

RESEARCH

Open Access



Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces

Anna Cuscó^{1*}, Daniel Pérez², Joaquim Viñes^{1,2}, Norma Fàbregas¹ and Olga Francino²

Abstract

Background: Long-read sequencing in metagenomics facilitates the assembly of complete genomes out of complex microbial communities. These genomes include essential biologic information such as the ribosomal genes or the mobile genetic elements, which are usually missed with short-reads. We applied long-read metagenomics with Nanopore sequencing to retrieve high-quality metagenome-assembled genomes (HQ MAGs) from a dog fecal sample.

Results: We used nanopore long-read metagenomics and frameshift aware correction on a canine fecal sample and retrieved eight single-contig HQ MAGs, which were > 90% complete with < 5% contamination, and contained most ribosomal genes and tRNAs. At the technical level, we demonstrated that a high-molecular-weight DNA extraction improved the metagenomics assembly contiguity, the recovery of the rRNA operons, and the retrieval of longer and circular contigs that are potential HQ MAGs. These HQ MAGs corresponded to *Succinivibrio*, *Sutterella*, *Prevotellamassilia*, *Phascolarctobacterium*, *Catenibacterium*, *Blautia*, and *Enterococcus* genera. Linking our results to previous gastrointestinal microbiome reports (metagenome or 16S rRNA-based), we found that some bacterial species on the gastrointestinal tract seem to be more canid-specific –*Succinivibrio*, *Prevotellamassilia*, *Phascolarctobacterium*, *Blautia_A sp900541345*–, whereas others are more broadly distributed among animal and human microbiomes –*Sutterella*, *Catenibacterium*, *Enterococcus*, and *Blautia sp003287895*. *Sutterella* HQ MAG is potentially the first reported genome assembly for *Sutterella stercoricanis*, as assigned by 16S rRNA gene similarity. Moreover, we show that long reads are essential to detect mobilome functions, usually missed in short-read MAGs.

Conclusions: We recovered eight single-contig HQ MAGs from canine feces of a healthy dog with nanopore long-reads. We also retrieved relevant biological insights from these specific bacterial species previously missed in public databases, such as complete ribosomal operons and mobilome functions. The high-molecular-weight DNA extraction improved the assembly's contiguity, whereas the high-accuracy basecalling, the raw read error correction, the assembly polishing, and the frameshift correction reduced the insertion and deletion errors. Both experimental and analytical steps ensured the retrieval of complete bacterial genomes.

Keywords: Long-read metagenomics, Gastrointestinal microbiome, Fecal microbiome, Long-reads, Nanopore, Canine microbiome, Dog microbiome, Metagenome-assembled genomes, *Sutterella*

* Correspondence: anna.cusco@vetgenomics.com

¹Vetgenomics, Ed Eureka, Parc de Recerca UAB, Barcelona, Spain
Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Metagenomics is a powerful and rapidly developing approach that can be used to unravel uncultured microbial diversity and expand the tree of life, and give new biological insights into the microbes inhabiting underexplored environments [1]. When applied to both the canine gastrointestinal (GI) and the fecal microbiomes, metagenomics provides information on health and disease as well as essential clues on how to prevent or treat specific pathologies.

Previous studies have reported similarities between canine and human GI microbiome. In general, different GI diseases relate to an altered GI microbiome that, on the other hand, can be modulated by diet and dietary complements (such as pre- and probiotics) (See [2–5] for extensive reviews). Besides the veterinarian interest itself, dogs are considered closer models to humans than other animal models for GI microbiome studies [6, 7].

Microbiome studies are predominantly either marker-specific (e.g., 16S rRNA gene for Bacteria) or whole metagenome sequencing [8]. To date, the canine GI microbiome studies available use next-generation sequencing –short-read sequencing– or earlier technologies and are mostly amplicon-based strategies (16S rRNA gene). Only three studies used shotgun metagenomics with short-read sequencing to characterize the whole microbial community and the gene content in dog feces [7, 9, 10].

The application of long-read sequencing to metagenomics enables retrieving metagenome-assembled genomes (MAGs) with high completeness. The most recent strategy in long-read metagenomics uses the long reads to obtain the draft metagenome assembly –ensuring the greatest contiguity of MAGs– and short reads to polish and improve the overall accuracy. This strategy was applied to assess the human GI microbiome [11], among others –such as mock communities [12], cow rumen [13], natural whey starter cultures [14], or wastewater [15]. Worthy of considering, some authors suggest that we may overcome the need for short reads to polish long-read data by either using correction software, such as frameshift-aware correction [16], or with ultra-deep coverage of the genomes [12].

In our previous work, we used long-read metagenomics to assess the taxonomy and reach species identification on the canine fecal microbiome. Even though we used a low-depth sequencing approach, we assembled a circular contig corresponding to an *uncultured CrAssphage* [17].

In the present study, we use nanopore long-read metagenomics and frameshift aware correction to overcome the need for polishing with short reads. As a result, we retrieve and characterize eight high-quality MAGs and gain new biological insights into the dog fecal microbiome.

Material and methods

DNA extraction and long-read sequencing

Our study focuses on the analysis of a single fecal sample of a healthy pet dog. The fecal sample used for the DNA extraction was collected when walking a healthy pet dog. We have neither altered nor manipulated the animal in any way. The dog was an adult male Beagle of 6 years and 8 months old with no recent antibiotics intake. The last time he was treated with antibiotics was three years before sampling, when he underwent a 15-day treatment with doxycycline –tetracycline-class antibiotic– due to excess secretion of mucus and saliva. A fresh sample was collected and stored at -80°C until further processing.

We used two different kits from Zymobiomics (Zymo Research) for DNA extraction following the manufacturer's instructions: the Quick-DNA HMW MagBead for High-Molecular Weight DNA (without bead-beating) and the DNA Miniprep Kit, which is with a classical bead-beating based microbiome DNA extraction. Throughout the manuscript, we use HMW-DNA (high-molecular weight DNA) extraction and non-HMW DNA (no high-molecular weight DNA) extraction terms.

Each DNA extraction was sequenced in a single Flowcell R9.4.1 using MinION™ (Oxford Nanopore Technologies). The Ligation Sequencing Kit 1D (SQK-LSK109; Oxford Nanopore Technologies) was used to prepare both libraries. For non-HMW DNA, we followed the manufacturer's protocol. For the HMW-DNA, we tuned few parameters: i) at DNA repair and end-prep step, we incubated at 20°C for 20 min and 65°C for 20 min; ii) we extended rotator mixer (Hula mixer) times to 10 min; iii) we extended elution time after AMPure XP beads to 10 min; iv) final incubation with elution buffer was performed at 37°C for 15 min (as recommended for HMW-DNA).

Raw reads: pre-processing, quality control, and taxonomic analyses

Raw fast5 files 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 7 were discarded. The detailed bioinformatics workflow can be found in Additional File 1.

To obtain the first taxonomic assignment directly from the raw reads, we processed the data using Kraken2 2.0.8 [18] with the maxikraken2 database (Loman Lab, from March 2019) that includes all the genomes from RefSeq. We visualized Kraken2 reports using Sankey diagrams with pavian 1.0.0 R package [19].

We used Nanoplot 1.28 (<https://github.com/wdecoster/NanoPlot>) to obtain the run summary statistics, Porechop 0.2.4 (<https://github.com/rrwick/Porechop>) for

adapters trimming, Nanofilt 2.6.0 (<https://github.com/wdecoster/nanofilt>) to discard reads shorter than 1000 bp, and different modules of seqkit 0.11.0 [20] to manipulate fastq and fasta files during the whole analysis.

Metagenomics assembly and polishing

Before proceeding with the metagenomics assembly, we performed an error-correction step of the raw nanopore reads using canu 2.0 [21], which performs all-versus-all overlapping of the reads to retrieve consensus reads reducing the overall error rate.

We used the corrected reads to perform metagenome assembly with Flye 2.7 [22] (options: `--nano-corr --meta, --genome-size 500 m, --plasmids`). We performed several assemblies with different random amounts of data (100, 75, 50%, and HMW dataset) to recover the maximum number of high-quality MAGs (HQ MAGs). We used Bandage 0.8.1 [23] to visualize the metagenome assemblies. We polished the Flye assembly with one round of medaka 1.0.1 (<https://nanoporetech.github.io/medaka/>), including all the raw fastq files as input.

The next step for the HQ MAGs was to correct the frameshift errors, as described in [16], using Diamond 0.9.32 [24] (options: `--range-culling --top 10 -F 15 --outfmt 100 -c1 -b12 -t /dev/shm`) and MEGAN-LR 6.19.1 [25] (options: `--longReads --lcaAlgorithm longReads --lcaCoveragePercent 51 --readAssignmentMode alignedBases --acc2tax prot_acc2tax-Nov2018X1.abin`). We used ideel (<https://github.com/mw55309/ideel>) to visualize the number of truncated open-reading frames (ORF).

To assess the quality of the MAGs, we used CheckM 1.1.1 [26] to retrieve completeness and contamination. Considering MIMAG criteria, MAGs are classified as: high-quality, with >90% completeness, <5% contamination, and presence of rRNAs genes and tRNAs; medium-quality, with >50% completeness and <10% contamination and low-quality, the remaining ones [27].

Characterization of the high-quality MAGs

GTDB-tk 1.3.0 [28] with GTDB taxonomy release 95 [29] was used to assess the novelty and the taxonomy of HQ MAGs. We used PROKKA 1.13.4 [30] to annotate the MAGs and an associated Perl script to predict the number of pseudogenes (*prokka-suggest_pseudogenes.pl*). We used FastANI 1.3 [31] to confirm a potentially new species by determining the average nucleotide identity (ANI) between the most related genomes.

