

# Capturing global pet dog gut microbial diversity and hundreds of near-finished bacterial genomes by using long-read metagenomics in a Shanghai cohort

Anna Cuscó<sup>1\*</sup>, Yiqian Duan<sup>1</sup>, Fernando Gil<sup>1</sup>, Alexei Chklovski<sup>2</sup>, Nithya Kruthi<sup>2</sup>, Shaojun Pan<sup>1</sup>, Sofia Forslund<sup>3,4,5,6,7</sup>, Susanne Lau<sup>5</sup>, Ulrike Löber<sup>3,4,5,6</sup>, Xing-Ming Zhao<sup>1</sup>, Luis Pedro Coelho<sup>2\*</sup>

## SUMMARY

Pet dogs are considered part of the family, and understanding their gut microbiomes can provide insights into both animal and household health. Most comprehensive studies, however, relied on short-read sequencing, resulting in fragmented MAGs that miss mobile elements, antimicrobial-resistance genes, and ribosomal genes. Here, we applied deep long-read metagenomics (polished with short-reads) to fecal samples from 51 urban pet dogs in Shanghai, generating 2,676 MAGs—representing 320 bacterial species—, of which ~72% achieved near-finished quality, often improving on the corresponding reference public genome. Comparisons with external datasets showed that our Shanghai-based MAG catalog is representative of pet dogs worldwide (median read mapping of >90%). Moreover, we recovered circular extrachromosomal elements, including those linked to antimicrobial resistance, which were also detected in external dog gut datasets. In conclusion, we provide a high-quality reference resource and demonstrate the power of deep long-read metagenomics to resolve microbial diversity in complex host-associated microbiomes.

## Authors' affiliations:

<sup>1</sup> Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

<sup>2</sup> Centre for Microbiome Research, School of Biomedical Sciences, Queensland University of Technology, Translational Research Institute, Woolloongabba, QLD, Australia

<sup>3</sup> Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany / <sup>4</sup> Experimental and Clinical Research Center, A Cooperation of Charité-Universitätsmedizin Berlin and Max Delbrück Center for Molecular Medicine, Berlin, Germany / <sup>5</sup> Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health. Dept of Pediatric Respiratory Medicine, Immunology and Critical Care Medicine, Berlin, Germany / <sup>6</sup> DZHK (German Centre for Cardiovascular Research), Partner Site, Berlin, Germany / <sup>7</sup> Structural and Computational Biology Unit, European Molecular Biology Laboratory, Structural and Computational Biology Unit, Heidelberg, Germany

\*Correspondence to: [anna@big-data-biology.org](mailto:anna@big-data-biology.org) and [luispedro.coelho@qut.edu.au](mailto:luispedro.coelho@qut.edu.au)

# INTRODUCTION

Pet dogs (*Canis lupus familiaris*) are commonly regarded as family members, living in close contact with their humans and often sharing not only physical spaces but also microbes<sup>1,2</sup>. Understanding dog microbiome provides insights not only directly linked to their health<sup>3</sup>, but also to household health<sup>4,5</sup> and global health<sup>6</sup>, being key within One Health frameworks<sup>2</sup>.

Studies on dog gut metagenomics are scarce and all comprehensive ones are based on short-read sequencing<sup>7–12</sup>. Recently, a large-scale project significantly expanded the dog gut metagenomes available<sup>12</sup>. However, most short-read derived metagenome-assembled genomes (MAGs) are heavily fragmented and fail to assemble ribosomal genes. Although the Minimum Information about a Metagenome-Assembled Genome (MIMAG) criteria for high-quality include recovering these (and other elements)<sup>13</sup>, most short-read based studies consider only (estimated) completeness and contamination. We will follow previous literature<sup>14</sup> and use the term near-finished for MAGs that fulfill all MIMAG high-quality criteria. For instance, in the Unified Human Gastrointestinal Genome catalog, only 12% of the species-level representative metagenome-assembled genomes (573 out of 4,644 MAGs) met these quality criteria<sup>15</sup>. Antibiotic resistance genes (ARGs), mobile genetic elements, and extrachromosomal elements are similarly often absent from short-read MAGs<sup>16–18</sup>. Long-read technologies can address these limitations<sup>19–23</sup>, but large cohorts with deeply sequenced samples using long reads are still scarce. For example, the dog gut microbiome using long-read metagenomics has been recently explored in three pilot studies, which included a single animal<sup>24–26</sup>.

Finally, most large dog gut microbiome studies characterize animals living in colonies, usually from nutritional companies<sup>7,9,12</sup> rather than pet dogs<sup>10,11</sup>. These cohorts comprise dogs with homogeneous characteristics, such as a limited number of breeds, individuals of similar ages, similar diets, and a shared environment. While these controlled conditions are ideal for certain types of studies (e.g., randomized controlled trials) as they reduce variability, they may not generalize to pet dogs living in households<sup>27</sup>.

Here, we aim to expand the dog gut microbiome data by using deep metagenomics sequencing (minimum per sample: 20 Gbp of long reads + 20 Gbp of short reads) from a relatively large and well-characterized cohort of 51 pet dogs living in an urban environment (Shanghai, China). We make available ~1,600 Gbp of Nanopore data and ~1,000 Gbp of Illumina data, 2,676 canine MAGs, 185 circular extrachromosomal elements, non-redundant gene and smORFs catalogs, as well as associated functional and ARGs annotation. By using external public datasets and a new

cohort of pet dogs from Germany (also available), we demonstrate that the Shanghai MAG catalog captures global pet dog diversity.

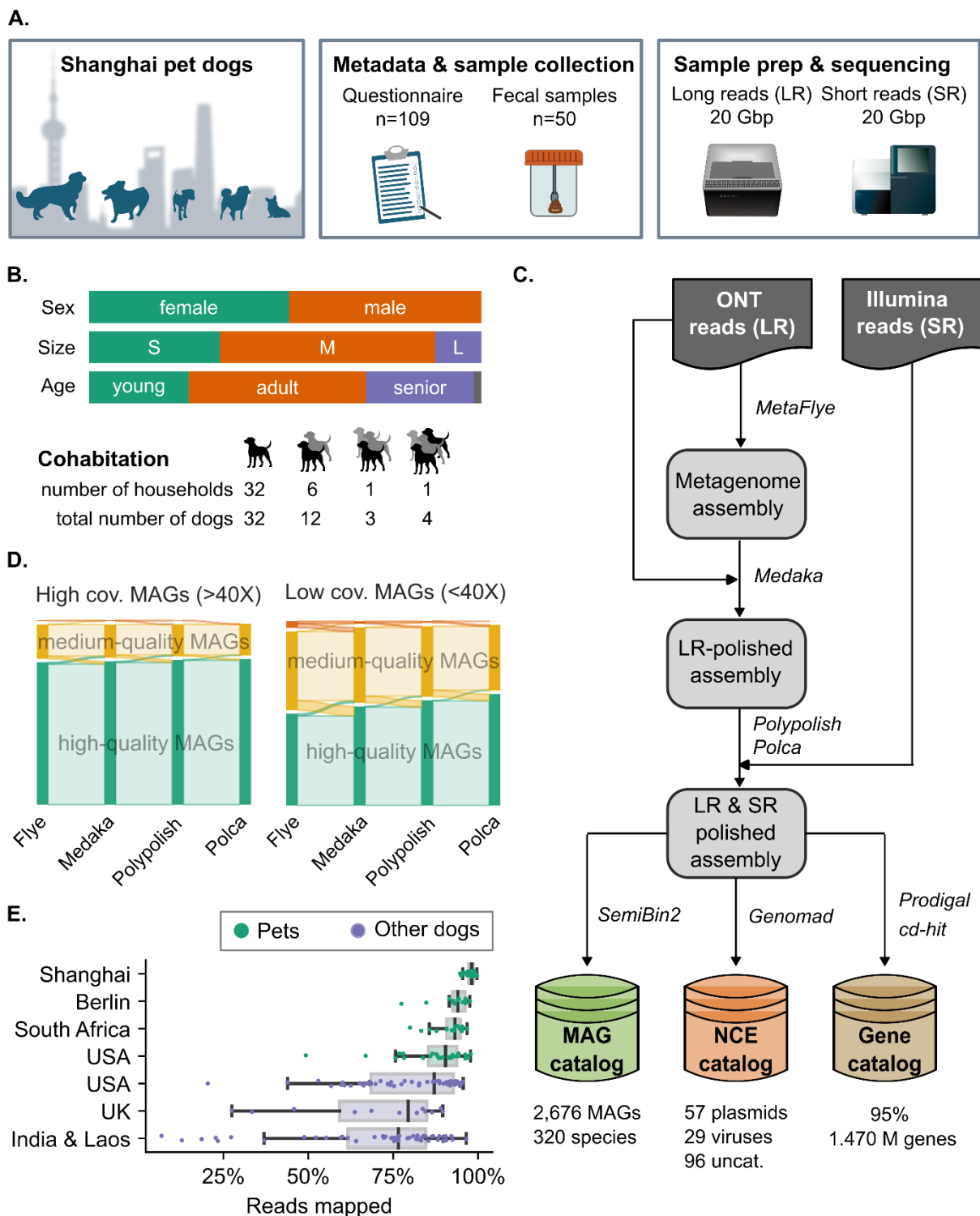
# RESULTS

## Near-finished MAGs are recovered from deep long- and short-read sequencing of a Shanghai dog cohort

We collected a fresh fecal sample from 51 Shanghai pet dogs, associated with extensive dog-associated information (gathered through a questionnaire). Each fecal sample was deeply sequenced using a minimum of 20 Gbp of long-read (Oxford Nanopore Technologies) and 20 Gbp of short-read (Illumina) DNA sequencing (Figure 1A, Supplementary Figure S1).

The pet dogs were from the same urban environment (Shanghai, China) and had an urban indoor lifestyle. The cohort comprises animals of different ages, sizes, breeds, diets, and overall habits (Figure 1B and Supplementary Table S1). In addition to the sequenced dogs, we make available 59 questionnaire responses from other Shanghai pet dog owners who did not donate a fecal sample (Supplementary Table S2).

In the MAG generation pipeline, we used the long reads for metagenome assembly and the short reads for polishing, followed by a combination of binning strategies and a final dereplication step, leading to 2,676 MAGs (Figure 1C, Supplementary Figure S2). Each polishing step significantly increased the completeness and decreased the contamination of the canine MAGs (Wilcoxon FDR corrected p-value < 0.05). Polishing converted 177 medium-quality MAGs into high-quality ones, and 48 low-quality MAGs into medium-quality ones (Figure 1D, Supplementary Figure S3), with the greatest benefits observed for MAGs with lower sequencing depth (<40X). Therefore, polishing with short reads was still relevant, especially when working with metagenomes that often include low-abundant species.



**Figure 1. Shanghai pet dog gut microbiome using long-read metagenomics. A)** Project design overview: we collected information from 109 pet dogs in Shanghai and 51 fecal samples. The samples were processed and sequenced to obtain  $\geq 20$  Gbp of both long- and short-read data. **B)** Overview of the Shanghai dog-associated metadata. Complete information for all the dogs can be

found in Supplementary Table S1. **C)** MAG generation pipeline: long-read (LR) sequencing data were used for metagenome assembly with MetaFlye. The metagenome assembly was first polished with long reads using Medaka, and later with short reads (SR) using Polypolish, followed by Polca. The polished metagenome assembly was binned using SemiBin2 to obtain the Shanghai Dog MAG catalog (the binning strategy is detailed in the Methods and represented in Supplementary Figure S2). **D)** Short-reads (SR) for polishing increased the quality of the MAGs, especially for those with low coverage (<40X). **E)** The Shanghai dog MAG catalog captures the vast majority of the microbial diversity of other pet dog cohorts living in households (median read mapping of >90%). The mapping is lower for non-pet cohorts (colony, shelter, or free-roaming dogs).

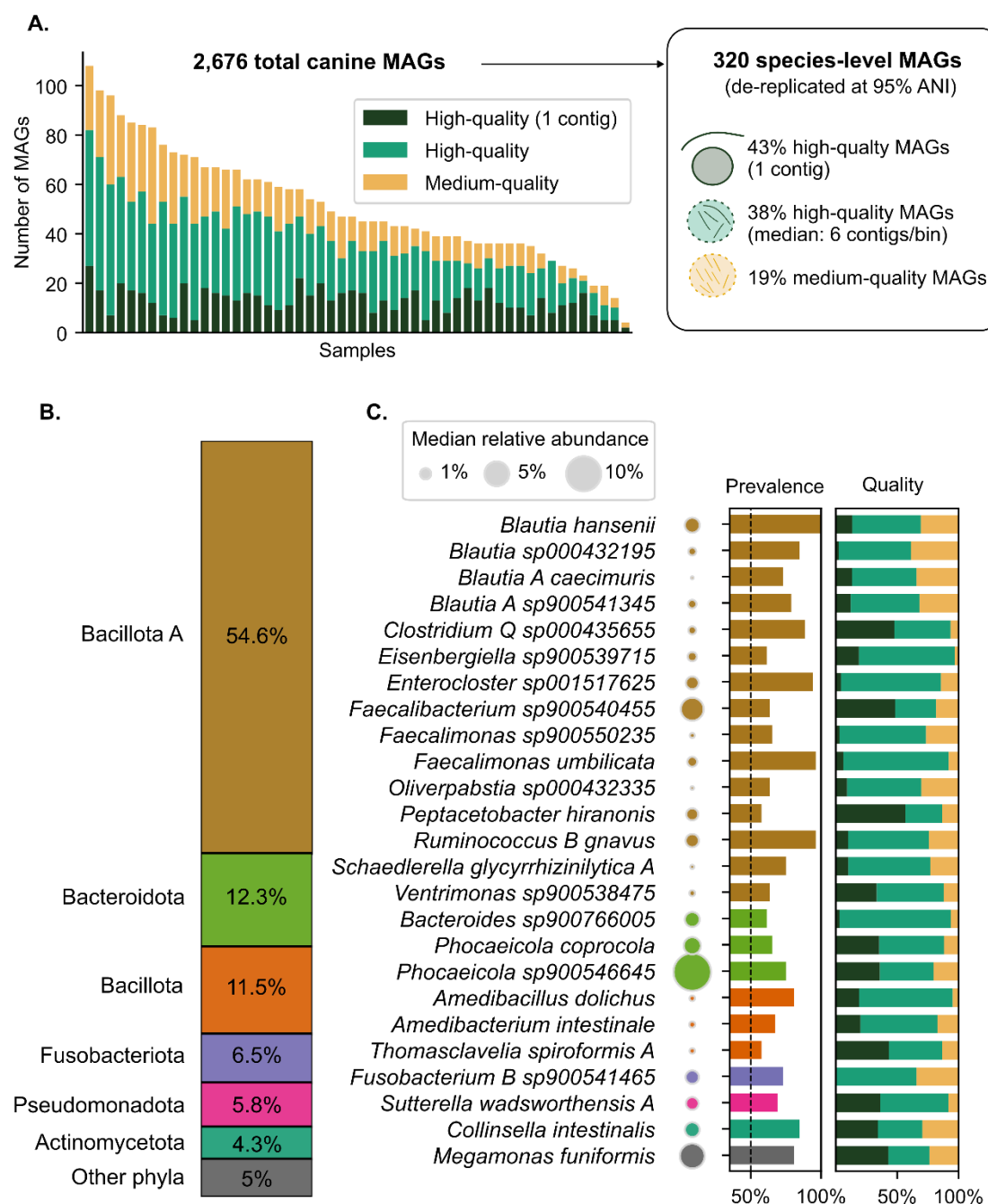
The final MAG collection consists of 2,676 MAGs from the Shanghai pet dogs, representing 320 bacterial species (Figure 2A, Supplementary Table S3). The Shanghai dog MAG catalog contains ~72% near-finished MAGs (n=1,928) (high-quality MAGs fulfilling all the MIMAG criteria: >90% completeness, <5% contamination, presence of ribosomal genes, and at least 18 tRNAs<sup>13</sup>). Of those, 34% were single-contig assemblies, whereas the remaining had a median of six contigs per genome assembly. Medium-quality MAGs represented ~27% of the MAGs in the catalog (n=732), and only ~1% (n=16) of the MAGs were high-quality MAGs (>90% completeness, <5% contamination) but did not fulfill some of the remaining MIMAG criteria. Apart from MAGs, we also identified 185 non-redundant circular extrachromosomal elements: 58 plasmids, 30 viruses, and 97 uncategorized elements.

