environmental microbiology

Environmental Microbiology (2019) **00**(00), 00–00 doi:10.1111/1462-2920.14813

Unravelling plasmidome distribution and interaction with its hosting microbiome

Aya Brown Kav,[1,2] Roye Rozov,[3] David Bogumil,[1] Søren Johannes Sørensen,[4] Lars Hestbjerg Hansen,[5] Itai Benhar,[2] Eran Halperin,[2,3,6] Ron Shamir[3] and Itzhak Mizrahi[1*]

[1]*Faculty of Natural Sciences, Department of Life Sciences, The National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, P.O.B. 653 Beer-Sheva, Building 41, Room 228, Beer-Sheva, Israel.*

[2]*School of Molecular Cell Biology and Biotechnology, Tel Aviv University, Tel Aviv, Israel.*

[3]*Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel.*

[4]*Department of Biology, University of Copenhagen, Copenhagen, Denmark.*

[5]*Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg, Denmark.*

[6]*Departments of Computer Science, Computational Medicine, Human Genetics, Anesthesiology and Perioperative Medicine, University of California, Los Angeles, California.*

Summary

Horizontal gene transfer via plasmids plays a pivotal role in microbial evolution. The forces that shape plasmidomes functionality and distribution in natural environments are insufficiently understood. Here, we present a comparative study of plasmidomes across adjacent microbial environments present in different individual rumen microbiomes. Our findings show that the rumen plasmidome displays enormous unknown functional potential currently unannotated in available databases. Nevertheless, this unknown functionality is conserved and shared with published rat gut plasmidome data. Moreover, the rumen plasmidome is highly diverse compared with the microbiome that hosts these plasmids, across both similar and different rumen habitats. Our analysis demonstrates that its structure is shaped more by stochasticity than selection. Nevertheless, the plasmidome is an active partner in its intricate relationship with the host microbiome with both interacting with and responding to their environment.

Received 22 May, 2019; revised 24 September, 2019; accepted 28 September, 2019. *For correspondence. E-mail imizrahi@bgu.ac.il; Tel. (+972) 8 647 9836; Fax: (+972) 8 647 9839.

Introduction

Horizontal gene transfer – the transfer of genetic material from one organism to another – is known to play a seminal role in microbial evolution. With microbial ecology research increasingly focused on the microbiome, it is of immense importance to understand horizontal gene-transfer processes within these complex microbial communities. Besides the uptake of naked DNA fragments, these gene-transfer processes are mediated by two main mechanisms – plasmids and viruses. These mobile genetic elements together create a mobile gene pool, which is available to microbiome members and enables the expansion of the microbial flexible genome repertoire. The plasmidome – the entire plasmid collection within a microbial community – is thought to be a cardinal member of this communal gene pool and to be a major driver of the horizontal gene-transfer process. Hence in order to understand the complete ecological and evolutionary impact of the plasmidome, one should understand what drives its composition and genetic repertoire.

We previously studied the plasmidome of the rumen microbiome by developing a molecular method to purify and sequence the plasmidome from environmental samples (Brown Kav *et al*., 2012, 2013a,b). We then overcame plasmid-assembly challenges by providing plasmidome-specific assembler software that allows for the proper assembly of plasmids from deep-sequencing reads (Rozov *et al*., 2017). The plasmidome was found to be a diverse selection of extrachromosomal DNA sequences, a portion of which had known and annotated functions, but an even larger portion of which remains unknown (Brown Kav *et al*., 2012; Jørgensen *et al*., 2014a,b). Of the annotated plasmidome sequences, we reported that the rumen plasmidome not only carries a wide array of accessory functions but also enriched in functions that seem beneficial to microbial prosperity within their niche, suggesting that the environment plays a role in plasmid selection. Similar results were also reported in activated sludge and soil

plasmidomes (Sentchilo *et al*., 2013; Luo *et al*., 2016). We also reported that the plasmids within the rumen environment transfer genes across large phylogenetic barriers, including gene-transfer events across members of two separate phyla.

The studies performed thus far on plasmidomes in different ecological niches provide important insights into the structure, complexity and content of the plasmidome, intertwined with its host microbiome. However, many central fundamental questions that are imperative to our understanding of the role of the plasmidome within the microbiome remain unanswered: What characterizes the relationship between the plasmidome and microbiome? What forces shape its composition and its genetic repertoire? How plastic is its genetic content and what drives it? Understanding these key questions is central to our understanding of the forces that drive the composition of the communal gene pool and its contribution to microbial evolution in natural environments. To answer these questions, we devised an experimental set-up in which we could apply different dietary regimes to induce changing environmental conditions and to induce fluctuations to the microbiome composition and structure. This enabled us to explore the plasmidome and its function within the host microbiome. Our study provides fundamental answers regarding the forces that shape the genetic capabilities and gene mobility of the plasmidome, its composition and its interaction with its hosting microbiome in different habitats and changing environments.

Results and discussion

Sampling and sequencing

We explored the stability and composition of the plasmidome across multiple ecological niches. We sampled 22 individual rumens, purified their plasmids and sequenced them using the Illumina HiSeq platform with the paired-end protocol. Following read trimming and quality control, performed by Trimmomatic, 6.81E6 $\pm$ 3.99E6 paired-end reads were generated for each sample. Reads were then subjected to our new plasmid-specific assembler (described in Rozov *et al*., 2017). From the assembled contigs, only those over 1 kb in length displaying a circular form were chosen for further analysis. The number of plasmids in each sample varied considerably, with an average of 397.3 plasmids per sample. The average length of the plasmids passing our stringent filters was ~1560 bps, and the maximum and average lengths for each of the rumen plasmidomes, as well as the number of plasmids assembled in each plasmidome and their size distribution are described in Table S1 and Fig. S1. In addition, we analysed metagenomic DNA extracted from the same rumen samples using 16S rRNA gene amplicon sequencing and the QIIME pipeline.

Plasmids and their annotation

We previously reported that over 60% of the open reading frames encoded in the rumen plasmidome are unannotated in the five queried databases (Brown Kav *et al*., 2012). In this work, roughly 16 000 ORFs were predicted in the plasmidome samples. Only 25.3% were annotated to a known protein function; the rest were either unannotated or assigned to hypothetical proteins. Over 90% of the annotated proteins were assigned to plasmid functions, including mobilization, replication and stabilization. The rest of the plasmid annotations included some minor representation of amino acid and carbohydrate metabolism proteins together with plasmid-persistence modules such as antibiotic resistance and restriction endonucleases. Notably, we found minor representation of phage proteins, which are characterized by a circular, cytoplasmic form within their microbiome hosts, appearing only rarely in our analyses (Fig. 1A).

Strikingly, 26.4% of the ORFs were found to be similar (*E*-value $<10^{-5}$) to protein sequences originating from the rat gut plasmidome (Jørgensen *et al*., 2014a,b), shedding some light on the universally conserved and functional gut plasmidome. To further elucidate this similarity, we explored the distribution of rat gut plasmidome protein hits within the rumen plasmidome. These seemed to follow a long tail distribution, where a few proteins were found to be very abundant in the rumen plasmidome, whereas most were quite rare (Fig. 1B). Functional analysis of the highly abundant proteins (over 100 hits within the rumen plasmidome) revealed enrichment in hypothetical proteins (70%), which was not the case for the rest of the rat gut proteins with rumen plasmidome homologues (only 35% hypotheticals) (Fig. 1B). This high abundance of a small set of proteins (most of which lack functional annotation) suggests that this group has a highly important, as-yet undiscovered role in the maintenance and sustainability of the plasmidome in gut systems and perhaps even universally. The existence of such mechanisms, propagating small mobile genetic elements lacking a transfer machinery, was also proposed by Xue and colleagues (2015) in the marine environment.

