

# Lawrence Berkeley National Laboratory

## Recent Work

### Title

Large Circular Plasmids from Groundwater Plasmidomes Span Multiple Incompatibility Groups and Are Enriched in Multimetal Resistance Genes.

### Permalink

<https://escholarship.org/uc/item/2jb7305g>

### Journal

mBio, 10(1)

### ISSN

2150-7511

### Authors

Kothari, Ankita  
Wu, Yu-Wei  
Chandonia, John-Marc  
et al.

### Publication Date

2019-02-01







### DOI

10.1128/mbio.02899-18

Peer reviewed



# Large Circular Plasmids from Groundwater Plasmidomes Span Multiple Incompatibility Groups and Are Enriched in Multimetal Resistance Genes

 Ankita Kothari,<sup>a</sup>  Yu-Wei Wu,<sup>a,b</sup>  John-Marc Chandonia,<sup>c,d</sup>  Marimikel Charrier,<sup>a</sup> Lara Rajeev,<sup>a</sup> Andrea M. Rocha,<sup>e,\*</sup> Dominique C. Joyner,<sup>f</sup>  Terry C. Hazen,<sup>e,f</sup> Steven W. Singer,<sup>a</sup>  Aindrila Mukhopadhyay<sup>a,c</sup>

<sup>a</sup>Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California, USA

<sup>b</sup>Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

<sup>c</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA

<sup>d</sup>Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA

<sup>e</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

<sup>f</sup>Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, Tennessee, USA

**ABSTRACT** Naturally occurring plasmids constitute a major category of mobile genetic elements responsible for harboring and transferring genes important in survival and fitness. A targeted evaluation of plasmidomes can reveal unique adaptations required by microbial communities. We developed a model system to optimize plasmid DNA isolation procedures targeted to groundwater samples which are typically characterized by low cell density (and likely variations in the plasmid size and copy numbers). The optimized method resulted in successful identification of several hundred circular plasmids, including some large plasmids (11 plasmids more than 50 kb in size, with the largest being 1.7 Mb in size). Several interesting observations were made from the analysis of plasmid DNA isolated in this study. The plasmid pool (plasmidome) was more conserved than the corresponding microbiome distribution (16S rRNA based). The circular plasmids were diverse as represented by the presence of seven plasmid incompatibility groups. The genes carried on these groundwater plasmids were highly enriched in metal resistance. Results from this study confirmed that traits such as metal, antibiotic, and phage resistance along with toxin-antitoxin systems are encoded on abundant circular plasmids, all of which could confer novel and advantageous traits to their hosts. This study confirms the ecological role of the plasmidome in maintaining the latent capacity of a microbiome, enabling rapid adaptation to environmental stresses.

**IMPORTANCE** Plasmidomes have been typically studied in environments abundant in bacteria, and this is the first study to explore plasmids from an environment characterized by low cell density. We specifically target groundwater, a significant source of water for human/agriculture use. We used samples from a well-studied site and identified hundreds of circular plasmids, including one of the largest sizes reported in plasmidome studies. The striking similarity of the plasmid-borne ORFs in terms of taxonomical and functional classifications across several samples suggests a conserved plasmid pool, in contrast to the observed variability in the 16S rRNA-based microbiome distribution. Additionally, the stress response to environmental factors has stronger conservation via plasmid-borne genes as marked by abundance of metal resistance genes. Last, identification of novel and diverse plasmids enriches the existing plasmid database(s) and serves as a paradigm to increase the repertoire of biological parts that are available for modifying novel environmental strains.

**KEYWORDS** antibiotic resistance, *mer*, mercury resistance, metal resistance, native plasmids, plasmidome

**Citation** Kothari A, Wu Y-W, Chandonia J-M, Charrier M, Rajeev L, Rocha AM, Joyner DC, Hazen TC, Singer SW, Mukhopadhyay A. 2019. Large circular plasmids from groundwater plasmidomes span multiple incompatibility groups and are enriched in multimetal resistance genes. *mBio* 10:e02899-18. <https://doi.org/10.1128/mBio.02899-18>.

**Editor** Stephen Carlyle Winans, Cornell University

**Copyright** © 2019 Kothari et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Aindrila Mukhopadhyay, [amukhopadhyay@lbl.gov](mailto:amukhopadhyay@lbl.gov).

\* Present address: Andrea M. Rocha, Geosyntec Consultants, Inc., Knoxville, Tennessee, USA.

This article is a direct contribution from a Fellow of the American Academy of Microbiology. Solicited external reviewers: Gary Saylor, University of Tennessee at Knoxville; Himadri Pakrasi, Washington University.

**Received** 31 December 2018

**Accepted** 11 January 2019

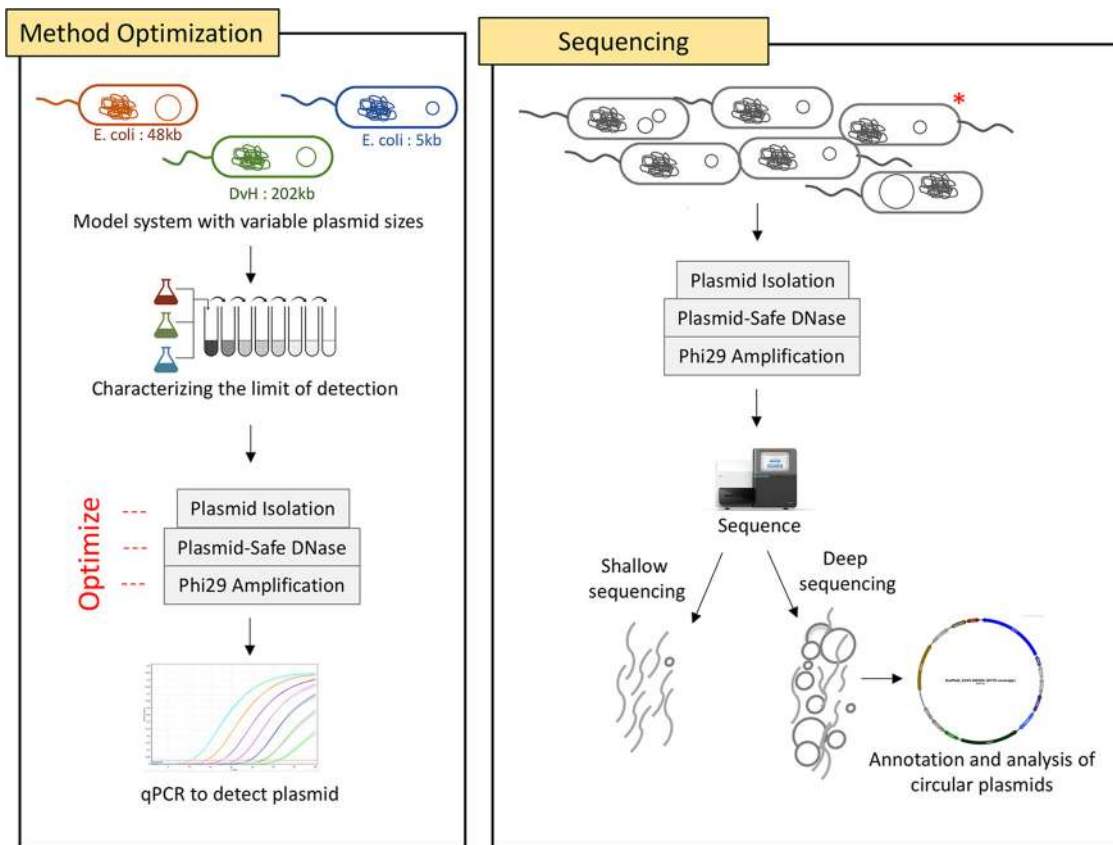
**Published** 26 February 2019

Plasmids are important in horizontal gene transfer and are critical in facilitating genome restructuring by providing a mechanism for distributing genes that provide a selective advantage to their host (1). Typically, plasmids have a modular structure, containing several functional genetic modules. Plasmids are known to vary from 5 to 500 kb in size, although plasmids as small as 2 kb (2–4) to as large as more than 1 Mb in size (5, 6) have been reported. Historically, environmental plasmid studies focused on exogenous isolation of plasmids studied via mating experiments (7, 8) or via plasmid isolation from bacterial strains that can be cultured (9–12). A previous study on groundwater samples revealed the presence of plasmids in strains that could be cultured, indicating the presence of plasmids in bacteria from low-cell-density environments (9). Given that it is well established that only 1% of bacteria on Earth can be readily cultivated (13), a lot remains unexplored, establishing the need to explore plasmids by a cultivation-independent method. More recently, with affordable DNA sequencing technologies, methods have been developed to specifically isolate and sequence circular plasmid DNA. The plasmidome is described to be the entire plasmid content in an environment that is resolved by metagenomic approaches during high-throughput-sequencing experiments (14) and thus circumvents the need to culture environmental bacteria. It identifies all plasmid types—conjugative, mobilizable, and nonmobilizable. Such plasmidome analyses have been performed in cow rumen (15, 16), rat cecum (17), soil (18), and activated sludge (19, 20), samples that are abundant in bacteria. To the best of our knowledge, plasmidomes have not been explored in low-cell-density environments. Due to the role of plasmids in environmental stress adaptation, we explore the plasmids in groundwater typically characterized by low cell counts but diverse and dynamic microbiomes (21). Here we examine samples from the Oak Ridge Field Research Center (ORFRC) site at the Y-12 Federal Security Complex in Oak Ridge, TN, which is a widely studied and characterized model groundwater environment (22–25). To explore the plasmidome of groundwater known to have a fluctuating microbial community along with lowered cell counts (26, 27), we modified known plasmid DNA isolation methods. The goal of this study was to discover the incidence, distribution, and function of plasmids from this site and to develop a foundation to explore the plasmidome of low-cell-density environments. We present the plasmidome analyses from groundwater samples that resulted in the identification of several hundred circular plasmids bearing genes involved in plasmid replication, mobilization, and maintenance along with those that code for metal, antibiotic, and phage resistance, thus bestowing beneficial traits on the host.

