Soil Biology & Biochemistry 101 (2016) 175–183

Contents lists available at ScienceDirect

Soil Biology & Biochemistry

journal homepage: www.elsevier.com/locate/soilbio

Diverse gene functions in a soil mobilome

Wenting Luo [a], Zhuofei Xu [a, b], Leise Riber [a], Lars H. Hansen [a, c], Søren J. Sørensen [a, *]

[a] Department of Biology, University of Copenhagen, Copenhagen, Denmark
[b] Stem Cell Institute, Catholic University Leuven, Leuven, Belgium
[c] Department of Environmental Science, Aarhus University, Aarhus, Denmark

ARTICLE INFO

Article history:
Received 23 November 2015
Received in revised form
9 July 2016
Accepted 20 July 2016
Available online 27 July 2016

Keywords:
Mobile genetic elements
Mobilome
Plasmid metagenome
Soil microbiology

ABSTRACT

Accessing bacterial mobilomes of any given environment enables the investigation of genetic traits encoded by circular genetic elements, and how their transfer drives the adaptation of microbial communities. Here we take advantage of Illumina HiSeq sequencing and report, for the first time, the soil mobilome sampled from a well-characterized field in Hygum, Denmark. Soil bacterial cells were obtained by Nycodenz extraction, total DNA was purified by removing sheared chromosomal DNA using exonuclease digestion, and the remaining circular DNA was amplified with the phi29 polymerase and finally sequenced. The soil mobilome represented a wide range of known bacterial gene functions and highlighted the enrichment of plasmids, transposable elements and phages when compared to a well-characterized soil metagenome that, on the other hand, was dominated by basic biosynthesis and metabolism functions. Approximately one eighth of the gene set was of plasmid-intrinsic traits, including replication, conjugation, mobilization and stability based on Pfam database analysis. Resistance determinants toward aminoglycosides, beta-lactams and glycopeptides as well as multi-drug functions indicated that a substantial fraction of the soil resistome is plasmid-encoded and potentially mobilizable. Additionally, we recovered more than half of all Pfam-listed plasmid replication protein families, of which the composition of both common and rare replication families was significantly different from a previously reported wastewater and rat cecum mobilome. This comprehensive analysis reveals a distinct plasmid gene pool in the soil environment and suggests the prevalence of specific plasmid groups and plasmid-encoded genetic traits in distinct ecological environments.

1. Introduction

In the last decade, the successful application of high-throughput sequencing platforms has led to the development of metagenomic characterization of distinct environments from the mammalian intestine to permafrost soil (Tringe et al., 2005; Yergeau et al., 2010; Hu et al., 2013). One crucial step of metagenome studies is to acquire enough pure genomic DNA for the construction of clone libraries or for direct sequencing, with the latter not being limited to traditional culture-dependent methods. Thus, metagenomic studies largely extend our current knowledge about microbial diversity and metabolic potential at a community level. However, complex environments with massive microbial diversity, such as soil, provide major challenges for current metagenomic approaches (Howe et al., 2014), and more focused strategies become advantageous.

The mobilome technique, focusing on plasmids and other circular mobile genetic elements (MGEs), including transposons, integrons and phages, is a derivative of the metagenome approach (Kav et al., 2012; Li et al., 2012; Jørgensen et al., 2014; Norman et al., 2014). Plasmids, the major component of the mobilome, represent extra-chromosomal DNA elements that can self-replicate and possibly transfer themselves among different prokaryotic hosts (Thomas and Nielsen, 2005). One example of plasmid mobility is the Ti plasmid that can transfer from virulent to non-virulent Agrobacteria (Hooykaas et al., 1977). The ability of plasmids to transfer antibiotic resistance genes has become the textbook example and has been documented numerous times since the spread of tetracycline resistance was reported to relate to the R factor (Datta et al., 1971). Other accessory genes, for instance, symbiotic nitrogen fixation genes located on megaplasmids of rhizobia (Hynes and McGregor, 1990), organic compound degradative genes (Friello et al., 1976), and genes coding virulent factors

* Corresponding author. Department of Biology, University of Copenhagen, Universitetsparken 15, Building 1, DK-2100, Copenhagen, Denmark.
E-mail address: sjs@bio.ku.dk (S.J. Sørensen).

http://dx.doi.org/10.1016/j.soilbio.2016.07.018
0038-0717/

of *Salmonella* (Williamson et al., 1988), can also be transferred. The transfer of such genes, mediated by mobilizable and conjugative plasmids in environments such as soil and wastewater, has been surveyed by exogenous isolation (Top et al., 1994; Smalla et al., 2000). Due to the implication in prokaryotic evolution (Sørensen et al., 2005), it is necessary to study the phylogenetic and functional diversity of the mobilome.

There are several studies on mobilome/plasmid metagenomes accessed by different methodologies. Based on a transposon aided capture method (TRACA), plasmids of Gram-positive bacterial origins as well as genes involved in plasmid mobilization were captured from bacterial cells of the human gut (Jones and Marchesi, 2007). A wastewater mobilome, built in our group, enriched genes involved in plasmid replication, mobilization and conjugation (Li et al., 2012). Finally, a rumen plasmidome and a rat cecum mobilome, respectively, were recently successfully established (Kav et al., 2012; Jørgensen et al., 2014). Although soil has been proposed as major reservoir of a vast mobilome (D'Costa et al., 2007), the genetic content of the soil mobilome is not easily assessed. This is possibly because of the complexity in acquiring intact plasmid DNA and in removing sheared and linear genomic DNA, and finally because of the presence of humic acids and other inhibitors.

In this study, circular genetic elements from a soil bacterial community were captured and sequenced following a culture-independent protocol previously applied in a wastewater study (Li et al., 2012). We aim to describe the general mobilome-encoded (mainly plasmid-encoded) gene functions and the capture of plasmid accessory genes, such as antibiotic resistance genes. We compare our dataset with public databases, including the ACLAME (A CLAssification of Mobile genetic Elements) database (Leplae et al., 2004, 2010) and the Pfam protein families database (Finn et al., 2014), to distinguish mobilome-determined traits from chromosomal functions. Also, we characterize the distribution of plasmid replication proteins and relate this to plasmid groups identified in a wastewater and a rat cecum mobilome, revealing distinct plasmid communities distributed in different environments.

2. Material and methods

2.1. Soil sampling

Soil samples were collected in biological triplicates (A, B and C) in December 2011 (0–20 cm depth) from Hygum, Vejle, Denmark (9.434647W, 55.776159N) and stored in sterile containers at 5 °C. The sampling site is adjacent to a copper-gradient polluted field previously being well described (Strandberg et al., 2006; Berg et al., 2010). The unaffected vegetation indicates that this area has not been contaminated by copper (Strandberg et al., 2006). The bacterial cell extraction, mobilome DNA purification and exonuclease digestion were operated for the triplicates. However, due to the overwhelming number of 16S rRNA gene copies in replicate C, indicating an inadequate removal of chromosomal DNA by exonuclease digestion (data not shown), replicate A and B were pooled and referred to as the Hygum soil sample for the following study. Thus, we will not evaluate any biological variance in this single soil mobilome sample because of the pooling.

