of gene transfer in host-associated microbiomes may be high. In experiments performed on *S.* Typhimurium in mice, phage transduction occurred during acute inflammation[104]. Bacteria carrying GFP plasmids introduced by gavage can be transferred to multiple hosts within days of inoculation[122]. Cas1–Cas2 engineered reporting systems show an initial burst of spacer incorporation within hours of introduction of an organism into the gut microbiome[110]. The forces that govern the rate of gene transfer, whether they are driven by contacts, environmental cues or cellular stress signals, have yet to be defined, but these experiments suggest that gene transfer

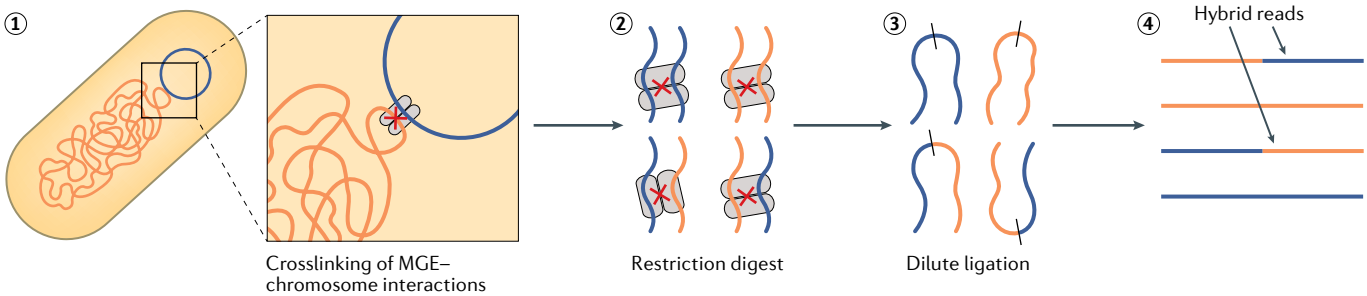

Fig. 4 | **Hi-C applications to identify bacterial host associations of mobile genetic elements.** The Hi-C workflow starts with crosslinking, creating linkages of extrachromosomal (blue) and integrated mobile genetic elements (MGEs) (integrated MGEs are not shown) with their associated genomes within cells in a microbial community (1). DNA is cut with restriction enzymes (2). A dilute ligation is performed to promote ligation between DNA crosslinked within the same complex. Proteins are digested and the remaining DNA is sequenced (3). Some read pairs, hybrids of chromosomal and extrachromosomal DNA, will be sufficient to link MGEs to genomes (4).

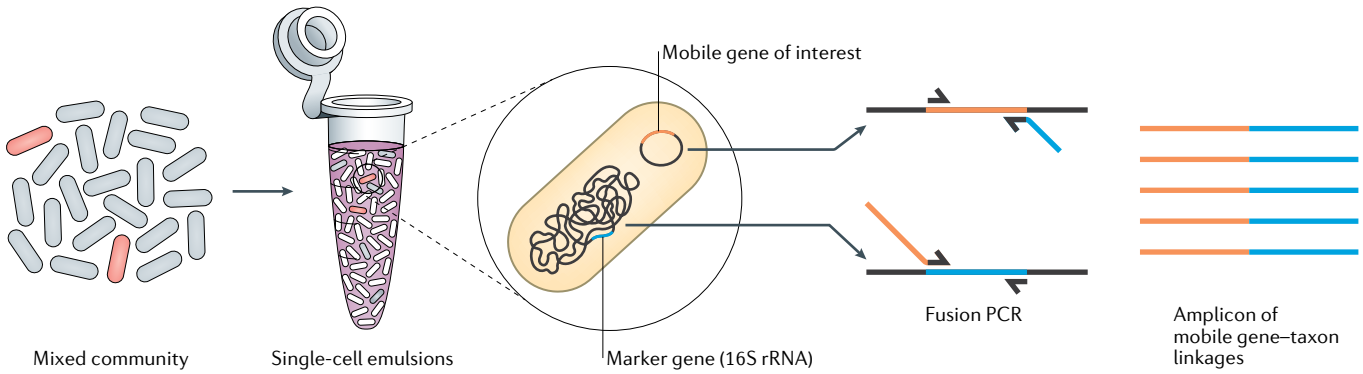

Fig. 5 | **PCR-based methods for examining single mobile genes and their genomic contexts.** Single cells from a mixed community are captured in an oil-in-water emulsion, where most of the bubbles are empty and those with cells contain only one cell. A standard PCR targets the mobile gene, and a second fusion PCR using bridging primers can physically link amplicons of a mobile gene (orange) with a taxonomic marker (blue). The final amplicons will therefore contain only PCR fusions with marker genes from cells containing the plasmid, providing information about the specific host or the co-occurrence of mobile genes. rRNA, ribosomal RNA.

may occur across a number of organisms synchronously, rather than at regular frequencies.

