Phylogenomic analysis of the understudied *Neisseriaceae* species reveals a poly- and paraphyletic *Kingella* genus

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 Philadelphia, Philadelphia PA.

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

*Contributed Equally

#Corresponding Author: Paul J. Planet (planetp@chop.edu)

Running title: Polyphyly in the genus *Kingella*

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, CA: common ancestor

Abstract

Taxonomic classification and phylogenetic analysis of the *Neisseriaceae* family has focused on the pathogens *Neisseria meningitidis* and *Neisseria gonorrhoeae*. Less is known about the relationships of commensal *Neisseria* species and other *Neisseriaceae* genera, raising the possibility that the phylogeny of this family may not agree with taxonomy. In this study we used available nucleotide sequences and a phylogenetic approach to assess the *Kingella* genus and its relatives. We found that this genus is both paraphyletic and polyphyletic. *Kingella potus* is more closely related to *Neisseria bacilliformis* than other *Kingella* species. The *Alysiella* and *Simonsiella* genera form a distinct clade within the *Kingella* genus that is closely related to the pathogens *K. kingae* and *K. negevensis*. We find a phylogenetic relationship between *Conchiformibius, Alysiella, Simonsiella,* and *Kingella,* which we name the CASK clade. Finally, we define the gene sets that differentiate each genus of the CASK clade from one another and from the rest of the *Neisseriaceae* family.

Importance

Understanding the evolutionary relationships between the species in the *Neisseriaceae* has been a persistent challenge in bacterial systematics due to high recombination rates in these species. Previous studies of this family have focused on *N. meningitidis* and *N. gonorrhoeae*. However, previously understudied *Neisseriaceae* species are gaining new attention, with *K. kingae* now recognized as a common human pathogen and with *Alysiella* and *Simonsiella* being unique in the bacterial world as multicellular organisms. A better understanding of the genomic evolution of the *Neisseriaceae* can lead to identification of the specific genes and traits that characterize the remarkable diversity of this family.

24 **Introduction**

25 Defining taxonomic relationships is a cornerstone of microbial systematics, but in species
26 that reproduce asexually and trade genetic information through active horizontal gene transfer
27 across the species boundary, defining the relationships between species can be challenging. As
28 in many taxonomic disciplines, the relationships amongst and within species have traditionally
29 been established using bacterial cell morphology and other phenotypic properties. With the
30 advent of molecular techniques such as DNA-DNA hybridization, 16S rRNA gene sequencing,
31 multilocus sequence typing (MLST), and matrix-assisted laser desorption/ionization time of flight
32 (MALDI-TOF), this system gave way to the so called polyphasic classification strategy that seeks
33 to consider both phenotypic and molecular data (1, 2). As whole genome sequences became
34 readily available, a newer version of this system, taxonomogenomics, seeks to incorporate this
35 rich genomic data into taxonomical descriptions (3–6). This approach also has the benefit of being
36 explicitly phylogenetic, providing a non-arbitrary approach to classification (5, 7, 8). While there
37 is still controversy about how best to integrate phenotypic and phylogenomic data in bacterial
38 taxonomy, there is a general consensus that taxonomy should be reflective of the evolutionary
39 relationships and that monophyly is a desirable characteristic of taxonomic groups.

40 The *Neisseriaceae* family is a group of Gram-negative, β-proteobacteria that currently
41 includes 20 genera: *Neisseria*, *Wielerella*, *Vitreoscilla*, *Uruburuella*, *Stenoxybacter*, *Snodgrassela*,
42 *Simonsiella*, *Rivicola*, *Prolinoborus*, *Paralysiella*, *Morococcus*, *Kingella*, *Eikenella*, *Crenobacter*,
43 *Craterilacuibacter*, *Conchiformibius*, *Bergeriella*, *Aquella*, *Amantichitinum*, and *Alysiella* (9–11).
44 The majority of prior work devoted to understanding the taxonomic relationships between these
45 genera has focused primarily on the pathogenic *Neisseria* species, *N. meningitidis* and *N.*

46 *gonorrhoeae* (2, 8, 12–17). Early work using 16S rRNA failed to properly resolve species
47 boundaries, resulting in several polyphyletic genera (18). In addition, pervasive horizontal gene
48 transfer (HGT) and recombination in the *Neisseriaceae* further increases taxonomic ambiguity,
49 with clonal lineages punctuated by large recombination events between distant relatives (14, 19).
50 HGT results in the formation of ill-defined species with ambiguous species boundaries (20). To
51 help resolve this ambiguity, recent studies of the commensal *Neisseria* species employed core
52 genome MLST (cgMLST) and proposed genus specific cut-offs for defining species using genome
53 relatedness indices (8, 19, 20). However, these studies did not comprehensively address other
54 polyphyletic genera within the *Neisseriaceae*.

55 The *Kingella* genus includes five taxa: *K. potus, K. oralis, K. denitrificans, K. negevensis,*
56 and *K. kingae* (21–27). Recently, two novel species have been proposed, namely *K. bonacorsii*
57 *and K. pumchi* (28, 29). All of the *Kingella* species are typically associated with the oral or
58 oropharyngeal microbiome, a common niche for other *Neisseriaceae*. Isolates of *K. kingae*, *K.*
59 *negevensis*, and *K. potus* have all been associated with invasive disease in immunocompetent
60 individuals (23, 25, 30). In contrast, *K. oralis* and *K. denitrificans* are present in dental plaque and
61 in gingival and periodontal samples and only rarely cause invasive disease, typically in
62 immunocompromised individuals (26, 31, 32). *Kingella* species are oxidase positive, catalase
63 negative, rods and typically form short chains (33).

64 A review of published phylogenetic relationships among the commensal *Neisseria* species
65 reveals a high level of variability in the proposed taxonomic relationships, primarily owing to the
66 genes selected and the method used to perform the analysis (15, 34–36). In many of these studies
67 the *Kingella* genus is polyphyletic, and the genera that interrupt a monophyletic clade with

68 *Kingella* species differ in different published analyses (for examples see 24, 33, 34, 37, 38). Most
69 isolates used for defining taxonomic nomenclature in this genus were collected prior to pervasive
70 whole genome sequencing and were assigned to *Kingella* as a result of phenotypic and/or
71 minimal sequence-based analyses. Additionally, as the frequency of culture-dependent and
72 culture-independent microbiome studies have become more prevalent, the number and
73 diversity of formally accepted species in the *Neisseriaceae* have increased (13). Here, we seek to
74 reassess the relatedness of the five species in the *Kingella* genus using whole genomes and a
75 phylogenomic approach.
76
77 **Results**
78 ***Kingella potus* is distinct from the other species in the *Kingella* genus**
79 To examine the *Kingella* species, 16S rRNA gene sequences for isolates from the
80 *Neisseriaceae* were collected for phylogenetic reconstruction (Fig. 1). By 16S rRNA relatedness
81 alone, it is clear that a major driver of the polyphyletic structure of *Kingella* is *K. potus.* This
82 species was recovered from an infected wound caused by an animal bite and is more closely
83 related to *Neisseria bacilliformis* than it is to any currently described *Kingella* species (25).
84 Phenotypically, *K. potus* and *N. bacilliformis* are characterized for a limited number of
85 characteristics, though both species are non-hemolytic, indole-negative coccobacilli that produce
86 acid from glucose, maltose, and fructose (Table S3).
87 To better understand the relationship between *K. potus* and *N. bacilliformis*, we
88 calculated average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) between
89 the type strains of these species. The ANI between *K. potus* and *N. bacilliformis* was calculated at

