

# Long-Read Metagenomics to Retrieve High-Quality Metagenome-Assembled Genomes from Canine Feces

Anna Cusco (✉ [anna.cusco@vetgenomics.com](mailto:anna.cusco@vetgenomics.com))

Vetgenomics, SL <https://orcid.org/0000-0002-9574-5755>

**Daniel Perez**

Molecular Veterinary Genetics Service (SVGM)

**Joaquim Viñes**

Vetgenomics, SL

**Olga Francino**

Molecular Genetics Veterinary Service (SVGM)

---

## Research

**Keywords:** Long-read metagenomics, gastrointestinal microbiome, fecal microbiome, long-reads, nanopore, canine microbiome, dog microbiome, metagenome-assembled genomes, Sutterella, Succinivibrio

**Posted Date:** August 19th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-60068/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background.** Metagenomics is a powerful and rapidly developing approach that provides new biological insights into the microbes inhabiting underexplored environments, such as canine fecal microbiome. We investigate long-read metagenomics with Nanopore sequencing to profile the fecal microbiome and to retrieve high-quality metagenome-assembled genomes (HQ MAGs) from a healthy dog.

**Results.** More than 99% of total classified reads corresponded to Bacteria. The most abundant phylum was Bacteroidetes (~80% of total reads), followed by Firmicutes, Proteobacteria, and Fusobacteria. *Prevotella* (>50%) and *Bacteroides* (>20%) are the more abundant genera, followed by *Fusobacterium*, *Megamonas*, *Sutterella*, and other fecal-related genera, (each representing <5% of the total bacterial composition). We retrieved eight single-contig HQ MAGs and three medium-quality MAGs, after combining several metagenome dataset assemblies. The HQ MAGs corresponded to *Succinivibrio*, *Sutterella*, *Prevotellamassilia*, *Phascolarctobacterium*, *Enterococcus*, *Blautia*, and *Catenibacterium* genera. *Succinivibrio* HQ MAG represents a novel candidate bacterial species. *Sutterella* HQ MAG is potentially the first reported genome assembly for *Sutterella stercoricanis*, as assigned by 16S rRNA gene similarity. *Prevotellamassilia*, *Phascolarctobacterium*, *Catenibacterium*, and *Blautia* sp900541345 HQ MAGs improved the contiguity of previously reported genome assemblies in their respective genera, and the number of rRNA genes and tRNA genes. Finally, *Enterococcus hirae* and *Blautia* sp003287895 HQ MAGs represented species that already have a complete reference genome. At the technical level, we demonstrated that a high-molecular weight DNA extraction improved the taxonomic classification of the raw unassembled reads, the metagenomics assembly contiguity, and the retrieval of longer and circular contigs, which are potential HQ MAGs.

**Conclusions.** Long-read metagenomics allowed us to recover HQ MAGs from canine feces of a healthy dog. The high-molecular weight DNA extraction to improve contiguity and the correction of the insertions and deletions to reduce frameshift errors ensure the retrieval of complete single-contig HQ MAGs.

## Background

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

Previous studies reported similarities between canine and human GI microbiome (See [2–5] for extensive reviews). 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). 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 either marker-specific (e.g., 16S rRNA gene for Bacteria) or whole metagenome sequencing [8]. The canine GI microbiome studies published until today (August 2020) use next-generation sequencing – short-read sequencing– or earlier technologies and are mostly amplicon-based strategies (16S rRNA gene). To date, only two studies have used shotgun metagenomics with short-read sequencing to characterize further the whole microbial community and the gene content in dog feces [7,9].

The application of long-read sequencing to metagenomics enables the retrieval of metagenome-assembled genomes (MAGs) with high completeness. The most recent strategy in long-read metagenomics uses the long reads to obtain the raw 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 [10], among others –such as mock communities [11], cow rumen [12], natural whey starter cultures [13] or wastewater [14]. 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 [15] or with ultra-deep coverage of the genomes [11].

In our previous work, we used long-read metagenomics to profile canine fecal microbiome taxonomically and reach species identification. Despite using a low-depth sequencing approach, we were able to assemble a circular contig corresponding to an *uncultured CrAssphage* [16].

Here, using nanopore long-read metagenomics, we aim to unravel potential new bacterial diversity from the feces of a healthy dog. We assembled and characterized high-quality MAGs and identified their antimicrobial resistance genes to gain new biological insights on dog fecal metagenome.

## Material And Methods

### DNA extraction and long-read sequencing

Our study focuses on the analysis of a single fecal sample of a healthy dog. 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: the Quick-DNA HMW MagBead for High-Molecular Weight DNA (without bead-beating) and the DNA Miniprep Kit, which is the standard microbiome DNA extraction with bead-beating. Throughout the manuscript, we use HMW-DNA extraction and non-HMW DNA extraction terms, respectively.

Each DNA extraction was sequenced in a single MinION 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 minutes and 65°C for 20 minutes; ii) we extended rotator mixer (Hula mixer) times to 10 minutes; iii) we extended

elution time after AMPure XP beads to 10 minutes; iv) final incubation with elution buffer was performed at 37°C and for 15 minutes (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.

To obtain the first taxonomic assignment directly from the raw reads, we processed the data using Kraken2 2.0.8 [17] 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 [18].

