Title
Genomes from bacteria associated with the canine oral cavity: A test case for automated genome-based taxonomic assignment.

Permalink
https://escholarship.org/uc/item/80w0x551

Journal
PloS one, 14(6)

ISSN
1932-6203

Authors
Coil, David A
Jospin, Guillaume
Darling, Aaron E
et al.

Publication Date
2019

DOI
10.1371/journal.pone.0214354

Peer reviewed
Genomes from bacteria associated with the canine oral cavity: A test case for automated genome-based taxonomic assignment

David A. Coil1, Guillaume Jospin1, Aaron E. Darling2, Corrin Wallis3, Ian J. Davis3, Stephen Harris3, Jonathan A. Eisen1,4, Lucy J. Holcombe3, Ciaran O’Flynn3*

1 Genome Center, University of California, Davis, CA, United States of America, 2 The Ithree Institute, University of Technology Sydney, Ultimo NSW, Australia, 3 The Waltham Centre for Pet Nutrition, Melton Mowbray, Leicestershire, United Kingdom, 4 Evolution and Ecology, Medical Microbiology and Immunology, University of California, Davis, Davis, CA, United States of America

* ciaran.oflynn@effem.com

Abstract

Taxonomy for bacterial isolates is commonly assigned via sequence analysis. However, the most common sequence-based approaches (e.g. 16S rRNA gene-based phylogeny or whole genome comparisons) are still labor intensive and subjective to varying degrees. Here we present a set of 33 bacterial genomes, isolated from the canine oral cavity. Taxonomy of these isolates was first assigned by PCR amplification of the 16S rRNA gene, Sanger sequencing, and taxonomy assignment using BLAST. After genome sequencing, taxonomy was revisited through a manual process using a combination of average nucleotide identity (ANI), concatenated marker gene phylogenies, and 16S rRNA gene phylogenies. This taxonomy was then compared to the automated taxonomic assignment given by the recently proposed Genome Taxonomy Database (GTDB). We found the results of all three methods to be similar (25 out of the 33 had matching genera), but the GTDB approach required fewer subjective decisions, and required far less labor. The primary differences in the non-identical taxonomic assignments involved cases where GTDB has proposed taxonomic revisions.

Introduction

With the ever-decreasing costs of DNA sequencing, it has become far easier and cheaper to sequence bacterial genomes than to analyze them. Understanding the gene content and metabolic pathways of a newly sequenced isolate is a time-consuming and knowledge-intensive process. Another, perhaps underappreciated, bottleneck is properly assigning taxonomy to a genome. This is most often seen with metagenome-assembled genomes (MAGs) which are unidentified prior to sequencing, but is even a problem with cultured isolates. Traditional morphological taxonomic assignment of bacterial isolates is tedious, and the more common approach of 16S rRNA gene PCR followed by Sanger sequencing is often uninformative beyond the genus level. Given the costs, some laboratories sequence isolate genomes directly,
with no attempt at prior identification. While some have argued against the need for taxonomic assignment for many aspects of microbial genome analysis (e.g., [1][2]), there are many situations where taxonomy is considered valuable for making use of genomic information (e.g., see [3] and [4]).

There have been a number of proposed attempts to move to a genome-based taxonomy for bacteria and archaea, instead of relying on traditional chemotaxonomic/morphological characteristics of isolates [5][6][7]. These include the use of average nucleotide identity (ANI) [8][9][10][11], concatenated marker gene phylogenies (e.g. SILVA (unpublished) and GTDB [12], and shared protein content [13]. Most of these approaches however rely on a provisional identification (at least to genus), followed by locating and downloading the genomes of close relatives for comparison.

In this work, we briefly describe the genome sequences of 33 bacterial isolates from the canine oral cavity. These isolates were collected as part of a larger project on canine oral health and had a preliminary taxonomy assigned through Sanger sequencing and a sequence identity comparison of the 16S rRNA gene [14,15]. After genome sequencing, we first assigned taxonomy to these isolates based on a manual examination of “whole genome” concatenated marker phylogenetic trees, average nucleotide identity (ANI), and 16S rRNA gene phylogenetic trees. We then compared this taxonomy to automated taxonomic assignments given by the recently proposed Genome Taxonomy Database (GTDB).

