bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Phylogenomic analyses of understudied *Neisseriaceae* species support the reclassification of the polyphyletic genera *Kingella, Simonsiella, and Alysiella*.

Daniel P. Morreale[1,2], Joseph W. St Geme III[1,2†], and Paul J Planet [1,2,3*†]

Affiliations:

[1] Perelman School of Medicine, University of Pennsylvania, Philadelphia PA.

[2]Division of Infectious Diseases, Children's Hospital of Pennsylvania, Philadelphia PA.

[3] Comparative Genomics, American Museum of Natural History, New York, New York, United States of America

*Corresponding Author: Paul J. Planet (pjplanet@chop.edu)

† Contributed equally

Keywords: ***Kingella, Simonsiella, Alysiella, Neisseriaceae,*** **Polyphyly,**

Abbreviations: ANI: Average Nucleotide identity; dDDH: digital DNA-DNA hybridization, OGRI: overall genome relatedness index, COGs: clusters of orthologous gene groups; MuLDi: Multicellular, longitudinally-dividing; DUS: DNA-uptake sequences, MALDI-TOF MS: Matrix-assisted laser desorption-ionization time of flight mass spectrometry

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract.

Taxonomic classification of the *Neisseriaceae* family has focused primarily on the pathogens *Neisseria meningitidis* and *Neisseria gonorrhoeae*. As a consequence, the commensal *Neisseria* species and other *Neisseriaceae* genera such as *Kingella* have been only loosely classified, resulting in several polyphyletic genera. In this study, using available 16S rRNA sequences and a phylogenetic approach, we found that the genus *Kingella* is polyphyletic, due to misclassification of *Kingella potus*. Calculation of multiple genome relatedness indices revealed that *K. potus* is more similar to *Neisseria bacilliformis* than it is to other *Kingella* species. To better understand the evolution of this clade, we examined the core genome of *Kingella* and closely related genera and discovered that the related genera *Simonsiella* and *Alysiella* form a distinct clade within the genus *Kingella* that is closely related to the pathogens *K. kingae* and *K. negevensis*. Based on these analyses, we propose incorporation of *Simonsiella* and *Alysiella* species into the genus *Kingella*.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Main text.

The *Neisseriaceae* family is a group of Gram-negative, β-proteobacteria that currently includes 20 genera: *Neisseria*, *Wielerella*, *Vitreoscilla*, *Uruburuella*, *Stenoxybacter*, *Snodgrassela*, *Simonsiella*, *Rivicola*, *Prolinoborus*, *Paralysiella*, *Morococcus*, *Kingella*, *Eikenella*, *Crenobacter*, *Craterilacuibacter*, *Conchiformibius*, *Bergeriella*, *Aquella*, *Amantichitinum*, and *Alysiella* [1–3]. The majority of prior work devoted to understanding the taxonomic relationships between these genera has focused primarily on the pathogenic *Neisseria* species, *N. meningitidis* and *N. gonorrhoeae*. To date, the commensal *Neisseria* species and related genera have been only loosely classified based on preliminary studies. The genus *Kingella* includes five taxa: *K. potus, K. oralis, K. denitrificans, K. negevensis,* and *K. kingae*. Recently, a new species was proposed, *K. bonacorsii* [4–10]. Each of the *Kingella* species is typically associated with the oral or oropharyngeal microbiome, a common niche of other *Neisseriaceae*. Isolates of *K. kingae*, *K. negevensis*, and *K. potus* have all been associated with bacteremia, osteoarticular infections, and/or endocarditis in immunocompetent individuals [6, 8, 11]. In contrast, *K. oralis* and *K. denitrificans* are present in dental plaque, gingival and periodontal samples, and only rarely cause invasive infections, typically in immunocompromised individuals [9, 12, 13].

A review of published phylogenetic relationships among the commensal *Neisseria* species reveals a high level of variability in the proposed taxonomic relationships, primarily owing to the genes selected and method used to perform the analysis [14–17]. The genus *Kingella* is polyphyletic, and the genera that interrupt a monophyletic *Kingella* clade differ between published analyses. Most isolates used for defining taxonomic nomenclature in this genus were collected prior to pervasive whole genome sequencing and were assigned to *Kingella* as a result

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of phenotypic and/or minimal sequence-based analyses. Additionally, as the frequency of culture-dependent and independent microbiome studies have become more prevalent, the number and diversity of formally accepted species in the *Neisseriaceae* has increased [18]. With this background, we sought to reassess the relatedness of the five species in *Kingella*.