We extracted the 16S rRNA genes from the HQ MAGs before the frameshift correction step using Anvi'o 6.1 [32]. The 16S rRNA genes were analyzed using MOLE-BLAST tool in NCBI website (<https://blast.ncbi.nlm.nih.gov/moleblast/moleblast.cgi>) to obtain a phylogenetic tree. Mafft [33] in the EBI website was used to

align 16S rRNA gene sequences from *Sutterella* HQ MAG and obtain an identity matrix.

Abricate 0.9.8 (<https://github.com/tseemann/abricate>) was used to detect antimicrobial resistance genes using CARD database [34]. OriTfinder [35] was used to identify the origin of transfer (*oriT*) and conjugative machinery of mobile genetic elements, and SnapGene Viewer 5.0.7 (<https://www.snapgene.com/snapgene-viewer/>) to visualize the results.

Functional and Pangenomics analysis of the HQ MAGs

We compared the HQ MAGs obtained to previously reported MAGs from two recent gastrointestinal collections: i) the animal gut metagenome [10] and ii) the Unified Human Gastrointestinal Genome (UHGG) [36].

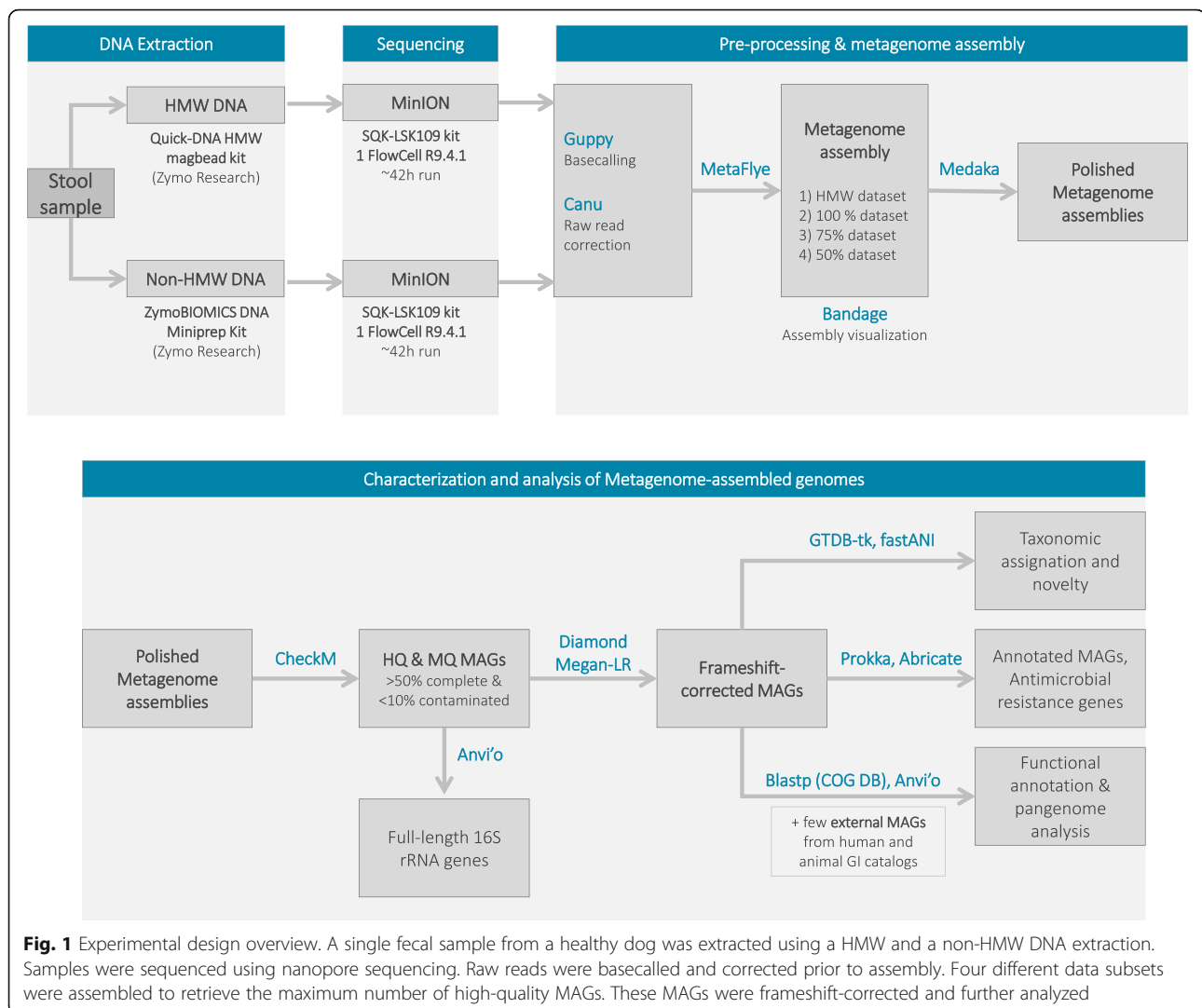
We retrieved MAGs that represented the same species as our HQ MAGs by keeping: i) those with >95% of ANI [31] for the animal gut metagenome; and ii) those with the same species-level taxonomy as stated by GTDB-tk for UHGG.

We performed a pangenome analysis for each bacterial species using Anvi'o 6.2 [32]. The pangenome included our HQ MAG and at least 10 genomes from public databases. If less than 10 genomes were available for a particular bacterial species, we did not perform a pangenome analysis. In contrast, when many genomes were available for a specific bacterial species, we chose high-quality representatives (>90% completeness and <5% contamination) that presented different ANI values against our HQ MAG. Within Anvi'o pangenomics workflow [37], Prodigal [38] was used as a gene caller to identify open reading frames, whereas genes were functionally annotated using blastp against NCBI Clusters of Orthologous Groups (COGs) database [39] (cog2003–2014). The pangenome database was created using NCBI's blastp to calculate each amino acid sequence's similarity in every genome against every other amino acid sequence across all genomes to resolve gene clusters. MCL inflation parameter was set to 10. pyANI was run in Anvi'o to retrieve the ANI within the genomes of a pangenome (Additional File 1 for detailed steps).

Results

We characterized the fecal microbiome of a healthy dog using long-read metagenomics with Nanopore sequencing. An overview of the complete experimental design is presented on Fig. 1. We obtained a total of 16.94 million reads (36.05 Gbp), after two runs corresponding to the HMW and non-HMW DNA extractions.

After high accuracy basecalling and error correction, we performed several metagenomics assembly strategies to retrieve eight single-contig high-quality MAGs (HQ MAGs), which were >90% complete with <5% contamination and contained most ribosomal genes and tRNAs,



and three medium-quality ones (MQ MAGs). We further corrected the HQ MAGs for frameshifts errors and compared them at the functional level with those previously identified in other gastrointestinal metagenome catalogs.

HMW DNA extraction for longer reads and larger contigs

HMW sequencing produced 5.81 million reads with N50 of 4369 bp and a median length of 2312 bp (total throughput: 18.76 Gbp), whereas non-HMW produced 11.13 million reads with N50 of 2102 bp and a median length of 1093 bp (total throughput: 17.29 Gbp).

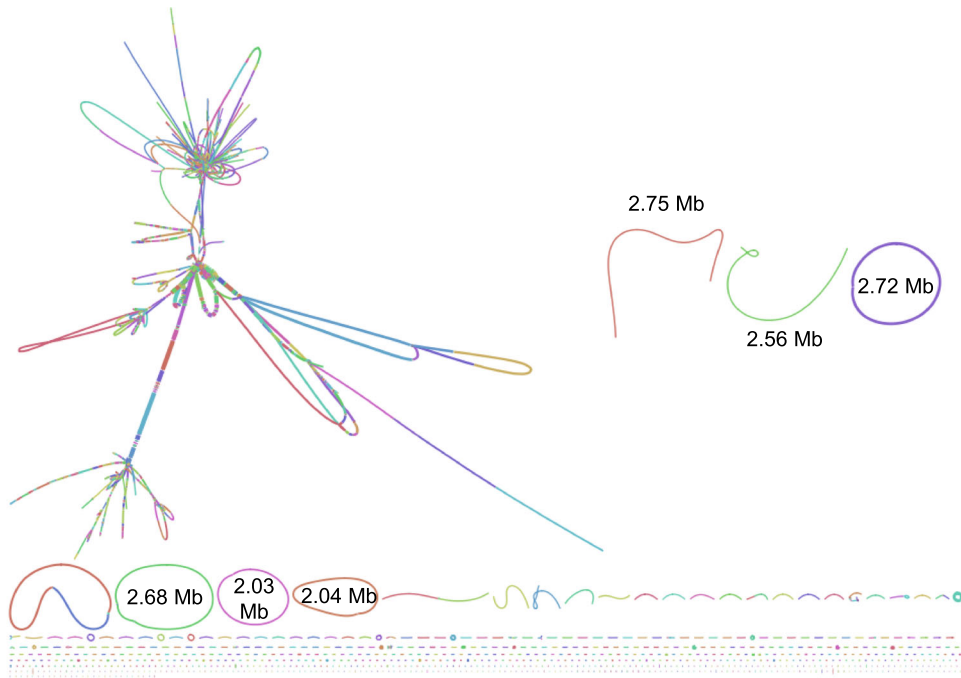
We taxonomically classified all the uncorrected raw reads with Kraken2 and found 81.8% of the classified reads in HMW vs. 70.8% in non-HMW. More than 99% of the total reads corresponded to Bacteria. The most abundant phylum was Bacteroidetes (~ 80% of total reads), followed in abundance by Firmicutes (12.5% in HMW vs. 8.9% in non-HMW), Proteobacteria (~ 5%), and Fusobacteria (1.9% in HMW vs. 3.9% in non-

HMW). At the genus level, this dog fecal microbiome was rich in *Prevotella* (> 50%) and *Bacteroides* (> 20%). Moreover, it also contained *Fusobacterium*, *Megamonas*, *Sutterella*, and other fecal-related genera, representing each one of them less than 5% of the total bacterial composition (Additional File 2).

