MICROBIAL GENOMICS

RESEARCH ARTICLE

Cuscó *et al.*, *Microbial Genomics* 2022;8:000802

DOI 10.1099/mgen.0.000802

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Novel canine high-quality metagenome-assembled genomes, prophages and host-associated plasmids provided by long-read metagenomics together with Hi-C proximity ligation

Anna Cuscó[1,2,*], Daniel Pérez[3], Joaquim Viñes[1,3], Norma Fàbregas[1] and Olga Francino[3]

Abstract

The human gut microbiome has been extensively studied, yet the canine gut microbiome is still largely unknown. The availability of high-quality genomes is essential in the fields of veterinary medicine and nutrition to unravel the biological role of key microbial members in the canine gut environment. Our aim was to evaluate nanopore long-read metagenomics and Hi-C (high-throughput chromosome conformation capture) proximity ligation to provide high-quality metagenome-assembled genomes (HQ MAGs) of the canine gut environment. By combining nanopore long-read metagenomics and Hi-C proximity ligation, we retrieved 27 HQ MAGs and 7 medium-quality MAGs of a faecal sample of a healthy dog. Canine MAGs (CanMAGs) improved genome contiguity of representatives from the animal and human MAG catalogues – short-read MAGs from public datasets – for the species they represented: they were more contiguous with complete ribosomal operons and at least 18 canonical tRNAs. Both canine-specific bacterial species and gut generalists inhabit the dog's gastrointestinal environment. Most of them belonged to *Firmicutes*, followed by *Bacteroidota* and *Proteobacteria*. We also assembled one *Actinobacteriota* and one *Fusobacteriota* MAG. CanMAGs harboured antimicrobial-resistance genes (ARGs) and prophages and were linked to plasmids. ARGs conferring resistance to tetracycline were most predominant within CanMAGs, followed by lincosamide and macrolide ones. At the functional level, carbohydrate transport and metabolism was the most variable within the CanMAGs, and mobilome function was abundant in some MAGs. Specifically, we assigned the mobilome functions and the associated mobile genetic elements to the bacterial host. The CanMAGs harboured 50 bacteriophages, providing novel bacterial-host information for eight viral clusters, and Hi-C proximity ligation data linked the six potential plasmids to their bacterial host. Long-read metagenomics and Hi-C proximity ligation are likely to become a comprehensive approach to HQ MAG discovery and assignment of extra-chromosomal elements to their bacterial host. This will provide essential information for studying the canine gut microbiome in veterinary medicine and animal nutrition.

DATA SUMMARY

An overview of the scripts used to analyse the data is available as Supplementary material (available with the online version of this article). The final CanMAGs (canine metagenome-assembled genomes) are available on Zenodo: **https://doi.org/10.5281/zenodo.5055248.** The raw fast5 files are available from the ENA (European Nucleotide Archive) under BioProject accession no. PRJEB42270.

Received 17 August 2021; Accepted 21 February 2022; Published 17 March 2022

Author affiliations: [1]Vetgenomics, Edificio Eureka, Parc de Recerca UAB, Barcelona, Spain; [2]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, PR China; [3]Molecular Genetics Veterinary Service (SVGM), Veterinary School, Universitat Autònoma de Barcelona, Barcelona, Spain.

***Correspondence:** Anna Cuscó, anna.cusco@vetgenomics.com

Keywords: canine metagenome; gut microbiome; Hi-C proximity ligation; long-read metagenomics; metagenome-assembled genomes; nanopore.

Abbreviations: ANI, average nucleotide identity; ARG, antimicrobial-resistance gene; CanMAG, canine metagenome-assembled genome; COG, Clusters of Orthologous Genes; GPD, Gut Phage Database; GTDB, Genome Taxonomy Database; Hi-C, high-throughput chromosome conformation capture; HQ MAG, high-quality metagenome-assembled genome; IS, insertion sequence; MAG, metagenome-assembled genome; MGE, mobile genetic element; MIMAG, Minimum Information about a Metagenome-Assembled Genome; MQ MAG, medium-quality metagenome-assembled genome; VC, viral cluster; WGS, whole-genome sequencing.

The final CanMAGs are available on Zenodo: https://doi.org/10.5281/zenodo.5055248. The raw fast5 files are available from the ENA (European Nucleotide Archive) under BioProject accession no. PRJEB42270.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Supplementary material is available with the online version of this article.

Impact Statement

Retrieval of high-quality genomes from metagenomes is a step towards creating niche-specific databases, which are needed in host-associated environments to understand the microbiome composition and functional capacity in health and disease, and better assess the impact of microbiome modulation strategies. We combined long-read nanopore metagenomics and Hi-C (high-throughput chromosome conformation capture) proximity ligation data as a proof-of-concept to retrieve metagenome-assembled genomes (MAGs) from the canine gut. Long-read metagenomics retrieves long contigs harbouring complete assembled ribosomal operons, antibiotic-resistance genes, prophages and other mobile genetic elements. Hi-C allowed the binning of the long contigs into high-quality MAGs (HQ MAGs) and medium-quality MAGs, some of them representing closely related species. Moreover, Hi-C also linked plasmids to their bacterial host. HQ MAGs improve the short-read MAGs of public datasets. Long-read metagenomics combined with proximity ligation binning is likely to become a comprehensive approach for the discovery of MAGs, which are essential to unravel the biological role of microbial members in multiple environments, such as the canine gut.

INTRODUCTION

The human gut microbiome has been extensively studied using metagenomics, and large catalogues of metagenome-assembled genomes (MAGs) are available to represent genomes of uncultured bacteria [1–3]. These MAG collections are used as references to assess differences between diseased and healthy states, as well as the effects of the diet or other environmental factors. The most recent human gut catalogue contains a total of 204 938 reference genomes, yet only 38 of them are high-quality MAGs (HQ MAGs) regarding the MIMAG (Minimum Information about a Metagenome-Assembled Genome) criteria [3].

The quality of the retrieved MAGs is assessed following the MIMAG standard criteria [4]. HQ MAGs are more comparable to complete genomes, and harbour key biological pieces such as complete rRNA and tRNA genes, as well as the mobile genetic elements (MGEs) and prophages that help understanding biological processes like horizontal gene transfer events. Most of the large-scale metagenomics studies rely on the use of short-read sequencing technologies. However, short-read-derived MAGs are usually fragmented and lack ribosomal gene sequences. Since these genes are repeated and highly conserved, short-read metagenomics collapses them together and cannot locate them in their respective bacterial genome [5].

The fields of veterinary medicine and nutrition have an increasing interest in the composition and function of the canine gut microbiome [6, 7]. Studies on this microbiome have generally been limited to 16S rRNA amplicon sequencing. Thus, they provided taxonomic and compositional information at the family or genus level, but no functional or antimicrobial-resistance information. To date, only one comprehensive metagenomics study is available [8], in which none of the 1525 released MAGs fulfils the high-quality MIMAG criteria [9]. However, high-quality genomes are essential to better understand the microbiome composition and functional capability in canine health and disease, and the impact of microbiome modulation strategies such as dietary interventions and pre- and probiotic supplementation. Recently, we tested long-read metagenomics for a canine faecal sample and retrieved eight single-contig HQ MAGs [10].

Long-read metagenomics uses long DNA stretches, solving many issues derived from short-read MAGs. Long-read sequencing spans complete ribosomal genes and their genomic context, bridging together microbiome insights obtained by short-read MAGs and 16S rRNA sequencing surveys [11]. In addition, it spans complete MGEs such as prophages or plasmids, which can harbour antimicrobial-resistance genes (ARGs) or virulence factors [12–16]. Sequencing full-length MGEs and locating them correctly in the chromosome or plasmid can unravel horizontal gene transfer events or the pathogenic potential of a specific micro-organism [17].

However, long-read sequencing needs to overcome two main issues: extracting long DNA fragments and reducing the sequencing error rate. For the first, high-molecular-weight DNA extractions suited for sample type work efficiently producing long reads, as previously demonstrated for faecal samples [18]. For the second, the higher error rate when compared to other technologies can be significantly reduced by deep sequencing [19] and by using error-specific correction software, such as frameshift-aware software for nanopore sequencing [20].

To further disentangle complex microbiomes, metagenomics can be complemented with high-throughput chromosome conformation capture (Hi-C) proximity ligation data. Hi-C proximity ligation cross-links DNA *in vivo* within intact cells to capture interactions between DNA molecules in close physical proximity [21, 22]. This approach further improves the contiguity of a metagenome assembly, and captures interactions between plasmids or viruses and their host genomes. To date, only two studies have combined long-read metagenomics with Hi-C proximity ligation data: in a cow rumen, to link viruses and ARGs to their microbial host [23]; and in a sheep gut, to generate 'lineage-resolved' MAGs [24].