To examine the *Kingella* species, we employed a phylogenetic approach. 840 of the available 16S rRNA gene sequences for isolates from the *Neisseriaceae* and in the RefNR database were downloaded from the SILVA r138.1 database on September 08, 2022 [19, 20]. 16S rRNA sequences were aligned with MAFFT, and a phylogeny was reconstructed using RAxML v. 8.4.2 with the general time reversible model of nucleotide substitution and gamma model of rate heterogeneity, and bootstrapped 100 times [21, 22]. The resulting tree was annotated using Figtree v.1.4.4 and ITOL and is shown in Fig. 1 [23]. By 16S rRNA relatedness alone, it is clear that a major driver of the polyphyletic structure of *Kingella* is *K. potus.* This species was recovered from an infected wound caused by an animal bite and is more closely related to *Neisseria bacilliformis* than it is to any currently described *Kingella* species [8, 24]. *N. bacilliformis* was identified shortly after *K. potus* was defined, is a part of the oral microbiome of non-human mammals, and is associated with bacteremia in humans after animal bites [24].

To better understand the relationship between the *K. potus* and *N. bacilliformis*, we calculated average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) between the type strains of these species using FastANI v.1.33 and the Type Strain Genome Server (TYGS), respectively [25–27]. The ANI between *K. potus* and *N. bacilliformis* was calculated at 86.75%, below the established species cut-off of 95% [28]. However, the ANI between *K. potus and N. bacilliformis* is appreciably greater than the ANI calculated between *K. potus* and other *Kingella*

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

species (Table 1). Similarly, while the calculated dDDH is below species thresholds (Table 2) [29], it is clear that *K. potus* is far more similar to *N. bacilliformis* than it is to other species within the genus *Kingella.*

To further define the relationships between *Neisseriaceae*, whole genome sequences for select species in this family were downloaded from NCBI (on September 09, 2022). These analyses were limited to only those species most closely related to *Kingella, by* 16S rRNA similarity and include *Neisseria elongata, Neisseria sheyganii, N. bacilliformis, Eikenella corrodens, Alysiella filiformis, Alysiella crassa, Simonsiella muelleri, Conchiformibius steedae, K. oralis, K. bonacorsii, K. potus, K. denitrificans, K. kingae,* and *K. negevensis* [7, 30]. In total, 134 genomes were reassembled with Prokka v1.14.6, using default settings for bacterial sequences [31]. Subsequent assemblies were analyzed with Roary v3.13.0. to identify the core genome [32]. Core genes were clustered at the 75%, 80%, 85%, 90%, and 95% sequence similarity level and were limited to only those clusters of orthologous groups (COGs) found in all genomes. Nucleotide sequences of core genes were compared in a codon-aware alignment by PRANK within Roary, and core gene alignments were used to reconstruct phylogenies using RAxML as above. Each phylogeny was overwhelmingly congruent with the others, and we therefore proceeded with the lowest similarity cutoff for detailed analysis (Fig. 2A). These phylogenies further support the close relationship between *K. potus* and *N. bacilliformis*.

The phylogenomic analyses also supported a close evolutionary relationship between *Kingella*, *Alysiella*, and *Simonsiella*. *Alysiella* and *Simonsiella* are commensal genera associated with the oral microbiome of mammals [14, 33, 34]. These two genera, along with the genus *Conchiformibius,* are distinctive for being multicellular, longitudinally-dividing (MuLDi) species

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that comprise the only known animal multicellular bacterial symbionts [15]. Based on a whole genome approach, *Alysiella* and *Simonsiella* diverged from the rest of the *Kingella* species more recently than *K. denitrificans* or *K. oralis* diverged. ANI, dDDH, and G+C% were calculated for type strains of *Kingella*, *Alysiella*, and *Simonsiella* and are shown in Tables 1 and 3. While there is no universally accepted ANI cutoff for demarcating a bacterial genus, ANI scores greater than 73% have been proposed [35]. All of these overall genome relatedness indices (OGRI) support the close relationship between these three genera, and the ANI falls above the proposed 73% genus cutoff score. Moreover, Nyongesa and co-workers [15] hypothesize that the MuLDi phenotype evolved twice in the *Neisseriaceae,* first in *Conchiformibius* species and then in the common ancestor of *Simonsiella and* Alysiella, a hypothesis that is strongly supported by our phylogenetic analyses.

