Antiviral defence systems in the rumen microbiome

Johan S. Sáenz[1,2,*], Bibiana Rios-Galicia[1,2] & Jana Seifert[1,2]

[1]Institute of Animal Science, University of Hohenheim, Emil-Wolff-Str. 6-10, 70593 Stuttgart, Germany

[2]HoLMiR—Hohenheim Center for Livestock Microbiome Research, University of Hohenheim, Leonore-Blosser-Reisen Weg 3, 70593 Stuttgart, Germany

*Corresponding author: johan.saenzmedina@uni-hohenheim.de

Abstract

The continuous interaction between phages and their respective hosts has resulted in the evolution of multiple bacterial immune mechanism. However, the diversity and prevalence of antiviral defence systems in complex communities is still unknown. We therefore investigated the diversity and abundance of viral defence systems in 3038 high-quality bacterial and archaeal genomes from the rumen. In total, 14,241 defence systems and 31,948 antiviral-related genes were identified. Those genes represented 114 unique system types grouped into 49 families. We observed a high prevalence of defence systems in the genomes. However, the number of defence systems, defence system families and systems density varied widely from genome to genome. Additionally, the number of defence system correlated positively with the number of defence systems families and the genome size. Restriction modification, Abi and cas system families were the most common, but many rare systems were present in only 1% of the genomes. Antiviral defence systems are prevalent and diverse in the rumen, but only a few are dominant, indicating that most systems are rarely present. However, the collection of systems throughout the rumen may

represent a pool of mechanisms that can be shared by different members of the community and modulate the phage-host interaction.

Importance

Phages may act antagonistically at the cell level but have a mutualistic interaction at the microbiome level. This interaction shapes the structure of microbial communities and is mainly driven by defence mechanism. However, the diversity of such mechanism is larger than previously thought. Because of that, we described the abundance and diversity of antiviral defence system of a collection of genomes, from the rumen. While defence mechanisms seem to be prevalent among bacteria and archaea, only a few were common. This suggests that most of these defence mechanisms are not present in many rumen microbes but could be shared among different members of the microbial community. This aligned with the 'pan-immune system' model, which appears to be common across different environments.

Keywords: defence-mechanism, phages, rumen microbiome, CRISPR-Cas

1. Introduction

It is estimated that demand for meat and milk will rise between 70-80% of current production by 2050 (1). Part of the current needs are met by raising cattle, which account for almost one third of all domesticated ruminants (2). However, sustainable beef and dairy production is challenged by economic and political issues, environmental factors, feed efficiency, and animal and human health (3, 4). Recent studies have shown that some of these problems are linked to the rumen microbiome.

The massive use of sequencing technologies has shown that the rumen is a complex ecosystem consisting of bacteria, archaea, protozoa, fungi, and viruses (5–7). This group of organisms contribute to the metabolism of complex plant materials and the production of volatile fatty acids (VFAs), which provide most of the energy required by the animal. Rumen bacteria and archaea have been associated with feed efficiency, methane production, animal health and fat and protein content of milk (8–11). Similarly, it has been shown that anaerobic fungi play a crucial role in the efficient degradation of fiber through enzymatic and physical processes (12). Rumen viruses are diverse, organism-specific, predominantly lytic, carry diverse auxiliary metabolic genes and have the potential to modulate nutrients recycling, fiber degradation and methanogenesis (13–16). In addition, dietary changes or perturbations can modulate viral richness and diversity, which has been shown to affect microbial metabolism (16).

Phages shape bacterial and archaeal populations and are a reservoir of accessory genes that facilitate the mobility of genetic material (17). The continuous interaction between phages and their respective hosts has resulted in the evolution of bacterial immune mechanism and phage counter-defence strategies (18, 19). This arms-race is a major driving force of molecular innovation and genetic diversification. Defence mechanism in prokaryotes are diverse and can be divided between innate and adaptive immune systems (20). The prevalence of defence mechanisms ranges from zero to several systems between cells. Defence systems can be compose by single or multiple genes, mainly associated with nucleases, helicases, proteases, kinases, ATPases and reverse transcriptases (21, 22). Antiviral defence mechanism are classify mainly in three groups: nucleic acid degradation, abortive infection and

inhibition of DNA/RNA synthesis, but the mechanism of the majority of systems remains unknow (23). Moreover, the effective immunity against phages seems to be driven by within-cell (24), presence of single or multiple system, and pan-genome dynamics (20). Previous studies have shown that rumen bacteria harbour multiple CRISPR-Cas and RM systems (14, 25), which can be associated with the modulation of phage-host interaction. However, a detail description of the diversity and prevalence of antiviral defence systems in the rumen is lacking. We explored the prevalence and diversity of antiviral defence mechanisms in 3038 high quality genomes from the rumen. This study shows that the repertoire of defence mechanisms in the rumen is broad and that the pan-immunity is mainly dominated by innate immune mechanisms.

2. Methods

2.1. Data collection and dataset description

In total 5990 genomes from the rumen microbiome were collected for the present study. Briefly, 5588 metagenomes assembled genomes (MAGs) from the MGnify (https://www.ebi.ac.uk/about/news/updates-from-data-resources/cow-rumen-catalogue-v10-released-mgnify/) cow rumen v1.0 MAG catalogue and 402 genomes of isolates from the Hungate 1000 catalogue of reference genomes were obtained (5). MAGs dataset is based on data collected from European cattle (26) and African cattle (27).

2.2. Quality control of MAGs

CheckM2 (28) was used to estimate the quality of the genomes and only the high-quality genomes with >90% completeness and <5% contamination were kept for

further analysis. Additionally, the N50, coding density, gene length, genome size and GC content of the genomes were calculated using CheckM2. The genomes were taxonomically classified using GTDB-tk v2.1 (29) and the GTDB database R207_v2 (30).

2.3. Antiviral defence prediction

Antiviral defence mechanisms were identified from all genomes using PADLOC v1.1 (22) and the PADLOC-DB. Systems labelled as other due to incomplete detection of essential genes were discarded. Defence systems were grouped by families using the PADLOC-DB metadata (https://github.com/padlocbio/padloc-db/blob/master/sys_meta.txt) and the references to relevant literature (https://github.com/padlocbio/padloc-db/blob/master/system_info.md). The density of defence systems per genome was calculated as the total defence systems per genome divided by the genome size (kbp). CRISPR-Cas genes and arrays were detected using CRISPRCasTyper v1.6 (31). All predicted CRISPR spacers from complete or incomplete CRISPR-Cas arrays were kept. Viral genomes from the rumen virome database (RVD) were matched against CRISPR spacers using Blast (32), only those with >95% identity, <= 1 mismatch and <0.0004 e-value were considered spacer to protospacer matches. Host-virus interaction network showing the spacer-protospacer match was created using Cytoscape (33).

2.4. Prophages detection and identification

Collected genomes were searched for prophages using Virsorter2 v2.2.3 (34). Only contigs longer than >10,000 were used. All contigs identified as prophages were kept for further analysis. Prophages were quality-checked using checkV (35). Prophages

with 0 viral genes and >1 host gene were discarded and prophages with 0 viral genes and 0 host genes were kept. A prophage was considered complete if it was predicted as such by checkV.

2.5. Phylogenetic placement

Phylogenetic tree of the whole 3038 genomes was created using Phylophlan (36) and 400 universal marker genes, with --medium diversity and --accurate options. Phylogenetic relation among species of *Butyrivibrio*, Ca. *Limimorpha*, *Prevotella* and *Ruminococcus* was calculated using a subset of genes present in at least 90% of all genomes (core genes) for each *Genus*. *Butyrivibrio*, *Prevotella* and *Ruminococcus* were selected due to their key role in the rumen and high prevalence of defence systems, while Ca. *Limimorpha due to its novelty*. Core genes were extracted, aligned, and concatenated using the tool: concatenated-gene-alignment-fasta from Anvi'o v 7.1 (37). The extracted genes (55 of *Butyrivibrio*, 71 of *Limimorpha*, 13 of *Prevotella* and 46 of *Ruminococcus*) were used to observe phylogenetic relatedness. Phylogenetic trees were inferred by maximum-likelihood using the concatenate of core genes of each genome with FastTree2 (38). Taxonomic information, number of defence families and numbers of prophages were displayed in a phylogenetic tree using iTOL (39).

2.6. Data wrangling and data availability

All data wrangling and statistical analysis were performed in R v4.1.2 (40) and the packages "tidyverse" (41), "data.table", "pheatmap", "ggdist", "patchwork", "ggtext", and "viridis" were used. Medians across groups were compared using the nonparametric Kruskal-Wallis test (*kruskall.test* function). Additionally, correlations were calculated using the Spearman's rank correlation coefficient (*cor.test*, method =

"spearman"). Data processing workflow can be found as a Jupyter-lab notebook at https://github.com/SebasSaenz/Papers_wf as well as other R scripts for data wrangling and visualisation. The MGnify cow rumen v1.0 MAG catalogue collection can be found at http://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes/cow-rumen/v1.0/ and the Hungate 1000 catalogue of reference genomes from the rumen microbiome is available at https://genome.jgi.doe.gov/portal/TheHunmicrobiome/TheHunmicrobiome.info.html.