In this context, our main objective was to evaluate nanopore long-read metagenomics and Hi-C proximity ligation to provide HQ MAGs as representatives of the canine gut environment. We retrieved 27 HQ MAGs and 7 MQ MAGs harbouring complete ribosomal genes, MGEs and prophages as well as ARGs from a single sample, increasing the number of our previous canine MAGs (CanMAGs) [10], and capturing new interactions between plasmids and their bacterial host genomes. More specifically, the high-quality genomes and the unique genomic information that we provide in this study will be key for future functional analysis of the canine gut microbiome.

METHODS

Long-read metagenomics: DNA extraction and nanopore sequencing

Our study focuses on the microbiome analysis of a single faecal sample of a healthy dog. Using the same faecal sample, we extracted high-molecular-weight DNA with a Quick-DNA HMW MagBead kit (Zymo Research) and non-high-molecular-weight DNA with a DNA miniprep kit (Zymo Research). We prepared a sequencing library for each DNA extraction using the Ligation Sequencing kit 1D (SQK-LSK109; Oxford Nanopore Technologies) and sequenced each of them in a flowcell R9.4.1 using MinION (Oxford Nanopore Technologies). After the two nanopore runs, we obtained a total of 16.94 million reads (36.05 Gb). Further details have been described previously [10].

Hi-C metagenome cross-linking and Illumina sequencing

The same faecal sample was used to generate a Hi-C library using the ProxiMeta Hi-C kit following the manufacturer's protocol (Phase Genomics). The Hi-C method cross-links DNA molecules that are in close physical proximity within intact cells. Hi-C libraries were sequenced on an Illumina HiSeq 4000 platform, generating 75 bp paired-end reads. The Proximeta Hi-C library produced 75.01 million paired-end reads (11.40 Gb).

Metagenome assembly and deconvolution

Raw fast5 files from nanopore sequencing were basecalled using Guppy 3.4.5 (Oxford Nanopore Technologies) with high accuracy basecalling mode (dna_r9.4.1_450bps_hac.cfg). During the basecalling, the reads with an accuracy lower than seven were discarded. Before proceeding with the metagenomics assembly, we performed an error-correction step of the raw nanopore reads using canu 2.0 [25]. We merged the data from the two nanopore runs and performed the metagenome assembly with Flye 2.7 [26] (options: --nano-corr --meta, --genome-size 500 m, --plasmids). We polished the metagenome assembly with one round of medaka 1.0.1 (https://github.com/nanoporetech/medaka), including all the raw nanopore fastq files as input. We uploaded the metagenome assembly and the raw Hi-C sequencing data to the ProxiMeta cloud-based pipeline (https://proximeta.phasegenomics.com/; Phase Genomics; December 2020) [21], where it was processed, and the final metagenomic bins were retrieved.

Characterization of the HQ MAGs and MQ MAGs

We further corrected the metagenomic bins by correcting the frameshift errors, as described elsewhere [20], using Diamond 0.9.32 [27] and MEGAN-LR 6.19.1 [28]. We classified the MAGs considering MIMAG criteria [4] as a HQ MAG, when it is >90% complete, and presents <5% contamination, rRNAs genes and tRNAs; and a MQ MAG, when it is >50% complete and presents <10% contamination.

To assess the novelty and the taxonomy of the metagenomic bins, we used GTDB-tk 1.3.0 [29] with GTDB (Genome Taxonomy Database) taxonomy release 95 [30]. FastANI 1.3 [31] was used to determine the average nucleotide identity (ANI) between related genomes.

We used Prokka 1.13.4 [32] to annotate the genomes and assess the number of coding sequences, ribosomal genes and tRNAs of the MAGs. Since the ribosomal genes are together within the *rrn* operon, when the number of 16S rRNAs, 23S rRNAs and 5S rRNAs was not the same within a MAG, we double-checked their presence using the RNAmmer 1.2 [33] server.

We compared the HQ MAGs obtained to previously reported MAGs from the most extensive and recent gastrointestinal collections: (i) the animal gut metagenome [9], which includes MAGs from the dog gut catalogue [8], and (ii) the Unified Human Gastrointestinal Genome (UHGG) [3]. We retrieved MAGs representing the same species as our HQ MAGs by keeping: (i) those with >95% of ANI [31] for the animal gut metagenome; and (ii) those with the equivalent species-level taxonomy as stated by GTDB-tk for the UHGG. A detailed overview of the bioinformatics process is provided in the Supplementary code.

Plasmid analysis

We assessed the metagenomic bins representing HQ MAGs and MQ MAGs with <5% contamination for any putative plasmids. The putative plasmids within the HQ MAGs and MQ MAGs were predicted using Plasflow 1.1.0 [34]. They were further annotated with Prokka 1.14.6 [35] to identify plasmid-associated genes, and with Abricate 0.8.13 (https://github.com/tseemann/

abricate) to identify potential ARGs with CARD (Comprehensive Antibiotic Resistance Database) [32] or virulence factors with VFDB (Virulence Factor Database) [36]. We further inspected the putative plasmids by assessing: (i) BLAST results against the nr/nt NCBI (National Center for Biotechnology Information) database; (ii) their relative coverage when compared to the associated bacterial host (from Flye 2.7 [26] output); (iii) their circularity (from Flye 2.7 output); and (iv) their annotation with Prokka [35].

Bacteriophage analysis

VirSorter2 2.1 [37] and Vibrant 1.2.1 [38] were used to detect viruses within the HQ MAGs and MQ MAGs. CheckV 0.7.0 (https://bitbucket.org/berkeleylab/checkv/) was used to assess the quality of single-contig viral genomes and remove potential host contamination within integrated viruses. If VirSorter2 and Vibrant redundantly detected a viral signal, we kept the one with the highest quality and completeness. We used vConTACT2 0.9.19 [39] to cluster viral sequences and provide taxonomic context. The results reported here are from high-quality and medium-quality predicted viruses. Low-quality predicted viruses were not included.

To perform vConTACT2, we used a subset of the Gut Phage Database (GPD) [40]. To create this subset, we mapped our predicted bacteriophages to the whole GPD (n=142809) using Minimap2 2.17 [41]. The GPD viral genomes that mapped with our predicted bacteriophages (n=682) and our predicted bacteriophages were included as input sequences into vConTACT2. Then, we predicted the proteins using Prodigal 2.6.3 [42] and ran vConTACT2 against its ProkaryoticViralRefSeq201-merged database. The resulting network was visualized using Cytoscape 3.8.2 [43]. We named these bacteriophages, regarding their CanMAG bacterial host, as follows: BPX-CanMAG_XX.

RESULTS

Metagenome characterization

We characterized the faecal metagenome of a healthy dog by combining a nanopore long-read metagenomics assembly with Hi-C proximity ligation data, and retrieved a total of 27 HQ MAGs and 7 MQ MAGs, according to the MIMAG criteria. We named the 34 MAGs described in this study as CanMAGs. The long reads provided long contigs that harboured non-collapsed repetitive regions, complete ribosomal genes, ARGs and MGEs. Hi-C data allowed the binning of these long contigs to create HQ MAGs and MQ MAGs. Five out of the 34 CanMAGs were single-contig genome assemblies, so they needed no Hi-C data for binning. The long contigs harboured 50 prophages, and the Hi-C data linked the 6 plasmids to their bacterial host. We did not describe free viral particles.

CanMAGs harboured ribosomal genes and improved genome contiguity of representatives from the animal and human MAG catalogues

We retrieved 34 CanMAGs from the faeces of a healthy dog, which were – according to the MIMAG criteria [3]: 27 high-quality with >90% completeness and <5% contamination, and presence of ribosomal genes and at least 18 canonical tRNAs; and 7 medium-quality with >50% completeness and <10% contamination. The frameshift correction step [20] applied to the initial genomic bins reduced insertion and deletion errors – the most common error type in nanopore sequencing – of the CanMAGs (Table S1). After this extra correction step, the completeness was either increased or maintained, transforming five MQ MAGs to HQ MAGs. Twenty-four of the CanMAGs improved previous genome assemblies for the bacterial species they represented, both by recovering more ribosomal genes and by improving the genomic contiguity (Fig. 1, Table S2).

We compared each genome assembly in this study (CanMAG) to the best genome assembly of the same species from public datasets (Table S2). These genome assemblies from public datasets were: (i) short-read MAGs (n=19; 10 from faecal catalogues and 9 from the GTDB); (ii) genome assemblies from pure cultures [whole-genome sequencing (WGS) assemblies, short-read; n=12]; and (iii) complete genomes (n=3).