Additionally, we calculated pairwise ANI scores for each of the genomes listed above. The ANI scores confirm the majority of the proposed species classifications with one exception. Recently, an unclassified *Kingella* isolate was proposed to be a new species and was named *K. bonacorsii* [4]. This isolate was recovered from a gingival sample, and the authors report a 16S rRNA similarity of 98.7%, ANI of 95.83%, and dDDH of 63.6% relative to *K. oralis* UB-38, as well as non-identification by MALDI-TOF mass-spectrometry [4]. We replicated each of the genomic calculations reported, as well as additional dDDH calculations using alternative formulae (Tables 1 and 4) [28, 36]. The 16S rRNA similarity and ANI score are both above the threshold for classification of *K. bonacorsii* and *K. oralis* as the same species, consistent with the very similar G+C% and their common niche. Importantly, we note that the algorithm used to calculate dDDH is critical in this instance. Analysis with either *d0* or *d6* yields dDDH above the threshold of 70%.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Only if dDDH is calculated with *d4,* which is considered most appropriate for highly fragmented or gapped genomes, the dDDH falls below the accepted threshold of 70% (Table 4). Very few isolates of *K. oralis* are publicly available for additional comparisons, and without additional isolates and sequencing data, it is difficult to appreciate the diversity of *K. oralis* or account for bias in strain selection. Based on the available data, we conclude that *K. bonacorsii* is likely not a unique species in the genus *Kingella* and instead may be a subspecies of *K. oralis.*

Finally, another recent evolutionary study of these taxa has used less stringent similarity and gene presence cutoff to define the core genome for species in the *Neisseriaceae* for subsequent phylogenetic analyses [15]. We replicated this analysis by reanalyzing the data using a sequence similarity cut off of 50% for genes present in 99% or 80% of analyzed genomes. By lowering the threshold for a gene to be considered part of the core genome, we increase the number of genes considered from 109 at the 75% similarity level up to 620 genes. Notably, as we increase the number of genes considered part of the core genome, the phylogenetic architecture remains constant (Fig. 3A).

The reproducibility of the core genome phylogenetic structure very strongly supports the inclusion of *Kingella*, *Alysiella*, and *Simonsiella* in the same genus, and the exclusion of *K. potus*. We hypothesized that prior misclassification could be due to high levels of horizontal gene transfer of accessory genes between *Kingella* (*sensu stricto*) genomes. To test this, we generated a presence/absence matrix generated by Roary at 75% sequence similarity level, a technique that emphasizes accessory genes over core genes, and we reconstructed a phylogeny using the RAxML binary model with gamma model of rate heterogeneity, and bootstrapped 100 times (Fig. 3B). This was the only analysis in which we observed a monophyletic *Kingella* genus, based on current

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

taxonomy, though *Kingella potus* was still excluded. *Simonsiella*, *K. oralis*, and *K. kingae* utilize specific DNA-uptake sequences (DUS) to limit horizontal gene transfer (HGT) events, with *Simonsiella*, *K. oralis* seemingly sharing the same DUS [37]. Given that members of the *Neisseriaceae* are highly recombinogenic and undergo significant HGT of accessory genes, we assert that the core genome is a higher fidelity evolutionary record for this family [38].

Based on these results, we propose several changes to the genera *Kingella, Alysiella, and Simonsiella.* As *K. potus* is more related to *N. bacilliformis* than other *Kingella* species by 16S rRNA, OGRI, and whole genome analyses, we propose revising the taxonomic nomenclature of this isolate to reflect this close relationship. Additionally, as phylogenomic analysis of the core genomes of *Alysiella* and *Simonsiella* supports their close relationship to the genus *Kingella,* we propose incorporating these genera into the genus *Kingella*, as *K. crassa*, *K. filiformis*, and *K. muelleri*. These changes would adjust the accepted taxonomic nomenclature to better reflect the phylogenomic relationships between these isolates and result in a monophyletic *Kingella* genus. Finally, further characterization will be required for confirm if *K. bonacorsii* constitutes its own species independent of *K. oralis* or if it would be better classified as a subspecies (*K. oralis* subsp. *bonacorsii*).

Acknowledgements.

We thank the members of the Planet and St. Geme Labs for their useful feedback and discussion.

Funding.

This work was supported by the National Institute of Allergy and Infectious Diseases under the training grant in Microbial Pathogenesis and Genomics (5T32141493) to DPM and under award 1R01AI121015 to JWS.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

147 **Conflicts of Interest.**

148 The authors report no conflicts of interest.

149

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References.

1. **Parte AC, Carbasse JS, Meier-Kolthoff JP, Reimer LC, Göker M**. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 2020;70:5607–5612.

2. **Parker CT, Tindall BJ, Garrity GM**. International code of nomenclature of Prokaryotes. *Int J Syst Evol Microbiol* 2019;69:S1.

3. **Oren A, Garrity G**. Notification of changes in taxonomic opinion previously published outside the IJSEM. List of changes in taxonomic opinion no. 36. *Int J Syst Evol Microbiol* 2022;72:005413.

4. **Antezack A, Boxberger M, Rolland C, Monnet-Corti V, La Scola B**. Isolation and Characterization of Kingella bonacorsii sp. nov., A Novel Kingella Species Detected in a Stable Periodontitis Subject. *Pathogens* 2021;10:1–13.