Most of the rumen plasmids were found to be small (up to 2 kb), a phenomenon also observed in the rat gut (Jørgensen *et al*., 2014a,b). It is important to note that phi29 DNA amplification step (see 'Materials and Methods' section) can introduce further bias towards small circular elements (Norman *et al*., 2014; Jørgensen *et al*., 2014a,b). Nonetheless, they are likely hitchhikers in this microbial environment, as their size does not permit them to carry any transfer machinery (67% of the plasmids in the current study). As such, the outcome of ORF prediction yielded one to two predicted ORFs per

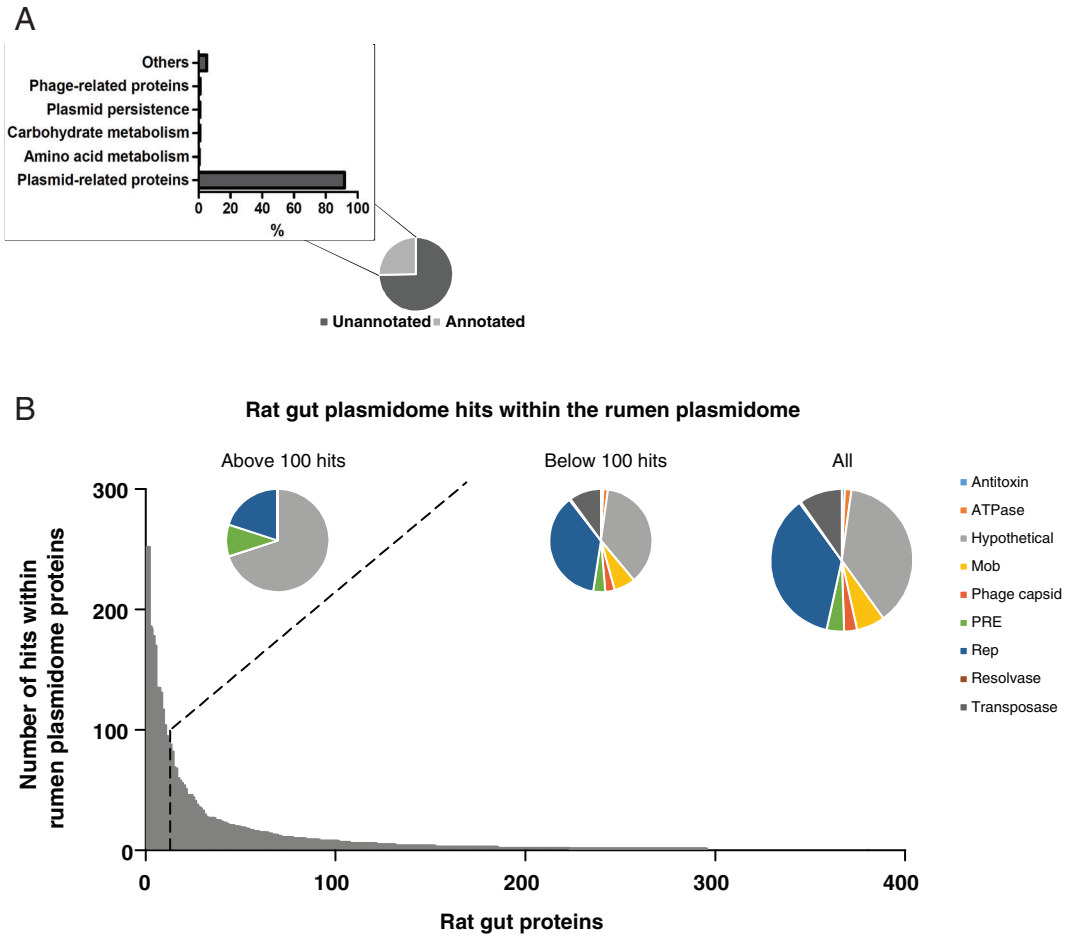

Fig. 1. The greater part of the plasmidome is left unannotated, yet displays conserved functionality.
A. Plasmidome functional annotation. Plasmidome's predicted ORFs were compared against the NCBI protein non-redundant database using a maximum *E*-value of $\leq 10^{-5}$. Most plasmidome ORFs (74.7%) were left unannotated (pie chart). The top left panel describes the prevalence of different functional groups among the annotated ORFs.
B. Analysis of the rumen plasmidome ORFs and the rat gut plasmidome. Each of the rat gut plasmidome open reading frames (ORFs) was compared to the rumen plasmidome ORFs and the number of hits is illustrated in the graph. Depicted in the far right pie chart is the functional annotation of each ORF. The data set was further broken down into rat gut ORFs which had less than 100 hits in the rumen plasmidome and those with 100 hits or more (see dashed line) and their functional annotations were, respectively, divided (middle and left pie charts).

plasmid. Most of these ORFs could not be assigned to any functional gene, leaving them presently cryptic, while only a small fraction carried sequences annotated as replication or mobilization proteins (Fig. S2). Moreover, structural similarity analysis yielded the same results, therefore suggesting that these proteins lack any known homologues and potentially indicating new mechanisms for plasmid maintenance inside their hosts (Table S2).

Horizontal gene transfer is mediated by favoured microbiome members

As plasmids facilitate horizontal gene transfer, allowing the movement of genes across different organisms, we could estimate the potential trail of plasmids across their hosts by determining the origin of their gene content from the potential microbial lineages. We therefore determined the extent of this phenomenon by examining phylogenetic associations of the ORFs carried by plasmids to study horizontal gene transfer within the rumen plasmidome. This analysis was carried out by the Phylogenie pipeline (Frickey and Lupas, 2004) for automated phylogeny generation and analysis (Price *et al*., 2009). We could confidently (80% bootstrap support) determine the phylum-level association of over 2400 plasmid contigs, out of 8741 overall plasmids (27.8%). This proportion was due to unclear determination of the phylogeny association by the Phylogenie pipeline (see 'Materials and Methods' section). This analysis revealed that 54 out of 498 (10.8%) of the contigs with at least two assigned ORFs were found to be mosaic at the phylum level, with ORFs originating from two different phyla.