Genomes and taxonomy

Genome selection

The study design, bacterial isolation, DNA extraction, isolation identification and genome sequencing/assembly have been previously described in our work on the Porphyromonas genus [16]. Briefly, bacterial isolates from the canine oral cavity were grown on supplemented Columbia Blood Agar containing 5% defibrinated horse blood (CBA; Oxoid, UK) with or without the addition of 5 mg/L hemin (catalog no. H9039; Sigma) and 0.5 mg/L menadione (catalog no. M5625; Sigma) or Heart Infusion Agar containing 5% defibrinated horse blood (HIA;Oxoid,UK) (Table 1). Aerobes were incubated at 38˚C under normal atmospheric conditions for 1–5 days. Microaerophilic and anaerobic strains were incubated at 38˚C for 1–21 days in a MACS1000 anaerobic workstation (Don Whitley, UK) with gas levels at 5% oxygen, 10% carbon dioxide, and 85% nitrogen for microaerophiles, and 10% hydrogen, 10% carbon dioxide, and 80% nitrogen for anaerobes. Following DNA extraction, library preparation, and Illumina sequencing, the reads were assembled using the A5-miseq assembly pipeline [17].

The remaining non-Porphyromonas genomes were further screened by a combination of assembly metrics and CheckM [18] to estimate completeness and contamination. We chose for further study those that appeared to be the highest quality using admittedly somewhat arbitrary cutoffs of (a) fewer than 350 contigs in the assembly, (b) CheckM contamination score of <3%, and (c) CheckM completeness score of >90%. A subset of 33 genomes meeting these criteria was chosen to study in more detail.

Preliminary taxonomic identification

All isolates were given a preliminary identification based on analysis of 16S rRNA gene sequences generated via Sanger sequencing (F24/Y36(9-29F/1525-1541R) primers). The 16S rRNA gene sequences were searched using BLAST [19] against an in-house canine oral microbiome database. This database contained 460 published 16S rRNA gene sequences obtained from canine oral taxa (Genbank accession numbers JN713151-JN713566 & KF030193-KF030235 [15]. In addition, sequences were searched against the RDP 16S rRNA database
Table 1. Comparative taxonomy of 33 bacterial strains by three different methods.