90 86.75%, below the established species cut-off of 95% (39). However, the ANI between *K. potus*
91 and *N. bacilliformis* is appreciably greater than the ANI calculated between *K. potus* and other
92 *Kingella* species, which ranges between <75% (*K. kingae*) and 78.8% (*K. denitrificans*) (Table 1,
93 Fig. 2). The calculated dDDH between these species falls below species thresholds at 30% (95%
94 confidence interval 27.6-32.5%) (Table 2).
95
96 **The *Kingella* genus is paraphyletic**
97 To further define the relationships among *Kingella* and closely related species, whole
98 genome sequences for select species in the *Neisseriaceae* were downloaded from NCBI. As 16S
99 rRNA gene sequences have been historically used to define species in this group, all following
100 analyses were limited to species most closely related to *Kingella* by 16S rRNA gene phylogeny
101 (Fig. 1). This group includes *Neisseria elongata, Neisseria sheyganii, N. bacilliformis, Eikenella*
102 *corrodens, Alysiella filiformis, Alysiella crassa, Simonsiella muelleri, Conchiformibius steedae, K.*
103 *oralis, K. bonacorsii, K. potus, K. denitrificans, K. kingae,* and *K. negevensis. N. elongata*, *N.*
104 *sheyganii,* and *E. corrodens* were included as outgroups for these analyses. In total, 134 whole
105 genome sequences were downloaded from NCBI and annotated with Prokka.
106 To minimize the impact of HGT and variability in the pangenome, Roary was used to
107 identify the core genome (40). Core genes were clustered at the 75%, 80%, 85%, 90%, and 95%
108 sequence similarity level and were limited to only those clusters of orthologous groups (COGs)
109 found in all genomes. Core gene alignments were generated and used to reconstruct maximum-
110 likelihood phylogenies (Fig. 3). Both the Robinson-Foulds distance and normalized Nye Similarity
111 metrics were used to quantify tree similarity. Each phylogeny was highly congruent with the

112 others based on these metrics, and each included a polyphyletic clade for the genus *Kingella* (Fig.
113 3B and C). We therefore proceeded with the lowest similarity cutoff of 75% for further analyses
114 (Fig. 3A). In each of these analyses, *Conchiformibius* spp., *Alysiella* spp., *Simonsiella* spp, and
115 *Kingella* spp. (with the exception of *K. potus*) form a closely related group that we will refer to as
116 the CASK clade. Phenotypic data on the species in the CASK clade are extremely limited, limiting
117 the utility of phenotypic properties in understanding relationships within the group (Table S3).

118 Based on a whole genome approach, *Alysiella* and *Simonsiella* form a clade within the
119 *Kingella* genus (Fig. 3A), dividing *K. oralis* and *K. denitrificans* from *K. negevensis* and *K. kingae.*
120 ANI, dDDH, and G+C% were calculated for type strains of *Kingella*, *Alysiella*, and *Simonsiella* and
121 are shown in Tables 1 and 3. All of these overall genome relatedness indices (OGRI) support a
122 close relationship between the species of these three genera.

123 Additionally, we calculated pairwise ANI scores for each of the genomes listed above. The
124 ANI scores confirm the majority of the proposed species classifications except for *K. bonacorsii*,
125 an unclassified *Kingella* isolate that was recently proposed to be a new species (28). This isolate
126 was recovered from a gingival sample, and the authors report a 16S rRNA similarity of 98.7%, ANI
127 of 95.83%, and dDDH of 63.6% relative to *K. oralis* UB-38, as well as non-identification by MALDI-
128 TOF mass-spectrometry (28). We replicated each of the genomic calculations reported as well as
129 additional dDDH calculations using alternative formulae and found that choice of formula is
130 essential, as only *d4* yielded a dDDH below the accepted values for species cutoffs (Tables 1 and
131 4, Fig. 2) (39, 41).

132 ***Kingella* species are unified by gene presence/absence in the pangenome**

133 Finally, another recent evolutionary study of these taxa has used less stringent similarity
134 and gene presence cutoffs to define the core genome for species in the *Neisseriaceae* for
135 subsequent phylogenetic analyses (34). We replicated this analysis by reanalyzing the data using
136 a sequence similarity cut off of 50% for genes present in 80% of analyzed genomes. By lowering
137 the threshold for a gene to be considered part of the core genome, we increase the number of
138 genes considered from 109 at the 75% similarity level up to 620 genes. Notably, as we increased
139 the number of genes considered part of the core genome by decreasing the stringency of our
140 Roary analysis, the presence of a polyphyletic *Kingella* clade remains constant (Fig. 4A),
141 demonstrating the robust relationships between these genera that are obtained when we
142 consider the core genome of species in the *Neisseriaceae*.

143 To better understand how the pangenome impacts the relationships described in the
144 above analyses, we reconstructed the phylogeny considering only gene presence and absence in
145 the pangenome. This analysis was performed with a binary COG presence/absence matrix as
146 defined by Roary with a 75% similarity cut-off with phylogenetic inference in IQ-Tree. For the first
147 time in any of our analyses, all named *Kingella* species except for *K. potus* are monophyletic,
148 excluding the *Simonsiella* and *Alysiella* genomes. To identify genes that drive the unification of
149 *Kingella sensu stricto* (*Kingella s.s.*), IQ-Tree was used to perform ancestral reconstructions of the
150 species in this tree. These ancestral reconstructions facilitate the identification of the genes that
151 drive the unification of *Kingella* into a monophyletic clade. Discounting statistically ambiguous
152 sites, monophyly is driven by the loss of eight genes in *Kingella s.s.* (Table 4), all of which fall in
153 the accessory genome. There are no gene gains that support this clade, and therefore this
154 monophyly is not well supported.

155 These data further suggest that horizonal gene transfer (HGT) can confound our ability to
156 define species relationships and that limits on HGT may be important to consider when defining
157 species in the *Neisseriaceae.* HGT is driven by DNA uptake sequences (DUS) in the *Neisseriaceae*
158 (42–44), which are largely thought to be species specific. The DUS in the CASK clade have been
159 previously characterized (42). Type genomes for each species were used to define DUS
160 preference by word search. *Alysiella filliformis*, *Alysiella crassa*, *Simonsiella muelleri*, and *Kingella*
161 *oralis* genomes are enriched for the AG-SimDUS (AGGCTGCCTGAA) and the AG-KingDUS
162 (AGGCAGCCTGAA). *Kingella kingae, Kingella negevensis,* and *Kingella denitrificans* genomes are
163 enriched for the AA-King3DUS (AAGCAGCCTGCA). *Conchiformibius steedae* and *Conchiformibius*
164 *kuhniae* do not appear to use any of these previously defined DUS.
165
166 **The CASK species are unique based on cellular morphology and lifestyle.**
167 To identify the genes that differentiate the species in the CASK clade and to define clusters of
168 orthologous groups (COGs) unique to groups of interest, Scoary was used to analyze the pangenome
169 analysis generated by Roary at the 75% sequence similarity level. The presence of genes is defined as
170 diagnostic for the in-group if they are 100% specific, are 100% sensitive, and have a $p > 0.001$. These
171 cutoffs reduce the impact of shared mobile genetic elements (MGEs) such as prophages or genomic
172 islands.
173 First, the CASK clade was compared against the rest of the species included in the phylogenetic
174 analyses (Fig. 5A, table S5). This clade is enriched for 1360 COGs that are absent in the rest of the analyzed
175 species, providing strong evidence supporting the evolutionary relatedness of this group. The average
176 genome size for the species in this group is approximately 2,000 coding sequences (CDS). To understand
177 putative roles of these COGs in the CASK species, COG groups and Gene Ontology (GO) functional

178 annotations were defined with eggnog mapper (Fig. 5B). Approximately half of the COGs were assigned
179 to the "function unknown" group (n = 656). Of the remaining COGs, the majority play putative roles in cell
180 wall, membrane, and envelope biogenesis (n = 276); DNA replication, recombination, and repair (n = 224);
181 or amino acid transport and metabolism (n = 205).