3. Results

3.1. Dataset description

A total of 5990 rumen microbiome genomes were collected from the MGnify genome resource (42) and the Hungate1000 collection (**Table S1**). In summary, 5578 of the collected genomes were metagenome-assembled genomes (MAGs) and 412 were bacterial and archaeal isolates. After checking the quality of the assemblies using CheckM2, 2930 genomes were classified as high-quality drafts, 3035 as medium quality drafts and zero as low-quality drafts. The remaining genomes were discarded because they had more than 10% contamination or completeness could not be calculated. Based on genome quality, 3038 genomes with ≥ 90% completeness were retained for further analysis. Briefly, the selected genomes had a mean 107,308 bp ± 274,286 N50, 0.901 ± 0.0263 coding density, 337 bp ± 27.8 gene length, 2,535,744 bp ± 794,061 genome size and 48.5 ± 8.6% GC content (**Table S2**). In terms of taxonomic classification, 62 genomes were classified as *Archaea* by GTDB-TK (29), while 2976 were classified as *Bacteria*. In addition, 22 phyla were identified across the dataset, of which 10 were represented by more than 20 genomes. Genomes

classified as *Bacillota* were the most abundant in the dataset representing 50.8% of the total genomes, followed by *Bacteroidota* (32.4%), *Pseudomonadota* (3.7%), *Actinomycetota* (3.4%), *Cyanobacteriota* (2.5%) and *Methanobacteriota* (1.7%). Moreover, genomes identified as the genus *Prevotella* (8.4%) were the most prevalent, followed by *Cryptobacteroides* (6.6%), uncultured CAG-791 (3.3%), *Ruminococcus* group E (3.0%) and *Sodaliphilus* (2.9%).

3.2. Prevalence of defence mechanism

Antiviral defence systems are prevalent in the rumen microbiome. In total, 14,241 defence systems and 31,948 antiviral related genes, representing 114 unique system types, were identified across all genomes. Out of the 3038 genomes, 89% of them carry at least one antiviral defence system. Approximately, 99% of isolates harboured at least one system, compared to 87% of the MAGs (**Figure 1A**). On average, isolates had a larger genome (**Figure S1**), number of defence system families (p-value<2.2e-16, **Figure 1B**), number of defence systems (p-value<2.2e-16, **Figure 1C**) and defence systems density (p-value=1.051e-15, **Figure 1H**) compared with the MAGs. Similarly, 96% of archaeal genomes harboured at least one defence system, compared to 88% of bacterial genomes **(Figure 1D, Table S3)**. On average, *Archaea* seem to carry 4.9 systems per genome while *Bacteria* 4.7 (p-value=0.1875, **Figure 1F**). However, this number highly varies, being zero the minimum (3.2% archaeal and 11.1% bacterial genomes) up to 34 in the Ca. *Limimorpha*. Approximately 92% of archaeal and bacterial genomes carry between one to ten systems but the remaining genomes harbour between 11 to 34 and 11 to 19 systems in *Bacteria* and *Archaea* respectively **(Figure 2A)**. The density of defence systems per genome significantly differed between *Archaea* and *Bacteria*

(p-value = 0.003752, **Figure 1G**), $2.12x10^{-3}$ and $1.75x^{-3}$, respectively. Additionally, after grouping the systems into families we found that the average number of unique antiviral defence families in archaeal and bacterial genomes is similar, 2.8 and 2.6 respectively (p-value = 0.1456, **Figure 1E**). The minimum number of families found was zero and 13 the maximum in the uncultured *Lachnospiraceae* CAG-603.

When compared by phyla, between 78% and 100% of the genomes carry at least one defence system (**Figures S2**), being *Verrucomicrobiota* and *Fibrobacterota* the highest and *Pseudomonadota* the lowest. *Bacillota* and *Bacteroidota* phyla are among the most prevalent in the dataset, with 84% and 96% of the genomes exhibiting defensive mechanisms respectively. Similarly, in average *Verrucomicrobiota* and *Fibrobacterota* genomes have the higher number of defence systems and defence systems density (**Figure S2**).

We then used Virsorter2 to detect proviral sequences longer than 10,000 bp in the genomes and CheckV to evaluate their quality. The results were compared with the number of defence mechanisms found in each genome. Forty percent of the genomes harboured at least one provirus, being zero the lowest and 10 the highest in one bacterial genome. A positive correlation was observed between the average number of defence systems and the number of proviruses per genome in archaea **(**Spearman ρ=0.575, p-value= 1.005e-06, **Figure 2C)**, however this was not observed in the bacterial genomes (Spearman ρ=0.080, p-value= 1.04e-05). In contrast, the genome size positively correlates with the number of defence systems in archaeal (Spearman, ρ=0.528, p-value=1.421e-05) and bacterial genomes (Spearman, ρ=0.525, p-value < 2.2e-16**, Figure 2B).** The relation between the

number of proviruses and the genome size has been previously evaluated as potential drivers of the number of antiviral systems in prokaryotic genomes and plasmids (23, 43).

3.3. Diversity and abundance of defence mechanism

To estimate the diversity of antiviral defence systems, all predicted systems were grouped into families using the metadata provided by PADLOC-DB. In total, 114 different system types were found across all genomes, and they were grouped in 49 families. All of them were found in bacterial genomes, whereas only 16 of them were found in the archaeal genomes. RM, Abi and cas system families were present in 72.1, 55.2 and 27.9 % of all genomes. Moreover, 28 of the families were present only in 1% of the genomes. In terms of relative abundance, systems belonging to RM were the most abundant in *Archaea* (51.1%) and *Bacteria* (44.5%) compared to the total systems found **(Figure 3A and 3B)**. However, not always the most abundant systems were equally present in both groups. For example, the abundance of the Abi system was 6.9 and 24.6% in *Archaea* and *Bacteria* respectively. Similarly, the family AVAST represented 5.6 and 0.4% of all systems found in *Archaea* and *Bacteria*. Additionally, the relative abundance of 34 of the total families was less than 1%.

3.4. Defence mechanism in key members of the rumen

To corroborate if the antiviral defence systems are enriched in a specific taxon, we selected all genera represented by more than 10 genomes (n=57). The number of defence systems and families of the selected genera were compared. Both total numbers of systems and number of different families vary between genera. On

average, these genomes (genera represented by at least 10 genomes) carried 4.6 defence systems and 2.6 defence system families. The uncultivated genus Ga6A1 (12.8) and Ca. *Limimorpha* (12.6) had the highest number of systems, followed by *Fibrobacter* (9.8) **(Figure 4B)**. Similar trends were observed when the number of families were compared **(Figure 4A)**. None of the genera were totally devoid of defence systems but *Enterousia* and several uncultivated bacteria had a low prevalence of systems. Additionally, we observed a positive correlation between the number of system and families, which indicates that *Bacteria* or *Archaea* carrying several defence systems will also harbour more defence system families (**Figure S3**)

Additionally, the diversity of defence system and system families varies across the dominant members of the rumen (**Figure 5A, S4**). For example, between 37-97% of genomes of the genera *Prevotella*, Ca. *Limimorpha*, *Butyrivibrio* and *Ruminococcus* carry RM and Abi systems (**Figure 5B**). However, 22, 31, 24, and 23 different defence system families are present in less than 20% of the same genomes respectively. The Ca. *Limimorpha* seems to be an example of a genus carrying a core group of defence genes, because at least 50% of its genomes harboured RM, Abi, Pycsar, cas and DRT systems (**Figure 5B**).

3.5. CRISPR-Cas arrays and phage interaction

Cas proteins and CRISPR arrays have been reported across complete or draft prokaryotic genomes from different environments (43, 44). However, the accurate automated prediction of CRISPR-Cas loci is challenging due to the variability in genome quality and the ambiguous prediction of the system subtypes. Therefore, we

used CRISPRCASTyper to identify complete loci and the system subtype based on both Cas genes and CRISPR repeats in the rumen genomes. In summary, 791 CRISPR-Cas complete locus were found in 668 genomes. This indicates that most of these genomes carry only one complete system. In total, 21.9% of all 3038 rumen microbiome genomes harbour at least one CRISPR-Cas complete locus. When compared, more bacterial genomes seem to carry more complete systems than archaeal genomes, 22.1 and 14.5 % respectively **(Figure 6A)**. The phylum with the highest prevalence of CRISPR-Cas was the uncultured UBP6 (42% of genomes), followed by *Actinomycetota* (31%), *Bacteroidota* (26%) and *Verrucomicrobiota* (26%) **(Figure 6B)**. Interestingly, of the phyla represented by at least 50 genomes, only *Cyanobacteriota* was almost devoid of complete CRISPR-Cas. In this case, one genome out of 77 carried an array. In total, 21 CRISPR-Cas subtypes were predicted across all genomes and among them the subtype I-C was the most abundant (21.9%), followed by II-C (15.1%), II-A (13.5%) and I-E (10.3%). The number of subtypes found per phyla greatly varied, being 15 the maximum in *Bacillota* followed by 13 in *Bacteroidota*. However, in the phylum *Spirochaetota*, *Eremiobacterota*, *Elusimicrobiota*, and *Desulfobacterota* only one subtype was found in all genomes **(Figure 6C)**.