2.2. Acquisition of mobilome DNA

We recovered bacterial cells by high speed centrifugation with Nycodenz solution (Nyco, Norway) (Burmølle et al., 2003). Cell pellets were subsequently harvested for DNA extraction by using Plasmid mini AX kit (A&A Biotechnology, Poland). To remove chromosomal and sheared DNA, the extracted total DNA (1–10 ng) was digested by using Plasmid-Safe™ ATP-Dependent DNAse (EPICENTRE® Biotechnologies, USA). The qPCR of 16s rRNA genes (Eub338/518) (Fierer et al., 2005) was performed to monitor the removal of chromosomal DNA during digestion. The digestion products were washed twice by Sigma water and condensed to 20 μL by using Amino Ultra 0.5 mL Centrifugal Filter Units (Merck Millipore, Germany). The multiple displacement amplification (MDA) of the digested DNA (Pan et al., 2008; Li et al., 2012) combined with the REPLI-g® Mini Kit (QIAGEN, USA) then acquired all intact circular genetic elements.

2.3. Bioinformatics

Illumina raw sequences were trimmed and filtered by using Biopieces (http://www.biopieces.org). Ribosomal RNA fragments were identified by using Meta-RNA (version H3) with default settings (Huang et al., 2009). Reads were assembled by using IDBA-UD assembler (version 1.0.9) (Peng et al., 2012), and the resulting contigs (>500 bp in length) were used for gene calling by Prodigal (Prokaryotic Dynamic Programming Gene finding Algorithm) (Hyatt et al., 2010). After gene calling, the non-redundant coding sequences (CDS) were filtered by using CD-HIT (Li and Godzik, 2006; Fu et al., 2012). The workflow above was repeated each time an increasing certain amount of clean reads was randomly extracted; the generated number of non-redundant CDS was plotted as the accumulative curve to test the saturation of sequencing data.

The local database ACLAME (version 0.4) is a collection of 2326 prokaryotic mobile genetic elements (MGEs) comprising all known transposons, phage genomes and plasmids (Leplae et al., 2004, 2010). Both the non-redundant CDS and the ACLAME database were submitted to the MG-RAST (Metagenomics Rapid Annotation using Subsystem Technology; version 3.3) server for functional analysis (Meyer et al., 2008). The public dataset of the Waseca county farm soil metagenome (accession number 4441091.3) (Tringe et al., 2005) was chosen to identify functional differences among these 3 datasets based on the SEED subsystems classification (Overbeek et al., 2005). The Waseca county farm soil was of low level organic matter content and contained primarily prokaryotes (Tringe et al., 2005), thus representing a typical prokaryotic genetic pool.

The homologs of protein sequences of six mobilome-determined traits, “conjugation”, “replication”, “mobilization”, “stability and maintenance”, “MGEs” and “phages”, were searched and screened for based on the Pfam database (Protein family, version 27.0) (Punta et al., 2012). We applied the same data-processing workflow to a wastewater (Li et al., 2012) and a rat cecum mobilome (Jørgensen et al., 2014) previously built in our group (these data of the two mobilomes were not published). To compare the composition of the “replication” category among these 3 mobilomes, we also simulated 16S rRNA gene diversity studies (Blaxter et al., 2005).

Antibiotic resistance protein families of the soil mobilome were screened by employing the HMMER package (version 3.1b1) (Finn et al., 2011) in the Resfams HMM database (e-value 10^{-5}) (core, version 1) (Gibson et al., 2015). A BLASTP search of the detected antibiotic resistance genes against the NCBI non-redundant database (NCBInr) (Acland et al., 2013) was performed to define their taxonomy. The minimum query coverage and sequence identity was both 70%, and with a threshold of e-value 10^{-5}.

The relative abundance of each gene was defined as the number of reads that aligned to this specific coding sequence divided by the total number of aligned reads (referred to as mobilome reads). We aligned high-quality reads to all non-redundant CDS by using Burrows-Wheeler aligner (version 0.7.5a) and SAMtools (version

0.1.18) (Li and Durbin, 2009). We filtered mapped reads with a mapping quality of 30 (-q30 option of SAMtools). If a pair of filtered reads mapped to the same CDS, only one read with the better mapping quality was included. The total number of reads aligned to all non-redundant CDS was recorded as 100%.

2.4. Statistical analysis

Proportional differences of the SEED subsystems among the soil mobilome, ACLAME and the Waseca soil metagenome, as well as proportional differences of plasmid replication protein families among the soil, a wastewater (Li et al., 2012) and a rat cecum mobilome (Jørgensen et al., 2014) were tested using a two-sided Fisher's exact test in the STAMP (Statistical Analysis of Metagenomic Profiles) software package (version 2.0.0) (Parks and Beiko, 2010). A standard asymptotic approach with continuity correction was used to calculate the confidence interval (95%), and a Bonferroni correction was used to modify the *P*-values. *P*-values <0.05 were considered to state the statistically significant difference between two datasets (Parks and Beiko, 2010).

2.5. Accession numbers

The non-redundant CDS of the soil mobilome and the ACLAME database has been deposited to the MG-RAST server under the accession numbers 4574032.3 and 4500447.3, respectively.

3. Results

3.1. Sequencing the soil mobilome

We captured approximately 0.8×10^6 cells per gram soil after Nycodenz extraction (Table S1). Following 40 h of exonuclease digestion, the copy number of 16S rRNA genes, representing sheared and linear genomic DNA, was at a level close to the detection limit (only 0.0014% of genomic DNA left, Fig. S1), which ensured the enrichment of exclusively circular DNA during MDA. The Illumina sequencing resulted in approximately 90 million clean paired-end reads with a mean length 95 bp. The average percentage of 16S rRNA genes in bacterial chromosomal DNA has previously been estimated to 0.12% (Jørgensen et al., 2014). Using this number and based on the abundance of 16S rRNA gene reads (0.0041%) found in this study, we estimated to have less than 4% of prokaryotic chromosomal DNA in the soil mobilome data (Fig. S1).

IDBA-UD assembler yielded 3,162 contigs (>500 bp) with an N_{50} of 10,267 bp and a total length of 11.46 Mbp. After gene calling and clustering, the redundant CDS were excluded and 14,717 non-redundant CDS of a mean length of 634 bp were applied for the downstream analysis. In the accumulative curve, both the length of contigs and the number of non-redundant CDS increased significantly until 20 million clean reads. The tendency of reaching a plateau at 60 million clean reads illustrated the sufficient amount of sequences for surveying the soil mobilome (Fig. 1).

To have an outline of the functions encoded by the soil mobilome, we annotated all non-redundant CDS against the NCBInr and Pfam database, and classified functional groups according to the Clusters of Orthologous Groups (COGs) (Tatusov et al., 1997) (Table S6) and SEED subsystems database, respectively. The proportion of CDS assigned to protein functions varied depending on the databases, representing 20.4% of SEED subsystems, 26.2% of COGs, 38.2% of Pfam and 47.0% of NCBInr.

