Description of two novel *Corynebacterium* species isolated from human nasal passages and skin

1.1   Author names

Elena B. Popowitch, BS, MHS[1,*]

Tommy H. Tran, BA[2,*]

Isabel Fernandez Escapa, PhD[2]

Eeshta Bhatt[3]

Alp K. Sozat[3]

Nashwa Ahmed[4]

Clayton Deming[4]

Ari Q. Roberts, MS[2]

NISC Comparative Sequencing Program[5]

Julia A. Segre, PhD[4]

Heidi H. Kong, MD, MHSc[6]

Sean Conlan, PhD[4]

Katherine P. Lemon, MD, PhD[2,7,*]

Matthew S. Kelly, MD, MPH[1,*]

1.2   Affiliation

[1]Division of Pediatric Infectious Diseases, Duke University School of Medicine, Durham, NC, USA

[2]Alkek Center for Metagenomics & Microbiome Research, Department of Molecular Virology &

Microbiology, Baylor College of Medicine, Houston, Texas, USA

[3]Duke University, Durham, NC, USA

[4]National Human Genome Research Institute, Bethesda, MD, USA

26   [5]National Institutes of Health Intramural Sequencing Center, National Human Genome Research

27   Institute, Rockville, MD, USA

28   [6]National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, MD, USA

29   [7]Division of Infectious Diseases, Texas Children's Hospital, Department of Pediatrics, Baylor College

30   of Medicine, Houston, Texas, USA

31   [*]These authors contributed equally to this work

32

1.3   Corresponding authors

34   Katherine P. Lemon, MD, PhD (katherine.lemon@bcm.edu)

35   Matthew S. Kelly, MD, MPH (matthew.kelly@duke.edu)

36

1.4   Keywords

38   *Corynebacterium*; upper respiratory tract microbiome; nasal microbiota; skin microbiota; whole-

39   genome sequencing

40

1.5   Repositories

42   The sequencing files supporting the conclusions of this study are available in the Sequence Read

43   Archive (PRJNA804245, PRJNA854648, PRJNA842433). The partial 16S ribosomal RNA gene

44   sequences from PCR amplification and Sanger sequencing are available in GenBank for

45   *Corynebacterium hallux* sp. nov. CTNIH22[T] (accession number: PQ252679) and *Corynebacterium*

46   *nasorum* sp. nov. KPL3804[T] (accession number: PQ149068). The annotated genomic sequences for

47   the strains characterized in this study have been deposited in GenBank with the following accession

48   numbers: *C. hallux* sp. nov. CTNIH22[T] (GCF_032821755.1), *C. nasorum* sp. nov. KPL3804[T]

49   (GCF_037908315.1), *C. nasorum* sp. nov. MSK185 (GCF_030229765.1), *C. yonathiae* KPL2619

50   (GCF_037908465.1), and *C. yonathiae* MSK136 (GCF_022288805.2).

Abstract

Strains of two novel *Corynebacterium* species were cultured from samples of human nostrils and skin collected in the United States and Botswana. These strains demonstrated growth on Columbia Colistin-Nalidixic Acid agar with 5% sheep blood and in liquid media (brain heart infusion and tryptic soy broth) supplemented with Tween 80, a source of the fatty acid oleic acid. Cells were Gram-positive, non-spore-forming, non-motile bacilli that showed catalase but not oxidase activity. Major fatty acids in both of these species were 18:1 ω9c (oleic acid), 16:0 (palmitic acid), and 18:0 (stearic acid). Analysis of the 16S ribosomal RNA gene sequences identified these strains as belonging to the genus *Corynebacterium* (family Corynebacteriaceae). Whole-genome sequencing revealed that these strains formed distinct branches on a phylogenomic tree, with *C. tuberculostearicum* being the closest relative but with average nucleotide identities of < 95% relative to all previously described species. These results indicate that these strains represent novel species of *Corynebacterium*, for which we propose the names *Corynebacterium hallux* sp. nov., with the type strain CTNIH22^T (=ATCC TSD-435^T=DSM 117774^T), and *Corynebacterium nasorum* sp. nov., with the type strain KPL3804^T (=ATCC TSD-439^T=DSM 117767^T). We also describe the characteristics of two strains isolated from human nasal passages that are members of the recently named species *Corynebacterium yonathiae*.

Introduction

The genus *Corynebacterium* belongs to the family *Corynebacteriaceae* and includes more than 150 validly published species. Most *Corynebacterium* species are only rarely associated with disease among humans and animals. Common pathogenic members include *C. diphtheriae*, the causative agent of the human disease diphtheria, and *C. pseudotuberculosis* and *C. ulcerans*, which are frequent causes of zoonotic infections. Historically, species of the genus *Corynebacterium* have been

76   differentiated based on their host, ecological niche, biochemical characteristics, spectrometric

77   analyses [e.g., matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-

78   TOF MS)], or sequencing of specific genetic loci [1-3]. With regard to the latter, phylogenies based

79   only on the 16S ribosomal RNA (rRNA) gene often have poor support within this genus, with

80   improved results for phylogenies based on full or partial *rpoB* gene sequences [4]. Multi-locus

81   sequence typing of housekeeping genes is frequently used to characterize common pathogens such

82   as *C. diphtheriae* [5-8]. As a result of decreasing sequencing costs and improved tools for genomic

83   analysis, whole-genome sequencing is increasingly being performed for taxonomic classification of

84   *Corynebacterium* strains.

85

86   *C. tuberculostearicum* is a lipid-requiring species that is a common inhabitant of human skin. This

87   species was first described in 1984 when Brown and colleagues identified 16 strains that had

88   biochemical properties that distinguished them from previously described members of the

89   *Corynebacterium* genus [9]. They called this novel species *C. tuberculostearicum* because strains

90   were noted to contain tuberculostearic acid on fatty acid profiling [9]. In 2004, Feurer and colleagues

91   emended the description of *C. tuberculostearicum* and formally proposed it as a new species [10]. In

92   the present study, we identified several *Corynebacterium* strains from human nasal and skin samples

93   that are most closely related to *C. tuberculostearicum*, but that represent distinct species based on

94   their biological properties, chemical structures, and genomic sequences. We propose classification

95   of these strains into novel species *Corynebacterium hallux* sp. nov. ("hallux" referring to the

96   innermost toe) and *C. nasorum* sp. nov. ("nasorum" referring to "of noses") to reflect the ecological

97   niches from which these strains were isolated. Finally, we provide a detailed characterization of two

98   additional strains isolated from human nasal passages that are members of the recently described

99   species *C. yonathiae* [11].

