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

Phages play a critical role in the health and function of the gut microbiome, but the competitive mechanisms between the gut microbiome and phages are poorly understood. 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 in 89% of the 3038 genomes. Defence systems in archaeal genomes were slightly more prevalent than in bacterial genomes. In addition, 114 system types were identified and grouped into 49 families, with the restriction modification, Abi and cas system families being the most common. We observed a positive correlation between the number of systems and families, which indicates that *Bacteria* or *Archaea* carrying several defence systems will also harbour more defence families. The genome size correlated positively with the number of defence systems, while the number of proviruses per genome had a low correlation. Complete CRISPR-Cas loci were found in 21.9% of the genomes, with subtype I-C being the most prevalent. On average, the uncultivated genus Ga6A1, *Limimorpha and Fibrobacter* carried the

most number of systems in the rumen. 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.

Importance

Phages and their respective host compete to counter each other. 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 on genomes, from the rumen. While defence mechanism seem to be prevalent among bacteria and archaea, only a few were really common. This suggests that most of these defence mechanisms are not present in many gut microbes but are shared among different members of the microbial community. This supports 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 strongly 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]. However, the study of phages has been quite limited, even though the rumen virome is highly diverse and encodes genes that could modulate microbial methane formation [13, 14]. Because of that the role of phages as potential modulators of the microbiome in the rumen is poorly described.

Bacteriophages shape the bacterial and archaeal populations and are a reservoir of accessory genes that facilitate the mobility of genetic material [15]. The continuous interaction between phages and their respective hosts has resulted in a variety of defence mechanisms, which are more diverse than previously thought [16]. The prevalence of defence mechanisms ranges from zero to several systems between cells. However, the effective immunity seems to be driven by within-cell [17] and pan-genome dynamics [16]. In this way, understanding the immunity of bacteria and archaea as a group of mechanisms could shed light on the interaction between the

phages and host. 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 [18] and African cattle [19].

2.2. Quality control of MAGs

CheckM2 [20] 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 [21] and the GTDB database R207_v2 [22].

2.3. Antiviral defence prediction

Antiviral defence mechanisms were identified from all genomes using PADLOC v1.1 [23] 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). CRISPR-Cas genes and arrays were detected using CRISPRCasTyper v1.6 [24]. All predicted CRISPR spacers from complete or incomplete CRISPR-Cas arrays were kept.

2.4. Prophages detection and identification

Collected genomes were searched for prophages using Virsorter2 v2.2.3 [25]. Only contigs longer than >10000 were used. All contigs identified as Prophages were kept for further analysis. Prophages were quality-checked using checkV [26]. 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. Predicted prophages were matched against CRISPR spacers using Blast [27], only those with >95% identity and <0.0004 e-value were considered spacer to protospacer matches. Network showing the phage host interaction was done using Cytoscape [28].

2.5. Phylogenetic placement

Phylogenetic tree of the whole 3038 genomes was created using Phylophlan [29] and 400 universal marker genes, with --medium diversity and --accurate options. Phylogenetic relation among species of *Butyrivibrio*, *Limimorpha*, *Prevotella* and

Ruminococcus was calculated using a subset of genes present in at least 90% of all genomes (core genes) for each *Genus*. Core genes were extracted, aligned, and concatenated using the tool: concatenated-gene-alignment-fasta from Anvi'o v 7.1 [30]. The extracted genes (55 of *Butirivibrio*, 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 FastTree [31]. Taxonomic information, number of defence families and numbers of prophages were displayed in a phylogenetic tree using iTOL [32].

2.6. Data wrangling and data availability

All data wrangling and statistical analysis were performed in R v4.1.2 [33] and the packages "tidyverse" [34], "data.table", "pheatmap", "ggdist", "patchwork", "ggtext", and "viridis" were used. 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 [35] 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. In terms of taxonomic classification, 62 genomes were classified as *Archaea* by GTDB-TK [21], 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 located in the rumen microbiome high quality genomes were identified using the Prokaryotic Antiviral Defence LOCator (PADLOC) tool [23]. Antiviral defence systems are prevalent in the rumen microbiome. In total, 14,241 systems and 31,948 antiviral related genes were identified across all genomes. Out of the 3038 genomes, 89% of them carry at least one antiviral defence system.

Almost all archaeal genomes harboured a system (96%), compared to 88% of bacterial genomes **(Figure 1A)**. On average, *Archaea* seem to carry 4.9 systems per genome while *Bacteria* 4.7 **(Figure 1C)**. 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 1D)**. 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 (**Figure 1B**). The minimum number of families found was zero and 13 the maximum in the uncultured *Lachnospiraceae* CAG-603. Between the phyla represented by more than 5 genomes, uncultured UBP6 carry the most number of systems on average (11.4), followed by *Fibrobacterota* (8.9), *Firmicutes* group C (7.3), *Verrucomicrobiota* and *Planctomycetota* (6.2).

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. 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. Twenty nine percent of the genomes harboured at least one provirus, being zero the lowest and five the highest number of proviruses. Even though, a positive trend can be observed between the average number of system and between one to three proviruses **(Figure 1F)**, there was a low correlation between the number of proviruses and the number of defence mechanisms

(Spearman, ρ=0.080 p-value = 8.055e-06). 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 1E).**