Complex selective pressures on MGEs. Of great concern is the spread of antibiotic resistance genes commonly found on MGEs. Travellers have been shown to accumulate more antibiotic resistance genes after travel[123], suggesting that the mobile gene pool of an individual is flexible and subject to selective pressures; however, the scale and time frame are not known. Mobile carbohydrate degradation genes have been shown to differ between the gut microbiota of global populations[1], despite shared culinary experiences in certain cases. Antibiotic resistance genes were also not maintained during the transit of microorganisms and MGEs through a wastewater treatment plant[120]. There is a need to understand the role of selection on mobile genes in different genomes. Often, mobile genes, including those providing auxiliary functions and those related to carriage of exogenous DNA, can be detrimental, and compensatory mutations are often found in genomes retaining an MGE[124]. One illustrative study showed differences between the mobility and function of resistance genes that convey resistance to antimicrobial peptides as compared with resistance genes that convey resistance to antibiotics across species in the gut. This work reveals a greater requirement for functional compatibility of the antimicrobial peptide-resistance genes with their host genomes[125]. In general, we are only beginning to understand the greater selective pressures that shape mobile gene pools.

Interplay between MGEs and their bacterial host genomes. An increasing body of evidence has pointed to interactions between mobile elements and their bacterial host genomes. This can include specific interactions related to the function of the host species, as in the case of plasmid-associated genes modulating *Acinetobacter baumannii* gene expression patterns to promote urinary tract colonization[24]. A phage within *Listeria monocytogenes* is transduced during phagocytosis and promotes bacterial survival within the phagosome[126]. Notably, this has also included anti-CRISPR systems encoded by phages[127]. How commonly these bacterial host-associated traits are carried on MGEs has not yet been assessed in a high-throughput manner, but they suggest co-evolution between MGEs and their host genomes, despite the horizontal transmission of MGEs.

Outlook

Despite a growing toolbox of techniques used to measure HGT, there are many questions that remain to which some of these tools can be applied. The underlying dynamics of HGT are not well understood. Experiments using phages in mouse models of inflammation show that HGT within a single microbial community can be rapid[128]. However, although we have witnessed the rapid spread of mobile genes worldwide[129], we do not know the ecological underpinnings that govern successful or lasting gene exchange and how frequently this occurs in natural settings. None of the aforementioned techniques can provide a complete or perfect picture of HGT in natural communities. The techniques differ in terms of their comprehensiveness and sensitivity, MGE length and overall throughput, and breadth and accuracy. With regard to the participants of HGT in these communities, only a few of these techniques can link MGEs to their bacterial hosts. None elucidates the direction of gene flow, and even the mechanism — transduction, transformation or conjugation — may be elusive. Additionally, there are other putative HGT mechanisms, such as through outer membrane vesicles[130–132], which we know little about. To answer these systems-level questions about HGT, these methods need to be more robust. They need to become of higher throughput and more cost-effective to be applied in the context of a large comparative case–control study or time course studies.

These techniques hold the promise of answering ecological questions about the role of HGT in microbial communities. When does HGT result in lasting genomic integration that may affect changes to the microbial ecosystem or host? There has been a