One interesting observation was the disproportionate participation of different phyla in cross-phylum horizontal gene-transfer events. One possible explanation for this is over and under-representation of certain phyla in the database. However, as the prevalence of each phylum in the rumen system can also influence their participation in cross-phylum events, we studied the relationship between phylum abundance and their participation in horizontal gene-transfer events over large phylogenetic distances. The ratio between phylum abundance in the rumen and that in cross-phylum horizontal gene-transfer events allowed us to explore this relationship by expecting that the abundance of a specific phylum would also explain its participation in such events (Fig. 2). However, this ratio varied between the different phyla and was not correlated to their abundance. While the prevalent phylum Firmicutes participated in the expected manner in cross-phylum horizontal gene transfer relative to its abundance, other major rumen phyla, such as Bacteroides, barely participated at all, despite being very abundant in the rumen microbiome. Two outstanding phyla which were found to be key participants in inter-phylum gene transfer were Actinobacteria and Proteobacteria, despite their low abundance in the rumen, indicating their increased participation in long-distance horizontal gene transfer via plasmids. These two phyla have been reported to be strongly affected by horizontal gene transfer (Jeong *et al*., 2016), suggesting that regardless of their environment, they engage in gene transfer more often than others. Another study also reported Actinobacteria and Proteobacteria as hubs of plasmid-borne gene transfer (Tamminen *et al*., 2012).

Plasmid distribution and dispersal across niches

Our experimental set-up enabled us to elucidate how different individual habitats (rumen ecosystems) affect plasmidome attributes. To investigate plasmid distribution across habitats, we developed a tool for clustering plasmid sequences. The need for this tool arose because plasmids lack a conserved entity such as rRNA genes, thereby requiring a search for complete sequence similarity. This tool uses the same rationale as for OTUs, which have become standard for classifying and grouping together microbes.

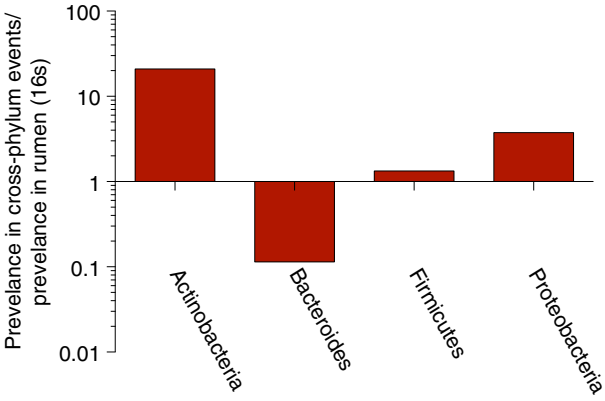

Fig. 2. Ratio between relative participation in cross-phylum horizontal gene-transfer events and the relative abundance of each phylum in the rumen, as calculated by 16S rRNA analysis.

Plasmidome diversity is higher than that of the microbiome

Once the plasmids were successfully clustered, we could explore plasmid diversity across multiple rumens. To assess plasmidome and species OTU diversity, we performed a sample rarefaction analysis. As shown in Fig. 3A, an asymptotic curve was obtained for the OTUs indicating that most of the species diversity was covered. However, this was not the case for the plasmids, as the slope declined only slightly across the curve, indicating that the plasmidome has higher diversity than the resident microbiome in this niche. Another study reported the same outcome for bacteriophages in this niche, displaying high diversity among animals of the same herd (Ross *et al*., 2013).

Next, we explored the relationship between the plasmidome and the microbiome across different individual rumen habitats. We calculated the pairwise similarity for each pair of rumen plasmidomes and microbiomes taking into account the presence/absence of plasmids and OTUs across the samples using the Dice similarity index. The average similarity between any two pairs was significantly lower in the plasmidome than in the microbiome (Fig. 3B). Similar findings were also reported by Brito and colleagues (2016) who reported that mobile gene content varied considerably while the microbiome composition across individual humans from nearby villages changed considerably less. The low similarity exhibited by the rumen plasmidomes, as compared with their hosting microbiomes, suggests that the forces shaping microbial composition within the rumen ecosystem do not necessarily apply to their plasmidomes, at least not to the same extent, resulting in much higher diversity of the plasmidome across habitats.

The recurrence of individual plasmids across samples was also examined. The occurrence of each OTU or plasmid across the samples was evaluated and binned into different consecutive prevalence bins. Figure 3C exhibits their percentages shared by each sample occurrence category. Interestingly, microbiome members tended to persist across more habitats than their plasmid counterparts (Fig. 3C). This is best portrayed by the presence of OTUs that were present among almost all samples (90% to 100% bin), whereas none of the plasmids