The metagenomics assembly with the HMW-DNA dataset was more contiguous, presenting fewer and longer contigs than the non-HMW DNA one (contigs: 1898 vs. 2944; N50: 187,680 vs. 94,109 bp) (Additional File 3). Moreover, HMW-DNA metagenomics assembly retrieved six HQ MAGs, yet only one HQ MAG was retrieved from the non-HMW DNA assembly (Fig. 2 and Additional File 3).

In summary, HMW-DNA extraction improved the taxonomic classification of the raw unassembled reads (less unclassified reads), the metagenomics assembly contiguity, and the retrieval of longer and circular contigs (potential HQ MAGs). Thus, HMW-DNA extraction becomes the

A) HMW DNA metagenome assembly



B) Non-HMW DNA metagenome assembly

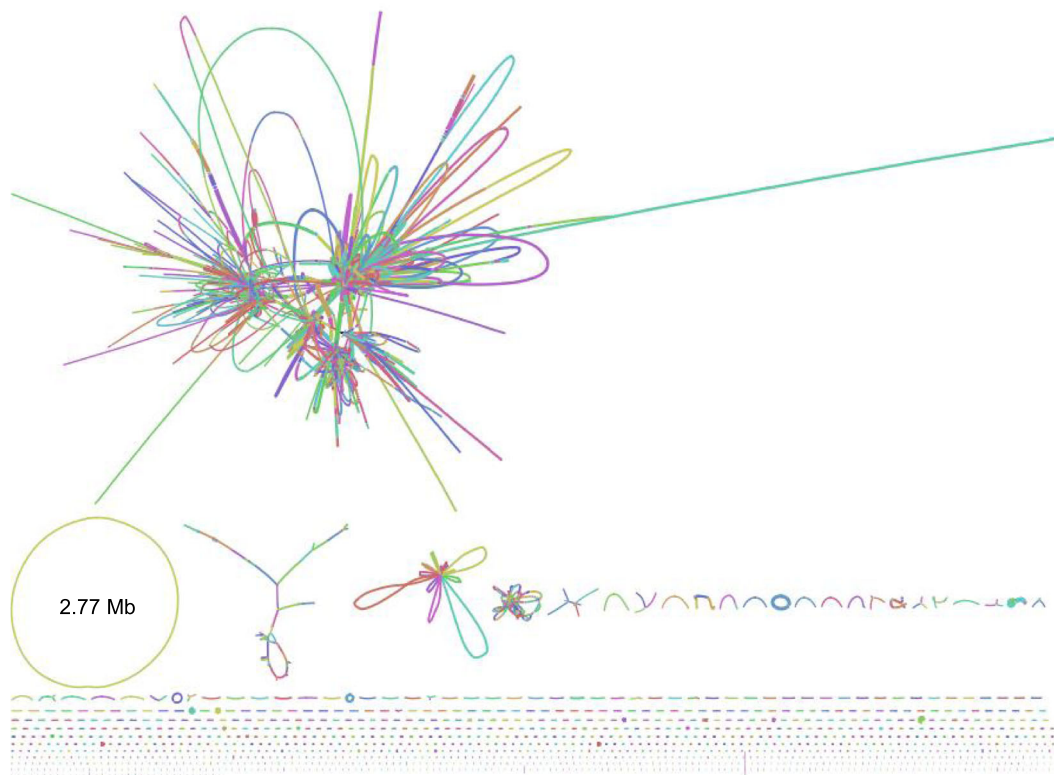


Fig. 2 HMW-DNA vs. non-HMW DNA metagenomics assembly from the fecal sample of a healthy dog. Bandage plots of **a)** HMW-DNA assembly and **b)** non-HMW DNA assembly. HMW-DNA allows the recovery of long, circular contigs, which can potentially represent complete closed MAGs. We report the longest contigs in both datasets (Mb)

preferred choice to recover HQ MAGs directly from complex metagenomics samples.

Metagenomics assembly with different subsets followed by frameshift aware correction retrieved eight high-quality MAGs

To ensure the highest coverage and consensus accuracies for the retrieved MAGs, we further merged and assembled the HMW and the non-HMW datasets (100% dataset; 16.94 million reads, 36.05 Gbp). As we aimed to retrieve the maximum number of HQ MAGs, we performed extra metagenomics assemblies using 75 and 50% data subsets from that merged dataset (Additional File 3).

After assigning taxonomy and comparing among assemblies, we identified non-redundant MAGs: eight HQ MAGs, and three MQ MAGs (Table 1). When compared to HMW assembly, we retrieved two new MQ MAGs from the 100% data assembly (the HMW and the non-HMW datasets together). Moreover, two MQ MAGs from HMW and 100% datasets were recovered as HQ MAGs from the 75% dataset. None of the performed assemblies alone retrieved the eight HQ MAGs.

For each HQ MAG, we selected the representative with the highest coverage –and subsequent highest consensus accuracy– for further analyses. We performed an extra step of frameshift aware correction that reduced the insertions and deletions (indels), which are the most abundant nanopore sequencing error type. The frameshift correction resulted in fewer predicted coding sequences (CDS) (Fig. 3, and Additional File 4). This correction step transformed two MQ MAGs into HQ

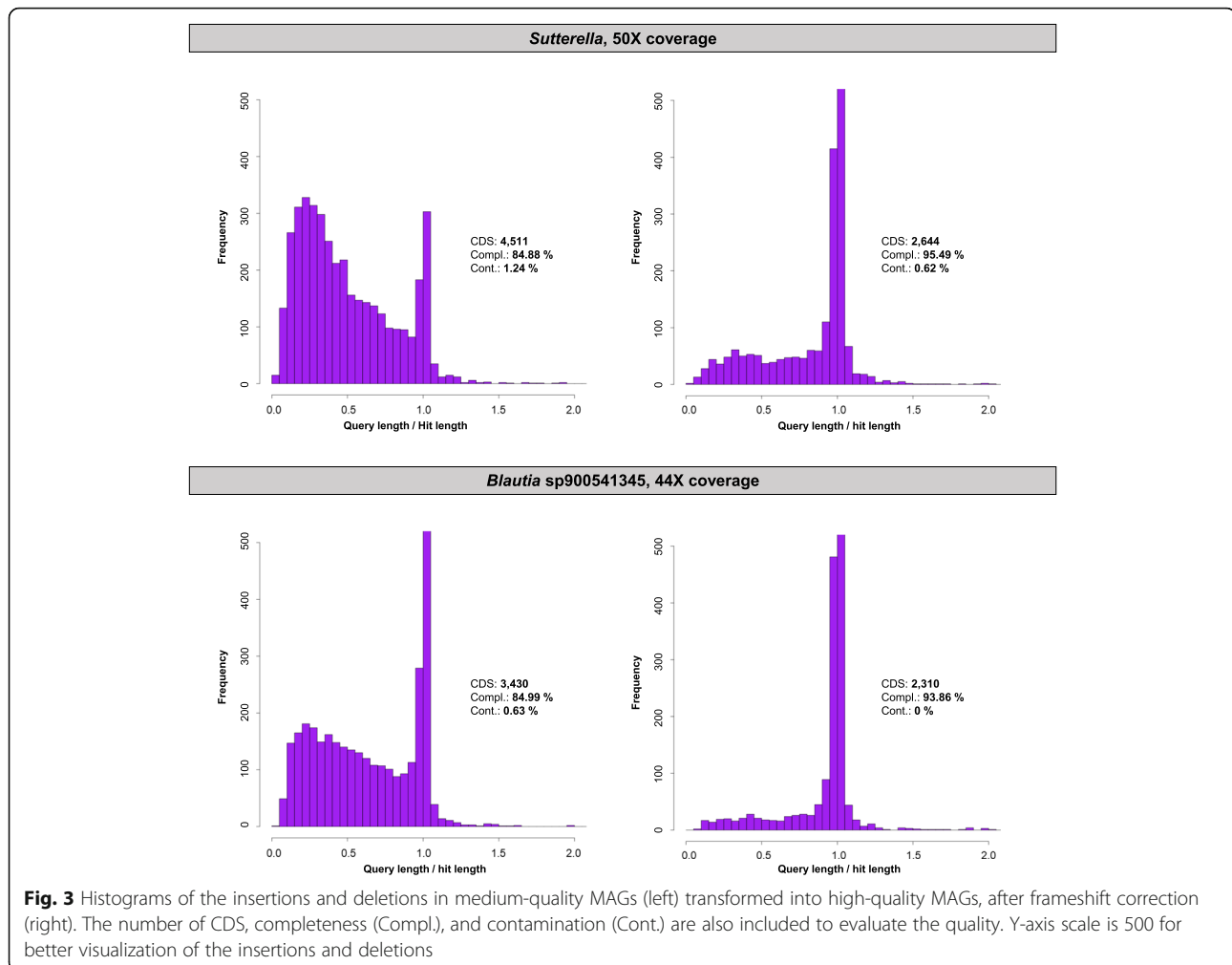
MAGs: *Blautia sp900541345* on the HMW-only assembly (from MQ MAG with 84.99% completeness to HQ MAG with 93.86% completeness) and the *Sutterella* MAG on the 75% assembly (from MQ MAG with 84.88% completeness to HQ MAG with 95.49% completeness) (Fig. 3). On the other HQ MAGs, completeness remained constant or increased after applying the frameshift correction, except for one of the contigs (*Enterococcus hirae*, 47X coverage; completeness of 99.69 to 99.13% after the indel correction). In all the HQ MAGs, contamination value was maintained or reduced after the frameshift correction step (Fig. 3, and Additional File 4). The differences in applying frameshift correction were more evident in contigs with low coverage than in those with high coverage.