Notably, 19 out of the 51 dogs in the Shanghai cohort came from 8 multi-dog households, each contributing at least two sampled dogs (Figure 1B). Within these households, dogs shared a higher proportion of bacterial strains and showed greater similarity in their extrachromosomal elements, as measured by presence-absence data using Jaccard similarity (Supplementary Figure S4).

## The Shanghai dog MAG catalog captures most microbial diversity found in pet dogs

The Shanghai dog MAG catalog captures the microbial diversity of pet dogs living in households globally, as demonstrated by a median read mapping of >90% against metagenome datasets from different geographical origins (Germany, newly sequenced for this study; South Africa<sup>11</sup>; and USA<sup>10</sup>). Despite mapping rates being lower for non-pet dog cohorts, the median was >86% for dogs living in colonies<sup>7,9</sup> and >75% for dogs living in shelters or the streets (as free-roaming dogs)<sup>11</sup>. Within the Shanghai cohort, 98.2% (median value) of the short-reads mapped to the MAG

collection, demonstrating that high- and medium-quality MAGs capture almost all of the diversity present in the samples (Figure 1E, Supplementary Table S4).



**Figure 2. The Shanghai pet dog MAG catalog contains 2,676 MAGs from 320 species. A)** Stacked bar plots show the total number of MAGs per sample, stratified by quality. Almost all the high-quality MAGs (all but 16) fulfilled the MIMAG criteria. The catalog contains a total of 2,676 MAGs, from which 320 species-level MAGs can be obtained (dereplicated at 95% ANI). **B)**

Distribution of phyla in the MAG catalog of Shanghai pet dogs. **C)** The most prevalent species in Shanghai pet dog gut microbiomes, defined as species-level genomes assembled in >30 different dogs (out of 51 total samples). Bubble size represents the median relative abundance of each bacterial species within the total microbiome composition, when present (i.e., zero values were excluded). Prevalence bars show the prevalence for each species. Quality bars show the MAG assembly quality distribution per bacterial species (as defined in **A**). The complete Shanghai Dog MAG catalog information is available in Supplementary Table S3.

More than half of the Shanghai dog MAGs belonged to the Bacillota A phylum, followed by Bacteroidota, Bacillota, Fusobacteriota, Pseudomonadota, and Actinomycetota (Figure 2B, Supplementary Table S3 for the full MAG catalog, annotations with the Genome Taxonomy Database (GTDB) taxonomy r214). Dog gut microbiome studies agree on these phyla being the main ones inhabiting the gut of healthy dogs<sup>3,7</sup>. The most prevalent bacterial species on the Shanghai dog cohort belonged to the Lachnospiraceae family (Bacillota phylum, n=13), followed by the Bacteroidaceae family (Bacteroidota phylum, n=3) (Figure 2C). The most prevalent species in the Shanghai dog cohort—present in almost every dog—were four Lachnospiraceae species: *Blautia hansenii*, *Enterocloster sp001517625*, *Faecalimonas umbilicata*, and *Ruminococcus B gnavus*. Except for *E. sp001517625*, these three species were reported as the most prevalent (prevalence >0.83) on a large cohort of pet dogs living in households in the USA (n=286) as screened with full-length 16S rRNA gene<sup>28</sup>. When focusing on species that are prevalent and also abundant (>2% of total median abundance, when present), we detected *Phocaeicola coprocola*, and *Phocaeicola sp900546645* (Bacteroidaceae family); *Fusobacterium A sp900555845*; *Megamonas funiformis*; and *Faecalibacterium sp900540455* (Figure 2C).

## Species-level long-read MAGs capture novel species and improve representative genome assemblies

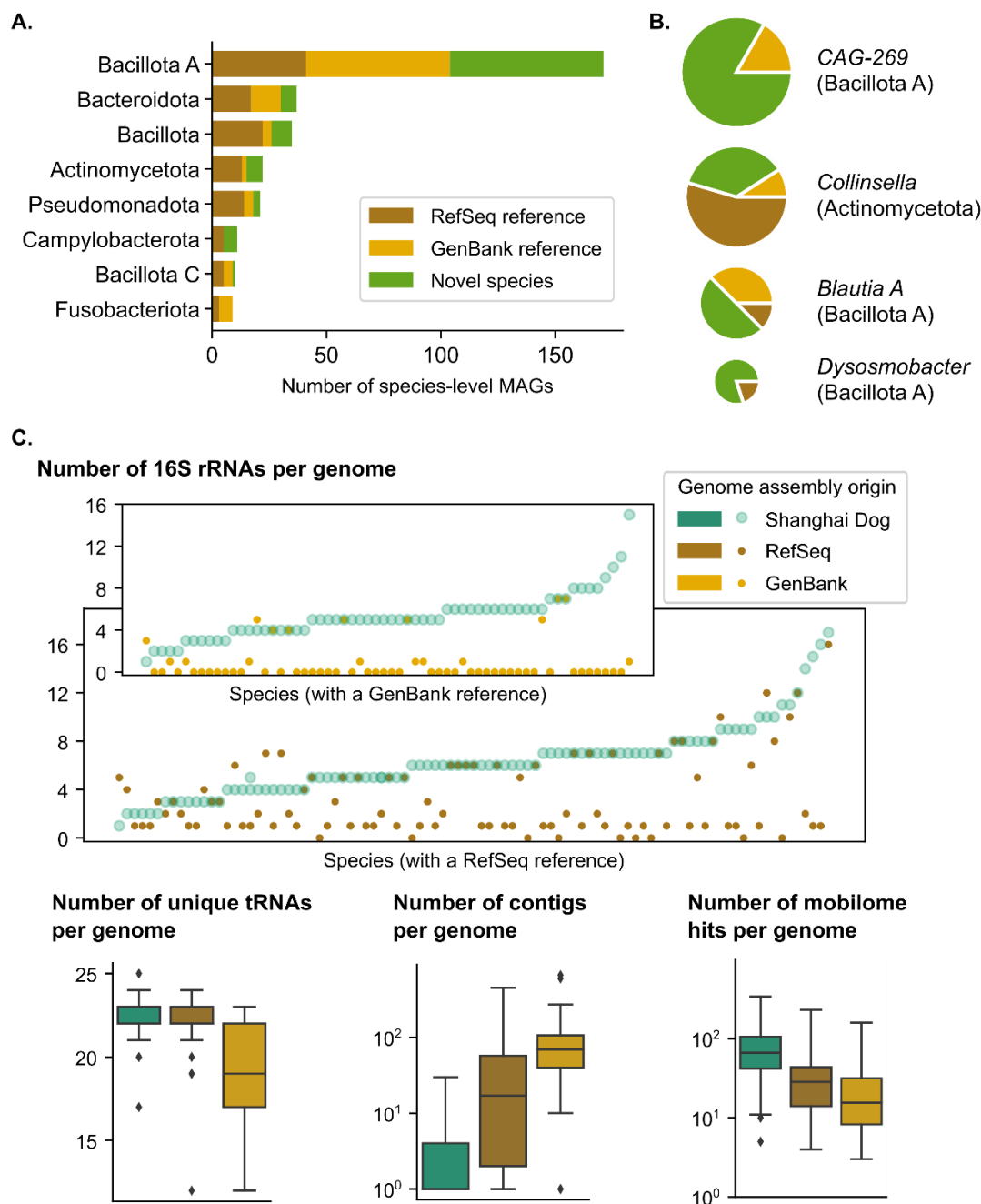
Two-thirds of the species-level MAGs represented known bacterial species (~68%, n=218), as their genomes are found in the GTDB database (r214) (Figure 3A). We compared the species-level high-quality MAG assemblies here to their specific high-quality representative genome in the public database (156 species with either GenBank or RefSeq assemblies) (Figure 3C and Supplementary Table S5). We found that the species-level long-read MAGs were significantly more contiguous, contained more full-length ribosomal genes and unique tRNA genes (all pairwise comparisons had a corrected p-value < 0.05, Supplementary Table S5). They usually fulfilled the MIMAG criteria, which was not true for most of the high-quality representative genomes in public databases. One challenge for short-read shotgun metagenomics is the

assembly of ribosomal genes, which are multi-copy and conserved within similar genomes, and consequently, they end up collapsed and unassembled<sup>16</sup>.

Moreover, compared to reference assemblies, canine long-read MAGs usually had a larger mobilome fraction—genes annotated as COG category X: Mobilome (corrected p-value < 0.05, Supplementary Table S5), consistent with previous reports showing that short-read assemblies were largely ineffective for mobile genetic elements analysis<sup>17</sup>, whereas long-reads enabled the capture of mobile elements within their genomic context<sup>19,29</sup>. Overall, our long-read MAGs improve representative genome assemblies on databases, especially those from short-read based assemblies.

One-third of the species-level MAGs represented bacterial species not present in the GTDB (r214) (~32%, n=102). The genera that contained the most novel species were *CAG-269* with 10 different new species-level MAGs, followed by *Dysosmobacter*, *Collinsella*, and *Blautia A* with four different new species-level MAGs each (Figure 3B, Supplementary Table S3). To further explore the novelty of the species and considering that 16S rRNA gene databases are more extensive than whole genomes, we aligned the full-length 16S rRNA genes of the high-quality MAGs against the 16S ribosomal RNA sequences database from NCBI using the blastn suite. The best hits corresponded to 16S rRNA genes from known species within the same genera, and in 11 cases, the identity of the best hit was >99% (Supplementary Table S6). In nine out of 11 cases, we confirmed the inability of this marker gene to distinguish species within certain genera<sup>30</sup>.

Several bacterial species lack representative genome assemblies in public databases but have phenotypic and 16S rRNA gene sequencing data. These include *Sutterella stercoricanis* and *Glucerbacter canis*, which were first isolated in dog feces<sup>31,32</sup>. A previous study showed that high-quality long-read MAGs can link 16S rRNA gene-based community profiling studies to the functional potential of the genomes within the same environment<sup>20</sup>. Long-read MAGs enabled us to putatively link genome assemblies to previously phenotypically characterized species in dogs through 16S rRNA gene sequences.



**Figure 3. Species-level canine MAGs characterization.** **A)** Counts of species-level representative MAGs by phylum in the Shanghai Dog cohort. Stacked bars are colored according to the genome assembly novelty for that species: novel genome (light green); representative species genome assembly present in RefSeq (brown); or representative species genome assembly present in GenBank (yellow). **B)** Top four genera with a larger number of novel species-level MAGs in the Shanghai dog cohort. Pie charts represent the number of species-level genome assemblies within each genus. The size of the pie charts represents the total number of species per genus. **C)**

Considering only high-quality genome assemblies (>90% completeness and <5% contamination criteria), comparison of the representative species-level genome assemblies: canine long-read MAGs vs. public database representative (RefSeq or GenBank). Scatterplot distribution of the number of 16S rRNA genes: on the x-axis, we represented all the species, and on the y-axis, the number of rRNA genes. Boxplots represent the distribution of the number of unique tRNA genes, number of contigs, and number of mobilome hits, stratified by the origin of the corresponding publicly available representative genome assembly for each species.

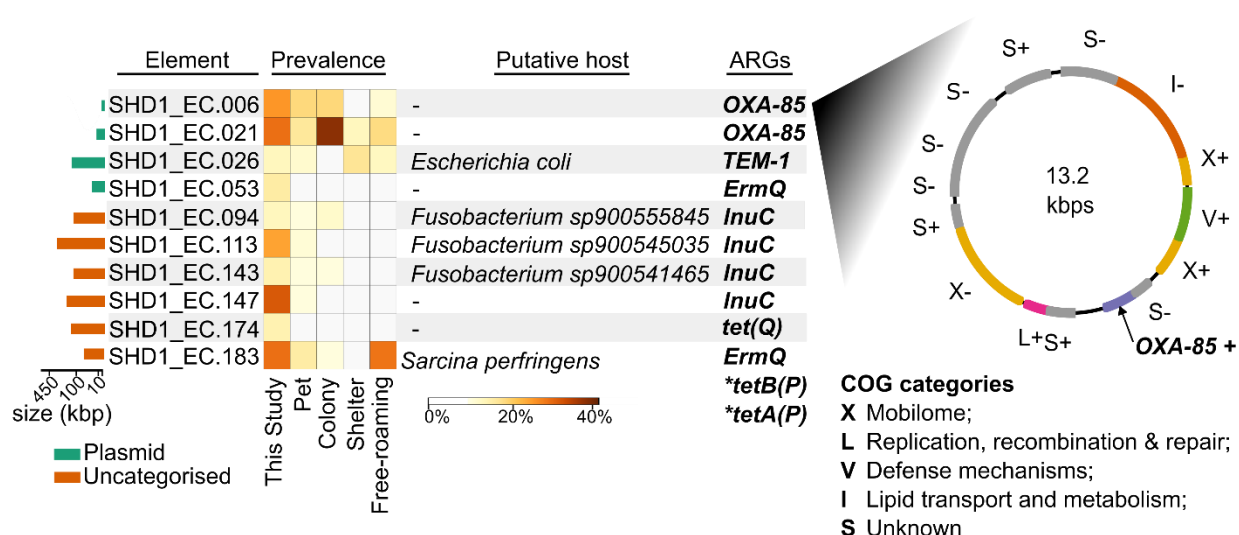
## Long-read metagenomics uncovers ARG-carrying circular extrachromosomal elements in the canine gut

We identified 253 unique antibiotic resistance genes (ARGs) in the Shanghai dog cohort (using RGI, see Methods). Tetracycline ARGs were the most common, representing ~32% of the total ARG counts (with 2,461 total hits, corresponding to 25 unique ARGs), followed by aminoglycoside and lincosamide ARGs, representing ~13% of the total ARG counts each (1,021 and 1,001 total hits, corresponding to 31 and 5 unique ARGs, respectively). This aligns with findings from a large-scale resistome analysis including >3,000 dog samples, where tetracyclines, lincosamide, and aminoglycosides ARGs were the most abundant ARGs on dogs<sup>33</sup>. Additionally, these ARGs correspond to antimicrobial drug classes that are essential and commonly used in small animal medicine<sup>34</sup>.

Interestingly, of the 253 unique ARGs, 43 were exclusively found in contigs that were not part of a MAG. Among ARGs, those located on extrachromosomal elements (ECEs) are particularly concerning due to their potential for horizontal gene transfer across bacterial species. This mobility facilitates rapid dissemination of antimicrobial resistance within microbial communities<sup>35</sup>. The use of long-read sequencing is not only crucial for contextualizing ARGs, but also effective at assembling circular ECEs that might harbour them<sup>19,21</sup>. In addition to the MAG catalog, we assembled 185 non-redundant circular ECEs: 58 plasmids, 30 viruses, and 97 uncategorized elements. Among those, four plasmids and six uncategorized elements harboured ARGs (Figure 4, Supplementary Table S7).

We identified two different, circular, and single-contig ECEs harboring the OXA-85 gene, which were present in more than 25% of the samples in our cohort, and had a prevalence of up to 45% in colony dogs (SHD1\_EC.021) and up to 14% in other pet dog cohorts (SHD1\_EC.006) (Figure 4). The OXA-85 gene encodes a class D  $\beta$ -lactamase that confers narrow-spectrum resistance primarily to penicillin-class antibiotics<sup>36</sup>. We could not assign these two ECEs to a putative

bacterial host; therefore, they could be present in multiple species. Previously, OXA-85 had only been reported on the chromosome of *Fusobacterium nucleatum*<sup>36</sup> and *Fusobacterium necrophorum*<sup>37</sup>, consistent with entries in the CARD database<sup>38</sup>, which additionally found it in *Campylobacter ureolyticus*. In fact, we also detected OXA-85 within the chromosomes of *Fusobacterium B sp900541465* (n=3), *Fusobacterium A sp900543175* (n=1), as well as *Phascolarctobacterium A sp900552855* (n=3). This finding is concerning due to the extrachromosomal location of the OXA-85 gene, which could enable horizontal gene transfer; its potential association with unreported bacterial hosts; and its prevalence across diverse dog cohorts. All these suggest broader dissemination and host range than previously observed.



**Figure 4. Circular extra-chromosomal elements (ECEs) harboring antimicrobial resistance genes (ARGs) in the Shanghai Dog cohort.** The heatmap represents the prevalence of each ECE in both our cohort and external dog cohorts (as confirmed by read mapping). For some ECEs, we predicted their putative bacterial host and identified their ARGs. The circular diagram represents the ORFs identified in EC.006 and their predicted COG functional category.