5. **Henricksen SD, Bovre K**. Transfer of *Moraxella kingae* Henriksen and Bovre to the genus *Kingella* gen. nov. in the family *Neisseriacea*e. *Int J Syst Bacteriol* 1976;26:447–450.

6. **El Houmami N, Bakour S, Bzdrenga J, Rathored J, Seligmann H, *et al.*** Isolation and characterization of *Kingella negevensis* sp. nov., a novel *Kingella* species detected in a healthy paediatric population. *Int J Syst Evol Microbiol* 2017;67:2370–2376.

7. **Dewhirst FE, Chen CKC, Paster BJ, Zambon JJ**. Phylogeny of species in the family Neisseriaceae isolated from human dental plaque and description of *Kingella oralis* sp. nov [corrected]. *Int J Syst Bacteriol* 1993;43:490–499.

8. **Lawson PA, Malnick H, Collins MD, Shah JJ, Chattaway MA, *et al.*** Description of Kingella potus sp. nov., an Organism Isolated from a Wound Caused by an Animal Bite. *J Clin*

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Microbiol 2005;43:3526.

9. **Snell JJS, Lapage SP**. Transfer of some saccharolytic *Moraxella* species to *Kingella* Henriksen and Bovre 1976, with descriptions of *Kingella indologenes* sp. nov. and *Kingella denitrificans* sp. nov. *Int J Syst Bacteriol* 1976;26:451–458.

10. **Hollis DG, Weaver RE, Riley PS**. Emended description of *Kingella denitrificans* (Snell and Lapage 1976): correction of the maltose reaction. *J Clin Microbiol* 1983;18:1174.

11. **Yagupsky P**. *Kingella kingae*: Carriage, transmission, and disease. *Clinical Microbiology Reviews* 2015;28:54–79.

12. **Chen C, Chen C**. Distribution of a newly described species, *Kingella oralis*, in the human oral cavity. *Oral Microbiol Immunol* 1996;11:425–427.

13. **Kopyt N, Kumar A, Agrawal V**. A Case of Suppurative Peritonitis by a Commensal Oral Organism, *Kingella denitrificans*, in an Adult Peritoneal Dialysis Patient. *Perit Dial Int* 2015;35:105.

14. **Adeolu M, Gupta RS**. Phylogenomics and molecular signatures for the order Neisseriales: proposal for division of the order Neisseriales into the emended family Neisseriaceae and Chromobacteriaceae fam. nov. *Antonie Van Leeuwenhoek* 2013;104:1–24.

15. **Nyongesa S, Weber PM, Bernet È, Pulido F, Nieves C, *et al.*** Evolution of longitudinal division in multicellular bacteria of the Neisseriaceae family. *Nat Commun 2022 131* 2022;13:1–18.

16. **Garrity GM, Bell JA, Lilburn T**. Class II. Betaproteobacteria class. nov. *Bergey's Manual®* *Syst Bacteriol* 2005;575–922.

17. **Enright MC, Carter PE, MacLean IA, McKenzie H**. Phylogenetic relationships between

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

some members of the genera *Neisseria*, *Acinetobacter, Moraxella*, and *Kingella* based on partial 16S ribosomal DNA sequence analysis. *Int J Syst Bacteriol* 1994;44:387–391.

18. **Maiden MCJ, Harrison OB**. Population and functional genomics of *Neisseria* revealed with gene-by-gene approaches. *J Clin Microbiol* 2016;54:1949–1955.

19. **Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, *et al.*** 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J Biotechnol* 2017;261:169–176.

20. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, *et al.*** The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;41:D590–D596.

21. **Stamatakis A**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312.

22. **Katoh K, Standley DM**. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 2013;30:772.

23. **Letunic I, Bork P**. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.

24. **Han XY, Hong T, Falsen E**. *Neisseria bacilliformis* sp. nov. Isolated from Human Infections. *J Clin Microbiol* 2006;44:474.

25. **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 91* 2018;9:1–8.

26. **Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M**. TYGS and LPSN: a database

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* 2022;50:D801–D807.

27. **Meier-Kolthoff JP, Göker M**. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun 2019 101* 2019;10:1–10.

28. **Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M**. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013;14:1–14.

29. **Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, *et al.*** Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Int J Syst Evol Microbiol* 1987;37:463–464.

30. **Dewhirst FE, Paster BJ, Bright PL**. *Chromobacterium, Eikenella, Kingella, Neisseria, Simensiella*, and *Vitreoscilla* species comprise a major branch of the beta group Proteobacteria by 16S ribosomal ribonucleic acid sequence comparison: Transfer of *Eikenella* and *Simonsiella* to the family Neisseriaceae (emend.). *Int J Syst Bacteriol* 1989;39:258–266.