Long-read CanMAG assemblies improved all the short-read MAGs for the same species (Table S2). Short-read MAG assemblies were highly fragmented (24 to 223 contigs, mean=144) and had from 0 to 2 ribosomal genes and from 6 to 19 canonical tRNAs (mean=15). In contrast, CanMAGs presented from 9 to 28 ribosomal genes (including 16S, 23S and 5S rRNA genes constituting complete ribosomal operons; different total counts depend on the bacterial species), were more contiguous (1 to 47 contigs, mean=14) and presented at least 18 canonical tRNAs (Fig. 1a). Long-read CanMAG assemblies also improved some of the WGS assemblies, especially regarding ribosomal genes (Fig. 1b). When comparing to complete reference genomes, we recovered a similar number of ribosomal genes (Fig. 1c). Specifically, for the *Enterococcus hirae* CanMAG, we recovered an identical number of rRNA genes, as well as a single-contig for the main genome and a single-contig for the main plasmid. This result further validates that long-read MAGs can recover high-quality genomes.

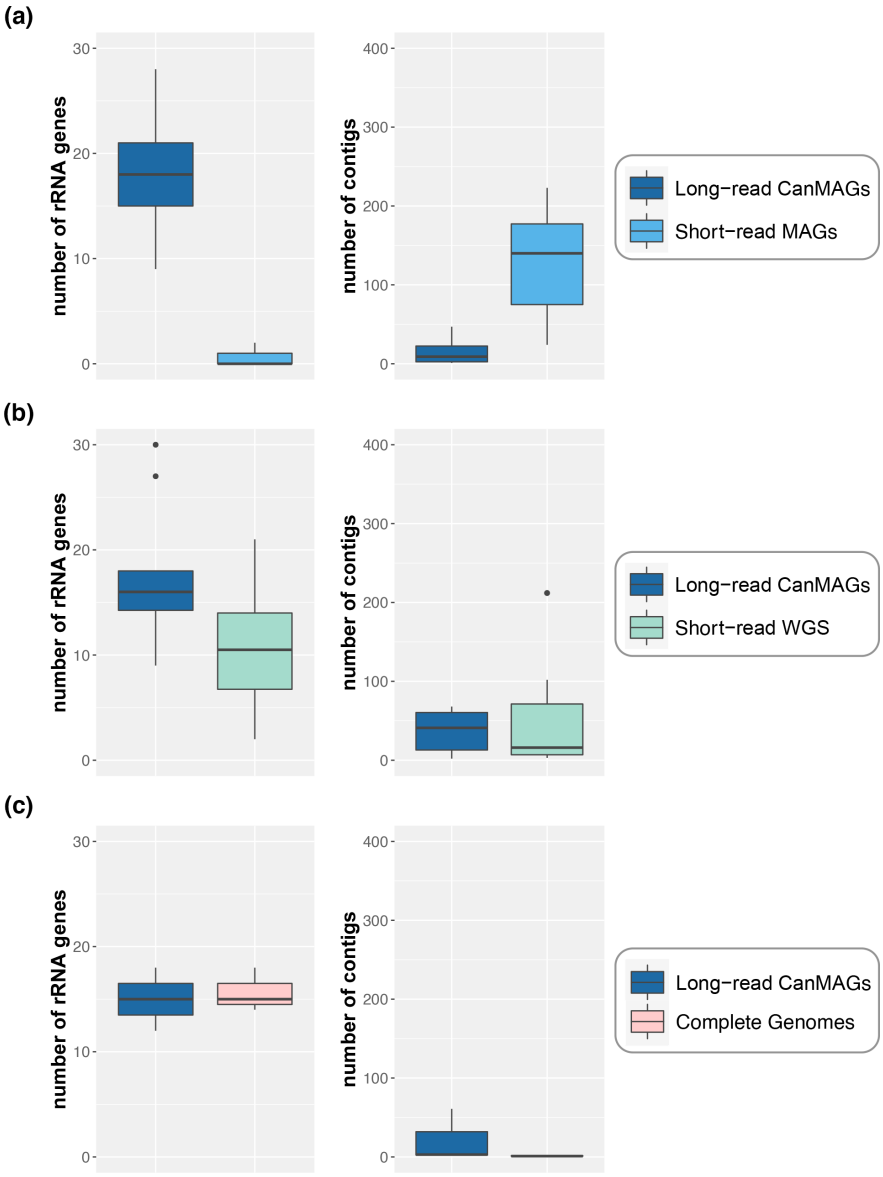

Fig. 1. Number of ribosomal genes and contigs between long-read CanMAGs and representative genomes from public datasets. Boxplots represent the distribution of the number of ribosomal genes (left) and contigs (right) for the bacterial species identified in this study. Other quality parameters assessments are detailed in Table S1. For each bacterial species, the best genome assembly available on public datasets was included for comparison. Representative genomes available from public database were: (a) short-read MAGs for 19 bacterial species, (b) WGS assemblies for 12 bacterial species and (c) complete genome assemblies for 3 bacterial species.

Both canine-specific bacterial species and gut generalists inhabit the dog gastrointestinal environment

We recovered, from the faeces of a healthy dog, 34 CanMAGs that belonged to the phylum *Firmicutes* (*n*=21), followed by the phyla *Bacteroidota* (*n*=8) and *Proteobacteria* (*n*=3). We also found one *Fusobacteriota* and one *Actinobacteriota* CanMAG. Overall, the most abundant genera recovered were: four *Blautia*, two *Blautia_A* (GTDB taxonomy considers they are different, despite being classically regarded as the same genera), and two *Clostridium* species (*Firmicutes*); four *Phocaeicola* (former *Bacteroides* species) [44], and two *Prevotellamassilia* (*Bacteroidota*); and two *Sutterella* (*Proteobacteria*) (Fig. 2).

We assigned taxonomy to the CanMAGs using GTDB-tk and GTDB taxonomy and nomenclature. Seven CanMAGs were predicted to be novel by GTDB-tk (g__ in Fig. 2) and were further compared against animal and human gut MAG catalogues. These CanMAGs presented an ANI >95% to previously reported MAGs in animal and human gut catalogues, so we considered them to be the same bacterial species [31] (Fig. 2, Table S3).