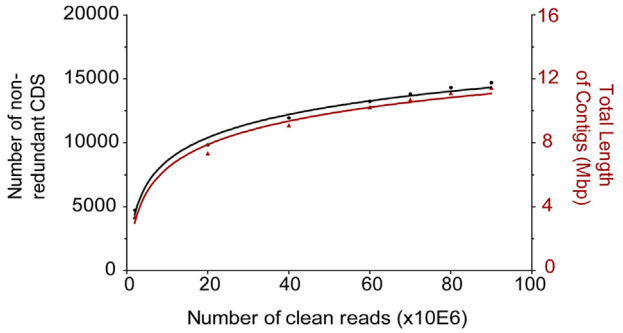

Fig. 1. Non-redundant coding sequences (CDS) accumulation curve. The number of non-redundant CDS (dots in black) and the length of contigs (triangles in red) corresponding to a certain amount of clean reads extracted randomly from all clean reads were plotted, respectively. A logarithmic trendline was used to fit a curve to each dataset (red filled curve: $R^2 = 0.9820$; black filled curve: $R^2 = 0.9879$.). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Functional comparison among the soil mobilome, the ACLAME and the soil metagenome

Based on outputs of the MG-RAST server, the annotated CDS of the soil mobilome, ACLAME and the soil metagenome were assigned into 28 SEED categories, respectively. Two sets of comparisons were made. First, we compared ACLAME with the soil metagenomic dataset (Fig. 2A). This comparison was built as the "background" functional distribution of a known mobile gene pool and a soil-specific chromosomal gene pool. Even though the ACLAME database might not reflect the actual microbial mobilome in every given environment, it is a functional indicator of MGEs. The well-characterized soil metagenome provided an overview of microbial functions corresponding to its large genome size and soil-specific phylogenetic composition (Tringe et al., 2005). In general, the pairwise comparison distinguished the mobile gene pool from the metagenomic gene pool.

In contrast to ACLAME, the soil mobilome was over-represented by the category "Phages, Prophages, Transposable elements and Plasmids" at the expense of "Membrane Transport" (Fig. 2B). Additionally, the soil mobilome was dominated by "Cell Wall and Capsule", "Cell Division and Cell Cycle" as well as 4 other categories (Fig. 2B). Compared with the soil mobilome, the soil metagenome had an increased distribution of basic biosynthesis and metabolism-related genes, including "Carbohydrates", "Fatty Acids, Lipids and Isoprenoids" and "Nucleosides and Nucleotides". The soil metagenome was clearly less represented by "Membrane Transport" and "Phages, Prophages, Transposable elements and Plasmids" functions (Fig. 2C). Approximately 17% of the non-redundant CDS fell into the poorly characterized category of "Clustering-based Subsystems", which was the most prevalent subsystem in the soil mobilome (Fig. 2B). The proportion of "Phages, Prophages, Transposable elements and Plasmids" ranked the second abundant category of the soil mobilome (about 13.5%) (Fig. 2B and C). The categories "Membrane Transport" and "DNA Metabolism" occupied 7.26 and 7.36%, respectively, of the soil mobilome (Fig. 2C).

3.3. Relative abundance of soil mobilome-encoded functions

Approximately 40% of all clean reads aligned to the non-redundant CDS, which indicated that a substantial fraction of sequences were represented by the predicted genes. The MG-RAST server does not only provide a platform to analyze gene functional distribution by assigning them to the SEED subsystems, but

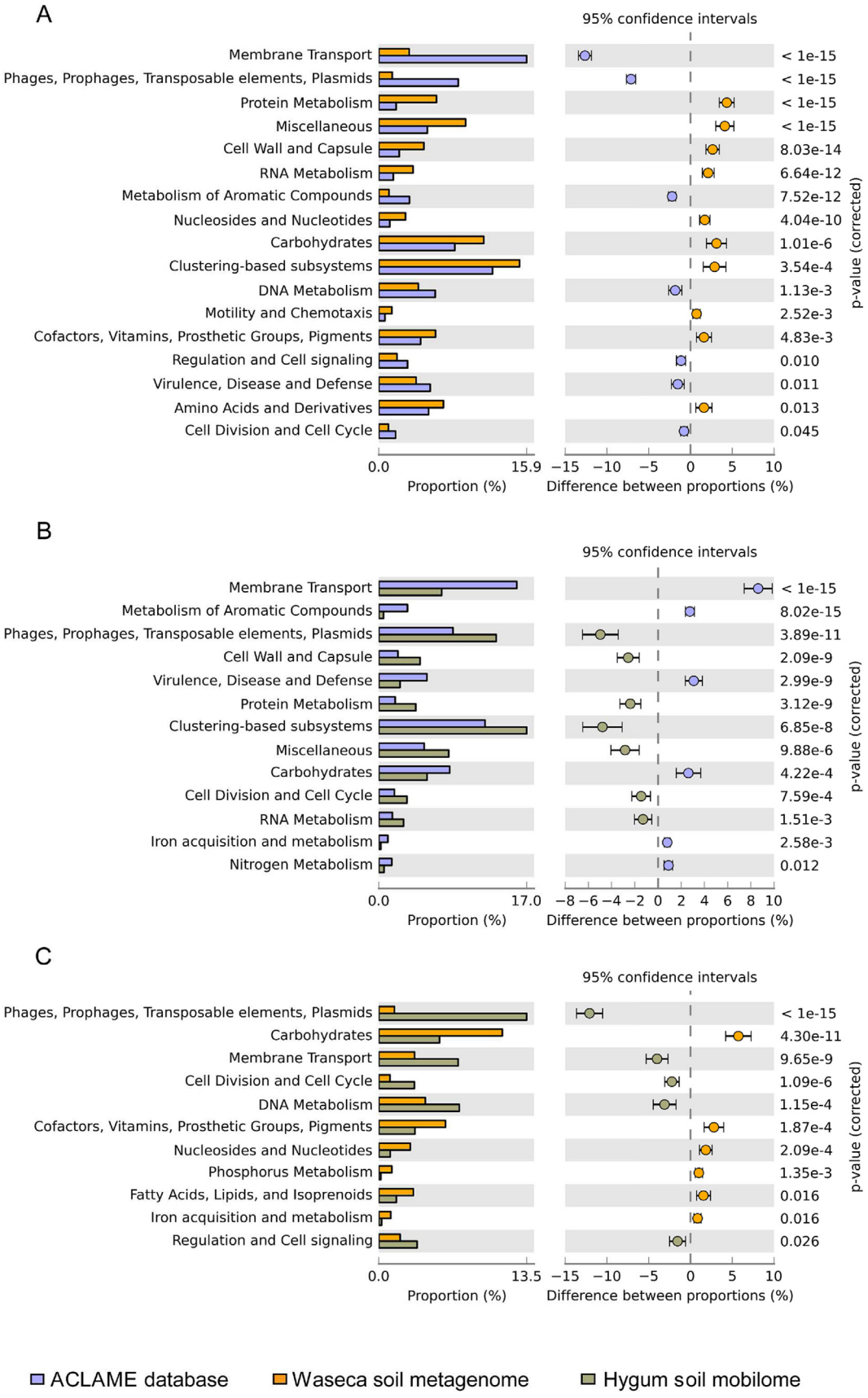

Fig. 2. Comparative functional distribution of the soil mobilome, the Waseca County Farm soil metagenome and the ACLAME database based on SEED subsystems. Pairwise proportional differences were calculated by using STAMP software and only statistically different ($P < 0.05$) SEED subsystems are presented here. **(A)** ACLAME database vs. Waseca soil Metagenome; **(B)** ACLAME database vs. Soil mobilome; **(C)** Waseca soil Metagenome vs. Soil mobilome.

also allow the calculation of relative abundance of each category of interest. The three most abundant categories were “Phages, Prophages, Transposable elements, Plasmids”, “Membrane Transport” and “Cofactors, Vitamins and Pigments” (Table 1). At more defined functional levels “phage integration and excision”, “plasmid conjugative transfer” and “folate and pterines” were the most enriched subsystems in the soil mobilome (Table S2).