All CRISPR spacers predicted by CRISPRCASTyper were collected, including those from incomplete CRISPR-Cas locus. In total, 27,717 spacers were found in 845 genomes, ranging from one to one hundred spacers per array. All collected spacers were aligned against a rumen virome database (RVD) using blast, containing 397,180 species-level viral operational taxonomic units (vOTUs). The vOTUs were mined from 975 public rumen metagenomes, including samples from four different

continents and 13 animal species. In total, 30.2% of the spacers matched against 4986 vOTUs, which could indicate a previous history of phage-bacterial and phage-archaeal infection (**Figure 7, Table S4**). Most of the interactions were found between members of the phylum *Bacillota*, *Bacteroidota* and *Actinomycetota* and different vOTUs. For example, two uncultured species from the *Firmicutes* group A have a history of interaction with 53 and 52 different vOTUs, which was the maximum found (**Table S5**). Similarly, the spacers of the archaea *Methanobrevibacter* and bacteria *Prevotella* matched with 49 and 39 different vOTUs, respectively. Additionally, approximately 7% of the collected phages showed intra-species host specificity, while 2.6% intra-genus host specificity (**Table S6**). Only 1.8% of the vOTUs that matched with the spacers have been classified as members of the *Siphoviridae* and *Myoviridae* families, while the rest remain unclassified.

4. Discussion

It is estimated that phages are 10-times more abundant than bacteria and their role in shaping microbial communities, influencing genetic diversity, and participating in biogeochemical cycles is a crucial aspect of the Earth's ecosystems (45, 46). By exerting evolutionary pressure on host populations, phages drive the evolution of their hosts. In the rumen, phages are diverse and infect the core microbiome, including *Prevotella*, *Ruminococcus*, *Fibrobacter*, *Butyrivibrio*, and Methanobrevibacter (13, 47). This is especially important because the arm-race between phages and host can modulate the functional traits in the rumen (14). This arm-race is driven by defence systems in prokaryotes and by anti-defence proteins in phages and other MGE (18, 19). In the rumen we found that defence systems are

diverse, mainly associated to innate immunity and dominated by RM, abortive systems and cas proteins. The high prevalence of defence systems with a broad genetic organisation suggests that selective pressure has been exerted by phages and other MGEs in the rumen. However, this pressure also triggers the evolution of antidefence encoded in phages. For example, proteins that inhibit CRISPR-Cas, Gabija, Thoeris and Hachiman systems are present in phages that infect taxonomically diverse bacterial species (19, 48, 49). The rumen is not the exception as it has been found that rumen phages carry anti-RM system, which represents a counter-defence mechanism of the most abundant prokaryotic defence (13). The emergence of different strategies to evade defence systems in phages and other MGEs could drive the rapid acquisition of diverse and novel defence systems in prokaryotes. This arm-race extends even between viruses, as phages are reservoir of antiviral systems which can inhibit competing viruses (50, 51).

The high variability of defence systems and defence system families across the genomes from different environments, including the rumen, could back the ‘pan-immune system’ model (20). This model suggests that genetically closely related strains could share defence systems via horizontal gene transfer, which indicates that the immunity could be driven by the collection of systems in the community. Mobile genetic elements (MGEs) play a main role in the mobilization of defence systems because they often carry defence genes, even more frequently compared to chromosomes (52, 53). Moreover, defence system seen to organize in defence island, accompanied by MGE, from which novel defence genetic architecture could emerge (21, 52). In the rumen, the pan-immunity could be driven mainly by the innate response, as CRISPR-Cas is only present in approximately 21% of the rumen

genomes, but they have been shown to be important in the evolutionary dynamic between phage and their host in the rumen (14). This seems to be common also in other environments like cheese and silage-associated bacterial communities (54, 55). The genomes analysed in the present study harbouring CRISPR-Cas systems showed a history of infection with phages mined from rumen samples. Some of this phages could infect not only several genomes but also species and genera, which has been previously reported in samples from ruminants (56), the human gut (57) and the ocean**.** A broad host range could possibly have arisen through adaptation to different receptors by mutations, which is also considered as an advantage for the arm-race without acquiring anti-defence systems (58). Such polyvalent phages can influence the rumen metabolic traits at the cellular and microbiome level. For example, our data showed that *Prevotella* and *Methanobrevibacter* with a wide repertoire of defence mechanism, involved in fiber degradation and methanogenesis, have a broad interaction with phages (59, 60). This was consistent with previous reports on the modulating role of the rumen viral community in key metabolic traits like nutrient recycling, fiber degradation and methanogenesis. (14, 61). This shows that even though in low proportion the adaptive immunity can participate in the response of specific infections that can modulate the microbial and animal metabolism.

Moreover, the present study adds to the growing evidence that highlights the prevalence and significance of antiviral defence systems in prokaryotic communities. We show that defence mechanisms in the rumen microbiome are prevalent, with almost 90% of the bacterial and archaeal genomes carrying at least one system. A similar number of defence systems has been observed within genomes from the

NCBI RefSeq (23) and MAGs from soil and human gut (**Figure S5**) (52). Additionally, the number of systems per genome from different environments seem to follow a binomial distribution, where most genomes carry between 2-5 systems. We observed a similar distribution within the families of the defence systems in the rumen. The combination of different systems can provided a broader resistance and resistance range (62, 63) and the widespread distribution of these systems suggests a crucial role in maintaining microbial diversity and ecosystem stability.

Also, our findings further reinforce the notion that defence mechanisms are diverse but just a few are dominant in the rumen environment. Consistently, RM, Abi and Cas are the most abundant defence system families across different genomes and environments (52, 64). At least in the rumen these three families represent approximately 70% of all defence systems. Furthermore, 28 out of the total 49 families found in the rumen were present only in 1% of the genomes. This shows that these systems are rare, but they are still present in diverse genomes. Such numbers have been observed in a collection of 21000 RefSeq genomes, which also suggests that the diversity of antiviral systems is bigger than previously thought (23).

Different environmental and genetic factors can contribute to the selection of genes (65, 66). Previous studies have shown that bacterial and phage genome, and plasmid size seems to drive the acquisition of defence mechanisms (23, 43, 67). We also found that this is true for bacteria and archaea in the rumen, as the number of defence systems positively correlated with genome size. Genomes would tend to be larger when defence systems such as CRISPR-Cas are present due to their size, and the same could apply when multiple small defence systems are recruited in the

same genome or defence islands (52). On the other hand, the number of prophages in the genome had a low or positive correlation with the number of defence systems. We were expecting a negative correlation between the number of defence systems and prophages, as more systems can provide a broader resistance, but it was not the case in our samples. However, other studies have found that the abundance of CRISPR-Cas systems can be linked with the viral abundance (68).

It is relevant to point that our study has a few limitations. Firstly, most of the used genomes in the study are product of MAG binning, which is known to poorly reconstruct closely related strains (69). This is especially important, because it could limit the detection of defence systems due to the variations of defensome composition between strains (21). Additionally, short-sequencing may not provide accurate genome-specific signals, low abundance, and could affect the sensitivity/specificity trade-off of the single-gene or multi-gene system detection specially in large CRISPR arrays (22). Secondly, the unbalanced number of isolates and MAGs shows that the rumen microbiome is still largely unculturable. In this way the description of the defensome is mostly computational dependant. Also, isolates tend to harbour more defence system, defence system families and a higher defence system density, which can be associated with the genome completes and low fragmentation. The accumulation of more defence mechanism in isolates compared to MAGs can be mainly associates to the methods used to recover them and not their lifestyle or role in the microbial community. However, a larger portion of isolates need to be collected to draw robust conclusion. Besides these limitations, we believe that this targeted data analysis is a starting point to expand the search for defence

system in the rumen using state of the art techniques such as long sequencing and high throughput culturomics.

5. Conclusion

Antiviral defence systems are prevalent and diverse in the rumen, but only a few are dominant, indicating that many systems are rarely present. However, the collection of systems throughout the rumen may represent a pool of mechanisms that can be shared by different members of the community. This could support the 'pan-immune system' model, which appears to be common across different environments.

Ethics approval and consent to participate

Not applicable

Availability of data and materials

Data processing workflow can be found as a Jupyter-lab notebook at https://github.com/SebasSaenz/Papers_wf as well as other R scripts for data wrangling and visualisation. The MGnify cow rumen v1.0 MAG catalogue collection can be found at http://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes/cow-rumen/v1.0/ and the Hungate 1000 catalogue of reference genomes from the rumen microbiome is available at https://genome.jgi.doe.gov/portal/TheHunmicrobiome/TheHunmicrobiome.info.html.

Competing interests

The authors declare that they have no competing interests.

Funding

German Research Foundation (DFG) project number 327953272 (SE2059/3-1) and project number 202989534 (SE2059/2-2).

Authors' contributions

JSS collected and analysed the data, prepared the figures and wrote the manuscript. BRG contributed to the analysis of the data. JS contributed to the discussion of the data, supervision and funding. All authors revised and approved the final version of the manuscript.

Acknowledgements

The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG.

References

1. Herrero M, Wirsenius S, Henderson B, Rigolot C, Thornton P, Havlík P, de Boer I, Gerber PJ. 2015. Livestock and the Environment: What Have We Learned in the Past Decade? Annu Rev Environ Resour 40:177–202.

2. Hackmann TJ, Spain JN. 2010. Invited review: ruminant ecology and evolution: perspectives useful to ruminant livestock research and production. J Dairy Sci 93:1320–1334.

3. Hubbart JA, Blake N, Holásková I, Mata Padrino D, Walker M, Wilson M. 2023. Challenges in Sustainable Beef Cattle Production: A Subset of Needed Advancements. Challenges 14:14.

4. FAO. 2023. Contribution of terrestrial animal source food to healthy diets for

improved nutrition and health outcomes – An evidence and policy overview on the state of knowledge and gaps. Rome.

5. Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, Eloe-Fadrosh EA, Pavlopoulos GA, Hadjithomas M, Varghese NJ, Paez-Espino D, Hungate1000 project collaborators, Perry R, Henderson G, Creevey CJ, Terrapon N, Lapebie P, Drula E, Lombard V, Rubin E, Kyrpides NC, Henrissat B, Woyke T, Ivanova NN, Kelly WJ. 2018. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. Nat Biotechnol 36:359–367.

6. Ishaq SL, AlZahal O, Walker N, McBride B. 2017. An Investigation into Rumen Fungal and Protozoal Diversity in Three Rumen Fractions, during High-Fiber or Grain-Induced Sub-Acute Ruminal Acidosis Conditions, with or without Active Dry Yeast Supplementation. Front Microbiol 8:1943.

7. Xie F, Jin W, Si H, Yuan Y, Tao Y, Liu J, Wang X, Yang C, Li Q, Yan X, Lin L, Jiang Q, Zhang L, Guo C, Greening C, Heller R, Guan LL, Pope PB, Tan Z, Zhu W, Wang M, Qiu Q, Li Z, Mao S. 2021. An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. Microbiome 9:137.

8. Tröscher-Mußotter J, Saenz JS, Grindler S, Meyer J, Kononov SU, Mezger B, Borda-Molina D, Frahm J, Dänicke S, Camarinha-Silva A, Huber K, Seifert J. 2021. Microbiome Clusters Disclose Physiologic Variances in Dairy Cows Challenged by Calving and Lipopolysaccharides. mSystems 6:e0085621.

9. Tapio I, Snelling TJ, Strozzi F, Wallace RJ. 2017. The ruminal microbiome associated with methane emissions from ruminant livestock. J Anim Sci Biotechnol 8:7.

10. Xue M-Y, Xie Y-Y, Zhong Y, Ma X-J, Sun H-Z, Liu J-X. 2022. Integrated meta-omics reveals new ruminal microbial features associated with feed efficiency in dairy cattle. Microbiome 10:32.

11. Amin N, Schwarzkopf S, Tröscher-Mußotter J, Camarinha-Silva A, Dänicke S, Huber K, Frahm J, Seifert J. 2023. Host metabolome and faecal microbiome shows potential interactions impacted by age and weaning times in calves. Anim Microbiome 5:12.

12. Hartinger T, Zebeli Q. 2021. The Present Role and New Potentials of Anaerobic Fungi in Ruminant Nutrition. J Fungi (Basel) 7.

13. Yan M, Pratama AA, Somasundaram S, Li Z, Jiang Y, Sullivan MB, Yu Z. 2023. Interrogating the viral dark matter of the rumen ecosystem with a global virome database. Nat Commun 14:5254.

14. Zhong Z-P, Du J, Köstlbacher S, Pjevac P, Orlić S, Sullivan MB. 2024. Viral potential to modulate microbial methane metabolism varies by habitat. Nat Commun 15:1857.

15. Wu Y, Gao N, Sun C, Feng T, Liu Q, Chen W-H. 2024. A compendium of ruminant gastrointestinal phage genomes revealed a higher proportion of lytic phages than in any other environments. Microbiome 12:69.

16. Anderson CL, Sullivan MB, Fernando SC. 2017. Dietary energy drives the dynamic response of bovine rumen viral communities. Microbiome 5:155.

17. Brown TL, Charity OJ, Adriaenssens EM. 2022. Ecological and functional roles of bacteriophages in contrasting environments: marine, terrestrial and human gut. Curr Opin Microbiol 70:102229.

18. Hampton HG, Watson BNJ, Fineran PC. 2020. The arms race between bacteria and their phage foes. Nature 577:327–336.

19. Yirmiya E, Leavitt A, Lu A, Ragucci AE, Avraham C, Osterman I, Garb J, Antine SP, Mooney SE, Hobbs SJ, Kranzusch PJ, Amitai G, Sorek R. 2024. Phages overcome bacterial immunity via diverse anti-defence proteins. Nature 625:352–359.

20. Bernheim A, Sorek R. 2020. The pan-immune system of bacteria: antiviral defence as a community resource. Nat Rev Microbiol 18:113–119.

21. Georjon H, Bernheim A. 2023. The highly diverse antiphage defence systems of bacteria. Nat Rev Microbiol 21:686–700.

22. Payne LJ, Todeschini TC, Wu Y, Perry BJ, Ronson CW, Fineran PC, Nobrega FL, Jackson SA. 2021. Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. Nucleic Acids Res https://doi.org/10.1093/nar/gkab883.

23. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, Bernheim A. 2022. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun 13:2561.

24. Arias CF, Acosta FJ, Bertocchini F, Herrero MA, Fernández-Arias C. 2022. The coordination of anti-phage immunity mechanisms in bacterial cells. Nat Commun 13:7412.

25. Gilbert RA, Kelly WJ, Altermann E, Leahy SC, Minchin C, Ouwerkerk D, Klieve AV. 2017. Toward understanding phage:Host interactions in the Rumen; Complete genome sequences of lytic phages infecting Rumen bacteria. Front Microbiol 8:2340.

26. Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. 2019. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. Nat Biotechnol 37:953–961.

27. Wilkinson T, Korir D, Ogugo M, Stewart RD, Watson M, Paxton E, Goopy J, Robert C. 2020. 1200 high-quality metagenome-assembled genomes from the rumen of African cattle and their relevance in the context of sub-optimal feeding. Genome Biol 21:229.

28. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. 2023. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods 20:1203–1212.

29. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics https://doi.org/10.1093/bioinformatics/btz848.

30. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol 38:1079–1086.

31. Russel J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ. 2020. CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci. CRISPR J 3:462–469.

32. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

33. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504.

34. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9:37.

35. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol 39:578–585.

36. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, Zhu Q, Bolzan M, Cumbo F, May U, Sanders JG, Zolfo M, Kopylova E, Pasolli E, Knight R, Mirarab S, Huttenhower C, Segata N. 2020. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun 11:2500.

37. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, Trigodet F, Watson AR, Esen ÖC, Moore RM, Clayssen Q, Lee MD, Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjödin A, Scott JJ, Vázquez-Campos X, McKay LJ, McDaniel EA, Stevens SLR, Anderson RE, Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6:3–6.

38. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.

39. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296.

40. R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

41. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the tidyverse. J Open Source Softw 4:1686.

42. Gurbich TA, Almeida A, Beracochea M, Burdett T, Burgin J, Cochrane G, Raj S, Richardson L, Rogers AB, Sakharova E, Salazar GA, Finn RD. 2023. MGnify Genomes: A Resource for Biome-specific Microbial Genome Catalogues. J Mol Biol 435:168016.

43. Pinilla-Redondo R, Russel J, Mayo-Muñoz D, Shah SA, Garrett RA, Nesme J, Madsen JS, Fineran PC, Sørensen SJ. 2021. CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids. Nucleic Acids Res https://doi.org/10.1093/nar/gkab859.

44. Burstein D, Sun CL, Brown CT, Sharon I, Anantharaman K, Probst AJ, Thomas BC, Banfield JF. 2016. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. Nat Commun 7:10613.

45. Correa AMS, Howard-Varona C, Coy SR, Buchan A, Sullivan MB, Weitz JS. 2021. Revisiting the rules of life for viruses of microorganisms. Nat Rev Microbiol 19:501–513.

46. Eppley JM, Biller SJ, Luo E, Burger A, DeLong EF. 2022. Marine viral particles reveal an expansive repertoire of phage-parasitizing mobile elements. Proc Natl Acad Sci U S A 119:e2212722119.

47. Yu Z, Yan M, Somasundaram S. 2024. Rumen protozoa and viruses: The predators within and their functions-A mini-review. JDS Commun 5:236–240.

48. Davidson AR, Lu W-T, Stanley SY, Wang J, Mejdani M, Trost CN, Hicks BT, Lee J, Sontheimer EJ. 2020. Anti-CRISPRs: Protein inhibitors of CRISPR-Cas systems. Annu Rev Biochem 89:309–332.

49. Samson JE, Magadán AH, Sabri M, Moineau S. 2013. Revenge of the phages: defeating bacterial defences. Nat Rev Microbiol 11:675–687.

50. Rousset F, Depardieu F, Miele S, Dowding J, Laval A-L, Lieberman E, Garry D, Rocha EPC, Bernheim A, Bikard D. 2022. Phages and their satellites encode hotspots of antiviral systems. Cell Host Microbe 30:740-753.e5.

51. Patel PH, Taylor VL, Zhang C, Getz LJ, Fitzpatrick AD, Davidson AR, Maxwell KL. 2024. Anti-phage defence through inhibition of virion assembly. Nat Commun 15:1644.

52. Beavogui A, Lacroix A, Wiart N, Poulain J, Delmont TO, Paoli L, Wincker P,

Oliveira PH. 2024. The defensome of complex bacterial communities. Nat Commun 15:2146.

53. Oliveira PH, Touchon M, Rocha EPC. 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res 42:10618–10631.

54. Somerville V, Schowing T, Chabas H, Schmidt RS, von Ah U, Bruggmann R, Engel P. 2022. Extensive diversity and rapid turnover of phage defense repertoires in cheese-associated bacterial communities. Microbiome 10:137.

55. Sáenz JS, Rios-Galicia B, Rehkugler B, Seifert J. 2023. Dynamic Development of Viral and Bacterial Diversity during Grass Silage Preservation. Viruses 15:951.