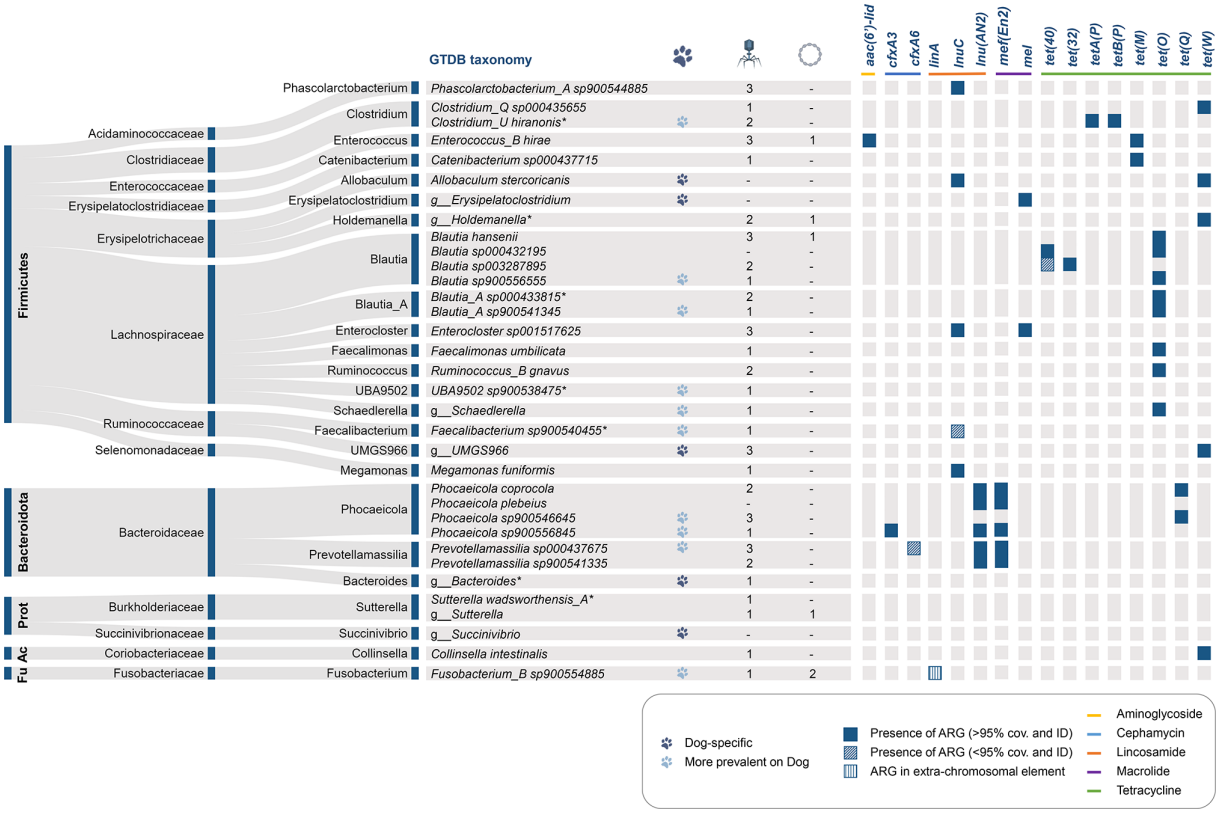

Fig. 2. CanMAGs overview: taxonomy, prevalence in canine gut, ARGs, bacteriophages and plasmids. Fu, *Fusobacteriota*; Ac, *Actinobacteriota*; Prot, *Proteobacteria*. Genome assemblies with a taxonomy of 'g__' are considered novel species by GTDB-tk. Those marked with an asterisk are MQ MAGs. A dark blue paw symbol indicates that the bacterial species has only been observed in dogs when assessing animal and human faecal MAG catalogues; a light blue paw symbol indicates that the bacterial species is more prevalent in dogs (see Table S3 for more details). All the predicted bacteriophages were integrated within the bacterial host chromosome. Plasmids were linked to the genome using Hi-C data. Cov., coverage; ID, identity. Coloured lines represent resistance to a specific antibiotic, as stated in the key.

Canine-specific species include g__*Erysipelatoclostridium*, g__UMGS966, g__*Succinivibrio*, *Allobaculum stercoricanis* and UBA9502 sp900538475. Moreover, g__*Holdemanella* CanMAG is only observed in dogs and cats. We also detected other bacterial species that are more prevalent in canine gut metagenomes compared to human or other animal guts (Fig. 2, Table S3).

Finally, 18 of the bacterial species represented by CanMAGs have been found in animal and human gastrointestinal microbiomes, suggesting they are more adaptable in different gastrointestinal environments, and probably represent gastrointestinal generalists (in Fig. 2, MAGs without the 'paw' symbol, Table S3).

CanMAGs harboured ARGs and prophages and were linked to plasmids

CanMAGs harboured ARGs, but no virulence factors. We detected 16 different ARGs spread among the different CanMAGs, most of them located in the bacterial chromosome (Fig. 2). Only one ARG was located in a plasmid: the *linA* gene in PL2-CanMAG_34 from *Fusobacterium_B* sp900554885 (Table S4).

The most prevalent antimicrobial resistance was to tetracycline, encoded by eight different ARGs and present in 19 out of 34 CanMAGs; followed by lincosamide, encoded by three different ARGs and present in 11 CanMAGs. Specifically, the most prevalent ARG was the *tet(O)* gene present in eight different *Lachnospiraceae* CanMAGs, which conferred resistance to tetracycline. *tet(W)* was also prevalent, and observed in five CanMAGs from different phyla. Finally, *mef(En2)* and *lnu(AN2)* were also present in five different CanMAGs from the genera *Phocaeicola* and *Prevotellamassilia* (not present in *Phocaeicola* sp900546645) (Fig. 2). They had exactly the same distribution pattern, since they were contiguous in the genome.

We also detected 50 bacteriophages in the CanMAGs, ranging from 0 to 3 per genome (Fig. 2). The bacteriophages were integrated within the bacterial chromosome (prophages) rather than in free viral particles. We further describe them in the following section.

Finally, Hi-C proximity ligation linked some potential plasmids to their bacterial host (Fig. 2, Table S4). We identified six potential plasmids linked to *Enterococcus hirae*, *g__Holdemanella*, *Blautia hansenii* and *g__Sutterella* CanMAGs, and two plasmids to *Fusobacterium_B* sp900554885 CanMAG. They presented an increased coverage compared to their bacterial host chromosome, and five of them were circular. Moreover, the plasmids contained typical plasmid- or mobilome-associated genes, and BLAST matched to previously identified plasmids – despite usually with a low coverage (Table S4).

CanMAG prophages provide novel bacterial host information

We detected 50 bacteriophages in the CanMAGs integrated within the bacterial chromosome (prophages) (Fig. 2, Table 1): 29 were high quality (>90% completeness), and 21 were genome-fragments with >50% completeness (as defined by MIUViG criteria) [29] (Table 1). Low-quality predicted bacteriophages (as determined by CheckV) were not included in this analysis.

CanMAGs harboured from 0 to 3 prophages with genome sizes ranging from 2515 to 191453 bp (Table 1). Fourteen out of 34 CanMAGs harboured two or more different prophages, which were at least >50% complete. Within each CanMAG genome, prophages were different among them, which could indicate co-infection events.

To assess the similarity of CanMAG prophages to previous datasets, we clustered our prophages together with a subset of the GPD [40], with 682 bacteriophage sequences. Thirty-three CanMAG prophages were clustered into 27 viral clusters (VCs) (Fig. 3, Tables 1 and S5). Each VC included from 3 to 27 bacteriophage sequences (derived from GPD and CanMAGs) and grouped bacteriophages with similar genome sizes (Fig. 3a) and bacterial hosts (Fig. 3b). Finally, 18 CanMAG prophages were further classified by vCONTACT2 as: outliers (n=9), when they were attached to a VC, but not statistically significant; overlap (n=7), when they presented overlapping genes between two or more VC; and singletons (n=2), when they did not cluster with anything else (Table 1).

All CanMAG prophages were embedded within a highly complete bacterial genome (HQ MAGs and MQ MAGs), so their bacterial host was clear. However, >75% of the GPD bacteriophages constituting the VCs lacked bacterial host information. This is due to the challenge of recovering genomic context with short-read sequencing data. More specifically, we provided novel bacterial host information for 8 out of the 27 VCs that included GPD viral genomes (ND in the GPD bacterial host column in Table 1): VC_241, VC_254, VC_553, VC_403, VC_554, VC_405, VC_488 and VC_257.

For the other VCs, CanMAG bacteriophages presented similar bacterial hosts as the GPD representatives that had this information within each cluster. Three VCs shared a specific bacterial host at the species level: VC_253 contained bacteriophages only observed in *Megamonas funiformis*; VC_342, in *Blautia hansenii*; and VC_347, in *Clostridium hiranonis*. Four VCs shared the same bacterial host at the genus level: VC_219, VC_545 and VC_318 contained bacteriophages only observed in the genus *Phocaeicola*; and VC_348, in *Fusobacterium*. The remaining VCs grouped bacteriophages with a broader range of bacterial hosts (family or above).

Finally, all the bacteriophages were predicted to be integrated except for BP3-CanMAG_15, which was circular and lytic. Despite harbouring only one viral protein, it clustered together with other GPD bacteriophages in VC_554, which was probably grouping another extra-chromosomal element rather than a lytic virus. In addition, most of the predicted prophages were dsDNA, except three that VirSorter2 predicted as ssDNA: BP1-CanMAG_17 (*Ruminococcus_B gnavus*) and BP2-CanMAG_09 (*Blautia hansenii*), which were clustering together in VC_552; and BP1-CanMAG_33 (*Collinsella intestinalis*), which was a singleton.

CanMAGs presented variable proportions of carbohydrate transport and metabolism, energy production and conversion, and mobilome functions

We assessed the functional potential of the CanMAGs by annotating them with the COG (Clusters of Orthologous Genes) database [45]. Heatmap hierarchical clustering of the relative abundances of the main COG functions showed two clusters revealing carbohydrate transport and metabolism functions as the most variable COG category across the CanMAGs (Fig. 4). Half of the CanMAGs were within the first cluster, showing a high percentage (12–18%) of carbohydrate transport and metabolism functions, whereas the other half belonged to the second cluster with a low percentage (2–11%) of this COG category (Table S6). Notably, the CanMAGs with a low percentage of carbohydrate transport and metabolism showed a high percentage of translation, ribosomal and biogenesis functions (Fig. 4, Table S6), which likely reflected a higher protein translation at the expense of the metabolic activity in these bacteria.

Within the first cluster, *Blautia*, *Blautia_A*, *UBA9502* (*Lachnospiraceae*), *Enterocloster* and *Rumnicoccus* CanMAGs showed the highest percentage (>15%) of carbohydrate transport and metabolism functions (Table S6). The most abundant subcategory COG functions for these CanMAGs were ABC-type sugar transport systems and sugar phosphorylation by kinases (ribokinase, 6-phosphofructokinase). In particular, the ABC-type glycerol-3-phosphate transport was the most abundant function. Within this cluster, *Succinivibrio* showed a unique abundance pattern of the TRAP-type C4-dicarboxylate transport system, and *Allobaculum* and *Catenibacterium* showed a high specific abundance of β-glucosidase/β-galactosidase genes (Table S7).