We manually extracted six groups of plasmid-associated proteins based on the Pfam output, and the overall abundance of these six groups represented 13.6% of mobilome reads. In addition to plasmid “selfish” genetic components such as conjugation and mobilization, we also predicted the existence of characters relating to plasmid stability and maintenance, for instance, control of copy number, active partitioning and post-segregational killing. Within the six groups, phage-related proteins, plasmid replication proteins and MGEs accounted for 4.07, 3.55 and 2.10% of the total aligned reads, respectively. The abundance of “conjugation”, “mobilization” and “stability and maintenance” was more evenly distributed, with proportions of 1.29, 1.22 and 1.36% of the total aligned reads, respectively.

3.4. Diversity of plasmid replication protein families

We detected 12 of 20 Pfam replication protein families in the soil mobilome, of which Rep_1 (*PF01446*), Rep_trans (*PF02486*) and Replicase (*PF03090*) were the three most abundant families (Table 2). The previously published wastewater (Li et al., 2012) and rat cecum mobilome (Jørgensen et al., 2014) displayed a significantly distinct distribution of plasmid replicon protein families compared with the present soil mobilome ($P < 0.05$, Table S9). However, Rep_1 replicons were the most abundant family occupying more than 40% in all three mobilomes (Table 2). Rep_trans and Rep_3 (*PF01051*) ranked as the second and third most abundant families in the wastewater mobilome, whereas Rep_2 (*PF01719*) and Rep_3 were the second and third most abundant in the rat cecum mobilome (Table 2). The Replicase family (*PF03090*) occupied 16.5% in the soil mobilome, whereas it was reduced to only 1.5% and less than 0.1% in the wastewater and rat cecum mobilome, respectively. The Rep_trans family was represented by 23.0% in the soil mobilome, and represented by 34.2% in the rat cecum mobilome, but only by 3.1% in the wastewater mobilome (Table 2). Primpol, RepA_C and PriCT_2 were present exclusively in the soil mobilome, whereas Rop, Rep_2, RepC and PriCT_1 were present in only the wastewater and the rat cecum mobilome.

Diversity indices provide information about rarity and commonness of replication protein families and enable us to evaluate differences and similarities among the three mobilomes. We observed the most diverse and evenly distributed plasmid replication protein families in the soil mobilome with a Shannon's evenness of 0.68 and a Simpson's index of 4.06, whereas these indices were 0.49 and 2.92 in the rat cecum mobilome, which indicates a reduced diversity and evenness for this latter mobilome (Table 3).

3.5. Detection of antibiotic resistance determinants

In addition to plasmid-intrinsic functions, we screened for antibiotic resistance (AR) determinants in the soil mobilome. We detected 36 AR protein families among all 166 Resfams families, including resistance towards 6 antibiotic classes: aminoglycosides, beta-lactams, glycopeptides, macrolides, quinolones and tetracyclines. The distribution of AR protein families is shown in Fig. 3 with a relative abundance of 0.46% of all aligned reads. The dominant Resfams AR families were VanR and VanS (transcriptional regulators of other vancomycin resistance genes) (Gibson et al., 2015), aminoglycoside phosphotransferase (type III), Class C beta-

Table 1
Relative abundance of the SEED functional systems (occupied 10.31% of the mobilome reads) in the soil mobilome. The abundance was the number of reads that aligned to genes in each subsystem divided by the total number of aligned reads. Subsystems that are less abundant than 0.01% are not shown.

The SEED subsystems	Relative abundance
Phages, Prophages, Transposable ele. and Plasmids	2.67%
Membrane Transport	1.61%
Cofactors, Vitamins, Prosthetic Groups, Pigments	0.93%
Carbohydrates	0.77%
Clustering-based subsystems	0.64%
Stress Response	0.54%
Miscellaneous	0.52%
DNA Metabolism	0.49%
Virulence, Disease and Defense	0.44%
Respiration	0.26%
Cell Division and Cell Cycle	0.24%
Sulfur Metabolism	0.22%
Amino Acids and Derivatives	0.20%
RNA Metabolism	0.18%
Cell Wall and Capsule	0.15%
Regulation and Cell signaling	0.12%
Protein Metabolism	0.12%
Nitrogen Metabolism	0.09%
Fatty Acids, Lipids, and Isoprenoids	0.05%
Metabolism of Aromatic Compounds	0.03%
Nucleosides and Nucleotides	0.02%
Motility and Chemotaxis	0.02%
Potassium Metabolism	0.01%
Beyond the SEED functional systems	89.69%

Table 2
Comparative distribution of detected plasmid replication protein families in the soil, wastewater and rat cecum mobilomes based on the Pfam analysis.[a]

Pfam protein families	Proportion (%)[b]		
	Hygum soil	Wastewater	Rat cecum
Rep_1	39.1	42.2	46.5
Rep_trans	23.9	34.2	0.0
Replicase	16.5	1.5	0.0
Rep_3	3.1	11.1	23.0
Prim-Pol	6.2	0.0	0.0
RPA	5.0	0.1	0.0
RepA_C	3.7	0.0	0.0
RHH_1	0.8	2.2	0.1
PriCT_2	0.8	0.0	0.0
IncFII_repA	0.4	0.0	~0.0
TrfA	0.4	0.8	0.0
RepL	~0.0	5.2	3.4
PriCT_1	0.0	0.2	~0.0
Rep_2	0.0	1.8	26.9
RepC	0.0	~0.0	~0.0
Rop	0.0	0.6	0.1

[a] Pairwise proportional differences were calculated by using STAMP and significant differences ($P < 0.05$) of replication families were detected. The proportions of all replication families between any two mobilomes were significantly different.
[b] Proportions of replication protein families that are less than 0.1% are shown as ~0.0%.

Table 3
Diversity indices of Pfam plasmid replication families in the soil, wastewater and rat cecum mobilomes.

Measure	Formula[a]	Hygum soil	Wastewater	Rat cecum
Shannon's index (H')	$H' = -\sum p_i \ln p_i$	1.68	1.44	1.18
Shannon evenness (E)	$E = H'/\ln S$	0.68	0.58	0.49
Richness (S)	–	12	12	11
Simpson's index (D)	$D = 1/\sum p_i^2$	4.06	3.21	2.92

[a] p_i is the proportion of each Pfam plasmid replication family.