31. **Seemann T**. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.

32. **Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, *et al.*** Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.

33. **Hedlund BP, Tønjum T**. *Simonsiella*. *Bergey's Man Syst Archaea Bact* 2015;1–12.

34. **Xie CH, Yokota A**. Phylogenetic analysis of *Alysiella* and related genera of Neisseriaceae: proposal of *Alysiella crassa* comb. nov., *Conchiformibium steedae* gen. nov., comb. nov.,

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conchiformibium kuhniae sp. nov. and *Bergeriella denitrificans* gen. nov., comb. nov. *J Gen Appl Microbiol* 2005;51:1–10.

35. **Barco RA, Garrity GM, Scott JJ, Amend JP, Nealson KH, *et al.*** A genus definition for bacteria and archaea based on a standard genome relatedness index. *MBio*;11. Epub ahead of print 2020. DOI: 10.1128/MBIO.02475-19/SUPPL_FILE/MBIO.02475-19-S0001.DOCX.

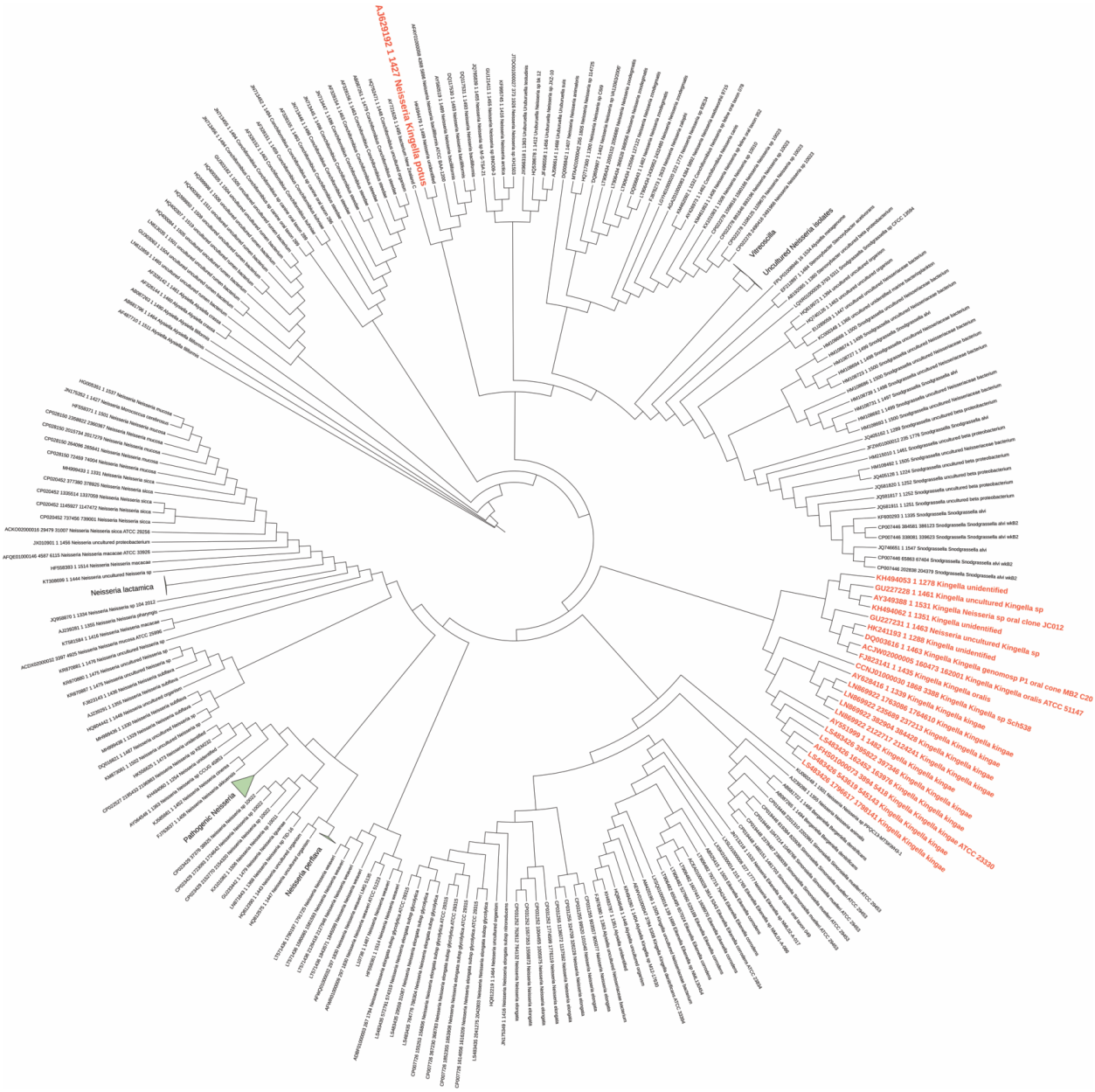
36. **Auch AF, von Jan M, Klenk HP, Göker M**. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2010;2:117.