We used Nanoplot 1.28 to obtain the run summary statistics [19], Porechop 0.2.4 [20] for adapters trimming, Nanofilt 2.6.0 [21] to discard reads shorter than 1,000 bp, and different modules of seqkit 0.11.0 [22] 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 [23].

We used the corrected reads to perform metagenome assembly with Flye 2.7 [24] (options: `-nano-corr -meta, -genome-size 500m, -plasmids`). We performed several assemblies with different random amounts of data (100%, 75%, 50%, and HMW) to recover the maximum number of high-quality MAGs (HQ MAGs). No binning software was applied. We used Bandage 0.8.1 to visualize the metagenome assemblies [25]. We polished the Flye assembly with one round of medaka 1.0.1, including all the raw fastq files as input [26].

The next step for the HQ MAGs was to correct the frameshift errors, as described in [15], using Diamond 0.9.32 [27] and MEGAN-LR 6.19.1 [28]. We used ideel [29] to visualize the number of truncated ORF.

To assess the quality of the MAGs, we used CheckM 1.1.1 [30] to retrieve completeness and contamination. MAGs can be 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 [31].

## Characterization of the high-quality MAGs

GTDB-tk 1.3.0 [32] with GTDB taxonomy release 95 [33] were used to assess the novelty and the taxonomy of HQ MAGs. We used PROKKA 1.13.4 to annotate the MAGs [34].

For the novel HQ MAGs, we used GtoTree 1.4.15 [42] to perform a *de novo* phylogenetic tree including the HQ MAG; the GTDB entries classified as the same genus; other NCBI assemblies of the same genus not included in GTDB; and a genome of a related taxon as an outgroup. We visualized the tree with iTOL 5.5.1 [35]. Abricate 0.9.8 [36] was used to detect antimicrobial resistance genes using CARD database [37]. OriTfinder was used to identify the origin of transfer (oriT) and conjugative machinery of mobile genetic elements [38] and SnapGene Viewer 5.0.7 [39] to visualize the results.

We used FastANI 1.3 [40] to confirm a potentially new species by determining the average nucleotide identity (ANI) between the most related genomes. One-to-one whole genome alignments were performed and visualized with dot plots using Mummer 4.0 [41].

We extracted the 16S rRNA genes from the HQ MAGs before the frameshift correction step using ANVIO 6.1 [42]. The 16S rRNA genes were analyzed using MOLE-BLAST tool in NCBI website [43] to obtain a phylogenetic tree. Mafft [44] in the EBI website was used to align 16S rRNA gene sequences from *Sutterella* HQ MAG and obtain an identity matrix.

## Results

We applied long-read nanopore sequencing to the fecal microbiome of a healthy dog. At the technical level, we compared the assembly results when choosing a HMW DNA extraction vs. a non-HMW one from the same fecal sample. Finally, using different metagenomics datasets, we retrieved and characterized eight high-quality single-contig draft metagenome-assembled genomes (HQ-MAGs) considering MIMAGs criteria (completeness >90%, contamination <5%, and presence of rRNA and tRNA genes) and three medium-quality draft metagenome-assembled genomes (MQ-MAGs; completeness >50%, contamination <10%) [31].

## HMW vs. non-HMW DNA: raw reads and metagenome assembly

HMW sequencing produced 5.81 million reads with N50 of 4,369 bp and a median length of 2,312 bp (total throughput: 18.76 Gb), whereas non-HMW produced 11.13 million reads with N50 of 2,102 bp and a median length of 1,093 bp (total throughput: 17.29 Gb).

We taxonomically classified the uncorrected raw reads with Kraken2 and found 81.8% of classified reads in HMW vs. 70.8% in non-HMW. More than 99% of 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 is rich in *Prevotella* (>50%) and *Bacteroides* (>20%).

Moreover, it also contains *Fusobacterium*, *Megamonas*, *Sutterella*, and other fecal-related genera, representing each one of them less than 5% of the total bacterial composition (Supplementary Figure S1).

The metagenomics assembly is more contiguous, presenting fewer and longer contigs, with HMW-DNA reads rather than non-HMW DNA one (n° of contigs: 1,898 vs. 2,944; N50: 187,680 vs. 94,109 bp). Moreover, HMW-DNA metagenomics assembly retrieves three circular contigs, and that could represent complete closed MAGs, for only one circular with de non-HMW DNA assembly (Figure 1).

So, 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).

## Metagenome assemblies, frameshift-aware correction, and retrieval of HQ and MQ MAGs

For the in-depth analyses, we assembled both the HMW only dataset and the HMW and non-HMW merged datasets (100% dataset; 16.94 million reads, 36.05 Gb) to ensure the highest coverage and consensus accuracies. 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 (Table 1).

The number of contigs ranged from 1,898 with HMW dataset to 2,639 when analyzing all the merged data together. N50 ranged from 187,680 bp (HMW dataset) to 149,125 bp (50% subset), and mean coverage ranged from 138X (100% dataset) to 95X (50% subset). The largest contig of 2.95 Mbp was retrieved when using 75% of the data.

**Table 1. 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.

	HMW data	100% data	75% data	50% data
Total length	125,567,322	141,997,441	131,702,503	119,187,600
Contigs	1,898	2,639	2,259	1,901
Contigs N50	187,680	150,083	162,895	149,125
Largest contig	2,751,144	2,769,659	2,950,218	2,846,287
Mean coverage	104X	138X	119X	95X
n° of HQ MAGs	6	4	6	3
n° of MQ MAGs	3	6	4	5

After assigning taxonomy and comparing among assemblies, we identified a total of eight different HQ MAGs, and three different MQ MAGs (Table 2). The different datasets retrieved redundant MAGs but with different degrees of quality. None of the performed assemblies alone retrieved all the HQ MAGs.