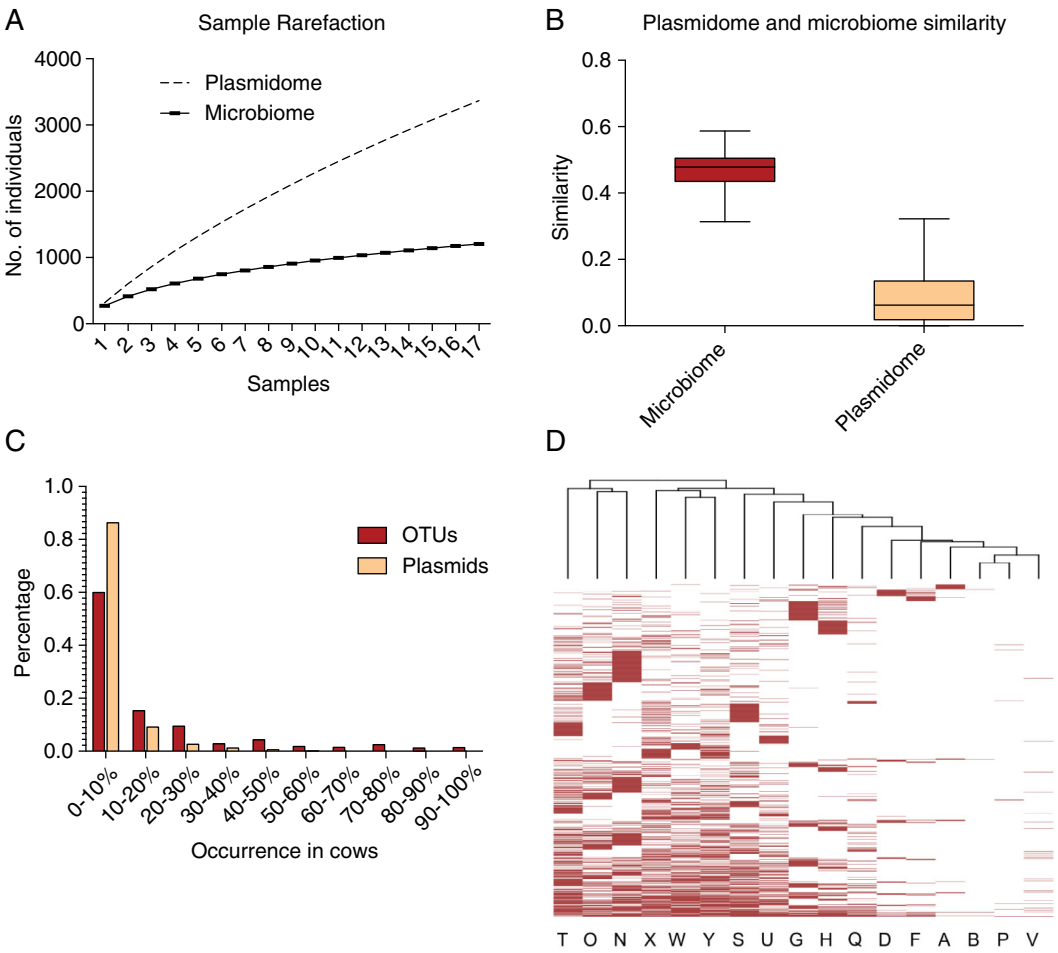

Fig. 3. The plasmidome is highly diverse as compared to its hosting microbiome.
A. Sample-based rarefaction analysis of the microbiome and plasmidome. Species-accumulation curves showing the increase in OTU (full line) and plasmid (dashed line) numbers as a function of the number of individuals sampled.
B. Average pairwise similarity of the plasmidome and microbiome between any two rumen samples. Using the Dice similarity index, we calculated the pairwise similarity of both the plasmidome and microbiome between any two samples. Whiskers illustrate minimum and maximum values. The two distributions were significantly different (*t* test, $p < 0.0001$).
C. Occurrence of plasmids and OTUs across different rumen habitats. Different plasmids and OTUs were summed into bins according to their frequency of occurrence across rumen samples and binned accordingly. *X*-axis depicts the percentage of cow rumens sharing a specific plasmid or out; *Y*-axis represents the percentage of plasmids or OTUs in each bin.
D. Presence of individual plasmids across rumen habitats. Heat map analysis of plasmids across individual rumen environments (letters) depicts plasmids shared by at least two samples. Burgundy lines denote individual plasmids and samples were clustered according to Euclidean distance matrix.

were present across all samples. We further explored the distribution of individual plasmids across similar habitats and found that individual plasmids are not shared by many of the rumen hosts (Fig. 3D). This intriguingly low congruence of the plasmidome across samples indicates lesser selective forces governing the presence of specific plasmids across habitats, perhaps highlighting that devoid of host-beneficial traits, the cost of carrying a plasmid is similar, as is its effect on microbial fitness. Taken together, it is highly possible that the persistence of plasmids across microbial communities is governed by neutral forces such as drift (given a constant fitness effect, in a small population, random genetic drift will overcome the effects of selection; Duret, 2008).

Plasmids favour interactions with plasmids from similar hosts

As plasmidome prevalence and diversity could directly affect DNA mobility via plasmids, we examined the genetic interaction between plasmidomes. We added another dimension to our study by introducing rumen samples of animals eating distinct diets. This enabled us to change the environment of the plasmidome and study horizontal gene-transfer dynamics as a function of ecology. We aligned the sequences of each plasmid pair and studied their similarity. Two plasmids were considered as interacting if they shared an aligned region of over 200 bps and at least 70% DNA sequence similarity. First, we explored the properties of the interactions by

assembling a network in which the nodes represent individual plasmids and the edges connecting them are the interactions (Fig. 4). The topology of the network generated by the plasmid interactions seemed to show an aggregated but disconnected pattern. This suggests a limitation in horizontal gene-transfer events, whereas a highly connected and central network would be expected with unlimited DNA-transfer events. To elucidate the limitations, the nodes in the network were coloured according to their determined phylogenetic phylum association (Fig. 4A), the sample from which they originated (Fig. 4C) and diet from which the samples were taken (Fig. 4B). As can be seen, the phylum association and diet seem to have a stronger effect on plasmid-interaction limitations than the sample from which they come. Hence, plasmids tend to exchange DNA with plasmids that are carried by more phylogenetically related hosts that share the same ecology. To ask whether plasmid size had an effect on these patterns, we divided the plasmids into two size bins (above 3000 bps and below 3000 bps) and assembled networks from each. Indeed, the results of these analyses show that the size of plasmid is connected to the genes that they carry (Fig. S2) and their distribution across diet and phylogeny (Fig. S3). Another interesting finding was that the plasmids defined as having cross-phylum events were mostly found in network clusters containing the phyla between which they were presumably mobilized (Fig. 4A). This further supports their central role in gene mobilization between hosting phyla.

Next, we asked whether there is a limitation to the proportion of plasmid sequences that can be shared between plasmids. To reduce biases arising from the different lengths of the two interacting plasmids, we normalized each interaction by calculating the per cent coverage of the larger plasmid. Analysis of the per cent coverage of all plasmid–plasmid interactions revealed a distribution with a peak at 20% plasmid coverage. From 20% to 80% there was a decrease in the prevalence of plasmid–plasmid interactions, indicating that the smaller sizes are favoured for recombination between plasmids (Fig. S4).

The plasmidome, albeit less conserved, is still tightly connected to its microbiome

The observed higher diversity led us to determine whether the microbiome has a role in shaping the plasmidome. We explored the relationship between plasmidome and microbiome similarity across multiple rumen habitats. We compared the pairwise similarity between microbiomes and plasmidomes from the same samples in two different environments shaped by different diets. This analysis revealed that despite the higher diversity within the plasmidome, its composition was tightly linked to that of its microbiome as the pairwise similarity of the plasmidome between any two samples was significantly correlated to the pairwise similarity of the microbiome between any two samples (Fig. 5A). To further elucidate this relationship, we explored whether any one of the major bacterial phyla influences plasmidome diversity by multiple linear regression analysis. We first determined how similar members of specific phyla are connected to each other over different cows by measuring the pairwise sequence similarity between each pair of OTUs within the phyla, and then calculated the average pairwise similarity of microbial composition for each phylum separately. The specific similarity of each phylum across the different hosts was incorporated into a multiple linear regression model together with the plasmidome similarity across the same cows. Model's overall R^2 was 0.58 (p-value = 10^{-20}). Bacteroides was found to be the most influential phylum within the model (p-value = $7*10^{-6}$ within the fitted model. The other major rumen phyla, Firmicutes, Actinobacteria and Proteobacteria, did not affect plasmidome diversity significantly. This finding that plasmidome diversity is highly influenced by Bacteroides phylum diversity across host communities which suggests that this phylum serves as host to a large proportion of the plasmidome, thus governing plasmidome diversity as Bacteroides species diversity changes across rumen environments.

In light of these findings, in which changes in the microbiome are directly reflected as changes in the plasmidome, we then examined how dramatic changes in habitat ecology, caused by dietary shifts, will affect plasmidome composition, because these shifts are known to have a profound effect on the rumen microbiome (Friedman *et al*., 2017). We used the aforementioned animal group fed distinct diets and analysed microbiome and plasmidome composition of each cow by principal component analysis. This analysis revealed that both microbiome and plasmidome cluster according to diet (Fig. 5B). This was confirmed by analysis of similarity for the different groups, resulting in significant separation between samples coming from the two diets (p = 0.0008 for the microbiome and 0.0185 for the plasmidome). Moreover, Procrustes analysis was also performed as described in Muegge and colleagues (2011), resulting in high compatibility between plasmidome and microbiome similarities across the changing conditions (Monte Carlo p value < 10^{-5}, M^2 = 0.279) (Fig. S5). Taken together, our findings further support the notion that although plasmidome similarity across samples is comparatively low, it is governed by microbiome composition. Finally, changes in environmental conditions (i.e. diet) are known to have an immense effect on microbial composition. Here we show that these changes consistently affect plasmidome composition.