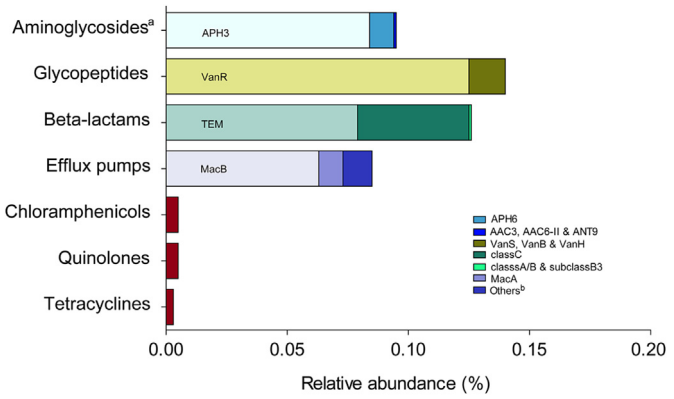

Fig. 3. Relative abundance of antibiotic resistance (AR) protein families within the soil mobilome based on the Resfam HMM database. The protein families occupied 0.46% of mobilome reads and are classified based on the antibiotic to which they confer resistance (except efflux pumps), and the dominant Resfams AR protein families are depicted in the bars. [a]AAC: aminoglycoside acetyltransferase; ANT: aminoglycoside nucleotidyl transferase; APH: aminoglycoside phosphotransferase. [b]Others includes MsbA, TolC, AdeA-AdeI, RND, ABC and MFS Antibiotic Efflux Pump families.

lactamase and TEM beta-lactamase, MacB and MacA (known to be parts of the ABC-type efflux for exporting macrolides) (Rouquette-Loughlin et al., 2005) (Fig. 3). The 36 AR protein families were represented by 203 AR genes (ARGs). These 203 ARGs are prevalent in a wide range of soil-dwelling bacterial species such as *Pseudomonas, Agrobacterium* and *Rhizobium* as well as uncultured bacteria from the candidate division TM7 phylum. One ARG was of high amino acid sequence identity (100%) to the putative ATP-transporter of *Klebsiella pneumonia* and another ARG displayed a similar amino acid sequence identity to the aminoglycoside 3'-phosphotransferase of *Corynebacterium diphtheria*. However, we only perceived four ARGs, encoding resistance to beta-lactams, aminoglycosides, chloramphenicols and tetracyclines by using BLAST tools against the ARDB database (Antibiotic Resistance genes DataBase) (Table S7) (Liu and Pop, 2009).

4. Discussion

This study represents the first mobilome of a soil environment. Since bead-beating or silica-column based extraction methods will disrupt the circular nature of plasmids in the soil, the preparation of mobilome DNA from environmental samples requires several difficult procedures, for example, detachment of bacterial cells from soil particles and exclusion of chromosomal DNA. Although the Nycodenz cell extraction method is biased, it is an efficient and gentle procedure to separate intact bacterial cells from other biological particles (Holmsgaard et al., 2011). By exonuclease digestion followed by MDA we were able to eliminate a significant amount of sheared and linear chromosomal DNA, and amplify a sufficient amount of pure circular DNA (presumably plasmids) for sequencing. These two steps are, however, biased towards the enrichment of smaller plasmids because larger sized plasmids are fragile and easily sheared, and also, the rolling-circle amplification during MDA naturally favors smaller circular elements over larger ones, which eventually leads to a reduction of larger plasmids (Jørgensen et al., 2014).

4.1. The novelty of gene functions encoded by the soil mobilome

The soil mobilome was functionally surprisingly diverse. All listed SEED database categories were captured including gene functions involving cellular processes and signaling, information storage, and metabolism. The novelty of sequences could lead to relatively fewer annotations of functions in the soil mobilome when we assigned them to the SEED categories. In contrast, the assignment of the ACLAME to the SEED database would overestimate the abundance of some functions because the recruited MGEs sequences were shared between these two public databases. Thus, only 20.4% of the gene set could be assigned to the SEED subsystems, while 67.0% of all 26,689 proteins in the ACLAME database could be assigned. Previous research shows that the proportion of assigned sequences varies significantly. For instance, 10–20% of pyro-sequencing reads for swine gut metagenomes, 41% reads for a marine metagenome, and 25% of Illumina reads for a soil metagenome were reported to be assigned to the SEED categories (Lamendella et al., 2011; Mitra et al., 2011; Wang et al., 2013). The annotation proportion may depend on the alignment sensitivity for highly fragmented genes as well as on the integrity of each database (Raes et al., 2007a, 2007b), but also on the complexity of the samples. Therefore, approximately 50% of the reads in the soil mobilome consisted of uncharacterized features, demonstrating that soil mobilomes, like soil metagenomes, are complex, but valuable sources of novel gene functions that are not yet listed in known databases (Daniel, 2004).

4.2. Soil is a genetic reservoir of phages and plasmids

The pairwise comparison of the soil mobilome to the soil metagenome demonstrated the abundance of phages, plasmids and transposable elements in the soil mobilome. The proportion of these functions in the mobilome was approximately 9 times higher than similar functions in the soil metagenome, especially regarding three subsystems “phage integration and excision”, “plasmid-encoded transferred-DNA” and “plasmid replication”. The comparison between the soil mobilome and ACLAME also highlighted the overwhelming population of plasmids and phages in the soil mobilome. The large plasmids of *Agrobacteria* and *Rhizobia* contributed to the largest number of proteins in ACLAME (Leplae et al., 2004). Thus, the portion of the indicative features of “Phages, Prophages, Transposable elements and Plasmids”, such as “plasmid replication”, was relatively small in ACLAME. In contrast, for the soil mobilome, the “Phages, Prophages, Transposable elements and Plasmids” occupied a relatively higher portion of all proteins. Consistent with the SEED analysis, the protein domain-based analysis, tuned on plasmid-intrinsic and phage-related functions, also predicted the MGE-rich nature of the soil mobilome. These parameters might assist soil bacteria in dealing with functions relating to plasmid replication, stability, maintenance and conjugation (see below) as well as phage replication.

4.3. Detection of gene functions associated with larger sized plasmids

We found that the abundance of “Membrane Transport” in the soil mobilome was significantly higher than that of the soil metagenome. The major subsystem “type IV secretion system” indicated the presence of plasmids larger than 10 Kbp, since such systems most likely are associated with conjugative plasmids, e.g. IncQ and IncF plasmids, generating filaments to facilitate bacterial cell-to-cell contact during conjugation (Clarke et al., 2008). As reported, the largest natural IncQ plasmid is ~15 Kbp (Rawlings and Tietze, 2001), whereas the smallest IncF plasmid is around 50 Kbp (Mshana et al., 2011). Other observed functional roles representing the type IV secretion system were pathogenic virulence Vir-like proteins encoded by the plasmid pVir (~35 Kbp). In addition, type II and VI secretion systems were detected, constituting only a small fraction of this category. With the help of the function “Membrane

Transport", soil bacteria might communicate with other soil microorganisms such as fungi (pathogenic, saprotrophic and mycorrhizal fungi) in the nutrient-cycling.