## RESULTS AND DISCUSSION

The optimized plasmid DNA isolation methodology used in this study yielded the largest plasmid sizes reported in plasmidome studies. Since a previous culture-based study of native plasmids from groundwater environment revealed the presence of a large plasmid (202 kb) (9), we used strains containing large plasmids for the method optimization. The key steps that aided in the optimization of standard protocols (15, 16, 19) were (i) using a model system that contained a large 202-kb plasmid in comparison to the 65-kb (28) and 56-kb (29) plasmids used earlier and (ii) optimization of Phi29 amplification to better represent large plasmids (details in Fig. S1 in the supplemental material). The optimized method was used to isolate plasmid DNA from seven groundwater samples followed by shallow (five plasmid DNA libraries from samples A to E pooled) and deep (two plasmid DNA libraries from samples F and G not pooled) sequencing (Fig. 1). The resulting scaffold distribution showed that the majority of the genes were annotated to be bacterial in origin (Fig. 2). Additionally, the number of raw reads generated was an order of magnitude higher in the deeply sequenced samples, enabling a more comprehensive analysis.

We found that the plasmidome sequences across the groundwater samples were more conserved in contrast to the corresponding bacterial taxonomic distribution of these samples (Fig. 3). This pattern might have ecological significance in the role of plasmids in maintaining and transferring conserved key latent functionalities in an

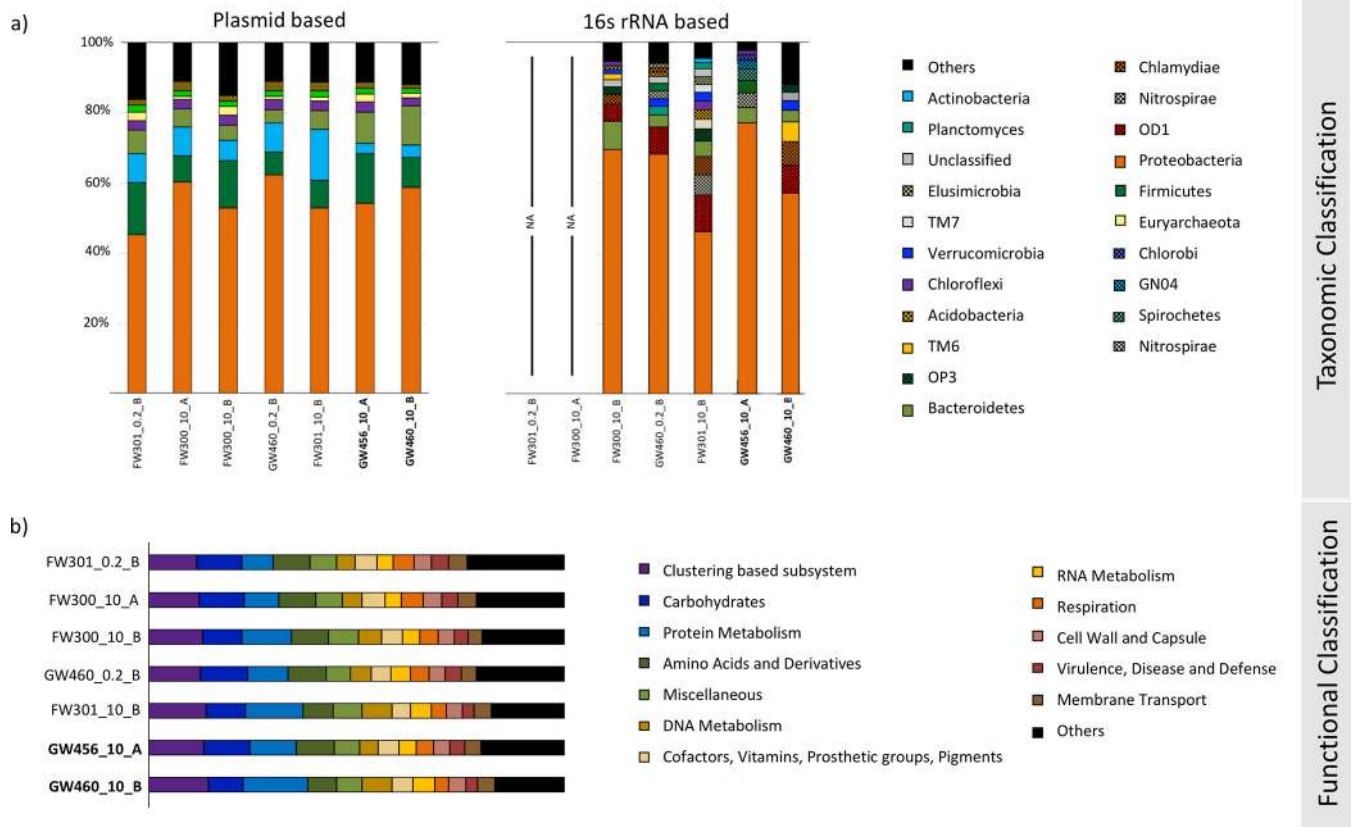


**FIG 1** An overview of plasmid DNA isolation and characterization. The plasmid DNA isolation method was optimized using a model system containing two *Escherichia coli* strains and one *Desulfovibrio vulgaris* Hildenborough (*DvH*) with 5-, 48-, and 202-kb plasmid sizes. Successful isolation of plasmid DNA was confirmed by qPCR against a specific plasmid-borne gene. The optimized method was used to isolate plasmid DNA from groundwater samples. The plasmid DNA libraries from certain samples were sequenced individually (deep sequencing), while others were pooled (shallow sequencing). Deep sequencing resulted in identification of several circular plasmids which were further analyzed. \*, the model system strains were spiked onto the environmental samples subjected to deep sequencing.

ecosystem and has also been reported for plasmidomes from soil and rumen environments (18). The most predominant bacterial phyla represented by ORFs from the groundwater plasmidome were *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, and *Actinobacteria*, with *Proteobacteria* being the most abundantly represented. Interestingly, these phyla are similar to the rumen plasmidome, albeit with a different order of predomi-

Sample Name	Shallow Sequencing					Deep Sequencing	
	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G
Number of raw reads	7,627,412	4,228,107	5,521,694	5,927,974	5,436,990	<b>34,376,686</b>	<b>76,385,700</b>
Number of assembled scaffolds	73,356	107,891	123,053	110,041	90,074	462,471	521,964
Longest scaffold size (bp)	5,485	5,485	1,538	1,285	1,925	1.16 Mb	0.63 Mb
Taxonomic breakdown							
Number of scaffolds > 2kb	1	1	0	0	0	42,440	32,232
MG-RAST ID	mgm4773096.3	mgm4773097.3	mgm4773098.3	mgm4773099.3	mgm4773100.3	mgm4690359.3	mgm4690360.3
Circular plasmids	-	-	-	-	-	67	548

**FIG 2** Plasmidome analysis after sequencing of seven groundwater samples from the ORFC.

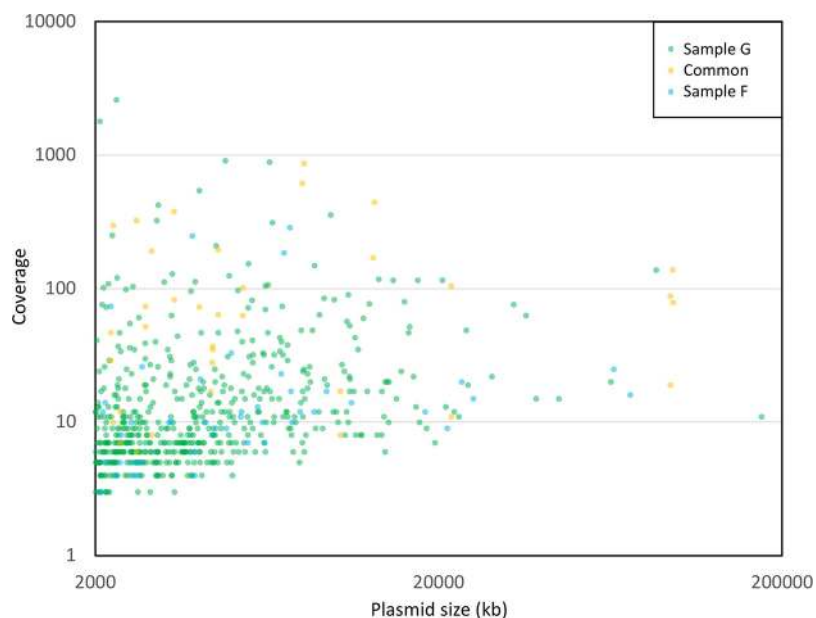


**FIG 3** (a) Taxonomic classification (normalized) of “all\_ scaffolds” into phyla based on plasmidome (MG-RAST) and 16S rRNA (QIIME) sequences. The MG-RAST analysis was based on the lowest common ancestor. (b) Functional classification (normalized) of “all\_ scaffolds” into SEED subsystem categories using MG-RAST (parameters:  $1e-5$  maximum E value cutoff, 60% maximum identity cutoff, 15-bp minimum alignment length cutoff). It is important to note that the phylogenetic assignment provides the taxonomic classification of the closest known gene homolog and is not indicative of the microbe that originally contained the plasmid at the ORFRC.

nance. The most highly represented functional categories (carbohydrates, amino acid metabolism, and clustering-based subsystems) were also similar to that reported in the plasmidome of rumen bacteria (28). This indicates that plasmids from diverse sources dominantly carry genes in similar phylogenetic and functional categories.