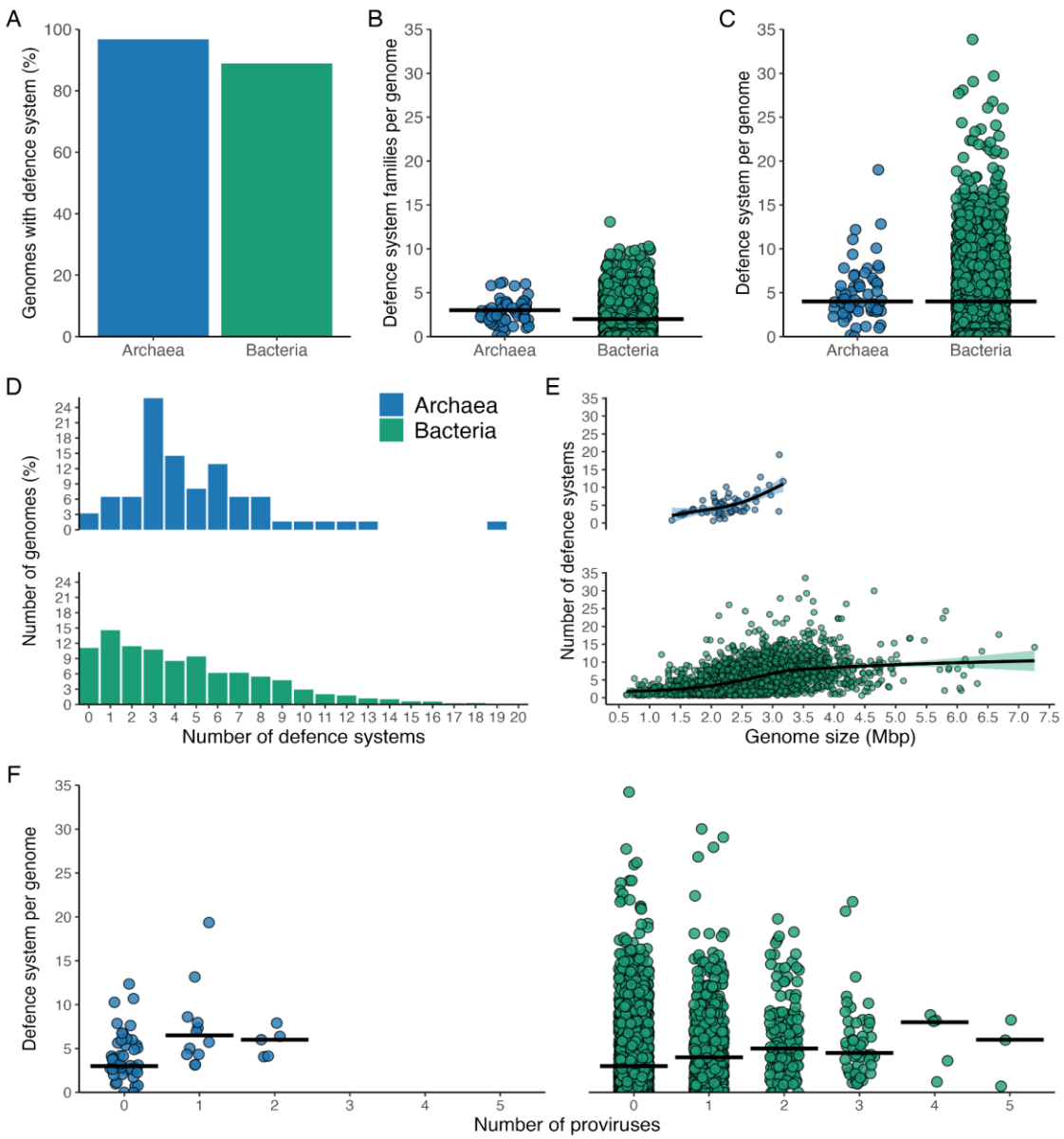

Figure 1. Prevalence of defence mechanisms within bacteria and archaea in the rumen. A) the percentage of genomes with at least one defence mechanism, B)

the number of defence systems per genome, C) the number of unique defence system families per genome, D) the percentage of genomes carrying different number of defence systems, E) correlation between the number of defence system per genome and the genome size and F) correlation between the number of defence system per genome and the number of proviruses per genome. All figures are based on a collection of 3038 genomes. Black horizontal bar indicates the mean.

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 [23]. 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 2A and 2B)**. 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%.

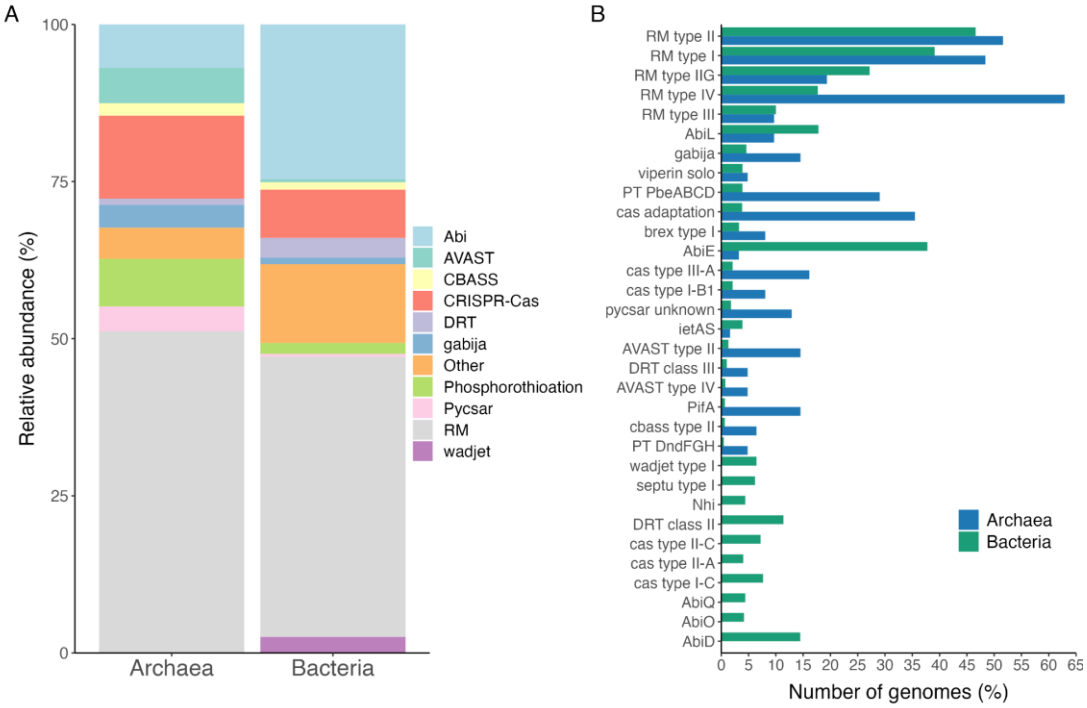

Figure 2. Abundance and diversity of defence mechanisms within *Bacteria* and *Archaea* in the rumen. A) The relative abundance of all defence systems grouped by families in *Bacteria* and *Archaea* and B) the percentage of genomes per defence system. System families with a relative abundance < 1.5 %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

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, the most prevalent genus in the dataset carried 4.6 defence systems and 2.6 defence families. The uncultivated genus Ga6A1 (12.8) and Ca. *Limimorpha*