In the soil mobilome, the other two abundant categories, "Carbohydrates" and "Co-factors, Vitamins and Pigments", were not as enriched as in the soil metagenome. They were mainly associated with the subsystems related to synthesis or assembly of "mono-, di- and oligo-saccharides" and "folate and pterines", respectively. These are common bacterial functions in soil, especially in agricultural soil, being influenced by high amounts of dissolved nutrients from crop residues, root secretions and microbial degradation (Jones et al., 2009). The slight amount of chromosomal DNA contamination could contribute to these macromolecular-related functions, as genes of folate synthesis are typically encoded on bacterial chromosomes (van Kranenburg et al., 1997; de Crecy-Lagard et al., 2007). However, the synthesis of oligosaccharides has also been identified to have a plasmid origin (van Kranenburg et al., 1997). All three categories indicate that this soil mobilome protocol is able to capture traits associated with relatively larger sized plasmids.

4.4. The occurrence of mobilome-encoded antibiotic resistance traits

Due to the general perception that soil is a major source of resistance genes and MGEs (D'Costa et al., 2007; Monier et al., 2011; Forsberg et al., 2012), we were interested in describing the antibiotic resistance profile of the soil mobilome. These analyses revealed mainly resistance determinants towards aminoglycosides, beta-lactams, glycopeptides and multidrugs in the soil mobilome. The detection of plasmid borne resistance to antibiotics without any known direct selective pressure in the Hygum soil is consistent with findings in remote Alaskan soil of diverse beta-lactamases (Allen et al., 2009). Seven soil-originated ARGs that had high nucleotide similarity to ARGs from clinical pathogens were found to be associated with mobile elements (Forsberg et al., 2012), and a mobilizable multi-resistant plasmid, pKLH80, has been identified in ancient permafrost (Petrova et al., 2014). The soil mobilome indeed harbors a notable amount of antibiotic resistance traits likely to cause resistance towards a broad selection of antimicrobial agents, and with the potential to be mobilized into indigenous soil-dwelling bacteria including opportunistic pathogenic species. Thus, plasmids are possibly involved in the maintenance of the soil ecosystem, yet the actual function of these genes is to be determined.

4.5. Diversity and composition of soil-specific plasmid replication protein families

In the soil mobilome, the most abundant replication protein families, Rep_1, Rep_trans and Replicase, occupied almost 4/5 of all detected replication protein families. The Rep_1 family is represented by the Rep75 protein, which catalyzes DNA synthesis and exhibits the nicking-closing activity in an archaeal plasmid, pGT5 (Marsin and Forterre, 1999). This family is also included by small bacterial plasmids, such as the *Bacillus subtilis* plasmid, pUH1 (Hara et al., 1991), and the mobilizable vancomycin-resistance plasmid pDT1 found in *Enterococcus faecium* (Todokoro et al., 2006), replicated by a rolling-circle mechanism. The Rep_trans family is represented by RepN, which likely functions to recognize the *ori* region of the staphylococcal plasmid, pCW7 (Balson and Shaw, 1990). This family is also included by other small resistance plasmids, such as pNS1 (Aoki et al., 1993) and pUB112 (probably an IncFII plasmid) (Ehret and Matzura, 1988), which are staphylococcal tetracycline-resistant and chloramphenicol-resistant plasmids, respectively. The representative protein from the Replicase family is RepA in *Pseudomonas* plasmid pPS10, which is a common RepA-type initiator in Gram-negative bacteria (Giraldo and Diaz-Orejas, 2001). This family has been found in the broad-host-range (BHR) plasmid pEF2 (Zhang et al., 1994) and some cryptic plasmids like pJD1 (Gauchatfeiss et al., 1985) and pAL5000 (Stolt and Stoker, 1996). The Rep_1 family is mostly found in *Firmicutes,* the Rep_trans family is from *Firmicutes* and *Proteobacteria,* whereas the typical host bacteria of the Replicase family belong to *Proteobacteria* and *Actinobacteria*. All together, the detected replication protein families covered a broad variety of bacterial phyla (Table S3). The dominant plasmid host phyla, *Firmicutes*, *Proteobacteria* and *Actinobacteria*, are consistent with a previous 16S rRNA gene pyrosequencing study from the same site, in which *Actinobacteria* and *Proteobacteria* represented around 40% and 25% of the bacterial sequences, respectively (Berg et al., 2012). This broad diversity of replication protein families clearly indicates that the soil mobilome comprises plasmids from a broad spectrum of different incompatibility groups and bacterial phyla.

In general, the wastewater and rat cecum mobilomes were expected to be compositionally more similar compared with the soil mobilome, since the rats originated from a hospital sewage system and the community of wastewater mainly originated from domestic wastewater enriched with human gut microbes. The diversity of the normalized types of replication protein families showed a tendency of being increased in the Hygun soil mobilome as compared to that of the rat cecum mobilome. This is consistent with the fact that soil, as the most complex ecological system, harbors the most diverse collections of plasmids, although the current dataset is too sparse to show any significant differences. Nonetheless, the different distribution of replication protein families among the three mobilomes suggests that different environmental communities select for various plasmid groups and plasmid-encoded genetic traits.

5. Conclusion

In this study, the mobilome approach applied for soil revealed a novel and diverse pool of circular genetic elements unique to this given environment. Also, we believe that the mobilome approach reveals several elements playing a key role in gene acquisition between the microbial and the viral metagenome. By means of horizontal gene transfer, presumably advanced mobilome functions distributed among the inner MGE community will, when encountering a selective pressure, spread to the outside bacterial host community. Thus, the soil mobilome technique is an essential supplement to the soil metagenome to interpret the genetic equilibrium between bacterial core genomes and MGEs at a community level. Besides, the combination of the mobilome and the 16S rRNA gene based taxonomical composition of the microbiome may reveal host specific functional shifts and gene transfer mediated events. In conclusion, the mobilome technique highly improves our understanding of the complexity of the pool of mobile genetic elements in natural environments.

Author contributions

Conceived and designed the experiments: WL, LR, LHH and SJS. Performed the experiments: WL. Analyzed the data: WL and ZX. Contributed reagents/materials/analysis tools: WL, ZX, LR, LHH and SJS. Wrote the paper: WL, ZX, LR, LHH and SJS.

Acknowledgments

This work was supported by grants from the Center for

Environmental and Agricultural Microbiology (CREAM), funded by The Villum Foundation and The Danish Council for Independent Research (DFF). Soil samples were kindly provided by Kristian Koefoed Brandt, Department of Plant and Environmental Sciences, University of Copenhagen. Sequencing data were produced by The Danish National High-Throughput DNA-Sequencing Centre (Copenhagen, Denmark). We thank Lili Li, Karin Vestberg and Anette Løth for excellent technical assistance during laboratory work, Martin Hansen for developing the Biopieces package, Tue S. Jørgensen for providing the rat cecum mobilome data, Analia Sannazzaro and Stefan Morberg for providing the bacterial cell number data and Annelise Kjøller and Urvish Trivedi for reading and correcting the manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.soilbio.2016.07.018.