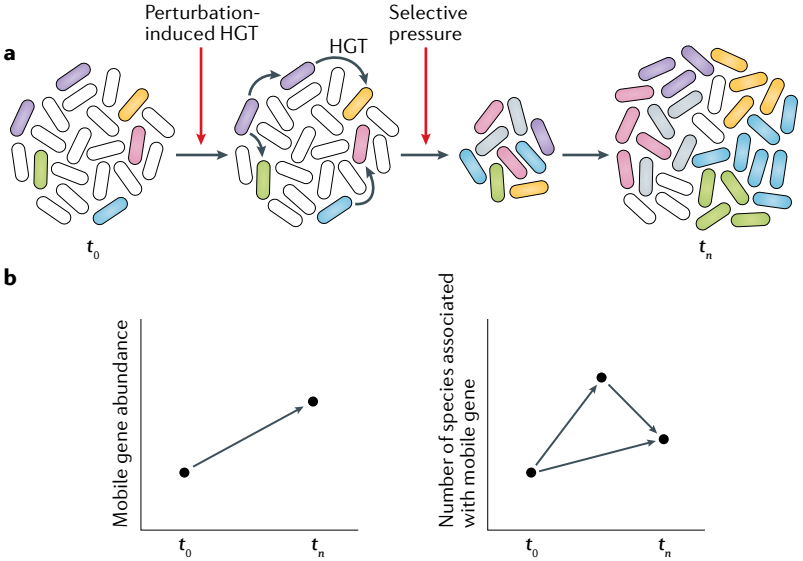

Fig. 6 | **Disentangling the processes of horizontal gene transfer and natural selection. a** | In a natural microbial community, horizontal gene transfer (HGT), likely due to a perturbation, and a subsequent selective pressure imposed on a mobile genetic element (MGE) or gene are observed in combination. From an initial starting time (t_0), there may be an expansion of organisms carrying a specific MGE and a subsequent imposed selective pressure that may limit the number of organisms carrying the MGE over time. Depending on the time at which a community is observed, it is therefore difficult to distinguish the contribution of these two processes: transfer and selection. **b** | Specifically, an increase in abundance of a mobile gene may be due to the replication of a small number of hosts (left panel) or the dispersal of a gene throughout a community (right panel). Disentangling the respective contributions of HGT and selection can be aided by more frequent sampling and by the methods discussed in the main text.

long-standing question about the role of antibiotics in not only selection but also in the promotion of HGT[133]. Which MGEs are under positive selection, and how do individual genes contribute? Which mode of HGT is dominant, and how does this differ across microbiome types? We need methods that can measure the utility of MGEs in novel genomic contexts to measure selective pressures to keep or lose a mobile gene, as gaining MGEs may be initially deleterious[134]. One way to examine whether recently acquired MGEs are expressed in communities is to examine metagenomic data alongside proteomic data[131].

Even the most robust methods described here do not comprehensively capture all HGT in microbial communities. The methods outlined in this Review all face the challenge of increasing the signal-to-noise ratio and increasing sensitivity. However, in addition to detection issues, other reasons for not observing HGT in communities are the myriad natural barriers to gene flow. Some of these are well known; for example, restriction endonuclease activity, incompatible HGT machinery, phage adsorption specificity and adaptive CRISPR-mediated immunity. However, there are complexities emerging that upend our assumptions; for example, phages that can circumvent host modification or CRISPR mechanisms[25,135,136]. The contributions of these barriers to the overall observed HGT in natural communities may require methods other than those discussed here. As microbiome research achieves better understanding of community assembly at the species level, greater attention will be focused on aspects of gene flow that have remained elusive. There will be increased attention to HGT as a driver of the spread of antibiotic resistance and in understanding how bacterial communities that provide important ecological services adapt under changing environmental conditions.