(12.6) had the highest number of systems, followed by *Fibrobacter* (9.8) **(Figure 3B)**. Similar trends were observed when the number of families were compared **(Figure 3A)**. 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 families (**Figure S2**)

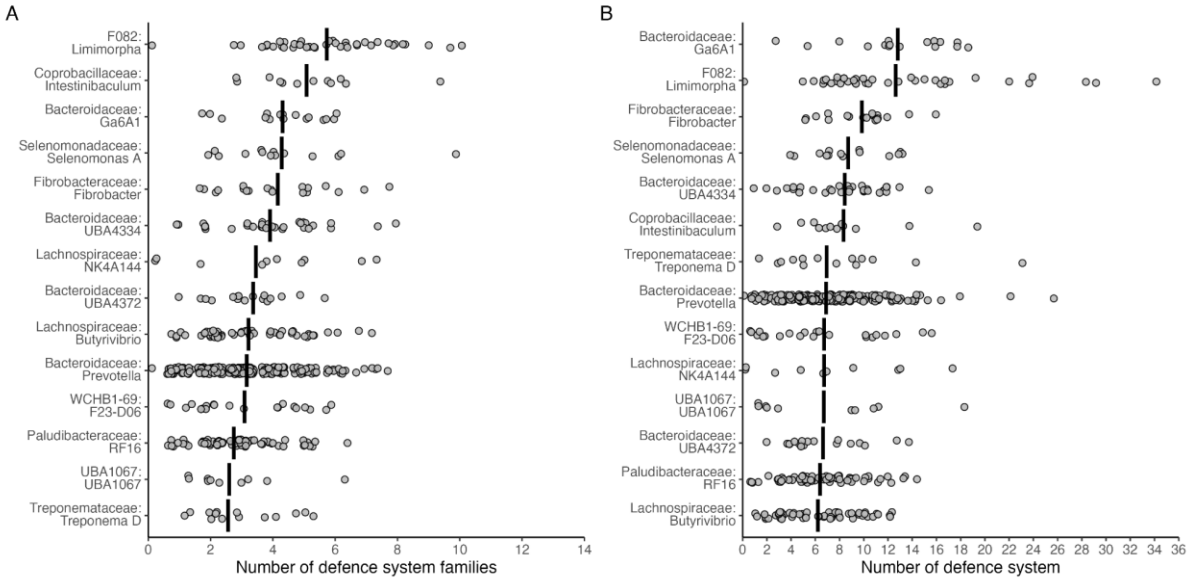

Figure 3. Number of defence systems in different genera from the rumen. A) Number of defence system families and B) number of defence systems. found in the top 14 genera with the highest average number of defence systems per genome. The vertical bar represents the average number of defence systems and system families. Genera with at least 10 genomes are shown. The labels in the x-axis represent the Family:Genus.

Additionally, the diversity of defence system families varies across the dominant members of the rumen. For example, between 37-97 % of genomes of the genus *Prevotella*, Ca. *Limimorpha*, *Butyrivibrio* and *Ruminococcus* carry RM and Abi

systems (**Figure 4b**). However, 22, 31, 24, and 23 different 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 4b**).