| Isolate Number | Media          | Aerobe/anaerobe | 16S taxonomy (Sanger, BLAST) | Manual Taxonomy (S, A, or W) | GTDB Taxonomy | Comments | Accession |
|----------------|----------------|-----------------|------------------------------|------------------------------|---------------|----------|-----------|
| ATCC 29435     | CBA H&M        | Aerobe          | Conchiformibius steedae      | Conchiformibius steedae      | Conchiformibius steedae | All species | RQYCO0000000 |
| OH1139         | CBA Microaerophile | Staphylococcus epidermidis | Staphylococcus epidermidis | Staphylococcus epidermidis | All species | RQYG0000000 |
| OH1877         | CBA Aerobe     | Actinomyces hordeovulnaris | Actinomyces hordeovulnaris | Actinomyces hordeovulnaris | All species | RQYO0000000 |
| OH2158         | CBA H&M       | Anaerobe        | Campylobacter rectus         | Campylobacter rectus         | Campylobacter_A rectus | All species | RQYP0000000 |
| OH1206         | HIA Anaerobe   | Bacillus licheniformis (100%) | Bacillus licheniformis     | Bacillus licheniformis      | All species | RQYT0000000 |
| OH3297         | CBA Aerobe     | Actinomyces hordeovulnaris | Actinomyces hordeovulnaris | Actinomyces hordeovulnaris | All species | RQYV0000000 |
| OH4621         | CBA Aerobe     | Streptococcus minor | Streptococcus minor          | Streptococcus minor         | All species | RQZA0000000 |
| OH953          | CBA Aerobe     | Streptococcus sanguinis | Streptococcus sanguinis     | Streptococcus sanguinis     | All species | RQZI0000000 |
| OH1186         | CBA H&M       | Anaerobe        | Desulfovibrio sp. COT-070    | Desulfovibrio             | Desulfovibrio | All genus | RQYH0000000 |
| OH1205         | HIA Anaerobe   | Alloprevotella sp. COT-284 | Alloprevotella             | F0040 [Alloprevotella]   | All genus | RQYT0000000 |
| OH2545         | CBA Aerobe     | Ottowia sp. COT-014 | Ottowia          | Ottowia                 | All genus | RQYR0000000 |
| OH741          | CBA H&M       | Microaerophile  | Erysipelotrichaceae [G-1]    | Erysipelotrichaceae       | Erysipelotrichaceae | All family | RQZE0000000 |
| OH1209         | CBA Aerobe     | Streptococcus sp. COT-279 | Desulfovibrio             | Desulfovibrio            | Mis-identification | RQYK0000000 |
| OH4692         | CBA Aerobe     | Streptococcus pneumoniae COT-348 | Streptococcus | Streptococcus           | 16S species, manual & GTDB genus | RQZB0000000 |
| OH2310         | CBA Aerobe     | Xenophillus sp. COT-174 | Lampropedia             | Lampropedia              | Related genus | RQYQ0000000 |
| OH3737         | CBA Aerobe     | Xenophillus sp. COT-264 | Lampropedia             | Lampropedia              | Related genus | RQX0000000 |
| OH1047         | CBA H&M       | Anaerobe        | Bacteroides heparinolyticus COT-310 | Bacteroides heparinolyticus | Bacteroides | Official nomenclature change | RQYF0000000 |
| OH1426         | CBA H&M       | Anaerobe        | Tannerella forsythia COT-023 | Tannerella forsythia     | Tannerella | GTDB genus, manual & 16S species | RQYN0000000 |
| OH2617         | CBA H&M       | Anaerobe        | Tannerella forsythia COT-023 | Tannerella forsythia     | Tannerella | GTDB genus, manual & 16S species | RQYS0000000 |
| OH4460         | CBA H&M       | Anaerobe        | Fusobacterium canalis COT-188 | Fusobacterium canalis     | Fusobacterium | GTDB genus, manual & 16S species | RQY0000000 |
| OH5050         | CBA Aerobe     | Actinomyces bowdennii COT-413 | Actinomyces bowdenni     | Actinomyces              | Actinomyces | GTDB genus, manual & 16S species | RQZC0000000 |

(Continued)
Taxonomic identifications were made using percent identity cutoffs of 98.5% (for species), 94% (for genus), and 92% (for family).

Traditional taxonomic identification

The contigs comprising the genomes were first uploaded to the RAST server [21] for annotation. The results archives were downloaded using the RAST API and then searched for full-length 16S rRNA gene sequences by scanning the annotations using the “SSU rRNA” tag and filtering for length greater than 1 kb. These sequences were uploaded to the Ribosomal Database Project (RDP) [20] and incorporated into their alignment. For first-pass taxonomic identification, an alignment and phylogenetic tree was generated using SSU-ALIGN [22] and FastTree 2 [23] of all ~12,000 type strain 16S rRNA gene sequences from RDP, along with our 33 sequences from RAST. This provided us with the general region of the tree for each isolate or group of isolates.

Next, we generated a series of concatenated “whole genome sequence” (WGS) marker trees for each isolate or group of isolates as described below. All available genome sequences from the genus or family of interest were downloaded from GenBank, except in cases where more than 500 genomes were available. In those genome-rich genera (Bacillus, Clostridium, Streptococcus, Staphylococcus, Campylobacter), the WGS trees were built with only a subset of genomes, selected from the type strain results. For example, Staphylococcus contains ~40 species, but only genomes from those species closest (by eye in the tree) to our isolate of interest were needed to build a useful WGS tree.

For all WGS trees, an outgroup genome(s) was chosen from nearby in the NCBI taxonomy (e.g. another genus in the same family). The file names and sequences were reformatted for easier visualization. The assemblies were then screened for 37 core maker genes [24] using PhyloSift [25] in the search and align mode using “isolate” and “besthit” flags. PhyloSift concatenates and aligns the hits of interest then the sequences are subsequently extracted from the PhyloSift output files and added to a single file for tree-building. An approximately maximum-likelihood tree was then inferred using FastTree2 with default parameters [23].