182 Next, *Kingella sensu stricto (Kingella s.s.)*, comprised of *K. oralis, K. denitrificans, K. negevensis,*
183 and *K. kingae*, was compared to the rest of the CASK clade. Only 67 COGs meet the criteria to be diagnostic
184 for *Kingella s.s* (Table S6). When *Alysiella* and *Simonsiella* or the pathogenic *Kingella* spp. were compared
185 to the rest of CASK clade, we identified only 41 or 79 diagnostic COGs, respectively (Fig. 5A). Again, we
186 assigned putative functional roles for the COGs in each of these groups (Fig. 5C). The COGs diagnostic for
187 *Kingella s.s*. and the pathogenic *Kingella* spp (*K. negevensis* and *K. kingae*) are primarily involved in
188 carbohydrate transport and metabolism or nucleotide transport and metabolism. By contrast, *Alysiella*
189 and *Simonsiella* diagnostic COGs are involved in DNA replication, recombination, and repair or cell wall,
190 membrane, and envelope biogenesis (Fig. 5C).

191 Finally, we compared the genomes of the type-strain for each species in the CASK clade using
192 BLASTn against *K. kingae* ATCC 23330 (Fig. 5D). We found high levels of nucleotide sequence similarity
193 across the *K. kingae* genome relative to the related species. There are few notable regions of dissimilarity,
194 which appear to be MGEs in the *K. kingae* genome.

195

196 **Discussion**

197 Classification of the species in the *Neisseriaceae* has been difficult using traditional metrics of
198 relatedness, both phenotypically and phylogenetically. As a result, the family has been restructured
199 several times since its inception to include or exclude genera (15, 24, 36, 45–47). Species in this family
200 resist robust phylogenetic classification due to pervasive horizontal gene transfer and recombination,
201 including in the 16S rRNA sequence (18–20, 42). Due to the large diversity of species and variability,

202 phenotypic classification using strict phenotypic classification is either difficult or impossible without
203 additional sampling. For example, presence of catalase is conserved among *Neisseria* isolates, and yet
204 isolates of *N. elongata* have been recovered that lack the catalase enzyme(33). Furthermore, as more
205 species are proposed using culture-independent methods or based on a single isolate, it will be essential
206 to establish phylogenomic species definitions for the *Neisseriaceae* (1, 18, 20).

207 In this study, we employed a phylogenetic approach to define the evolutionary relationships
208 within a major subgroup of the *Neisseriaceae*, defined by the genera *Eikenella, Conchiformibius, Alysiella,*
209 *Simonsiella, Kingella*, and select *Neisseria* species. Examination by 16S rRNA sequence unveiled a very
210 recent common ancestor between *K. potus* and *N. bacilliformis*, which was further supported by several
211 other genome relatedness indices (Fig. 1, Tables 1,3). *K. potus* is a catalase-negative species collected from
212 an infected wound that resulted from a bite by a captive South American Kinkajou (*Potus flavus*) (25). *N.*
213 *bacilliformis* is an opportunistic pathogen that has been recovered from human infections, varies in the
214 production of catalase (48), and is phenotypically identical to *N. elongata* subsp. *elongata.* These species
215 were reported nearly simultaneously, preventing their direct comparison. Our reanalysis suggests that *K.*
216 *potus* is much more closely related to *N. bacilliformis* than to any other *Kingella* species, though it is likely
217 a distinct species from *N. bacilliformis*. Inclusion in the genus *Kingella* was likely due to a negative catalase
218 test, underscoring the importance of a genomic approach to classification of these isolates.

219 The phylogenomic analyses also uncovered close evolutionary relationships between
220 *Conchiformibius* sp., *Alysiella* sp., *Simonsiella* sp., and *Kingella* sp., which we have named the
221 "CASK" clade (Fig. 3). The CASK clade is a major branch of the *Neisseriaceae* that includes both
222 invasive pathogens (*K. negevensis* and *K. kingae*) and oral-associated commensals. *Alysiella* and
223 *Simonsiella* are genera associated with the oral microbiome of non-human mammals (15, 47, 49).
224 *Conchiformibius* spp., *Alysiella* spp., and *Simonsiella* spp. are remarkable for their multicellularity,

225 which is the result of incomplete septation by the forming peptidoglycan during bacterial division
226 (34).

227 The reproducibility of the core genome phylogenetic structure very strongly supports the
228 close relationship between the species of the CASK clade and the exclusion of *K. potus* (Fig. 3-4).
229 We hypothesized that prior misclassification could be due to high levels of horizontal gene
230 transfer of accessory genes between *Kingella* (*sensu stricto*) genomes. To test this hypothesis, we
231 generated a presence/absence matrix generated by Roary at 75% sequence similarity level, a
232 technique that emphasizes accessory genes over core genes, and we reconstructed a phylogeny
233 using the RAxML binary model with gamma model of rate heterogeneity and bootstrapped 100
234 times (Fig. 3B). This was the only analysis in which we observed a monophyletic *Kingella* genus,
235 though *Kingella potus* was still excluded. We favor the analyses that are based on nucleotide
236 sequence of the core genome because they contain much more genomic information (~76,500
237 variable sites compared to ~29,000 variable genes) and likely lessen confounding signal from
238 recombined areas. A separate whole gene-based analysis also identified more diagnostic genes
239 uniting *Alysiella*, *Simonsiella,* and the pathogenic *Kingella* species (n=79) than genes uniting the
240 *Kingella* genus *sensu strictu* (n=69). Additionally, as *Simonsiella*, *K. oralis*, and *K. kingae* utilize
241 specific DNA-uptake sequences (DUS), suggesting they may be able to share DNA more easily
242 among themselves as compared to the rest of the *Neisseriaceae*. Given that members of the
243 *Neisseriaceae* are highly recombinogenic and undergo significant HGT of accessory genes, we
244 believe that the core genome is a higher fidelity evolutionary record for this family (20).

245 The CASK clade has a large set of diagnostic genes that differentiate these species from
246 the remainder of the *Neisseriaceae*. Interestingly, this group is enriched for processes that impact

247 cell wall and envelope biogenesis (Fig. 5B), likely driven by the unusual cellular morphologies of
248 *Conchiformibius* spp., *Alysiella* spp., and *Simonsiella* spp. These genera are distinctive for being
249 multicellular, longitudinally-dividing (MuLDi) species that comprise the only known animal
250 multicellular bacterial symbionts (34). Nyongesa and co-workers hypothesize that the MuLDi
251 phenotype evolved twice in the *Neisseriaceae,* first in *Conchiformibius* species and then in the
252 common ancestor of *Alysiella* and *Simonsiella*, a hypothesis that is strongly supported by our
253 phylogenetic analyses. This hypothesis is further supported by consideration of the diagnostic
254 genes that differentiate *Alysiella* and *Simonsiella* from the rest of this group (Fig. 5C). We again
255 find an enrichment for COGs implicated in cellular morphology, as we would expect if these
256 genera evolved the MuLDi phenotype independently of *Conchiformibius* spp. By contrast,
257 *Kingella s.s.* is enriched for COGs associated with nutrient transport and metabolism (Fig. 5C).
258 Our study suggests that the current taxonomic classifications of species in the *Kingella* genus
259 may not reflect evolutionary history. Based on our results, *K. potus* is more related to *N.*
260 *bacilliformis* than to other *Kingella* species by 16S rRNA, OGRI, and whole genome analyses and
261 should probably be assigned to the genus *Neisseria* or another genus entirely. Additionally,
262 phylogenomic analysis of the core genomes of *Alysiella* and *Simonsiella* indicates their close
263 relationship to the *Kingella* genus and supports their reassignment to *Kingella*, resulting in a
264 monophyletic genus.
265
266 **Methods**
267 *Phylogenetic Analysis*

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 (50, 51). 16S sequence accession numbers are listed in table S1. 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 (52, 53).

Whole genome sequences for *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* were downloaded from the NCBI Genome database on September 09, 2022. Sequence accession numbers are listed in table S2. In total, 134 genomes were reassembled with Prokka v1.14.6, using default settings for bacterial sequences (54). Phenotypic characteristics for each of these species were collected by literature review and are listed in table S3. Assemblies were analyzed with Roary v3.13.0. to identify the core genome (40). 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 (-cd 100), unless otherwise noted. Additionally, other 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 (8, 34). 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. Nucleotide sequences of core genes were compared in a codon-aware alignment by PRANK within Roary, and core gene

289 alignments were used to reconstruct phylogenies using RAxML as above. Phylogenetic trees were
290 annotated using Figtree v.1.4.4 and ITOL (55). Bootstrap values >70 are shown.