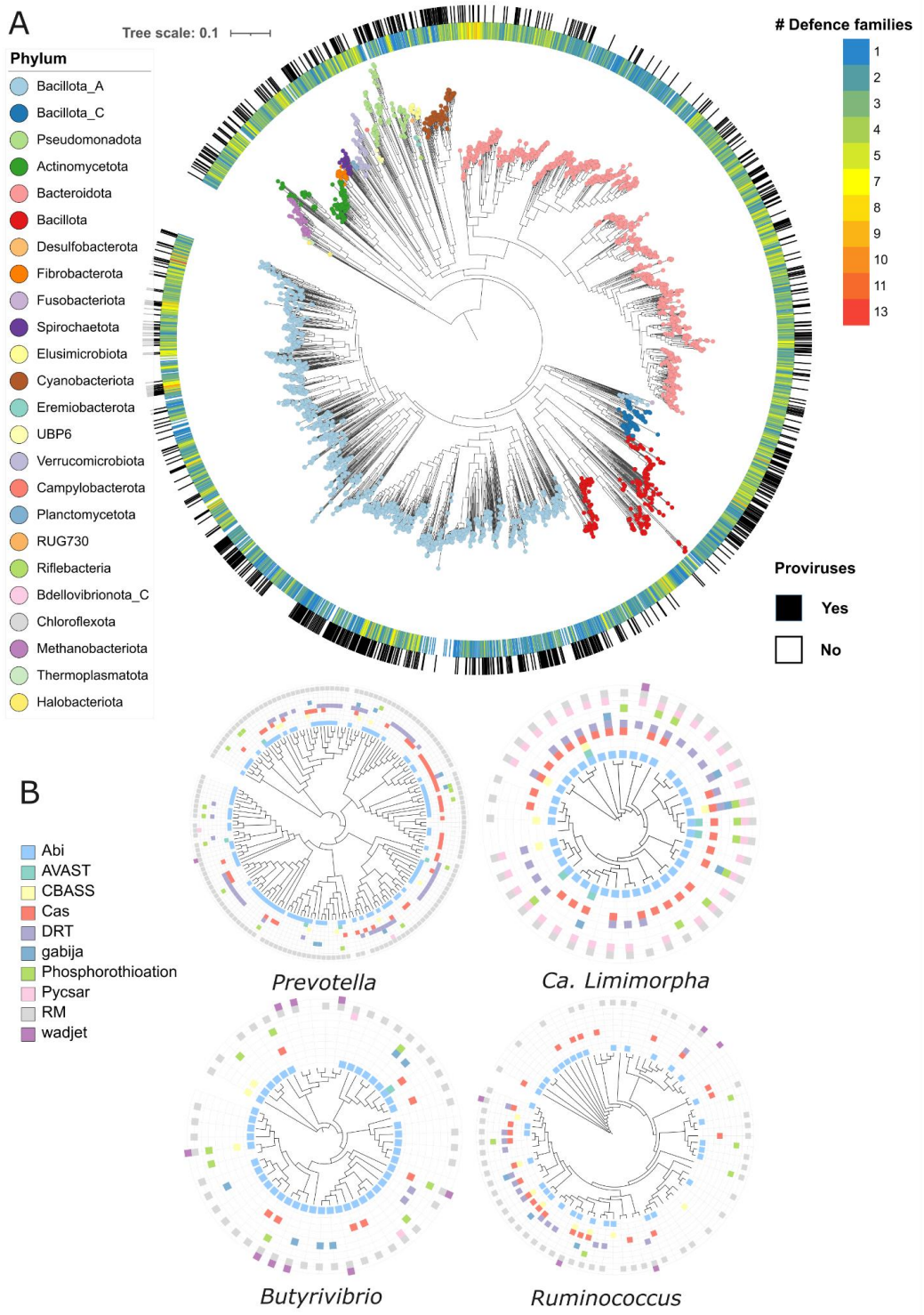

Figure 4. Distribution of defence mechanisms across the rumen. A) Phylogenetic tree of the rumen microbiome (3038 genomes) based on 400 universal markers. b). Phylogenetic tree based on core genes of four members of the rumen microbiome: *Prevotella*, Ca. *Limimorpha*, *Butyrivibrio* and *Ruminococcus*.

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 [36, 37]. 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 [24] 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 5A)**. 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 5B)**. 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 5C)**.

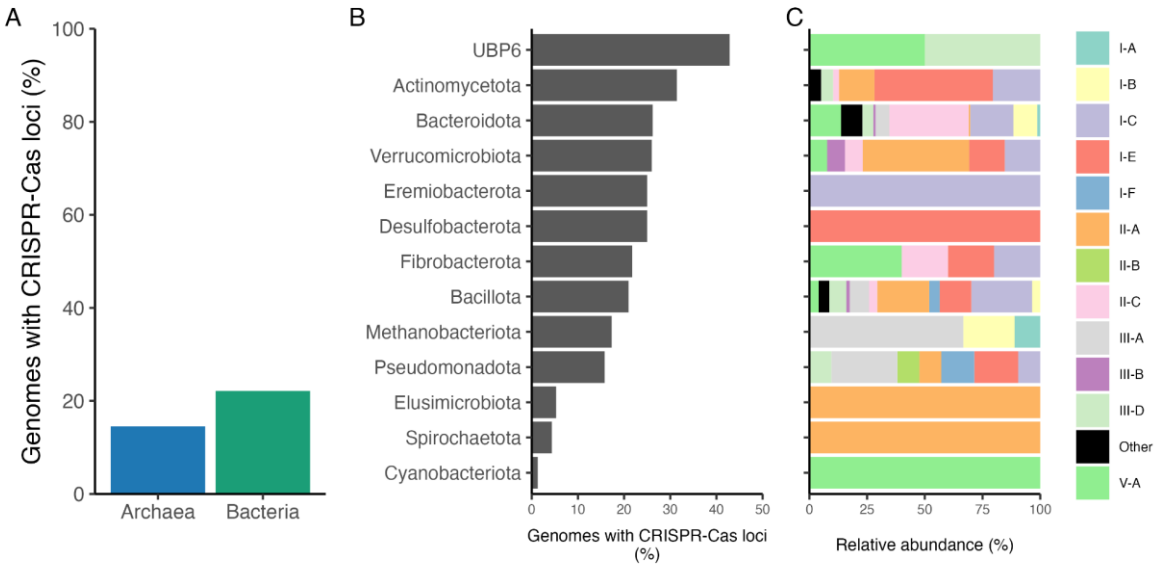

Figure 5. Abundance and diversity of CRISPR-Cas systems within bacteria and archaea in the rumen. A) Percentage of genomes harbouring at least one complete CRISPR-Cas loci, B) percentage of genomes by phylum harbouring at least one complete CRISPR-Cas loci, and C) the relative abundance of CRISPR-Cas system types in each phylum.

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) [13]. The vOTUs were mined from 975 public rumen metagenomes, including samples from 4 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 6**). Most of the interactions were found between members of the phylum *Bacillota*, *Bacteroidota* and *Actinomycetota* and different vOTUs (Figure 5). 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. Similarly, the spacers of a species of the archaeal *Methanobrevibacter* matched with 49 different vOTUs. 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.

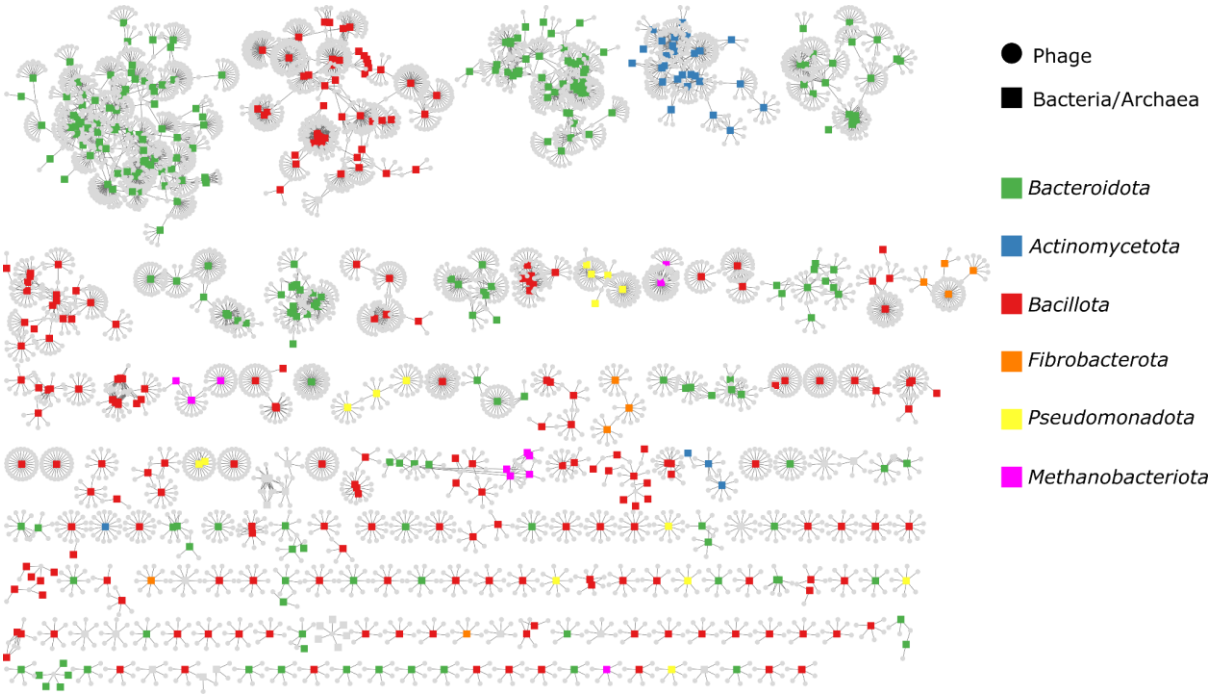

Figure 6. Interaction between rumen phages and their host . The network shows the match between the identified CRISPR spacers from this study and species-level viral operational taxonomic units (vOTU) obtained from the global rumen virome

database (RVD). Round grey circles indicate a vOTU while squares represent bacterial/archaeal genomes coloured based on their phylum.