Metal and antibiotic resistance genes are one of the most frequently found phenotypic modules carried by bacterial plasmids (30). We found a high abundance of genes annotated to provide resistance to metals—copper, triclosan, arsenic, and mercury—with a large majority being proteobacterial in origin (Table S1). A previous study noted a high abundance of metal resistance genes in these groundwater samples (24) and hypothesized that they might be present on plasmids. The present study confirms that hypothesis. Among antibiotic resistance genes, those providing resistance to aminocoumarin, elfamycin, and bacitracin were predominant (Table S2). Overall, the plasmids were enriched in metal resistance genes compared to the antibiotic resistance genes. There was no detectable metal contamination in our groundwater samples, but given that these groundwater samples are close to a metal-contaminated site at ORFRC (22, 31), it is possible that the microbiome was exposed to metal stress at some point and/or that the dynamic nature of groundwater flow coupled with weather changes might lead to sporadic exposure to low levels of metal contaminants. Accordingly, we find metal resistance genes to be the most predominant, rather than antibiotic resistance genes predominant in activated sludge/wastewater (19, 20), genes required for survival under dairy conditions reported in *Lactococcus* (32), or genes providing an advantage in rumen environments (16).

The deeply sequenced groundwater samples resulted in identification of hundreds

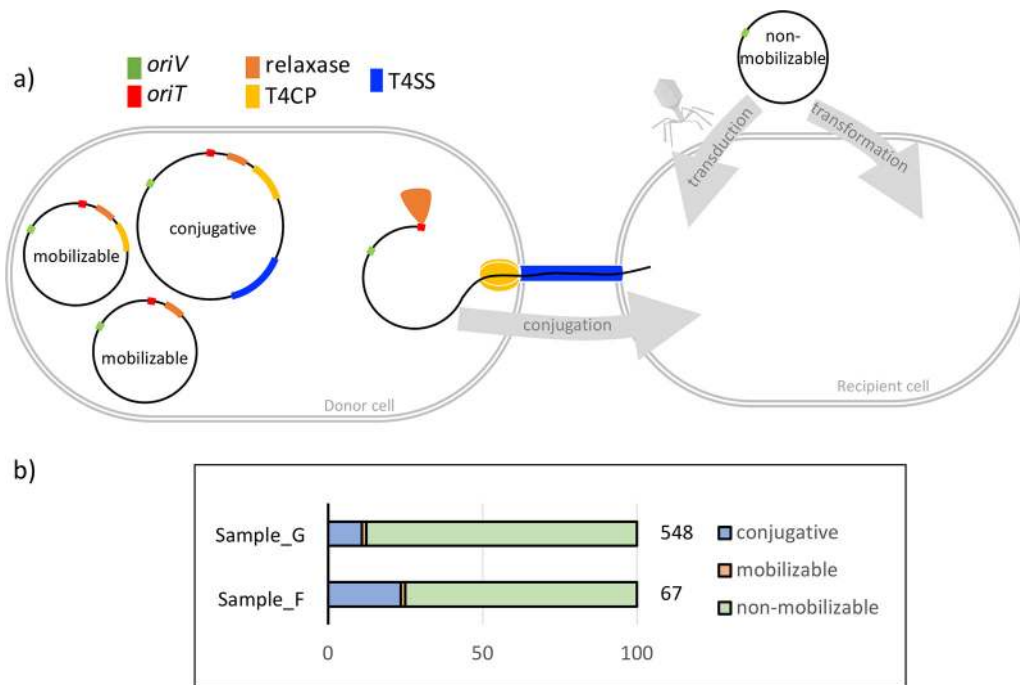


**FIG 4** Distribution of circular plasmids from groundwater samples F and G, along with plasmids common to both, based on sequence coverage and plasmid size.

of complete circular plasmid units of various sizes (67 from sample F and 548 from sample G, Fig. 4). Comparison of the circular plasmids with the ACLAME plasmid database resulted in about 70 to 80% of the ORFs having hits (Table S3), providing further confirmation of the presence of known plasmid-associated genes in our plasmidome data set. Toxin-antitoxin systems important in plasmid maintenance (33, 34) were also reported (Table S4). This study identified several circular plasmids encoding an interesting mix of features such as those providing advantageous traits to the hosts (metal and phage resistance), along with those that help in plasmid maintenance, replication, mobilization, and conjugation (a detailed list of the most abundant plasmids is found in Table S5a and b). Surprisingly, plasmids also carried ORFs annotated to possibly enable phages to invade bacteria (antirestriction protein and a putative phage protein [35]). Additionally, several plasmids were cryptic and could potentially serve as an important source for discovery of novel functional genes and replication systems (17).

The circular plasmids from groundwater were diverse in terms of the plasmid type classifications. They were classified as conjugative, mobilizable, or nonmobilizable (Fig. 5). As observed previously (36), the nonmobilizable plasmids were highly predominant in both samples. The circular plasmids from groundwater were diverse in encoding five out of six different relaxase groups and seven incompatibility groups into which plasmids are classified. Based on the relaxase classification (37),  $MOB_Q$  and  $MOB_P$  were the most abundant (Fig. 6a). In fact, the relaxase type follows a plasmid size-based distribution (Fig. 6b) as reported earlier. Based on incompatibility classification, plasmids belonging to group Inca/C were highly abundant (Fig. 6c). Interestingly, all the multimetal-resistant plasmids identified in this study were classified as the Inca/C  $cgPMLST$ , a group commonly associated with multidrug resistance plasmids (38).

It is noteworthy that this study identified several large plasmids (Fig. 4). The largest circular scaffolds identified were 2.96 and 1.74 Mb from the groundwater samples F and G, respectively. Curiously, the 2.96-Mb scaffold had similarity to known *Ralstonia* and *Pseudomonas* circular phage DNA. Given that this phage DNA was isolated because it was circular, it was removed from further plasmidome analysis but nevertheless emphasizes our methods being optimized for isolation of large circular DNA molecules. The 1.74-Mb plasmid p67 (Table S6) carried several metal resistance genes (for cobalt, zinc, cadmium, and copper) along with genes for plasmid mobilization and conjuga-