**Table 2. High quality (HQ) and medium quality (MQ) single-contig MAGs retrieved in each metagenome assembly.** Taxonomy assigned using the GTDB database release 95. Cov. is the coverage from Flye.

\**Blautia\_A* sp900541345 and \*g\_\_*Sutterella* HQ MAGs after correction of the indels.

	HMW data		100% data		75% data		50% data	
Taxonomy (GTDB)	MAG quality	Cov.						
<b>HQ MAG</b>								
<i>Prevotellamassilia</i> sp900541335	HQ	394X	HQ	577X	HQ	430X	HQ	282X
<i>Phascolarctobacterium</i> sp900544885	HQ	87X	HQ	205X	HQ	155X	MQ	98X
<i>Catenibacterium</i> sp000437715	HQ	13X	MQ	24X	HQ	17X	MQ	11X
<i>Enterococcus_B hirae</i>	HQ	17X	HQ	42X	HQ	31X	HQ	22X
<i>Blautia_A</i> sp900541345*	HQ	44X	-	-	MQ	45X	-	-
<i>Blautia_A</i> sp003287895	-	-	MQ	38X	HQ	31X	MQ	18X
<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
<b>MQ MAG</b>								
<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	-	-	-	-

The eight HQ MAGs corresponded to the genera *Prevotellamassilia*, *Phascolarctobacterium*, *Catenibacterium*, *Enterococcus*, *Succinivibrio*, *Blautia*, and *Sutterella* (Table 2). The HMW dataset and the 75% subset assemblies recovered six out of the eight HQ MAGs. Four of them were redundant and corresponded to *Prevotellamassilia* sp900541335, *Phascolarctobacterium\_A* sp900544885, *Catenibacterium* sp000437715, and *Enterococcus\_B hirae*. The remaining two from the HMW dataset were *g\_\_Succinivibrio* (found in all the datasets except for the 75% subset) and *Blautia\_A*

*sp900541345* (recovered after frameshift correction). Finally, the remaining two from the 75% subset were *Blautia\_A sp003287895* and *g\_\_Sutterella\** (recovered after frameshift correction).

For each HQ MAG, we chose the representative with the highest coverage – and subsequent highest consensus accuracy – to continue the analysis. We performed an extra correction step to reduce the insertions and deletions (indels), the most abundant error of nanopore sequencing. The indels correction reduced the frameshift errors and, consequently, the number of predicted coding sequences (CDS) (Supplementary Figure S2). This correction step transformed two MQ MAGs to HQ MAGs: the *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). On the other HQ MAGs, completeness remained constant or increased after applying the indel correction, except for one of the contigs (*Enterococcus hirae*, 47X coverage; completeness of 99.69% to 99.13% after the indel correction). The differences in applying indels correction were more evident in contigs with low coverage than in those with high coverage.

## Characterization of the HQ MAGs of the canine fecal microbiome

The eight HQ MAGs obtained are single-contig and represent complete HQ draft MAGs, without gaps or unplaced scaffolds (Table 3). GTDB-tk uses average nucleotide identity (ANI) values to identify potential novel taxa (>95% ANI is considered as the same species [40,45]). From the eight HQ MAGs, two corresponded to potentially new species (*Succinivibrio sp.* and *Sutterella sp.*); four, represented the first contiguous draft genome assembly for their genus (*Prevotellamassilia*, *Phascolarctobacterium*, and *Catenibacterium* HQ MAGs) or their species (*Blautia sp900541345* HQ MAG); and the remaining two, corresponded to complete reference genomes (*Enterococcus\_B hirae* and *Blautia sp003287895*).

**Table 3. High-quality MAGs comparison to references.** Completeness (% Compl.) values come from CheckM; tRNAs and rRNA from PROKKA; genome reference (Ref.) for the bacterial species from GTDB-tk.

HQ MAG	Length (Mbp)	% Compl.	tRNAs	rRNAs	Contiguity level
<i>Succinivibrio</i> sp.	2.04	98.68	77	22	Complete - new sp
<i>Succinivibrio</i> genus	1.38 - 3.96	51.33 – 100	10 - 66	0 - 24	10 - 320 scaffolds
<i>Sutterella</i> sp.	2.70	95.49	67	18	Complete
<i>Sutterella</i> genus	2.28 - 2.99	74.48 – 100	15 - 67	0 - 24	1 - 298 scaffolds
<i>Prevotellamassilia</i> sp900541335	2.72	97.65	72	21	Complete
Ref: GCA_900541335.1	2.42	96.13	16	0	95 contigs
<i>Phascolarctobacterium</i> _sp900544885	2.09	99.85	58	15	Complete
Ref: GCA_900544885.1	1.75	98.65	18	1	87 contigs
<i>Catenibacterium</i> sp000437715	2.53	98.50	76	21	Complete
Ref: GCF_004168205.1	2.54	100	20	2	212 contigs
<i>Blautia</i> sp900541345	2.44	93.86	53	18	Complete
Ref: GCA_900541345.1	2.69	95.85	16	0	160 contigs
<i>Enterococcus</i> _B_hirae	2.78	99.13	69	18	Complete
Ref: GCF_000271405.2	2.83	99.63	71	18	Complete
<i>Blautia</i> sp003287895	2.96	92.78	58	10	Complete
Ref: GCF_003287895.1	3.30	97.64	57	14	Complete