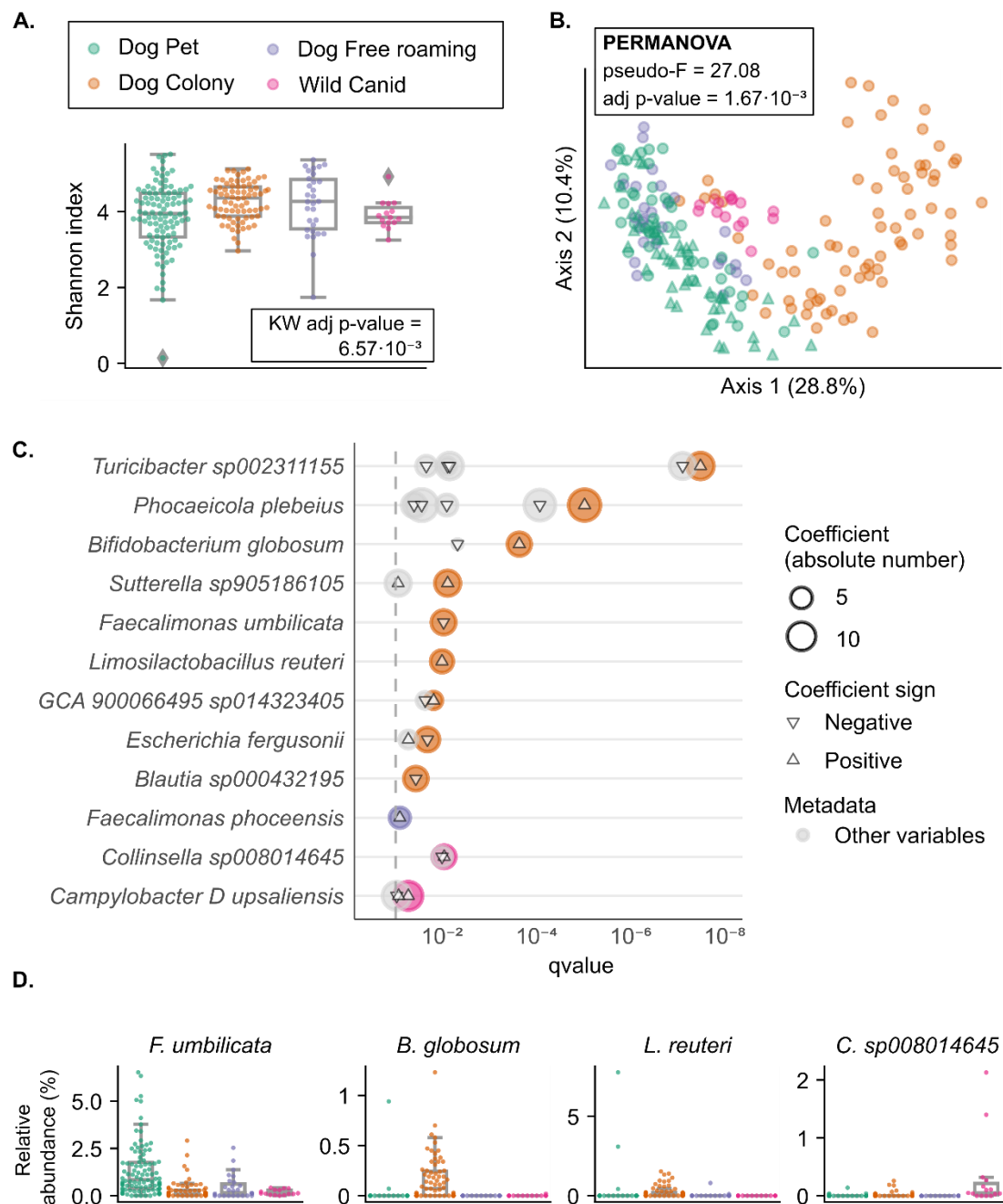
We found one circular ECE harboring *TEM-1*, which was linked to *E. coli* as its putative host. *TEM-1* is a plasmid-encoded beta-lactamase—found in many Gram-negative bacteria—that confers resistance to penicillins and first-generation cephalosporins<sup>39</sup>. Its prevalence across other dog cohorts reinforces concerns about the circulation of clinically relevant ARGs in non-clinical reservoirs. We also identified two different, circular, and single-contig ECEs carrying the *ErmQ* gene, one with *Sarcina perfringens* as its putative host (SHD1\_EC.183) —consistent with the original *ErmQ* report<sup>40</sup>. The *ErmQ* gene encodes a ribosomal RNA methyltransferase that confers

resistance to macrolide, lincosamide, and streptogramin B antibiotics. Interestingly, a recent paper characterizing pet-derived *S. perfringens* in China found that 91% of them showed a high resistance rate to erythromycin associated with the presence of the *ErmQ* gene<sup>41</sup>. Finally, and coinciding with the most abundant and prevalent ARGs, lincosamide (*InuC*) and tetracycline resistance genes (*TetQ*, *TetA*, *TetB*) were the ARGs detected in the remaining five ECEs (Figure 4).

## Global analysis detected that dogs living in colonies present a different gut microbiome

To contextualize our data and investigate factors that structure the dog gut microbiome, we integrated the Shanghai dog gut microbiome with a set of external dog samples (Berlin cohort, and publicly available canid cohorts until February 2023). The largest dog gut metagenome studies focused on colony dogs rather than pet dogs living in households<sup>7,9</sup>. Only three studies involved pet dogs: one examined dogs with recurrent diarrhea<sup>8</sup>, while the other two analyzed healthy pets from South Africa<sup>11</sup> and the USA<sup>10</sup>. Using representative samples from these public datasets (Supplementary Table S8), we assessed the influence of the living environment (household, colony, or free roaming), age, size, and sex of the animal. Since these studies are short-read studies, for consistency, we used the short-read Illumina component of our data and performed taxonomic profiling on the combined dataset using a pipeline based on singleM<sup>42</sup> (see Methods).

The living environment was the variable structuring the microbial community taxonomic composition at the beta diversity level (PERMANOVA adjusted p-value =  $1.67 \times 10^{-3}$ ), with a clearly separated cluster for dogs living in colonies. Pet dogs overlapped and clustered in the ordination plot, even though they were from different geographies (China, USA, Germany, and South Africa) (Figure 5B). Analogously, alpha diversity (Shannon index) was significantly higher in colony dogs when compared to pet dogs (Mann-Whitney U corrected p-value =  $4.90 \times 10^{-3}$ ) and wild canids (Mann-Whitney U corrected p-value = 0.025; Figure 5A). Ancient coprolite samples presented the lowest alpha diversity values and clustered independently on the beta diversity plot, but considering that these samples all originate from the same study<sup>43</sup>, it was not possible to exclude a study effect (Supplementary Figure S5).



**Figure 5. The living environment is the major determinant of dog gut microbiome composition.** The colored legend applies to the whole figure. **A)** Boxplots representing alpha diversity (Shannon index). **B)** PCoA plot representing beta diversity (Bray-Curtis on log-transformed data). Green triangles indicate pet dogs in this study. **C)** Differentially abundant bacterial species in each dog group when compared to pet dogs (Maaslin2 and Kruskal-Wallis, see Methods). The x-axis shows the q-value of each species computed by Maaslin2. Circles are colored according to

the group to which the dogs belong. Grey circles represent significant differences in that species, for variables other than the living environment. Diversity analysis with all dog groups can be found in Supplementary Figure S5. **D)** Boxplots of total relative abundances (%) of selected differentially abundant bacterial species, considering the living environment.

When looking at differential species abundance, six species were more abundant in colony dogs compared to pet dogs and other dog cohorts, including *Bifidobacterium globosum* and *Limosilactobacillus reuteri* (Figure 5C and 5D). *Lactobacillus* and *Bifidobacterium* species are commonly regarded as probiotics in dogs<sup>44</sup>, so this could represent a more common intake of probiotics in dogs living in colonies compared to other dog cohorts. Moreover, we identified *Faecalimonas umbilicata* as a pet-associated species, as it was significantly more abundant in pet dogs compared to all the other dog groups. Finally, we also detected two bacterial species more abundant in wild canids and one in free-roaming dogs (Figure 5C).

Differential abundance analysis also detected bacterial species associated with: age of the animal (n=13), body condition (n=2), and size of the animal (n=2) (Supplementary Figure S6). However, no significant differences were found regarding sex. Differences related to dogs' age were mostly significant when comparing senior to young animals, *Dorea B phocaeensis*, *Slackia A piriformis*, *Eisenbergiella sp900539715*, and *Faecalibacterium sp900540455* being the most significant. Interestingly, *Faecalibacterium* and *Slackia* were among those genera more abundant in healthy dogs when compared to those having cancer<sup>45</sup>, and cancer is an age-related disease. We also found *Butyricicoccus pullicaecorum* to be more abundant in young dogs. Interestingly, this species has been assessed as a potential probiotic and has been shown to extend lifespan in a *C. elegans* model<sup>46</sup>.

## Generated resources description

In this study, we generated several high-quality resources from pet dogs using long-read metagenomics, polished with short reads. We make these datasets available to be further explored and used:

1. **Comprehensive dog-associated information of Shanghai pet dogs:** includes responses from 107 dog owners in Shanghai regarding their pets' health, diet, and lifestyle (among others) collected through an extensive questionnaire (Supplementary Table S2). Considering the most relevant variables, we created a simplified table for the 51 sequenced dogs included in the study (Supplementary Table S1).

2. **Deeply sequenced samples (raw data):** includes 52 fecal samples belonging to 51 dogs (1 biological replicate) with at least 20 Gbp of data for long-read and 20 Gbp of data for short-read. Long-read datasets have up to 65 Gbp of data per sample (available on ENA under Bioproject ID: PRJEB85799).
3. **Metagenome-assembled genomes (MAG) catalog:** includes 2,676 MAGs that belong to 320 different bacterial species. 72% are near-finished MAGs, meaning they are >90% complete, <5% contaminated, harbor ribosomal genes, and at least 18 unique tRNAs. 665 of the MAGs are single-contig. 32% of the species-level representatives are novel (Supplementary Table S3). In addition to the sequences (in FASTA format), we provide extracted rRNA genes (with the 16S rRNA linked to MicrobeAtlas to enable users to link to 16S amplicon-based studies<sup>47</sup>), and annotations using eggno-mapper and RGI.
4. **Extrachromosomal elements (ECEs) catalog:** includes 185 single-contig circular elements (non-redundant), as assembled by MetaFlye from our cohort. We identified 58 plasmids, with sizes ranging from 4,333 bp to 154,111 bp (median length: 30,562 bp); 30 viruses; and 97 uncategorized elements. Moreover, we predicted the putative bacterial host for 28% of them (n=52).
5. **Gene catalog:** includes 1,470,802 non-redundant genes (95% nucleotide identity, representing species-level unigenes<sup>48</sup>). Of these, 98.7% are complete genes, and 73.5% are contained in MAGs or ECEs.
6. **Small ORFs catalog:** includes 403,491 non-redundant smORFs, which produce 273,065 clusters at 90% identity.
7. **External dog datasets:** raw data for a newly sequenced short-read metagenomics pet dog cohort from Berlin (Bioproject ID: PRJNA881055), as well as standardized and manually curated metadata for dog gut metagenomics datasets in public databases up to February 2023<sup>7–11,43,49–58</sup> (Supplementary Table S9). We also created a metadata subset including a single representative sample per animal for longitudinal or repeated measures studies (Supplementary Table S8).

We created a website (<https://sh-dog-mags.big-data-biology.org/>) to dynamically explore the MAG catalog, and all the derived resources can be downloaded on Zenodo: <https://doi.org/10.5281/zenodo.16356977>

# DISCUSSION

Here, we present the Shanghai dog gut microbiome resource, comprising samples from 51 pet dogs. This dataset includes extensive metadata –over 50 lifestyle, health, and dietary variables collected through owner questionnaires– and deep metagenomics sequencing. From this, we derived multiple resources, including a comprehensive pet dog MAG catalog with 2,676 MAGs from 320 different bacterial species. Our Shanghai dog MAG catalog captures most of the global microbial diversity of pet dogs as highly contiguous, near-finished MAGs, as shown by the high mapping rates to the catalog (>90% median) of external pet dog cohorts from different geographies (Germany, South Africa<sup>11</sup>, USA<sup>10</sup>).

Previous short-read metagenomics studies on the dog gut have focused on functional and taxonomic profiling<sup>8–11</sup> or building a gene catalog<sup>7</sup>, without assembling MAGs. Recently, a large-scale study on companion animals retrieved 7,275 high-quality and 21,706 medium-quality MAGs from canine and feline samples using short-read metagenomics<sup>12</sup>. However, short-read metagenome assemblies typically fail to recover ribosomal RNA genes<sup>16</sup>. They also often miss other important features such as antimicrobial resistance genes (ARGs)<sup>18</sup> and mobile genetic elements<sup>17</sup>. In this study, we present a high-quality canine MAG catalog built from long-read nanopore sequencing, polished by short reads. Most (~72%) of the MAGs are near-finished, with completeness >90%, contamination <5%, and include ribosomal RNA genes and tRNAs, thus fulfilling the full MIMAG criteria<sup>13</sup>. Moreover, we retrieved 185 non-redundant circular extrachromosomal elements; 1,470,802 non-redundant unigenes; and 403,491 non-redundant small open reading frames (smORFs). Our resource includes 320 species, and over 80% have a near-finished MAG as a representative genome. Many of our species-level representative MAGs improve existing reference genomes in the GTDB r214 database (including ones that were obtained from cultures, but assembled with short reads), offering higher contiguity, and including the ribosomal genes and tRNAs. Looking forward, we expect that many species in multiple habitats will have higher-quality genomes available in the next few years as long-read metagenomics becomes widely used.

A previous study suggested that >40X coverage with Nanopore R10.4 chemistry can yield near-finished MAGs from metagenomes<sup>14</sup>. Here, we observed further improvements in completeness and contamination values when incorporating short reads, suggesting additive benefits. While we cannot rule out differences from basecalling models—our data were basecalled using the high-accuracy model versus the super-accuracy model used in that study—particularly considering

that complex microbiomes contain a plethora of low abundance genomes, it is evident that long-read sequencing is key to generating near-finished MAGs, and, with current ONT chemistry, short read sequencing is still valuable.

Within a One Health framework—which recognizes the interconnectedness of global human, animal, and environmental health—the close interaction between pets and humans, particularly in shared household environments, creates ideal conditions for the bidirectional transmission of bacteria and their associated ARGs<sup>59</sup>. Although we did not target the human gut microbiome, we observed that dogs living in the same household shared significantly more bacterial strains and extrachromosomal elements (ECEs) compared to dogs from different households, confirming the existence of intra-household transmission patterns.

Notably, 10 of the assembled ECEs carried ARGs, which play a central role in horizontal gene transfer and the dissemination of resistance within shared environments<sup>35</sup>. Among these, we identified a putative *E. coli* plasmid carrying the *TEM-1* gene, underscoring concerns about plasmid-mediated antibiotic resistance in a bacterial species that is commonly shared between humans and animals, and known for its ability to mediate gene exchange across mammalian hosts<sup>60,61</sup>. Additionally, the detection of the *OXA-85* gene in two ECEs and across multiple anaerobic species (e.g., *Fusobacterium*, *Phascolarctobacterium*) suggests a mobilization event and an expanded host range for this ARG. Although these specific ARGs are not currently considered major clinical threats, their detection in mobile genetic elements within the pet dog gut microbiome highlights the value of metagenomics surveillance using long-reads to identify ARG reservoirs and monitor ARG flow across taxa and environments<sup>62</sup>.