100

Isolation and Ecology

102

103   The new *Corynebacterium* strains described in this study are as follows by species: *C. hallux* sp. nov.

104   (CTNIH22[T]=ATCC TSD-435[T]=DSM 117774[T]), *C. nasorum* sp. nov. (KPL3804[T]=ATCC TSD-439[T]=DSM

105   117767[T], MSK185), and *C. yonathiae* (KPL2619, MSK136). *C. hallux* sp. nov. strain CTNIH22[T] was

106   isolated from the toe web of a healthy adult volunteer in 2019. This sample was collected using an

107   ESwab (Copan, Murrieta, CA), placed in liquid Amies transport medium, and grown on brain heart

108   infusion (BHI) agar with 10% Tween 80 at 35°C under aerobic conditions [12]. The KPL strains of *C.*

109   *nasorum* and *C. yonathiae* were isolated from nasal samples collected from a child and an adult

110   participating in scientific outreach events in Massachusetts in 2017 and 2018, respectively [13].

111   These samples were inoculated onto plates containing either BBL Columbia Colistin-Nalidixic Acid

112   (CNA) agar with 5% sheep blood or BHI agar with 1% Tween 80 and 25 µg/mL fosfomycin. Cultures

113   were incubated aerobically for 48 hours at 37°C in either atmospheric conditions or 5% carbon

114   dioxide (CO_2). For suspected *Corynebacterium*, Sanger sequencing was performed on a colony-

115   polymerase chain reaction (PCR) amplicon of the V1-V3 region of the 16S rRNA gene (primers 27F

116   and 519R). The MSK strains of *C. nasorum* and *C. yonathiae* were isolated from nasopharyngeal

117   samples collected from infants enrolled in a prospective cohort study that was conducted in

118   Gaborone, Botswana between February 2016 and January 2021 [14]. These samples were inoculated

119   onto plates containing Columbia CNA agar with 5% sheep blood, BHI agar supplemented with 50

120   µg/mL fosfomycin, and BHI agar with 1% Tween 80 and 50 µg/mL fosfomycin, and incubated

121   aerobically at 37°C in a 5% CO_2-enriched environment for 48 hours. Preliminary identification of

122   suspected *Corynebacterium* was performed using MALDI-TOF MS or Sanger sequencing on a colony-

123   PCR amplicon of the V1-V3 region of the 16S rRNA gene (primers 27F and 534R). The type strains of

124   *C. accolens* (ATCC 49725[T]) [15], *C. macginleyi* (ATCC 51787[T]) [16], and *C. tuberculostearicum* (ATCC

125   35692[T]) [10] were obtained from the American Type Culture Collection (Manassas, Virginia).

126

127 **Genome Features**

128

129 Draft genomes of *C. hallux* sp. nov. CTNIH22[T] and *C. nasorum* sp. nov. KPL3804[T] were generated

130 through assembly of Illumina sequencing reads, as described previously [12, 17]. Based on four

131 strain genomes, *C. hallux* sp. nov. has an average G+C content of 58.5 mol% and an estimated

132 average genome length of 2.49 Mb, with between 2,293 and 2,427 coding sequences. Based on 13

133 strain genomes, *C. nasorum* sp. nov. has an average G+C content of 58.5 mol% and an estimated

134 average genome length of 2.46 Mb, with between 2,303 and 2,451 coding sequences (**Table S1**).

135

16S rRNA Gene Phylogeny

137

138 PCR amplification of the V1-V9 regions of the 16S rRNA gene of *C. hallux* sp. nov. CTNIH22[T] and *C.*

139 *nasorum* sp. nov. KPL3804[T] was conducted by ACGT, Inc. (Wheeling, IL) and Azenta Life Sciences

140 (South Plainfield, NJ), respectively. Enzymatic cleanup of the PCR products was performed before

141 bidirectional, dye-terminator sequencing on a 3730xl DNA Analyzer (Applied Biosystems, Waltham,

142 MA). For both strains, the corresponding portions of the genome assembly-extracted 16S rRNA gene

143 sequence were determined to be 100% identical to the near-complete gene sequence obtained from

144 PCR amplification and Sanger sequencing. Thus, the full-length genome assembly-extracted 16S

145 rRNA gene sequences were used for subsequent phylogenetic analyses (**Figures 1** and **S1A**).

146

147 We note that we initially assigned strains of *C. hallux* sp. nov. to the species *C. tuberculostearicum*

148 based on their 16S rRNA gene sequences, and denoted these as ribotype B strains based on

149 differences with the 16S rRNA gene sequences of other known *C. tuberculostearicum* strains [12].

150 We subsequently used the Type (Strain) Genome Server to determine that genome-sequenced

151 strains of *C. hallux* sp. nov. were most closely related to *C. tuberculostearicum*, although with

152 average nucleotide identity calculations based on the BLAST+ algorithm (ANIb) values of <95% in

153   comparisons to *C. tuberculostearicum* reference genomes [18]. Similarly, we used the Genome

154   Taxonomy Database Toolkit (GTDB-Tk) to determine that *C. nasorum* sp. nov. was closely related to

155   *C. tuberculostearicum* [19]. Also, because the 16S rRNA gene sequences of *C. tuberculostearicum*

156   strain ATCC 35692[T] and the proposed *C. nasorum* sp. nov. strains are 99.9% identical, it is highly

157   probable that *C. nasorum* sp. nov. sequences were misassigned to *C. tuberculostearicum* in past 16S

158   rRNA gene-based microbiome studies.