56. Yan M, Yu Z. 2024. Viruses contribute to microbial diversification in the rumen ecosystem and are associated with certain animal production traits. Microbiome 12:82.

57. Zeng S, Almeida A, Li S, Ying J, Wang H, Qu Y, Paul Ross R, Stanton C, Zhou Z, Niu X, Mu D, Wang S. 2024. A metagenomic catalog of the early-life human gut virome. Nat Commun 15:1864.

58. Holtappels D, Alfenas-Zerbini P, Koskella B. 2023. Drivers and consequences of bacteriophage host range. FEMS Microbiol Rev 47:fuad038.

59. Danielsson R, Dicksved J, Sun L, Gonda H, Müller B, Schnürer A, Bertilsson J. 2017. Methane production in dairy cows correlates with Rumen methanogenic and bacterial community structure. Front Microbiol 8:226.

60. Betancur-Murillo CL, Aguilar-Marín SB, Jovel J. 2022. Prevotella: A Key Player in Ruminal Metabolism. Microorganisms 11.

61. Gilbert RA, Townsend EM, Crew KS, Hitch TCA, Friedersdorff JCA, Creevey CJ, Pope PB, Ouwerkerk D, Jameson E. 2020. Rumen Virus Populations: Technological Advances Enhancing Current Understanding. Front Microbiol 11:450.

62. Piel D, Bruto M, Labreuche Y, Blanquart F, Goudenège D, Barcia-Cruz R, Chenivesse S, Le Panse S, James A, Dubert J, Petton B, Lieberman E, Wegner KM, Hussain FA, Kauffman KM, Polz MF, Bikard D, Gandon S, Rocha EPC, Le Roux F. 2022. Phage-host coevolution in natural populations. Nat Microbiol 7:1075–1086.

63. Costa AR, van den Berg DF, Esser JQ, Muralidharan A, van den Bossche H, Bonilla BE, van der Steen BA, Haagsma AC, Fluit AC, Nobrega FL, Haas P-J, Brouns SJJ. 2024. Accumulation of defense systems in phage-resistant strains of Pseudomonas aeruginosa. Sci Adv 10:eadj0341.

64. Payne LJ, Meaden S, Mestre MR, Palmer C, Toro N, Fineran PC, Jackson SA. 2022. PADLOC: a web server for the identification of antiviral defence systems in microbial genomes. Nucleic Acids Res 50:W541–W550.

65. Bobay L-M, Ochman H. 2018. Factors driving effective population size and pan-genome evolution in bacteria. BMC Evol Biol 18:153.

66. Douglas GM, Shapiro BJ. 2021. Genic Selection Within Prokaryotic Pangenomes. Genome Biol Evol 13.

67. Al-Shayeb B, Sachdeva R, Chen L-X, Ward F, Munk P, Devoto A, Castelle CJ, Olm MR, Bouma-Gregson K, Amano Y, He C, Méheust R, Brooks B, Thomas A, Lavy A, Matheus-Carnevali P, Sun C, Goltsman DSA, Borton MA, Sharrar A, Jaffe AL, Nelson TC, Kantor R, Keren R, Lane KR, Farag IF, Lei S, Finstad K, Amundson R, Anantharaman K, Zhou J, Probst AJ, Power ME, Tringe SG, Li W-J, Wrighton K, Harrison S, Morowitz M, Relman DA, Doudna JA, Lehours A-C, Warren L, Cate JHD, Santini JM, Banfield JF. 2020. Clades of huge phages from across Earth's ecosystems. Nature 578:425–431.

68. Meaden S, Biswas A, Arkhipova K, Morales SE, Dutilh BE, Westra ER, Fineran PC. 2021. High viral abundance and low diversity are associated with increased CRISPR-Cas prevalence across microbial ecosystems. Curr Biol https://doi.org/10.1016/j.cub.2021.10.038.

69. Yue Y, Huang H, Qi Z, Dou H-M, Liu X-Y, Han T-F, Chen Y, Song X-J, Zhang Y-H, Tu J. 2020. Evaluating metagenomics tools for genome binning with real metagenomic datasets and CAMI datasets. BMC Bioinformatics 21:334.

Figure 1. Prevalence of defence mechanisms in the rumen. A) the percentage of genomes with at least one defence mechanism, B) the number of defence systems per genome, C) the number of unique defence system families per genome in isolates and MAGs. D) The percentage of genomes with at least one defence mechanism, E) the number of defence systems per genome, F) the number of unique defence system families per genome in *Archaea* and *Bacteria*. The defence system density (per genome per kbp) in G) *Archaea* and *Bacteria* and H) isolate and MAGs

Figure 2. Distribution of the total number of antiviral systems per genome and their correlation with genomes size and prophage prevalence. A) the percentage of genomes carrying different number of defence systems, B) correlation between the number of defence system per genome and the genome size and C) correlation between the number of defence system per genome and the number of proviruses per genome.

Figure 3. Abundance and diversity of defence mechanisms within *Bacteria* and *Archaea* in the rumen. A) The relative abundance of all defence systems grouped by families in *Bacteria* and *Archaea* and B) the percentage of genomes per defence system. Defence system families with a relative abundance < 1.5 % were grouped in the “other” category. The defence systems that were found in < 2% of genomes are not shown.

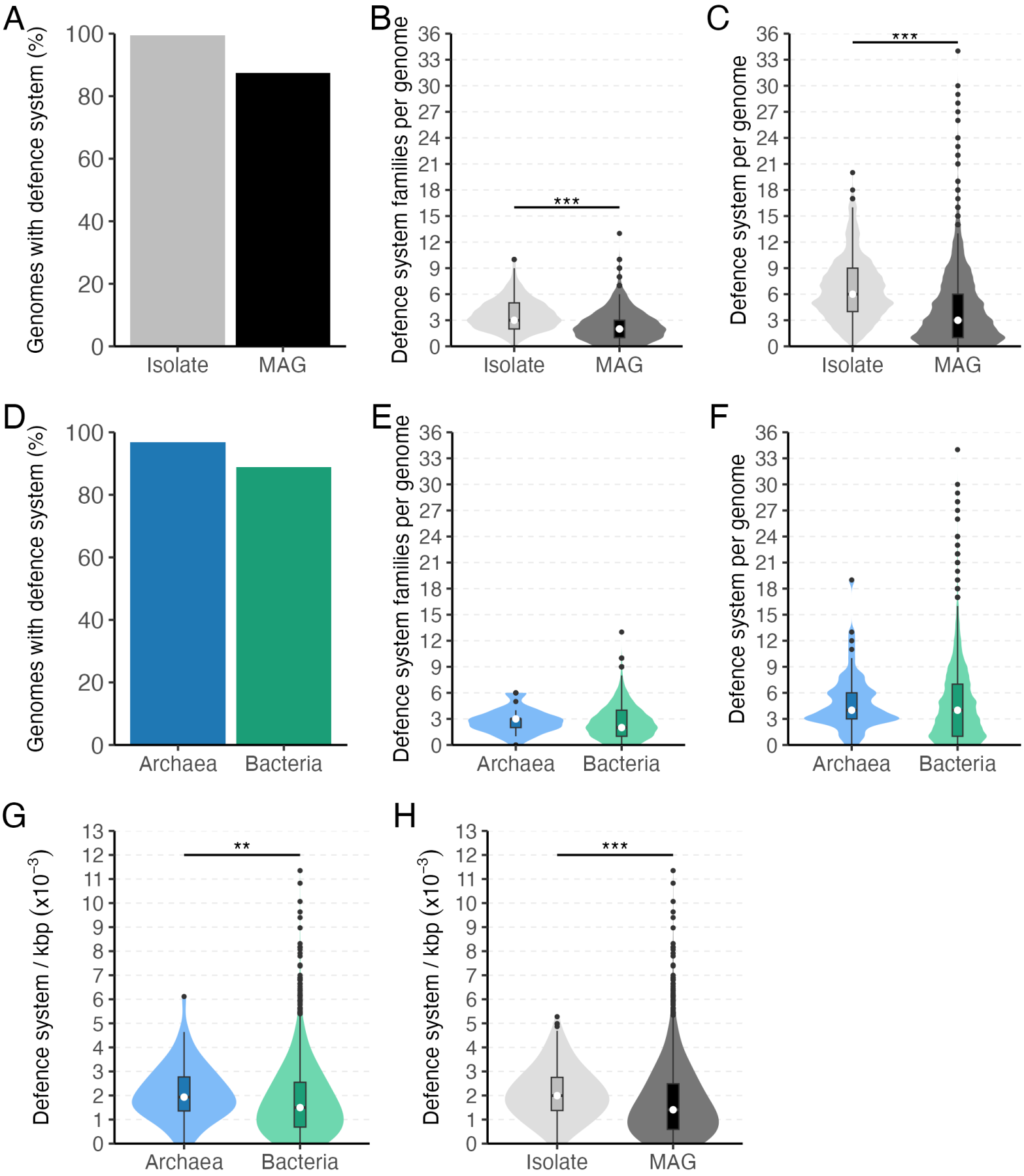
Figure 4. Number of defence systems in different genera from the rumen. A) Number of defence system families, B) number of defence systems and C) defence