291 To calculate phylogenetic tree similarity, final trees were analyzed in R using the Ape v5.6-
292 2 (56), ggtree v3.1.2. (57), and TreeDist v.2.5.0 (58, 59) packages . The Robinson-Fould distance
293 and normalized Nye Simillarity Score were determined for each pair of trees (58–61).

294 Ancestral reconstructions were performed using the Roary gene presence/absence (PA)
295 matrix generated at the 75% sequence similarity cut-off. The matrix was converted into a FASTA
296 formatted file of binary data using R. This file was used for phylogenetic reconstruction in IQ-Tree
297 v 1.6.12 with automatic model calling (62, 63). Branch support was determined via ultrafast
298 bootstrapping (-bb 1000) (64), and ancestral states (-asr) were calculated with default cut-offs.
299 Full inferred ancestral state matrices are available in Supplementary File 2. To determine the
300 COGs that unite *Kingella s.s.*, the inferred PA matrix of common ancestor of *Alysiella* spp.,
301 *Simonsiella* spp., and *Kingella s.s.* (Node 7), was compared to the inferred PA matrix of common
302 ancestor of *Kingella* spp. Statistically ambiguous sites present in either PA matrix were removed
303 from this analysis.

304 *Gene Set Enrichment*

305 Scoary v 1.6.16 was used to further analyze the gene presence absence matrix generated
306 by Roary at the 75% sequence similarity cutoff (65). The -r flag was used to limit the set of strains
307 considered in this analysis. COGs were eliminated from this analysis if the Bonferroni-corrected
308 p-value >0.0001. Diagnostic genes are only those genes which 1) meet this statistical threshold,
309 2) are enriched in the clade of interest, and 3) have specificity and sensitivity of >80%.

310 To identify pathways that are enriched in each phylogenetic group, the representative
311 COG sequence was collected from the Roary output files and compiled into a gene set. Gene sets
312 were analyzed with the COG database using eggNOG-mapper v 2.1.9 (66, 67), with a DIAMOND
313 alignment (68) and default settings for bacterial species (Table S4-9). Results were plotted using
314 Graphpad Prism.

315 *Genome Relatedness Indices*

316 Average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) between the
317 type strains of these species using FastANI v.1.33 and the Type Strain Genome Server (TYGS),
318 respectively (69–71).

319

320 **References**

321 1. Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, Raoult D, Fournier PE.
322 2014. A polyphasic strategy incorporating genomic data for the taxonomic description of novel
323 bacterial species. Int J Syst Evol Microbiol 64:384–391.

324 2. Vandamme P, Pot B, Gillis M, Vos P de, Kersters K, Swings J. 1996. Polyphasic taxonomy, a
325 consensus approach to bacterial systematics. Microbiol Rev 60:407.

326 3. Fournier PE, Lagier JC, Dubourg G, Raoult D. 2015. From culturomics to taxonomogenomics: A
327 need to change the taxonomy of prokaryotes in clinical microbiology. Anaerobe 36:73–78.

328 4. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species
329 definition. Proc Natl Acad Sci U S A 106:19126.

330 5. Chung M, Munro JB, Tettelin H, Hotopp JCD. 2018. Using Core Genome Alignments To Assign
331 Bacterial Species. mSystems 3.

332 6. Caputo A, Fournier PE, Raoult D. 2019. Genome and pan-genome analysis to classify emerging
333 bacteria. Biol Direct 14.

334 7. Moldovan MA, Gelfand MS. 2018. Pangenomic definition of prokaryotic species and the
335 phylogenetic structure of *Prochlorococcus spp.* Front Microbiol 9:428.

336 8. Khoder M, Osman M, Kassem II, Rafei R, Shahin A, Fournier PE, Rolain J-M, Hamze M. 2022.
337 Whole Genome Analyses Accurately Identify *Neisseria* spp. and Limit Taxonomic Ambiguity. Int J
338 Mol Sci 23:13456.

339 9. Parte AC, Carbasse JS, Meier-Kolthoff JP, Reimer LC, Göker M. 2020. List of prokaryotic names
340 with standing in nomenclature (LPSN) moves to the DSMZ. Int J Syst Evol Microbiol 70:5607–
341 5612.

342 10. Parker CT, Tindall BJ, Garrity GM. 2019. International code of nomenclature of Prokaryotes. Int J
343 Syst Evol Microbiol 69:S1.

344 11. Oren A, Garrity G. 2022. Notification of changes in taxonomic opinion previously published
345 outside the IJSEM. List of changes in taxonomic opinion no. 36. Int J Syst Evol Microbiol
346 72:005413.
347 12. Bennett JS, Jolley KA, Earle SG, Corton C, Bentley SD, Parkhill J, Maiden MCJ. 2012. A genomic
348 approach to bacterial taxonomy: An examination and proposed reclassification of species within
349 the genus *Neisseria*. Microbiol (United Kingdom) 158:1570–1580.
350 13. Maiden MCJ, Harrison OB. 2016. Population and functional genomics of *Neisseria* revealed with
351 gene-by-gene approaches. J Clin Microbiol 54:1949–1955.
352 14. Vidovic S, Horsman GB, Liao M, Dillon JAR. 2011. Influence of Conserved and Hypervariable
353 Genetic Markers on Genotyping Circulating Strains of *Neisseria gonorrhoeae.* PLoS One 6:e28259.
354 15. Adeolu M, Gupta RS. 2013. Phylogenomics and molecular signatures for the order *Neisseriales*:
355 proposal for division of the order Neisseriales into the emended family *Neisseriaceae* and
356 *Chromobacteriaceae fam. nov.* Antonie Van Leeuwenhoek 104:1–24.
357 16. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA
358 hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol
359 Microbiol 57:81–91.
360 17. Diallo K, MacLennan J, Harrison OB, Msefula C, Sow SO, Daugla DM, Johnson E, Trotter C,
361 MacLennan CA, Parkhill J, Borrow R, Greenwood BM, Maiden MCJ. 2019. Genomic
362 characterization of novel *Neisseria* species. Sci Rep 9.
363 18. Rossi-Tamisier M, Benamar S, Raoult D, Fournier PE. 2015. Cautionary tale of using 16s rRNA
364 gene sequence similarity values in identification of human-associated bacterial species. Int J Syst
365 Evol Microbiol 65:1929–1934.
366 19. Tibayrenc M, Ayala FJ. 2015. How clonal are *Neisseria* species? The epidemic clonality model
367 revisited. Proc Natl Acad Sci U S A 112:8909.

368 20. Hanage WP, Fraser C, Spratt BG. 2005. Fuzzy species among recombinogenic bacteria. BMC Biol
369 3.
370 21. Dewhirst FE, Paster BJ, Bright PL. 1989. *Chromobacterium*, *Eikenella*, *Kingella*, *Neisseria*,
371 *Simensiella*, and *Vitreoscilla* species comprise a major branch of the beta group Proteobacteria by
372 16S ribosomal ribonucleic acid sequence comparison: Transfer of *Eikenella* and *Simonsiella* to the
373 family *Neisseriaceae* (emend.). Int J Syst Bacteriol 39:258–266.
374 22. Henricksen SD, Bovre K. 1976. Transfer of *Moraxella kingae* Henriksen and Bovre to the genus
375 *Kingella gen. nov.* in the family *Neisseriaceae*. Int J Syst Bacteriol 26:447–450.
376 23. El Houmami N, Bakour S, Bzdrenga J, Rathored J, Seligmann H, Robert C, Armstrong N, Schrenzel
377 J, Raoult D, Yagupsky P, Fournier PE. 2017. Isolation and characterization of *Kingella negevensis*
378 *sp. nov.*, a novel *Kingella* species detected in a healthy paediatric population. Int J Syst Evol
379 Microbiol 67:2370–2376.
380 24. Dewhirst FE, Chen CKC, Paster BJ, Zambon JJ. 1993. Phylogeny of species in the family
381 *Neisseriaceae* isolated from human dental plaque and description of K*ingella oralis sp. nov*
382 [corrected]. Int J Syst Bacteriol 43:490–499.
383 25. Lawson PA, Malnick H, Collins MD, Shah JJ, Chattaway MA, Bendall R, Hartley JW. 2005.
384 Description of *Kingella potus sp. nov.*, an Organism Isolated from a Wound Caused by an Animal
385 Bite. J Clin Microbiol 43:3526.
386 26. Snell JJS, Lapage SP. 1976. Transfer of some saccharolytic *Moraxella* species to *Kingella Henriksen*
387 *and Bovre 1976*, with descriptions of *Kingella indologenes sp. nov.* and *Kingella denitrificans sp.*
388 *nov.* Int J Syst Bacteriol 26:451–458.
389 27. Hollis DG, Weaver RE, Riley PS. 1983. Emended description of *Kingella denitrificans* (Snell and
390 Lapage 1976): correction of the maltose reaction. J Clin Microbiol 18:1174.
391 28. Antezack A, Boxberger M, Rolland C, Monnet-Corti V, La Scola B. 2021. Isolation and