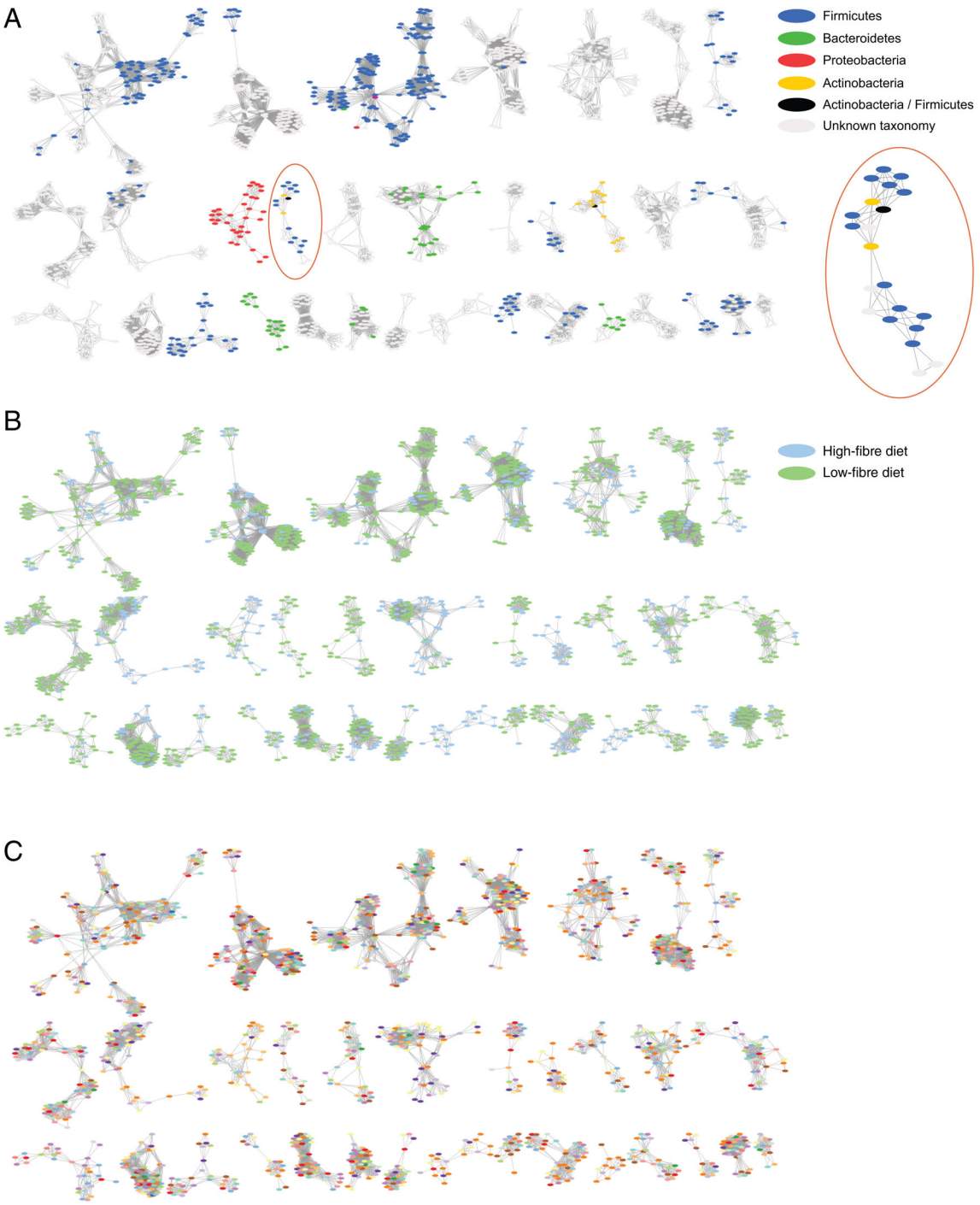

Fig. 4. The 33 biggest clusters of the plasmid interaction network. Nodes represent plasmids and are coloured according to taxonomy (A), diet (B) and the sample they were assembled from (C).
A. Interactions between plasmids are more often found within the same phylum, while some cross-phylum events are visible inside a cluster involving both related phyla (see magnification of orange circle).
B. Plasmid interactions are also more pronounced under similar diet.
C. No association observed between plasmid interactions and plasmids stemming from the same sample.

Plasmidome beta diversity is attributed mostly to plasmid replacement as opposed to plasmid loss

Biogeography research points to the need to disentangle two major processes affecting beta diversity among sites: spatial turnover and nestedness. The distinction between them could potentially shed light on the causality of biological processes leading to diversity between sites. In the first, high species turnover implies species

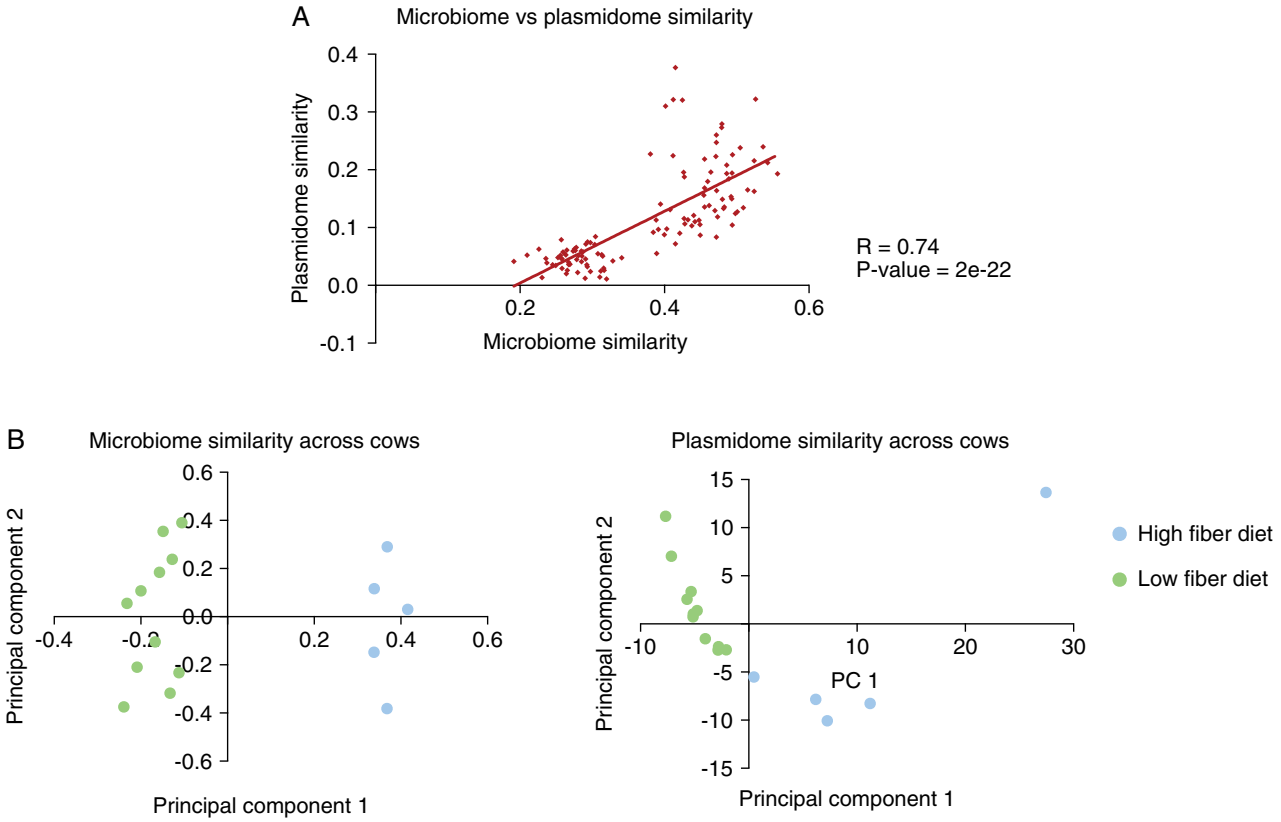

Fig. 5. OTU and plasmidome similarities.
A. Although the microbial OTUs of different samples are more similar to each other than the plasmidomes, there is a strong positive correlation between OTU and plasmidome similarity.
B. Principal component analysis shows microbiomes and plasmidomes to be grouped according to the host animal's dietary fibre content.