Apart from the clear pattern driven by carbohydrate transport and metabolism functions, we also detected that *Sutterella* CanMAGs presented a high proportion of energy production and conversion (11–12%), while the rest of the CanMAGs showed

Table 1. Predicted bacteriophages in CanMAGs: main characteristics and clustering information

Most of the predicted bacteriophages (BPs) were integrated into the CanMAG bacterial genome and dsDNA. We clustered them together with the GPD subset to create VCs. BP sequences were classified as: clustered (C), when confidently grouping in a VC; outlier (Out), when despite some links to a VC, the association was not statistically significant; overlap (Ovl), when the BP was linked to two or more VCs; or singleton (S), when it did not match any VC. % Compl., % completeness as assessed by CheckV. Details on the VCs can be found in Table S5.

Bacterial host (in this study)	BP ID	VC	VC status	VC size	BP length	% Compl.	Gene count	No. of viral genes	No. of host genes	GPD bacterial host*
Firmicutes										
Enterocloster sp001517625	BP1-CanMAG_15	VC_183	C	11	25 334	65.49	35	14	0	*Lachnospiraceae*
UBA9502 sp900538475	BP1-CanMAG_18	VC_183	C	11	39 523	100	66	16	0	*Lachnospiraceae*
Blautia sp003287895	BP1-CanMAG_11	VC_301	C	5	28 487	83.11	38	16	0	*Lachnospiraceae*
Blautia sp900556555	BP1-CanMAG_12	VC_344	C	9	36 598	100	50	10	0	*Lachnospiraceae*
Blautia_A sp000433815	BP1-CanMAG_13	VC_241	C	7	26 155	74.49	53	11	0	ND
Blautia hansenii	BP1-CanMAG_09	VC_342	C/S	–	151 986	89.15	226	40	7	*Blautia hansenii*
g__UMGS966; s__	BP1-CanMAG_21	VC_267	C	8	47 108	100	65	21	2	*Ruminococcaceae*
Clostridium_Q sp000435655	BP1-CanMAG_02	VC_254	C	5	53 237	100	74	15	3	ND
Clostridium_U hiranonis	BP1-CanMAG_03	VC_347	C	3	34 195	51.54	55	22	1	*Clostridium_U hiranonis*
Blautia_A sp000433815	BP2-CanMAG_13	VC_553	C	3	150 650	100	143	1	66	ND
Megamonas funiformis	BP1-CanMAG_22	VC_253	C	5	35 900	100	57	16	1	*Megamonas funiformis*
Catenibacterium sp000437715	BP1-CanMAG_05	VC_217	C	27	45 860	97.95	53	23	1	*Firmicutes*
g__Holdemanella; s__	BP1-CanMAG_08	VC_217	C	27	44 640	88.4	60	21	3	*Firmicutes*
Enterocloster sp001517625	BP2-CanMAG_15	VC_217	C	27	27 920	59.55	33	19	2	*Firmicutes*
Phascolarctobacterium_A sp900544885	BP2-CanMAG_01	VC_555	C	4	39 056	95.43	59	22	0	*Negativicutes*
Phascolarctobacterium_A sp900544885	BP1-CanMAG_01	VC_036	C	4	57 434	100	92	44	0	*Negativicutes*
Faecalibacterium sp900540455	BP1-CanMAG_20	VC_403	C	16	34 244	97.95	53	22	1	ND
Blautia hansenii	BP2-CanMAG_09	VC_552	C	3	3767	90.52	5	1	0	*Lachnospiraceae*
Ruminococcus_B gnavus	BP1-CanMAG_17	VC_552	C	3	6213	100	10	2	0	*Lachnospiraceae*
Enterocloster sp001517625	BP3-CanMAG_15	VC_554	C	7	191 453	68.4	258	1	13	ND
Blautia hansenii	BP3-CanMAG_09	–	Out	–	25 525	51.57	20	1	3	–
Blautia sp003287895	BP2-CanMAG_11	–	Out	–	19 133	100	17	7	0	–
Blautia_A sp900541345	BP1-CanMAG_14	–	Out	–	27 724	58.84	40	12	0	–
g__UMGS966; s__	BP2-CanMAG_21	–	Out	–	40 694	89.88	61	27	1	–
g__UMGS966; s__	BP3-CanMAG_21	–	Out	–	29 305	66.4	45	12	1	–
Clostridium_U hiranonis	BP2-CanMAG_03	–	Ovl	–	41 047	75.27	70	23	1	–
Phascolarctobacterium_A sp900544885	BP3-CanMAG_01	–	Ovl	–	22 711	54.16	34	13	1	–

Continued

Table 1. Continued

Bacterial host (in this study)	BP ID	VC	VC status	VC size	BP length	% Compl.	Gene count	No. of viral genes	No. of host genes	GPD bacterial host*
Ruminococcus_B gnavus	BP2-CanMAG_17	–	Ovl	–	36 619	95.92	67	20	0	–
Enterococcus_B hirae	BP1-CanMAG_04	–	S	–	32 704	50.36	38	8	3	–
Enterococcus_B hirae	BP2-CanMAG_04	–	Ovl	–	41 858	100	58	34	0	–
Enterococcus_B hirae	BP3-CanMAG_04	–	Out	–	34 545	90.55	50	9	3	–
Faecalimonas umbilicata	BP1-CanMAG_16	–	Ovl	–	33 688	83.74	57	24	0	–
g__Schaedlerella; s__	BP1-CanMAG_19	–	Out	–	37 489	93.41	51	13	1	–
g__Holdemanella; s__	BP2-CanMAG_08	–	Ovl	–	33 282	96.2	62	21	0	–
Bacteroidota										
Phocaeicola sp900546645	BP1-CanMAG_25	VC_219	C	12	34 229	92.18	44	18	2	*Phocaeicola*
Phocaeicola sp900556845	BP1-CanMAG_26	VC_318	C	10	47 132	100	63	9	1	*Phocaeicola*
Phocaeicola coprocola	BP2-CanMAG_23	VC_544	C	4	57 738	100	63	4	3	*Bacteroidaceae*
Phocaeicola sp900546645	BP2-CanMAG_25	VC_544	C	4	44 212	98.3	54	7	1	*Bacteroidaceae*
Phocaeicola sp900546645	BP3-CanMAG_25	VC_545	C	18	58 284	91.08	55	9	2	*Phocaeicola*
Prevotellamassilia sp000437675	BP3-CanMAG_27	VC_547	C	3	31 919	75.45	43	2	1	*Bacteroidaceae*
Phocaeicola coprocola	BP1-CanMAG_23	VC_508	C	11	59 043	100	74	10	3	*Bacteroidales*
Prevotellamassilia sp000437675	BP2-CanMAG_27	VC_510	C	13	44 671	92.29	47	5	5	*Bacteroidaceae*
Prevotellamassilia sp000437675	BP1-CanMAG_27	VC_405	C	5	37 057	74.49	49	8	0	ND
g__Bacteroides; s__	BP1-CanMAG_29	VC_488	C	3	6365	100	9	3	0	ND
Prevotellamassilia sp900541335	BP1-CanMAG_28	–	Out	–	20 636	54.51	16	1	4	–
Prevotellamassilia sp900541335	BP2-CanMAG_28	–	Out	–	37 022	57.64	18	2	2	–
Fusobacteriota										
Fusobacterium_B sp900554885	BP1-CanMAG_34	VC_348	C	12	43 899	100	75	13	2	*Fusobacterium*
Proteobacteriota										
g__Sutterella; s__	BP1-CanMAG_31	VC_257	C	7	42 692	90.3	72	27	0	ND
Sutterella wadsworthensis_A	BP1-CanMAG_30	–	Ovl	–	45 521	95.47	65	27	2	–
Actinobacteriota										
Collinsella intestinalis	BP1-CanMAG_33	–	S	–	2515	60.44	2	1	0	–

*Predicted bacterial host for GPD representatives within a specific VC; if variable taxa, we state the lowest shared taxonomic information. ND, Not determined – no reported bacterial host in GPD.