4. Discussion

It is estimated that phages are 10 times more abundant than bacteria and could be 100 times more abundant in the sea [38, 39]. Despite their abundance, viruses are often overlooked in discussions about biodiversity and ecosystem dynamics. However, their role in shaping microbial communities, influencing genetic diversity, and participating in biogeochemical cycles is increasingly recognized as a crucial aspect of the Earth's ecosystems. Because of that we described the number, abundance, and diversity of antiviral defence mechanisms in the rumen using a collection of metagenomes assembled genomes and isolates.

Our 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 [40] and MAGs from soil and human gut [41]. Additionally, the number of systems per genomes 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 systems. The widespread distribution of these systems suggests a crucial role in maintaining microbial diversity and ecosystem stability.

Also, our findings in the rumen microbiome further reinforce the notion that defence mechanisms are diverse but just a few are dominant in the environment. Consistently, RM, Abi and Cas are the most abundant defence families across different genomes [42]. 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 [40].

Different environmental and genetic factors can contribute to the selection of genes [43, 44]. Previous studies have shown that bacterial and phage genome, and plasmid size seems to drive the acquisition of defence mechanisms [37, 40, 45]. 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. On the other hand, the number of prophages in the genome had a lower correlation with the number of defence systems. We were expecting that a higher number of interactions with phages could drive the acquisition of more systems, but it is not the case in the rumen. However, other studies have found that the abundance of CRISPR-Cas systems can be linked with the viral abundance [46].

The high variability of defence systems and defence families across the genomes from different environments seem to back the ‘pan-immune system’ model [16]. This

model suggests that closely genetically related strains could share defence systems, which indicates that the immunity is driven by the collection of systems in the community. In the rumen, the pan-immune system seems to be driven by the innate response, as CRISPR-Cas is only present in approximately 21% of the rumen genomes. This seems to be common also in other environments like cheese and silage-associated bacterial communities [47, 48]. However, the genomes harbouring CRISPR-Cas systems showed a history of infection with viral particles mined from rumen samples. This shows that even though in low proportion the adaptive immunity can participate in the response of specific infections.

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 supports 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, et al. Livestock and the Environment: What Have We Learned in the Past Decade? Annu Rev Environ Resour. 2015;40:177–202.

2. Hackmann TJ, Spain JN. Invited review: ruminant ecology and evolution: perspectives useful to ruminant livestock research and production. J Dairy Sci. 2010;93:1320–34.

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

4. FAO. 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; 2023.

5. Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, et al. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. Nat Biotechnol. 2018;36:359–67.

6. Ishaq SL, AlZahal O, Walker N, McBride B. 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. 2017;8:1943.

7. Xie F, Jin W, Si H, Yuan Y, Tao Y, Liu J, et al. An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome

of ruminants. Microbiome. 2021;9:137.

8. Tröscher-Mußotter J, Saenz JS, Grindler S, Meyer J, Kononov SU, Mezger B, et al. Microbiome Clusters Disclose Physiologic Variances in Dairy Cows Challenged by Calving and Lipopolysaccharides. mSystems. 2021;6:e0085621.

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

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

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

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

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

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

15. Brown TL, Charity OJ, Adriaenssens EM. Ecological and functional roles of bacteriophages in contrasting environments: marine, terrestrial and human gut. Curr

Opin Microbiol. 2022;70:102229.

16. Bernheim A, Sorek R. The pan-immune system of bacteria: antiviral defence as a community resource. Nat Rev Microbiol. 2020;18:113–9.

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

18. Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. Nat Biotechnol. 2019;37:953–61.

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

20. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.

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

22. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol. 2020;38:1079–86.

23. Payne LJ, Todeschini TC, Wu Y, Perry BJ, Ronson CW, Fineran PC, et al.

Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. Nucleic Acids Res. 2021. https://doi.org/10.1093/nar/gkab883.

24. Russel J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ. CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci. CRISPR J. 2020;3:462–9.

25. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome. 2021;9:37.

26. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021;39:578–85.

27. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

28. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

29. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun. 2020;11:2500.

30. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, et al. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol. 2021;6:3–6.

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

32. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.

33. R Core Team. R: A Language and Environment for Statistical Computing. 2020.

34. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. J Open Source Softw. 2019;4:1686.

35. Gurbich TA, Almeida A, Beracochea M, Burdett T, Burgin J, Cochrane G, et al. MGnify Genomes: A Resource for Biome-specific Microbial Genome Catalogues. J Mol Biol. 2023;435:168016.

36. Burstein D, Sun CL, Brown CT, Sharon I, Anantharaman K, Probst AJ, et al. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. Nat Commun. 2016;7:10613.

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

38. Correa AMS, Howard-Varona C, Coy SR, Buchan A, Sullivan MB, Weitz JS. Revisiting the rules of life for viruses of microorganisms. Nat Rev Microbiol. 2021;19:501–13.

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

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

41. Beavogui A, Lacroix A, Wiart N, Poulain J, Delmont TO, Paoli L, et al. The defensome of complex bacterial communities. bioRxiv. 2023. https://doi.org/10.1101/2023.08.12.553040;

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

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

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

45. Al-Shayeb B, Sachdeva R, Chen L-X, Ward F, Munk P, Devoto A, et al. Clades of huge phages from across Earth's ecosystems. Nature. 2020;578:425–31.