replacement by others; this influences similarity between any two niches by resulting in a differential composition among sites. In the second, the lower richness samples are subsets of the larger communities, indicating that a non-random process is governing this species assemblage (Baselga, 2010). Using the Baselga method for beta diversity partitioning, on average, 95% of the plasmidome beta diversity among rumens could be attributed to spatial turnover, implying that plasmids are very often replaced as opposed to simply being lost. This pattern of plasmidome beta diversity between rumen habitats strongly suggests that plasmids are efficient dispersers within this environment. Furthermore, the spatial turnover-related beta diversity of the plasmidome was highly correlated to that of the microbiome ($R = 0.706$, $p\text{-value} = 2.02{*}10^{-19}$) (Fig. S6), possibly implying that in this niche, plasmids are carried more by the highly replaceable species, resulting in the observed high plasmid diversity.

The microbial gut ecosystem is suggested to result from co-evolution of the host and its microbial communities (Ley *et al*., 2006; Mizrahi, 2013). The plasmidome, as a carrier of natural selection and genetic conversation among microbes, plays a key role in the co-evolutionary processes that shape this natural habitat (Mizrahi, 2012; Brown Kav *et al*., 2013b). Here we explored the plasmidome and its relationship with its hosting microbiome. The rumen plasmidome was found to greatly differ among individual rumen habitats of different cows living under the same conditions and fed the same diet. These differences were even more pronounced when the cows were fed two distinct diets, which are known to change both environment and microbial composition. Environmental changes, together with the selective pressure that they impose on microbial composition, have also been shown to have a substantial impact on the mobile genes found within the microbiome in a study of two distant human populations (Brito *et al*., 2016). However, even with these immense differences between plasmidomes, the hosting microbial environment seems to play a pivotal role in shaping the plasmid population within the rumen environment, as changes in the microbial communities corresponded to changes in their plasmidomes. Whether these changes occur due to the changes in microbial composition, environmental changes or both still remains to be answered.

Discussion

Overall, the results of this study support the notion that the plasmid community residing within the rumen microbial environment is highly diverse in many of its aspects. The hosting microbial community has a key but only partial role in shaping the plasmidome within it and across different communities. The high diversity attributed to the plasmidome was portrayed by the low congruence of plasmids across rumen microbiomes, suggesting that these are subjected to weaker or perhaps different selective forces than their residing microbial community, and that their distribution across individual rumens is affected by their high dispersal rates and perhaps plasmid–plasmid interactions. Low congruence among plasmids, even within hosts of similar phylogenetic backgrounds, has also been reported in the marine environment (Xue *et al*., 2015). It is tempting to speculate that the plasmid diversity reported here and the bacteriophage diversity reported by Ross and colleagues (2013) are somehow intertwined. It is plausible that these two are stuck in a continuous loop for the attention of their hosts. In phages, the term arms race is frequently used, emphasizing a pronounced effect of selective forces (Stern and Sorek, 2011). Here we propose that with plasmids, a more subtle interaction exists, one governed mostly by neutral forces and only in part by selective ones. We suggest that most plasmids are subjected to constant pressure to decrease their genome size so as not to burden their hosts, as seen by the apparently small proportion that rumen plasmids allow to be accepted by horizontal gene-transfer events (Fig. 4B) and total plasmid size distribution (Table S1, Fig. S1); while being in this condition, they are subjected to neutral forces such as drift and therefore need to function as very efficient dispersers, creating a pattern of high turnover that results in low similarity between different plasmidomes (Fig. S5). Nevertheless, although this applies to most of the plasmidome (80% in our analysis), when a beneficial gene is acquired by the plasmid that makes it more attractive to its host, positive selection comes into play and overcomes the burden caused by the increased plasmid length. Similarly, the presence of plasmid addiction systems with antitoxins (Fig. 1A) suggests another strategy than stealth. Such interactions will positively select a small proportion of plasmids that carry beneficial or forced-essential genes and allow mostly natural forces to govern the majority of plasmidome existence. Future studies, examining this model at both the community level and the single host level should be conducted to elucidate this model.

Another interesting notion is the fact that even within the same ecological niche, under the same selective environmental pressures, the accessory genome of rumen plasmids is highly diverse among rumen microbiomes. Not only do the plasmids themselves differ (Fig. 3D), but there is also little similarity in their content (low connectivity network in Fig. 4). Do different microbes from different rumen microbiomes require different accessory functions? Or is this the outcome of random drift as well? In addition, rumen plasmids seem to prefer transferring proportionally small fragments among them, possibly as a mechanism to reduce plasmid carriage fitness costs. This was also shown by Brito and colleagues (2016), where most mobile genes in the human microbiome were transferred without any memory of their original genetic context, indicating final transfer and recombination of only single-gene-size elements. We previously reported that the plasmidome carries a wide array of functions that are favourable to their hosting microbial community (Brown Kav *et al*., 2012). Together with the high diversity of plasmids and their low congruency across niches, as observed in the current study, we argue that the plasmidome serves as a genetic reservoir, carrying genes which are intermediately favourable, while alleviating the burden of their carriage from most bacteria, as has been proposed previously (Bergstrom *et al*., 2000; Heuer *et al*., 2008).

Finally, many of the current studies on microbial ecology have attempted to borrow or implement ecological theories originating from macroorganisms. Plasmids and more generally, lateral gene transfer, is a feature so unique to microorganisms that it remains completely disregarded by most of the currently available ecological theories. Many of the elements required for understanding the contribution of plasmids to the microbial habitat are still lacking regarding origin, propagation rate and selection processes, leaving only the ability to postulate, given our results, on the many undiscovered forces that shape plasmid communities and their interplay with their host microbiome.

Experimental procedures

Rumen sampling and bacterial extraction

The experimental procedures used in this study were approved by the Faculty Animal Policy and Welfare Committee at the Agricultural Research Organization's (ARO) Volcani Center and were in accordance with the guidelines of the Israel Council of Animal Care.

Israeli Holstein cows (n = 22) were housed at the ARO dairy farm in one shaded corral with free access to water and their diet. The low-fibre diet (n = 17) consisted of 70% concentrated food, mineral and vitamin mix and 30% roughage, 17% of which was dietary neutral detergent fibre (NDF); the high-fibre diet (n = 5) consisted of 95% roughage, 17% of which was NDF. Samples were taken 1 h after the morning feeding and microbial fraction collected as previously described (Jami *et al*., 2013).