159

160   A 16S rRNA gene maximum-likelihood phylogeny was constructed using the following species: 1) all

161   of the validly named hits with ≥ 95.7% identity to the 16S rRNA gene of *C. hallux* sp. nov. CTNIH22[T]

162   and/or *C. nasorum* sp. nov. KPL3804[T] using the 16S-based ID service on EZBioCloud [20]; 2)

163   additional species that were closely related based on genomes in GTDB-Tk; and 3) *Mycobacterium*

164   *tuberculosis*[T] as an outgroup (**Figure 1**). In addition, a larger unrooted 16S rRNA gene phylogeny was

165   constructed to set the two proposed novel species in a broader context within the genus

166   *Corynebacterium* (**Figure S1A**). The 16S rRNA gene-based phylogenies had a large number of poorly

167   supported branches based on ultrafast bootstrap values (**Figures 1** and **S1A**) [21]. This was expected

168   given that the inadequacy of using the 16S rRNA gene alone for constructing reliable phylogenies

169   within the genus *Corynebacterium* is well described [4]. Among *Corynebacterium* species, the *rpoB*

170   gene has more polymorphisms [4], and a phylogeny based on the *rpoB* gene (**Figure S1B**) had a

171   branching pattern with higher support than the 16S rRNA gene-based phylogeny (**Figure S1A**).

172

Average Nucleotide Identity

174

175   ANIb calculations performed using the Python package pyani v0.2.9 [22, 23] indicated that the

176   genome sequences for the new strains and genomes described in this study were < 95% identical to

177   the type strain of *C. tuberculostearicum,* and to other closely related species (**Figure 2**) [10, 11]. In

178   general, an ANIb threshold of 95-96% accurately represents the boundary between prokaryotic

179   species [24]. Strains of the proposed *C. nasorum* sp. nov. had ANIb values above 95% in comparisons

180   to the genome currently called '*Corynebacterium kefirresidentii*' (**Figure 2**, purple box) [25].

181   However, we were unable to find a type strain bearing this name listed in the publicly available

182   catalogs of major strain repositories. In metagenomic analyses, Kalan and colleagues demonstrate

183   genomic material mapping to the genome called '*C. kefirresidentii*' is found on human skin but has

184   higher prevalence and relative abundance in human nasal samples [26]. Based on these findings, and

185   the isolation of a number of nasal strains with ANIb values of >95% to this genome by our

186   laboratories, we agree with the assertion by Kalan and colleagues that the human nasal passages are

187   one of the primary habitats of this species. To reflect this, we propose the species name

188   *Corynebacterium nasorum* sp. nov.. We further propose that strains and genomes previously

189   classified as '*C. kefirresidentii*' belong to this proposed novel species (**Table S2**).

190

Phylogenomic Analysis

192

193   A maximum-likelihood phylogenomic tree including all of the genome-sequenced strains from Figure

194   1 with *Mycobacterium tuberculosis* as an outgroup (**Figure 3**) and a phylogenomic tree including the

195   68 *Corynebacterium* species from Figure S1A (**Figure S1C**) each showed that both *C. hallux* sp. nov.

196   and *C. nasorum* sp. nov. belong to a larger clade that includes *C. tuberculostearicum* and the recently

197   named species *C. curieae*, *C. marquesiae,* and *C. yonathiae* [10, 11]. Kalan and colleagues refer to

198   this monophyletic clade as the "*C. tuberculostearicum* species complex," and it is most closely

199   related to the clade containing *C. accolens* and *C. macginleyi* (**Figures 3** and **S1C**) [26]. Of note, the

200   genomes currently labeled in GTDB-Tk [19] as "*C. aurimucosum*_E" are misassigned at the species

201   level based on the phylogenomic analyses shown here and those performed by Kalan and colleagues

202   [26], since *C. aurimucosum*_E 620_CAUR [27] clusters far from the genome for the type strain *C.*

203   *aurimucosum* DSM 44827 (**Figures 3** and **S1C**). Based on an ANIb threshold of 95%, genomes labeled

204   in GTDB-Tk as "*C. aurimucosum*_E" assign to the recently named species *C. marquesiae* (**Figures 2**

205   and **S2A**) [11].

206

207   Together, ANIb calculations and the phylogenomic trees confirm that the strains labeled *C. hallux* sp.

208   nov. and *C. nasorum* sp. nov. represent novel species belonging to the genus *Corynebacterium*.

209

Comparative Genomic Analysis

211

212   The metabolic capabilities of more divergent *Corynebacterium* species sharing the common habitat

213   of the human nasal passages are highly conserved [28]. Although members of the *C.*

214   *tuberculostearicum* species complex are known to inhabit different human body site habitats

215   including skin [12], the nasal passages [26], and the female urinary tract [11], we hypothesized that

216   these would exhibit conserved metabolic capabilities based on their close phylogenetic relationship

217   to each other (**Figures 3, S1C,** and **S2A**). Indeed, metabolic estimation on genomes of these species

218   using the anvi-run-kegg-kofams and the anvi-estimate-metabolism programs of anvi'o v8 [29, 30],

219   which rely on Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic annotations [31],

220   revealed largely shared metabolic capabilities with some strain-level variation within specific species

221   (**Figure 4**, **Table S3**; see https://klemonlab.github.io/NovCor_Manuscript/Methods_Anvio.html for

222   detailed methods). This analysis estimated these 30 strain genomes covering six species (**Figures 2**

223   and **S2**) all shared 48 stepwise complete KEGG modules, with most strains also sharing an additional

224   six complete KEGG modules (**Table 1**). These included many modules for amino acid biosynthesis,

225   which is typical of *Corynebacterium* species. We estimated that all 30 strain genomes also encode a

226   complete tricarboxylic acid cycle, consistent with their preference for aerobic growth, along with a

227   number of other modules involved in central carbohydrate metabolism (**Table 1**).