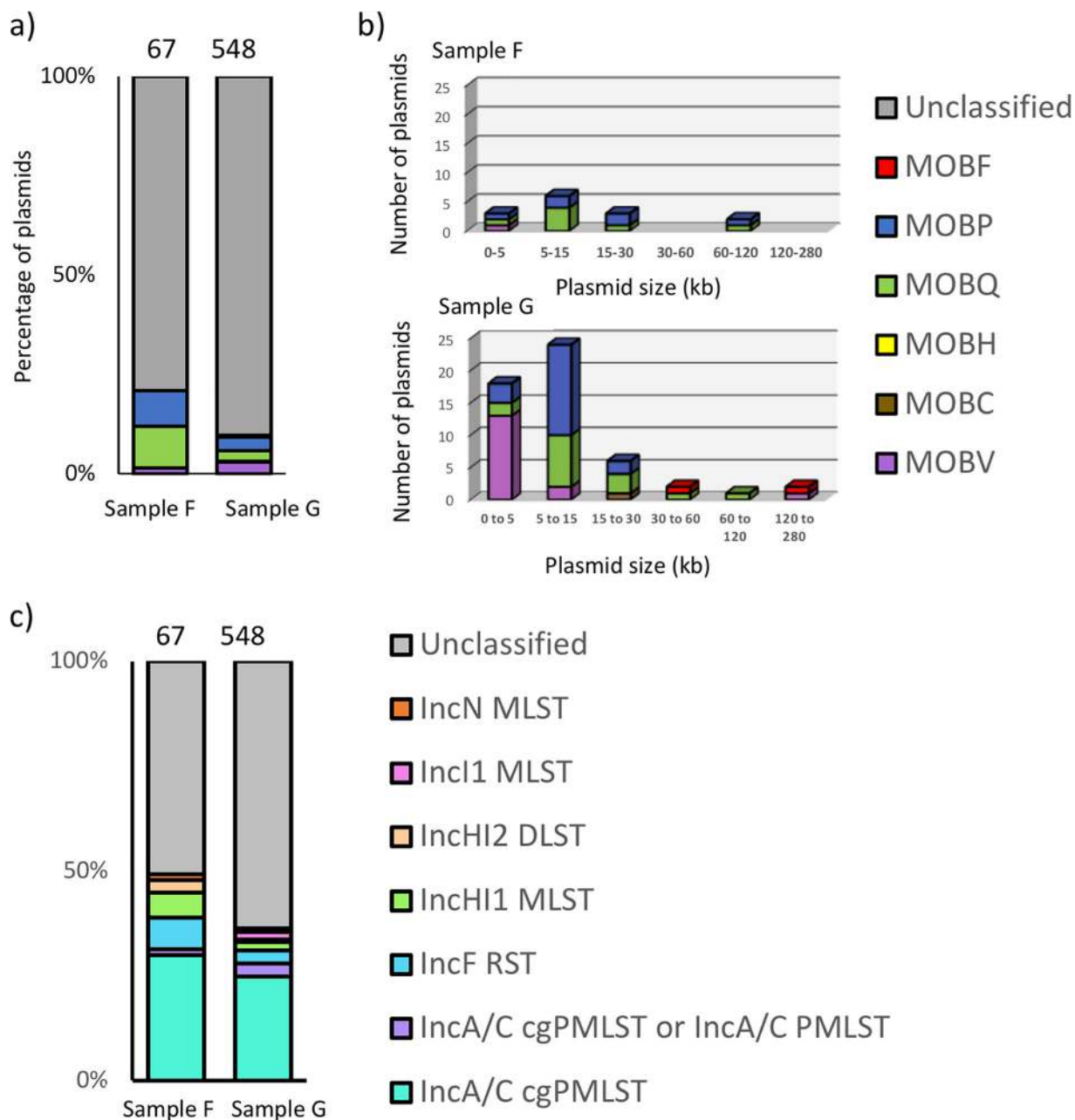


**FIG 5** (a) Schematic view of the genetic constitution of plasmids. A nonmobilizable plasmid typically encodes just the origin of vegetative replication (*OriV*), and the mobilizable plasmid encodes the origin of transfer (*OriT*) along with relaxase and possibly type IV coupling protein (T4CP). The conjugative plasmid (also called self-transmissible) codes for the type IV secretion system (T4SS) in addition to the above. The nonmobilizable plasmid relies on transformation or transduction for propagation whereas the mobilizable and conjugative plasmids can be propagated via conjugation. The latter process involves relaxase-based cleaving of the plasmid at *OriT*, followed by interactions with T4CP and T4SS which enable pumping of DNA into the recipient cell. Plasmid types are as described in reference 36. (b) Mobility of circular plasmids from samples F and G depicted as a percentage of the total circular plasmids in that sample.

tion. The plasmid was novel with the closest plasmid reported in literature (*Sphingobium baderi* DE-13 plasmid pDE1 from an herbicide-manufacturing factory in Kunshan, China [39]) depicting 94% identity but only 10% query coverage. This is one of the highest reported plasmid sizes captured by plasmidome studies (14), further providing evidence that our optimized method was better suited for isolation of large plasmid DNA molecules despite Phi29 amplification biases (29).

To gain a comprehensive insight into the diversity of the plasmidome with potential similarity observed between groundwater samples, a subset of plasmids encoding features of interest were graphically compared for further analysis (Fig. 7). Overall, there was higher similarity between the two groundwater samples than within the sample itself. Of the circular plasmids from samples F and G, 18 plasmids shared almost the exact same sequence (plasmid maps of a selected few are depicted in Fig. 8; a detailed list is in Table S5c). One explanation for the plasmidome similarity between the samples could be the geographical proximity of the sampling sites and the fact that the groundwater flow is continuous and dynamic. Additionally, it may be that there are limited variations in the genetic modules that can constitute a plasmid. This is suggested by the presence of modules on the circular plasmids (e.g., plasmids p5343 and p67 described in this study) that show very high similarity to other plasmids reported from diverse geographical locations across the globe. Additional explanations include bias in plasmid DNA extraction methodology which might preferentially isolate plasmid DNA from certain bacterial subpopulations or the Phi29 amplification bias, which preferentially amplifies a subset of plasmids.

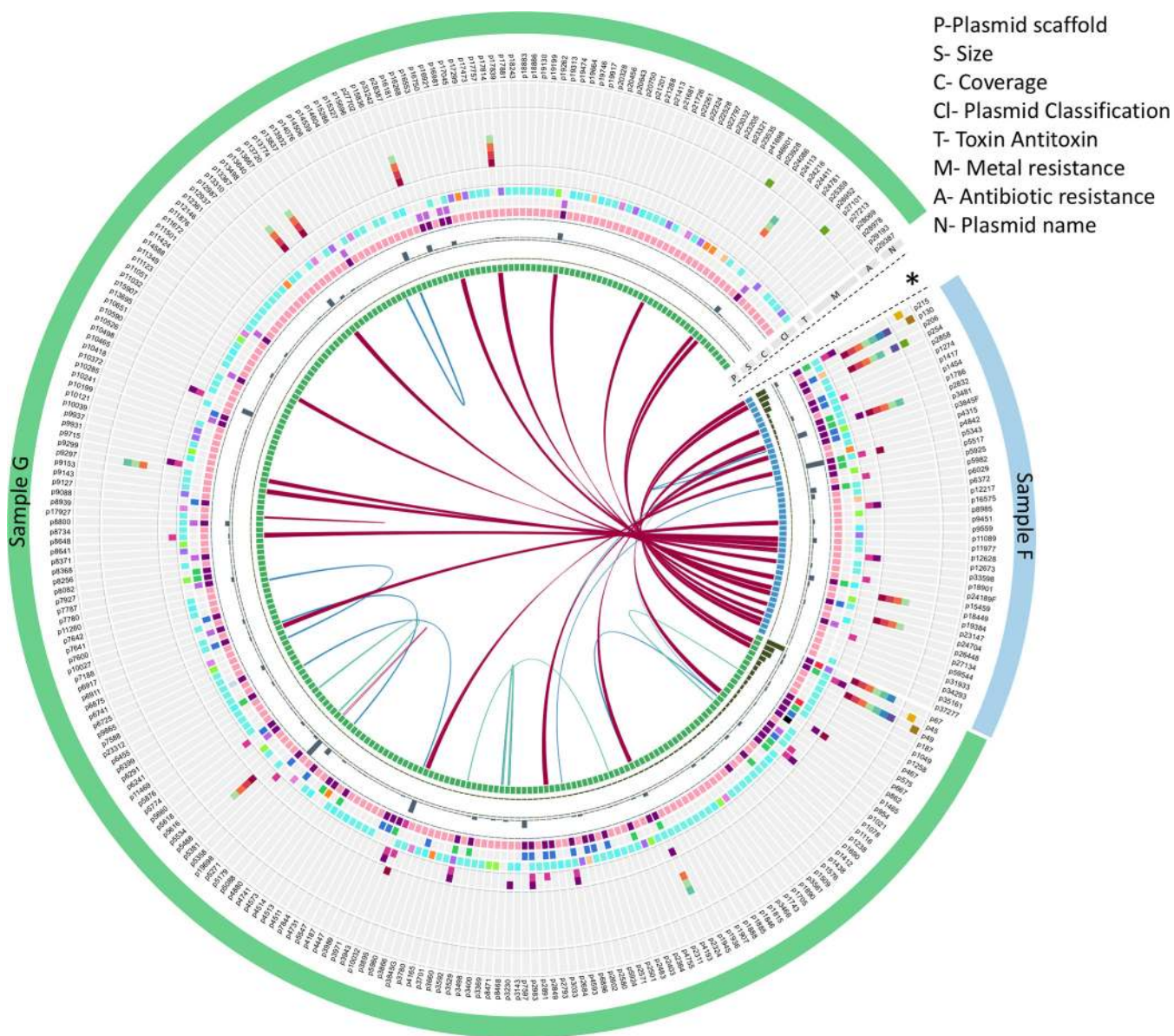
One of the most interesting plasmids identified was an 8-kb plasmid (p5343), highly abundant across samples, carrying genes annotated to be involved in mercury resistance along with plasmid mobilization and replication genes (Fig. 9). Most of the genes on this plasmid have homologs in the genus *Paracoccus*. The abundance of the genus



**FIG 6** (a and b) Classification of circular plasmids into relaxase/MOB type (a) with the relaxase type plotted as a function of plasmid size (b). (c) Classification of circular plasmids into incompatibility groups.