We observed that the mapping rate to our catalog for samples of dogs living in colonies was slightly lower than pets, which hinted at a difference in their microbiome structure. Previous studies showed that factors such as industrialization and feralization influence the gut microbiome composition and function in companion animals like dogs<sup>11</sup> and cats<sup>63</sup>. However, in a study by Yargaladaa and collaborators, industrialization was confounded with geography, as each industrialization level corresponded to a distinct geographic region. In contrast, a large-scale consumer study including 192 US pet dogs (16S rRNA gene sequencing) did not detect effects of urbanization on the gut microbiome, though some geographic signatures were reported<sup>64</sup>. The largest companion animal microbiome study to date<sup>12</sup> reported that housing facilities (research facilities vs. household vs. strays) significantly influenced the dog gut microbiome. Beta diversity analysis confirmed that pet dogs from different geographic regions (China, Germany, US<sup>10</sup>, and South Africa<sup>11</sup>) presented similar gut microbiome compositions, clustering together and

overlapping within the same region of the PCoA plot. In contrast, colony dogs from nutritional research facilities formed a distinct cohesive cluster despite their different geographic origins (US<sup>7</sup> and UK<sup>9</sup>). Taking all these observations together, we conclude that the living environment has a stronger influence than geography on the composition and structure of the canine gut microbiome. These living environment differences likely reflect differences in multiple factors such as diet, environmental exposure, human interaction, and veterinary care. Thus, while colonies have definite benefits for controlled studies, their microbiome differs from the general pet population, and results from colony studies should be complemented with studies in pets.

In summary, our study significantly expands the existing resources on the dog gut microbiome by generating and sharing a comprehensive multi-terabyte dataset derived from deep long- and short-read metagenomic sequencing of a well-defined cohort of urban pet dogs. With over 2,600 canine MAGs, hundreds of circular elements, and extensive gene and smORF catalogs, this dataset provides an unprecedented foundation for future research. Moreover, by validating our Shanghai-derived MAG catalog against both public data and an independent German cohort, we show that it captures a broad spectrum of global pet dog microbiome diversity, establishing it as a valuable reference for comparative and functional microbiome studies.

## STAR METHODS

### DATA AVAILABILITY

Raw reads and MAG catalog for the Shanghai pet dog cohort were deposited at the European Nucleotide Archive under Bioproject ID PRJEB85799. Berlin pet dog raw data can be found under Bioproject ID PRJNA881055.

MAG annotation files, ARGs prediction, ECEs catalog, gene catalog, and smORFs catalog can be found at Zenodo: <https://doi.org/10.5281/zenodo.16356977>. Finally, all original code is publicly available at GitHub: <https://github.com/BigDataBiology/ShanghaiDogs>

### EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

In this study, we analyzed the fecal microbiome of pet dogs living in households in Shanghai. The Animal Welfare and Ethics Group (Department of Experimental Animal Science, Fudan University) approved this study under the approval reference 022JSISTBI-003. All the dog owners who donated a fecal sample provided informed consent to participate in the study.

## METHOD DETAILS

### Dog-associated metadata collection

We designed a comprehensive questionnaire that covers various aspects of dogs, including their basic information (breed, age, sex, weight) and other relevant variables related to lifestyle, habits, diet, health, and more. Some of the questions in our questionnaire were adapted from Lehtimäki et al., 2018<sup>65</sup>. The questionnaire was entered into and administered through the *wjx* mini app on WeChat and shared across relevant WeChat groups. 107 dog owners answered the questionnaire, and 51 agreed to donate a fecal sample from their dogs (Supplementary Table S2).

For the 51 dogs sampled, updates in the metadata, such as health status, were done on the sampling date and adapted for further analysis (Supplementary Table S1). We grouped the dogs into three age categories, as previously considered<sup>66</sup>: Young (< 2 years old), Adult (2-6 years old), and Senior ( $\geq 7$  years old). Finally, we classified the dogs according to size, as previously described<sup>67</sup>: I) in five clusters, and II) in size categories: Small (< 10 Kg), Medium ( $\geq 10$  to < 25 Kg), and Large ( $\geq 25$  Kg). The dog cohort was balanced by sex.

### Fecal sample collection

The owners collected the fecal samples when walking their dogs per their regular habits, which were given to us 0h-24h after the dog's deposition. Therefore, we have neither altered nor manipulated the animals in any way.

Once we had the fresh dog fecal sample, we used a fresh pair of gloves to cut the feces with a plastic spatula and expose the inner part. We collected a spoon-sized sample from the inner part of the feces using a tube with a spoon. We did this process twice per sample, so we have a biological duplicate as a backup. Immediately after processing, the tubes were stored in dry ice and transported to a -80 °C freezer until further processing.

### DNA extraction and sequencing

Fecal samples were processed and sequenced by Novogene (Beijing, China). The DNA was extracted using the CTAB method. Genomic DNA quality was verified by: 1% agarose gels to detect DNA degradation; Nanodrop OD 260/280 ratio to check the purity of DNA; and Qubit® DNA Assay Kit (in Qubit® 3.0 Fluorometer) to measure DNA concentration.

For short-read Illumina sequencing library preparation, 0.2 µg of DNA per sample was used. The DNA was sonicated to an average fragment size of 350 bp. It was generated using NEB Next®

Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations, and index codes were added to each sample. Clustering of the index-coded samples was performed on a cBot Cluster Generation System using the Illumina PE Cluster Kit (Illumina, USA), following the manufacturer's instructions. The final DNA libraries were sequenced on an Illumina NovaSeq 6000, generating 150 bp paired-end reads.

For long-read Nanopore sequencing library preparation, 1-2 µg of DNA per sample was used. The DNA was size selected using magnetic beads, and the library was constructed following the Oxford Nanopore Technologies (ONT) 1D Genomic DNA by Ligation protocol (v14) with the Ligation Sequencing Kit SQK-LSK114 following the manufacturer's recommendations. The final DNA libraries were sequenced on a PromethION P48 instrument using R10.4.1 flow cells. Raw signal data (fast5 files) were basecalled using Guppy v6.4.6 with the High-Accuracy (HAC) model.

### **Selection of external dog gut metagenomes**

We selected all the available dog gut metagenomes until February 2023. We used the GMGC catalog<sup>48</sup> to collect the initial entries and their associated metadata, and we added new studies by searching at PubMed and the SRA database. Dog-associated metadata was manually curated from each sample's associated literature and Biosample or SRA-associated metadata. We additionally included a new shotgun metagenomic dataset of 15 pet dogs from Berlin (Germany).

The collection of external dog metagenomes used in this study and their associated metadata can be found in the Supplementary Table S9. Whenever multiple samples from the same animal were available, we selected a single representative sample (Supplementary Table S8). For choosing the representative samples, we prioritized that the dogs were on a non-interventional 'baseline' diet, the healthiest status (e.g., absence of clinical signs after treatment, in cases of chronic enteropathies), or the sample with the highest throughput. Samples that could not be linked to their metadata were excluded from the representative set of samples.

In addition to publicly available cohorts, we included a newly sequenced cohort from Berlin with 15 pet dogs from households in the context of a food allergy and tolerance study<sup>68</sup>.

### **Reads pre-processing**

Before the read pre-processing step, we run MinIONQC<sup>69</sup> on ONT raw reads to evaluate the quality of the sequencing runs, including the q-score, sequence length, and total throughput.

Short and long reads were pre-processed by quality, length, and presence of host reads. For Illumina short reads, we used NGLess<sup>70</sup>, and raw reads were discarded if they presented a quality

score lower than 25, were less than 45 bp in length, or mapped to the dog genome. For Nanopore long reads, we used Chopper<sup>71</sup>, and raw reads were discarded if they had a quality score lower than 10, were less than 500 bp in length, or mapped to dog genomes. After Chopper, we ran Porechop\_abi<sup>72</sup> to remove any remaining adapters and discard reads with middle adapters.

## **Metagenome assembly and binning**

Pre-processed long reads were used to perform metagenomics assembly with MetaFlye<sup>73</sup>. We performed two to three metagenome assemblies per sample, depending on the initial sequencing depth: all data, 20 Gb, and 10 Gb. Metagenomics assemblies using the whole sequencing data (all data) are the norm throughout the manuscript. Subset metagenomics assemblies were only used to obtain MAGs representing species that were missed with the ‘all data’ strategy (Supplementary Figure S2). The data subsets were generated with Rasusa<sup>74</sup>, indicating the desired sequencing depth.

The MetaFlye metagenome assemblies were further polished using three consecutive polishing steps: Medaka, which used long reads for polishing; followed by Polypolish<sup>75</sup>, which used short reads for polishing; and a final round using Polca<sup>76</sup>, which also used short reads for polishing. After polishing, we binned the contigs using SemiBin2<sup>77</sup> with three different strategies for ‘all data’ assemblies: single-sample binning with short reads, single-sample binning with long reads, and multi-against-multi binning<sup>78</sup>. For the subset assemblies, we exclusively used single-sample binning with short reads.

## **MAG catalog generation and characterization**

To characterize the generated metagenome bins, we used CheckM2<sup>79</sup> to evaluate contamination and completeness, followed by GTDB-tk<sup>80</sup> with the GTDB r214 database to assign taxonomy to the genomic bins. We used DASTool<sup>81</sup> to dereplicate the metagenome bins originating from the same initial read set (all data) with different binning strategies. Finally, we manually included metagenome bins from the subset assemblies if they represented a different taxonomy not captured with the ‘all data’ assembly.

The final collection of MAGs includes only those metagenome bins meeting high- or medium-quality standards based on the MIMAG criteria<sup>13</sup> (Supplementary Table S3). According to these criteria, high-quality MAGs have >90% completeness, <5% contamination, and must contain ribosomal genes and at least 18 out of the 20 tRNAs. Medium-quality MAGs are defined by >50% completeness and <10% contamination. To distinguish between MAGs meeting the full MIMAG high-quality criteria and those meeting only completeness and contamination thresholds (>90%

completeness, <5% contamination), we refer to the former as *near-finished* MAGs and the latter as high-quality MAGs throughout this study.

Apart from checking completeness and contamination, we used barrnap to identify the ribosomal genes, and for cases where the 5S rRNA genes were not predicted, we also used RNAmmer<sup>82</sup>. The extracted 16S rRNA genes were mapped to MicrobeAtlas OTUs<sup>47</sup> using their online API. We predicted tRNAs using tRNAscan<sup>83</sup>. Finally, the species-level MAG representatives were obtained by running dRep<sup>84</sup> at 95% ANI.

### **Manual quality curation on the Shanghai Dog MAG catalog**

During the quality assessment of *Allobaculum stercoricanis* MAGs, we observed that high-quality ones were downgraded to medium-quality after polishing. This shift in quality classification coincided with a change in the default model chosen by CheckM2 during its evaluation (moving from a general to a specific model). Several of these MAGs were single-contig, circular genome assemblies –which typically indicate high completeness and low contamination– but still classified as medium-quality. To address this discrepancy, we manually re-evaluated the quality of *A. stercoricanis* MAGs using the general CheckM2 model throughout all the polishing steps.

Four out of the total six *Klebsiella pneumoniae* MAGs identified in our cohort, derived from consecutive canine fecal samples (D023–D026), presented an identical ARGs pattern and exhibited extremely high ANI values. Samples D023–D026 are all from different households and different neighbourhoods of Shanghai; therefore, we suspected a cross-contamination event had happened during the sample processing. Dog D024 was on antibiotics, and its microbiome consisted of a *K. pneumoniae* MAG with 6,000X coverage and some other very low abundant bacteria potentially contaminating nearby samples.

### **Bacterial strain and extrachromosomal elements sharing**

We compared the sharing patterns of bacterial strains and extrachromosomal elements in the Shanghai dog cohort, considering pairwise comparisons between dogs within the same households vs. between households.

To analyze the sharing patterns of bacterial strains, we divided the number of shared bacterial strains (considering ‘same strain’ those MAGs with ANI > 99%) by the number of shared bacterial species (considering ‘same species’ as those MAGs with ANI > 95%). Two samples were only compared if they presented at least 10 shared bacterial species.

To analyze the sharing patterns of extrachromosomal elements, we used the ‘covered fraction’ table generated by CoverM<sup>85</sup> –based on short-read mapping results to the MAG and extrachromosomal element catalogs. We converted this table into a presence-absence matrix by considering elements with >80% coverage as present. Considering the extrachromosomal elements present in at least 10% of the samples, we calculated sample similarity using Jaccard Similarity.

To evaluate the statistical significance of shared bacterial strains or extrachromosomal elements between dogs from the same household vs. different households, we performed a Mann–Whitney U test.

### **Gene catalog generation**

Open reading frames were predicted with Prodigal in metagenomics mode (-p meta)<sup>86</sup>. ORFs were then clustered at 100% nucleotide identity (with 100% coverage of the shorter sequence) by comparing shorter sequences to longer sequences. This catalog was then clustered at 95% nucleotide identity (with 90% coverage of the shorter sequence) using CD-HIT-EST (command line parameters, -c 0.95 -d 0 -g 1 -G 0 -aS 0.9)<sup>87</sup>.

### **smORF catalog generation**

We ran GMSC-Mapper 0.1.0<sup>88</sup> on the polished metagenome assemblies and recovered the non-redundant small ORFs (with 100% amino acid identity) by removing duplicated sequences. These smORFs were subsequently clustered at 90% identity using CD-HIT with parameters -c 0.9 -n 5 -d 0.

### **Functional annotation of MAGs**

Using polished metagenome assemblies, we predicted protein-coding genes using Prodigal 2.6.3 with ‘-p meta’ option. Using as input the predicted proteins file, we run eggNOG mapper 2.1.12 with the following options: ‘-m diamond --itype proteins’.

### **Antimicrobial-resistant genes prediction**

We predicted the ARGs using two input file types: polished metagenome assemblies (contig-level fasta files) and the MAG catalog. We ran RGI v6.0.3 with the CARD database<sup>38</sup>. We filtered the output table by only keeping those ARGs predicted as ‘strict’ or ‘perfect’ and that presented at least 90% identity and 90% coverage to the hit on the database.

For the website (<https://sh-dog-mags.big-data-biology.org/>), we only included ‘clinically-relevant’ ARGs, which were obtained by filtering the output ARGs to include those also present in the ResFinder v4.0 database<sup>89</sup> by using the database mappings from argNorm 1.0<sup>90</sup>.

### **Putative plasmid mining and identification**

For all samples, contigs identified by Flye as circular were extracted and RecurM v0.2.8 was run on them with the options ‘-c 2 -collapse\_against\_assembly’. Only circular clusters identified by RecurM were considered for further analysis. To label the generated circular clusters, GeNomad v1.1.0<sup>91</sup> was run with default parameters, and any contigs with at least 1 plasmid hallmark gene as defined by GeNomad were considered a plasmid. Putative plasmids were annotated with DRAM v1.5.0<sup>92</sup> with default settings to determine functional gene encoding. To identify putative viruses, VirSorter2 v2.2.4<sup>93</sup> was run with the options ‘--include\_groups dsDNAphage, NCLDV, RNA, ssDNA, lavidaviridae --high-confidence-only’ and any viral sequence identified as ‘full’ in the final-viral-score.tsv file was labelled as virus. All other sequences were considered ‘non-categorised’. Any contigs that were binned into a MAG were discarded.

To identify putative hosts for plasmids, a co-abundance approach was used. MAG and putative plasmid abundance was calculated by mapping metagenomic reads using CoverM with the trimmed mean option. The trimmed mean of the relative abundances of both MAGs and putative plasmids was used as input into FlashWeave v0.19.2<sup>94</sup> using the sensitive, heterogeneous mode to construct a co-abundance table. Relationships between a MAG and a circular cluster with a FlashWeave correlation strength above 0.5 were kept as putative plasmid-host relationships.

## **QUANTIFICATION AND STATISTICAL ANALYSIS**

### **MAG polishing evaluation**

As mentioned above, the final MAG collection was generated by binning the polished contigs, which were generated by metagenome assembly with MetaFlye, followed by three consecutive polishing steps (Medaka - long-read polishing, Polypolish - first round short-read polishing, and Polca - second round short-read polishing). To evaluate the effect of each polishing step, we generated “artificial” genomic bins containing the unpolished versions of the contig of the final MAG catalog for each one of the previous polishing steps (Flye, Medaka, Polypolish).

We evaluated polishing effects by looking at completeness and contamination values obtained by CheckM2. To assess statistical differences in the MAGs completeness and contamination values, we used SciPy library implementation of the Friedman test, followed by the Wilcoxon test on

paired samples. Wilcoxon test on paired samples assessed significant differences between polishing steps (Flye - unpolished, Medaka - long-read polished, Polypolish - long-read + short-read polished, and Polca - long-read + 2x short-read polished), we evaluated if completeness between the polishing steps was higher (alternative='greater') and if contamination was lower (alternative='less'). The final results were considered significant if the q-score obtained after Benjamini-Hochberg FDR correction was  $<0.05$ .