37. **Frye SA, Nilsen M, Tønjum T, Ambur OH**. Dialects of the DNA Uptake Sequence in Neisseriaceae. *PLOS Genet* 2013;9:e1003458.

38. **Hanage WP, Fraser C, Spratt BG**. Fuzzy species among recombinogenic bacteria. *BMC Biol* 2005;3:1–7.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

253 **Figures.**



254

255 **Figure 1.** High quality 16S rRNA sequences were downloaded from the SILVA database for the

256 family Neisseriaceae (taxid: 408). Sequences were aligned with MAFFT, and a maximum

257 likelihood phylogeny was constructed using RAxML. Species that belong to the genus *Kingella* are

258 colored red. By 16S rRNA relatedness, *Kingella* forms a polyphyletic clade with *K. potus* falling

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

259 outside of the rest of the genus. *K. potus* forms a clade with the oral-associated *Neisseria*

260 *bacilliformis*. Bootstrap values greater than 70% are shown.

261

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

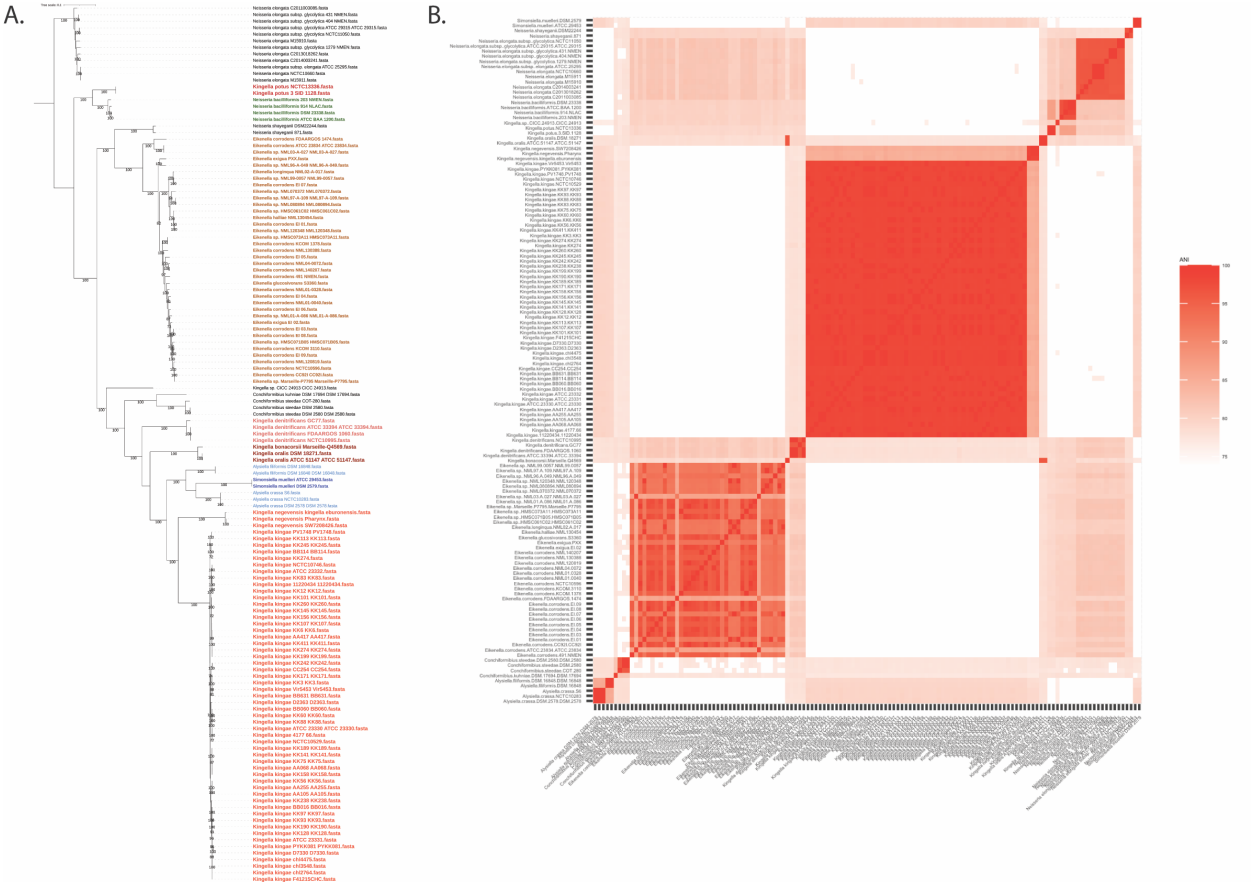


Figure 2. A. The core genome of the strains from the 13 species most closely related to *Kingella kingae* was determined using Roary with a minimum homology cutoff of 75%. The core genomes, which include the sequences of 109 genes total, which were aligned with PRANK in Roary and RAxML was used to reconstruct a maximum likelihood phylogeny. *K. potus* is only distantly related to the rest of the species in the genus *Kingella.* The genera *Alysiella* and *Simonsiella* fall within the *Kingella* clade, suggesting that the core genome of the species in each of these genera is very closely related to that of *Kingella.* Additionally, *K. bonacorsii* is very closely related to *K. oralis,* and these two species form a well-supported clade. The tree is rooted by *N. elongata,* and bootstraps greater than 70% are shown. **B.** ANI score between each isolate is shown on the tree.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

272 The ANI score when comparing *K. oralis* and *K. bonacorsii* suggests that differentiating between

273 these isolates as two species may not reflect speciation events.

274

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3. A. The core genome of the strains from the 13 species most closely related to *Kingella kingae* was determined using Roary with a minimum protein similarity cutoff of 50%, effectively reducing the stringency for genes to be considered a part of the core genome. The core genomes, which include the sequences of 620 CDS total, were aligned with MAFFT, and RAxML was used to reconstruct a maximum likelihood phylogeny with 100 bootstraps. Bootstrap values >70% are shown, and the tree is rooted by *N. elongata*. Unlike previous analyses with