*Paracoccus* was less than 0.5% based on 16S rRNA distribution, perhaps indicating that the plasmid might have consequently horizontally transferred into other hosts and/or is maintained in the original host in high copy numbers. Alternatively, it is possible that even at 0.5% *Paracoccus* is the primary genus in this environment that hosts the plasmid, given that not all bacteria host plasmids, and the numbers can be explained by plasmid DNA extraction and amplification biases along with a high copy number of the plasmid. Performing a nucleotide BLAST search reveals that this plasmid can be broken into three modules. The first module spans from *mobA* to the helix-turn-helix-containing protein and exhibits homology to a rat gut plasmid in GenBank, accession no. [LN852881.1](https://www.ncbi.nlm.nih.gov/nuclbase/LN852881.1) (total size 12.9 kb). The original rat gut plasmid codes for certain hypothetical proteins in addition to the *ccdA/ccdB* type II toxin-antitoxin system genes. The second and third modules contain mercury resistance genes and depict homology



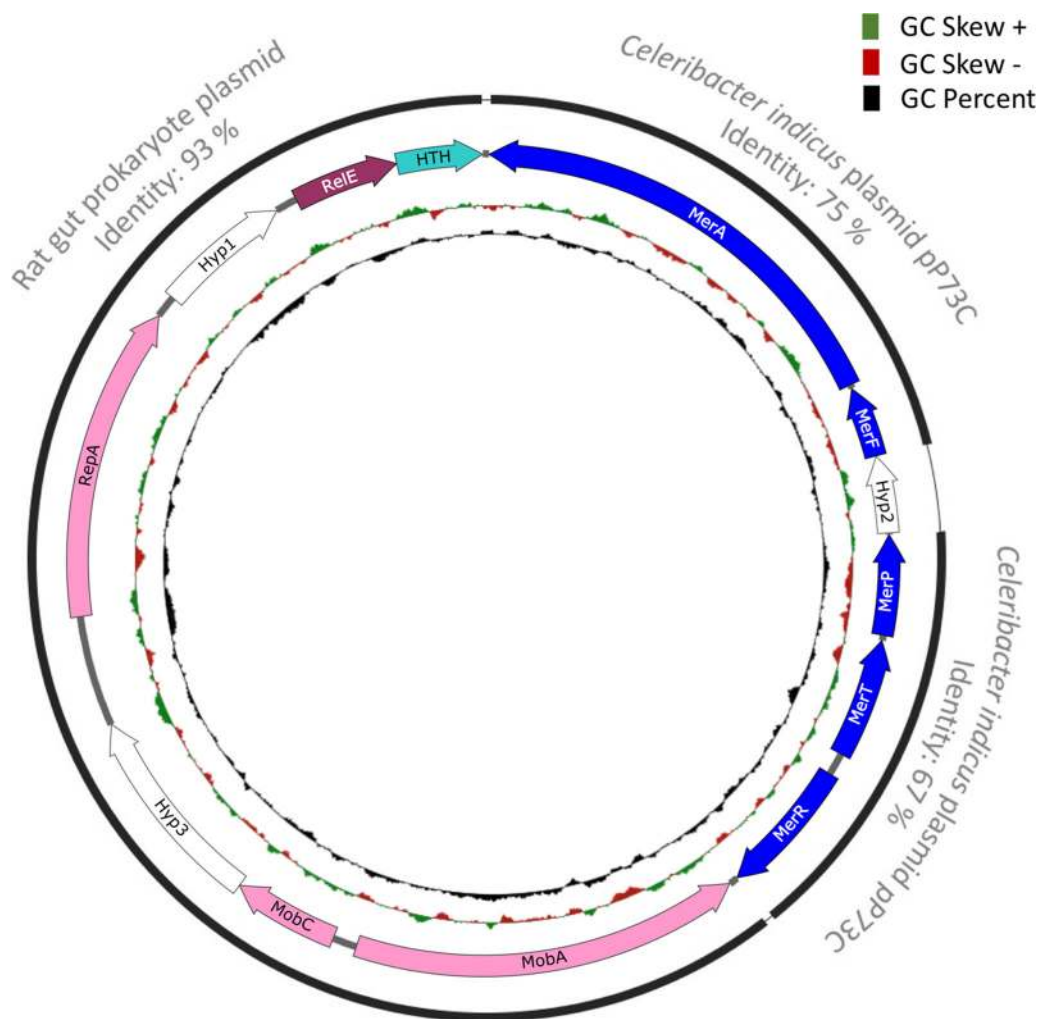


**FIG 7** Visualization of circular plasmids of interest from the groundwater samples F (blue) and G (green). Cutoff values for blast are E value of  $e^{-5}$  and minimum match of 1,000 bp (97%> = magenta, 90%> = blue, and 80%> = teal). Rings 2 and 3 depict histograms corresponding to the scaffold size and sequence coverage, respectively. Ring 4 depicts plasmid mobility (the colors light pink, pink, and violet represent nonmobilizable, mobilizable, and conjugative, respectively), ring 5 depicts MOB type (the colors blue, green, purple, red, black, orange, and light blue represent MOBp, MOBq, MOBv, MOBF, MOBC, MOBB, and MOBT, respectively), ring 6 depicts incompatibility groups (the colors teal, purple, blue, green, cream, pink, and orange represent IncA/C cgPMLST, IncA/C cgPMLST or IncA/C PMLST, IncF RST, Inc HI1 MLST, Inc HI2 DLST, IncI1 MLST, and IncN MLST, respectively), and rings 7 (pink) and 8 (purple) depict presence of toxin and antitoxin, respectively. Rings 9 to 15 depict presence of genes annotated (by KBase) to provide resistance to metals (the colors carmine, red, orange, light green, light blue, dark blue, and purple represent mercury, lead, cadmium, zinc, cobalt, copper, and arsenic, respectively), and rings 16 to 18 depict presence of genes annotated (by KBase) to provide resistance to antibiotics (the colors green, yellow, and brown represent acriflavine, polymyxin, and fosfomycin, respectively). The outermost ring depicts the plasmid name. The plot was created using Circos and Circoletto (64, 65). The asterisk indicates the circular plasmids not depicted on the figure because of lack of similarity with other plasmids and the lack of identifying features in rings 5 through 18.

to the native plasmid pP73c (total size, 122 kb) in *Celeribacter indicus* P73T. The P73T strain was isolated from a deep-sea sediment in the Indian Ocean (40). This study reveals that plasmids contain modules which are remarkably conserved in microbes in strikingly different environments, across the globe.

Intriguingly, even though mercury contamination is reported in certain sites nearby (31), these specific groundwater samples did not contain any detectable mercury, pointing to an interesting question—why do plasmids with metal resistance genes persist in such environments? Other circular plasmids also show the presence of genes





**FIG 9** Plasmid map of p5343. Genes encode the following proteins: MerA, mercuric ion reductase; MerF, mercuric ion uptake protein; Hyp, hypothetical protein; MerP, mercuric transport protein; MerR, regulator of mercury resistance genes; MobA, mobilization protein A; MobC, mobilization protein C; RepA, plasmid replication protein; HTH, helix-turn-helix domain protein; ReE, ReE toxin. The black lines indicate that the plasmid can be broken into different modules that show similarity to other previously reported plasmids (the closest NCBI BLAST hits with more than 92% query coverage are labeled in gray).

selective pressure (i.e., no use of antibiotics) helps in curtailing the spread of antibiotic resistance. Together, the plasmidome analysis of this site provides a broad insight into plasmid-borne functions and provides evidence that plasmid-mediated horizontal gene transfer plays a role in driving the evolution of this groundwater microbial community. Although certain observations made were unique to this site, the method to examine native plasmid DNA in low-cell-density environments and the broad trends observed are generalizable to all microbial communities.

## MATERIALS AND METHODS

**Sample collection.** Water samples were collected from groundwater wells of the Department of Energy's ORFRC, Tennessee (22) (well locations provided in Fig. S3 in the supplemental material). Given the difficulties in groundwater sampling, coupled with the fact that groundwater represents continuous dynamic water flow below the Earth's surface, these samples serve as survey snapshots rather than replicates. Prior to collection of samples, approximately 5 to 20 liters of groundwater was pumped until temperature, pH, conductivity, and oxidation-reduction (redox) values were stabilized to purge the well and the line of standing groundwater. Bulk groundwater measurements and geochemical sample collections (21) were conducted (Table S7). For 16S rRNA analysis and plasmid DNA isolation, a total of 8 and 5 liters of water, respectively, was filtered through 10- $\mu$ m and 0.2- $\mu$ m nylon filters (Sterlitech

Corporation, Kent, WA, USA). Filters were immediately stored on dry ice in 50-ml Falcon tubes until being transported to the  $-80^{\circ}\text{C}$  freezer.

**Geochemical measurements.** Temperature, pH, conductivity, redox, and dissolved oxygen were measured at the wellhead using an In-Situ Troll 9500 (In-situ Inc., Fort Collins, CO, USA). Sulfide and ferrous ion groundwater concentrations were determined using the USEPA methylene blue method (Hach 8131) and 1,10-phenanthroline method (Hach 8146), respectively, and analyzed with a field spectrophotometer (Hach DR 2800). All other biological and geochemical parameters were measured as previously described (21). Mercury analysis was performed on samples containing 25 ml groundwater and 25 ml glycerol by oxidation, purge, trap, and cold vapor atomic fluorescence spectrometry 1631E at ALS Environmental, Kelso, WA, USA.