Published online 12 April 2021

1. Hehemann, J.-H. et al. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908–912 (2010).
2. Summers, A. O. et al. Mercury released from dental 'silver' fillings provokes an increase in mercury- and antibiotic-resistant bacteria in oral and intestinal floras of primates. *Antimicrob. Agents Chemother.* **37**, 825–834 (1993).
3. Hemme, C. L. et al. Lateral gene transfer in a heavy metal-contaminated-groundwater microbial community. *mBio* **7**, e02234-15 (2016).
4. Böhm, M.-E., Huptas, C., Krey, V. M. & Scherer, S. Massive horizontal gene transfer, strictly vertical inheritance and ancient duplications differentially shape the evolution of Bacillus cereus enterotoxin operons hbl, cytK and nhe. *BMC Evol. Biol.* **15**, 246 (2015).
5. Ikuma, K. & Gunsch, C. K. Genetic bioaugmentation as an effective method for in situ bioremediation: functionality of catabolic plasmids following conjugal transfers. *Bioengineered* **3**, 236–241 (2012).
6. Nagpal, S., Haque, M. M. & Mande, S. S. Vikodak — a modular framework for inferring functional potential of microbial communities from 16S metagenomic datasets. *PLoS ONE* **11**, e0148347 (2016).
7. Langille, M. G. I. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821 (2013).
8. Loper, J. E. et al. Comparative genomics of plant-associated Pseudomonas spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet.* **8**, e1002784 (2012).
9. Rasko, D. A. et al. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. *J. Bacteriol.* **190**, 6881–6893 (2008).
10. Waldor, M. K. & Mekalanos, J. J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910–1914 (1996).
11. Putze, J. et al. Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infect. Immun.* **77**, 4696–4703 (2009).
12. Fischer, S. et al. Indication of horizontal DNA gene transfer by extracellular vesicles. *PLoS ONE* **11**, e0163665 (2016).
13. McDaniel, L. D. et al. High frequency of horizontal gene transfer in the oceans. *Science* **330**, 50 (2010).
14. Aune, T. E. V. & Aachmann, F. L. Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. *Appl. Microbiol. Biotechnol.* **85**, 1301–1313 (2010).
15. Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N. & Wuertz, S. Studying plasmid horizontal transfer in situ: a critical review. *Nat. Rev. Microbiol.* **3**, 700–710 (2005).
16. Hausner, M. & Wuertz, S. High rates of conjugation in bacterial biofilms as determined by quantitative in situ analysis. *Appl. Environ. Microbiol.* **65**, 3710–3713 (1999).
17. Keen, N. T., Tamaki, S., Kobayashi, D. & Trollinger, D. Improved broad-host-range plasmids for DNA cloning in gram-negative bacteria. *Gene* **70**, 191–197 (1988).
18. Eisen, J. A. Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Curr. Opin. Genet. Dev.* **10**, 606–611 (2000).
19. Koonin, E. V., Makarova, K. S. & Aravind, L. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **55**, 709–742 (2001).
20. Bertelli, C., Tilley, K. E. & Brinkman, F. S. L. Microbial genomic island discovery, visualization and analysis. *Brief. Bioinform.* **20**, 1685–1698 (2019).
21. Gao, F. & Zhang, C.-T. GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. *Nucleic Acids Res.* **34**, W686–W691 (2006).
22. Baran, R. H. & Ko, H. Detecting horizontally transferred and essential genes based on dinucleotide relative abundance. *DNA Res. Int. J. Rapid Publ. Rep. Genes. Genomes* **15**, 267–276 (2008).
23. Pal, C., Bengtsson-Palme, J., Kristiansson, E. & Larsson, D. G. J. Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC Genomics* **16**, 964 (2015).
24. Di Venanzio, G. et al. Urinary tract colonization is enhanced by a plasmid that regulates uropathogenic Acinetobacter baumannii chromosomal genes. *Nat. Commun.* **10**, 2763 (2019).
25. Landsberger, M. et al. Anti-CRISPR phages cooperate to overcome CRISPR-Cas immunity. *Cell* **174**, 908–916.e12 (2018).
26. Brito, I. L. et al. Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435–439 (2016).
27. Ling, A. & Cordaux, R. Insertion sequence inversions mediated by ectopic recombination between terminal inverted repeats. *PLoS ONE* **5**, e15654 (2010).
28. Pop, M. Genome assembly reborn: recent computational challenges. *Brief. Bioinform.* **10**, 354–366 (2009).
29. Zeevi, D. et al. Structural variation in the gut microbiome associates with host health. *Nature* **568**, 43–48 (2019).
Structural variants within bacterial genomes, largely due to integrated MGEs, found within the gut microbiota are associated with host phenotypes.
30. Kingsford, C., Schatz, M. C. & Pop, M. Assembly complexity of prokaryotic genomes using short reads. *BMC Bioinforma.* **11**, 21 (2010).