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

similar datasets, *Alysiella* and *Simonsiella* are still more closely related to *K. kingae* than to either *K. oralis* or *K. denitrificans*. **B.** The pangenome of the species closely related to *Kingella* was determined using Roary, at the 75% similar level and ignoring genes that occur in all isolates. RAxML was used to reconstruct a phylogeny of gene presence/absence. Bootstrap values >70% are shown, and the tree is rooted by *N. elongata*. This phylogeny is different than the 75% sequence similarity tree shown in panel A, suggesting that the pangenome gene content may heavily influence the phylogenetic relation relationship between these species, clouding the evolutionary history demonstrated by core genes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Average ANI scores calculated among taxa of interest. ANI scores were calculated using FastANI and are shown only if the ANI >75%. Empty cells denote data excluded from the table.

	N. bacilliformis	*K. denitrificans*	*K. oralis*	*K. negevensis*	*K. kingae*	*N. elongata*	*A. crassa*	*A. filiformis*	*S. muelleri*
K. potus	86.75	78.88	78.74	<75	<75	80.7	-	-	-
K. bonacorsii	-	79.84	96.41	78.06	77.9	78.76	77.98	77.87	77.14
A. crassa	-	77.81	77.96	79.37	79.74	<75	100	85.43	79.79
A. filiformis	-	77.81	77.96	79.37	79.74	<75	85.43	100	78.75
S. muelleri	-	<75	76.9	79.16	79.96	<75	79.79	78.75	100

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

295 **Table 2. dDDH values of *K. potus* calculated against closely related taxa.** dDDH was calculated
296 using TYGS using each of the possible algorithms, as well as the difference in G+C content
297 between *K. potus* and each of the subject type strains.

Query strain	Subject strain	dDDH (d0, in %)	C.I. (d0, in %)	dDDH (d4, in %)	C.I. (d4, in %)	dDDH (d6, in %)	C.I. (d6, in %)	G+C content difference (in %)
K. potus NCTC13336	*N. bacilliformis* ATCC BAA-1200	45.6	[42.2 - 49.0]	30	[27.6 - 32.5]	41	[38.0 - 44.0]	1.86
K. potus NCTC13336	*N. elongata* subsp. *glycolytica* ATCC 29315	20.3	[17.1 - 23.9]	22.5	[20.2 - 25.0]	19.7	[17.0 - 22.8]	3.71
K. potus NCTC13336	*K. negevensis* Sch538	12.9	[10.2 - 16.2]	23.5	[21.2 - 25.9]	13.3	[10.9 - 16.0]	12.24

298

299

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

300 **Table 3. OGRI of *Kingella*, *Alysiella*, and *Simonsiella*.** dDDH was calculated using TYGS using each
301 of the possible algorithms, as well as the difference in G+C content between type strains of
302 *Alysiella* or *Simonsiella* species and each of the closest related subject type strains.