## Potential novel *Succinivibrio* species

*Succinivibrio* HQ MAG represents a new *Succinivibrio* species without any described representative, as confirmed by an ANI of 80% to its closest genome assembly GCA\_900552905.1 (<80% to *Succinivibrio dextrinosolvens* representatives). Moreover, all the *Succinivibrio* genome assemblies in NCBI are fragmented ('contig' or 'scaffold' level). So, this is the first contiguous assembly for the *Succinivibrio* genus. In GTDB taxonomy, several genome assemblies from the *Succinatimonas* genus and others are reclassified as *Succinivibrio*, so we included representatives of these genera in the phylogenetic tree (Figure 2).

Further genome characterization detected a total of 22 ribosomal genes. Among these, its seven 16S rRNA genes presented the highest identity to *uncultured bacterium clone CL\_F\_057* (GeneBank:

FJ978526.1) (Supplementary Figure S3), previously identified in wolves' distal gut microbiome [46]. *Succinivibrio* HQ MAG did not harbor antimicrobial-resistant genes.

## Potential genome for *Sutterella stercoricanis*

*Sutterella* HQ MAG is probably the genome assembly for *Sutterella stercoricanis*, as suggested by identities >98% with the previously reported 16S rRNA gene reference (Figure 3A), since its whole-genome sequence is absent in the public databases. *Sutterella stercoricanis* was first isolated in feces from a healthy dog and was characterized using microbiological methods and 16S rRNA gene sequencing (NR\_025600.1) [47].

Here, we retrieved a potential complete genome assembly for *Sutterella stercoricanis* in a single-contig HQ MAG. *Sutterella* HQ MAG is 2.70 Mbp and contains 18 ribosomal genes, including nine 16S rRNA and nine 23S rRNA genes (Prokka did not predict 5S rRNA genes). Moreover, the number of tRNAs detected is concordant to other complete *Sutterella* species (Table 3). The closest genome assemblies –including a representative of *Sutterella wadsworthensis*– presented ANI values around 80% (Figure 3B). No antimicrobial-resistant genes were identified within this HQ MAG.

## Single-contig HQ MAGs for *Prevotellamassilia*, *Phascolarctobacterium*, *Catenibacterium* and *Blautia* sp900541345

*Prevotellamassilia* sp900541335, *Phascolarctobacterium*\_A sp900544885, *Catenibacterium* sp000437715, and *Blautia* sp900541345 HQ MAGs are draft genomes with high completeness values that improve the contiguity of previous assemblies of their respective bacterial species. The species representative genomes in GTDB are also MAGs obtained from gastrointestinal or fecal human microbiome and retrieved using short-read technologies. In consequence, they are highly fragmented and fail to recover all ribosomal genes and transfer RNAs (Table 3).

Moreover, *Prevotellamassilia*, *Phascolarctobacterium*, and *Catenibacterium* HQ MAGs are the first single-contig representative for their genus since all the other assemblies of these genera are fragmented ('scaffold' or 'contig' level).

Their 16S rRNA genes were close to others previously identified in wolves' distal gut microbiome [46] (*Prevotellamassilia* HQ MAG), canine intestinal microbiome [48] (*Phascolarctobacterium* HQ MAG), and human GI microbiome [49] (*Catenibacterium* and *Blautia* sp900541345 HQ MAG) (Supplementary Figure S3).

We further characterized the HQ MAG to assess the potential antimicrobial resistance. Firstly, *Prevotellamassilia* HQ MAG harbored *Mef(En2)* gene, which encodes for an efflux pump that exports macrolides. *Phascolarctobacterium* HQ MAG harbored two copies of *Inu(C)* gene conferring resistance to lincosamide. Each *Inu(C)* gene was located in an *ISSag10* mobile element, allowing it to transpose.

*Catenibacterium* HQ MAG harbored *tet(M)* and *Blautia* sp900541345 HQ MAG harbored *tet(O)*. Both genes confer resistance to tetracycline.

## Known genomes from metagenomes: *Enterococcus hirae* and *Blautia argii*

The HQ MAGs representing known genomes were *Enterococcus hirae*, and *Blautia* sp003287895 (Figure 4) –proposed name *Blautia argii*, first isolated and characterized on dog feces [50]. Both representative genomes in GTDB are already complete and reference genomes.

*Enterococcus hirae* HQ MAG presented a genome size similar to its reference and the same number of rRNA genes. It harbored *aac(6′)-IId* and *tet(M)* genes conferring resistance to aminoglycosides and tetracycline, respectively. Specifically, the *tetM* gene 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 (Supplementary Figure S4).

*Blautia* HQ MAG presented a smaller genome size than its reference genome (2,959,590 bp vs. 3,297,975). When aligning both genomes, we observed some gaps in our HQ MAG that are identifying those differences (Figure 4). Moreover, the completeness of this HQ MAG was the lowest (92.78%) among all the HQ MAGs retrieved. Further MAG characterization identified 5 *rrn* operons (10 ribosomal genes, since Prokka missed five 5S rRNA genes), which coincided with the reference. Moreover, *Blautia* HQ MAG harbored *tet(32)* and *tet(40)* genes conferring resistance to tetracycline.