392 Characterization of *Kingella bonacorsii sp. nov.,* A Novel *Kingella* Species Detected in a Stable
393 Periodontitis Subject. Pathogens 10:1–13.
394 29. Xiao M, Liu R, Du J, Liu R, Zhai L, Wang H, Yao S, Xu YC. 2022. *Kingella pumchi sp. nov.*, an
395 organism isolated from human vertebral puncture tissue. Antonie Van Leeuwenhoek
396 https://doi.org/10.1007/S10482-022-01786-W.
397 30. Yagupsky P. 2015. *Kingella kingae:* Carriage, transmission, and disease. Clin Microbiol Rev.
398 American Society for Microbiology https://doi.org/10.1128/CMR.00028-14.
399 31. Chen C, Chen C. 1996. Distribution of a newly described species, *Kingella oralis*, in the human oral
400 cavity. Oral Microbiol Immunol 11:425–427.
401 32. Kopyt N, Kumar A, Agrawal V. 2015. A Case of Suppurative Peritonitis by a Commensal Oral
402 Organism, *Kingella denitrificans*, in an Adult Peritoneal Dialysis Patient. Perit Dial Int 35:105.
403 33. Tønjum T. 2005. Order IV. *Neisseriales ord. nov.*, p. 575–922. *In* Garrity, GM, Bell, JA, Lilburn, T
404 (eds.), Bergey's Manual of Systematic Bacteriology, 2nd ed. Springer, Boston.
405 34. Nyongesa S, Weber PM, Bernet È, Pulido F, Nieves C, Nieckarz M, Delaby M, Viehboeck T, Krause
406 N, Rivera-Millot A, Nakamura A, Vischer NOE, vanNieuwenhze M, Brun Y V., Cava F, Bulgheresi S,
407 Veyrier FJ. 2022. Evolution of longitudinal division in multicellular bacteria of the *Neisseriaceae*
408 family. Nat Commun 2022 131 13:1–18.
409 35. Garrity GM, Bell JA, Lilburn T. 2005. Class II. Betaproteobacteria class. nov. Bergey's Manual® Syst
410 Bacteriol 575–922.
411 36. Enright MC, Carter PE, MacLean IA, McKenzie H. 1994. Phylogenetic relationships between some
412 members of the genera *Neisseria*, *Acinetobacter*, *Moraxella*, and *Kingella* based on partial 16S
413 ribosomal DNA sequence analysis. Int J Syst Bacteriol 44:387–391.
414 37. Adeolu M, Gupta RS. 2013. Phylogenomics and molecular signatures for the order *Neisseriales*:
415 proposal for division of the order *Neisseriales* into the emended family *Neisseriaceae* and

416 *Chromobacteriaceae fam. nov.* Antonie Van Leeuwenhoek 104:1–24.

417 38. Lawson PA, Malnick H, Collins MD, Shah JJ, Chattaway MA, Bendall R, Hartley JW. 2005.
418 Description of *Kingella potus sp. nov.*, an Organism Isolated from a Wound Caused by an Animal
419 Bite. J Clin Microbiol 43:3526.

420 39. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013. Genome sequence-based species
421 delimitation with confidence intervals and improved distance functions. BMC Bioinformatics
422 14:1–14.

423 40. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA,
424 Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics
425 31:3691–3693.

426 41. Auch AF, von Jan M, Klenk HP, Göker M. 2010. Digital DNA-DNA hybridization for microbial
427 species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci
428 2:117.

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

431 43. Weir S, Lee LIW, Marrs C. 1996. Identification of four complete type 4 pilin genes in a single
432 *Kingella denitrificans* genome. Infect Immun 64:4993–4999.

433 44. Treangen TJ, Ambur O, Tonjum T, Rocha EPC. 2008. The impact of the *Neisserial* DNA uptake
434 sequences on genome evolution and stability. Genome Biol 9.

435 45. Xie CH, Yokota A. 2005. Phylogenetic analysis of *Alysiella* and related genera of *Neisseriaceae*:
436 proposal of *Alysiella crassa comb. nov., Conchiformibium steedae gen. nov., comb. nov.,*
437 *Conchiformibium kuhniae sp. nov.* and *Bergeriella denitrificans gen. nov., comb. nov.* J Gen Appl
438 Microbiol 51:1–10.

439 46. Kuhnert P, Brodard I, Bock S, Hemphill A, Akarsu H, Engelhardt A, Kutzer P. 2022. *Wielerella bovis*

440 *gen. nov., sp. nov.* a member of the family *Neisseriaceae* associated with bovine endocarditis. Int
441 J Syst Evol Microbiol 72.
442 47. Xie CH, Yokota A. 2005. Phylogenetic analysis of *Alysiella* and related genera of *Neisseriaceae*:
443 proposal of A*lysiella crassa comb. nov., Conchiformibium steedae gen. nov., comb. nov.,*
444 *Conchiformibium kuhniae sp. nov.* and *Bergeriella denitrificans gen. nov., comb. nov.* J Gen Appl
445 Microbiol 51:1–10.
446 48. Han XY, Hong T, Falsen E. 2006. *Neisseria bacilliformis sp. nov.* Isolated from Human Infections. J
447 Clin Microbiol 44:474.
448 49. Hedlund BP, Tønjum T. 2015. *Simonsiella* . Bergey's Man Syst Archaea Bact 1–12.
449 50. Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, Bruns G, Yarza P, Peplies J,
450 Westram R, Ludwig W. 2017. 25 years of serving the community with ribosomal RNA gene
451 reference databases and tools. J Biotechnol 261:169–176.
452 51. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The
453 SILVA ribosomal RNA gene database project: improved data processing and web-based tools.
454 Nucleic Acids Res 41:D590–D596.
455 52. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
456 phylogenies. Bioinformatics 30:1312.
457 53. Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
458 Improvements in Performance and Usability. Mol Biol Evol 30:772.
459 54. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.
460 55. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
461 display and annotation. Nucleic Acids Res 49:W293–W296.
462 56. Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary
463 analyses in R. Bioinformatics 35:526–528.

464 57. Yu G. 2020. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinforma
465 69:e96.
466 58. Smith MR. 2022. Robust Analysis of Phylogenetic Tree Space. Syst Biol 71:1255–1270.
467 59. Smith MR. 2020. Information theoretic generalized Robinson-Foulds metrics for comparing
468 phylogenetic trees. Bioinformatics 36:5007–5013.
469 60. Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. Math Biosci 53:131–147.
470 61. Nye TMW, Liò P, Gilks WR. 2006. A novel algorithm and web-based tool for comparing two
471 alternative phylogenetic trees. Bioinformatics 22:117–119.
472 62. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R,
473 Teeling E. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the
474 Genomic Era. Mol Biol Evol 37:1530–1534.
475 63. Wang HC, Minh BQ, Susko E, Roger AJ. 2018. Modeling Site Heterogeneity with Posterior Mean
476 Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. Syst Biol 67:216–235.
477 64. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the
478 Ultrafast Bootstrap Approximation. Mol Biol Evol 35:518–522.
479 65. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-
480 genome-wide association studies with Scoary. Genome Biol 17:1–9.
481 66. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper
482 v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic
483 Scale. Mol Biol Evol 38:5825–5829.
484 67. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR,
485 Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally
486 and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses.
487 Nucleic Acids Res 47:D309–D314.