Plasmidome extraction

For each purification, 2.25 g of bacterial pellet was used. The pellet was resuspended in 81 ml of 25% (wt/vol) sucrose and 50 mM Tris (pH 8), and then 6 ml of lysozyme (10 mg/ml in 250 mM Tris, pH 8) was added. The tube was mixed by inversion and incubated on ice for 5 min. Ethylenediaminetetraacetic acid (EDTA) (30 ml of 250 mM, pH 8) was then added mixed by inversion and then ice-chilled for 5 min. SDS was added (30 ml of a 20% wt/vol in TE), mixed by inversion, then subjected to eight cycles of heat pulse and mixing (one cycle consisted of 15 s in a 55°C water bath, removal from the bath and then inversions for 15 s). At ambient temperature, 30 ml of 3 M NaOH was added, immediately followed by 3 min of inversions. Then, 60 ml of 2 M Tris (pH 7.0) was added, followed by 39.6 ml of 20% SDS and immediately after that, 75 ml of 5 M NaCl was added. After inversions, the tubes were chilled on ice and refrigerated (4°C) for 6 h or overnight. The reactions were centrifuged at 10,000*g* (4°C, 30 min), the pellet was discarded and the supernatant was precipitated in isopropanol. The DNA pellet was resuspended in 1 ml of double distilled water (DDW) as previously described (Brown Kav *et al*., 2012).

Plasmid-safe DNase

The plasmids purified from rumen bacteria were subjected to plasmid-safe DNase (PSD; Epicentre) digestion. This adenosine triphosphate (ATP)-dependent PSD is known to digest linear double-stranded DNA. For each of the rumen plasmid purifications, 10 μg DNA was used as template in triplicate. The reactions were incubated overnight (12 h) at 37°C, and the DNase was inactivated at 70°C for 30 min as previously described (Jones and Marchesi, 2007). The presence of genomic DNA was tested by polymerase chain reaction (PCR) with 16S universal primers (BAC338 F 5′-ACTCCTACGGGAGGCAG-3′ and BAC805 R 5′-GACTACCAGGGTATCTAATCC-3′). Rumen samples which showed no band corresponding to the 16S rRNA gene were subjected to isopropanol precipitation and resuspended in 15 μl of DDW.

Amplification of plasmid DNA

For the selective amplification of circular plasmid DNA, phi29 polymerase was utilized. Amplification reactions contained 5 μl of PSD-digested, 16S rRNA gene-free, plasmid DNA as template, 1 μl of 10 μM exonuclease-resistant random hexamers (Fermentas), 2 μl of phi29 DNA polymerase reaction buffer (New England Biolabs) and 8.2 μl of DDW. Reactions were incubated at 95°C for 5 min and immediately chilled on ice. Then 1.6 μl of phi29 DNA polymerase (New England Biolabs) was added along with 0.02 μl inorganic pyrophosphatase (New England Biolabs) and 2 μl deoxyribonucleotide triphosphates (dNTPs) (10 mM). Reactions were then incubated at 30°C for 16 h. Finally, 3 μl of each reaction mixture was loaded onto a 1% agarose gel stained with ethidium bromide for analysis.

Quantitative PCR measurements

Quantitative real-time PCR analysis was performed to investigate the relative abundance of the 16S rRNA gene as a marker for genomic DNA content using the 16S universal primers (5′-ACTCCTACGGGAGGCAGCAGT-3′ and 5′-GTATTACCGCGGCTGCTGGCAC-3′) (Walter *et al*., 2000) as previously described (Brown Kav *et al*., 2013b).

Sequencing and assembly

The pure amplified plasmid DNA was subjected to deep sequencing via Illumina paired-end protocol (Illumina GAIIX sequencer and Illumina HiSeq). Reads were subjected to trimming and quality control using Trimmomatic (Bolger *et al*., 2014), and then assembled using SOAPdenovo (Li *et al*., 2008), OMEGA (Haider *et al*., 2014) and SPAdes (Bankevich *et al*., 2012). After testing these assemblers, we devised a new assembly method termed plasmid two-step assembly. This procedure uses the ability of SPAdes to leverage libraries of different insert lengths. Reads were first assembled using SPAdes, then aligned to the assembled contigs using bioinformatic tool (BWA) (Li and Durbin, 2009), then split into groups; these groups were treated as separate sequenced libraries on which a second round of SPAdes assembly was performed. The splitting of reads was based on BAM file alignment property flags. These flags separate read pairs into 'proper' pairs, that is, those having correct orientation and expected insert size, and 'improper' pairs that fail to meet at least one of these criteria. These two read groups were used as separate inputs to the second execution of SPAdes, with the aim of benefiting from improved repeat resolution and formation of more cycles. The last step of this procedure re-examined mapping of read pairs, while focusing on contig ends this time. The purpose of this step was the identification of self-loops closed by read pairs where these loops did not close in the assembly graph. Finally, the selection of only circular contigs was carried out by the Recycler tool (Rozov *et al*., 2017), developed together with our collaborators in the groups of Prof. Eran Halperin and Prof. Ron Shamir at Tel Aviv University.

Phylogenetic assignment of annotation of ORFs

ORF prediction was carried out using MetaGeneMark (Zhu *et al*., 2010). For the phylogenetic assignment of

ORFs, we used the previously described Phylogenie pipeline (Frickey and Lupas, 2004), which was implemented with maximum likelihood programs RXaML and FastTree (Stamatakis, 2006; Price *et al*., 2009) using the NCBI-NR database and 100 bootstrap replicates. The taxonomic assignment of each ORF was determined using bootstrap support of 80% or higher. We excluded ORFs that failed to generate a tree, had bootstrap support of less than 80% or ORFs whose taxonomic assignments were ambiguous.

ORFs were also annotated with the NCBI-NR protein database (*E*-value cut-off 10^{-5}). A subset of ORFs, originating from small plasmids (up to 2 KB), was further analysed; ORFs were further clustered into families with CD-HIT (at 80% similarity) (Huang *et al*., 2010) and representative sequences from each family were chosen for structural similarity analysis with PHYRE2 (at 90% confidence) (Kelley *et al*., 2015).

16S amplicon sequencing and data analyses

Samples from the same cows were also used for metagenomic DNA extraction as previously described (Dowd *et al*., 2008). The tagging and sequencing protocol was as described previously using the Miseq platform (Illumina). Data quality control and analyses were performed using the QIIME pipeline (Caporaso *et al*., 2010) as previously described (Jami *et al*., 2014)

Plasmid clustering

Although circular DNA elements do not have linear edges, the assemblers used in metagenomics studies produce a linear contig. This can result in two or more different linear representations of the same plasmid created by assembling the same plasmids in different samples. Thus, we developed a tool that clusters plasmid sequences based on their similarity while overcoming the technical challenges of handling plasmid sequences and enabling analysis of the plasmids' distribution across individual animals.

Plasmid clustering began by aligning each pair of plasmids in the data using BLASTn. Then, a set of in-house PERL scripts handled the next pipeline steps. Next, hits of less than 200 bps were removed and for each pair of plasmids, we summed the length of their overall alignments to remove bias introduced by the random location of linear edge (of a circle) output by the assembler for different samples. Finally, the pipeline collects the clusters that are in agreement with the pre-defined parameters and creates a matrix of presence/absence of the plasmid clusters across the samples.