Table 1. (Continued)

| Isolate Number | Media  | Aerobe/anaerobe | 16S taxonomy (Sanger, BLAST) | Manual Taxonomy (S, A, or W) | GTDB Taxonomy | Comments | Accession |
|----------------|--------|-----------------|-----------------------------|-----------------------------|---------------|----------|----------|
| OH770          | CBA    | Aerobe          | Actinomyces canis COT-409   | Actinomyces canis           | Actinomyces_B | GTDB genus, manual & 16S species | RQZFO0000000 |
| OH5060         | CBA H&M| Anaerobe        | Fusobacterium sp. COT-236   | Fusobacterium              | Fusobacterium | GTDB & 16S genus, manual species | RZXDO0000000 |
| OH887          | CBA H&M| Anaerobe        | Propionibacterium sp. COT-365| Propionibacterium propionicum (S) | Pseudopropionibacterium | GTDB proposed taxonomic change | RQZGO0000000 |
| 6824 SB        | HIA    | Anaerobe        | Lachnospiraceae XIVa [G-6] sp. COT-073 | Clostridiales (S) | Lachnospirales | GTDB proposed taxonomic change | RQY DO0000000 |
| OH1046         | CBA H&M| Anaerobe        | Atopobium parvulum          | Coriobacterineae (W)       | Atopobacteriae | GTDB proposed taxonomic change | RQYE0000000 |
| OH2822         | CBA H&M| Anaerobe        | Propionibacterium sp. COT-296| Propionibacterium propionicum (S) | Pseudopropionibacterium | GTDB proposed taxonomic change | RQYT0000000 |

CBA: Columbia Blood Agar, HIA: Heart Infusion Agar, H&M: Hemin and Menadione.

"F0400" is actually a strain name, used as a placeholder by GTDB when the authors believe that genome belongs in a new genus, for which no type representative is present. (S) is for strains in the manual approach whose taxonomy was assigned using 16S rRNA gene trees, (A) for those for which ANI was used, and (W) for those for which a WGS tree was used. The comments column describes the way in which each of the three methods differs or is the same.

https://doi.org/10.1371/journal.pone.0214354.1001

v10.31 [20]. Taxonomic identifications were made using percent identity cutoffs of 98.5% (for species), 94% (for genus), and 92% (for family).
For all isolates where the WGS tree indicated a possible placement into a well-defined clade with sequenced genomes, we downloaded the type strain genome sequences for every member of that genus/family from Refseq at NCBI. These were used to create an ANI (average nucleotide identity) matrix using FastANI [26] for each group and anything having an ANI > 95% to a type strain was considered to belong to that species [8][27].

For the majority of taxa, the WGS tree and ANI matrix were still inadequate to assign species level taxonomy, due to a paucity of genome sequences for many groups. In those cases, we relied solely on the 16S rRNA gene for taxonomy, but using phylogeny instead of just sequence identity as in the first pass. This was again accomplished through RDP, by downloading all sequences for a given genus/family, inferring a phylogenetic tree, and looking for well-supported placement into a monophyletic clade. For all groups with fewer than 3000 sequences, the alignment was downloaded directly from RDP, for larger groups the sequences were downloaded and the alignment generated with SSU-ALIGN. All alignments were cleaned to remove problematic characters in the headers using a custom script [28] and all trees were inferred using FastTree2 with default parameters.

Final taxonomic assignments were based on a taxa-dependent combination of the WGS trees, the ANI results, and the 16S rRNA gene trees (Table 1). First priority was given to the ANI results, then to placement within a well-supported WGS tree with monophyletic clades, and then finally to 16S rRNA gene-based results. As a result of inadequate mapping between phylogeny and taxonomy, some isolates were only assigned to the genus or family level, and in one case the order level. Examples of both informative and non-informative WGS/16S trees, along with a sample ANI matrix, can be found in the Supplemental Materials. Of the remaining WGS/16S trees and ANI results, those that were informative in determining taxonomy can be found at (doi: 10.25338/B8801B).