## Overview of the MQ MAGs

Apart from the HQ MAGs, we identified three MQ MAGs (>50% completeness and <10% contamination). They corresponded to *Phocaeicola plebeius* and potentially novel species from *Phocaeicola* and *Bacteroides* genera (Table 2).

The closest genome reference for *Phocaeicola plebeius* (previously named *Bacteroides plebeius*) MQ MAGs was GCF\_000187895.1. When looking at the 100% dataset, we had a longer (2.37 Mb) and a more contiguous assembly for *Phocaeicola plebeius* MQ MAG (single-contig vs. 19 scaffolds), but with lower completeness (74.81% vs. 99.25% completeness). This MQ MAG lacks the *Mef(En2)* gene, conferring resistance to macrolides found in the 75% MQ MAG.

The longest *Phocaeicola* spp. MQ MAG (2.56 Mb, 75% subset assembly) had a completeness of 83.39% and 0.19% of contamination. It also harboured *Mef(En2)* gene, conferring resistance to macrolides. Finally, the longest *Bacteroides* MQ MAG (2 Mb, 100% dataset assembly) had a completeness of 61.66% and 0.37% of contamination. It harbored *CfxA2* gene conferring resistance to beta-lactams.

## Discussion

We applied long-read metagenomics to a fecal sample of a healthy dog and retrieved eight HQ MAGs and three MQ MAGs, all of them single-contigs.

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 gave less unclassified reads. For metagenomics assembly, it improved the contiguity and increased the retrieval of longer and circular contigs (potential HQ MAGs).

For the subsequent analyses, we used both the HMW data and the whole merged dataset to ensure the highest consensus accuracy. Moreover, we assessed different amounts of total data (75% and 50% data subsets) to retrieve the maximum number of HQ MAGs. None of the performed assemblies alone retrieved the eight HQ MAGs. The HQ MAGs were representatives of the *Succinivibrio*, *Sutterella*, *Prevotellamassilia*, *Phascolarctobacterium*, *Enterococcus*, *Blautia*, and *Catenibacterium* genera.

*Succinivibrio* HQ MAG is the first single-contig genome assembled in the genus. It represents a novel candidate bacterial species, with ANI of 80% to its closest genome assembly GCA\_900552905.1 (<80% to *Succinivibrio dextrinosolvens* representatives). Its full-length 16S rRNA genes cluster with the 16S rRNA gene from uncultured bacteria amplified in a wolves' GI microbiome study [51].

*Sutterella* HQ MAG is potentially the first reported genome assembly for *Sutterella stercoricanis* that was first isolated in feces from a healthy dog [47]. Since the reference isolate lacks additional genome information to confirm that the *Sutterella* HQ MAG represents the same species, 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 [52]: the nine 16S rRNA genes presented identities from 99.04% to 98.69% against *Sutterella stercoricanis* 16S ribosomal RNA (NR\_025600.1). Whole-genome sequencing of the reference isolate and comparison to the HQ MAG could confirm if they represent the same species.

*Prevotellamassilia*, *Phascolarctobacterium*, *Catenibacterium*, and *Blautia* sp900541345 HQ MAGs improved the contiguity of previously reported genome assemblies in their respective genera (single-contig assembly vs. multiple scaffolds), and the number of rRNA genes and tRNA genes. Finally, *Enterococcus* *shirae* and *Blautia* HQ MAGs represented species with complete reference genomes. *Blautia* HQ MAG was *Blautia* sp003287895 (proposed species name *Blautia argii*) and was first isolated from feces of a mature dog [50].

The genera *Succinivibrio*, *Prevotella*, *Phascolarctobacterium*, *Catenibacterium*, and *Blautia*, are recognized short-chain fatty acid (SCFA) producers [53–55]. These molecules provide multiple gut health benefits from reducing inflammation and tumorigenesis to increasing gut motility and secretory activity [2,54,56]. In dog GI microbiome, different diets and dietary interventions can modulate their abundances aiming to promote gut health [7,57–62]. 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 [63–66]. So, in-depth characterization of these genera is of most relevance to defining a healthy GI microbiome in dogs.

*Sutterella stercoricanis* was isolated from the feces of a healthy dog [47]. However, the increase of the genus *Sutterella* was associated with detrimental effects rather than health. Dogs with acute hemorrhagic diarrhea presented higher *Sutterella* [63], and some diets aiming to promote health benefits observed its decrease [67,68]. Further whole-genome sequencing studies are needed to identify the different *Sutterella* species on dog feces and correlate their abundances to health or disease status.

Finally, *Enterococcus 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 genes reservoirs that could be transferred to people [69–73].

*Enterococcus* HQ MAG harbors *aac(6′)-lid* gene, which was first detected in *Enterococcus durans* and conferred resistance to aminoglycosides [74]. Besides, it harbors a *tetM* gene within the Tn916 conjugative element, which was first reported in *Enterococcus faecalis* [75,76].