### **Taxonomic profiling and diversity analysis**

To perform read-level analysis, we used short-read sequencing data from canid studies in public databases (up to February 2023) together with our short-read data. We kept the SingleM OTU tables generated by the 13 universal marker genes that target Bacteria and Archaea. For each OTU table, we filtered out OTUs that did not have a total count across all samples of at least 10 and that were not found in at least 2% of the samples. Moreover, we filtered out very low-abundant OTUs by removing any OTU with a mean relative abundance lower than 0.005%. Finally, we discarded samples that did not have at least 200 OTUs detected.

For each OTU table, we computed alpha diversity using the Shannon index and beta diversity using Bray-Curtis dissimilarity metrics (on log-transformed data). We obtained the final alpha and beta diversity objects by computing the median values for the selected marker genes. To assess statistical significance, we used Kruskal-Wallis for alpha diversity and PERMANOVA for beta diversity. We evaluated the effect of living environment, age, size, and sex. For each variable, if a specific category was only represented by a single study, the associated samples were filtered out to avoid batch effects in the statistical analysis.

### **Differential abundance analysis**

Using SingleM species-level taxonomic profiles, we conducted differential abundance analysis using MaAsLin2<sup>95</sup>. For each metadata variable analyzed (living environment, age, size, and sex), *unknown* samples were discarded, and we adjusted for the other variables as relevant covariates. Features with a q-value  $< 0.1$  were considered significant. Additionally, a Kruskal-Wallis test with Benjamini-Hochberg FDR correction ( $q < 0.05$ ) was also conducted. Only taxa and functional features identified as significant by both methods and thresholds were considered robust associations and were further analyzed.

## Comparison of canine long-read MAGs to representative genome assemblies

Using high-quality species-level genome assemblies, we compared our species representatives' long-read MAGs to those from the public databases (GenBank or RefSeq). Firstly, we downloaded the reference genome for each species and ran CheckM2 on them to include only those species with >90% completeness and <5% contamination. We characterized them using the previously described workflow for counting tRNAs and identifying ribosomal genes.

To annotate the Mobilome function (COG category "X") for both reference and long-read canine MAGs, we used eggNOG COG IDs and manually linked them to the COG X category according to NCBI (<https://www.ncbi.nlm.nih.gov/research/cog/cogcategory/X/>). Wilcoxon test on paired samples assessed if the count of mobilome gene functions was higher (alternative = 'greater') in long-read species-level MAG here compared with representative genomes in public datasets. The final results were considered significant if the q-score obtained after Benjamini-Hochberg FDR correction was <0.05.

## Extrachromosomal elements in canids

We mapped the short reads from a total of 406 dog samples from our Shanghai cohort, Berlin cohort, and selected external canid cohorts<sup>7,9-11</sup> to the species-level MAGs and non-redundant ECEs. Reads were first quality filtered (trimmed to min. base quality of 25 using the 'subtrim' strategy, and min. post-trim read size of 45) and then mapped to the dog reference genome with bwa-mem<sup>96</sup> through NGLess<sup>70</sup>, keeping only hits with at least 45bp coverage at 90% identity. Reads mapping to the dog genome were then discarded and the remaining reads mapped to our catalog using the same software and filtering parameters. Using CoverM<sup>85</sup>, we obtained the "covered fraction" table and considered an EC element present if it was covered at least 80%.

Using the Chi-square test on the presence/absence table, we assessed differential prevalence patterns of ECEs across selected canid cohorts, considering several metadata variables: study, living environment, age, size, sex, and body condition. If the Chi-square test was significant, we ran a post-hoc Fisher's Exact test for pairwise comparisons. As a final step, we performed multiple testing correction to all the combined results using the Benjamini-Hochberg FDR test.

Pairwise comparisons that included an unknown or unclassified group were filtered out for further analysis. For each ECE, we considered only the most significant metadata category driving the difference (the lowest corrected p-value). If that was 'Study', the difference is considered technical rather than biological, and is not further discussed.

## Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Critical commercial assays</b>		
Flow Cell R10.4.1 and Ligation Sequencing Kit	Oxford Nanopore Technologies	R10.4.1 SQK-LSK114
NEBNext® Ultra™ DNA Library Prep Kit for Illumina	New England Biolabs	E7370L
<b>Deposited data</b>		
Raw sequencing data	This study	Bioproject ID: PRJEB85799
MAG catalog	This study	Bioproject ID: PRJEB85799 <a href="https://sh-dog-mags.big-data-biology.org/">https://sh-dog-mags.big-data-biology.org/</a>
ECEs catalog	This study	<a href="https://doi.org/10.5281/zenodo.16356977">https://doi.org/10.5281/zenodo.16356977</a>
Gene catalog	This study	<a href="https://doi.org/10.5281/zenodo.16356977">https://doi.org/10.5281/zenodo.16356977</a>
smORFs catalog	This study	<a href="https://doi.org/10.5281/zenodo.16356977">https://doi.org/10.5281/zenodo.16356977</a>
Functional annotations	This study	<a href="https://doi.org/10.5281/zenodo.16356977">https://doi.org/10.5281/zenodo.16356977</a>
RGI annotations	This study	<a href="https://doi.org/10.5281/zenodo.16356977">https://doi.org/10.5281/zenodo.16356977</a>
<b>Software and algorithms</b>		
All original code	This study	<a href="https://github.com/BigDataBiology/ShanghaiDogs">https://github.com/BigDataBiology/ShanghaiDogs</a>

MinIONQC 1.4.1	Lanfear et al., 2019 <sup>69</sup>	<a href="https://github.com/roblanf/minion_qc">https://github.com/roblanf/minion_qc</a>
Chopper 0.5.0	De Coster et al., 2023 <sup>71</sup>	<a href="https://github.com/wdecoster/chopper">https://github.com/wdecoster/chopper</a>
Porechop_abi 0.5.1	Bonenfant et al., 2023 <sup>72</sup>	<a href="https://github.com/bonsai-team/Porechop_ABI">https://github.com/bonsai-team/Porechop_ABI</a>
Rasusa 0.7.1	Hall et al., 2023 <sup>74</sup>	<a href="https://github.com/mbhall88/rasusa">https://github.com/mbhall88/rasusa</a>
Ngless 1.5.0	Coelho et al., 2019 <sup>70</sup>	<a href="https://github.com/ngless-toolkit/ngless">https://github.com/ngless-toolkit/ngless</a>
Flye 2.9.2	Kolmogorov et al., 2020 <sup>73</sup>	<a href="https://github.com/mikolmogorov/Flye">https://github.com/mikolmogorov/Flye</a>
SemiBin2 1.5.1	Pan et al., 2023 <sup>77</sup>	<a href="https://github.com/BigDataBiology/SemiBin">https://github.com/BigDataBiology/SemiBin</a>
Minimap2 2.26-r1175	Li, 2018 <sup>97</sup>	<a href="https://github.com/lh3/minimap2">https://github.com/lh3/minimap2</a>
SAMtools 1.17	Danecek et al., 2021 <sup>98</sup>	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
BWA v0.7.19	Li, 2013 <sup>96</sup>	<a href="https://github.com/lh3/bwa">https://github.com/lh3/bwa</a>
Strobealign v0.15	Shalin, 2022 <sup>99</sup> Pan et al., 2025 <sup>78</sup>	<a href="https://github.com/ksahlin/strobealign">https://github.com/ksahlin/strobealign</a>
MaAsLin2 1.16.0	Mallick et al., 2021 <sup>95</sup>	<a href="https://github.com/biobakery/Maaslin2">https://github.com/biobakery/Maaslin2</a>
Medaka 1.8.1	Oxford Nanopore Technologies	<a href="https://github.com/nanoporetech/medaka">https://github.com/nanoporetech/medaka</a>
Polypolish 0.5.0	Wick et al., 2022 <sup>75</sup>	<a href="https://github.com/rrwick/Polypolish">https://github.com/rrwick/Polypolish</a>

polca.py from Masurca 4.1.0	Zimin et al., 2017 <sup>76</sup>	<a href="https://github.com/alekseyzimin/masurca">https://github.com/alekseyzimin/masurca</a>
DASTool 1.1.6	Sieber et al., 2018 <sup>81</sup>	<a href="https://github.com/cmks/DAS_Tool">https://github.com/cmks/DAS_Tool</a>
dRep 3.5.0	Olm et al., 2017 <sup>84</sup>	<a href="https://github.com/MrOlm/drep">https://github.com/MrOlm/drep</a>
GTDBtk 2.3.0	Chaumeil et al., 2020 <sup>80</sup>	<a href="https://github.com/Ecogenomics/GTDBTk">https://github.com/Ecogenomics/GTDBTk</a>
GTDB database r214	Parks et al., 2018 <sup>100</sup>	<a href="https://gtdb.ecogenomic.org/">https://gtdb.ecogenomic.org/</a>
CheckM2 1.0.1	Chklovski et al., 2023 <sup>79</sup>	<a href="https://github.com/chklovski/CheckM2">https://github.com/chklovski/CheckM2</a>
RGI 6.0.3	Alcock et al., 2023 <sup>38</sup>	<a href="https://github.com/arpcard/rgi/">https://github.com/arpcard/rgi/</a>
Prodigal 2.6.3	Hyatt et al., 2010 <sup>86</sup>	<a href="https://github.com/hyattpd/Prodigal">https://github.com/hyattpd/Prodigal</a>
Barrnap 0.9	Seemann, 2013	<a href="https://github.com/tseemann/barrnap">https://github.com/tseemann/barrnap</a>
RNAmmmer v1.2	Lagesen et al., 2007 <sup>82</sup>	<a href="https://services.healthtech.dtu.dk/services/RNAmmmer-1.2/">https://services.healthtech.dtu.dk/services/RNAmmmer-1.2/</a>
tRNAscan-SE 2.0.12	Chan et al., 2021 <sup>83</sup>	<a href="https://github.com/UCSC-LoweLab/tRNAscan-SE">https://github.com/UCSC-LoweLab/tRNAscan-SE</a>
EggNOG mapper 2.1.12-cbe169f	Cantalapiedra et al., 2021 <sup>101</sup> Huerta-Cepas et al., 2019 <sup>102</sup>	<a href="https://github.com/eggnogetdb/eggnoget-mapper">https://github.com/eggnogetdb/eggnoget-mapper</a>
CoverM v0.7.0	Aroney S, et al., 2025 <sup>85</sup>	<a href="https://github.com/wwood/CoverM">https://github.com/wwood/CoverM</a>
RecurM 0.2.8	-	<a href="https://github.com/dn-ra/RecurM">https://github.com/dn-ra/RecurM</a>

SingleM 0.16.0	Woodcroft et al., 2025 <sup>42</sup>	<a href="https://github.com/wwood/singlem">https://github.com/wwood/singlem</a>
CD-HIT v4.8.1	Fu et al., 2012 <sup>87</sup>	<a href="https://sites.google.com/view/cd-hit">https://sites.google.com/view/cd-hit</a>
Other		
Public dog genome	ENSEMBL	GCA_000002285.4
Public raw reads from canids	ENA	Supplementary Table S9 for Bioproject IDs

**Acknowledgements:** The authors thank Alexandre Areias Castro (QUT) for assistance with sequence deposition, Chengkai Zhu (Fudan University) for support with logistics, Marion Draheim and Richard Lo-Man (Institut Pasteur Shanghai) for support with sample logistics, and the dog owners for their participation. Members of the Coelho group and the red herons group (QUT) are thanked for their feedback and suggestions throughout the project.

**Funding:** National Natural Science Foundation of China (RFIS-I, 32250410281) (A.C.), the Australian Research Council (#FT230100724) (L.P.C.), the National Health and Medical Research Council of Australia (under the framework of JPI AMR, #2031902, SEARCHER) (L.P.C.), and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the clinical research unit (CRU339): Food allergy and tolerance (FOOD@) – 409525714 (U.L.).

**Declaration of interests:** A.Cu. is a partner at Nano1Health, SL, and has previously been invited by Oxford Nanopore Technologies (ONT) to participate in conferences. These activities did not influence the results or conclusions of this work. All other authors declare no competing interests.

**Author contributions:** A.Cu. and L.P.C. designed the study. A.Cu. and F.G. collected the samples and the dog associated information. A.Cu., Y.D., A.Ch., S.P., X.-M.Z., and L.P.C. analyzed data. A.Cu., N.K., and L.P.C. visualized data. S.F., S.L., and U.L. provided the data from the Berlin dog cohort. A.Cu., X.-M.Z., and L.P.C. provided funding. L.P.C. supervised the project. A.Cu. wrote the first draft of the manuscript. All authors contributed to the revision of the manuscript prior to submission and approved the final version.

# REFERENCES

1. Song, S.J., Lauber, C., Costello, E.K., Lozupone, C.A., Humphrey, G., Berg-Lyons, D., Caporaso, J.G., Knights, D., Clemente, J.C., Nakielny, S., et al. (2013). Cohabiting family members share microbiota with one another and with their dogs. *eLife* 2, e00458. <https://doi.org/10.7554/eLife.00458>.
2. Tomasulo, A., Simionati, B., and Facchin, S. (2024). Microbiome One Health model for a healthy ecosystem. *Sci One Health* 3, 100065. <https://doi.org/10.1016/j.soh.2024.100065>.
3. Pilla, R., and Suchodolski, J.S. (2020). The Role of the Canine Gut Microbiome and Metabolome in Health and Gastrointestinal Disease. *Front. Vet. Sci.* 6. <https://doi.org/10.3389/fvets.2019.00498>.
4. Tun, H.M., Konya, T., Takaro, T.K., Brook, J.R., Chari, R., Field, C.J., Guttman, D.S., Becker, A.B., Mandhane, P.J., Turvey, S.E., et al. (2017). Exposure to household furry pets influences the gut microbiota of infant at 3-4 months following various birth scenarios. *Microbiome* 5, 40. <https://doi.org/10.1186/s40168-017-0254-x>.
5. Mäki, J.M., Kirjavainen, P.V., Täubel, M., Piippo-Savolainen, E., Backman, K., Hyvärinen, A., Tuoresmäki, P., Jayaprakash, B., Heinrich, J., Herberth, G., et al. (2021). Associations between dog keeping and indoor dust microbiota. *Sci Rep* 11, 5341. <https://doi.org/10.1038/s41598-021-84790-w>.
6. Bryce, C.M. (2021). Dogs as Pets and Pests: Global Patterns of Canine Abundance, Activity, and Health. *Integr. Comp. Biol.* 61, 154–165. <https://doi.org/10.1093/icb/icab046>.
7. Coelho, L.P., Kultima, J.R., Costea, P.I., Fournier, C., Pan, Y., Czarnecki-Maulden, G., Hayward, M.R., Forslund, S.K., Schmidt, T.S.B., Descombes, P., et al. (2018). Similarity of the dog and human gut microbiomes in gene content and response to diet. *Microbiome* 6, 72. <https://doi.org/10.1186/s40168-018-0450-3>.
8. Wang, S., Martins, R., Sullivan, M.C., Friedman, E.S., Misic, A.M., El-Fahmawi, A., De Martinis, E.C.P., O'Brien, K., Chen, Y., Bradley, C., et al. (2019). Diet-induced remission in chronic enteropathy is associated with altered microbial community structure and synthesis of secondary bile acids. *Microbiome* 7, 126. <https://doi.org/10.1186/s40168-019-0740-4>.
9. Allaway, D., Haydock, R., Lonsdale, Z.N., Deusch, O.D., O'Flynn, C., and Hughes, K.R. (2020). Rapid Reconstitution of the Fecal Microbiome after Extended Diet-Induced Changes Indicates a Stable Gut Microbiome in Healthy Adult Dogs. *Appl Environ Microbiol* 86, e00562-20. <https://doi.org/10.1128/AEM.00562-20>.
10. Tanprasertsuk, J., Shmalberg, J., Maughan, H., Tate, D.E., Perry, L.M., Jha, A.R., and Honaker, R.W. (2021). Heterogeneity of gut microbial responses in healthy household dogs transitioning from an extruded to a mildly cooked diet. *PeerJ* 9, e11648. <https://doi.org/10.7717/peerj.11648>.