**Plasmid DNA isolation optimization.** A model system of a 1:1:1 mixture of *Desulfovibrio vulgaris* Hildenborough (ATCC 29579) containing a 202-kb native plasmid (pDV1), *Escherichia coli* DH1 (ATCC 33849) containing a 48-kb fosmid (fSCF#19) (42), and *E. coli* strain J-2561 containing a 5-kb (pBbS5c) plasmid was prepared using cells grown to an optical density (at 600 nm) of 1. *Desulfovibrio* was grown in LS4D supplemented with 0.1% (wt/vol) yeast extract (43) while *E. coli* was grown in LB medium. This mixture was serially diluted 10-fold; stored at  $-80^{\circ}\text{C}$ ; and used to test, compare, and optimize plasmid detection via quantitative PCR (qPCR). Two alkaline hydrolysis methods were compared to preferentially isolate plasmid DNA (44, 45). Residual linear chromosomal DNA fragments were minimized by plasmid-safe ATP-dependent DNase (Epicentre, Madison, WI, USA) treatment for 24 to 48 h at  $37^{\circ}\text{C}$ . The presence of chromosomal DNA was tested by PCR using 16S rRNA universal primers (BAC338F, 5'-ACTCTACGG GAGGCAG-3', and BAC805R, 5'-GACTACCAGGGTATCTAATCC-3') (46). If 16S rRNA PCR product was visible on a 1% agarose gel, another overnight digestion reaction was performed until the product could no longer be visualized. The DNase was inactivated at  $70^{\circ}\text{C}$  for 30 min. The DNA was then amplified with Phi29 DNA polymerase (New England Biolabs, Ipswich, MA, USA) (16) at 4, 18, or  $30^{\circ}\text{C}$  for 168, 25, and 24 h, respectively. Plasmid DNA isolation was checked via qPCR against a specific plasmid-borne gene on all three plasmids. qPCR was performed using the SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA) per the manufacturer's protocol. Total DNA from *D. vulgaris* Hildenborough was used as a control for the 202-kb primers, and the plasmid DNA coding for pBbS5c was used as a control for the 5-Kb primers. Additionally, since our samples were essentially contained on filters, we tested whether the presence of filter interfered with plasmid DNA isolation. The filters were cut into smaller pieces and vortexed with beads in an attempt to improve plasmid recovery.

**Plasmid DNA isolation from environmental samples.** To extract DNA from bacteria on a filter, we modified an alkaline hydrolysis plasmid DNA isolation method (45) as described below. The filters from the groundwater samples A to E were thawed to room temperature, cut into pieces in a sterile petri dish using sterilized forceps and scissors, and split into two 50-ml Falcon tubes. The volumes of all reagents were multiplied 20 times to immerse each half-filter. Before the addition of lysozyme (Sigma-Aldrich, St. Louis, MO, USA), the samples were heated to  $37^{\circ}\text{C}$  with gentle inversion for 10 min and vortexed with 0.1-mm disrupter beads (Scientific Industries, Bohemia, NY, USA) at medium setting for 5 min. After the addition of sodium chloride, the liquid was transferred into 50-ml phase lock gel heavy tubes. A 14.5-ml amount of 25:24:1 phenol-chloroform-isoamyl alcohol was added to each tube, thoroughly mixed, and centrifuged for 5 min at  $1,500 \times g$  (Beckman Coulter Allegra 25R centrifuge). The upper phase was transferred to a fresh phase lock tube. A 14.5-ml amount of 24:1 chloroform-isoamyl alcohol was added and centrifuged for 5 min at  $1,500 \times g$ . The upper phase was transferred to a 50-ml Falcon tube and precipitated with an equal volume of isopropanol. The extractions from each half of the filter were recombined and incubated on ice for 1 h, followed by centrifugation for 5 min at  $8,000 \times g$ . The excess isopropanol was removed, and the pellet was resuspended in 1 ml of 10 mM Tris-1 mM EDTA, pH 7, transferred to a 1.6-ml tube, and dehydrated down to 50  $\mu\text{l}$  with a Vacufuge Plus (Eppendorf; V-AQ,  $45^{\circ}\text{C}$ ). The remnant linear DNA fragments were removed by plasmid-safe ATP-dependent DNase (Epicentre Biotechnologies, Madison, WI) at  $37^{\circ}\text{C}$  for 48 h with double the recommended ATP and enzyme amounts. The lack of chromosomal DNA contamination was confirmed by PCR with degenerate 16S rRNA primers. The plasmid DNA was amplified with Phi29 DNA polymerase (New England Biolabs, Ipswich, MA) as previously described (16) for 6 days at  $18^{\circ}\text{C}$ . This was followed by ethanol precipitation and use of a NanoDrop instrument to concentrate and quantify the DNA. For the plasmid DNA isolation from groundwater samples F and G, a variation to the method was that about  $1.33 \times 10^5$  cells of each plasmid-containing control strain were added to the filters to assess the efficiency of plasmid DNA isolation. The lack of chromosomal DNA contamination in plasmid DNA extracted from groundwater samples F and G was confirmed by PCR with degenerate 16S rRNA primers.

**Plasmid sequencing and bioinformatics.** For the groundwater samples A to E, the plasmid DNA libraries were pooled and sequenced on a single flow cell of the Illumina MiSeq reagent v3 kit (paired-end protocol), resulting in shallow sequencing. Plasmid DNA from two additional groundwater samples, F and G, was sequenced using the Illumina MiSeq reagent v3 kit (paired-end protocol) at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley (Berkeley, CA). An entire flow cell was used (no pooling of plasmid libraries), resulting in deep sequencing. The numbers of raw reads obtained by shallow and deep sequencing are depicted in Fig. 2. Trimmomatic 0.36 (<http://www.usadellab.org/cms/?page=trimmomatic>) was used to trim the reads with the following parameters: IlluminaClip: TruSeq3-PE.fa:2:30:10, Leading:3 Trailing:3 SlidingWindow:4:15 MinLen:36. As reported previously (17), IDBA-UD (47) was used for *de novo* read assembly with the parameter “-pre\_correction.” Assembled sequences were searched against the SILVA 16S rRNA database (48) using BLASTN; all scaffolds with  $>200$ -bp identity to 16S rRNA were removed from further analysis. The proportion of reads that mapped to scaffolds with 16S rRNA coding genes was 0.65% and 1.07% for samples F and G, respectively. We

removed the entire scaffolds and not just the reads mapping to 16S rRNA coding genes, to eliminate all potential chromosomal DNA contamination. To exclude the control plasmids in groundwater samples F and G, all sequences with more than 95% identity to these plasmids (minimum alignment length, 1,000 bp) were also removed. The resulting data set, referred to as “all\_scaffolds,” was analyzed with the MG-RAST server (49) using similarity to the SEED database (with a maximum E value of  $\leq 10^{-5}$ ) (50), generating taxonomic and functional assignments. All the sequence data generated are available via MG-RAST (IDs available in Fig. 2).

The sequencing coverage of plasmid DNA from groundwater samples F and G allowed additional analyses. We modified a pipeline method for postassembly detection of circularity among scaffolds (17) with the following criteria to identify the complete closed circular scaffolds referred to as “circular\_scaffolds” or simply circular plasmids: (i) scaffold length of  $>2$  kb, (ii)  $>34$ -bp homology (E value  $>1e-5$ ) at the ends of the scaffold in the correct direction, and (iii) at least two read pairs mapping on opposite ends of the contig, a maximum of 500 bp from the end. The complete pipeline with Perl scripts can be found at <https://github.com/yuwu/detect-circ-plasmid>. The “circular\_scaffolds” were subjected to annotation using components from the RAST (Rapid Annotations using Subsystems Technology) toolkit (RASTtk) with the Department of Energy Systems Biology Knowledgebase, KBase (<http://kbase.us>) (51). Annotation of circular plasmids are available through <https://narrative.kbase.us/narrative/ws.40055.obj.11>.

The resulting “all\_scaffolds” and “circular\_scaffolds” plasmid sequences were compared with (i) A Classification of Mobile genetic Elements (ACLAME) (52, 53), (ii) the antiBacterial biocide and Metal resistance genes database (BacMet) (54), (iii) the Toxin Antitoxin DataBase (TADB) (55), (iv) the Antibiotic Resistance genes DataBase (ARDB) (56), and (v) the Comprehensive Antibiotic Resistance Database (CARD) (57). The analyses were performed as follows. (i) ACLAME plasmid proteins and MGE (Mobile Genetic Elements) families were downloaded from the ACLAME website. The plasmid genes from both samples were mapped against the plasmid proteins using BLAST with an E value cutoff of  $1e-3$ . The BLAST tabulated results were parsed to obtain the taxonomic distributions of the plasmid genes by mapping the BLAST results to the MGE families, which consist of the taxonomic information. (ii) For BacMet, the Perl script BacMet-Scan.pl version 1.1, the predicted resistance gene data sets, and the experimentally confirmed resistance gene data set were downloaded from [bacmet.biomedicine.gu.se](http://bacmet.biomedicine.gu.se). The BacMet-Scan.pl was executed using default parameters (-blast -e 1 -l 30 -p 90) to generate the tabulated report against both predicted and experimentally confirmed data sets. (iii) For TADB, the database was downloaded from the TADB website version 1.1 (<http://202.120.12.135/TADB2/>) followed by BLAST with the following parameters: E value of  $1e-3$ , min\_target\_seqs 1. (iv) For ARDB, the Perl script ardbAnno.pl and ardbAnno.pm were downloaded from the ARDB website along with the resistance gene data set. The plasmid genes from both samples were mapped against the resistance gene data set using the scripts with default parameters. (v) For CARD, CARD and software RGI (Resistant Gene Identifier) databases were downloaded from the CARD website (<https://card.mcmaster.ca/home>). The script rgi.py was used to search the predicted plasmid genes against the CARD database with default parameters followed by parsing using a customized Perl script. The “circular\_scaffolds” were also categorized into incompatibility groups (58) and the relaxase/MOB types (37).