Tetracycline resistance genes were found not only in the genome of *Enterococcus hirae*, but also in *Catenibacterium* and both *Blautia* HQ MAGs and could be linked to a previous antimicrobial exposure that selected the resistant bacteria [77]. Three years before sampling, this dog was treated with doxycycline –tetracycline-class antibiotic– during 15 days for excess secretion of mucus and saliva. Whole resistome analyses are needed to determine the AMR genes within the fecal microbiome in healthy dogs and to evaluate all the bacterial species together with their mobile genetic elements that could act as a reservoir for AMR genes.

Despite humans and dogs share similar microbial composition on the GI microbiome [6,7], *Succinivibrio*, *Blautia*, and *Sutterella* HQ MAGs seem to be canine-related fecal bacterial species. This fact highlights the need for building and using niche-specific databases to accurately map and classify new reads from a particular environment, as well as understand the overall biological significance [12,78].

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, as well as to link antimicrobial resistance genes, mobile genetic elements, and bacteriophages to their bacterial host [79].

A limitation of this study is the use of nanopore-only data since it can compromise the accuracy of the HQ MAGs. To reduce the insertion and deletion error type, we applied a frameshift-aware correction step [15] that improved the completeness and reduced the number of CDS. On the other hand, long-read metagenomics improved the contiguity of the MAGs even for reference assemblies. Long-read metagenomics sequencing could harness short-read metagenomics data to build comprehensive and curated databases to in-depth characterize novel bacterial diversity in the canine fecal microbiome.

## Conclusions

To conclude, we recovered and characterized eight HQ MAGs and three MQ MAGs from a fecal sample of a healthy dog using long-read metagenomics. Among them, one potential novel species for *Succinivibrio* and the first genome assembly for *Sutterella stercoricanis*. Overall, long-read metagenomics allowed us to recover HQ MAGs from a complex microbiome. The high-molecular weight DNA extraction to improve contiguity and the correction of the insertions and deletions to reduce frameshift errors ensured the retrieval of complete single-contig HQ MAGs.

## 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

AMR: antimicrobial resistance

## Declarations

**Ethics approval and consent to participate:** The owner gave written consent to the collection of samples.

**Consent for publication:** not applicable.

**Availability of data and materials:** The raw assemblies, the metagenome-assembled genomes, and an overview of the scripts used are available on Zenodo: [10.5281/zenodo.3982645](https://zenodo.org/record/3982645). An overview of the scripts used to analyze the data is at Additional File 5.

**Competing interests:** AC works for Vetgenomics, SL. The other authors declare that they have no competing interests.

**Funding:** Vetgenomics and Molecular Genetics Veterinary Service (SVGm), Universitat Autònoma de Barcelona.

**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 final data. JV analyzed the

antimicrobial-resistance genes. AC drafted the manuscript. OF, JV and DP substantially revised the work. All the authors have approved the submitted version.

**Acknowledgments:** We would like to acknowledge Amanda Warr from the University of Edinburgh for her advice and support on the long-read metagenomics approach, and Norma Fabregas for reading and revising the draft manuscript.

## Bibliography

1. 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;2:1533–42.
2. Pilla R, Suchodolski JS. The Role of the Canine Gut Microbiome and Metabolome in Health and Gastrointestinal Disease. *Front Vet Sci.* 2020;6.
3. Redfern A, Suchodolski J, Jergens A. Role of the gastrointestinal microbiota in small animal health and disease. *Vet Rec.* 2017;181:370–370.
4. 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;2:71–94.
5. Honneffer JB, Minamoto Y, Suchodolski JS. Microbiota alterations in acute and chronic gastrointestinal inflammation of cats and dogs. *World J Gastroenterol.* 2014;20:16489–97.
6. 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.
7. 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;6:72.
8. 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;469:967–77.
9. 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;5:639–49.
10. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol.* 2020;38:701–7.
11. Nicholls SM, Quick JC, Tang S, Loman NJ. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience.* 2019;8.
12. Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol.* 2019;37:953–61.

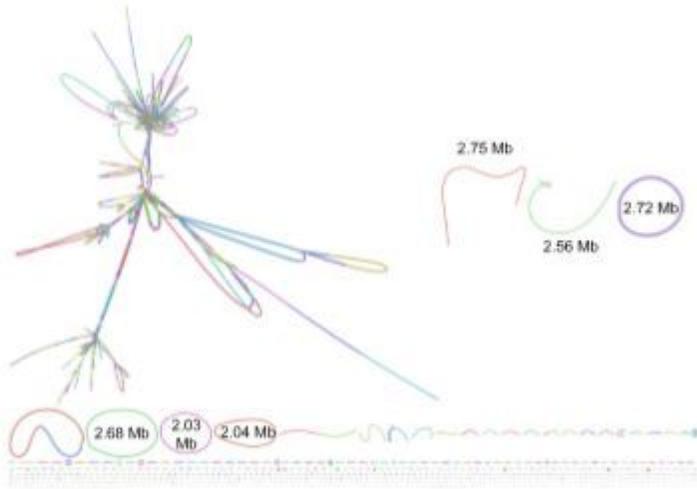
13. Somerville V, Lutz S, Schmid M, Frei D, Moser A, Irmeler S, et al. Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *BMC Microbiology*. 2019;19:143.
14. Singleton CM, Petriglieri F, Kristensen JM, Kirkegaard RH, Michaelsen TY, Andersen MH, et al. Connecting structure to function with the recovery of over 1000 high-quality activated sludge metagenome-assembled genomes encoding full-length rRNA genes using long-read sequencing. *Biorxiv*; 2020 May.
15. Arumugam K, Bağcı C, Bessarab I, Beier S, Buchfink B, Górska A, et al. Annotated bacterial chromosomes from frame-shift-corrected long-read metagenomic data. *Microbiome*. 2019;7:61.
16. 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 Mar.
17. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019;20:257.
18. Breitwieser FP, Salzberg SL. Pavian: Interactive analysis of metagenomics data for microbiomics and pathogen identification. *bioRxiv*. 2016;084715.
19. Coster WD. NanoPlot [Internet]. 2020. Available from: <https://github.com/wdecoster/NanoPlot>
20. Wick R. Porechop [Internet]. 2020. Available from: <https://github.com/rrwick/Porechop>
21. Coster WD. Nanofilt [Internet]. 2020. Available from: <https://github.com/wdecoster/nanofilt>
22. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE*. 2016;11:e0163962.
23. 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;27:722–36.
24. Kolmogorov M, Rayko M, Yuan J, Pevnikov E, Pevzner P. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Bioinformatics*; 2019 May.
25. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. Oxford Academic; 2015;31:3350–2.
26. Medaka – Medaka 0.12.1 documentation [Internet]. Available from: <https://nanoporetech.github.io/medaka/>
27. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*. 2015;12:59–60.
28. Huson DH, Albrecht B, Bağcı C, Bessarab I, Górska A, Jolic D, et al. MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. *Biology Direct*. 2018;13:6.
29. Watson M. ideel [Internet]. 2020. Available from: <https://github.com/mw55309/ideel>
30. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*.