11. Yarlagadda, K., Zachwieja, A.J., de Flamingh, A., Phungviwatnikul, T., Rivera-Colón, A.G., Roseman, C., Shackelford, L., Swanson, K.S., and Malhi, R.S. (2022). Geographically diverse canid sampling provides novel insights into pre-industrial microbiomes. *Proc. R. Soc. B.* 289, 20220052. <https://doi.org/10.1098/rspb.2022.0052>.
12. Branck, T., Hu, Z., Nickols, W.A., Walsh, A.M., Bhosle, A., Short, M.I., Nearing, J.T., Asnicar, F., McIver, L.J., Maharjan, S., et al. (2024). Comprehensive profile of the companion animal gut microbiome integrating reference-based and reference-free methods. *ISME J.* 18, wræ201. <https://doi.org/10.1093/ismejo/wrae201>.
13. Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloë-Fadrosh, E.A., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35, 725–731. <https://doi.org/10.1038/nbt.3893>.
14. Sereika, M., Kirkegaard, R.H., Karst, S.M., Michaelsen, T.Y., Sørensen, E.A., Wollenberg, R.D., and Albertsen, M. (2022). Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods* 19, 823–826. <https://doi.org/10.1038/s41592-022-01539-7>.
15. Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z.J., Pollard, K.S., Sakharova, E., Parks, D.H., Hugenholtz, P., et al. (2020). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* 39, 1–10. <https://doi.org/10.1038/s41587-020-0603-3>.
16. Yuan, C., Lei, J., Cole, J., and Sun, Y. (2015). Reconstructing 16S rRNA genes in metagenomic data. *Bioinformatics* 31, i35-43. <https://doi.org/10.1093/bioinformatics/btv231>.
17. Maguire, F., Jia, B., Gray, K.L., Lau, W.Y.V., Beiko, R.G., and Brinkman, F.S.L. (2020). Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic Islands. *Microb Genom* 6, e000436. <https://doi.org/10.1099/mgen.0.000436>.
18. Abramova, A., Karkman, A., and Bengtsson-Palme, J. (2024). Metagenomic assemblies tend to break around antibiotic resistance genes. *BMC Genomics* 25, 959. <https://doi.org/10.1186/s12864-024-10876-0>.
19. Bertrand, D., Shaw, J., Kalathiyappan, M., Ng, A.H.Q., Kumar, M.S., Li, C., Dvornicic, M., Soldo, J.P., Koh, J.Y., Tong, C., et al. (2019). Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat. Biotechnol.* 37, 937–944. <https://doi.org/10.1038/s41587-019-0191-2>.
20. Singleton, C.M., Petriglieri, F., Kristensen, J.M., Kirkegaard, R.H., Michaelsen, T.Y., Andersen, M.H., Kondrotaitė, Z., Karst, S.M., Dueholm, M.S., Nielsen, P.H., et al. (2021). Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated

- sludge using long-read sequencing. *Nat Commun* 12, 2009. <https://doi.org/10.1038/s41467-021-22203-2>.
21. Suzuki, Y., Nishijima, S., Furuta, Y., Yoshimura, J., Suda, W., Oshima, K., Hattori, M., and Morishita, S. (2019). Long-read metagenomic exploration of extrachromosomal mobile genetic elements in the human gut. *Microbiome* 7, 119. <https://doi.org/10.1186/s40168-019-0737-z>.
22. Arumugam, K., Bessarab, I., Haryono, M.A.S., Liu, X., Zuniga–Montanez, R.E., Roy, S., Qiu, G., Drautz–Moses, D.I., Law, Y.Y., Wuertz, S., et al. (2021). Recovery of complete genomes and non-chromosomal replicons from activated sludge enrichment microbial communities with long read metagenome sequencing. *npj Biofilms Microbiomes* 7, 1–13. <https://doi.org/10.1038/s41522-021-00196-6>.
23. Stewart, R.D., Auffret, M.D., Warr, A., Walker, A.W., Roehe, R., and Watson, M. (2019). Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* 37, 953–961. <https://doi.org/10.1038/s41587-019-0202-3>.
24. Moss, E.L., Maghini, D.G., and Bhatt, A.S. (2020). Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat. Biotechnol.* 38, 701–707. <https://doi.org/10.1038/s41587-020-0422-6>.
25. Cuscó, A., Pérez, D., Viñes, J., Fàbregas, N., and Francino, O. (2021). Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces. *BMC Genomics* 22, 330. <https://doi.org/10.1186/s12864-021-07607-0>.
26. Cuscó, A., Pérez, D., Viñes, J., Fàbregas, N., and Francino, O. (2022). Novel canine high-quality metagenome-assembled genomes, prophages and host-associated plasmids provided by long-read metagenomics together with Hi-C proximity ligation. *Microb Genom* 8, 000802. <https://doi.org/10.1099/mgen.0.000802>.
27. Jarett, J.K., Kingsbury, D.D., Dahlhausen, K.E., and Ganz, H.H. (2021). Best Practices for Microbiome Study Design in Companion Animal Research. *Front. Vet. Sci.* 8. <https://doi.org/10.3389/fvets.2021.644836>.
28. Rojas, C.A., Park, B., Scarsella, E., Jospin, G., Entrolezo, Z., Jarett, J.K., Martin, A., and Ganz, H.H. (2024). Species-level characterization of the core microbiome in healthy dogs using full-length 16S rRNA gene sequencing. *Front. Vet. Sci.* 11. <https://doi.org/10.3389/fvets.2024.1405470>.

29. Maghini, D.G., Moss, E.L., Vance, S.E., and Bhatt, A.S. (2021). Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome. *Nat Protoc* 16, 458–471. <https://doi.org/10.1038/s41596-020-00424-x>.
30. Větrovský, T., and Baldrian, P. (2013). The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLOS ONE* 8, e57923. <https://doi.org/10.1371/journal.pone.0057923>.
31. Greetham, H.L., Collins, M.D., Gibson, G.R., Giffard, C., Falsen, E., and Lawson, P.A. (2004). *Sutterella stercoricanis* sp. nov., isolated from canine faeces. *Int J Syst Evol Microbiol* 54, 1581–1584. <https://doi.org/10.1099/ijss.0.63098-0>.
32. Kawata, M., Tsukamoto, A., Isozaki, R., Nobukawa, S., Kawahara, N., Akutsu, S., Suzuki, M., and Asanuma, N. (2018). *Glucerbacter canisensis* gen. nov., sp. nov., isolated from dog feces and its effect on the hydrolysis of plant glucosylceramide in the intestine of dogs. *Arch Microbiol* 200, 505–515. <https://doi.org/10.1007/s00203-017-1463-1>.
33. Martiny, H.-M., Munk, P., Brinch, C., Aarestrup, F.M., Calle, M.L., and Petersen, T.N. (2024). Utilizing co-abundances of antimicrobial resistance genes to identify potential co-selection in the resistome. *Microbiol Spectr* 12, e04108-23. <https://doi.org/10.1128/spectrum.04108-23>.
34. Steagall, P.V., Pelligand, L., Page, S., Granick, J.L., Allerton, F., Bęczkowski, P.M., Weese, J.S., Hrček, A.K., Queiroga, F., and Guardabassi, L. (2023). The 2023 World Small Animal Veterinary Association (WSAVA): List of essential medicines for cats and dogs. *J Small Anim Pract* 64, 731–748. <https://doi.org/10.1111/jsap.13673>.
35. Partridge, S.R., Kwong, S.M., Firth, N., and Jensen, S.O. (2018). Mobile Genetic Elements Associated with Antimicrobial Resistance. *Clin Microbiol Rev* 31, 10.1128/cmr.00088-17. <https://doi.org/10.1128/cmr.00088-17>.
36. Voha, C., Docquier, J.-D., Rossolini, G.M., and Fosse, T. (2006). Genetic and Biochemical Characterization of FUS-1 (OXA-85), a Narrow-Spectrum Class D  $\beta$ -Lactamase from *Fusobacterium nucleatum* subsp. polymorphum. *Antimicrob Agents Chemother* 50, 2673–2679. <https://doi.org/10.1128/aac.00058-06>.
37. Perry, M.D., Vranckx, K., Copsey-Mawer, S., Scotford, S., Anderson, B., Day, P., Watkins, J., Corden, S., Hughes, H., and Morris, T.E. (2023). First large-scale study of antimicrobial susceptibility data, and genetic resistance determinants, in *Fusobacterium necrophorum* highlighting the importance of continuing focused susceptibility trend surveillance. *Anaerobe* 80, 102717. <https://doi.org/10.1016/j.anaerobe.2023.102717>.

38. Alcock, B.P., Huynh, W., Chalil, R., Smith, K.W., Raphenya, A.R., Wlodarski, M.A., Edalatmand, A., Petkau, A., Syed, S.A., Tsang, K.K., et al. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res* 51, D690–D699. <https://doi.org/10.1093/nar/gkac920>.
39. Salverda, M.L.M., De Visser, J.A.G.M., and Barlow, M. (2010). Natural evolution of TEM-1  $\beta$ -lactamase: experimental reconstruction and clinical relevance. *FEMS Microbiology Reviews* 34, 1015–1036. <https://doi.org/10.1111/j.1574-6976.2010.00222.x>.
40. Berryman, D.I., Lyristis, M., and Rood, J.I. (1994). Cloning and sequence analysis of ermQ, the predominant macrolide-lincosamide-streptogramin B resistance gene in *Clostridium perfringens*. *Antimicrob Agents Chemother* 38, 1041–1046. <https://doi.org/10.1128/aac.38.5.1041>.
41. Fang, M., Yuan, Y., Tian, X., Liu, Y., Wu, K., Zhu, Y., Zhang, L., Fox, E.M., Li, R., Bai, L., et al. (2025). Genomic and antimicrobial resistance profiles of *Clostridium perfringens* isolated from pets in China. *Vet Microbiol* 304, 110490. <https://doi.org/10.1016/j.vetmic.2025.110490>.
42. Woodcroft, B.J., Aroney, S.T.N., Zhao, R., Cunningham, M., Mitchell, J.A.M., Nurdiansyah, R., Blackall, L., and Tyson, G.W. (2025). Comprehensive taxonomic identification of microbial species in metagenomic data using SingleM and Sandpiper. *Nat Biotechnol*, 1–6. <https://doi.org/10.1038/s41587-025-02738-1>.
43. Witt, K.E., Yarlagadda, K., Allen, J.M., Bader, A.C., Simon, M.L., Kuehn, S.R., Swanson, K.S., Cross, T.-W.L., Hedman, K.M., Ambrose, S.H., et al. (2021). Integrative analysis of DNA, macroscopic remains and stable isotopes of dog coprolites to reconstruct community diet. *Sci Rep* 11, 3113. <https://doi.org/10.1038/s41598-021-82362-6>.
44. Jang, H.-J., Son, S., Kim, J.-A., Jung, M.Y., Choi, Y., Kim, D.-H., Lee, H.K., Shin, D., and Kim, Y. (2021). Characterization and Functional Test of Canine Probiotics. *Front. Microbiol.* 12. <https://doi.org/10.3389/fmicb.2021.625562>.
45. Herstad, K.M.V., Moen, A.E.F., Gaby, J.C., Moe, L., and Skancke, E. (2018). Characterization of the fecal and mucosa-associated microbiota in dogs with colorectal epithelial tumors. *PLOS ONE* 13, e0198342. <https://doi.org/10.1371/journal.pone.0198342>.
46. Kwon, G., Lee, J., Koh, J.-H., and Lim, Y.-H. (2018). Lifespan Extension of *Caenorhabditis elegans* by *Butyricicoccus pullicaecorum* and *Megasphaera elsdenii* with Probiotic Potential. *Curr Microbiol* 75, 557–564. <https://doi.org/10.1007/s00284-017-1416-6>.

47. Rodrigues, J.F.M., Tackmann, J., Malfertheiner, L., Patsch, D., Perez-Molphe-Montoya, E., Näpflin, N., Gaio, D., Rot, G., Danaila, M., Peluso, M.E., et al. (2025). The MicrobeAtlas database: Global trends and insights into Earth's microbial ecosystems. Preprint at bioRxiv, <https://doi.org/10.1101/2025.07.18.665519> <https://doi.org/10.1101/2025.07.18.665519>.
48. Coelho, L.P., Alves, R., del Río, Á.R., Myers, P.N., Cantalapiedra, C.P., Giner-Lamia, J., Schmidt, T.S., Mende, D.R., Orakov, A., Letunic, I., et al. (2022). Towards the biogeography of prokaryotic genes. *Nature* 601, 252–256. <https://doi.org/10.1038/s41586-021-04233-4>.
49. Lyu, T., Liu, G., Zhang, H., Wang, L., Zhou, S., Dou, H., Pang, B., Sha, W., and Zhang, H. (2018). Changes in feeding habits promoted the differentiation of the composition and function of gut microbiotas between domestic dogs (*Canis lupus familiaris*) and gray wolves (*Canis lupus*). *AMB Expr* 8, 123. <https://doi.org/10.1186/s13568-018-0652-x>.
50. Liu, Y., Liu, B., Liu, C., Hu, Y., Liu, C., Li, X., Li, X., Zhang, X., Irwin, D.M., Wu, Z., et al. (2021). Differences in the gut microbiomes of dogs and wolves: roles of antibiotics and starch. *BMC Vet Res* 17, 112. <https://doi.org/10.1186/s12917-021-02815-y>.
51. Rampelli, S., Turrone, S., Debandi, F., Alberdi, A., Schnorr, S.L., Hofman, C.A., Taddia, A., Helg, R., Biagi, E., Brigidi, P., et al. (2021). The gut microbiome buffers dietary adaptation in Bronze Age domesticated dogs. *iScience* 24. <https://doi.org/10.1016/j.isci.2021.102816>.
52. Hagan, R.W., Hofman, C.A., Hübner, A., Reinhard, K., Schnorr, S., Lewis Jr, C.M., Sankaranarayanan, K., and Warinner, C.G. (2020). Comparison of extraction methods for recovering ancient microbial DNA from paleofeces. *American Journal of Physical Anthropology* 171, 275–284. <https://doi.org/10.1002/ajpa.23978>.
53. Bai, H., Liu, T., Wang, S., Shen, L., and Wang, Z. (2023). Variations in gut microbiome and metabolites of dogs with acute diarrhea in poodles and Labrador retrievers. *Arch Microbiol* 205, 97. <https://doi.org/10.1007/s00203-023-03439-6>.
54. Xu, H., Zhao, F., Hou, Q., Huang, W., Liu, Y., Zhang, H., and Sun, Z. (2019). Metagenomic analysis revealed beneficial effects of probiotics in improving the composition and function of the gut microbiota in dogs with diarrhoea. *Food Funct.* 10, 2618–2629. <https://doi.org/10.1039/C9FO00087A>.
55. Chen, L., Sun, M., Xu, D., Gao, Z., Shi, Y., Wang, S., and Zhou, Y. (2022). Gut microbiome of captive wolves is more similar to domestic dogs than wild wolves indicated by metagenomics study. *Front. Microbiol.* 13. <https://doi.org/10.3389/fmicb.2022.1027188>.