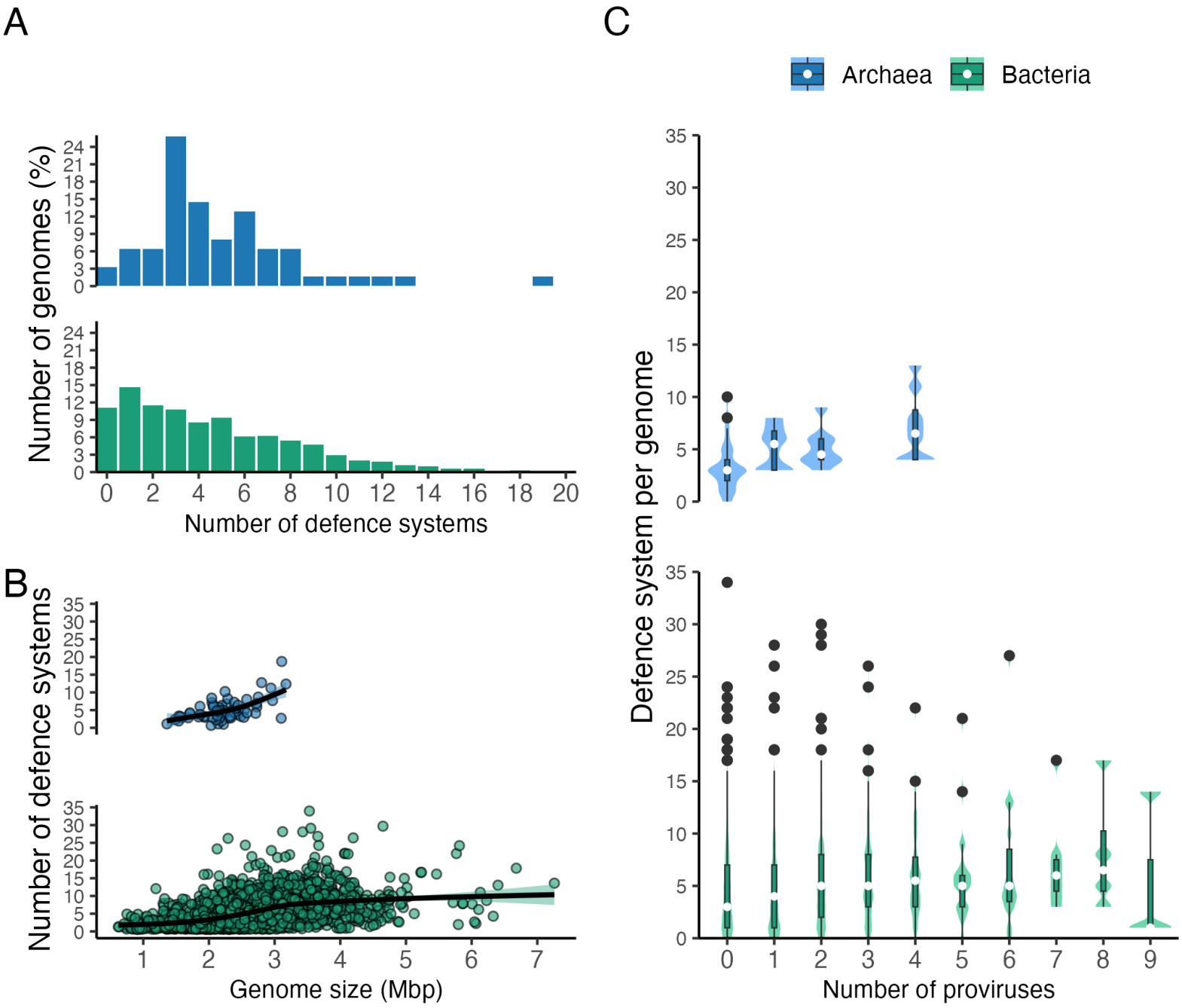
system density (per genome per kbp) found in the top 14 genera with the highest average number of defence systems per genome. The vertical bar represents the average number of defence systems and system families. Genera with at least 10 genomes are shown. The labels in the x-axis represent the Family:Genus.

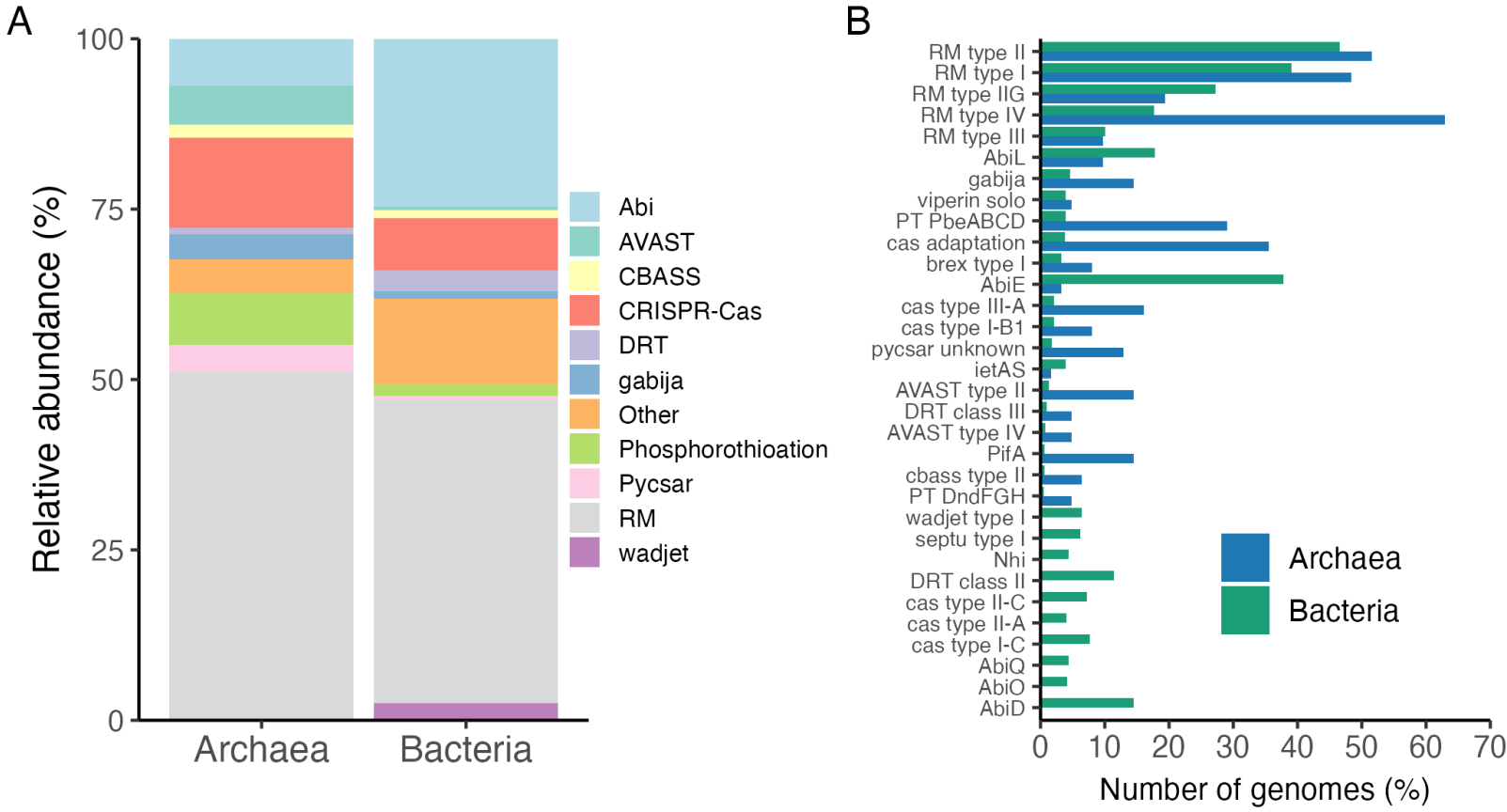
Figure 5. Distribution of defence mechanisms across the rumen. A) Phylogenetic tree of the rumen microbiome (3038 genomes) based on 400 universal markers. b). Phylogenetic tree based on core genes of four members of the rumen microbiome: *Prevotella*, Ca. *Limimorpha*, *Butyrivibrio* and *Ruminococcus*.