Query strain	Subject strain	dDDH (d0, in %)	C.I. (d0, in %)	dDDH (d4, in %)	C.I. (d4, in %)	dDDH (d6, in %)	C.I. (d6, in %)	G+C content difference (in %)
A. crassa NCTC10283	*A. filiformis* DSM 16848	21.7	[18.5 - 25.3]	34.4	[31.9 - 36.9]	21.9	[19.1 - 25.0]	1.29
A. crassa NCTC10283	*W. bovis* CCUG 44465T	18.4	[15.3 - 22.0]	33.4	[31.0 - 35.9]	18.8	[16.1 - 21.8]	2.67
A. crassa NCTC10283	*S. muelleri* ATCC 29453	16.9	[13.8 - 20.4]	23.2	[20.9 - 25.6]	16.9	[14.3 - 19.8]	3.83
A. crassa NCTC10283	*K. kingae* ATCC 23330	16.3	[13.3 - 19.8]	23.8	[21.5 - 26.3]	16.4	[13.8 - 19.3]	1.45
A. crassa NCTC10283	*K. negevensis* Sch538	15.6	[12.7 - 19.0]	23.3	[21.0 - 25.8]	15.7	[13.2 - 18.6]	0.17
A. crassa NCTC10283	*K. bonacorsii* Marseille-Q4569	14.2	[11.4 - 17.6]	22.7	[20.5 - 25.2]	14.5	[12.0 - 17.3]	8.71
A. crassa NCTC10283	*K. oralis* ATCC 51147	14.2	[11.4 - 17.6]	22	[19.7 - 24.4]	14.5	[12.1 - 17.3]	9.01
A. filiformis DSM 16848	*A. crassa* NCTC10283	21.8	[18.6 - 25.5]	34.1	[31.7 - 36.6]	22	[19.2 - 25.1]	1.25
A. filiformis DSM 16848	*W. bovis* CCUG 44465T	17.2	[14.2 - 20.8]	30.5	[28.1 - 33.0]	17.5	[14.9 - 20.5]	3.95
A. filiformis DSM 16848	*K. kingae* ATCC 23330	16.5	[13.5 - 20.0]	24	[21.7 - 26.5]	16.5	[14.0 - 19.5]	0.17
A. filiformis DSM 16848	*K. negevensis* Sch538	15.1	[12.2 - 18.6]	24.1	[21.8 - 26.6]	15.3	[12.8 - 18.2]	1.11
A. filiformis DSM 16848	*S. muelleri* ATCC 29453	15.1	[12.2 - 18.6]	23.7	[21.4 - 26.1]	15.3	[12.8 - 18.2]	5.12
A. filiformis DSM 16848	*C. steedae* DSM 2580	15.1	[12.2 - 18.5]	22.6	[20.3 - 25.0]	15.3	[12.8 - 18.2]	4.17
A. filiformis DSM 16848	*K. bonacorsii* Marseille-Q4569	14.6	[11.7 - 18.0]	22.8	[20.5 - 25.2]	14.8	[12.3 - 17.6]	7.43
A. filiformis DSM 16848	*K. oralis* ATCC 51147	14.6	[11.8 - 18.0]	22.2	[19.9 - 24.7]	14.8	[12.4 - 17.7]	7.72

303

304

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.13.512104; this version posted October 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4. dDDH calculations for the proposed species *K. bonacorsii.* dDDH was calculated using TYGS using each of the possible algorithms, as well as the difference in G+C content between *K. bonacorsii* and each of the subject type strains.

Query strain	Subject strain	dDDH (d0, in %)	C.I. (d0, in %)	dDDH (d4, in %)	C.I. (d4, in %)	dDDH (d6, in %)	C.I. (d6, in %)	G+C content difference (in %)
K. bonacorsii Marseille-Q4569	*K. oralis* ATCC 51147	84.3	[80.5 - 87.4]	63.6	[60.7 - 66.4]	83.1	[79.8 - 86.0]	0.29
K. bonacorsii Marseille-Q4569	*K. denitrificans* ATCC 33394	15.9	[13.0 - 19.4]	23.2	[20.9 - 25.7]	16	[13.5 - 18.9]	0.06
K. bonacorsii Marseille-Q4569	*N. bacilliformis* ATCC BAA-1200	15.1	[12.2 - 18.5]	25.3	[23.0 - 27.8]	15.3	[12.8 - 18.2]	5.55
K. bonacorsii Marseille-Q4569	*K. potus* NCTC13336	14.4	[11.6 - 17.8]	23.2	[20.9 - 25.7]	14.7	[12.2 - 17.5]	3.84
K. bonacorsii Marseille-Q4569	*K. negevensis* Sch538	14.4	[11.6 - 17.8]	22	[19.7 - 24.4]	14.7	[12.2 - 17.5]	8.54