488 68. Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using
489 DIAMOND. Nat Methods 2021 184 18:366–368.
490 69. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI
491 analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 2018 91 9:1–
492 8.
493 70. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. 2022. TYGS and LPSN: a database
494 tandem for fast and reliable genome-based classification and nomenclature of prokaryotes.
495 Nucleic Acids Res 50:D801–D807.
496 71. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-
497 the-art genome-based taxonomy. Nat Commun 2019 101 10:1–10.
498 72. Kuhn DA, Gregory DA, Buchanan GE, Nyby MD, Daly KR. 1978. Isolation, characterization, and
499 numerical taxonomy of *Simonsiella* strains from the oral cavities of cats, dogs, sheep, and
500 humans. Arch Microbiol 118:235–241.
501 73. Yagupsky P, Peled N, Katz O. 2002. Epidemiological Features of Invasive *Kingella kingae* Infections
502 and Respiratory Carriage of the Organism. J Clin Microbiol 40:4180.
503 74. Hedlund BP, Kuhn DA. 2006. The Genera *Simonsiella* and *Alysiella*. The Prokaryotes 828–839.
504 75. Hedlund BP, Staley JT. 2002. Phylogeny of the genus *Simonsiella* and other members of the
505 *Neisseriaceae*. Int J Syst Evol Microbiol 52:1377–1382.
506 76. Andersen BM, Weyant RS, Steigerwalt AG, Moss CW, Hollis DG, Weaver RE, Ashford D, Brenner
507 DJ. 1995. Characterization of *Neisseria elongata subsp. glycolytica* isolates obtained from human
508 wound specimens and blood cultures. J Clin Microbiol 33:76.
509 77. Kuhn DA, Gregory DA, Buchanan GE, Nyby MD, Daly KR. 1978. Isolation, characterization, and
510 numerical taxonomy of *Simonsiella* strains from the oral cavities of cats, dogs, sheep, and
511 humans. Arch Microbiol 118:235–241.

512 78. Bernard KA, Burdz T, Wiebe D, Bernier AM. 2020. Description of *Eikenella halliae sp. nov.* and
513 *Eikenella longinqua sp. nov.,* derived from human clinical materials, emendation of *Eikenella*
514 *exigua* Stormo et al. 2019 and emendation of the genus *Eikenella* to include species which are
515 strict anaerobes. Int J Syst Evol Microbiol 70:3167–3178.
516

517 **Tables.**

518 **Supplemental Tables are attached.**

519 **Table 1. Average ANI scores calculated among taxa of interest.** ANI scores were calculated using
520 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

521

522

523 **Table 2. dDDH values of *K. potus* calculated against closely related taxa.** dDDH was calculated
524 using TYGS using each of the possible algorithms, as well as the difference in G+C content
525 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

526

527

528 **Table 3. OGRI of *Kingella*, *Alysiella*, and *Simonsiella*.** dDDH was calculated using TYGS using each
529 of the possible algorithms, as well as the difference in G+C content between type strains of
530 *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

531

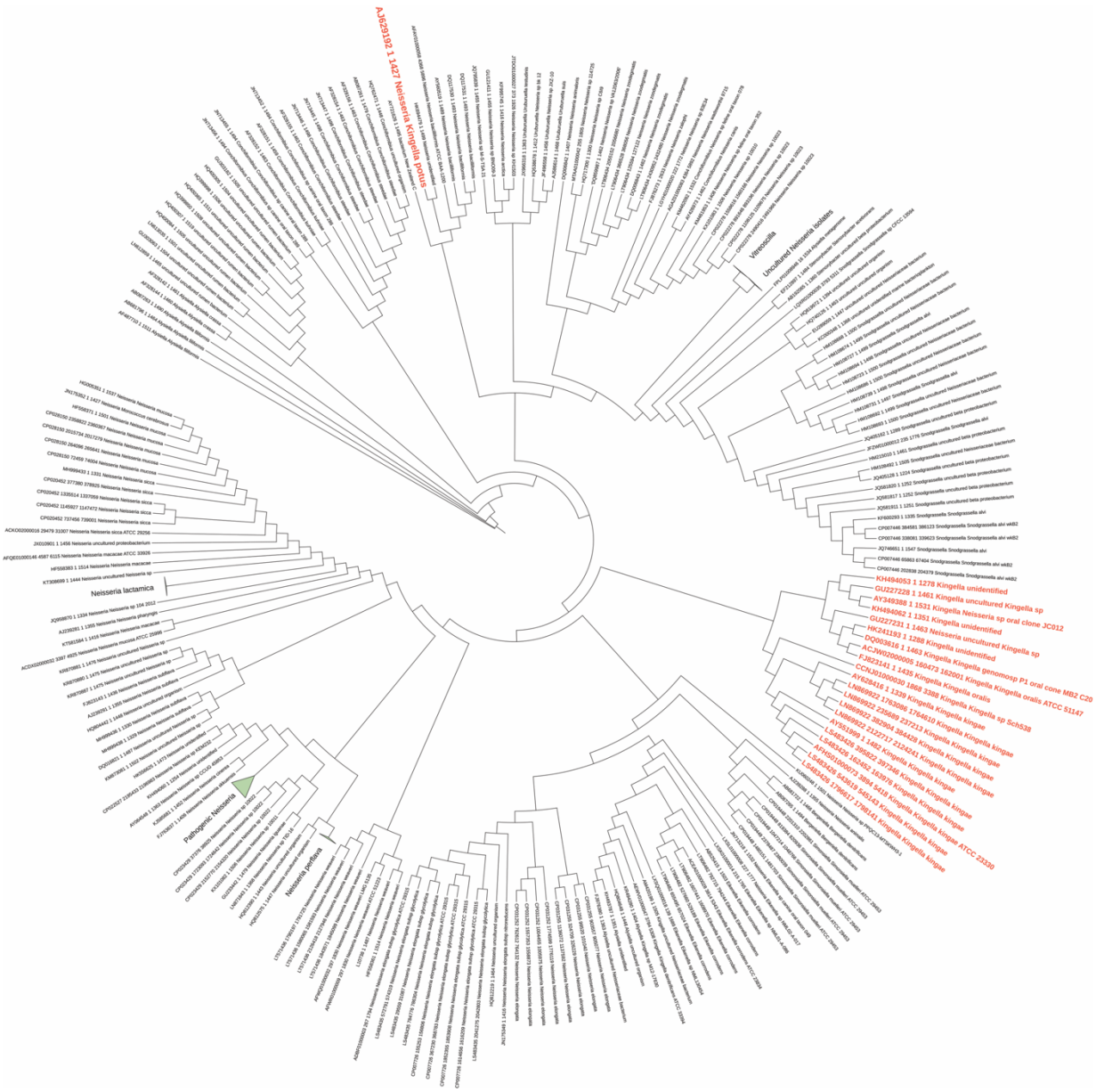
532

533 **Table 4. COGs responsible for the unification of *Kingella s.s.* in the accessory genome.**