References

Acland, A., et al., 2013. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 41, D8–D20. http://dx.doi.org/10.1093/Nar/Gks1189.

Allen, H.K., et al., 2009. Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. ISME J. 3, 243–251. http://dx.doi.org/10.1038/ismej.2008.86.

Aoki, T., et al., 1993. Detection of tetracycline resistance protein encoded by *Bacillus subtilis* plasmid pNS1. Biol. Pharm. Bull. 16, 81–83.

Balson, D.F., Shaw, W.V., 1990. Nucleotide sequence of the *rep* gene of Staphylococcal plasmid pCW7. Plasmid 24, 74–80. http://dx.doi.org/10.1016/0147-619x(90)90027-A.

Berg, J., et al., 2012. Selection for Cu-tolerant bacterial communities with altered composition, but unaltered richness, via long-term Cu exposure. Appl. Environ. Microbiol. 78, 7438–7446. http://dx.doi.org/10.1128/AEM.01071-12.

Berg, J., et al., 2010. Cu exposure under field conditions co selects for antibiotic resistance as determined by a novel cultivation-independent bacterial community tolerance assay. Environ. Sci. Technol. 44, 8724–8728. http://dx.doi.org/10.1021/es101798r.

Blaxter, M., et al., 2005. Defining operational taxonomic units using DNA barcode data. Philos. Trans. R. Soc. B Biol. Sci. 360, 1935–1943. http://dx.doi.org/10.1098/rstb.2005.1725.

Burmølle, M., et al., 2003. Presence of N-acyl homoserine lactones in soil detected by a whole-cell biosensor and flow cytometry. Microb. Ecol. 45, 226–236. http://dx.doi.org/10.1007/s00248-002-2028-6.

Clarke, M., et al., 2008. F-pili dynamics by live-cell imaging. Proc. Natl. Acad. Sci. U. S. A. 105, 17978–17981. http://dx.doi.org/10.1073/pnas.0806786105.

D'Costa, V.M., et al., 2007. Expanding the soil antibiotic resistome: exploring environmental diversity. Curr. Opin. Microbiol. 10, 481–489. http://dx.doi.org/10.1016/j.mib.2007.08.009.

Daniel, R., 2004. The soil metagenome - a rich resource for the discovery of novel natural products. Curr. Opin. Biotechnol. 15, 199–204. http://dx.doi.org/10.1016/j.copbio.2004.04.005.

Datta, N., Hedges, R.W., Shaw, E.J., Sykes, R.B., Richmond, M.H., 1971. Properties of an R factor from *Pseudomonas aeruginosa*. J. Bacteriol. 108, 1244–1249.

de Crecy-Lagard, V., et al., 2007. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. BMC Genom. 8, 245. http://dx.doi.org/10.1186/1471-2164-8-245.

Ehret, M., Matzura, H., 1988. Replication control of the *Staphylococcus aureus* chloramphenicol resistance plasmids pC223 and pUB112 in *Bacillus subtilis*. Nucleic Acids Res. 16, 2045–2062. http://dx.doi.org/10.1093/nar/16.5.2045.

Fierer, N., et al., 2005. Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. Appl. Environ. Microbiol. 71, 4117–4120. http://dx.doi.org/10.1128/Aem.71.7.4117-4120.2005.

Finn, R.D., et al., 2014. Pfam: the protein families database. Nucleic Acids Res. 42, D222–D230. http://dx.doi.org/10.1093/Nar/Gkt1223.

Finn, R.D., et al., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37. http://dx.doi.org/10.1093/Nar/Gkr367.

Forsberg, K.J., et al., 2012. The shared antibiotic resistome of soil bacteria and human pathogens. Science 337, 1107–1111. http://dx.doi.org/10.1126/science.1220761.

Friello, D.A., et al., 1976. XYL, a nonconjugative xylene-degradative plasmid in *Pseudomonas Pxy*. J. Bacteriol. 127, 1217–1224.

Fu, L.M., et al., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152. http://dx.doi.org/10.1093/bioinformatics/bts565.

Gauchatfeiss, D., et al., 1985. Cloning of genes involved in myoinositol transport in a *Pseudomonas* sp. J. Bacteriol. 162, 324–327.

Gibson, M.K., et al., 2015. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J. 9, 207–216. http://dx.doi.org/10.1038/ismej.2014.106.

Giraldo, R., Diaz-Orejas, R., 2001. Similarities between the DNA replication initiators of Gram-negative bacteria plasmids (RepA) and eukaryotes (Orc4p)/archaea (Cdc6p). Proc. Natl. Acad. Sci. U. S. A. 98, 4938–4943. http://dx.doi.org/10.1073/pnas.081079298.

Hara, T., et al., 1991. Molecular-structure of the replication origin of a *Bacillus subtilis* (Natto) plasmid, pUH1. Appl. Environ. Microbiol. 57, 1838–1841.

Holmsgaard, P.N., et al., 2011. Bias in bacterial diversity as a result of Nycodenz extraction from bulk soil. Soil Biol. Biochem. 43, 2152–2159. http://dx.doi.org/10.1016/j.soilbio.2011.06.019.

Hooykaas, P.J.J., et al., 1977. Transfer of the *Agrobacterium tumefaciens* Ti plasmid to avirulent agrobacteria and to Rhizobium *ex planta*. J. General Microbiol. 98, 477–484. http://dx.doi.org/10.1099/00221287-98-2-477.

Howe, A.C., et al., 2014. Tackling soil diversity with the assembly of large, complex metagenomes. Proc. Natl. Acad. Sci. U. S. A. 111, 4904–4909. http://dx.doi.org/10.1073/pnas.1402564111.

Hu, Y.F., et al., 2013. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. Nat. Commun. 4 http://dx.doi.org/10.1038/Ncomms3151.

Huang, Y., et al., 2009. Identification of ribosomal RNA genes in metagenomic fragments. Bioinforma. 25, 1338–1340. http://dx.doi.org/10.1093/bioinformatics/btp161.

Hyatt, D., et al., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinforma. 11, 119. http://dx.doi.org/10.1186/1471-2105-11-119.

Hynes, M.F., McGregor, N.F., 1990. Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen fixing nodules by *Rhizobium leguminosarum*. Mol. Microbiol. 4, 567–574. http://dx.doi.org/10.1111/j.1365-2958.1990.tb00625.x.

Jones, B.V., Marchesi, J.R., 2007. Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. Nat. Methods 4, 55–61. http://dx.doi.org/10.1038/Nmeth964.

Jones, D.L., et al., 2009. Carbon flow in the rhizosphere: carbon trading at the soil-root interface. Plant Soil 321, 5–33. http://dx.doi.org/10.1007/s11104-009-9925-0.

Jørgensen, T.S., et al., 2014. Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metamobilome. Plos One 9, e87924. http://dx.doi.org/10.1371/journal.pone.0087924.

Kav, A.B., et al., 2012. Insights into the bovine rumen plasmidome. Proc. Natl. Acad. Sci. U. S. A. 109, 5452–5457. http://dx.doi.org/10.1073/pnas.1116410109.