High-quality MAGs of the canine fecal microbiome improved previous genome assemblies

From a single canine fecal sample, we obtained eight HQ MAGs regarding MIMAG criteria [27]: > 90% completeness, < 5% contamination, and contained the ribosomal genes (presence of 16S, 23S and 5S rRNA genes) and at least 18 canonical tRNAs. Moreover, all the HQ MAGs were single-contig, and two of them predicted to be circular (Table 2). We used GTDB-tk to assign the taxonomy and assess the potential novelty. The ANI values serve to identify potential novel taxa (> 95% ANI are considered as the same species [31, 40]). Despite *Sutterella* and *Succinivibrio* were considered novel by GTDB-tk, we found one MAG for each in human and dog GI datasets, respectively, that presented > 95% ANI to our HQ MAGs. Similarly, *Prevotellamassilia*

Table 1 High quality (HQ) and medium quality (mq) single-contig MAGs retrieved in each metagenome assembly. Taxonomy assigned using the GTDB database release 95. Q is the MAG quality. Cov. is the coverage from Flye. **Blautia_A sp900541345* and **g_Sutterella* HQ MAGs after correction of the indels

Taxonomy (GTDB)	HMW data		100% data		75% data		50% data	
	Q	Cov.	Q	Cov.	Q	Cov.	Q	Cov.
HQ MAG								
<i>g_Succinivibrio</i>	HQ	47X	HQ	101X	mq	82X	HQ	50X
<i>g_Sutterella*</i>	mq	95X	mq	159X	HQ	123X	mq	87/80X
<i>Prevotellamassilia sp900541335</i>	HQ	394X	HQ	577X	HQ	430X	HQ	282X
<i>Phascolarctobacterium sp900544885</i>	HQ	87X	HQ	205X	HQ	155X	mq	98X
<i>Catenibacterium sp000437715</i>	HQ	13X	mq	24X	HQ	17X	mq	11X
<i>Blautia_A sp003287895</i>	–	–	mq	38X	HQ	31X	mq	18X
<i>Enterococcus_B hirae</i>	HQ	17X	HQ	42X	HQ	31X	HQ	22X
<i>Blautia_A sp900541345*</i>	HQ	44X	–	–	mq	45X	–	–
MQ MAG								
<i>Phocaeicola plebeius</i>	mq	126X	mq	234X	mq	168X	–	–
<i>g_Bacteroides</i>	mq	206X	mq	368X	mq	282X	mq	196X
<i>g_Phocaeicola</i>	–	–	mq	271X	–	–	–	–



sp900541335, *Phascolarctobacterium sp900544885*, *Cate-nibacterium sp000437715*, and *Blautia sp900541345* HQ MAGs were representing bacterial species previously retrieved from metagenomes. In contrast, *Enterococcus_B hirae* and *Blautia sp003287895* HQ MAGs were representing bacterial species that have complete reference genomes. In fact, *Blautia sp003287895* –proposed name *Blautia argii*– was first isolated and characterized from dog feces [41]. *Enterococcus_B hirae* and *Blautia sp003287895* HQ MAGs were aligned against their respective reference genomes to prove and validate the results (Additional File 5).

As we are working with nanopore-only assemblies, we can expect some uncorrected frameshift errors that lead to a larger number of pseudogenes. When compared to each representative genome, our HQ MAGs presented a higher percentage of pseudogenes in all the cases but *B. argii* HQ MAG (Table 2). More pseudogenes can be linked to the higher insertion and deletion errors from Nanopore sequencing that lead to frameshift mutations, when compared to short-read derived MAGs. It is worth

to note that in *Prevotellamassilia* HQ MAG, the % of pseudogenes is highly similar to that found in its representative genome, suggesting that a higher coverage, provides a better consensus with less frameshift errors. For *B. argii* HQ MAG, the frameshift-aware correction software may have over-corrected some real pseudogenes.

Screening of previous microbiome studies revealed the first potential genome assembly for *Sutterella stercoricanis*

We assessed the prevalence of the HQ MAGs retrieved in the present study among several GI microbiome surveys, either using whole-genome data (metagenome surveys) or the 16S rRNA genes data (amplicon surveys).

On the one hand, we assessed the prevalence of our HQ MAGs in humans' [36] and animals' [10] gastrointestinal metagenome catalogs (Table 3). We identified that some of the bacterial species represented by the HQ MAGs from this study seem to be more canid-specific – *Blautia_A sp900541345*, *Phascolarctobacterium sp900544885*, *Prevotellamassilia sp900541335*, *Succinivibrio* –, whereas

Table 2 Summary of genome statistics for High-quality MAGs compared to representatives on the public datasets. Coverage (Cov.) and circularity (Circ.) retrieved from Flye; completeness (% Compl.), from CheckM; tRNAs and rRNA values, from PROKKA. tRNAs count refers to unique canonical tRNAs. GTDB species representative are used as the references for comparison. The two exceptions are *Succinivibrio* and *Sutterella* since they were potential novel species regarding GTDB, but we found a MAG > 95% ANI on the animal gut metagenome and UHGG catalog, respectively. ^apartial gene call by PROKKA. ^bnot detected by PROKKA, but the GTDB/NCBI reference for *Blautia argii* (GCF_003287895.1) is described to contain five 5S rRNA genes

	Length (Mbp)	Cov.	Circ.	% Compl.	CDS	tRNAs	16S rRNA	23S rRNA	5S rRNA	% pseudo genes	Contiguity level
<i>Succinivibrio</i> sp.	2.04	101X	N	98.68	1683	20	7	7	8	0.83%	1 contig
Animal Gut Mg: Freddie_038 MAG	1.74	–	N	97.50	1490	14	0	0	0	0.20%	185 contigs
<i>Sutterella</i> sp.	2.70	123X	N	95.49	2646	19	9	9	0	3.86%	1 contig
UHHG: MGYG_HGUT_01574 MAG	1.14	–	N	78.72	1012	14	0	0	0	0.69%	24 contigs
<i>Prevotellamassilia</i> sp900541335	2.72	577X	Y	97.65	2037	20	7	7	7	0.15%	1 contig
GTDB Rep: GCA_900541335.1	2.42	13X	N	96.13	1842	16	0	0	0	0.11%	95 contigs
<i>Phascolarctobacterium</i> sp900544885	2.09	205X	N	99.85	1985	20	5	5	5	1.61%	1 contig
GTDB Rep: GCA_900544885.1	1.75	20X	N	98.65	1610	18	1 ^a	0	0	0.68%	87 contigs
<i>Catenibacterium</i> sp000437715	2.53	17X	N	98.50	2565	20	7	7	7 ^a	1.83%	1 contig
GTDB Rep: GCF_004168205.1	2.54	45X	N	100	2003	20	1 ^a	0	0	0.90%	212 contigs
<i>Blautia</i> sp900541345	2.44	44X	N	93.86	2313	19	6	6	6	1.30%	1 contig
GTDB Rep: GCA_900541345.1	2.69	8X	N	95.85	2499	16	0	0	0	0.72%	160 contigs
<i>Enterococcus_B</i> hiraе	2.78	42X	Y	99.13	2511	20	6	6	6	1.27%	1 contig
GTDB Rep: GCF_000271405.2	2.83	–	Y	99.63	2670	20	6	6	6	0.97%	Complete
<i>Blautia</i> sp003287895 (<i>Blautia argii</i>)	2.96	31X	N	92.78	2698	20	5	5	0 ^b	0.59%	1 contig
GTDB Rep: GCF_003287895.1	3.30	217X	Y	97.64	3203	20	5	5	0 ^b	2.06%	Complete

others are more broadly distributed among animal microbiomes –*Catenibacterium* sp000437715, *Enterococcus_B* hiraе, *Blautia* sp003287895, and *Sutterella*–.

On the other hand, we took advantage of the fact that long-read sequencing allows retrieving complete ribosomal genes, which are universal taxonomic markers for Bacteria. So, we further extracted the 16S rRNA genes of

the HQ MAGs to link them to 16S rRNA gene-based microbiome studies (Fig. 4, and Additional File 6) –most of the microbiome studies use this genetic marker. We found out that the *Sutterella* HQ MAG is potentially the first high-quality genome assembly for *Sutterella stercoricanis* since its 16S rRNA genes presented identities > 98% with the previously reported 16S rRNA gene

Table 3 Prevalence of the bacterial species identified in public microbiome surveys. For human-derived MAGs, the Unified Human Gut Genome database was used [36]. For animal-derived MAGs, the animal gut metagenome catalog [10] was used. If no MAG belonged to that bacterial species, we further screened GTDB [29]. For further detail on 16S rRNA gene phylogenies, see Additional File 6.

HQ MAG	Dog	Human	Other animals	Closest 16S	Main host
<i>Blautia_A</i> sp900541345	35	1	0	Human gut	Dog
<i>Phascolarctobacterium</i> sp900544885	12	1	0	Dog gut	Dog
<i>Prevotellamassilia</i> sp900541335	7	1	0	Wolves' gut	Canids
<i>g__Succinivibrio</i>	1	0	0	Wolves' gut	Canids
<i>Catenibacterium</i> sp000437715	27	691	2	Human gut	Human, animal
<i>Enterococcus_B</i> hiraе	1	35	3	Multiple	Human, animal
<i>Blautia</i> sp003287895	1	6	1	Dog gut	Human, animal
<i>g__Sutterella</i>	0	1	0	Multiple carnivora	Human, animal