**Genome taxonomy database**

The Genome Taxonomy Database (GTDB) is a recently proposed system attempting to create a genomic-based, standardized bacterial taxonomy [12]. The system uses 120 single-copy conserved marker genes to generate a phylogenetic tree of 94,759 genomes, and analysis of the topology of this tree was then used to propose large-scale revisions to bacterial taxonomy. During our work on this project, a tool based on GTDB was release; “GTDB-Tk” [29] attempts to automate taxonomic assignment for genomes, based on their revised GTDB taxonomy. The tool uses a combination of phylogenetic tree topology, ANI values, and relative evolutionary divergence (RED). We screened our 33 isolates using this tool and compared the taxonomic assignments to the manual process above (Table 1). Note that an alphabetic suffix in the GTDB taxonomy (e.g. “Actinomyces_B”) indicates that, within the GTDB tree, the genus does not belong to a monophyletic group with the type species of that genus.

**Discussion**

Here we present the genomes of 33 bacterial isolates from the canine oral cavity. Some are from groups for which some of the members of these groups are known to be involved in human and canine oral health (e.g. *Actinomyces* and *Fusobacterium*) and others have not been previously suggested to play such a role. A few of our isolates appear to be phylogenetically novel and potentially of interest for further study. For example Canine Oral Taxon number 073 (COT073) and bacterial isolate number OH741 were only identified to the order and family level respectively.

The difference in labor required by the two genomic methods of taxonomic assignment was noticeable, with the manual method having taken two weeks of daily downloads,
alignments, ANI comparisons and tree-building whereas the GTDB automated method ran overnight. The latter also is much less subjective (for the user) than the former. A comparison of the three approaches to taxonomic assignments shows a very high degree of similarity (Table 1).

For 51% (17/33) of isolates the three methods gave identical results; for 18.2% (6/33) of isolates the 16S and manual methods gave the same results; and in one case the 16S and GTDB methods gave the same results. Finally, for 12% (4/33) of isolates all three methods gave different results to each other i.e. different genera within same family or different levels of assignment (granularity/resolution) within the same branch of the tree. Most of the differences between the GTDB taxonomy and the other approaches appears to be due to proposed taxonomic revisions by the GTDB group. For example COT073 was placed within the Clostridiales order in the manual taxonomy, but that order was subdivided into several new groups in the GTDB taxonomy, based on calculations of relative evolutionary divergence (RED) within that group.

Our results suggest that the use of the automated GTDB tool to assign taxonomy to unidentified bacterial isolates is less subjective and much faster than manual assignment, while giving very similar results (e.g. identical taxonomy, difference only in taxonomic rank, or proposed taxonomic changes). Both genomic methods offer improved taxonomic resolution relative to analysis of just 16S rRNA gene sequences, at the additional cost/burden of requiring the entire genome. Finally, we expect these genomes will be of use to researchers studying canine oral health since the vast majority of closely related isolates so far sequenced have been collected from humans.

Accession numbers
The whole-genome shotgun projects have been deposited at DDBJ/ENA/GenBank under the accession numbers RQYC00000000-RQZI00000000. The versions described in this paper are versions RQYC01000000-RQZI01000000. These strains were submitted to NCBI using our “manual taxonomy” assignments, since the GTDB taxonomy is not yet widely accepted. However, for some of the isolates, NCBI made their own minor taxonomic revisions.

Supporting information
S1 Fig. A taxonomically inconclusive WGS tree for isolate OH7770. The isolate is not found within a clade of other sequenced isolates with species-level taxonomy. (PDF)

S2 Fig. A taxonomically inconclusive 16S rRNA gene tree for isolate OH1287. Taxonomy is not congruent with phylogeny and many neighboring sequences are only identified to the genus level. (PDF)

S3 Fig. A taxonomically informative 16S rRNA gene tree for isolate OH5050. The isolate is found in a monophyletic clade and the name given to the closest relatives is not found elsewhere in the tree. (PDF)

S1 Table. A sample ANI result for isolate OH3297. The isolate is ~99% identical to both OH1877 and to Actinomyces hordeovulneris. The ANI drops off quite rapidly to other members of the genus. (DOCX)
Acknowledgments

The authors would like to acknowledge the laboratory staff at WALTHAM Centre for Pet Nutrition for growing the bacterial isolates, extracting genomic DNA, and for initial species identification, and Zoe Lonsdale for figure creation and bioinformatics support. The authors would also like to acknowledge John Zhang from UC Davis for the sequencing library preparation of isolates. The sequencing was carried by the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01.