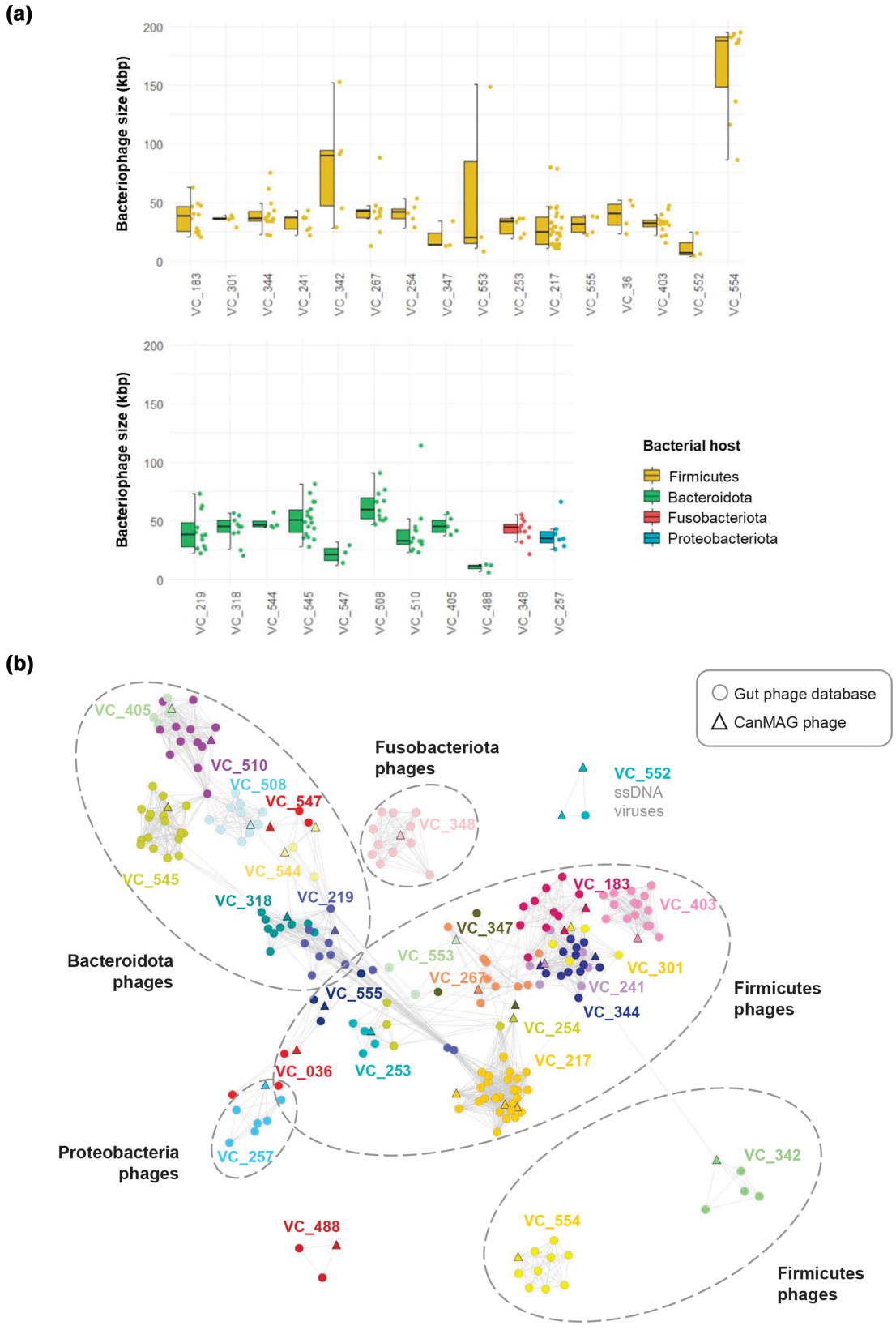

Fig. 3. Analysis of the 27 VCs that included CanMAG bacteriophages. Both parts of the figure contain data from the 33 clustered CanMAG bacteriophages and the representatives from GPD grouping together within the same VC. (a) Boxplots representing the bacteriophage genome sizes within each VC coloured by bacterial host phylum. (b) VCs network. For visualization purposes, each VC is coloured differently.

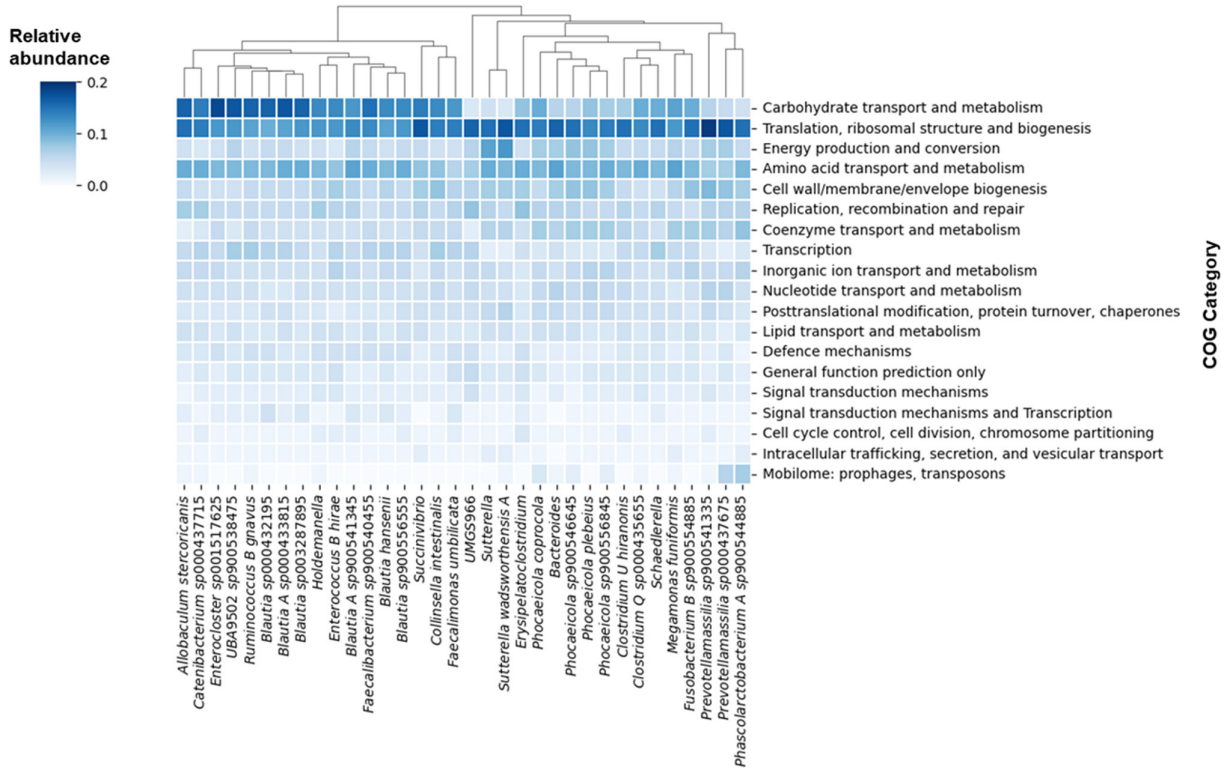

Fig. 4. Heatmap hierarchical clustering of the most abundant COG functions for CanMAGs. The CanMAGs are divided in two main clusters driven by carbohydrate transport and metabolism relative abundances. Only the most abundant COG functions are represented in the plot, for detailed COG functions see Tables S6 and S7.

2–8% of this function (Table S6). Within the functional subcategories, we observed that the most abundant was the succinate dehydrogenase/fumarate reductase flavoprotein subunit, followed by the anaerobic selenocysteine-containing dehydrogenase and the C4-dicarboxylate transporter DcuC (Table S7).

Finally, another interesting function that presented divergent patterns among CanMAGs was the mobilome, a COG function that is usually missed with short-reads metagenomes [12, 13, 16]. In *Phascolarctobacterium_A* sp900544885 and *Prevotellamassilia* sp000437675, the mobilome COG category was abundant and represented >5% of the total abundance; also in *Phocaeicola coprocola*, *Phocaeicola* sp900546645 and *Phocaeicola* sp900556845 represented >2% (Table S6). In CanMAGs, most of the genes inside this category were transposases. The most abundant transposases were the IS*5* family in *Prevotellamassilia*, the IS*30* in *Phascolarcobacterium* and the IS*4* family in the three *Phocaeicola* CanMAGs (Table S7).

DISCUSSION

Long-read metagenomics retrieved long contigs harbouring complete assembled ribosomal operons, prophages and other MGEs. Hi-C allowed the binning of the long contigs into HQ MAGs and MQ MAGs, some of them representing closely related species. Moreover, Hi-C data also linked plasmids to their bacterial host. By combining nanopore long-read metagenomics and Hi-C proximity ligation, we provided 27 HQ MAGs and 7 MQ MAGs from a single sample of the canine gut environment.

To date, only one comprehensive study has used shotgun metagenomics (short-read sequencing) to retrieve MAGs of the canine gut microbiome rather than the 16S rRNA gene to obtain a taxonomic profile [8], and none of the retrieved MAGs fulfilled the high-quality MIMAG criteria [9]. Recently, we characterized the same faecal sample using only long-read metagenomics and recovered eight single-contig HQ MAGs by combining metagenome assemblies from different data subsets (all data, 75% data, high-molecular-weight DNA data), demonstrating the potential of long-read metagenomics to retrieve HQ MAGs [10]. Here, we added Hi-C proximity ligation data to allow the binning of the long-read contigs, improving the contiguity of the long-read metagenomics assembly and retrieving 34 CanMAGs.