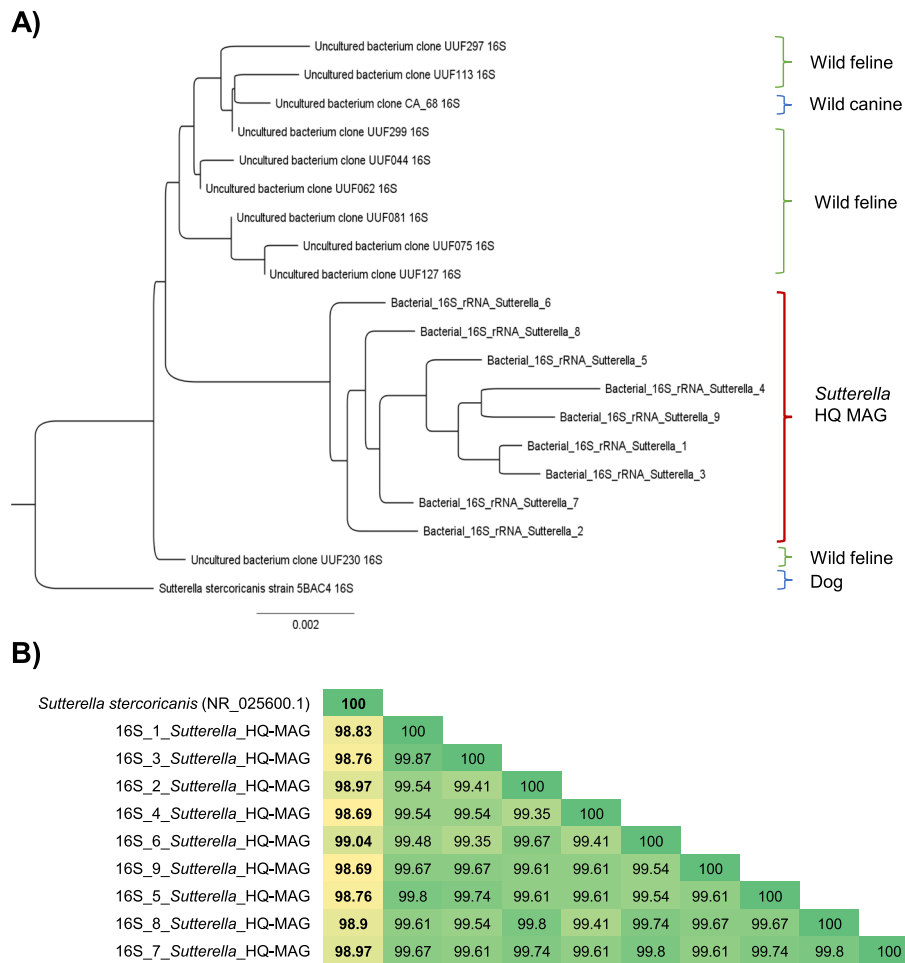


Fig. 4 Similarity of 16S rRNA gene from Sutterella HQ MAGs to public datasets. The 16S rRNA gene comparison from Sutterella HQ MAGs suggested it is the genome assembly for Sutterella stercoricanis. **a** Phylogenetic 16S rRNA gene tree of Sutterella HQ MAGs. It presents high similarity to uncultured bacterium clone with codes UUF from Panthera uncia (wild feline); uncultured bacterium clone CA_68 from Cuon alpinus (wild canid) (JN559525.1), and S. stercoricanis from dog feces [42]. **b** Identity matrix of 16S rRNA genes of Sutterella HQ MAG against S. stercoricanis (NR_025600.1). Sutterella HQ MAG contained nine 16S rRNA genes that were more than 98% identical to NR_025600.1 (reference). Specifically, 16S_6 presented more than 99% of identity

reference (NR_025600.1) (Fig. 4). S. stercoricanis was first isolated in feces from a healthy dog and was characterized using microbiological methods and 16S rRNA gene sequencing [42].

For the other five HQ MAGs without a reference genome, we identified that their 16S rRNA genes were closely related to others previously identified in wolves' distal gut microbiome [43] (Succinivibrio HQ MAG and Prevotellamassillia HQ MAG), canine intestinal microbiome [44] (Phascolarctobacterium HQ MAG), and human GI microbiome [45] (Catenibacterium and Blautia sp900541345 HQ MAG) (Additional File 6).

Finally, we performed a pangenome analysis among the HQ MAGs from our study and other genomes from the same bacterial species inhabiting different hosts to assess functional and genomic similarities (Additional File 7). We included only those in which more

than 10 representative genomes were available: Blautia_A sp900541345 (Additional File 7A), Catenibacterium sp000437715 (Additional File 7B), Enterococcus_B hirae (Additional File 7C), and Phascolarctobacterium sp900544885 (Additional File 7D). Based on the ANI values, the HQ MAGs clustered with dog MAGs for Blautia, with a human MAG for Phascolarctobacterium, and with MAGs from mixed host origins for Catenibacterium and Enterococcus hirae (Additional File 7). The number of gene clusters belonging to the accessory genome was the highest for Catenibacterium (84%) when compared to Enterococcus hirae (66%), Phascolarctobacterium sp900544885 (60%), and Blautia_A sp900541345 (50%). Altogether, these results coincide with the fact that Catenibacterium and Enterococcus hirae seem to be more broadly distributed among different hosts (Table 3).

Long reads provide genomic context and enable capturing mobilome functions and antimicrobial-resistant genes

Long reads enable to retrieve complete genes and their genomic context within a single read. Therefore, both the mobile genetic elements and the antimicrobial resistance genes assemble easily within the correct MAG.

We compared each HQ MAG's functional potential to previously published MAGs from the same bacterial species found in GI microbiome of dogs, humans, or other animals (Fig. 5). The main difference between the long-read HQ MAGs and other genomes from the same species in the public database is the overrepresentation of the COG category corresponding to Mobilome, except for *B. argii* and *E. hirae*, both with a reference genome in the database (Fig. 5b). Conversely to the MAGs from both UHGG and the animal gut metagenome catalogs obtained using exclusively short reads, the long-read metagenomic approach can retrieve mobile genetic elements and assemble them to the proper contig.

Finally, we further characterized the HQ MAGs to assess their potential antimicrobial resistance. Tetracycline resistance genes were detected in *E. hirae* (*tetM* gene), *Catenibacterium* sp000437715 (*tetM* gene), *Blautia* sp900541345 (*tet(O)* gene), and *Blautia* sp003287895 (*tet(32)* and *tet(40)* genes). Moreover, *E. hirae* also harbored *aac(6)-Iid* gene conferring resistance to aminoglycosides. *Prevotellamassilia* HQ MAG harbored *Mef* (*En2*) gene, which encodes for an efflux pump that exports macrolides. *Phascolarctobacterium* HQ MAG harbored two copies of *lnu(C)* gene conferring resistance to lincosamide. Each *lnu(C)* gene was located in an *ISSag10* mobile element, allowing it to transpose. *Succinivibrio* and *Sutterella* HQ MAGs did not contain any antimicrobial resistance genes.

As an example of the potential of long-reads for providing genomic context, we were able to identify that *tetM* gene in *E. hirae* was in a region identified as a conjugative element (Tn916) integrated into the chromosome. This region encoded for a transposase, type 4 secretion system (T4SS), type 4 coupling protein, *oriT*, and relaxase (Additional File 8).

Discussion

Metagenomics approaches can provide new biological insights into the microbes inhabiting underexplored environments, such as the canine fecal microbiome. Here, we applied nanopore long-read metagenomics and frameshift aware correction to a fecal sample of a healthy dog and retrieved eight HQ MAGs and three MQ MAGs.

At the technical level, we compared a HMW and non-HMW DNA extraction to perform long-read metagenomics and confirmed that a HMW-DNA extraction

was the best choice. For analyses using unassembled raw reads, it improved the taxonomic classification and displayed less unclassified reads. For metagenomics assembly, it improved the contiguity and increased the retrieval of longer and circular contigs (potential HQ MAGs). This is in line with recent studies on human fecal microbiome, where they used HMW-DNA extraction together with long-read metagenomics to recover high-quality MAGs [11, 46].

We tested several metagenomics assembly strategies (using HMW data only, 100, 75, and 50% of the total data) to retrieve the highest number of different HQ MAGs. The HMW data and the 75% data retrieved the highest number of HQ MAGs, but none of the performed assemblies alone retrieved the eight HQ MAGs.

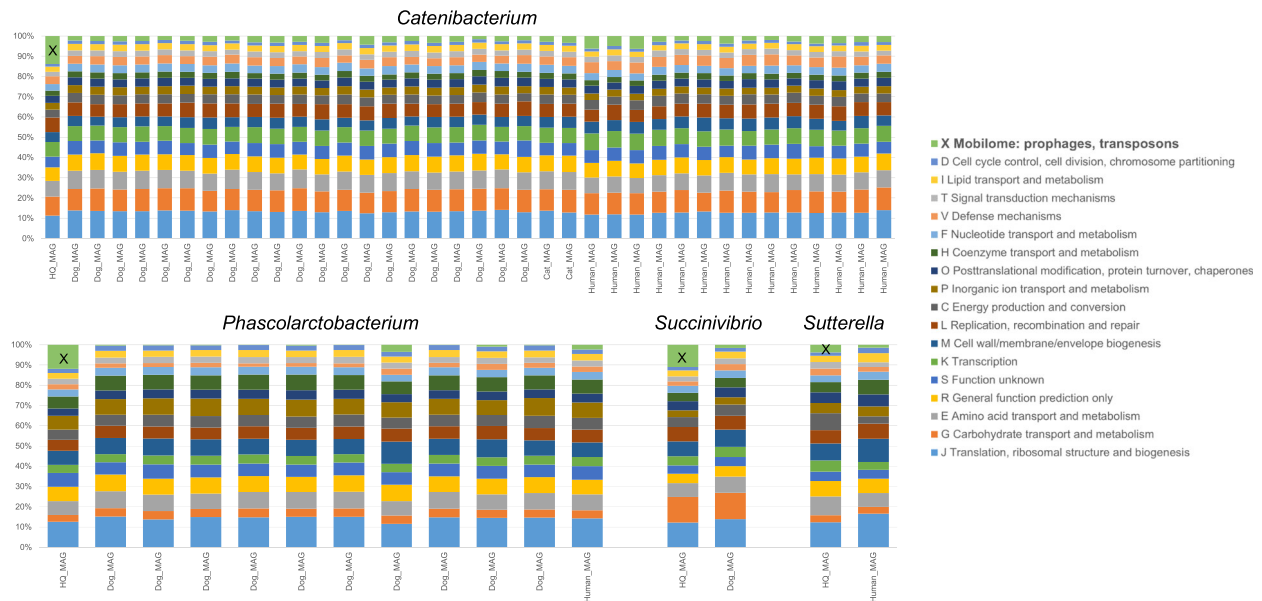
The HQ MAGs belonged to *Succinivibrio*, *Sutterella*, *Prevotellamassilia*, *Phascolarctobacterium*, *Enterococcus*, *Blautia*, and *Catenibacterium* genera. The HQ MAGs presented > 90% completeness and < 5% contamination, improved the contiguity of previous MAGs in databases (single-contigs vs. multiple contigs), and contained the full-length ribosomal genes. Thus, our MAGs met MIMAG criteria for high-quality [27]. This fact is challenging when using shotgun metagenomics (with short-read technologies).