Author Contributions

Conceptualization: Aaron E. Darling, Corrin Wallis, Ian J. Davis, Stephen Harris, Jonathan A. Eisen, Lucy J. Holcombe, Ciaran O’Flynn.

Data curation: Guillaume Jospin, Aaron E. Darling, Corrin Wallis, Ian J. Davis, Lucy J. Holcombe, Ciaran O’Flynn.

Formal analysis: David A. Coil, Guillaume Jospin, Jonathan A. Eisen, Lucy J. Holcombe, Ciaran O’Flynn.

Funding acquisition: Aaron E. Darling, Corrin Wallis, Jonathan A. Eisen, Ciaran O’Flynn.

Investigation: David A. Coil, Corrin Wallis, Ian J. Davis, Stephen Harris, Jonathan A. Eisen, Lucy J. Holcombe, Ciaran O’Flynn.

Methodology: David A. Coil, Guillaume Jospin, Ciaran O’Flynn.

Project administration: David A. Coil, Jonathan A. Eisen, Lucy J. Holcombe, Ciaran O’Flynn.

Software: Guillaume Jospin, Aaron E. Darling.

Supervision: David A. Coil, Corrin Wallis, Ian J. Davis, Stephen Harris, Jonathan A. Eisen, Lucy J. Holcombe, Ciaran O’Flynn.

Validation: David A. Coil, Aaron E. Darling.

Visualization: David A. Coil.

Writing – original draft: David A. Coil, Jonathan A. Eisen, Lucy J. Holcombe, Ciaran O’Flynn.

Writing – review & editing: David A. Coil, Guillaume Jospin, Aaron E. Darling, Corrin Wallis, Ian J. Davis, Stephen Harris, Jonathan A. Eisen, Lucy J. Holcombe, Ciaran O’Flynn.

References

1. Sedlar K, Kupkova K, Provaznik I. Bioinformatics strategies for taxonomy independent binning and visualization of sequences in shotgun metagenomics. Comput Struct Biotechnol J. 2017; 15: 48–55. https://doi.org/10.1016/j.csbj.2016.11.005 PMID: 27980708

2. Sun Y, Cai Y, Huse SM, Knight R, Farmerie WG, Wang X, et al. A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. Brief Bioinform. 2012; 13: 107–121. https://doi.org/10.1093/bib/bbr009 PMID: 21525143

3. Chun J, Rainey FA. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int J Syst Evol Microbiol. 2014; 64: 316–324. https://doi.org/10.1099/ijsem.0.054171-0 PMID: 24505069

4. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol. 2018; 68: 461–466. https://doi.org/10.1099/ijsem.0.002516 PMID: 29292687
5. Sentausa E, Fournier P-E. Advantages and limitations of genomics in prokaryotic taxonomy. Clin Microbiol Infect. 2013; 19: 790–795. https://doi.org/10.1111/1469-0691.12181 PMID: 23490121

6. Garrity GM. A New Genomics-Driven Taxonomy of Bacteria and Archaea: Are We There Yet? J Clin Microbiol. 2016; 54: 1956–1963. https://doi.org/10.1128/JCM.02000-16 PMID: 27194877

7. Klenk H-P, Göker M. En route to a genome-based classification of Archaea and Bacteria? Syst Appl Microbiol. 2010; 33: 175–182. https://doi.org/10.1016/j.syapm.2010.03.003 PMID: 20409658

8. Chan JZ-M, Halachev MR, Loman NJ, Constantinidou C, Pallen MJ. Defining bacterial species in the genomic era: insights from the genus Acinetobacter. BMC Microbiol. 2012; 12: 302. https://doi.org/10.1186/1471-2180-12-302 PMID: 23259572