Lamendella, R., et al., 2011. Comparative fecal metagenomics unveils unique functional capacity of the swine gut. BMC Microbiol. 11, 103. http://dx.doi.org/10.1186/1471-2180-11-103.

Leplae, R., et al., 2004. ACLAME: a CLAssification of mobile genetic elements. Nucleic Acids Res. 32, D45–D49. http://dx.doi.org/10.1093/Nar/Gkh084.

Leplae, R., et al., 2010. ACLAME: a classification of mobile genetic elements, update 2010. Nucleic Acids Res. 38, D57–D61. http://dx.doi.org/10.1093/Nar/Gkp938.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.

Li, L.L., et al., 2012. Metamobilomics - expanding our knowledge on the pool of plasmid encoded traits in natural environments using high-throughput sequencing. Clin. Microbiol. Infect. 18, 8–11. http://dx.doi.org/10.1111/j.1469-0691.2012.03862.x.

Li, W.Z., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659. http://dx.doi.org/10.1093/bioinformatics/btl158.

Liu, B., Pop, M., 2009. ARDB-Antibiotic resistance genes database. Nucleic Acids Res. 37, D443–D447. http://dx.doi.org/10.1093/Nar/Gkn656.

Marsin, S., Forterre, P., 1999. The active site of the rolling circle replication protein Rep75 is involved in site-specific nuclease, ligase and nucleotidyl transferase activities. Mol. Microbiol. 33, 537–545. http://dx.doi.org/10.1046/j.1365-2958.1999.01498.x.

Meyer, F., et al., 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinforma. 9, 386. http://dx.doi.org/10.1186/1471-2105-9-386.

Mitra, S., et al., 2011. Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. BMC Bioinforma. 12, S21. http://dx.doi.org/10.1186/1471-2105-12-S1-S21.

Monier, J.M., et al., 2011. Metagenomic exploration of antibiotic resistance in soil. Curr. Opin. Microbiol. 14, 229–235. http://dx.doi.org/10.1016/j.mib.2011.04.010.

Mshana, S.E., et al., 2011. Multiple ST clonal complexes, with a predominance of ST131, of *Escherichia coli* harbouring blaCTX-M-15 in a tertiary hospital in Tanzania. Clin. Microbiol. Infect. 17, 1279–1282.

Norman, A., et al., 2014. An improved method for including upper size range plasmids in metamobilomes. PLoS One 9, e104405. http://dx.doi.org/10.1371/journal.pone.0104405.

Overbeek, R., et al., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 33, 5691–5702. http://dx.doi.org/10.1093/Nar/Gki866.

Pan, X.H., et al., 2008. A procedure for highly specific, sensitive, and unbiased whole-genome amplification. Proc. Natl. Acad. Sci. U. S. A. 105, 15499–15504.

http://dx.doi.org/10.1073/pnas.0808028105.
Parks, D.H., Beiko, R.G., 2010. Identifying biologically relevant differences between metagenomic communities. Bioinformatics 26, 715–721. http://dx.doi.org/10.1093/bioinformatics/btq041.
Peng, Y., et al., 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28, 1420–1428. http://dx.doi.org/10.1093/bioinformatics/bts174.
Petrova, M., et al., 2014. Genetic structure and biological properties of the first ancient multiresistance plasmid pKLH80 isolated from a permafrost bacterium. Microbiology 160, 2253–2263. http://dx.doi.org/10.1099/Mic.0.079335-0.
Punta, M., et al., 2012. The Pfam protein families database. Nucleic Acids Res. 40, D290–D301. http://dx.doi.org/10.1093/Nar/Gkr1065.
Raes, J., et al., 2007a. Get the most out of your metagenome: computational analysis of environmental sequence data. Curr. Opin. Microbiol. 10, 490–498. http://dx.doi.org/10.1016/j.mib.2007.09.001.
Raes, J., et al., 2007b. Protein function space: viewing the limits or limited by our view? Curr. Opin. Struct. Biol. 17, 362–369. http://dx.doi.org/10.1016/j.sbi.2007.05.010.
Rawlings, D.E., Tietze, E., 2001. Comparative biology of IncQ and IncQ-like plasmids. Microbiol. Mol. Biol. Rev. 65, 481–496.
Rouquette-Loughlin, C.E., et al., 2005. Characterization of the MacA-MacB efflux system in *Neisseria gonorrhoeae*. J. Antimicrob. Chemother. 56, 856–860. http://dx.doi.org/10.1093/Jac/Dki333.
Smalla, K., et al., 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. http://dx.doi.org/10.1128/AEM.66.11.4854-4862.2000.
Sørensen, S.J., et al., 2005. Studying plasmid horizontal transfer in situ: a critical review. Nat. Rev. Microbiol. 3, 700–710. http://dx.doi.org/10.1038/Nrmicro1232.
Stolt, P., Stoker, N.G., 1996. Functional definition of regions necessary for replication and incompatibility in the *Mycobacterium fortuitum* plasmid pAL5000. Microbiology 142, 2795–2802.
Strandberg, B., et al., 2006. Effect of a copper gradient on plant community structure. Environ. Toxicol. Chem. 25, 743–753.
Tatusov, R.L., et al., 1997. A genomic perspective on protein families. Science 278, 631–637. http://dx.doi.org/10.1126/science.278.5338.631.
Thomas, C.M., Nielsen, K.M., 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat. Rev. Microbiol. 3, 711–721. http://dx.doi.org/10.1038/Nrmicro1234.
Todokoro, D., et al., 2006. Genetic analysis of bacteriocin 43 of vancomycin-resistant *Enterococcus faecium*. Appl. Environ. Microbiol. 72, 6955–6964. http://dx.doi.org/10.1128/Aem.00934-06.
Top, E., et al., 1994. Exogenous isolation of mobilizing plasmids from polluted soils and sludges. Appl. Environ. Microbiol. 60, 831–839.
Tringe, S.G., et al., 2005. Comparative metagenomics of microbial communities. Science 308, 554–557. http://dx.doi.org/10.1126/science.1107851.
van Kranenburg, R., et al., 1997. Molecular characterization of the plasmid-encoded *eps* gene cluster essential for exopolysaccharide biosynthesis in *Lactococcus lactis*. Mol. Microbiol. 24, 387–397.
Wang, Z., et al., 2013. Metagenomic profiling of antibiotic resistance genes and mobile genetic elements in a tannery wastewater treatment plant. PLoS One 8, e76079. http://dx.doi.org/10.1371/journal.pone.0076079.
Williamson, C.M., et al., 1988. A common virulence region on plasmids from 11 serotypes of *Salmonella*. J. General Microbiol. 134, 975–982. http://dx.doi.org/10.1099/00221287-134-4-975.
Yergeau, E., et al., 2010. The functional potential of high Arctic permafrost revealed by metagenomic sequencing, qPCR and microarray analyses. ISME J. 4, 1206–1214. http://dx.doi.org/10.1038/ismej.2010.41.
Zhang, Y.M., et al., 1994. Molecular analysis and characterization of a broad-host-range plasmid, pEP2. J. Bacteriol. 176, 5718–5728.