For *Sutterella* HQ MAG, we suggest that it is potentially the first reported high-quality genome assembly for *Sutterella stercoricanis*, which can be used as a representative genome for this bacterial species. It was first isolated in feces from a healthy dog, and it was defined as a novel species phenotypically and with full-length 16S rRNA sequencing [42]. Since the reference isolate lacks additional genome information, we compared the full-length 16S rRNA gene sequences to identify the bacterial species. Both the classical threshold of 97% identity and the updated one of 99% identity were met in this case [47]: the nine 16S rRNA genes presented identities from 99.04 to 98.69% against *S. stercoricanis* 16S rRNA (NR_025600.1). Whole-genome sequencing of the reference isolate and comparison to the HQ MAG could confirm if they represent the same species.

Despite humans and dogs share similar microbial signatures on the GI microbiome [6, 7], we found that *Succinivibrio*, *Prevotellamassilia* sp900541335, *Phascolarctobacterium* sp900544885, *Blautia_A* sp90054134 seem more candidate-specific, whereas *Sutterella*, *Catenibacterium* sp000437715, *Enterococcus_B hirae*, and *Blautia* sp003287895 are more broadly distributed among human and animal gastrointestinal microbiomes. These findings highlight the need for building and using niche-specific databases to accurately map and classify new reads from a particular environment and understand the overall biological significance [13, 48].

The genera *Succinivibrio*, *Prevotella*, *Phascolarctobacterium*, *Catenibacterium*, and *Blautia*, are recognized

A) Functional profile considering the more abundant COG categories



B) Mobilome COG category distribution

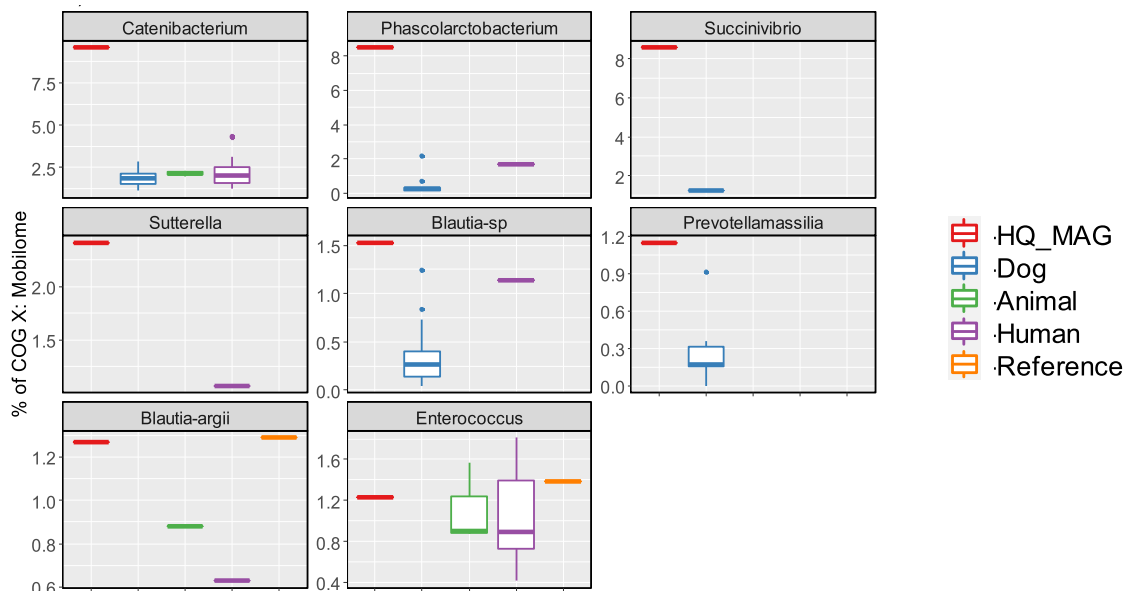


Fig. 5 Functional analysis and comparison of HQ MAGs and published bacterial species using the COG database. **a** Stacked bar plots representing the 18 more abundant COG categories for *Catenibacterium*, *Phascolarctobacterium*, *Succinivibrio*, and *Sutterella* representatives (HQ MAGs with more Mobilome functions). Y-axis is escalated to 100% for visualization proposes. **b** Boxplots representing the actual percentage of mobilome COG category in each of the HQ MAGs and the bacterial species present in the databases grouped by origin. Y-axis represents different scales

short-chain fatty acid (SCFA) producers [49–51]. These molecules provide multiple gut health benefits, from reducing inflammation and tumorigenesis to increasing gut motility and secretory activity [2, 50, 52]. In the dog GI microbiome, different diets and dietary interventions

can modulate their abundances to promote gut health [7, 53–58]. Moreover, several studies on dog GI microbiome identified *Blautia* genus –among others– as a microbial marker for health and had targeted it to assess differences with disease status [59–62]. Thus, in-depth

characterization of these genera is of most relevance to defining a healthy GI microbiome in dogs.

S. stercoricanis was isolated from the feces of a healthy dog [42]. However, the increase of the genus *Sutterella* was associated with detrimental effects rather than health. Dogs with acute hemorrhagic diarrhea presented higher *Sutterella* [59], and some diets aiming to promote health benefits observed its decrease [63, 64]. Further metagenomics studies are needed to identify the different *Sutterella* species on dog feces and correlate their abundances to health or disease status.

Finally, *E. hirae* is a prevalent Enterococci species of the GI microbiome of healthy dogs. However, Enterococci species usually carry antimicrobial-resistant genes and virulence factors and are potential antimicrobial-resistant gene reservoirs that could be transferred to people [65–69]. *Enterococcus* HQ MAG harbors *aac* (6′)-*lid* gene, which conferred resistance to aminoglycosides. Besides, it harbors a *tetM* gene within the Tn916 conjugative element, which was first reported in *Enterococcus faecalis* [70, 71]. The use of long-reads enables the retrieval of complete genes and their genomic context within a single read, facilitating the location of antimicrobial resistance genes within the proper MAG and the evaluation of its mobilization mechanisms [72, 73].

Tetracycline resistance genes were found not only in the genome of *E. hirae*, but also in *Catenibacterium* and both *Blautia* HQ MAGs and could be linked to a previous antimicrobial exposure that selected the resistant bacteria [74]. Three years before sampling, this dog was treated with doxycycline –tetracycline-class antibiotic– for 15 days due to excess secretion of mucus and saliva. Whole resistome analyses are needed to determine the antimicrobial-resistant genes within the fecal microbiome in healthy dogs and evaluate all the bacterial species and their mobile genetic elements that could act as a reservoir for these genes.

At the functional level, we detected an overrepresentation of the Mobilome COG category within most of the HQ MAGs retrieved here when compared to other MAGs –not when compared to reference genomes. Long-reads allow retrieving complete mobile genetic elements together with their genomic context facilitating its assembly to the proper MAG. This advantage was also reported in metagenomics studies that include short- and long-reads in their assemblies (hybrid assemblies) [11, 13, 75].

Apart from eight HQ MAGs, we recovered three different MQ MAGs from potentially new species of the *Bacteroides* and *Phocaeicola* genera and *Phocaeicola plebeius*. Our next step is to apply proximity ligation to link all contigs among them and recover new HQ MAGs and MQ MAGs and link antimicrobial resistance genes,

mobile genetic elements, and bacteriophages to their bacterial host [76].

A limitation of this study is the use of nanopore-only data since it can compromise the accuracy of the HQ MAGs, and the use of short-read polishing could have further improved the sequence accuracy. However, the combination of high-accuracy basecallers and raw reads correction, followed by further polishing of the metagenome assemblies increased the consensus accuracy to levels suitable to retrieve high-quality MAGs from a single fecal sample. In our case, we applied Guppy for basecalling, Canu for raw reads correction, and Medaka for polishing the assembled metagenomes. To reduce the insertion and deletion error type, we further applied a frameshift-aware correction step [16] that improved the completeness, and decreased contamination and number of CDS. Despite more pseudogenes –caused by frameshift mutations– are observed when compared to representative genomes, the MAGs retrieved here were high-quality regarding MIMAG criteria [27], presenting at least 18 unique tRNAs, the ribosomal genes (16S, 23S, and 5S rRNAs), mobilome functions, altogether within a single contig.

Abbreviations

GI: Gastrointestinal; HMW: High-molecular weight DNA; Non-HMW: Non high-molecular weight DNA; MAG: Metagenome-assembled genome; HQ MAG: High-quality metagenome-assembled genome; MQ MAG: Medium-quality metagenome-assembled genome; Indels: Insertions and deletions

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07607-0>.

Additional File 1. Bioinformatics workflow overview. The file contains information on the software used and their versions and the commands and the options to perform the bioinformatics analysis used here.

Additional File 2. Genus-level taxonomic classification of a canine fecal metagenome using HMW reads and non-HMW reads. Taxonomic classification was performed with Kraken2 and MAXI_DB, as stated in materials and methods. The Sankey diagram represents the data in the Table below.

Additional File 3. Flye assembly summary statistics and the number of the final number of HQ and MQ MAGs for each metagenome assembly. HQ: high-quality; MQ: medium-quality.