9. Zhang W, Du P, Zheng H, Yu W, Wang L, Chen C. Whole-genome sequence comparison as a method for improving bacterial species definition. J Gen Appl Microbiol. 2014; 60: 75–78. PMID: 24859865

10. Kim M, Oh H-S, Park S-C, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol. 2014; 64: 346–351. https://doi.org/10.1099/ijsem.0.059774-0 PMID: 24505072

11. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyriides NC, et al. Microbial species delineation using whole genome sequences. Nucleic Acids Res. 2015; 43: 6761–6771. https://doi.org/10.1093/nar/gkv657 PMID: 26150420

12. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018; 36: 996–1004. https://doi.org/10.1038/nbt.4229 PMID: 30148503

13. Qin Q-L, Xie B-B, Zhang X-Y, Zhang X-L, Zhou B-C, Zhou J, et al. A proposed genus boundary for the bacterial taxon “Lambdena”. PLoS One. 2013; 8: e59616. https://doi.org/10.1371/journal.pone.0059616 PMID: 23225412

14. Dewhirst FE, Klein EA, Thompson EC, Blanton JM, Chen T, Milella L, et al. The canine oral microbiome. PLoS One. 2012; 7: e36067. https://doi.org/10.1371/journal.pone.0036067 PMID: 22558330

15. Davis IU, Wallis C, Deusch O, Colyer A, Milella L, Loman N, et al. A cross-sectional survey of bacterial species in plaque from client owned dogs with healthy gingiva, gingivitis or mild periodontitis. PLoS One. 2013; 8: e83158. https://doi.org/10.1371/journal.pone.0083158 PMID: 24349448

16. Coil DA, Alexiev A, Wallis C, O’Flynn C, Deusch O, Davis I, et al. Draft genome sequences of 26 porphyromonas strains isolated from the canine oral microbiome. Genome Announc. 2015; 3. https://doi.org/10.1101/gna.00187-15 PMID: 25858832

17. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics. 2015; 31: 587–589. https://doi.org/10.1093/bioinformatics/btu661 PMID: 25338718

18. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015; 25: 1043–1055. https://doi.org/10.1101/gr.186072.114 PMID: 25977477

19. Allschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 22317122

20. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014; 42: D633–42. https://doi.org/10.1093/nar/gkt1244 PMID: 24288368

21. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008; 9: 75. https://doi.org/10.1186/1471-2164-9-75 PMID: 18261238

22. Nawrocki EP. Structural RNA homology search and alignment using covariance models [Internet]. Eddy SR, editor. Washington University in St. Louis. 2009. Available: https://search.proquest.com/docview/305017419

23. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One. 2010; 5: e9490. https://doi.org/10.1371/journal.pone.0009490 PMID: 20224823

24. Wu D, Jospin G, Eisen JA. Systematic Identification of Gene Families for Use as “Markers” for Phylogenetic and Phylogeny-Driven Ecological Studies of Bacteria and Archaea and Their Major Subgroups. PLoS One. Public Library of Science; 2013; 8: e70733. https://doi.org/10.1371/journal.pone.0070733 PMID: 24146854

25. Darling AE, Jospin G, Lowe E, Matsen FA 4th, Bik HM, Eisen JA. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ. 2014; 2: e243. https://doi.org/10.7717/peerj.243 PMID: 24482762

26. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018; 9: 5114. https://doi.org/10.1038/s41467-018-07641-9 PMID: 30504858
27. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 2009; 106: 19126–19131. https://doi.org/10.1073/pnas.0906412106 PMID: 19855009

28. Dunitz MI, Lang JM, Jospin G, Darling AE, Eisen JA, Coil DA. Swabs to genomes: a comprehensive workflow. PeerJ. 2015; 3: e960. https://doi.org/10.7717/peerj.960 PMID: 26020012

29. Ecogenomics. Ecogenomics/GtdbTk. In: GitHub [Internet]. [cited 24 May 2018]. Available: https://github.com/Ecogenomics/GtdbTk