Phenotypic and Chemotaxonomic Characterisation

Growth of strains included in this study was first determined on the following solid media: tryptic soy agar (TSA) with 5% sheep blood (Remel, Lenexa, KS), BHI agar (Becton Dickinson, Franklin Lakes, NJ), and BHI with 1% Tween 80 (MilliporeSigma, Burlington, MA). Isolates were suspended in sterile phosphate buffered saline (Genesee Scientific, El Cajon, CA) to an OD_{600} of 0.10–0.15 and plated in a quadrant pattern using a sterile 10-µl loop. Growth was judged by two authors (EBP, MSK) based on the size and density of colonies on these plates. Given that all strains grew well on BHI with 1% Tween 80 plates, growth on this medium was further evaluated for up to 14 days at various temperature (4°C, 20°C, 30°C, 37°C, 42°C, 50°C) and atmospheric (aerobic/5% CO_2, microaerophilic, anaerobic) conditions. Microaerophilic and anaerobic conditions were generated using the AnaeroPack system with MicroAero and Anaerobic gas generators (Thermo Fisher Scientific, Waltham, MA). Growth of all strains was observed at temperatures between 30°C and 42°C, with *C. accolens* ATCC 49725[T], *C. tuberculostearicum* ATCC 35692[T], *C. hallux* sp. nov. CTNIH22[T], and strains of *C. yonathiae* additionally demonstrating growth at 20°C (**Table 2**). Growth was also observed for all strains in microaerophilic conditions, with *C. accolens* ATCC 49725[T], *C. macginleyi* ATCC 51787[T], and strains of *C. nasorum* sp. nov. and *C. yonathiae* additionally demonstrating weak growth in anaerobic conditions (**Table 2**).

C. hallux sp. nov. CTNIH22[T] typically grew as creamy, white colonies measuring 3-5 mm in diameter on BHI with 1% Tween 80 agar; growth was weaker on TSA with 5% sheep blood or on BHI agar without Tween 80, with colonies measuring 1-2 mm in diameter that were non-hemolytic on blood-containing agar. *C. nasorum* sp. nov. strains grew optimally on BHI with 1% Tween 80 agar, yielding large white colonies 5-10 mm in diameter. Colonies of *C. nasorum* sp. nov. were non-hemolytic on

253   TSA with 5% sheep blood agar. *C. yonathiae* strains appeared as raised, creamy colonies between 5-

254   10 mm in diameter when grown on BHI with 1% Tween 80 agar. Colonies on TSA with 5% sheep

255   blood were flat, translucent, non-hemolytic, and approximately 2-3 mm in diameter.

256

257   Growth in liquid media was assessed using BHI broth, BHI broth with 0.2% Tween 80, tryptic soy

258   broth (TSB), and TSB with 0.2% Tween 80. For these assays, culture tubes were inoculated with

259   bacterial cells washed twice with phosphate buffered saline to remove traces of Tween 80 retained

260   from the solid media, then resuspended to an OD_{600} of 0.10–0.15. All cultures reached an OD_{600} at or

261   above 2.0 at 48 hours after inoculation into BHI broth with 0.2% Tween 80. Therefore, this liquid

262   medium was used for subsequent assays testing for growth at varying pH and (2.0, 4.0, 6.0, 7.0, 8.0,

263   10.0, 12.0) and salinity (0%, 3%, 5%, 7%, 10%, 14%, 20%). Liquid culture tubes were incubated

264   aerobically at 37°C with shaking at 200 rpm for 72 hours or until the cultures exceeded an OD_{600} of

265   2.0. The pH range for growth of most strains was 7–8, with optimal growth observed at pH 8 (**Table**

266   **2**). Growth of all strains was observed at salinity at or below 10%, with *C. accolens* ATCC 49725[T] and

267   *C. hallux* sp. nov. CTNIH22[T] demonstrating growth at salinity up to 14% (**Table 2**). All strains also

268   demonstrated growth in TSB with 0.2% Tween 80. No growth was observed in BHI or TSB broth in

269   the absence of Tween 80.

270

271   Gram stains were performed using fresh cultures grown on BHI with 1% Tween 80 agar and using a

272   commercial kit (Hardy Diagnostics, Santa Maria, CA). Cells of *C. hallux* sp. nov. CTNIH[T] and strains of

273   *C. nasorum* sp. nov. and *C. yonathiae* were Gram-positive, non-spore-forming irregular rods or

274   coccoid. For visualization by scanning electron microscopy, cells were fixed with a solution

275   containing 2% glutaraldehyde and 4% formaldehyde prior to transfer to the Duke University Shared

276   Materials Instrumentation Facility. In scanning electron microscopy images, *C. hallux* sp. nov. cells

277   **Figure 5A**) appeared as pleomorphic rods to coccoid, with most cells being 0.6-2.0 µm long and 0.4-

278   0.6 µm wide. *C. nasorum* sp. nov. cells (**Figure 5B**) similarly appeared as heterogenous rods to

279 coccoid, with most cells measuring between 0.6-2.2 µm long and 0.4-0.6 µm wide. *C. yonathiae* cells

280 (**Figure 5C**) were coccoid to elongated rods measuring up to 5 µm in length and 0.4-0.6 µm wide.

281

282 Enzymatic (including catalase) and fermentation activities were tested using API CORYNE strips

283 (bioMérieux, Marcy-l`Etoile, France) according to the manufacturer's instructions. Oxidase testing

284 was performed using OxiStrips (Hardy Diagnostics, Santa Maria, CA) according to the package insert.

285 Motility was assessed by stabbing culture tubes containing BHI with 1% Tween 80 and 0.5% agar

286 with a fresh culture of each *Corynebacterium* strain; *Pseudomonas aeruginosa, Escherichia coli*, and

287 *Staphylococcus aureus* were used as comparators for this assay. All strains were catalase-positive,

288 oxidase-negative, and non-motile (**Table 3**). On biochemical testing, positive reactions for alkaline

289 phosphatase were observed for *C. macginleyi* ATCC 51787^T, *C. hallux* sp. nov. CTNIH22^T, and strains

290 of *C. nasorum* sp. nov. and *C. yonathiae* (**Table 3**). *C. hallux* sp. nov. CTNIH22^T and strains of *C.*

291 *yonathiae* sp. nov. had a positive reaction for pyrrolidonyl arylamidase, while this activity was

292 variable for strains of *C. nasorum* sp. nov. Nitrate reduction was observed for *C. accolens* ATCC

293 49725^T and *C. macginleyi* ATCC 51787^T, while strains of *C. nasorum* sp. nov. had a weakly positive

294 reaction for pyrazinamidase. Fermentation of D-glucose and D-ribose was noted for strains *C.*

295 *accolens* ATCC 49725^T and *C. macginleyi* ATCC 51787^T, while *C. macginleyi* ATCC 51787^T additionally

296 fermented D-mannitol and D-saccharose (**Table 3**). No carbohydrate fermentation was noted for *C.*

297 *hallux* sp. nov. CTNIH22^T or strains of *C. nasorum* sp. nov. or *C. yonathiae*.