31. Leplae, R., Lima-Mendez, G. & Toussaint, A. ACLAME: a classification of mobile genetic elements, update 2010. *Nucleic Acids Res.* **38**, D57–D61 (2010).
32. Jiang, X., Hall, A. B., Xavier, R. J. & Alm, E. J. Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. *PLoS ONE* **14**, e0223680 (2019).
33. Carattoli, A. et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).
34. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).
35. Wagner, A., Lewis, C. & Bichsel, M. A survey of bacterial insertion sequences using IScan. *Nucleic Acids Res.* **35**, 5284–5293 (2007).
36. Rozov, R. et al. Recycler: an algorithm for detecting plasmids from de novo assembly graphs. *Bioinformatics* **33**, 475–482 (2017).
37. Krawczyk, P. S., Lipinski, L. & Dziembowski, A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res.* **46**, e35 (2018).
38. Arredondo-Alonso, S., Willems, R. J., van Schaik, W. & Schürch, A. C. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb. Genomics* **3**, e000128 (2017).
39. Jørgensen, T. S., Xu, Z., Hansen, M. A., Sørensen, S. J. & Hansen, L. H. Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metamobilome. *PLoS ONE* **9**, e87924 (2014).
40. Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
41. Gregory, A. C. et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **177**, 1109–1123.e14 (2019).
42. Bin Jang, H. et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
43. Song, W., Wemheuer, B., Zhang, S., Steensen, K. & Thomas, T. MetaCHIP: community-level horizontal gene transfer identification through the combination of best-match and phylogenetic approaches. *Microbiome* **7**, 36 (2019).
44. Franzosa, E. A. et al. Identifying personal microbiomes using metagenomic codes. *Proc. Natl Acad. Sci. USA* **112**, E2930–E2938 (2015).
45. Brito, I. L. et al. Transmission of human-associated microbiota along family and social networks. *Nat. Microbiol.* **4**, 964–971 (2019).
46. Seiler, E., Trappe, K. & Renard, B. Y. Where did you come from, where did you go: refining metagenomic analysis tools for horizontal gene transfer characterisation. *PLOS Comput. Biol.* **15**, e1007208 (2019).
47. Sieber, C. M. K. et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018).
48. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
49. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
50. Sharon, I. et al. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res.* **23**, 111–120 (2013).
51. Brooks, B. et al. Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. *Microbiome* **2**, 1 (2014).
52. Cleary, B. et al. Detection of low-abundance bacterial strains in metagenomic datasets by eigengenome partitioning. *Nat. Biotechnol.* **33**, 1053–1060 (2015).
53. Lim, E. S. et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* **21**, 1228–1234 (2015).
54. Duerkop, B. A. et al. Murine colitis reveals a disease-associated bacteriophage community. *Nat. Microbiol.* **3**, 1023–1031 (2018).
55. Norman, J. M. et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **160**, 447–460 (2015).
56. Reyes, A. et al. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc. Natl Acad. Sci. USA* **112**, 11941–11946 (2015).
57. Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L. & Rohwer, F. Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* **4**, 470–483 (2009).
58. Roux, S. et al. Minimum information about an uncultivated virus genome (MIUViG). *Nat. Biotechnol.* **37**, 29–37 (2019).
This article provides a coherent outline for evaluating metagenomic assembled phage genomes.
59. d'Humières, C. et al. A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach. *Sci. Rep.* **9**, 11331 (2019).
60. Castro-Mejía, J. L. et al. Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome* **3**, 64 (2015).
61. Kleiner, M., Hooper, L. V. & Duerkop, B. A. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics* **16**, 7 (2015).
62. Schulz, F. et al. Giant viruses with an expanded complement of translation system components. *Science* **356**, 82–85 (2017).
63. Dutilh, B. E. et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* **5**, 4498 (2014).
64. Devoto, A. E. et al. Megaphages infect Prevotella and variants are widespread in gut microbiomes. *Nat. Microbiol.* **4**, 693–700 (2019).
65. Tamminen, M., Virta, M., Fani, R. & Fondi, M. Large-scale analysis of plasmid relationships through gene-sharing networks. *Mol. Biol. Evol.* **29**, 1225–1240 (2012).
66. Brown Kav, A. et al. Insights into the bovine rumen plasmidome. *Proc. Natl Acad. Sci. USA* **109**, 5452–5457 (2012).
67. Li, A.-D., Li, L.-G. & Zhang, T. Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants. *Front. Microbiol.* **6**, 1025 (2015).
68. Jones, B. V. & Marchesi, J. R. Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nat. Methods* **4**, 55–61 (2007).
69. Delaney, S., Murphy, R. & Walsh, F. A comparison of methods for the extraction of plasmids capable of conferring antibiotic resistance in a human pathogen from complex broiler cecal samples. *Front. Microbiol.* **9**, 1731 (2018).
70. Hidalgo-Cantabrana, C., Sanozky-Dawes, R. & Barrangou, R. Insights into the human virome using CRISPR spacers from microbiomes. *Viruses* **10**, 479 (2018).
71. Paez-Espino, D. et al. Uncovering Earth's virome. *Nature* **536**, 425–430 (2016).
72. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**, 1047–1050 (2008).
73. Stern, A., Mick, E., Tirosh, I., Sagy, O. & Sorek, R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res.* **22**, 1985–1994 (2012).
This article shows an examination of CRISPR spacers across a population can be used to identify prevalent and individual-specific MGEs.
74. Minot, S. et al. Rapid evolution of the human gut virome. *Proc. Natl Acad. Sci. USA* **110**, 12450–12455 (2013).
75. Pärnänen, K. et al. Evaluating the mobility potential of antibiotic resistance genes in environmental resistomes without metagenomics. *Sci. Rep.* **6**, 35790 (2016).
76. Shintani, M., Sanchez, Z. K. & Kimbara, K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Front. Microbiol.* **6**, 242 (2015).
77. Kuleshov, V. et al. Synthetic long-read sequencing reveals intraspecies diversity in the human microbiome. *Nat. Biotechnol.* **34**, 64–69 (2016).
78. Zlitni, S. et al. Strain-resolved microbiome sequencing reveals mobile elements that drive bacterial competition on a clinical timescale. *Genome Med.* **12**, 50 (2020).
79. Suzuki, Y. et al. Long-read metagenomic exploration of extrachromosomal mobile genetic elements in the human gut. *Microbiome* **7**, 119 (2019).
80. Amarasinghe, S. L. et al. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 30 (2020).
81. Moss, E. L., Maghini, D. G. & Bhatt, A. S. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat. Biotechnol.* **38**, 701–707 (2020).
82. Beaulaurier, J., Schadt, E. E. & Fang, G. Deciphering bacterial epigenomes using modern sequencing technologies. *Nat. Rev. Genet.* **20**, 157–172 (2019).
83. Rand, A. C. et al. Mapping DNA methylation with high throughput nanopore sequencing. *Nat. Methods* **14**, 411–413 (2017).
84. Somerville, V. et al. Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *BMC Microbiol.* **19**, 143 (2019).
85. Cohen, N. R. et al. A role for the bacterial GATC methylome in antibiotic stress survival. *Nat. Genet.* **48**, 581–586 (2016).
86. Leonard, M. T. et al. The methylome of the gut microbiome: disparate Dam methylation patterns in intestinal Bacteroides dorei. *Front. Microbiol.* **5**, 361 (2014).
87. Smillie, C. S. et al. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**, 241–244 (2011).
88. Zhao, S. et al. Adaptive evolution within gut microbiomes of healthy people. *Cell Host Microbe* **25**, 656–667.e8 (2019).
89. Bonham, K. S., Wolfe, B. E. & Dutton, R. J. Extensive horizontal gene transfer in cheese-associated bacteria. *eLife* **6**, e22144 (2017).
90. Forster, S. C. et al. A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat. Biotechnol.* **37**, 186–192 (2019).
91. Poyet, M. et al. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nat. Med.* **25**, 1442–1452 (2019).
92. De Gelder, L., Ponciano, J. M., Joyce, P. & Top, E. M. Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. *Microbiol. Read. Engl.* **153**, 452–463 (2007).
93. del Campo, I. et al. Determination of conjugation rates on solid surfaces. *Plasmid* **67**, 174–182 (2012).
94. Arevalo, P., VanInsberghe, D., Elsherbini, J., Gore, J. & Polz, M. F. A reverse ecology approach based on a biological definition of microbial populations. *Cell* **178**, 820–834.e14 (2019).
95. Durrant, M. G., Li, M. M., Siranosian, B. & Bhatt, A. S. A Bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. *Cell Host Microbe* **27**, 140–153 (2020).
96. Daubin, V., Lerat, E. & Perrière, G. The source of laterally transferred genes in bacterial genomes. *Genome Biol.* **4**, R57 (2003).
97. Stepanauskas, R. et al. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat. Commun.* **8**, 84 (2017).
98. de Bourcy, C. F. A. et al. A quantitative comparison of single-cell whole genome amplification methods. *PLoS ONE* **9**, e105585 (2014).
99. Xu, L., Brito, I. L., Alm, E. J. & Blainey, P. C. Virtual microfluidics for digital quantification and single-cell sequencing. *Nat. Methods* **13**, 759–762 (2016).
100. Labonté, J. M. et al. Single cell genomics indicates horizontal gene transfer and viral infections in a deep subsurface Firmicutes population. *Front. Microbiol.* **6**, 349 (2015).
101. Licht, T. R., Christensen, B. B., Krogfelt, K. A. & Molin, S. Plasmid transfer in the animal intestine and other dynamic bacterial populations: the role of community structure and environment. *Microbiol. Read. Engl.* **145**, 2615–2622 (1999).
102. Haagensen, J. A. J., Hansen, S. K., Johansen, T. & Molin, S. In situ detection of horizontal transfer of mobile genetic elements. *FEMS Microbiol. Ecol.* **42**, 261–268 (2002).
103. Stalder, T. & Top, E. Plasmid transfer in biofilms: a perspective on limitations and opportunities. *NPJ Biofilms Microbiomes* **2**, 16022 (2016).
104. Diard, M. et al. Inflammation boosts bacteriophage transfer between Salmonella spp. *Science* **355**, 1211–1215 (2017).
Barcoded phages are used to monitor transfer between two *Salmonella* phages transduced as a result of inflammation.
105. Klümper, U. et al. Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. *ISME J.* **9**, 934–945 (2015).
Using fluorescence-activated cell sorting, the authors track a GFP-tagged plasmid between species within a soil community.