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

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

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

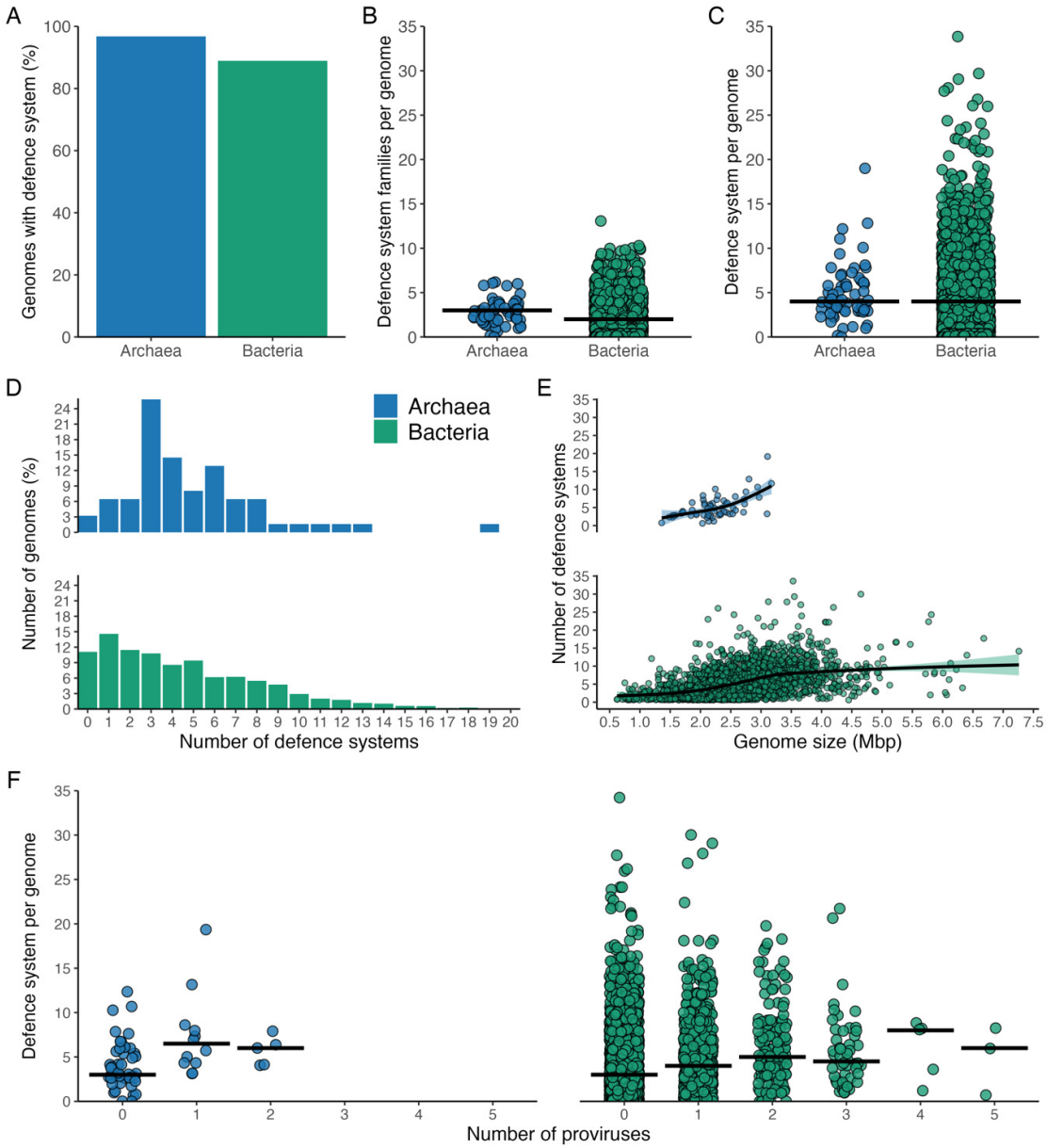

Figure 1. Prevalence of defence mechanisms within bacteria and archaea in the rumen. A) the percentage of genomes with at least one defence mechanism, B) the number of defence systems per genome, C) the number of unique defence system families per genome, D) the percentage of genomes carrying different number of defence systems, E) correlation between the number of defence system per genome and the genome size and F) correlation between the number of defence system per

genome and the number of proviruses per genome. All figures are based on a collection of 3038 genomes. Black horizontal bar indicates the mean.

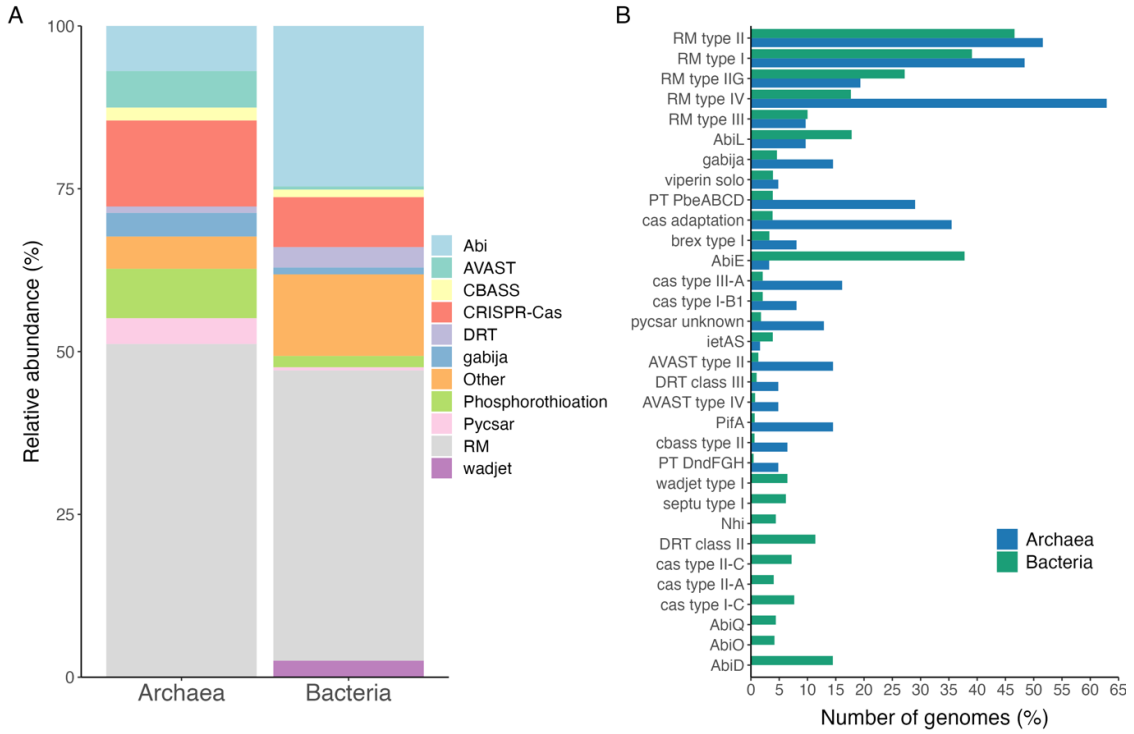

Figure 2. Abundance and diversity of defence mechanisms within *Bacteria* and *Archaea* in the rumen. A) The relative abundance of all defence systems grouped by families in *Bacteria* and *Archaea* and B) the percentage of genomes per defence system. System families with a relative abundance < 1.5 %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