We tested a range of coverage cut-offs to study the behaviour of clustering for various coverage spans and found 75% coverage to be ideal. When increasing the coverage cut-off to 80%, there was a drop in the number of plasmids in each cluster and an increase in the number of clusters, resulting in a larger number of small clusters. Below 75%, the number of clusters increased moderately, as expected when the stringency is lowered. Hence, in this pipeline, two plasmids were clustered together if their overall alignment spanned at least 75% of the longer sequence (coverage cut-off), and the DNA sequence alignment was at least 80% identical.

Calculating the distance between and within phyla in the rumen

A 16S rRNA tree was created using the amplicon sequences generated for the cows. 16S rRNA sequences were aligned using MUSCLE (Edgar, 2004), and the tree was created using FastTree (Price *et al*., 2009). We then calculated the branch-length distance between each pair of OTUs. The distance between the phyla was the average branch-length distances of all pairs of OTUs between the two phyla.

Plasmid networks

Plasmid networks in this study were based on interactions between plasmids. An interaction between plasmids was defined as sharing a sequence longer than 200 bps. We therefore compared the sequences of all plasmid pairs using the BLASTn algorithm. Plasmid interactions were then collected and the network was built using Cytoscape software.

References

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al*. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.

Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Glob Ecol Biogeogr* **19**: 134–143.

Bergstrom, C.T., Lipsitch, M., and Levin, B.R. (2000) Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* **155**: 1505–1519.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Brito, I.L., Yilmaz, S., Huang, K., Xu, L., Jupiter, S.D., Jenkins, A.P., *et al*. (2016) Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**: 435–439.

Brown Kav, A., Benhar, I., and Mizrahi, I. (2013a) Rumen plasmids. In *Lateral Gene Transfer in Evolution*, Gophna, U. (ed). New York, NY: Springer New York, pp. 105–120.

Brown Kav, A., Benhar, I., and Mizrahi, I. (2013b) A method for purifying high quality and high yield plasmid DNA for metagenomic and deep sequencing approaches. *J Microbiol Method* **95**: 272–279.

Brown Kav, A., Sasson, G., Jami, E., Doron-Faigenboim, A., Benhar, I., and Mizrahi, I. (2012) Insights into the bovine rumen plasmidome. *Proc Natl Acad Sci USA* **109**: 5452–5457.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al*. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Method* **7**: 335–336.

Dowd, S.E., Callaway, T.R., Wolcott, R.D., Sun, Y., McKeehan, T., Hagevoort, R.G., and Edrington, T.S. (2008) Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol* **8**: 125.

Duret, L. (2008) Neutral theory: the null hypothesis of molecular evolution. *Nat Educat* **1**: 218.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Frickey, T., and Lupas, A.N. (2004) PhyloGenie: automated phylome generation and analysis. *Nucleic Acids Res* **32**: 5231–5238.

Friedman, N., Shriker, E., Gold, B., Durman, T., Zarecki, R., Ruppin, E., and Mizrahi, I. (2017) Diet-induced changes of redox potential underlie compositional shifts in the rumen archaeal community. *Environ Microbiol* **19**: 174–184.

Haider, B., Ahn, T.-H., Bushnell, B., Chai, J., Copeland, A., and Pan, C. (2014) Omega: an overlap-graph de novo assembler for metagenomics. *Bioinformatics* **30**: 2717–2722.

Heuer, H., Abdo, Z., and Smalla, K. (2008) Patchy distribution of flexible genetic elements in bacterial populations mediates robustness to environmental uncertainty. *FEMS Microbiol Ecol* **65**: 361–371.

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010) CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**: 680–682.

Jami, E., Israel, A., Kotser, A., and Mizrahi, I. (2013) Exploring the bovine rumen bacterial community from birth to adulthood. *ISME J* **7**: 1069–1079.

Jami, E., White, B.A., and Mizrahi, I. (2014) Potential role of the bovine rumen microbiome in modulating Milk composition and feed efficiency. *PloS One* **9**: e85423.

Jeong, H., Sung, S., Kwon, T., Seo, M., Caetano-Anollés, K., Choi, S.H., *et al*. (2016) HGTree: database of horizontally transferred genes determined by tree reconciliation. *Nucleic Acids Res* **44**: D610–D619.

Jones, B.V., and Marchesi, J.R. (2007) Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nat Methods* **4**: 55–61.

Jørgensen, T.S., Kiil, A.S., Hansen, M.A., Sørensen, S.J., and Hansen, L.H. (2014a) Current strategies for mobilome research. *Front Microbiol* **5**: 750.

Jørgensen, T.S., Xu, Z., Hansen, M.A., Sørensen, S.J., and Hansen, L.H. (2014b) Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metamobilome. *PloS One* **9**: e87924.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**: 845–858.

Ley, R.E., Peterson, D.A., and Gordon, J.I. (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**: 837–848.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**: 713–714.

Luo, W., Xu, Z., Riber, L., Hansen, L.H., and Sørensen, S.J. (2016) Diverse gene functions in a soil mobilome. *Soil Biol Biochem* **101**: 175–183.

Mizrahi, I. (2012) The rumen plasmidome: a genetic communication hub for the rumen microbiome. *Mob Genet Elem* **2**: 152–153.

Mizrahi, I. (2013) Rumen symbioses. In *The prokaryotes*, Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. (eds). Berlin, Heidelberg: Springer, pp. 533–544.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., *et al*. (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**: 970–974.

Norman, A., Riber, L., Luo, W., Li, L.L., Hansen, L.H., and Sørensen, S.J. (2014) An improved method for including upper size range plasmids in metamobilomes. *PloS One* **9**: e104405.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**: 1641–1650.

Ross, E.M., Petrovski, S., Moate, P.J., and Hayes, B.J. (2013) Metagenomics of rumen bacteriophage from thirteen lactating dairy cattle. *BMC Microbiol* **13**: 242.

Rozov, R., Brown Kav, A., Bogumil, D., Shterzer, N., Halperin, E., Mizrahi, I., and Shamir, R. (2017) Recycler: an algorithm for detecting plasmids from de novo assembly graphs. *Bioinformatics* **33**: 475–482.

Sentchilo, V., Mayer, A.P., Guy, L., Miyazaki, R., Green Tringe, S., Barry, K., *et al*. (2013) Community-wide plasmid gene mobilization and selection. *ISME J* **7**: 1173–1186.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Stern, A., and Sorek, R. (2011) The phage-host arms race: shaping the evolution of microbes. *Bioessays* **33**: 43–51.

Tamminen, M., Virta, M., Fani, R., and Fondi, M. (2012) Large-scale analysis of plasmid relationships through gene-sharing networks. *Mol Biol Evol* **29**: 1225–1240.

Walter, J., Tannock, G.W., Tilsala-Timisjarvi, A., Rodtong, S., Loach, D.M., Munro, K., and Alatossava, T. (2000) Detection and identification of gastrointestinal *Lactobacillus* species by using denaturing gradient gel electrophoresis and species-specific PCR primers. *Appl Environ Microbiol* **66**: 297–303.

Xue, H., Cordero, O.X., Camas, F.M., Trimble, W., Meyer, F., Guglielmini, J., *et al*. (2015) Eco-evolutionary dynamics of episomes among ecologically cohesive bacterial populations. *mBio* **6**: e00552–15.

Zhu, W., Lomsadze, A., and Borodovsky, M. (2010) Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* **38**: e132–147.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1: Supporting Information