106. de Jonge, P. A. et al. Adsorption sequencing as a rapid method to link environmental bacteriophages to hosts. *iScience* **23**, 101439 (2020).
107. Deng, L. et al. Viral tagging reveals discrete populations in Synechococcus viral genome sequence space. *Nature* **513**, 242–245 (2014).
108. Deng, L. et al. Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *mBio* **3**, e00373 (2012).
109. Džunková, M. et al. Defining the human gut host-phage network through single-cell viral tagging. *Nat. Microbiol.* **4**, 2192–2203 (2019).
This article presents a community-level variation on viral tagging, in which phages are isolated from a gut microbiome sample, fluorescently labelled and restored with the gut bacteria to allow sorting and single-cell sequencing, thereby linking phages with their host genomes.
110. Munck, C., Sheth, R. U., Freedberg, D. E. & Wang, H. H. Real-time capture of horizontal gene transfers from gut microbiota by engineered CRISPR-Cas acquisition. Preprint at *bioRxiv* https://doi.org/10.1101/492751 (2018).
111. Burton, J. N., Liachko, I., Dunham, M. J. & Shendure, J. Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3* **4**, 1339–1346 (2014).
112. Beitel, C. W. et al. Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ* **2**, e415 (2014).
113. Bickhart, D. M. et al. Assignment of virus and antimicrobial resistance genes to microbial hosts in a complex microbial community by combined long-read assembly and proximity ligation. *Genome Biol.* **20**, 153 (2019).
114. Stalder, T., Press, M. O., Sullivan, S., Liachko, I. & Top, E. M. Linking the resistome and plasmidome to the microbiome. *ISME J.* **13**, 2437–2446 (2019).
115. Press, M. O. et al. Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. Preprint at *bioRxiv* https://doi.org/10.1101/198713 (2017).
116. Yaffe, E. & Relman, D. A. Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. *Nat. Microbiol.* https://doi.org/10.1038/s41564-019-0625-0 (2019).
117. Kent, A. G., Vill, A. C., Shi, Q., Satlin, M. J. & Brito, I. L. Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. *Nat. Commun.* **11**, 4379 (2020).
118. Ignacio-Espinoza, J. C. et al. Ribosome-linked mRNA-rRNA chimeras reveal active novel virus host associations. Preprint at *bioRxiv* https://doi.org/10.1101/2020.10.30.332502 (2020).
119. Spencer, S. J. et al. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J.* **10**, 427–436 (2016).
120. Hultman, J. et al. Host range of antibiotic resistance genes in wastewater treatment plant influent and effluent. *FEMS Microbiol. Ecol.* **94**, fiy038 (2018).
121. Diebold, P. J., New, F. N., Hovan, M., Satlin, M. J. & Brito, I. L. Linking plasmid-based beta-lactamases to their bacterial hosts using single-cell fusion PCR. Preprint at *bioRxiv* https://doi.org/10.1101/2021.01.22.427834 (2021).
The authors develop an easy-to-use, versatile platform for linking specific plasmid genes to taxonomic markers.
122. Ronda, C., Chen, S. P., Cabral, V., Yaung, S. J. & Wang, H. H. Metagenomic engineering of the mammalian gut microbiome in situ. *Nat. Methods* **16**, 167–170 (2019).
123. Langelier, C. et al. Microbiome and antimicrobial resistance gene dynamics in international travelers. *Emerg. Infect. Dis.* **25**, 1380–1383 (2019).
124. Zwanzig, M. et al. Mobile compensatory mutations promote plasmid survival. *mSystems* https://doi.org/10.1128/mSystems.00186-18 (2019).