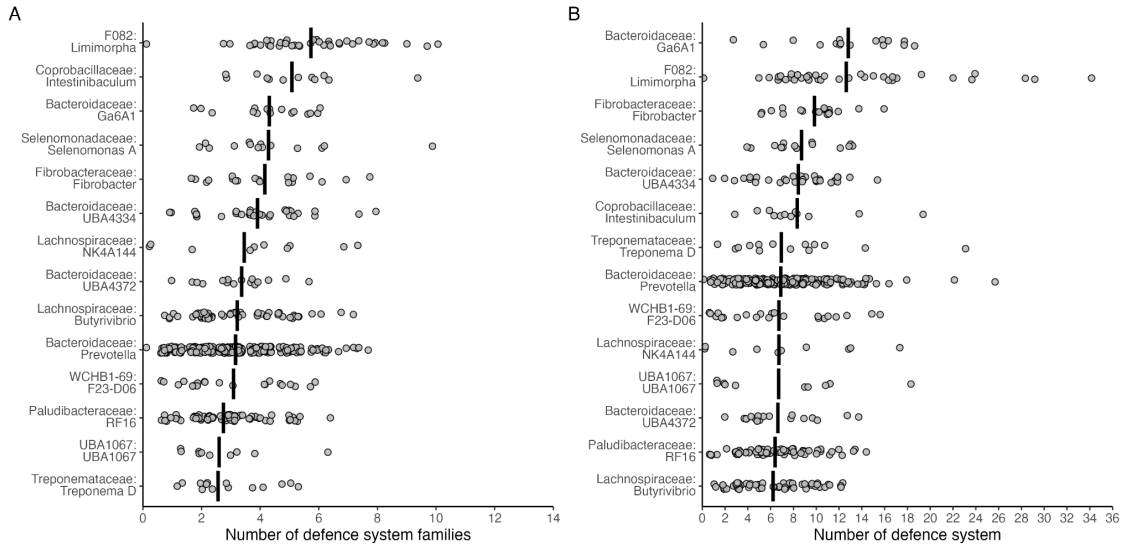

Figure 3. Number of defence systems in different genera from the rumen. A) Number of defence system families and B) number of defence systems. found in the top 14 genera with the highest average number of defence systems per genome. The vertical bar represents the average number of defence systems and system families. Genera with at least 10 genomes are shown. The labels in the x-axis represent the Family:Genus.

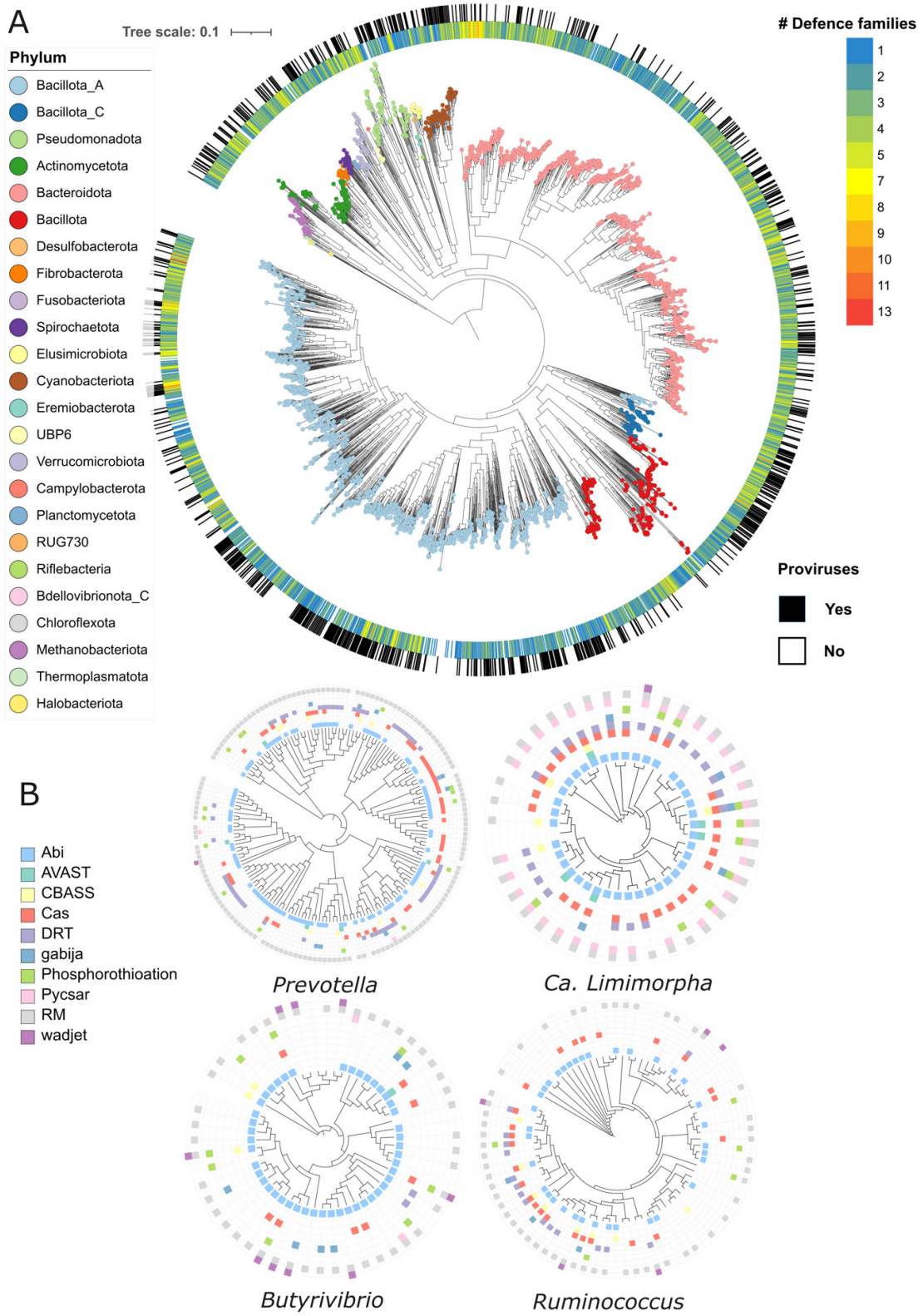
A
Tree scale: 0.1
Phylum
Bacillota_A
Bacillota_C
Pseudomonadota
Actinomycetota
Bacteroidota
Bacillota
Desulfobacterota
Fibrobacterota
Fusobacteriota
Spirochaetota
Elusimicrobiota
Cyanobacteriota
Eremiobacterota
UBP6
Verrucomicrobiota
Campylobacterota
Planctomycetota
RUG730
Riflebacteria
Bdellovibrionota_C
Chloroflexota
Methanobacteriota
Thermoplasmatota
Halobacteriota
Defence 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