2015;25:1043–55.

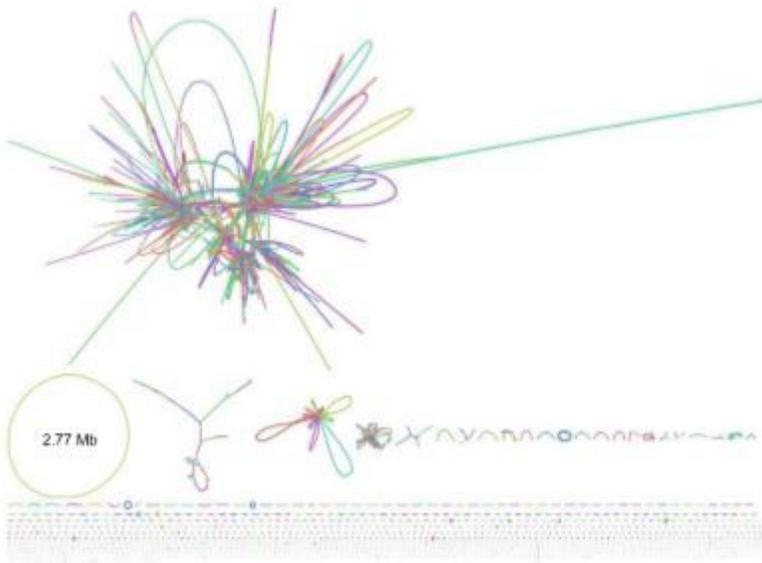
31. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol.* 2017;35:725–31.
32. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2020;36:1925–7.
33. 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;1–8.
34. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30:2068–9.
35. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 2019;47:W256–9.
36. Seemann T. Abricate [Internet]. 2020. Available from: <https://github.com/tseemann/abricate>