125. Kintses, B. et al. Phylogenetic barriers to horizontal transfer of antimicrobial peptide resistance genes in the human gut microbiota. *Nat. Microbiol.* **4**, 447–458 (2019).
126. Rabinovich, L., Sigal, N., Borovok, I., Nir-Paz, R. & Herskovits, A. A. Prophage excision activates Listeria competence genes that promote phagosomal escape and virulence. *Cell* **150**, 792–802 (2012).
127. Faure, G. et al. CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.* **17**, 513–525 (2019).
128. Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C. & Segata, N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res.* **27**, 626–638 (2017).
129. Wang, R. et al. The global distribution and spread of the mobilized colistin resistance gene mcr-1. *Nat. Commun.* **9**, 1179 (2018).
130. Fulsundar, S. et al. Gene transfer potential of outer membrane vesicles of Acinetobacter baylyi and effects of stress on vesiculation. *Appl. Environ. Microbiol.* **80**, 3469–3483 (2014).
131. Tran, F. & Boedicker, J. Q. Genetic cargo and bacterial species set the rate of vesicle-mediated horizontal gene transfer. *Sci. Rep.* **7**, 8813 (2017).
132. Domingues, S. & Nielsen, K. M. Membrane vesicles and horizontal gene transfer in prokaryotes. *Curr. Opin. Microbiol.* **38**, 16–21 (2017).
133. Li, J. et al. Antibiotic treatment drives the diversification of the human gut resistome. *Genomics Proteom. Bioinforma.* **17**, 39–51 (2019).
134. Baltrus, D. A. Exploring the costs of horizontal gene transfer. *Trends Ecol. Evol.* **28**, 489–495 (2013).
135. van Houte, S. et al. The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532**, 385–388 (2016).
136. Chaikeeratisak, V. et al. Assembly of a nucleus-like structure during viral replication in bacteria. *Science* **355**, 194–197 (2017).
137. Trappe, K., Marschall, T. & Renard, B. Y. Detecting horizontal gene transfer by mapping sequencing reads across species boundaries. *Bioinforma. Oxf. Engl.* **32**, i595–i604 (2016).
138. Durrant, M. G., Li, M. M., Siranosian, B. A., Montgomery, S. B. & Bhatt, A. S. A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. *Cell Host Microbe* **28**, 767 (2020).
139. Sentchilo, V. et al. Community-wide plasmid gene mobilization and selection. *ISME J.* **7**, 1173–1186 (2013).
140. Beaulaurier, J. et al. Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. *Nat. Biotechnol.* **36**, 61–69 (2018).
141. Marbouty, M., Baudry, L., Cournac, A. & Koszul, R. Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay. *Sci. Adv.* **3**, e1602105 (2017).
142. Marbouty, M. et al. Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *eLife* **3**, e03318 (2014).

Acknowledgements
The author is a Packard Fellow of Science and Engineering, a Pew Biomedical Scholar and a Sloan Foundation Research Fellow. This research was funded by the Sloan Foundation Microbes in the Built Environment programme (2018-11009), the US National Sciences Foundation (ABI 1661338, EAGER 1650122), the US National Institutes of Health (1DP2HL141007), the US Department of Agriculture (BRAG:2017-03796) and the Bill & Melinda Gates Grand Challenges Program (OPP1161064). The author thanks members of the Brito laboratory for comments on the manuscript.

Competing interests
The author declares no competing interests.

Peer review information
Nature Reviews Microbiology thanks H. Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.