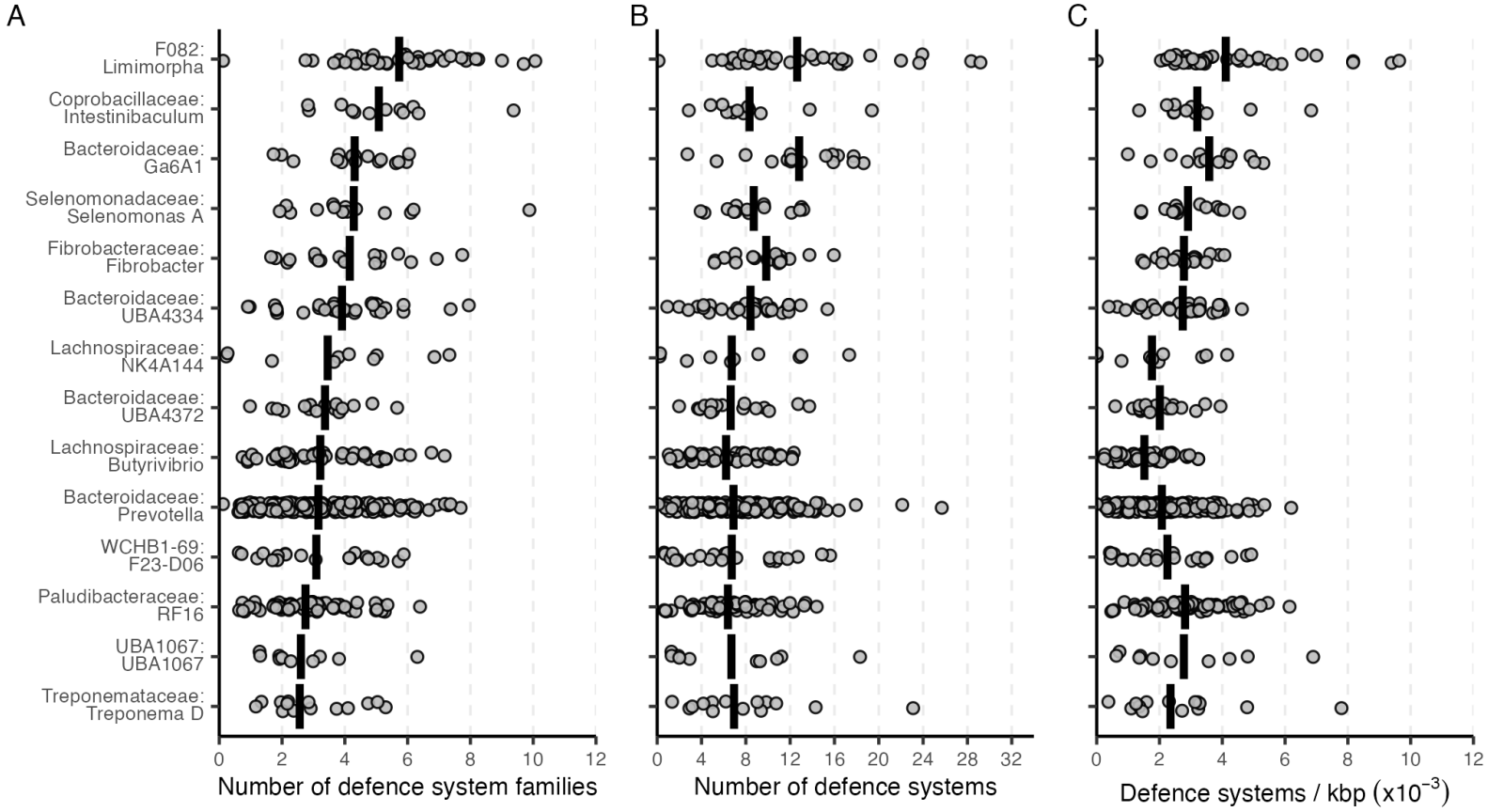
Figure 6. Abundance and diversity of CRISPR-Cas systems within bacteria and archaea in the rumen. A) Percentage of genomes harbouring at least one complete CRISPR-Cas loci, B) percentage of genomes by phylum harbouring at least one complete CRISPR-Cas loci, and C) the relative abundance of CRISPR-Cas system types in each phylum.

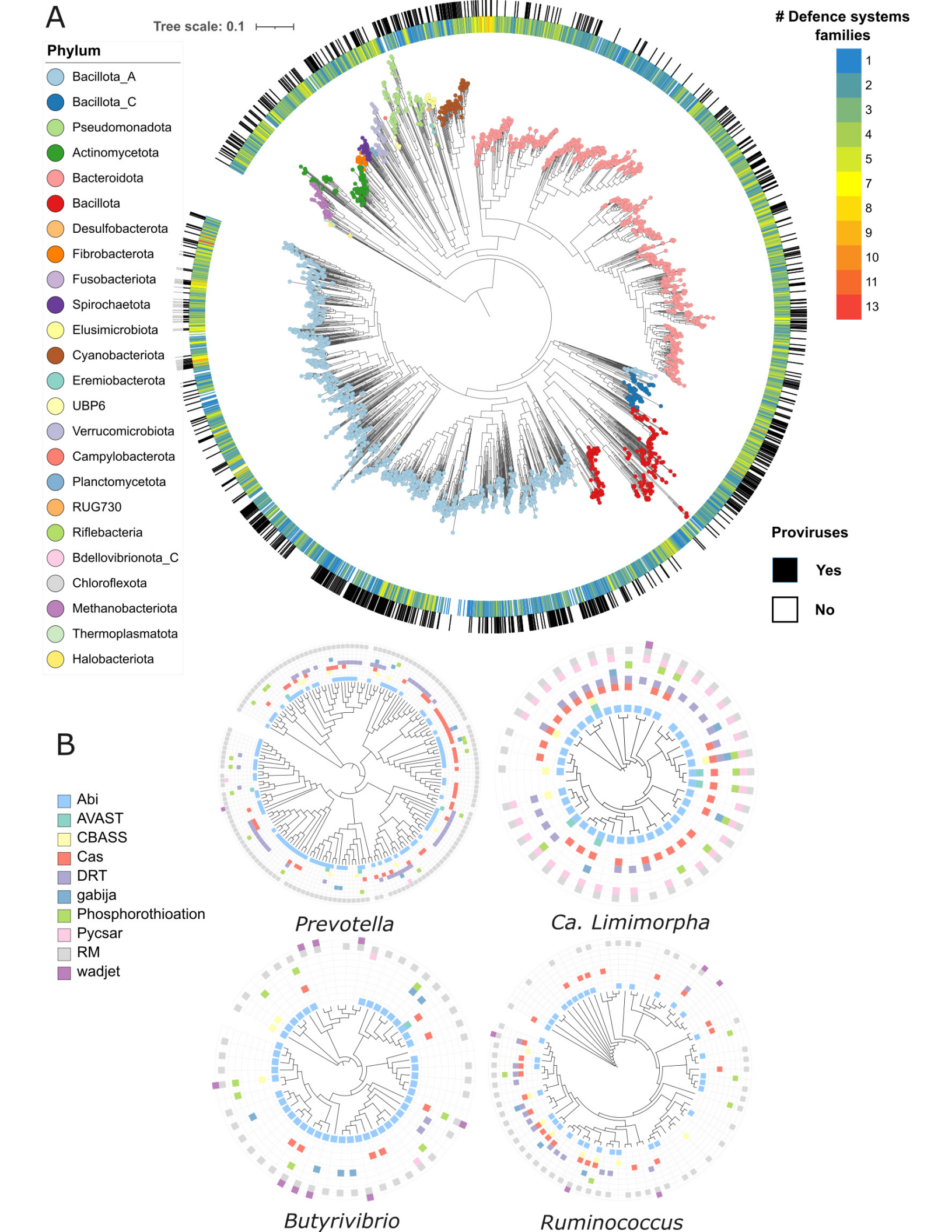
Figure 7. Interaction between rumen phages and their host. The host-virus interaction network shows the match between the identified CRISPR spacers from this study and species-level viral operational taxonomic units (vOTU) obtained from the global rumen virome database (RVD). Grey circles indicate a vOTU while squares represent bacterial/archaeal genomes coloured based on their phylum.