**16S rRNA gene sequencing.** Genomic DNA was extracted using the modified Miller DNA extraction method (21) followed by purification and concentration using a Genomic DNA Clean & Concentrator kit (Zymo Research, Irvine, CA). DNA quality was determined using the NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA), and concentration was determined using a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA). The V4 region of both bacterial and archaeal 16S rRNA genes was amplified using a two-step PCR approach. The primers (515F, 5'-GTGCCAGCMGCCGCGTAA-3', and 806R, 5'-GGACTACHVGGGTWTCTAAT-3') were used without added sequencing components in the first step to avoid additional bias. To increase the base diversity in sequences of sample libraries, phasing primers were used in the second-step PCR. Spacers of different lengths (0 to 7 bases) were added before the forward and reverse primers, which shifts sequencing phases among different community samples from both directions. Sequencing was performed on the Illumina MiSeq platform (21).

The resulting 16S rRNA gene sequence data were processed using custom python scripts (<https://github.com/almlab/SmileTrain>) that call USEARCH for quality filtering and overlapping paired-end reads and Biopython (59) for file format input and output. The sequences were then progressively clustered to 90% with UCLUST (60), aligned to the SILVA database with mothur and align.seqs, and processed with distribution-based clustering as previously described (61) with k\_fold 10 to remove sequencing errors. Chimeras were identified with UCHIME (62) and removed. Taxonomic identification was performed with the Ribosomal Database Project (63) using 0.50 as a confidence threshold for taxonomic classification at every level. The OTU table data were then converted to a biom format to analyze diversity and taxon summaries in Qiime.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.02899-18>.

**FIG S1**, DOCX file, 0.2 MB.

**FIG S2**, DOCX file, 0.4 MB.

**FIG S3**, DOCX file, 3.2 MB.

**TABLE S1**, DOCX file, 1 MB.

**TABLE S2**, DOCX file, 0.02 MB.

**TABLE S3**, DOCX file, 0.01 MB.

**TABLE S4**, DOCX file, 0.01 MB.

**TABLE S5**, DOCX file, 0.02 MB.

**TABLE S6**, DOCX file, 0.03 MB.

**TABLE S7**, DOCX file, 0.01 MB.

## ACKNOWLEDGMENTS

This work was part of the ENIGMA-Ecosystems and Networks Integrated with Genes and Molecular Assemblies (<http://enigma.lbl.gov>), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, and is supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research, under contract number DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

The authors do not have any conflict of interest.

We thank Nurgul Kaplan (Lawrence Berkeley National Lab [LBNL]), Jennifer Chiniquy (LBNL), and Garima Goyal (LBNL) for help with Nexterra library prep on the five plasmidome samples via the DIVA service at EmeryStationEast, LBNL, and Nathan Hillson (LBNL) for managing this resource. We thank Dylan Chivian and Paramvir Dehal (LBNL) for the invaluable assistance and guidance in analysis of the plasmidome via KBase (<https://kbase.us/>). We also thank Ferran Garcia-Pichel (Arizona State University) for his valuable comments on the manuscript.

## REFERENCES

- Aminov R. 2011. Horizontal gene exchange in environmental microbiota. *Front Microbiol* 2:158. <https://doi.org/10.3389/fmicb.2011.00158>.
- Biet F, Cenatiempo Y, Fremaux C. 2002. Identification of a replicon from pTXL1, a small cryptic plasmid from *Leuconostoc mesenteroides* subsp. *mesenteroides* Y110, and development of a food-grade vector. *Appl Environ Microbiol* 68:6451–6456. <https://doi.org/10.1128/AEM.68.12.6451-6456.2002>.
- Kobori H, Sullivan CW, Shizuya H. 1984. Bacterial plasmids in Antarctic natural microbial assemblages. *Appl Environ Microbiol* 48:515–518.
- Rozhon W, Petutschnig E, Khan M, Summers DK, Poppenberger B. 2010. Frequency and diversity of small cryptic plasmids in the genus *Rhizobium*. *BMC Microbiol* 10:56. <https://doi.org/10.1186/1471-2180-10-56>.
- Harrison PW, Lower RP, Kim NK, Young JP. 2010. Introducing the bacterial ‘chromid’: not a chromosome, not a plasmid. *Trends Microbiol* 18:141–148. <https://doi.org/10.1016/j.tim.2009.12.010>.
- Finan TM, Kunkel B, De Vos GF, Signer ER. 1986. Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. *J Bacteriol* 167:66–72. <https://doi.org/10.1128/jb.167.1.66-72.1986>.
- Smalla K, Heuer H, Gotz A, Niemeyer D, Krogerrecklenfort E, Tietze E. 2000. Exogenous isolation of antibiotic resistance plasmids from piggery manure slurries reveals a high prevalence and diversity of IncQ-like plasmids. *Appl Environ Microbiol* 66:4854–4862. <https://doi.org/10.1128/AEM.66.11.4854-4862.2000>.
- Dahlberg C, Linberg C, Torsvik VL, Hermansson M. 1997. Conjugative plasmids isolated from bacteria in marine environments show various degrees of homology to each other and are not closely related to well-characterized plasmids. *Appl Environ Microbiol* 63:4692–4697.
- Ogunseit OA, Tedford ET, Pacia D, Sirotkin KM, Saylor GS. 1987. Distribution of plasmids in groundwater bacteria. *J Ind Microbiol* 1:311–317. <https://doi.org/10.1007/BF01569309>.
- Richmond M. 1968. The plasmids of *Staphylococcus aureus* and their relation to other extrachromosomal elements in bacteria. *Adv Microb Physiol* 2:43–88. [https://doi.org/10.1016/S0065-2911\(08\)60259-3](https://doi.org/10.1016/S0065-2911(08)60259-3).
- Hada HS, Sizemore RK. 1981. Incidence of plasmids in marine *Vibrio* spp. isolated from an oil field in the northwestern Gulf of Mexico. *Appl Environ Microbiol* 41:199–202.
- Burton NF, Day MJ, Bull AT. 1982. Distribution of bacterial plasmids in clean and polluted sites in a South Wales river. *Appl Environ Microbiol* 44:1026–1029.
- Vartoukian SR, Palmer RM, Wade WG. 2010. Strategies for culture of ‘unculturable’ bacteria. *FEMS Microbiol Lett* 309:1–7. <https://doi.org/10.1111/j.1574-6968.2010.02000.x>.
- Dib JR, Wagenknecht M, Farias ME, Meinhardt F. 2015. Strategies and approaches in plasmidome studies—uncovering plasmid diversity disregarding of linear elements? *Front Microbiol* 6:463. <https://doi.org/10.3389/fmicb.2015.00463>.
- Mizrahi I. 2012. The rumen plasmidome: a genetic communication hub for the rumen microbiome. *Mob Genet Elements* 2:152–153. <https://doi.org/10.4161/mge.20793>.
- Brown Kav A, Sasson G, Jami E, Doron-Faigenboim A, Benhar I, Mizrahi I. 2012. Insights into the bovine rumen plasmidome. *Proc Natl Acad Sci U S A* 109:5452–5457. <https://doi.org/10.1073/pnas.1116410109>.
- Jorgensen TS, Xu Z, Hansen MA, Sorensen SJ, Hansen LH. 2014. Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metaplastidome. *PLoS One* 9:e87924. <https://doi.org/10.1371/journal.pone.0087924>.
- Luo WT, Xu ZF, Riber L, Hansen LH, Sorensen SJ. 2016. Diverse gene functions in a soil mobilome. *Soil Biol Biochem* 101:175–183. <https://doi.org/10.1016/j.soilbio.2016.07.018>.
- Zhang T, Zhang XX, Ye L. 2011. Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. *PLoS One* 6:e26041. <https://doi.org/10.1371/journal.pone.0026041>.
- Shi YH, Zhang H, Tian Z, Yang M, Zhang Y. 2018. Characteristics of ARG-carrying plasmidome in the cultivable microbial community from wastewater treatment system under high oxytetracycline concentration. *Appl Microbiol Biotechnol* 102:1847–1858. <https://doi.org/10.1007/s00253-018-8738-6>.
- Smith MB, Rocha AM, Smillie CS, Olesen SW, Paradis C, Wu LY, Campbell JH, Fortney JL, Mehlhorn TL, Lowe KA, Earles JE, Phillips J, Techtmann SM, Joyner DC, Elias DA, Bailey KL, Hurt RA, Preheim SP, Sanders MC, Yang J, Mueller MA, Brooks S, Watson DB, Zhang P, He ZL, Dubinsky EA, Adams PD, Arkin AP, Fields MW, Zhou JZ, Alm EJ, Hazen TC. 2015. Natural bacterial communities serve as quantitative geochemical biosensors. *mBio* 6:e00326-15. <https://doi.org/10.1128/mBio.00326-15>.
- Watson D, Kostka J, Fields M, Jardine P. 2004. The Oak Ridge field research center conceptual model. NABIR Field Research Center, Oak Ridge, TN.
- Hemme CL, Deng Y, Gentry TJ, Fields MW, Wu LY, Barua S, Barry K, Tringe SG, Watson DB, He ZL, Hazen TC, Tiedje JM, Rubin EM, Zhou JZ. 2010. Metagenomic insights into evolution of a heavy metal-contaminated