	COG	State	p_0[1]	p_1[2]		State	p_0[1]	p_1[2]	COG name[3]	Predicted Function[3]
Last Common Ancestor *Alysiella, Simonsiella,* and *Kingella s.s.*	742	1	0.15414	0.84586	Last Common Ancestor of *Kingella s.s.*	0	0.98197	0.01803	group_2532	Protein IscX
	766	1	0.02993	0.97007		0	0.97924	0.02076	group_2994	hypothetical protein
	771	1	0.02993	0.97007		0	0.97924	0.02076	*purU*	hypothetical protein
	923	1	0.24673	0.75327		0	0.97821	0.02179	group_3650	Sulfur carrier protein TusA
	3159	1	0.06887	0.93113		0	0.96883	0.03117	group_6852	Isocitrate lyase
	3277	1	0.35969	0.64031		0	0.96819	0.03181	*thiG*	Thiazole synthase
	3284	1	0.35969	0.64031		0	0.96819	0.03181	group_5996	Imidazole glycerol phosphate synthase subunit HisH
	3586	1	0.13975	0.86025		0	0.95545	0.04455	group_5655	Isocitrate dehydrogenase [NADP]

534
535 1 – Maximum likelihood posterior probability that this COG was absent at this site in the inferred
536 ancestral reconstruction. Genes were called as absent if p_0 >= 0.95.
537 2 - Maximum likelihood posterior probability that this COG was present at this site in the inferred
538 ancestral reconstruction. Genes were called as present if p_0 >= 0.95.
539 3 – Predicted names and functions of COGs of interest were recovered from the associated Roary
540 matrix. Sequential COGs are not necessarily syntenic in any given genome.
541
542

543 **Figures and Legends.**



544

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

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

547 likelihood phylogeny was constructed using RAxML. Species that belong a species in the genus

548 *Kingella* are colored red. By 16S rRNA relatedness, *Kingella* forms a polyphyletic clade with *K.*

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

550 *Neisseria bacilliformis*. Bootstrap values greater than 70% are shown.

551

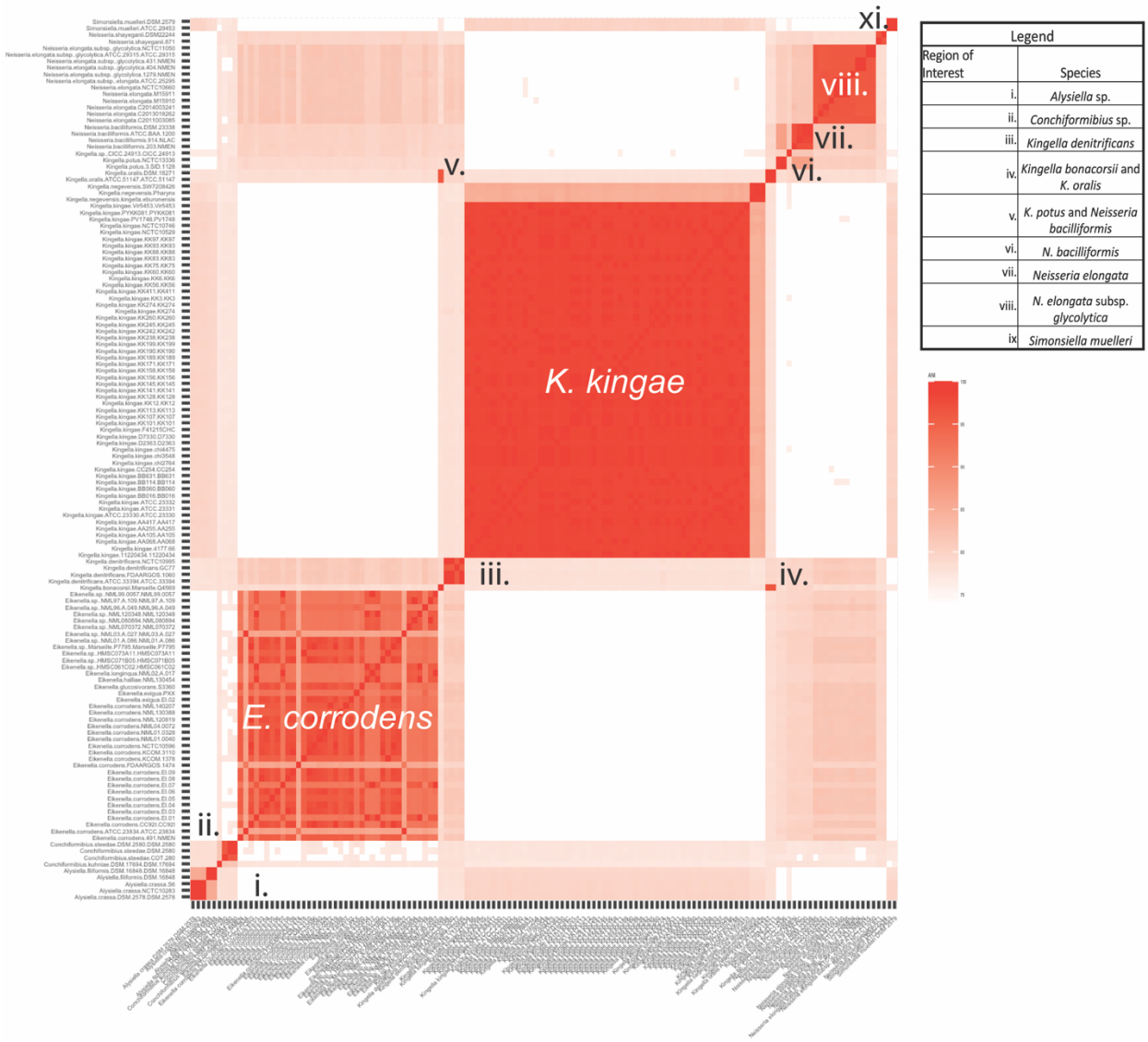
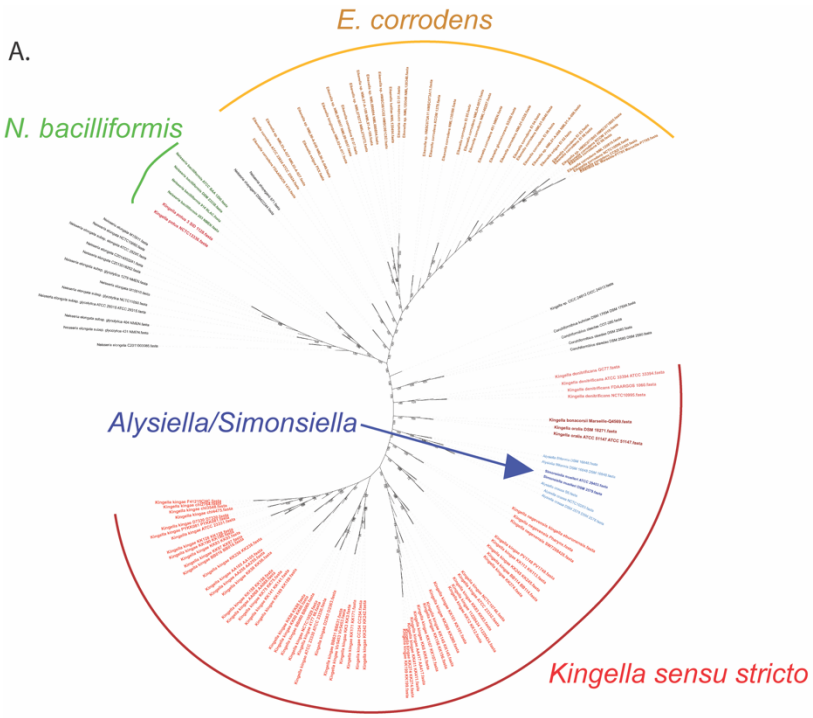


Figure 2. ANI scores comparing sequenced isolates from species in the *Neisseriaceae* most closely related to the genus *Kingella*. ANI scores were calculated between publicly available genomes downloaded from NCBI using FastANI. Several regions of interest are indicated on the heatmap, including the species delimiters, and regions of interest comparing *K. bonacorsii* and *K. oralis* (iv) and *K. potus* and *N. bacilliformis* (v). The resulting ANI matrix was rendered as a heatmap using R.



B.