56. Youngblut, N.D., Cuesta-Zuluaga, J. de la, Reischer, G.H., Dauser, S., Schuster, N., Walzer, C., Stalder, G., Farnleitner, A.H., and Ley, R.E. (2020). Large-Scale Metagenome Assembly Reveals Novel Animal-Associated Microbial Genomes, Biosynthetic Gene Clusters, and Other Genetic Diversity. *mSystems* 5, e01045-20. <https://doi.org/10.1128/mSystems.01045-20>.
57. Alessandri, G., Milani, C., Mancabelli, L., Mangifesta, M., Lugli, G.A., Viappiani, A., Duranti, S., Turroni, F., Ossiprandi, M.C., van Sinderen, D., et al. (2019). Metagenomic dissection of the canine gut microbiota: insights into taxonomic, metabolic and nutritional features. *Environ Microbiol* 21, 1331–1343. <https://doi.org/10.1111/1462-2920.14540>.
58. Pehrsson, E.C., Tsukayama, P., Patel, S., Mejía-Bautista, M., Sosa-Soto, G., Navarrete, K.M., Calderon, M., Cabrera, L., Hoyos-Arango, W., Bertoli, M.T., et al. (2016). Interconnected microbiomes and resistomes in low-income human habitats. *Nature* 533, 212–216. <https://doi.org/10.1038/nature17672>.
59. Guardabassi, L., Schwarz, S., and Lloyd, D.H. (2004). Pet animals as reservoirs of antimicrobial-resistant bacteria: Review. *Journal of Antimicrobial Chemotherapy* 54, 321–332. <https://doi.org/10.1093/jac/dkh332>.
60. Manges, A.R., Geum, H.M., Guo, A., Edens, T.J., Fibke, C.D., and Pitout, J.D.D. (2019). Global Extraintestinal Pathogenic Escherichia coli (ExPEC) Lineages. *Clin Microbiol Rev* 32, e00135-18. <https://doi.org/10.1128/CMR.00135-18>.
61. Benavides, J.A., Salgado-Caxito, M., Opazo-Capurro, A., González Muñoz, P., Piñeiro, A., Otto Medina, M., Rivas, L., Munita, J., and Millán, J. (2021). ESBL-Producing Escherichia coli Carrying CTX-M Genes Circulating among Livestock, Dogs, and Wild Mammals in Small-Scale Farms of Central Chile. *Antibiotics* 10, 510. <https://doi.org/10.3390/antibiotics10050510>.
62. Ko, K.K.K., Chng, K.R., and Nagarajan, N. (2022). Metagenomics-enabled microbial surveillance. *Nat Microbiol* 7, 486–496. <https://doi.org/10.1038/s41564-022-01089-w>.
63. Aizpurua, O., Botnen, A.B., Eisenhofer, R., Odriozola, I., Santos-Bay, L., Bjørnsen, M.B., Gilbert, M.T.P., and Alberdi, A. (2025). Functional Insights Into the Effect of Feralisation on the Gut Microbiota of Cats Worldwide. *Mol Ecol* 34, e17695. <https://doi.org/10.1111/mec.17695>.
64. Jha, A.R., Shmalberg, J., Tanprasertsuk, J., Perry, L., Massey, D., and Honaker, R.W. (2020). Characterization of gut microbiomes of household pets in the United States using a direct-to-consumer approach. *PLOS ONE* 15, e0227289. <https://doi.org/10.1371/journal.pone.0227289>.

65. Lehtimäki, J., Sinkko, H., Hielm-Björkman, A., Salmela, E., Tiira, K., Laatikainen, T., Mäkeläinen, S., Kaukonen, M., Uusitalo, L., Hanski, I., et al. (2018). Skin microbiota and allergic symptoms associate with exposure to environmental microbes. *PNAS* 115, 4897–4902. <https://doi.org/10.1073/pnas.1719785115>.
66. Harvey, N.D. (2021). How Old Is My Dog? Identification of Rational Age Groupings in Pet Dogs Based Upon Normative Age-Linked Processes. *Front. Vet. Sci.* 8. <https://doi.org/10.3389/fvets.2021.643085>.
67. Salt, C., Morris, P.J., German, A.J., Wilson, D., Lund, E.M., Cole, T.J., and Butterwick, R.F. (2017). Growth standard charts for monitoring bodyweight in dogs of different sizes. *PLOS ONE* 12, e0182064. <https://doi.org/10.1371/journal.pone.0182064>.
68. Worm, M., Alexiou, A., Höfer, V., Birkner, T., Jeanrenaud, A.C.S.N., Fauchère, F., Pazur, K., Steinert, C., Arnau-Soler, A., Banerjee, P., et al. (2022). An interdisciplinary approach to characterize peanut-allergic patients—First data from the FOOD@ consortium. *Clin Transl Allergy* 12, e12197. <https://doi.org/10.1002/ctt2.12197>.
69. Lanfear, R., Schalamun, M., Kainer, D., Wang, W., and Schwessinger, B. (2019). MinIONQC: fast and simple quality control for MinION sequencing data. *Bioinformatics* 35, 523–525. <https://doi.org/10.1093/bioinformatics/bty654>.
70. Coelho, L.P., Alves, R., Monteiro, P., Huerta-Cepas, J., Freitas, A.T., and Bork, P. (2019). NG-meta-profiler: fast processing of metagenomes using NGLess, a domain-specific language. *Microbiome* 7, 84. <https://doi.org/10.1186/s40168-019-0684-8>.
71. De Coster, W., and Rademakers, R. (2023). NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 39, btad311. <https://doi.org/10.1093/bioinformatics/btad311>.
72. Bonenfant, Q., Noé, L., and Touzet, H. (2023). Porechop\_ABI: discovering unknown adapters in Oxford Nanopore Technology sequencing reads for downstream trimming. *Bioinform Adv* 3, vbac085. <https://doi.org/10.1093/bioadv/vbac085>.
73. Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., et al. (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* 17, 1103–1110. <https://doi.org/10.1038/s41592-020-00971-x>.
74. Hall, M.B. (2022). Rasusa: Randomly subsample sequencing reads to a specified coverage. *J Open Source Softw* 7, 3941. <https://doi.org/10.21105/joss.03941>.

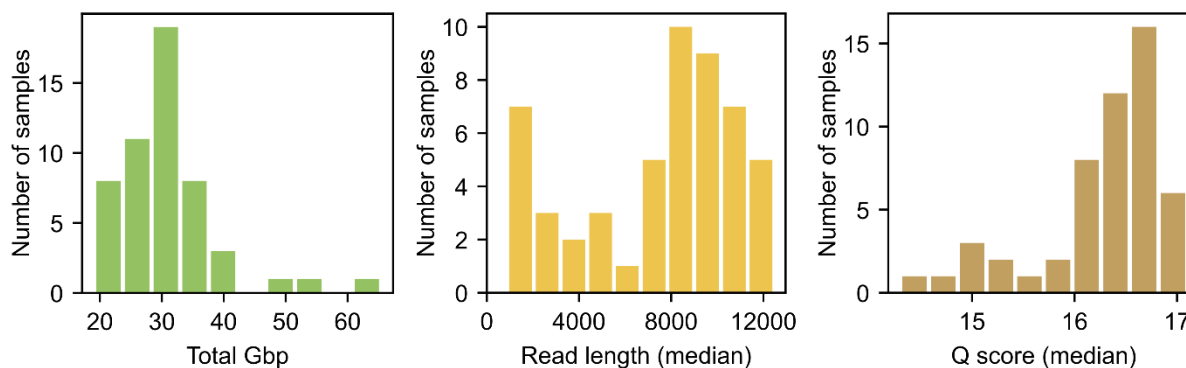
75. Wick, R.R., and Holt, K.E. (2022). Polypolish: Short-read polishing of long-read bacterial genome assemblies. *PLoS Comput Biol* 18, e1009802. <https://doi.org/10.1371/journal.pcbi.1009802>.
76. Zimin, A.V., Puiu, D., Luo, M.-C., Zhu, T., Koren, S., Marçais, G., Yorke, J.A., Dvořák, J., and Salzberg, S.L. (2017). Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* 27, 787–792. <https://doi.org/10.1101/gr.213405.116>.
77. Pan, S., Zhao, X.-M., and Coelho, L.P. (2023). SemiBin2: self-supervised contrastive learning leads to better MAGs for short- and long-read sequencing. *Bioinformatics* 39, i21–i29. <https://doi.org/10.1093/bioinformatics/btad209>.
78. Pan, S., Tolstoganov, I., Sahlin, K., Martin, M., Zhao, X.-M., and Coelho, L.P. (2025). AEMB: efficient abundance estimation for metagenomic binning. Preprint at bioRxiv, <https://doi.org/10.1101/2025.07.30.667338> <https://doi.org/10.1101/2025.07.30.667338>.
79. Chklovski, A., Parks, D.H., Woodcroft, B.J., and Tyson, G.W. (2023). CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 20, 1203–1212. <https://doi.org/10.1038/s41592-023-01940-w>.
80. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36, 1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
81. Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., and Banfield, J.F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 3, 836–843. <https://doi.org/10.1038/s41564-018-0171-1>.
82. Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35, 3100–3108. <https://doi.org/10.1093/nar/gkm160>.
83. Chan, P.P., Lin, B.Y., Mak, A.J., and Lowe, T.M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research* 49, 9077–9096. <https://doi.org/10.1093/nar/gkab688>.
84. Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 11, 2864–2868. <https://doi.org/10.1038/ismej.2017.126>.

85. Aroney, S.T.N., Newell, R.J.P., Nissen, J.N., Camargo, A.P., Tyson, G.W., and Woodcroft, B.J. (2025). CoverM: read alignment statistics for metagenomics. *Bioinformatics* *41*, btaf147. <https://doi.org/10.1093/bioinformatics/btaf147>.
86. Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* *11*, 119. <https://doi.org/10.1186/1471-2105-11-119>.
87. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* *28*, 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>.
88. Duan, Y., Santos-Júnior, C.D., Schmidt, T.S., Fullam, A., de Almeida, B.L.S., Zhu, C., Kuhn, M., Zhao, X.-M., Bork, P., and Coelho, L.P. (2024). A catalog of small proteins from the global microbiome. *Nat Commun* *15*, 7563. <https://doi.org/10.1038/s41467-024-51894-6>.
89. Florensa, A.F., Kaas, R.S., Clausen, P.T.L.C., Aytan-Aktug, D., and Aarestrup, F.M. (2022). ResFinder – an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb Genom* *8*, 000748. <https://doi.org/10.1099/mgen.0.000748>.
90. Ugarcina Perovic, S., Ramji, V., Chong, H., Duan, Y., Maguire, F., and Coelho, L.P. (2025). argNorm: normalization of antibiotic resistance gene annotations to the Antibiotic Resistance Ontology (ARO). *Bioinformatics* *41*, btaf173. <https://doi.org/10.1093/bioinformatics/btaf173>.
91. Camargo, A.P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P.S.G., Nayfach, S., and Kyrpides, N.C. (2024). Identification of mobile genetic elements with geNomad. *Nat Biotechnol* *42*, 1303–1312. <https://doi.org/10.1038/s41587-023-01953-y>.
92. Shaffer, M., Borton, M.A., McGivern, B.B., Zayed, A.A., La Rosa, S.L., Solden, L.M., Liu, P., Narrowe, A.B., Rodríguez-Ramos, J., Bolduc, B., et al. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* *48*, 8883–8900. <https://doi.org/10.1093/nar/gkaa621>.
93. Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O., Pratama, A.A., Gazitúa, M.C., Vik, D., Sullivan, M.B., et al. (2021). VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* *9*, 37. <https://doi.org/10.1186/s40168-020-00990-y>.

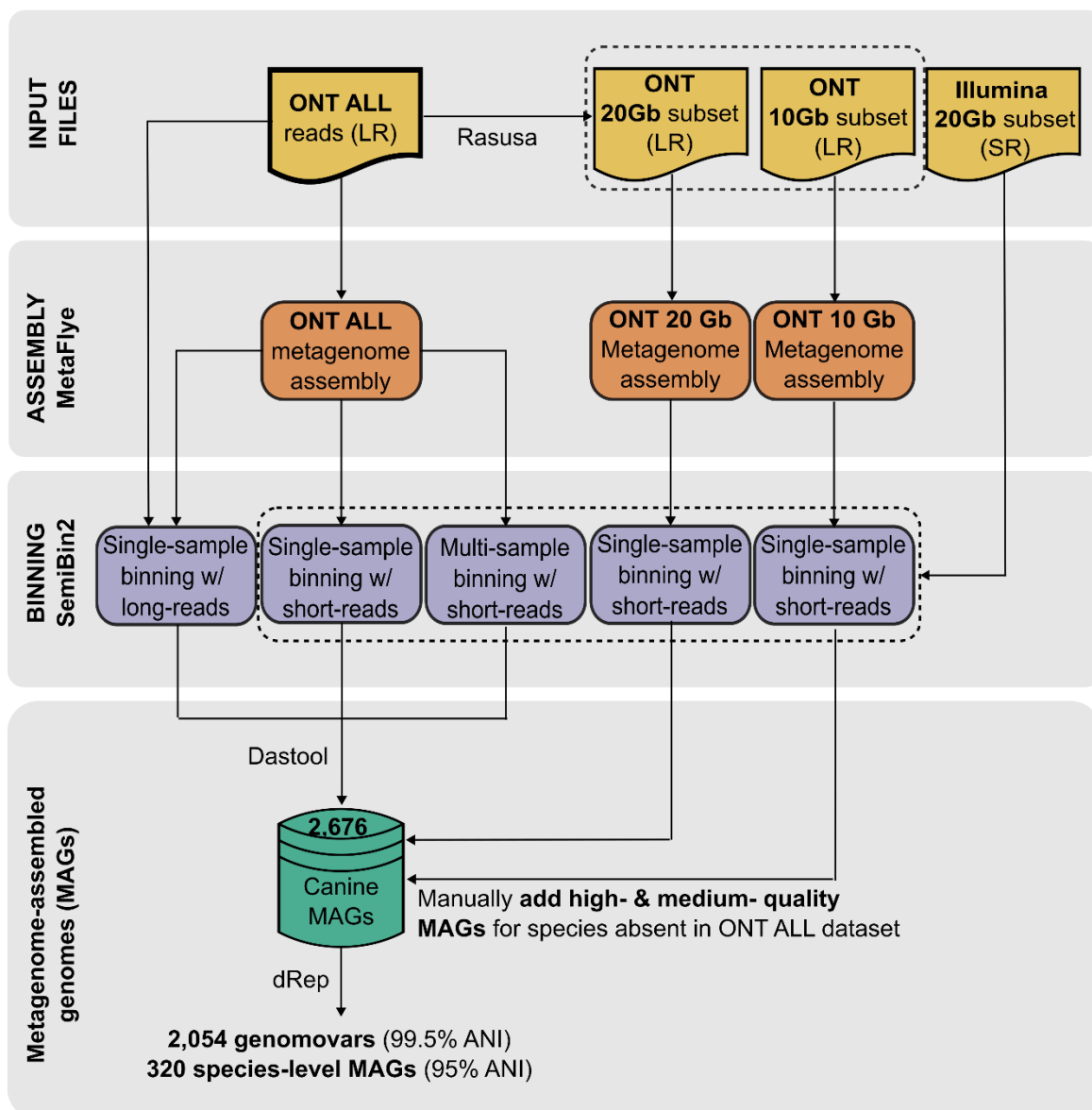
94. Tackmann, J., Matias Rodrigues, J.F., and von Mering, C. (2019). Rapid Inference of Direct Interactions in Large-Scale Ecological Networks from Heterogeneous Microbial Sequencing Data. *Cell Systems* 9, 286-296.e8. <https://doi.org/10.1016/j.cels.2019.08.002>.
95. Mallick, H., Rahnavard, A., McIver, L.J., Ma, S., Zhang, Y., Nguyen, L.H., Tickle, T.L., Weingart, G., Ren, B., Schwager, E.H., et al. (2021). Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol* 17, e1009442. <https://doi.org/10.1371/journal.pcbi.1009442>.
96. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv, <https://doi.org/10.48550/arXiv.1303.3997> <https://doi.org/10.48550/arXiv.1303.3997>.
97. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
98. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>.
99. Sahlin, K. (2022). Strobealign: flexible seed size enables ultra-fast and accurate read alignment. *Genome Biology* 23, 260. <https://doi.org/10.1186/s13059-022-02831-7>.
100. Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarszewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36, 996–1004. <https://doi.org/10.1038/nbt.4229>.
101. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* 38, 5825–5829. <https://doi.org/10.1093/molbev/msab293>.
102. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47, D309–D314. <https://doi.org/10.1093/nar/gky1085>.