Figure 4. 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*.

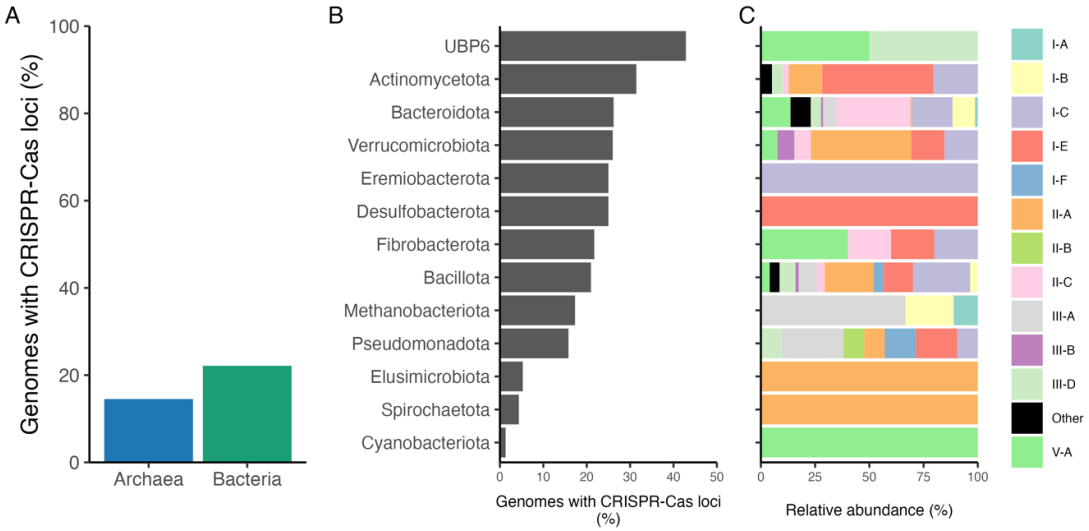

Figure 5. Abundance and diversity of CRISPR-Cas systems within bacteria and archaea in the rumen. A) Percentage of genomes harbouring at least one complete CRISPR-Cas loci, B) percentage of genomes by phylum harbouring at least one complete CRISPR-Cas loci, and C) the relative abundance of CRISPR-Cas system types in each phylum.

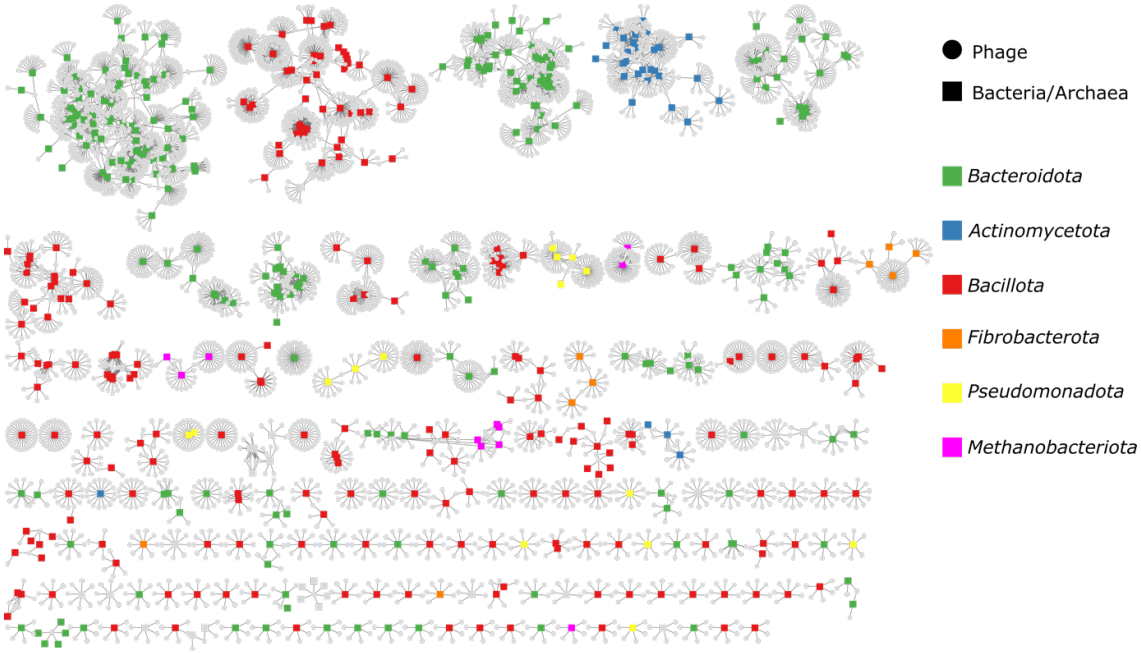

Figure 6. Interaction between rumen phages and their host . The network shows the match between the identified CRISPR spacers from this study and species-level viral operational taxonomic units (vOTU) obtained from the global rumen virome database (RVD). Round grey circles indicate a vOTU while squares represent bacterial/archaeal genomes coloured based on their phylum.