298

299 For fatty acid analysis, cells of all strains were harvested from the same culture conditions during the

300 late log phase (at 37 °C in a 5% CO2-enriched environment for 2 days on TSA with 5% sheep blood

301 agar). Fatty acids were extracted from cells using the standard midi protocol (Sherlock Microbial

302 Identification System, v6.0B), analysed with a gas chromatograph (6890 Series GC System, Hewlett

303 Packard), and identified using the TSBA6 database of the Microbial Identification System [32]. The

304 cellular fatty acid profiles of all strains included saturated, unsaturated, and branched-chain fatty

305   acids (**Table 4**). The major fatty acids identified in *C. tuberculostearicum* ATCC 35692[T], *C. accolens*

306   ATCC 49725[T], *C. hallux* sp. nov. CTNIH22[T], and strains of *C. nasorum* sp. nov. and *C. yonathiae* were

307   18:1 ω9c (oleic acid), 16:0 (palmitic acid), and 18:0 (stearic acid). *C. hallux* sp. nov. CTNIH22[T] had a

308   higher amount of 18:1 ω9c (oleic acid) than other strains tested, and this fatty acid was absent from

309   the composition of *C. macginleyi* ATCC 51787[T]. Several fatty acids were uniquely present in lesser

310   amounts in *C. hallux* sp. nov. CTNIH22[T] (e.g., 12:0, 15:1 ω8c, 20:1 ω9c).

311

Description of *Corynebacterium hallux* sp. nov.

313

314   *Corynebacterium hallux* sp. nov. (hal'lux. N.L. neut. n. *hallux* referring to the innermost toe, the skin

315   site representing the source of this isolate).

316

317   Cells of *C. hallux* sp. nov. CTNIH22[T] are Gram-positive, catalase-positive, oxidase-negative, non-

318   spore-forming, non-motile bacilli (0.6-2.0 μm long and 0.4-0.6 μm wide). Optimal growth on solid

319   medium was observed on BHI with 1% Tween 80 agar with aerobic incubation at 37 °C in a 5% CO_2-

320   enriched environment. Colonies on this medium were creamy white and measure approximately 3-5

321   mm in diameter; growth is weaker on TSA with 5% sheep blood agar, with non-hemolytic colonies

322   measuring 1-2 mm in diameter. In liquid culture, *C. hallux* sp. nov. CTNIH22[T] requires the addition of

323   0.2% Tween 80 for growth in BHI or TSB and tolerates salinity up to 14%. However, it has more

324   stringent requirements for pH, with growth only observed at pH between 7.0 and 8.0. On

325   biochemical testing, a positive reaction is observed for alkaline phosphatase with a weakly positive

326   reaction for pyrrolidonyl arylamidase. No carbohydrate fermentation is noted in testing performed

327   using API CORYNE strips. The major fatty acids identified are oleic (C18:1 ω9c; 32.5%), palmitic

328   (C16:0; mean of 26.9%), and stearic (C18:0; 12.9%) acids. The genome size and DNA G+C content of

329   the type strain are 2.53 Mb and 58.4 mol%, respectively.

330

331 The type strain, CTNIH22^T (=ATCC TSD-435^T=DSM 117774^T), was isolated from the toe web of a

332 healthy adult. The partial 16S rRNA gene sequence of strain CTNIH22^T is available in GenBank

333 (accession number: PQ252679). The GenBank accession number for the genomic sequence of this

334 strain is GCF_032821755.1.

335

336 ## Description of *Corynebacterium nasorum* sp. nov.

337

338 *Corynebacterium nasorum* sp. nov. (nas'or.um L. gen. adj. *nasorum* referring to "of noses", the

339 human body site that is the source of the isolates).

340

341 Cells are Gram-positive, catalase-positive, oxidase-negative, non-spore-forming, non-motile bacilli

342 (0.6-2.2 µm long and 0.4-0.6 µm wide). Optimal growth on solid medium is observed on BHI with 1%

343 Tween 80 agar with aerobic incubation at 37 °C in a 5% CO_2-enriched environment. Colonies on this

344 medium are creamy white and measure 5-10 mm in diameter; growth is weaker on TSA with 5%

345 sheep blood agar with non-hemolytic colonies. Optimal growth of *C. nasorum* sp. nov. in liquid

346 medium is observed in BHI broth with 0.2% Tween 80; growth is also observed in TSB with 0.2%

347 Tween 80. Growth of strains of *C. nasorum* sp. nov. occurs at pH between 6.0 and 8.0 and at salinity

348 up to 10%. On biochemical testing, a positive reaction is observed for alkaline phosphatase, with a

349 weakly positive reaction for pyrazinamidase and variable pyrrolidonyl arylamidase activity. No

350 carbohydrate fermentation is noted in testing performed using API CORYNE strips. The major fatty

351 acids are palmitic (C16:0; mean of 35.0%), oleic (C18:1 ω9c; mean of 19.5%), and stearic (C18:0;

352 mean of 13.9%) acids. The genome size and DNA G+C content of the type strain are 2.46 Mb and

353 58.5 mol%, respectively.

354

355   The type strain, KPL3804[T] (=ATCC TSD-439[T]=DSM 117767[T]), was isolated from a swab of the nostrils

356   of a healthy adult aged between 31 and 60 years in Massachusetts, USA. The partial 16S rRNA gene

357   sequence of strain KPL3804[T] is available in GenBank (accession number: PQ149068). The GenBank

358   accession numbers for the genomic sequences of the *C. nasorum* sp. nov. strains described in this

359   study are GCF_037908315.1 (KPL3804[T]) and GCF_030229765.1 (MSK185).