## Figures

A) HMW DNA metagenome assembly

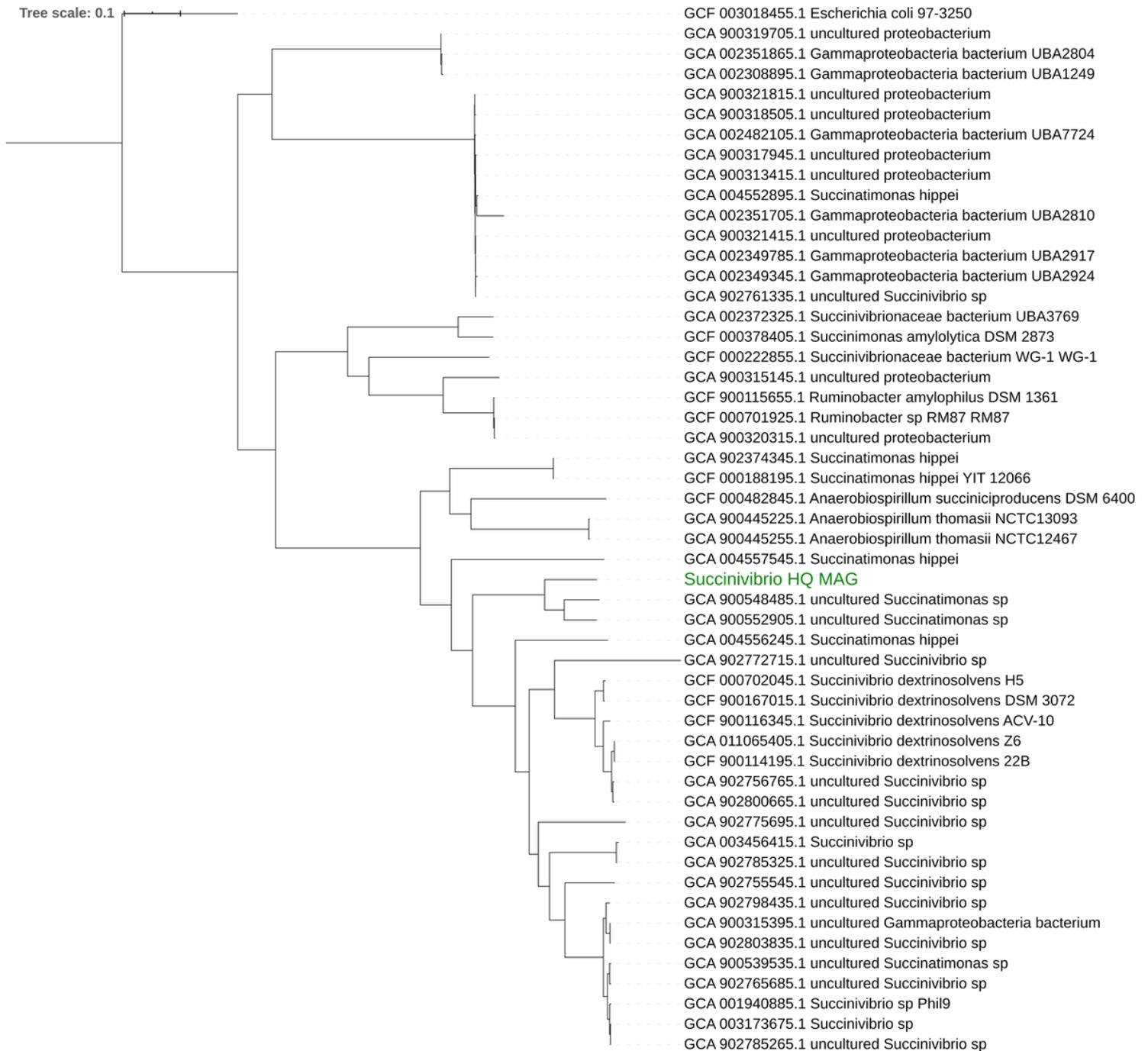


B) Non-HMW DNA metagenome assembly



**Figure 1**

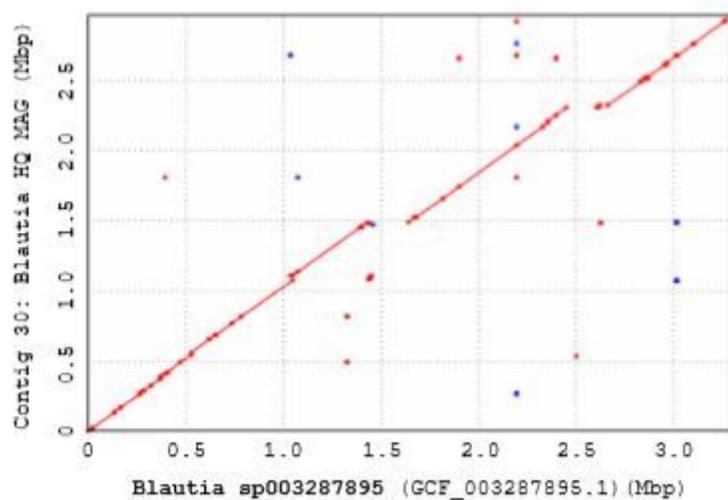
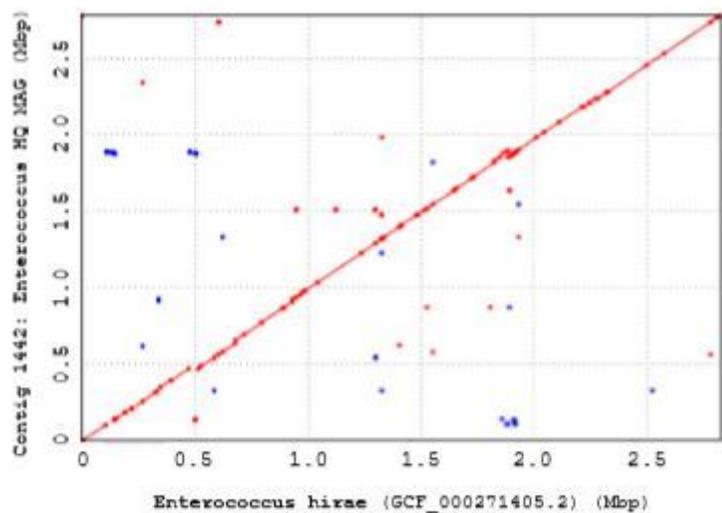
HMW-DNA vs. non-HMW DNA metagenomics assembly from the fecal sample of a healthy dog. Bandage plots of A) HMW-DNA assembly and B) non-HMW DNA assembly. HMW-DNA allows the recovery of long, circular contigs, which can potentially represent complete closed MAGs. We report the longest contig in both datasets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



**Figure 2**

Phylogenetic de novo tree for *Succinivibrio* HQ MAG. The tree included: i) the HQ MAG; ii) the GTDB entries classified as *Succinivibrio*; iii) extra NCBI assemblies of the same genus not included in GTDB; iv) a genome of a related taxon as an outgroup (for rooting the tree).





**Figure 4**

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).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFile5.pdf](#)
- [AdditionalFile4.pdf](#)
- [AdditionalFile3.pdf](#)
- [AdditionalFile2.pdf](#)
- [AdditionalFile1.pdf](#)