Additional File 4. Histograms of the indels in high-quality MAGs before (left) and after (right) correction. The number of CDS, completeness, and contamination are also included to evaluate the quality. Y-axis scale is 500 for better visualization of the indels.

Additional File 5 Whole-genome alignment dot plots for HQ MAGs against its ‘complete’ reference genome. *Enterococcus* HQ MAG against *Enterococcus hirae* str. ATCC 9790 (GCF_000271405.2) and *Blautia* HQ MAG against *Blautia* N6H1–15 (GCF_003287895.1).

Additional File 6. Phylogenetic 16S rRNA gene tree of HQ MAGs. The phylogenetic trees were computed using MOLE-BLAST against nt/nr database and including uncultured and environmental taxa. The 16S rRNA genes from our HQ MAGs are indicated with a green line.

Additional File 7 Pangenome visualization of gastrointestinal microbes from different hosts. In A) *Blautia_A* sp900541345; in B) *Catenibacterium* sp000437715; in C) *Enterococcus_B* *hirae*; and in D) *Phascolarctobacterium*

sp900544885. Blue: Dog_MAG from [10], Violet: Human_MAG from [36], Green: Animal_MAG from [10], Pink: Dog_HQ_MAG (this study). The dendrogram in the center is ordered by gene cluster presence/absence. The dendrogram in the right up corner clustering is ordered by ANI percentage identity. CORE: gene clusters shared by all the representatives. ACCESSORY: gene clusters shared by some of the representatives. SING LETON: unique gene clusters, exclusive to a single representative.

Additional File 8 *Enterococcus hirae* conjugative element: transposon Tn916. Genetic elements identified by OriTFinder, which coincided with predicted ORFs by Prokka, were highlighted in different colors: orange for the transposase (tnp) of Tn916 element; red for antibiotic resistance genes (*tet(M)*); blue for conjugative elements (T4SS, type IV secretion system); pink for the relaxase; green for the type IV coupling protein (T4CP); and grey for hypothetical proteins (hp).

Acknowledgments

We would like to acknowledge Amanda Warr from the University of Edinburgh for her advice and support on the long-read metagenomics pipeline.

Authors' contributions

OF and AC conceptualized the study. OF and AC designed the experiment. DP extracted the DNA, performed the sequencing libraries and the nanopore sequencing. AC performed the metagenome assembly and correction. AC analyzed and interpreted the data. JV analyzed the antimicrobial-resistance genes. NF and DP performed the pangenome analysis. AC wrote the main manuscript text. OF, NF, DP and JV substantially revised the work. All the authors have approved the submitted version.

Funding

Vetgenomics and Molecular Genetics Veterinary Service (SVGM), Universitat Autònoma de Barcelona.

Availability of data and materials

The draft metagenome assemblies, the HQ MAGs, and an overview of the scripts used are available on Zenodo: <https://zenodo.org/record/3982645>. The raw fast5 files and the HQ MAGs are available on ENA under bioproject PRJEB42270. An overview of the scripts used to analyze the data is at Additional File 1.

Declarations

Ethics approval and consent to participate

The owner gave written consent to the collection of samples.

Consent for publication

not applicable.

Competing interests

AC, NF, and JV work for Vetgenomics, SL. The other authors declare that they have no competing interests.

Author details

¹Vetgenomics, Ed Eureka, Parc de Recerca UAB, Barcelona, Spain. ²Molecular Genetics Veterinary Service (SVGM), Veterinary School, Universitat Autònoma de Barcelona, Barcelona, Spain.

Received: 25 December 2020 Accepted: 12 April 2021

Published online: 06 May 2021

References

- Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol*. 2017;2(11):1533–42. <https://doi.org/10.1038/s41564-017-0012-7>.
- Pilla R, Suchodolski JS. The role of the canine gut microbiome and Metabolome in health and gastrointestinal disease. *Front Vet Sci*. 2020;6. <https://doi.org/10.3389/fvets.2019.00498>.
- Redfern A, Suchodolski J, Jergens A. Role of the gastrointestinal microbiota in small animal health and disease. *Vet Rec*. 2017;181(14):370. <https://doi.org/10.1136/vr.103826>.
- Schmitz S, Suchodolski J. Understanding the canine intestinal microbiota and its modification by pro-, pre- and synbiotics – what is the evidence? *Vet Med Sci*. 2016;2(2):71–94. <https://doi.org/10.1002/vms3.17>.
- Honneffer JB, Minamoto Y, Suchodolski JS. Microbiota alterations in acute and chronic gastrointestinal inflammation of cats and dogs. *World J Gastroenterol*. 2014;20(44):16489–97. <https://doi.org/10.3748/wjg.v20.i44.16489>.
- Vázquez-Baeza Y, Hyde ER, Suchodolski JS, Knight R. Dog and human inflammatory bowel disease rely on overlapping yet distinct dysbiosis networks. *Nat Microbiol*. 2016;1:1–5.
- Coelho LP, Kultima JR, Costea PI, Fournier C, Pan Y, Czarnecki-Maulden G, et al. Similarity of the dog and human gut microbiomes in gene content and response to diet. *Microbiome*. 2018;6(1):72. <https://doi.org/10.1186/s40168-018-0450-3>.
- Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun*. 2016;469(4):967–77. <https://doi.org/10.1016/j.bbrc.2015.12.083>.
- Swanson KS, Dowd SE, Suchodolski JS, Middelbos IS, Vester BM, Barry KA, et al. Phylogenetic and gene-centric metagenomics of the canine intestinal microbiome reveals similarities with humans and mice. *ISME J*. 2011;5(4):639–49. <https://doi.org/10.1038/ismej.2010.162>.
- Youngblut ND, de la Cuesta-Zuluaga J, Reischer GH, Dauser S, Schuster N, Walzer C, et al. Large-Scale Metagenome Assembly Reveals Novel Animal-Associated Microbial Genomes, Biosynthetic Gene Clusters, and Other Genetic Diversity. *mSystems*. 2020;5:e01045–20.
- Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol*. 2020;38(6):701–7. <https://doi.org/10.1038/s41587-020-0422-6>.
- Nicholls SM, Quick JC, Tang S, Loman NJ. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience*. 2019;8:giz043.
- 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(8):953–61. <https://doi.org/10.1038/s41587-019-0202-3>.
- Somerville V, Lutz S, Schmid M, Frei D, Moser A, Irmeler S, et al. Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *BMC Microbiol*. 2019;19(1):143. <https://doi.org/10.1186/s12866-019-1500-0>.
- Singleton CM, Petriglieri F, Kristensen JM, Kirkegaard RH, Michaelsen TY, Andersen MH 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(1):1–13. <https://doi.org/10.1038/s41467-021-22203-2>.
- Arumugam K, Bağcı C, Bessarab I, Beier S, Buchfink B, Górka A, et al. Annotated bacterial chromosomes from frame-shift-corrected long-read metagenomic data. *Microbiome*. 2019;7(1):61. <https://doi.org/10.1186/s40168-019-0665-y>.
- Cuscó A, Salas A, Torre C, Francino O. Shallow metagenomics with Nanopore sequencing in canine fecal microbiota improved bacterial taxonomy and identified an uncultured CrAssphage. *Biorxiv*. 2019:585067.
- Wood DE, Lu J, Langmead B. Improved metagenomic analysis with kraken 2. *Genome Biol*. 2019;20(1):257. <https://doi.org/10.1186/s13059-019-1891-0>.
- Breitwieser FP, Salzberg SL. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics*. 2020; 36(4): 1303–1304.
- Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*. 2016;11(10):e0163962. <https://doi.org/10.1371/journal.pone.0163962>.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017;27(5):722–36. <https://doi.org/10.1101/gr.215087.116>.
- Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods*. 2020;17(11):1103–10. <https://doi.org/10.1038/s41592-020-00971-x>.

23. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics Oxford Academic*. 2015;31(20):3350–2. <https://doi.org/10.1093/bioinformatics/btv383>.
24. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60. <https://doi.org/10.1038/nmeth.3176>.
25. Huson DH, Albrecht B, Bağcı C, Bessarab I, Górška A, Jolic D, et al. MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. *Biol Direct*. 2018;13(1):6. <https://doi.org/10.1186/s13062-018-0208-7>.
26. 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(7):1043–55. <https://doi.org/10.1101/gr.186072.114>.
27. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol*. 2017;35(8):725–31. <https://doi.org/10.1038/nbt.3893>.
28. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics*. 2020;36:1925–7.
29. 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(9):1079–86. <https://doi.org/10.1038/s41587-020-0501-8>.
30. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068–9. <https://doi.org/10.1093/bioinformatics/btu153>.
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(1):5114. <https://doi.org/10.1038/s41467-018-07641-9>.
32. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*. 2015;3:e1319. <https://doi.org/10.7717/peerj.1319>.
33. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
34. Jia B, Raphenya AR, Alcock B, Wagglechner N, Guo P, Tsang KK, et al. CARD 2014: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res*. 2017;45(D1):D566–73. <https://doi.org/10.1093/nar/gkw1004>.
35. Li X, Xie Y, Liu M, Tai C, Sun J, Deng Z, et al. oriTfinder: a web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. *Nucleic Acids Res*. 2018;46:W229–34.
36. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotech*. 2021;39(1):105–14. <https://doi.org/10.1038/s41587-020-0603-3>.
37. Delmont TO, Eren AM. Linking pangenomes and metagenomes: the Prochlorococcus metapangenome. *PeerJ*. 2018;6:e4320. <https://doi.org/10.7717/peerj.4320>.
38. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11(1):119. <https://doi.org/10.1186/1471-2105-11-119>.
39. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Kooin EV, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*. 2003;4(1):41. <https://doi.org/10.1186/1471-2105-4-41>.
40. Kim M, Oh H-S, Park S-C, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol*. 2014;64(Pt_2):346–51. <https://doi.org/10.1099/ijs.0.059774-0>.
41. Paek J, Shin Y, Kook J-K, Chang Y-H. *Blautia argi* sp. nov., a new anaerobic bacterium isolated from dog faeces. *Int J Syst Evol Microbiol*. 2019;69(1):33–8. <https://doi.org/10.1099/ijsem.0.002981>.
42. Greetham HL, Collins MD, Gibson GR, Giffard C, Falsen E, Lawson PA. *Sutterella stercoricanis* sp. nov., isolated from canine faeces. *Int J Syst Evol Microbiol*. 2004;54(5):1581–4. <https://doi.org/10.1099/ijs.0.63098-0>.
43. Zhang H, Chen L. Phylogenetic analysis of 16S rRNA gene sequences reveals distal gut bacterial diversity in wild wolves (*Canis lupus*). *Mol Biol Rep*. 2010;37(8):4013–22. <https://doi.org/10.1007/s11033-010-0060-z>.
44. 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(3):567–78. <https://doi.org/10.1111/j.1574-6941.2008.00521.x>.
45. Li E, Hamm CM, Gulati AS, Sartor RB, Chen H, Wu X, et al. Inflammatory Bowel Diseases Phenotype, *C. difficile* and NOD2 Genotype Are Associated with Shifts in Human Ileum Associated Microbial Composition. *PLOS ONE*. 2012;7:e26284.
46. 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 Protocols*. Nature Publishing Group. 2021;16(1):458–71. <https://doi.org/10.1038/s41596-020-00424-x>.
47. Edgar RC. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics*. 2018;34(14):2371–5. <https://doi.org/10.1093/bioinformatics/bty113>.
48. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*. 2019;176:649–62 e20.
49. Kageyama A, Benno Y. *Catenibacterium mitsuokai* gen. nov., sp. nov., a gram-positive anaerobic bacterium isolated from human faeces. *Int J Syst Evol Microbiol*. 2000;50(4):1595–9. <https://doi.org/10.1099/00207713-50-4-1595>.
50. Koh A, Vadder FD, Kovatcheva-Datchary P, Bäckhed F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell*. 2016;165(6):1332–45. <https://doi.org/10.1016/j.cell.2016.05.041>.
51. O'Herrin SM, Kenealy WR. Glucose and carbon dioxide metabolism by *Succinivibrio dextrinosolvens*. *Appl Environ Microbiol*. 1993;59(3):748–55. <https://doi.org/10.1128/AEM.59.3.748-755.1993>.
52. Parada Venegas D, De la Fuente MK, Landskron G, González MJ, Quera R, Dijkstra G, et al. Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Front Immunol*. 2019;10:277. <https://doi.org/10.3389/fimmu.2019.00277>.
53. Jarett JK, Carlson A, Rossoni Seroa M, Strickland J, Serfilippi L, Ganz HH. Diets with and without edible cricket support a similar level of diversity in the gut microbiome of dogs. *PeerJ*. 2019;7:e7661. <https://doi.org/10.7717/peerj.7661>.
54. Li Q, Lauber CL, Czarnecki-Maulden G, Pan Y, Hannah SS. Effects of the Dietary Protein and Carbohydrate Ratio on Gut Microbiomes in Dogs of Different Body Conditions. *mBio*. 2017;8:e01703–16. <https://doi.org/10.1128/mBio.01703-16>.
55. Myint H, Iwahashi Y, Koike S, Kobayashi Y. Effect of soybean husk supplementation on the fecal fermentation metabolites and microbiota of dogs. *Anim Sci J*. 2017;88(11):1730–6. <https://doi.org/10.1111/asj.12817>.
56. Schauf S, de la Fuente G, Newbold CJ, Salas-Mani A, Torre C, Abecia L, et al. Effect of dietary fat to starch content on fecal microbiota composition and activity in dogs1. *J Anim Sci*. 2018;96(9):3684–98. <https://doi.org/10.1093/jas/sky264>.
57. Sandri M, Dal Monego S, Conte G, Sgorlon S, Stefanon B. Raw meat based diet influences faecal microbiome and end products of fermentation in healthy dogs. *BMC Vet Res*. 2017;13:65. <https://doi.org/10.1186/s12917-017-0981-z>.
58. Scarsella E, Cintio M, Iacumin L, Ginaldi F, Stefanon B. Interplay between Neuroendocrine Biomarkers and Gut Microbiota in Dogs Supplemented with Grape Proanthocyanidins: Results of Dietary Intervention Study. *Animals (Basel)*. 2020;10:531.
59. Suchodolski JS, Markel ME, Garcia-Mazcorro JF, Unterer S, Heilmann RM, Dowd SE, et al. The fecal microbiome in dogs with acute diarrhea and idiopathic inflammatory bowel disease. *PLoS ONE*. 2012;7(12):e51907. <https://doi.org/10.1371/journal.pone.0051907>.
60. AlShawaqfeh MK, Wajid B, Minamoto Y, Markel M, Lidbury JA, Steiner JM, et al. A dysbiosis index to assess microbial changes in fecal samples of dogs with chronic inflammatory enteropathy. *FEMS Microbiol Ecol*. 2017;93(11):fix136. <https://doi.org/10.1093/femsec/fix136>.
61. Ziese AL, Suchodolski JS, Hartmann K, Busch K, Anderson A, Sarwar F, et al. Effect of probiotic treatment on the clinical course, intestinal microbiome, and toxigenic *Clostridium perfringens* in dogs with acute hemorrhagic diarrhea. *PLoS ONE*. 2018;13(9):e0204691. <https://doi.org/10.1371/journal.pone.0204691>.
62. Minamoto Y, Minamoto T, Isaiah A, Sattasathuchana P, Buono A, Rangachari VR, et al. Fecal short-chain fatty acid concentrations and dysbiosis in dogs

- with chronic enteropathy. *J Vet Intern Med.* 2019;33(4):1608–18. <https://doi.org/10.1111/jvim.15520>.
63. Xu H, Huang W, Hou Q, Kwok L-Y, Laga W, Wang Y, et al. Oral Administration of Compound Probiotics Improved Canine Feed Intake, weight gain, Immunity and Intestinal Microbiota. *Front Immunol.* 2019;10: 666. <https://doi.org/10.3389/fimmu.2019.00666>.
 64. Ide K, Shinohara M, Yamagishi S, Endo A, Nishifuji K, Tochio T. Kestose supplementation exerts bifidogenic effect within fecal microbiota and increases fecal butyrate concentration in dogs. *J Vet Med Sci.* 2020;82(1):1–8. <https://doi.org/10.1292/jvms.19-0071>.
 65. Devriese LA, Colque JIC, Herdt PD, Haesebrouck F. Identification and composition of the tonsillar and anal enterococcal and streptococcal flora of dogs and cats. *J Appl Bacteriol.* 1992;73(5):421–5. <https://doi.org/10.1111/j.1365-2672.1992.tb04998.x>.
 66. Poeta P, Costa D, Rodrigues J, Torres C. Antimicrobial resistance and the mechanisms implicated in faecal enterococci from healthy humans, poultry and pets in Portugal. *Int J Antimicrob Agents.* 2006;27(2):131–7. <https://doi.org/10.1016/j.ijantimicag.2005.09.018>.
 67. Jia W, Li G, Wang W. Prevalence and antimicrobial resistance of Enterococcus species: a hospital-based study in China. *Int J Environ Res Public Health.* 2014;11(3):3424–42. <https://doi.org/10.3390/ijerph110303424>.
 68. Kubašová I, Stropňová V, Lauková A. Safety assessment of commensal enterococci from dogs. *Folia Microbiol.* 2017;62(6):491–8. <https://doi.org/10.007/s12223-017-0521-z>.
 69. Ben Said L, Dziri R, Sassi N, Lozano C, Ben Slama K, Ouzari I, et al. Species distribution, antibiotic resistance and virulence traits in canine and feline enterococci in Tunisia. *Acta Vet Hung.* 2017;65(2):173–84. <https://doi.org/10.1556/004.2017.018>.
 70. Franke AE, Clewell DB. Evidence for a chromosome-borne resistance transposon (Tn916) in *Streptococcus faecalis* that is capable of “conjugal” transfer in the absence of a conjugative plasmid. *J Bacteriol.* 1981;145(1): 494–502. <https://doi.org/10.1128/JB.145.1.494-502.1981>.
 71. Flannagan SE, Zitzow LA, Su YA, Clewell DB. Nucleotide sequence of the 18-kb conjugative transposon Tn916 from *Enterococcus faecalis*. *Plasmid.* 1994; 32(3):350–4. <https://doi.org/10.1006/plas.1994.1077>.
 72. van der Helm E, Imamovic L, Hashim Ellabaan MM, van Schaik W, Koza A, Sommer MOA. Rapid resistome mapping using nanopore sequencing. *Nucleic Acids Res.* 2017;45:e61. <https://doi.org/10.1093/nar/gkw1328>.
 73. Che Y, Xia Y, Liu L, Li A-D, Yang Y, Zhang T. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. *Microbiome.* 2019;7(1):44. <https://doi.org/10.1186/s40168-019-0663-0>.
 74. Casals-Pascual C, Vergara A, Vila J. Intestinal microbiota and antibiotic resistance: perspectives and solutions. *Hum Microbiome J.* 2018;9:11–5. <https://doi.org/10.1016/j.humic.2018.05.002>.
 75. Bertrand D, Shaw J, Kalathiyappan M, Ng AHQ, Kumar MS, Li C, et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat Biotechnol.* 2019;37(8):937–44. <https://doi.org/10.1038/s41587-019-0191-2>.
 76. Stalder T, Press MO, Sullivan S, Liachko I, Top EM. Linking the resistome and plasmidome to the microbiome. *ISME J.* 2019;13(10):2437–46. <https://doi.org/10.1038/s41396-019-0446-4>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