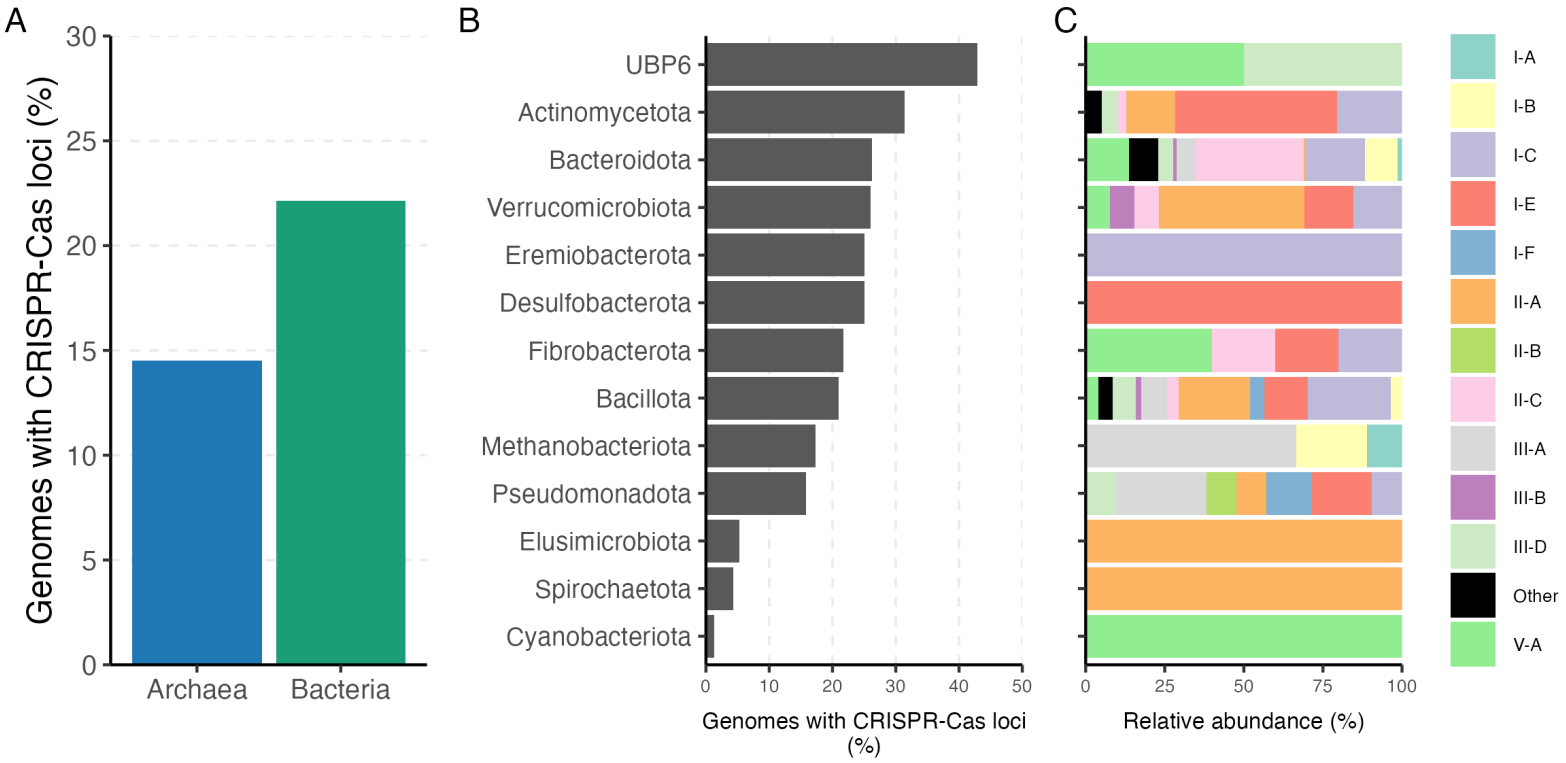

A
Genomes with defence system (%)
100
80
60
40
20
0
Isolate
MAG
B
Defence system families per genome
36
33
30
27
24
21
18
15
12
9
6
3
0
***
Isolate
MAG
C
Defence system per genome
***
Isolate
MAG
D
Genomes with defence system (%)
Archaea
Bacteria
E
Defence system families per genome
Archaea
Bacteria
F
Defence system per genome
Archaea
Bacteria
G
Defence system / kbp (x10^{-3})
13
12
11
10
9
8
7
6
5
4
3
2
1
0
**
Archaea
Bacteria
H
Defence system / kbp (x10^{-3})
***
Isolate
MAG

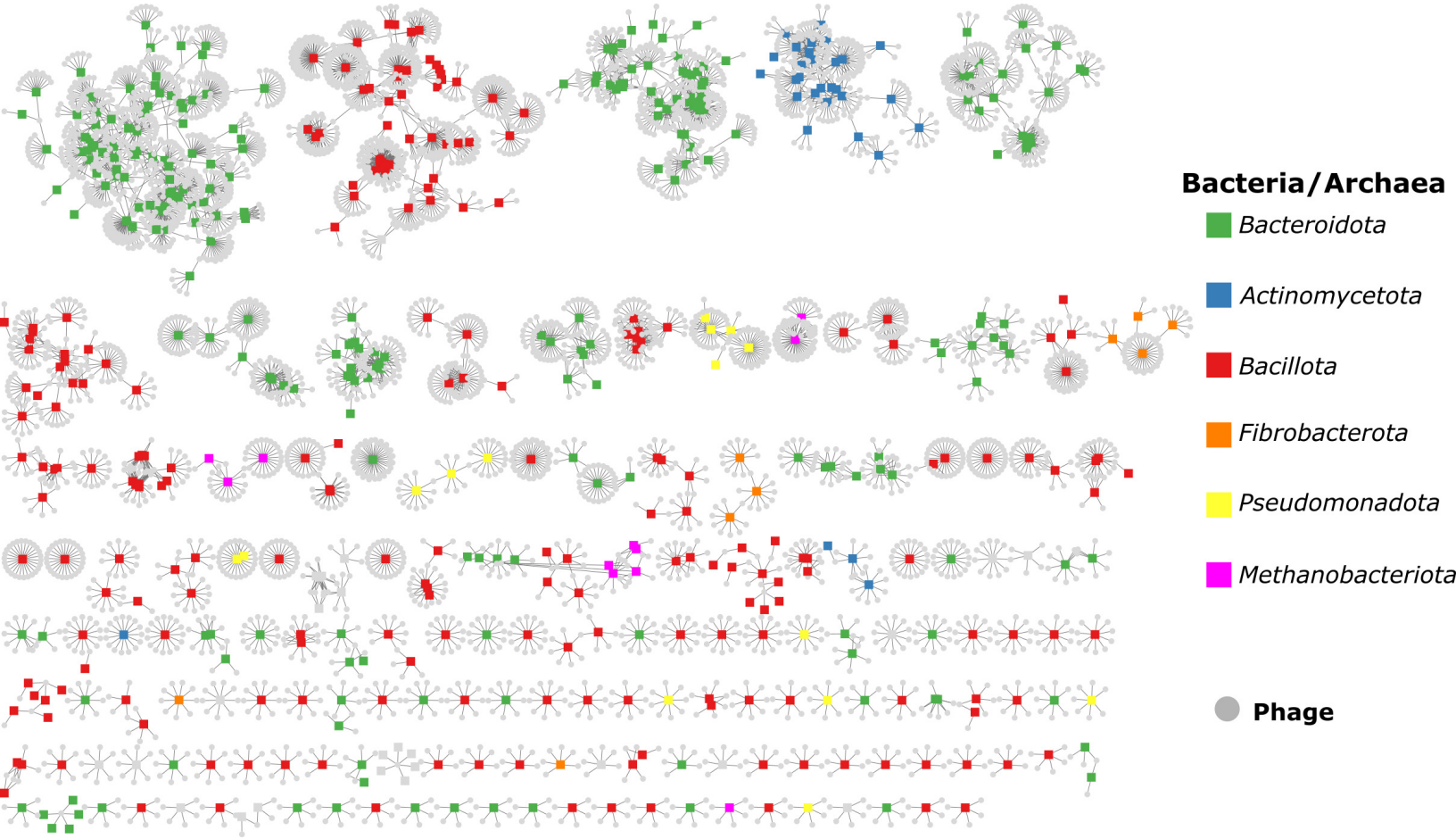

A
Number of genomes (%)
Number of defence systems
B
Number of defence systems
Genome size (Mbp)
C
Archaea
Bacteria
Defence system per genome
Number of proviruses


A
100
75
50
25
0
Relative abundance (%)
Archaea
Bacteria
Abi
AVAST
CBASS
CRISPR-Cas
DRT
gabija
Other
Phosphorothioation
Pycsar
RM
wadjet
B
RM type II
RM type I
RM type IIG
RM type IV
RM type III
AbiL
gabija
viperin solo
PT PbeABCD
cas adaptation
brex type I
AbiE
cas type III-A
cas type I-B1
pycsar unknown
ietAS
AVAST type II
DRT class III
AVAST type IV
PifA
cbass type II
PT DndFGH
wadjet type I
septu type I
Nhi
DRT class II
cas type II-C
cas type II-A
cas type I-C
AbiQ
AbiO
AbiD
Archaea
Bacteria
0
10
20
30
40
50
60
70
Number of genomes (%)


A
B
C
F082:
Limimorpha
Coprobacillaceae:
Intestinibaculum
Bacteroidaceae:
Ga6A1
Selenomonadaceae:
Selenomonas A
Fibrobacteraceae:
Fibrobacter
Bacteroidaceae:
UBA4334
Lachnospiraceae:
NK4A144
Bacteroidaceae:
UBA4372
Lachnospiraceae:
Butyrivibrio
Bacteroidaceae:
Prevotella
WCHB1-69:
F23-D06
Paludibacteraceae:
RF16
UBA1067:
UBA1067
Treponemataceae:
Treponema D
0 2 4 6 8 10 12
Number of defence system families
0 4 8 12 16 20 24 28 32
Number of defence systems
0 2 4 6 8 10 12
Defence systems / kbp (x10^{-3})


A
Tree scale: 0.1
Phylum
Bacillota_A
Bacillota_C
Pseudomonadota
Actinomycetota
Bacteroidota
Bacillota
Desulfobacterota
Fibrobacterota
Fusobacteriota
Spirochaetota
Elusimicrobiota
Cyanobacteriota
Eremiobacterota
UBP6
Verrucomicrobiota
Campylobacterota
Planctomycetota
RUG730
Riflebacteria
Bdellovibrionota_C
Chloroflexota
Methanobacteriota
Thermoplasmatota
Halobacteriota
Defence systems families
1
2
3
4
5
7
8
9
10
11
13
Proviruses
Yes
No
B
Abi
AVAST
CBASS
Cas
DRT
gabija
Phosphorothioation
Pycsar
RM
wadjet
Prevotella
Ca. Limimorpha
Butyrivibrio
Ruminococcus


A
Genomes with CRISPR-Cas loci (%)
30
25
20
15
10
5
0
Archaea
Bacteria
B
UBP6
Actinomycetota
Bacteroidota
Verrucomicrobiota
Eremiobacterota
Desulfobacterota
Fibrobacterota
Bacillota
Methanobacteriota
Pseudomonadota
Elusimicrobiota
Spirochaetota
Cyanobacteriota
0
10
20
30
40
50
Genomes with CRISPR-Cas loci (%)
C
0
25
50
75
100
Relative abundance (%)
I-A
I-B
I-C
I-E
I-F
II-A
II-B
II-C
III-A
III-B
III-D
Other
V-A


Bacteria/Archaea
Bacteroidota
Actinomycetota
Bacillota
Fibrobacterota
Pseudomonadota
Methanobacteriota
Phage