Robinson-Foulds distance

Roary Cutoff Scores	95	90	85	80	75
95	0	0.133	0.183	0.166	0.270
90	0.133	0	0.128	0.151	0.263
85	0.183	0.128	0	0.174	0.271
80	0.166	0.151	0.174	0	0.232
75	0.270	0.263	0.271	0.232	0

C.

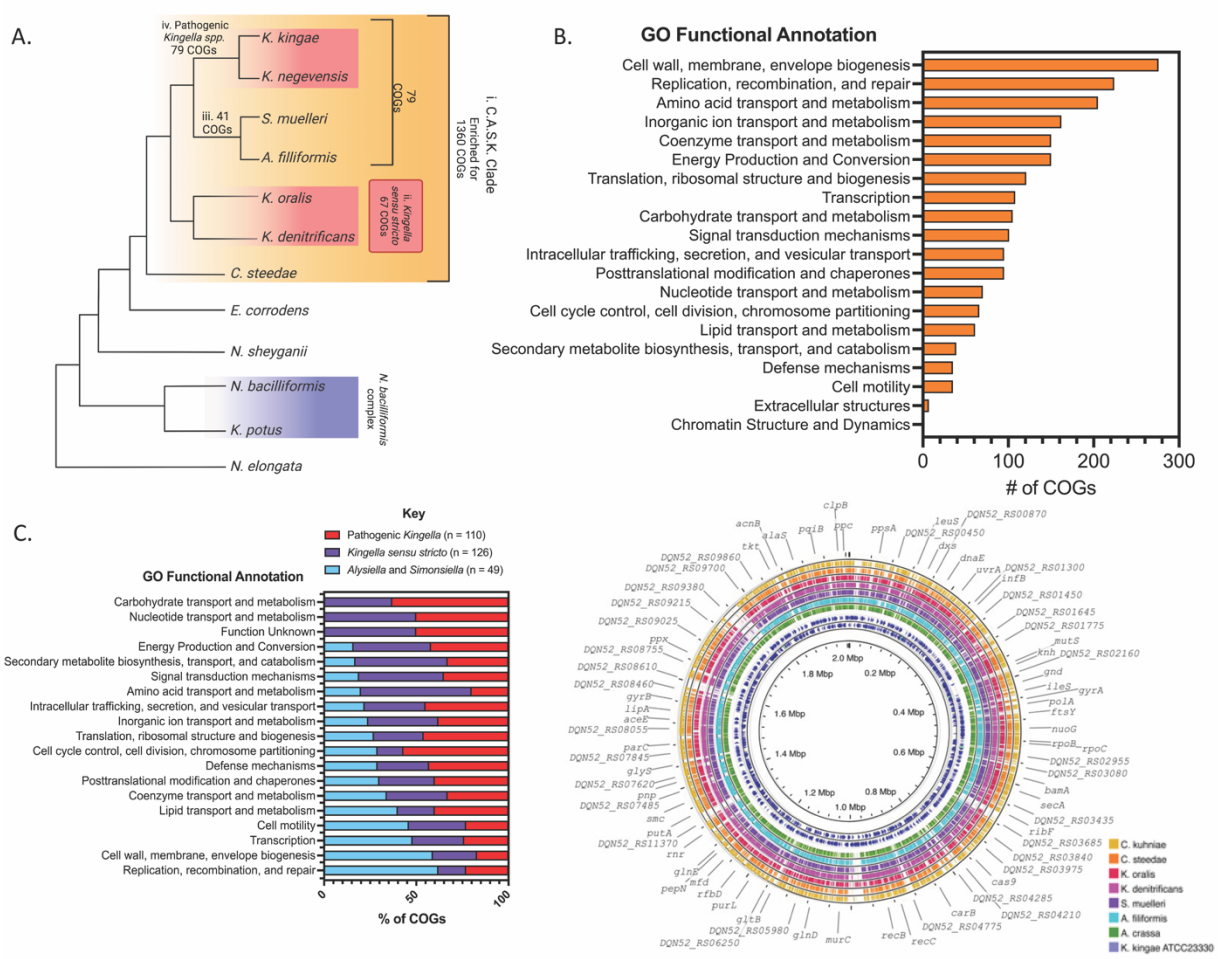
Normalized Nye Similarity Score

Roary Cutoff Scores	95	90	85	80	75
95	1	0.8685	0.8182	0.8271	0.7307
90	0.8685	1	0.8714	0.8282	0.7407
85	0.8182	0.8714	1	0.8216	0.7384
80	0.8271	0.8282	0.8216	1	0.7685
75	0.7307	0.7407	0.7384	0.7685	1

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 homology cutoff of 75%. The core genomes, which encompass approximately two million sites, were aligned with MAFFT and used to reconstruct a maximum likelihood phylogeny. As above, *K. potus* is only distantly related to the rest of the species in the genus *Kingella.* Additionally, the genera *Alysiella* and *Simonsiella* fall within the *Kingella* clade, suggesting the core genome of the species in each of these genera is very closely related to that of *Kingella.* Bootstraps greater than 70% are shown. **B-C.** The Robinson-Foulds and Normalized Nye similarity score congruency metrics were used to quantify changes in tree architecture between phylogenies reconstructed using core genome alignments. The Nye similarity score was normalized against the number of possible rearrangements given the number of leaves in the tree.

Figure 4. 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 50% with COGs present in at least 80% of the species. The core genomes, which encompass approximately two million

577 sites, were aligned with MAFFT and used to reconstruct a maximum likelihood phylogeny. As
578 above, *K. potus* is only distantly related to the rest of the species in the genus *Kingella.*
579 Additionally, the genera *Alysiella* and *Simonsiella* fall within the *Kingella* clade, suggesting the
580 core genome of the species in each of these genera is very closely related to that of *Kingella.*
581 Bootstraps greater than 70% are shown. The phylogeny is rooted by the *Neisseria elongata* clade.
582 **B.** A binary gene presence/absence matrix was generated using Roary at the 75% similarity cut-
583 off, and includes both the core and accessory genes of all species. A phylogenetic tree based on
584 this matrix was constructed using RAxML via the GAMMA model of rate heterogeneity. This
585 analysis, based solely on gene content unites *Kingella sensu stricto* into a monophyletic group.
586 Bootstraps greater than 70% are shown. The phylogeny is rooted by the *Neisseria elongata* clade.
587



588

589 **Figure 5. A**. Scoary was used to generate lists of genes differentially encoded genes comparing

590 (i) CASK species against the *Neisseriaceae*, (ii) *Kingella sensu stricto* (iii) *Alysiella* and *Simonsiella*

591 or (iii) pathogenic *Kingella* against the CASK clade. Finally, *Alysiella* and *Simonsiella* COGs were

592 compared against the pathogenic *Kingella.* Diagnostic COGs are defined as COGs with 100%

593 sensitivity and specificity to the in-group, and filtered to $p < 0.001$. Please see Supplemental

594 tables 5-9 list for full COGs sets. **B-C.** Diagnostic gene sets for each group were annotated using

595 eggNOG mapper, and classified by Gene Ontology functional groups using the NCBI COG

596 database. B. The CASK clade is enriched for unique COGs with putative roles in cell wall,

597 membrane and envelope biogenesis, as well as DNA replication, recombination and repair. COGs

598 with no functional annotations were omitted from this graphic (n = 656 COGs). **C.** Similarly, the
599 majority of the unique COGs present in *Alysiella* and *Simonsiella* may function as part of DNA
600 replication, recombination and repair pathways and cell wall, membrane and envelope
601 biogenesis. *Kingella sensu stricto* and the pathogenic *Kingella* species encode more unique COGs
602 implicated in carbohydrate, amino acid, or nucleotide transport and metabolism. **D.** Type strains
603 of each species in the CASK clade was compared to the *K. kingae* ATCC 23330 genome (inner
604 circle) using BLASTn. Bars in each ring represent regions of sequence similarity between
605 genomes. Each of the CASK species shows significant sequence similarity to nearly all of the *K.*
606 *kingae* reference, further demonstrating how closely related each of these species are. Regions
607 of dissimilarly map to regions that encode putative MGEs, particularly prophages.
608