- groundwater microbial community. *ISME J* 4:660–672. <https://doi.org/10.1038/ismej.2009.154>.
24. Hemme CL, Tu Q, Shi Z, Qin Y, Gao W, Deng Y, Nostrand JD, Wu L, He Z, Chain PS, Tringe SG, Fields MW, Rubin EM, Tiedje JM, Hazen TC, Arkin AP, Zhou J. 2015. Comparative metagenomics reveals impact of contaminants on groundwater microbiomes. *Front Microbiol* 6:1205. <https://doi.org/10.3389/fmicb.2015.01205>.
  25. King AJ, Preheim SP, Bailey KL, Robeson MS, Chowdhury TR, Crable BR, Hurt RA, Mehlhorn T, Lowe KA, Phelps TJ, Palumbo AV, Brandt CC, Brown SD, Podar M, Zhang P, Lancaster WA, Poole F, Watson DB, Fields MW, Chandonia JM, Alm EJ, Zhou JZ, Adams MWW, Hazen TC, Arkin AP, Elias DA. 2017. Temporal dynamics of in-field bioreactor populations reflect the groundwater system and respond predictably to perturbation. *Environ Sci Technol* 51:2879–2889. <https://doi.org/10.1021/acs.est.6b04751>.
  26. Hemme CL, Green SJ, Rishishwar L, Prakash O, Pettenato A, Chakraborty R, Deutschbauer AM, Van Nostrand JD, Wu L, He Z, Jordan IK, Hazen TC, Arkin AP, Kostka JE, Zhou J. 2016. Lateral gene transfer in a heavy metal-contaminated-groundwater microbial community. *mBio* 7:e02234–15. <https://doi.org/10.1128/mBio.02234-15>.
  27. He Z, Zhang P, Wu L, Rocha AM, Tu Q, Shi Z, Wu B, Qin Y, Wang J, Yan Q, Curtis D, Ning D, Van Nostrand JD, Wu L, Yang Y, Elias DA, Watson DB, Adams MWW, Fields MW, Alm EJ, Hazen TC, Adams PD, Arkin AP, Zhou J. 2018. Microbial functional gene diversity predicts groundwater contamination and ecosystem functioning. *mBio* 9:e02435-17. <https://doi.org/10.1128/mBio.02435-17>.
  28. Brown Kav A, Benhar I, Mizrahi I. 2013. A method for purifying high quality and high yield plasmid DNA for metagenomic and deep sequencing approaches. *J Microbiol Methods* 95:272–279. <https://doi.org/10.1016/j.mimet.2013.09.008>.
  29. Norman A, Riber L, Luo WT, Li LL, Hansen LH, Sorensen SJ. 2014. An improved method for including upper size range plasmids in metagenomes. *PLoS One* 9:e104405. <https://doi.org/10.1371/journal.pone.0104405>.
  30. Foster TJ. 1983. Plasmid-determined resistance to antimicrobial drugs and toxic metal ions in bacteria. *Microbiol Rev* 47:361–409.
  31. Rothschild ER, Turner R, Stow SH, Bogle MA, Hyder LK, Sealand OM, Wyrick HJ. 1984. Investigation of subsurface mercury at the Oak Ridge Y-12 plant. (no. ORNL/TM-9092). Oak Ridge National Laboratory, Oak Ridge, TN.
  32. van Mastrigt O, Di Stefano E, Hartono S, Abee T, Smid EJ. 2018. Large plasmidome of dairy *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis* FM03P encodes technological functions and appears highly unstable. *BMC Genomics* 19:620. <https://doi.org/10.1186/s12864-018-5005-2>.
  33. Jaffe A, Ogura T, Hiraga S. 1985. Effects of the *ccd* function of the F plasmid on bacterial growth. *J Bacteriol* 163:841–849.
  34. Gerdes K, Rasmussen PB, Molin S. 1986. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proc Natl Acad Sci U S A* 83:3116–3120. <https://doi.org/10.1073/pnas.83.10.3116>.
  35. Spoerel N, Herrlich P, Bickle TA. 1979. A novel bacteriophage defence mechanism: the anti-restriction protein. *Nature* 278:30–34. <https://doi.org/10.1038/278030a0>.
  36. Smillie C, Garcillan-Barcia MP, Francia MV, Rocha EP, de la Cruz F. 2010. Mobility of plasmids. *Microbiol Mol Biol Rev* 74:434–452. <https://doi.org/10.1128/MMBR.00020-10>.
  37. Garcillan-Barcia MP, Alvarado A, de la Cruz F. 2011. Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS Microbiol Rev* 35:936–956. <https://doi.org/10.1111/j.1574-6976.2011.00291.x>.
  38. Hancock SJ, Phan MD, Peters KM, Forde BM, Chong TM, Yin WF, Chan KG, Paterson DL, Walsh TR, Beatson SA, Schembri MA. 2017. Identification of IncA/C plasmid replication and maintenance genes and development of a plasmid multilocus sequence typing scheme. *Antimicrob Agents Chemother* 61:e01740-16. <https://doi.org/10.1128/AAC.01740-16>.
  39. Li Y, Chen Q, Wang CH, Cai S, He J, Huang X, Li SP. 2013. Degradation of acetochlor by consortium of two bacterial strains and cloning of a novel amidase gene involved in acetochlor-degrading pathway. *Bioresour Technol* 148:628–631. <https://doi.org/10.1016/j.biortech.2013.09.038>.
  40. Cao J, Lai Q, Yuan J, Shao Z. 2015. Genomic and metabolic analysis of fluoranthene degradation pathway in *Celeribacter indicus* P73T. *Sci Rep* 5:7741. <https://doi.org/10.1038/srep07741>.
  41. San Millan A, Pena-Miller R, Toll-Riera M, Halbert ZV, McLean AR, Cooper BS, MacLean RC. 2014. Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nat Commun* 5:5208. <https://doi.org/10.1038/ncomms6208>.
  42. Ruegg TL, Kim EM, Simmons BA, Keasling JD, Singer SW, Lee TS, Thelen MP. 2014. An auto-inducible mechanism for ionic liquid resistance in microbial biofuel production. *Nat Commun* 5:3490. <https://doi.org/10.1038/ncomms4490>.
  43. Ray JK, Keller KL, Catena M, Juba TR, Zemla M, Rajeev L, Knierim B, Zane GM, Robertson JJ, Auer M, Wall JD, Mukhopadhyay A. 2014. Exploring the role of CheA3 in *Desulfovibrio vulgaris* Hildenborough motility. *Front Microbiol* 5:77. <https://doi.org/10.3389/fmicb.2014.00077>.
  44. Birnboim HC, Doly J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7:1513–1523. <https://doi.org/10.1093/nar/7.6.1513>.
  45. Anderson DG, McKay LL. 1983. Simple and rapid method for isolating large plasmid DNA from lactic streptococci. *Appl Environ Microbiol* 46:549–552.
  46. Stevenson DM, Weimer PJ. 2007. Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Appl Microbiol Biotechnol* 75:165–174. <https://doi.org/10.1007/s00253-006-0802-y>.
  47. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428. <https://doi.org/10.1093/bioinformatics/bts174>.
  48. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>.
  49. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA. 2008. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. <https://doi.org/10.1186/1471-2105-9-386>.
  50. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Ruckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33:5691–5702. <https://doi.org/10.1093/nar/gki866>.
  51. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology* 36:566. <https://doi.org/10.1038/nbt.4163>.
  52. Leplae R, Lima-Mendez G, Toussaint A. 2010. ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res* 38:D57–D61. <https://doi.org/10.1093/nar/gkp938>.
  53. Leplae R, Hebrant A, Wodak SJ, Toussaint A. 2004. ACLAME: a CLAssification of Mobile genetic Elements. *Nucleic Acids Res* 32:D45–D49. <https://doi.org/10.1093/nar/gkh084>.
  54. Pal C, Bengtsson-Palme J, Rensing C, Kristiansson E, Larsson DG. 2014. BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res* 42:D737–D743. <https://doi.org/10.1093/nar/gkt1252>.
  55. Shao YC, Harrison EM, Bi DX, Tai C, He XY, Ou HY, Rajakumar K, Deng ZX. 2011. TADB: a web-based resource for type 2 toxin-antitoxin loci in bacteria and archaea. *Nucleic Acids Res* 39:D606–D611. <https://doi.org/10.1093/nar/gkq908>.
  56. Liu B, Pop M. 2009. ARDB—Antibiotic Resistance Genes Database. *Nucleic Acids Res* 37:D443–D447. <https://doi.org/10.1093/nar/gkn656>.
  57. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang WL, Yan M, Yu T, Wright DG. 2013. The Comprehensive Antibiotic Resistance Database. *Antimicrob Agents Chemother* 57:3348–3357. <https://doi.org/10.1128/AAC.00419-13>.
  58. Carattoli A, Zankari E, Garcia-Fernandez A, Larsen MV, Lund O, Villa L, Aarestrup FM, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58:3895–3903. <https://doi.org/10.1128/AAC.02412-14>.
  59. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg

- I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.
60. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
61. Preheim SP, Perrotta AR, Martin-Platero AM, Gupta A, Alm EJ. 2013. Distribution-based clustering: using ecology to refine the operational taxonomic unit. *Appl Environ Microbiol* 79:6593–6603. <https://doi.org/10.1128/AEM.00342-13>.
62. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200. <https://doi.org/10.1093/bioinformatics/btr381>.
63. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267. <https://doi.org/10.1128/AEM.00062-07>.
64. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: aAn information aesthetic for comparative genomics. *Genome Res* 19:1639–1645. <https://doi.org/10.1101/gr.092759.109>.
65. Darzentas N. 2010. Circoletto: visualizing sequence similarity with Circos. *Bioinformatics* 26:2620–2621. <https://doi.org/10.1093/bioinformatics/btq484>.