360

361   ## Description of strains of the recently described species *C. yonathiae*

362

363   Cells are Gram-positive, catalase-positive, oxidase-negative, non-spore-forming, non-motile bacilli

364   (up to 5 μm in length and 0.4-0.6 μm wide). Optimal growth on solid medium is observed on BHI

365   with 1% Tween 80 agar with incubation aerobically at 37 °C in a 5% CO_2-enriched environment.

366   Colonies on this medium are raised, creamy colonies between 5-10 mm in diameter; growth is

367   weaker on TSA with 5% sheep blood agar, with non-hemolytic colonies measuring 2-3 mm in

368   diameter. In liquid culture, strains require the addition of 0.2% Tween 80 for growth in BHI or TSB

369   and tolerate salinity up to 10% and pH between 7.0 and 8.0. On biochemical testing, positive

370   reactions are observed for alkaline phosphatase and pyrrolidonyl arylamidase. No carbohydrate

371   fermentation is noted in testing performed using API CORYNE strips. The major fatty acids are

372   palmitic (C16:0; mean of 37.6%), oleic (C18:1 ω9c; mean of 21.1%), and stearic (C18:0; mean of

373   14.5%) acids. The genome sizes of *C. yonathiae* strains MSK136 and KPL2619 are 2.47 Mb and 2.35

374   Mb, with DNA G+C content of 58.5 and 58.6 mol%, respectively.

AUTHOR STATEMENTS

375

376

1.6   Authors and contributors

377

378   Writing – original draft: E.B.P., M.S.K., K.P.L., T.H.T.

379   Writing – review and editing: E.B.P., E.B., A.K.S., J.A.S., H.H.K., S.C., T.H.T., I.F.E., A.Q.R., K.P.L., M.S.K.

380   Investigation: E.B.P., T.H.T., I.F.E., E.B., A.K.S., N.A., A.Q.R., J.A.S., H.K., S.C., K.P.L., M.S.K.

381   Formal analysis: E.B.P., T.H.T., I.F.E., A.Q.R., M.S.K.

382   Resources: J.A.S., H.H.K., K.P.L., M.S.K.

383   Funding acquisition: M.S.K., K.P.L.

384

1.7   Conflicts of interest

385

386   The authors declare that there are no conflicts of interest.

387

1.8   Funding information

388

395

1.9   Ethical approval

The study protocol for collection of nasopharyngeal samples from infants in Botswana was approved

by the Botswana Ministry of Health, the Princess Marina Hospital ethics committee, and institutional

review boards at the University of Pennsylvania, Children's Hospital of Philadelphia, McMaster

University, and Duke University. Healthy volunteers were sampled as part of a prospective natural

history study at the NIH Clinical Center approved by the NIH Institutional Review Board (

www.clinicaltrials.gov/ct2/show/NCT00605878). The Forsyth Institutional Review Board approved

the protocol (FIRB #17-02) used to collect nasal bacteria KPL strains in Massachusetts.

1.10   Consent for publication

Not applicable

1.11   Acknowledgements

ABBREVIATIONS

412

413

414   ANIb, average nucleotide identity based on BLAST+

415   ATCC, American Type Culture Collection

416   BHI, brain heart infusion

417   CO_2, carbon dioxide

418   CNA, Colistin-Nalidixic Acid

419   FAME, fatty acid methyl ester

420   GTDB-Tk, Genome Taxonomy Database Toolkit

421   KEGG, Kyoto Encyclopedia of Genes and Genomes

422   MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry

423   OD_{600}, optimal density at a wavelength of 600 nanometers

424   PCR, polymerase chain reaction

425   rpm, revolutions per minute

426   rRNA, ribosomal RNA

427   TSA, tryptic soy agar

428   TSB, tryptic soy broth

REFERENCES

1. Bolt, F., et al., *Multilocus sequence typing identifies evidence for recombination and two distinct lineages of Corynebacterium diphtheriae.* Journal of Clinical Microbiology, 2010. **48**(11): p. 4177-4185.

2. Dangel, A., et al., *Geographically diverse clusters of nontoxigenic Corynebacterium diphtheriae infection, Germany, 2016–2017.* Emerging infectious diseases, 2018. **24**(7): p. 1239.

3. Pascual, C., et al., *Phylogenetic analysis of the genus Corynebacterium based on 16S rRNA gene sequences.* International journal of systematic and evolutionary microbiology, 1995. **45**(4): p. 724-728.

4. Khamis, A., D. Raoult, and B. La Scola, *rpoB gene sequencing for identification of Corynebacterium species.* J Clin Microbiol, 2004. **42**(9): p. 3925-31.

5. Khamis, A., D. Raoult, and B. La Scola, *Comparison between rpoB and 16S rRNA gene sequencing for molecular identification of 168 clinical isolates of Corynebacterium.* Journal of clinical microbiology, 2005. **43**(4): p. 1934-1936.

6. Khamis, A., D. Raoult, and B. La Scola, *rpoB gene sequencing for identification of Corynebacterium species.* Journal of clinical microbiology, 2004. **42**(9): p. 3925-3931.

7. Czajka, U., et al., *Changes in MLST profiles and biotypes of Corynebacterium diphtheriae isolates from the diphtheria outbreak period to the period of invasive infections caused by nontoxigenic strains in Poland (1950–2016).* BMC Infectious Diseases, 2018. **18**: p. 1-8.

8. Hoefer, A., et al., *Molecular and epidemiological characterization of toxigenic and nontoxigenic Corynebacterium diphtheriae, Corynebacterium belfantii, Corynebacterium rouxii, and Corynebacterium ulcerans isolates identified in Spain from 2014 to 2019.* Journal of clinical microbiology, 2021. **59**(3): p. 10.1128/jcm. 02410-20.

454   9.   Brown, S., et al. *Description of Corynebacterium tuberculostearicum sp. nov., a leprosy-*

455   *derived Corynebacterium*. in *Annales de l'Institut Pasteur/Microbiologie*. 1984. Elsevier.