# SUPPLEMENTARY MATERIAL

## Supplementary figures

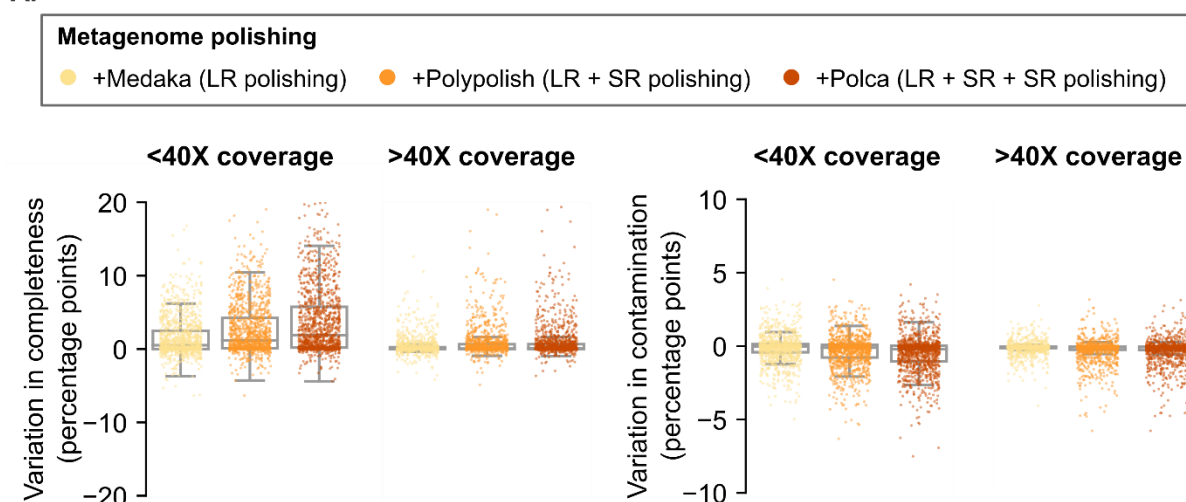


**Supplementary Figure S1. Read-level summary statistics for ONT raw reads.**

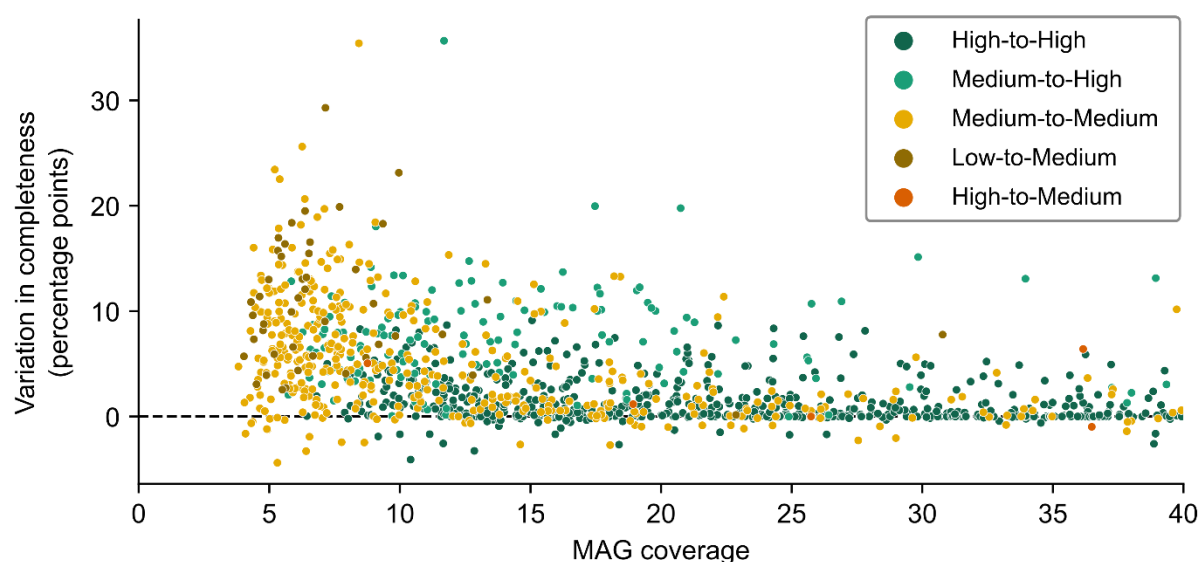


**Supplementary Figure S2. MAG generation pipeline detailed.** Depending on the initial throughput of ONT reads, we created one or two subsets of data (20 Gbp and 10 Gbp). We performed metagenome assemblies on all the datasets (2-3 metagenome assemblies per sample). Using 'ONT all' assemblies, we performed three different binning strategies: single-sample binning using long reads for abundance estimation, single-sample binning using short reads for abundance estimation, and multi-sample binning using short reads for abundance estimation. Using subset assemblies, we performed single-sample binning with short reads for abundance estimation. We used Dastool to keep the best MAGs for each species from each sample. After assigning taxonomy using GTDB-tk, if subset assemblies contained a medium- or high-quality MAG representing a species absent in the MAG catalog, we manually added it.

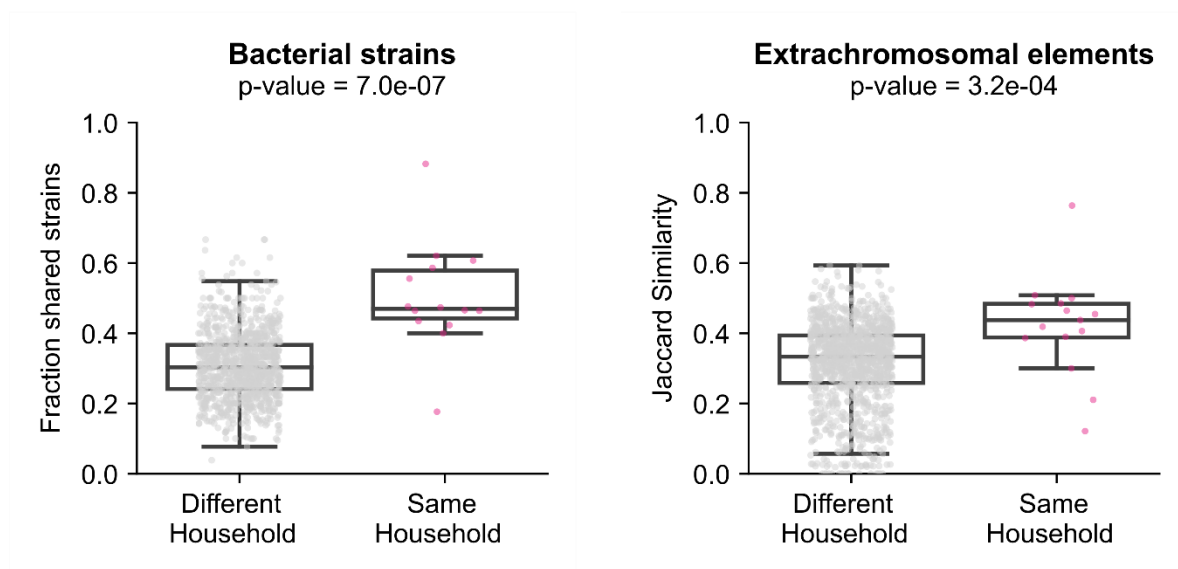
**A.**



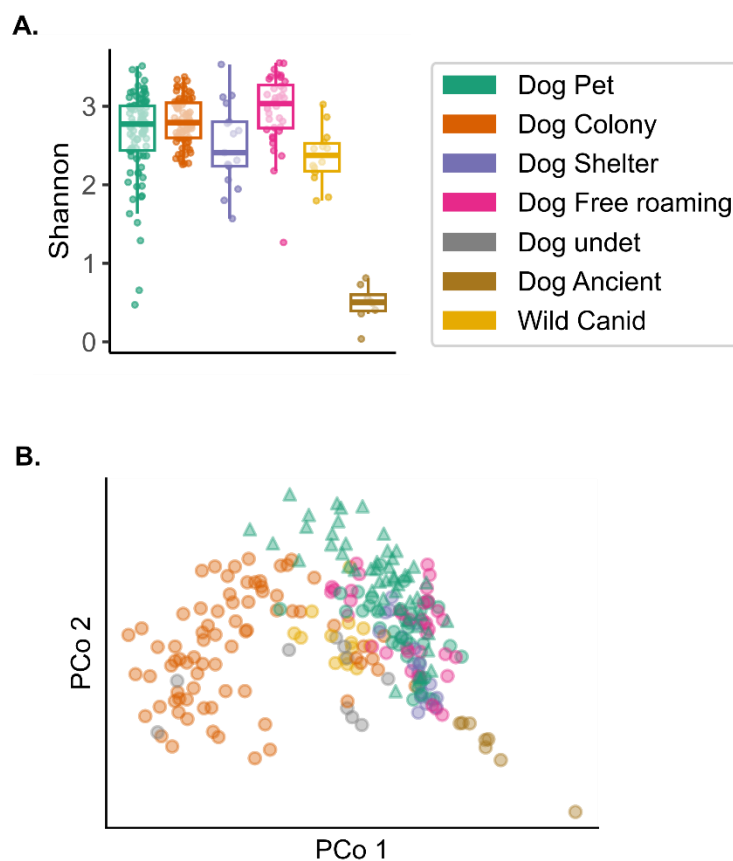
**B.**



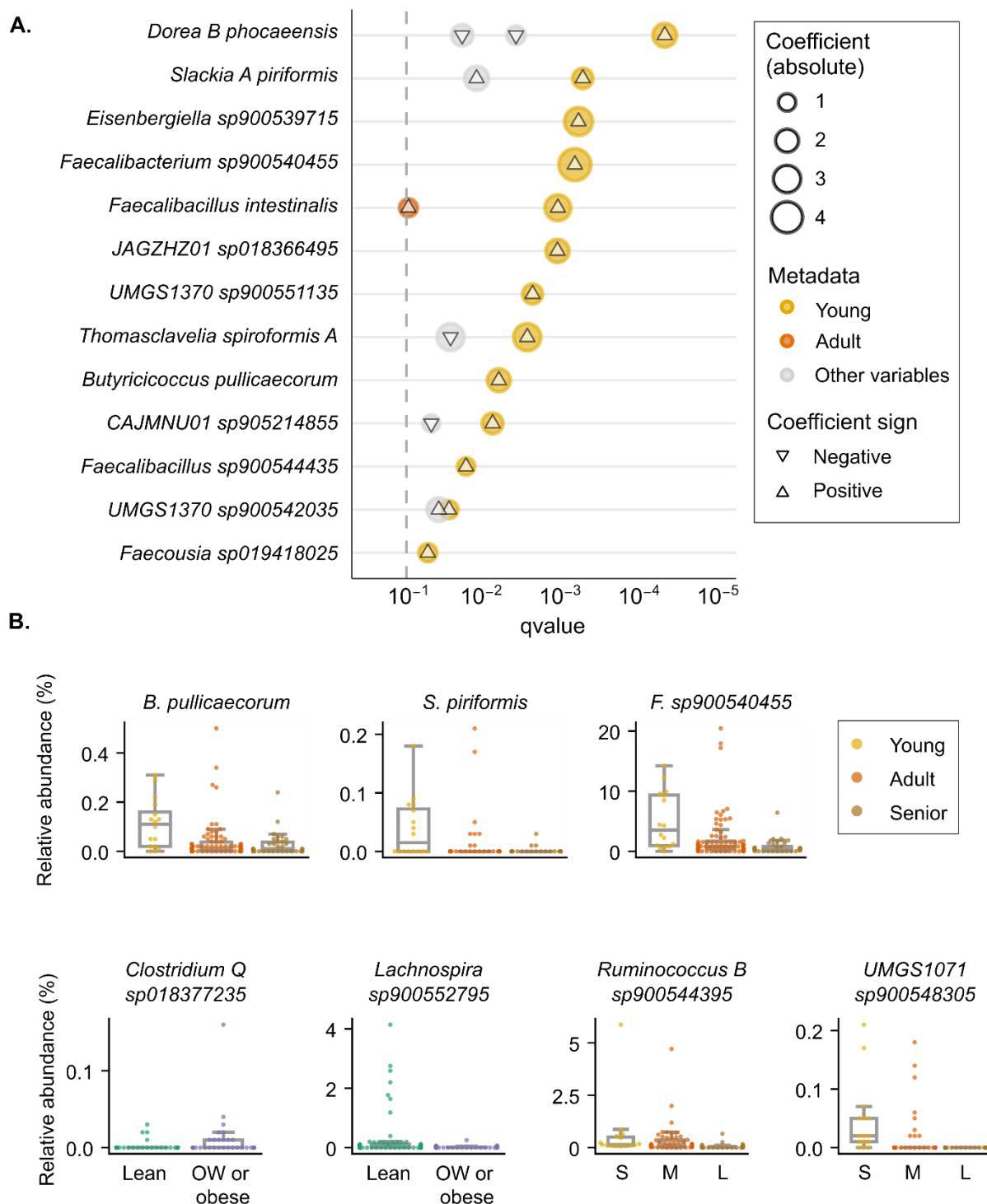
**Supplementary Figure S3. Polishing strategy evaluation in MAGs. A)** Each polishing step, and especially those using short reads (SR), increased the quality of the MAGs by increasing completeness and reducing contamination (estimated with CheckM2). The effect was more significant for MAGs with lower coverage (<40X). **B)** Unpolished to (most) polished MAG evolution for low-covered MAGs: we can observe a general increase in completeness values, which may lead to transforming medium-quality MAGs to high-quality MAGs.



**Supplementary Figure S4. Strain and extrachromosomal elements sharing within the Shanghai Dog cohort.** We performed pairwise comparisons (Mann–Whitney U test, one-sided test) between dogs within the same households vs. between households, and found significantly more shared strains (left) and EC elements (right) in dogs sharing households.



**Supplementary Figure S5. Living environment effect on the dog gut microbiome, including all Canid groups.** Analogous to Fig. 5A-B, here, we also included free-roaming dogs, ancient canid samples, and unclassified canid samples. The colored legend applies to the whole figure. **A)** Boxplots representing alpha diversity (Shannon index). **B)** PCoA plot representing beta diversity (Bray-Curtis on log-transformed data). Green triangles indicate pet dogs in this study.



**Supplementary Figure S6. Differential abundance analysis results by age, size, and body condition. A)** Differentially abundant bacterial species associated with dog age (concordant results, Maaslin2 and Kruskal-Wallis). In the y-axis, the significant taxa, and in the x-axis, their

associated q-value in Maaslin2. We used senior dogs as the reference group in Maaslin2, so circles are colored according to the group that differs from them, and they are scaled in concordance with the absolute coefficient value. Grey circles represent significant differences found in that species when comparing groups in another variable, rather than age (age being the most significant variable). Triangles within the circles indicate if the coefficient is negative (inverted triangle, less abundant species) or positive (more abundant). **B)** Boxplots of total relative abundances (%) of selected differentially abundant bacterial species, considering age (young, adult, and senior); body condition (lean, overweight/obese); and size (small, medium, large).

## Supplementary tables

**Supplementary Table S1. Dog-associated information for the Shanghai cohort (n=51).** A summarized version of the most relevant questionnaire results, with appropriate metadata categories considering the questionnaire output (with the latest information regarding animal health status).

**Supplementary Table S2. Shanghai Dog owner questionnaire results (n=107).** From questionnaire to final sample collection, several days to weeks might have passed, so questionnaire results regarding health status might slightly differ from the final metadata (Supplementary Table S1).

**Supplementary Table S3. Shanghai Dog Metagenome-assembled genomes catalog metadata.** Includes taxonomic classification, quality information, and other descriptive characteristics of the MAGs.

**Supplementary Table S4. Read mapping values of Canid datasets to Shanghai Dog catalogs.** Values include mapping to the whole MAG catalog ('aligned'), to the representative species MAGs ('aligned\_sp'), and to the catalog and extrachromosomal elements ('aligned\_EC'). It also includes read mapping summary statistics by study, and by living environment.

**Supplementary Table S5. Comparison of representative species-level genome assemblies: Shanghai Dogs vs. reference.** For each bacterial species, we compared the representative genome assembly from Shanghai Dog cohort to that found in a public database (RefSeq or GenBank). Only species that had a high-quality genome assembly (>90% completeness and <5% contamination) were included. Comparison included the counts of: contigs, ribosomal genes, and mobilome COG hits (COG category X), and statistical differences were assessed using Wilcoxon pairwise test.

**Supplementary Table S6. The 16S rRNA highest identity BLAST hit for species-level MAGs representing a novel species.** The table only includes high-quality MAGs that have at least one of their 16S rRNA hits with >99% identity to 16S rRNA NCBI databases.

**Supplementary Table S7. Shanghai Dog circular extrachromosomal elements catalog metadata (n=185).** The table includes basic descriptive information (size, sample origin, and category of the ECE).

Uncategorized elements might include genetic markers for both viruses and plasmids, or may lack identifiable markers. Finally, for 52 of them, we computationally predicted the putative bacterial host.

**Supplementary table S8. Metadata for the representative Canid samples: one sample per dog.** This metadata includes Shanghai Dogs, and external shotgun metagenomics studies from: public datasets (up to February 2023); a new pet dog cohort in Germany. For longitudinal studies, we chose a single representative sample per dog based on: the baseline diet (*e.g.*, in dietary intervention studies); or the 'healthiest' status (*e.g.*, absence of clinical signs after treatment, in cases of chronic enteropathies).

**Supplementary Table S9. Metadata for all the Canid datasets used.** For external public datasets, it includes dog shotgun metagenomics studies up to February 2023. Finally, it also includes metadata for a pet dog cohort in Germany.