The CanMAGs improved the short-read-based genome assemblies on public datasets for the species they represented, which mainly derived from shotgun metagenomics and WGS studies. These HQ CanMAGs serve as a proof-of-concept that, extended to more microbiome members and to larger cohorts, will provide biological insights to better understand the canine gut environment in health and disease, such as the impact of microbiome modulation strategies (dietary interventions, or prebiotic and probiotic supplementation).

Animal gut microbiomes have specific taxonomic profiles and specific gene functions associated with the animal's diet, taxonomy and gut morphology, among other factors [46, 47]. Half of the CanMAGs were more prevalent in dog gut than in human and other animal guts, suggesting a certain degree of specialization and a need for a niche-specific database. Canine-specific microbes might be a more appropriate probiotics source, rather than extending the use of human probiotics directly to dogs [48]. As an example, *Succinivibrio* is more prevalent in dog gut than other animal guts. This CanMAG showed a uniquely high proportion of the TRAP-type C4-dicarboxylate transport system, allowing C4-dicarboxylates like succinate, fumarate and malate (tricarboxylic acid (TCA) cycle intermediates) to be moved. *Succinivibrio* ferments organic matter produced by the TCA cycle to generate acetate and succinate [49]. Acetate is a short-chain fatty acid (SCFA) that reduces whole-body lipolysis and pro-inflammatory cytokine levels by increasing energy expenditure and fat oxidation [50]. Thus, evaluating canine *Succinivibrio* as a potential probiotic in breeds with overweight or obesity problems would be interesting.

Most of the CanMAGs belonged to *Firmicutes*, followed by *Bacteroidota*, which agrees with the most abundant phyla described for the healthy canine gastrointestinal microbiome [8, 51]. We also recovered MAGs for several *Clostridiales* members (families *Lachnospiraceae*, *Clostridiaceae*, *Ruminococcaceae*), which are among the most prevalent in the large intestine of dogs [52]. At the functional level, the most prevalent and abundant COG function of CanMAGs was translation, ribosomal structure and biogenesis. In contrast, carbohydrate transport and metabolism presented a highly variable proportion, being the most abundant function in thirteen *Firmicutes* CanMAGs. A previous study reported an overrepresentation of carbohydrate metabolism in the domestic dog gut microbiome compared to wolves, probably due to dog diets containing complex polysaccharides [53].

Long-read HQ MAGs are more comparable to complete genomes since they harbour key biological elements such as ARGs, MGEs and prophages that help in the understanding of biological processes like horizontal gene transfer events. The CanMAGs harboured 16 ARGs related to resistance to five different types of antibiotics: tetracycline (8 ARGs in 19 CanMAGs); lincosamides (3 ARGs in 11 CanMAGs); macrolides (2 ARGs in 7 CanMAGs); cephamycin (2 ARGs in 2 CanMAGs); and aminoglycosides (1 ARG in 1 CanMAG). In agreement with our results, ARGs conferring resistance to tetracycline were the most prevalent, followed by lincosamides and macrolides, in the gut of healthy dogs [54]. Specifically, among the tetracycline ARGs, the most prevalent in CanMAGs were *tet(O)* and *tet(W)*, in agreement with previous studies on healthy dogs [54, 55]. Most of the detected ARGs in the CanMAGs were shared among very similar taxa at the genera or family levels. The two main exceptions for this dog were *tet(W)* and *lnu(C)*. They were both shared among different family members of the phylum *Firmicutes*. We even detected *tet(W)* in the *Actinobacteria* CanMAG. Previous work on dog gut also described a broad range of hosts for these two ARGs, which should be carefully monitored [54].

Long-read metagenomics identifies transposases and MGEs that are missed with short-read metagenomics studies [12–14, 16]. Insertion sequences (ISs) are among the simplest MGEs and are widespread in all domains of life. CanMAGs harbour both ISs and integrated prophages. For example, the *Phascolarcobacterium* CanMAG harboured abundant functions linked to the IS*30* family, and three *Phocaeicola* CanMAGs harboured abundant transposase InsG linked to the IS*4* family. ISs can move within a genome or horizontally between different bacterial genomes as part of other MGE vectors such as phages and plasmids. Thus, screening and controlling MGEs is a key step since IS elements can affect antibiotic-resistance patterns [56].

The most common approach to determine a bacteriophage's bacterial host is by bioinformatically screening CRISPR spacers of bacterial genomes and then further confirming the prediction by analysing co-occurrence patterns between bacterial host and prophages. In the GPD (~142000 non-redundant viral genomes), only 28% of the bacteriophages can be linked to a bacterial host [40]. Here, we provide experimental evidence by long-read metagenomics of some of these predicted bacteria–bacteriophage interactions and report novel bacterial host information for eight VCs. We identified a total of 50 different bacteriophages (with >50% completeness) integrated within the CanMAG genomes and clustered them together with a subset of the GPD to identify their host range.

Overall, identifying the bacterial host and the co-infections with multiple bacteriophages is critical to understanding the biological impact on the bacterial host metabolism and function, and the global effect on microbiome dynamics, and for the development of phage therapies [57]. We described three species-specific VCs containing bacteriophages that infected *Megamonas funiformis*, *Blautia hansenii* and *Clostridium hiranonis* exclusively; and four genus-specific VCs, three for the genus *Phocaeicola* and one for *Fusobacterium*. However, most of the VCs included bacteriophages with a broad spectrum of bacterial hosts, contrasting with GPD findings, where most of the VCs were predicted to be species-specific – note that most of the bacteriophages lacked bacterial host information [40] – and agreeing with some other studies that suggest that most bacteriophages have a broad range [58].

Apart from the experimental binning of the long contigs to retrieve MAGs, Hi-C proximity ligation cross-links extra-chromosomal elements within a single cell [22, 23, 59–61]. We linked the six potential plasmids to their bacterial host. We might have missed some plasmids as we did not use the rapid sequencing kit for the nanopore library preparation, preferred for this objective [62]. Since we aimed to retrieve longer reads, we used the ligation sequencing kit rather than the rapid sequencing kit, which produces shorter reads because it uses transposase fragmentation to insert the adapters. If aiming to assess links between extra-chromosomal elements and their hosts, we would also recommend evaluating the use of the rapid sequencing kit despite the shorter read length, which should be compensated with Hi-C binning data.

The technical approach used included a high-molecular-weight DNA extraction, which provided long reads that facilitated the assembly of closely related bacterial species. In fact, we retrieved different bacterial species within the same genera, as seen for *Phocaeicola* and *Blautia* species. A recent study combining long-read metagenomics with Hi-C proximity ligation data confirmed that around 50% of long-read MAGs within a sheep faecal sample were polymorphic and collapsed different lineages within a single MAG [24]. We cannot rule out that this might have happened to some of the CanMAGs since we did not perform SNP-level analysis and haplotype phasing. Further steps will aim to retrieve more MAGs using binning bioinformatics tools or performing lineage-resolved metagenomics.

In conclusion, the HQ MAGs improve the short-read-based genome assemblies in public datasets, which mainly derive from shotgun metagenomics and WGS studies. These HQ MAGs present a high added value to better understand the microbiome composition and functional capacity in health and disease and better assess the impact of microbiome modulation strategies with niche-specific databases for non-model organisms. Nanopore sequencing is affordable for any lab, and recent advances in sequencing chemistry and basecalling software have improved the raw read quality, allowing nearly perfect bacterial genomes from metagenomes [63]. Nanopore long-read metagenomics and Hi-C binning are likely to become a comprehensive approach to discovering HQ MAGs and assigning extra-chromosomal elements to the bacterial host.

Funding information
This work was supported by Vetgenomics and the Molecular Genetics Veterinary Service (SVGM), Universitat Autònoma de Barcelona, Spain. The Spanish Ministry of Science and Innovation granted a Torres Quevedo Project to Vegenomics, reference no. PTQ2018-009961, that was cofinanced by the European Social Fund.

Acknowledgements
We would like to thank Justa Martín, from Vetgenomics, for the initial support with the Hi-C procedure. We would also like to thank Ivan Liahcko and Gherman Uritskiy from Phase Genomics (Seattle, USA) for their support with the Hi-C data analysis. Finally, we would like to thank the anonymous reviewers for their insights and suggestions that have helped to improve the article.