456   10.   Feurer, C., et al., *Taxonomic characterization of nine strains isolated from clinical and*

457   *environmental specimens, and proposal of Corynebacterium tuberculostearicum sp. nov.* Int J

458   Syst Evol Microbiol, 2004. **54**(Pt 4): p. 1055-1061.

459   11.   Cappelli, E.A., et al., *Expanding the bacterial diversity of the female urinary microbiome:*

460   *description of eight new Corynebacterium species.* Microorganisms, 2023. **11**(2): p. 388.

461   12.   Ahmed, N., et al., *Genomic characterization of the C. tuberculostearicum species complex, a*

462   *prominent member of the human skin microbiome.* mSystems, 2023. **8**(6): p. e0063223.

463   13.   Flores Ramos, S., et al., *Genomic stability and genetic defense systems in dolosigranulum*

464   *pigrum, a candidate beneficial bacterium from the human microbiome.* Msystems, 2021.

465   **6**(5): p. e00425-21.

466   14.   Kelly, M.S., et al., *Non-diphtheriae Corynebacterium species are associated with decreased*

467   *risk of pneumococcal colonization during infancy.* The ISME journal, 2021: p. 1-11.

468   15.   Neubauer, M., et al., *Corynebacterium accolens sp. nov., a gram-positive rod exhibiting*

469   *satellitism, from clinical material.* Systematic and applied microbiology, 1991. **14**(1): p. 46-

470   51.

471   16.   Riegel, P., et al., *Genomic diversity and phylogenetic relationships among lipid-requiring*

472   *diphtheroids from humans and characterization of Corynebacterium macginleyi sp. nov.*

473   International Journal of Systematic and Evolutionary Microbiology, 1995. **45**(1): p. 128-133.

474   17.   Tran, T.H., et al., *Metabolic capabilities are highly conserved among human nasal-associated*

475   *Corynebacterium species in pangenomic analyses.* bioRxiv, 2024: p.

476   2023.06.05.543719.

477   18.   Meier-Kolthoff, J.P. and M. Göker, *TYGS is an automated high-throughput platform for state-*

478   *of-the-art genome-based taxonomy.* Nature communications, 2019. **10**(1): p. 2182.

479   19.   Chaumeil, P.A., et al., *GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy*

480   *Database.* Bioinformatics, 2019. **36**(6): p. 1925-7.

481   20.   Chalita, M., et al., *EzBioCloud: a genome-driven database and platform for microbiome*

482   *identification and discovery.* Int J Syst Evol Microbiol, 2024. **74**(6).

483   21.   Hoang, D.T., et al., *UFBoot2: Improving the Ultrafast Bootstrap Approximation.* Mol Biol Evol,

484   2018. **35**(2): p. 518-522.

485   22.   Pritchard, L., et al., *Genomics and taxonomy in diagnostics for food security: soft-rotting*

486   *enterobacterial plant pathogens.* Analytical methods, 2016. **8**(1): p. 12-24.

487   23.   Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**: p.

488   421.

489   24.   Riesco, R. and M.E. Trujillo, *Update on the proposed minimal standards for the use of*

490   *genome data for the taxonomy of prokaryotes.* International Journal of Systematic and

491   Evolutionary Microbiology, 2024. **74**(3): p. 006300.

492   25.   Blasche, S., Y. Kim, and K.R. Patil, *Draft Genome Sequence of Corynebacterium kefirresidentii*

493   *SB, Isolated from Kefir.* Genome Announc, 2017. **5**(37).

494   26.   Salamzade, R., M.H. Swaney, and L.R. Kalan, *Comparative Genomic and Metagenomic*

495   *Investigations of the Corynebacterium tuberculostearicum Species Complex Reveals Potential*

496   *Mechanisms Underlying Associations To Skin Health and Disease.* Microbiol Spectr, 2023.

497   **11**(1): p. e0357822.

498   27.   Roach, D.J., et al., *A Year of Infection in the Intensive Care Unit: Prospective Whole Genome*

499   *Sequencing of Bacterial Clinical Isolates Reveals Cryptic Transmissions and Novel Microbiota.*

500   PLoS Genet, 2015. **11**(7): p. e1005413.

501   28.   Tran, T.H., et al., *Metabolic capabilities are highly conserved among human nasal-associated*

502   *Corynebacterium species in pangenomic analyses.* mSystems, 2024: p. e0113224.

503   29.   Eren, A.M., et al., *Community-led, integrated, reproducible multi-omics with anvi'o.* Nat

504   Microbiol, 2021. **6**(1): p. 3-6.

505 30. Delmont, T.O. and A.M. Eren, *Linking pangenomes and metagenomes: the Prochlorococcus*
506   *metapangenome.* PeerJ, 2018. **6**: p. e4320.
507 31. Kanehisa, M., et al., *KEGG for taxonomy-based analysis of pathways and genomes.* Nucleic
508   Acids Res, 2023. **51**(D1): p. D587-d592.
509 32. Sasser, M., *Bacterial identification by gas chromatographic analysis of fatty acids methyl*
510   *esters (GC-FAME).* Newark, NY: Microbial ID, 2006.
511 33. Shen, W., B. Sipos, and L. Zhao, *SeqKit2: A Swiss army knife for sequence and alignment*
512   *processing.* Imeta, 2024. **3**(3): p. e191.
513 34. Larsson, A., *AliView: a fast and lightweight alignment viewer and editor for large datasets.*
514   Bioinformatics, 2014. **30**(22): p. 3276-3278.
515 35. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: the European Molecular Biology Open Software*
516   *Suite.* Trends Genet, 2000. **16**(6): p. 276-7.
517 36. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.*
518   Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
519 37. Minh, B.Q., et al., *IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in*
520   *the Genomic Era.* Mol Biol Evol, 2020. **37**(5): p. 1530-1534.
521 38. Contreras-Moreira, B. and P. Vinuesa, *GET_HOMOLOGUES, a versatile software package for*
522   *scalable and robust microbial pangenome analysis.* Appl Environ Microbiol, 2013. **79**(24): p.
523   7696-701.
524 39. Vinuesa, P., L.E. Ochoa-Sánchez, and B. Contreras-Moreira, *GET_PHYLOMARKERS, a*
525   *Software Package to Select Optimal Orthologous Clusters for Phylogenomics and Inferring*
526   *Pan-Genome Phylogenies, Used for a Critical Geno-Taxonomic Revision of the Genus*
527   *Stenotrophomonas.* Front Microbiol, 2018. **9**: p. 771.
528 40. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p.
529   2068-2069.