Author contributions
O.F. and A.C., conceptualized the study and designed the experiments. D.P., extracted the DNA, and worked on the sequencing libraries, the nanopore sequencing and the Hi-C proximity ligation protocol. A.C., performed the metagenome assembly and correction, and analysed and interpreted the data. J.V. and N.F., analysed the plasmid data. J.V., analysed the ARGs. N.F., analysed the functional profile. A.C., wrote the main manuscript text. O.F., N.F., D.P. and J.V., substantially revised the work. All the authors have approved the submitted version.

Conflicts of interest
A.C., N.F. and J.V. work for Vetgenomics. The other authors declare that there are no conflicts of interest.

References

1. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, *et al*. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* 2019;176:649–662.
2. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, *et al*. A new genomic blueprint of the human gut microbiota. *Nature* 2019;568:499–504.
3. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, *et al*. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* 2021;39:105–114.
4. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, *et al*. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017;35:725–731.
5. Yuan C, Lei J, Cole J, Sun Y. Reconstructing 16S rRNA genes in metagenomic data. *Bioinformatics* 2015;31:i35–i43.
6. Pilla R, Suchodolski JS. The role of the canine gut microbiome and metabolome in health and gastrointestinal disease. *Front Vet Sci* 2019;6:498.
7. Wernimont SM, Radosevich J, Jackson MI, Ephraim E, Badri DV, *et al*. The effects of nutrition on the gastrointestinal microbiome of cats and dogs: impact on health and disease. *Front Microbiol* 2020;11:1266.
8. Coelho LP, Kultima JR, Costea PI, Fournier C, Pan Y, *et al*. Similarity of the dog and human gut microbiomes in gene content and response to diet. *Microbiome* 2018;6:72.
9. Youngblut ND, de la Cuesta-Zuluaga J, Reischer GH, Dauser S, Schuster N, *et al*. Large-scale metagenome assembly reveals novel animal-associated microbial genomes, biosynthetic gene clusters, and other genetic diversity. *mSystems* 2020;5:e01045-20.
10. Cuscó A, Pérez D, Viñes J, Fàbregas N, Francino O. Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces. *BMC Genomics* 2021;22:330.
11. Singleton CM, Petriglieri F, Kristensen JM, Kirkegaard RH, Michaelsen TY, *et al*. Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nat Commun* 2021;12:2009.
12. Suzuki Y, Nishijima S, Furuta Y, Yoshimura J, Suda W, *et al*. Long-read metagenomic exploration of extrachromosomal mobile genetic elements in the human gut. *Microbiome* 2019;7:119.

13. Bertrand D, Shaw J, Kalathiyappan M, Ng AHQ, Kumar MS, *et al*. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat Biotechnol* 2019;37:937–944.
14. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol* 2020;38:701–707.
15. Yahara K, Suzuki M, Hirabayashi A, Suda W, Hattori M, *et al*. Long-read metagenomics using PromethION uncovers oral bacteriophages and their interaction with host bacteria. *Nat Commun* 2021;12:27.
16. Che Y, Xia Y, Liu L, Li A-D, Yang Y, *et al*. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. *Microbiome* 2019;7:44.
17. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile genetic elements associated with antimicrobial resistance. *Clin Microbiol Rev* 2018;31:e00088-17.
18. Maghini DG, Moss EL, Vance SE, Bhatt AS. Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome. *Nat Protoc* 2021;16:458–471.
19. Nicholls SM, Quick JC, Tang S, Loman NJ. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience* 2019;8:giz043.
20. Arumugam K, Bağcı C, Bessarab I, Beier S, Buchfink B, *et al*. Annotated bacterial chromosomes from frame-shift-corrected long-read metagenomic data. *Microbiome* 2019;7:61.
21. Burton JN, Liachko I, Dunham MJ, Shendure J. Species-level deconvolution of metagenome assemblies with Hi-C–based contact probability maps. *G3* 2014;4:1339–1346.
22. Beitel CW, Froenicke L, Lang JM, Korf IF, Michelmore RW, *et al*. Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ* 2014;2:e415.
23. Bickhart DM, Watson M, Koren S, Panke-Buisse K, Cersosimo LM, *et al*. Assignment of virus and antimicrobial resistance genes to microbial hosts in a complex microbial community by combined long-read assembly and proximity ligation. *Genome Biol* 2019;20:153.
24. Bickhart DM, Kolmogorov M, Tseng E, Portik DM, Korobeynikov A, *et al*. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat Biotechnol* 2022.
25. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, *et al*. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 2017;27:722–736.
26. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, *et al*. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* 2020;17:1103–1110.
27. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;12:59–60.
28. Huson DH, Albrecht B, Bağcı C, Bessarab I, Górska A, *et al*. MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. *Biol Direct* 2018;13:6.
29. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2019;36:1925–1927.
30. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, *et al*. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 2020;38:1079–1086.
31. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.
32. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, *et al*. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2017;45:D566–D573.
33. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, *et al*. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007;35:3100–3108.
34. Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res* 2018;46:e35.
35. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
36. Chen L, Yang J, Yu J, Yao Z, Sun L, *et al*. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 2005;33:D325–D328.
37. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, *et al*. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 2021;9:37.
38. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 2020;8:90.
39. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, *et al*. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol* 2019;37:632–639.
40. Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. Massive expansion of human gut bacteriophage diversity. *Cell* 2021;184:1098-1109.
41. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
42. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, *et al*. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;11:119.
43. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–2504.
44. García-López M, Meier-Kolthoff JP, Tindall BJ, Gronow S, Woyke T, *et al*. Analysis of 1,000 type-strain genomes improves taxonomic classification of *Bacteroidetes*. *Front Microbiol* 2019;10:2083.
45. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, *et al*. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res* 2021;49:D274–D281.
46. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol* 2008;6:776–788.
47. Levin D, Raab N, Pinto Y, Rothschild D, Zanir G, *et al*. Diversity and functional landscapes in the microbiota of animals in the wild. *Science* 2021;372:eabb5352.
48. Grześkowiak Ł, Endo A, Beasley S, Salminen S. Microbiota and probiotics in canine and feline welfare. *Anaerobe* 2015;34:14–23.
49. Santos EO, Thompson F. The family *Succinivibrionaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E and Thompson F (eds). *The Prokaryotes*. Berlin, Heidelberg: Springer; 2014. pp. 639–648.
50. Hernández MAG, Canfora EE, Jocken JWE, Blaak EE. The short-chain fatty acid acetate in body weight control and insulin sensitivity. *Nutrients* 2019;11:E1943.
51. Swanson KS, Dowd SE, Suchodolski JS, Middelbos IS, Vester BM, *et al*. Phylogenetic and gene-centric metagenomics of the canine intestinal microbiome reveals similarities with humans and mice. *ISME J* 2011;5:639–649.
52. Suchodolski JS, Camacho J, Steiner JM. Analysis of bacterial diversity in the canine duodenum, jejunum, ileum, and colon by comparative 16S rRNA gene analysis. *FEMS Microbiol Ecol* 2008;66:567–578.
53. Lyu T, Liu G, Zhang H, Wang L, Zhou S, *et al*. Changes in feeding habits promoted the differentiation of the composition and function of gut microbiotas between domestic dogs (Canis lupus familiaris) and gray wolves (Canis lupus). *AMB Express* 2018;8:123.
54. Kim Y, Leung MHY, Kwok W, Fournié G, Li J, *et al*. Antibiotic resistance gene sharing networks and the effect of dietary nutritional

content on the canine and feline gut resistome. *Anim Microbiome* 2020;2:4.

55. Pillai DK, Peterson G, Zurek L. Insights into the diversity of gut microbiota and associated antibiotic resistance genes in healthy dogs. *Vet Sci Med* 2019;2:210.
56. Vandecraen J, Chandler M, Aertsen A, Van Houdt R. The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit Rev Microbiol* 2017;43:709–730.
57. Díaz-Muñoz SL. Viral coinfection is shaped by host ecology and virus-virus interactions across diverse microbial taxa and environments. *Virus Evol* 2017;3:vex011.
58. Ross A, Ward S, Hyman P. More is better: selecting for broad host range bacteriophages. *Front Microbiol* 2016;7:1352.
59. Marbouty M, Baudry L, Cournac A, Koszul R. Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay. *Sci Adv* 2017;3:e1602105.
60. Stalder T, Press MO, Sullivan S, Liachko I, Top EM. Linking the resistome and plasmidome to the microbiome. *ISME J* 2019;13:2437–2446.
61. Marbouty M, Thierry A, Millot GA, Koszul R. MetaHiC phage-bacteria infection network reveals active cycling phages of the healthy human gut. *Elife* 2021;10:e60608.
62. Wick RR, Judd LM, Wyres KL, Holt KE. Recovery of small plasmid sequences via Oxford Nanopore sequencing. *Microb Genom* 2021;7:000631.
63. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, *et al*. Oxford Nanopore R10.4 long-read sequencing enables near-perfect bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Biorxiv* 2021.