530   41.   Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site*

531   *identification.* BMC Bioinformatics, 2010. **11**: p. 119.

532   42.   Kristensen, D.M., et al., *A low-polynomial algorithm for assembling clusters of orthologous*

533   *groups from intergenomic symmetric best matches.* Bioinformatics, 2010. **26**(12): p. 1481-7.

534   43.   Li, L., C.J. Stoeckert, and D.S. Roos, *OrthoMCL: identification of ortholog groups for*

535   *eukaryotic genomes.* Genome research, 2003. **13**(9): p. 2178-2189.

536   44.   Kalyaanamoorthy, S., et al., *ModelFinder: fast model selection for accurate phylogenetic*

537   *estimates.* Nature methods, 2017. **14**(6): p. 587-589.

538   45.   Chernomor, O., A. von Haeseler, and B.Q. Minh, *Terrace Aware Data Structure for*

539   *Phylogenomic Inference from Supermatrices.* Syst Biol, 2016. **65**(6): p. 997-1008.

540   46.   Anisimova, M., et al., *Survey of branch support methods demonstrates accuracy, power, and*

541   *robustness of fast likelihood-based approximation schemes.* Systematic biology, 2011. **60**(5):

542   p. 685-699.

543   47.   Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree*

544   *display and annotation.* Nucleic Acids Res, 2021. **49**(W1): p. W293-w296.

545   48.   Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database*

546   *search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.

547

Tables and Figures

549

550 **Table 1.** Estimated stepwise complete KEGG modules shared by at least 26 of the 30

551 *Corynebacterium* strain genomes belonging to species closely related to *C. tuberculostearicum,*

552 including *C. nasorum* sp. nov. and *C. hallux* sp. nov. (# represents number of genomes)

Module ID	Module Name	Module Subcategory	#
M00015	Proline biosynthesis, glutamate => proline	Arginine and proline metabolism	30
M00844	Arginine biosynthesis, ornithine => arginine	Arginine and proline metabolism	30
M00970	Proline degradation, proline => glutamate	Arginine and proline metabolism	30
M00022	Shikimate pathway, phosphoenolpyruvate + erythrose-4P => chorismite	Aromatic amino acid metabolism	30
M00019	Valine/isoleucine biosynthesis, pyruvate => valine / 2-oxobutanoate => isoleucine	Branched-chain amino acid metabolism	30
M00432	Leucine biosynthesis, 2-oxoisovalerate => 2-oxoisocaproate	Branched-chain amino acid metabolism	30
M00570	Isoleucine biosynthesis, threonine => 2-oxobutanoate => isoleucine	Branched-chain amino acid metabolism	30
M00017	Methionine biosynthesis, aspartate => homoserine => methionine	Cysteine and methionine metabolism	30
M00021	Cysteine biosynthesis, serine => cysteine	Cysteine and methionine metabolism	30
M00045	Histidine degradation, histidine => N-formiminoglutamate => glutamate	Histidine metabolism	30
M00018	Threonine biosynthesis, aspartate => homoserine => threonine	Serine and threonine metabolism	30
M00621	Glycine cleavage system	Serine and threonine metabolism	30
M00793	dTDP-L-rhamnose biosynthesis	Polyketide sugar unit biosynthesis	30
M00096	C5 isoprenoid biosynthesis, non-mevalonate pathway	Terpenoid backbone biosynthesis	30
M00364	C10-C20 isoprenoid biosynthesis, bacteria	Terpenoid backbone biosynthesis	30
M00365	C10-C20 isoprenoid biosynthesis, archaea	Terpenoid backbone biosynthesis	30
M00001	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	Central carbohydrate metabolism	30

M00002	Glycolysis, core module involving three-carbon compounds	Central carbohydrate metabolism	30
M00003	Gluconeogenesis, oxaloacetate => fructose-6P	Central carbohydrate metabolism	30
M00004	Pentose phosphate pathway (Pentose phosphate cycle)	Central carbohydrate metabolism	30
M00005	PRPP biosynthesis, ribose 5P => PRPP	Central carbohydrate metabolism	30
M00006	Pentose phosphate pathway, oxidative phase, glucose 6P => ribulose 5P	Central carbohydrate metabolism	30
M00007	Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P	Central carbohydrate metabolism	30
M00009	Citrate cycle (TCA cycle, Krebs cycle)	Central carbohydrate metabolism	30
M00010	Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate	Central carbohydrate metabolism	30
M00011	Citrate cycle, second carbon oxidation, 2-oxoglutarate => oxaloacetate	Central carbohydrate metabolism	30
M00549	Nucleotide sugar biosynthesis, glucose => UDP-glucose	Other carbohydrate metabolism	30
M00554	Nucleotide sugar biosynthesis, galactose => UDP-galactose	Other carbohydrate metabolism	30
M00632	Galactose degradation, Leloir pathway, galactose => alpha-D-glucose-1P	Other carbohydrate metabolism	30
M00909	UDP-N-acetyl-D-glucosamine biosynthesis, prokaryotes, glucose => UDP-GlcNAc	Other carbohydrate metabolism	30
M00151	Cytochrome bc1 complex respiratory unit	ATP synthesis	30
M00155	Cytochrome c oxidase, prokaryotes	ATP synthesis	30
M00157	F-type ATPase, prokaryotes and chloroplasts	ATP synthesis	30
M00579	Phosphate acetyltransferase-acetate kinase pathway, acetyl-CoA => acetate	Carbon fixation	30
M00086	beta-Oxidation, acyl-CoA synthesis	Fatty acid metabolism	30
M00120	Coenzyme A biosynthesis, pantothenate => CoA	Cofactor and vitamin metabolism	30
M00121	Heme biosynthesis, plants and bacteria, glutamate => heme	Cofactor and vitamin metabolism	30
M00125	Riboflavin biosynthesis, plants and bacteria, GTP => riboflavin/FMN/FAD	Cofactor and vitamin metabolism	30
M00126	Tetrahydrofolate biosynthesis, GTP => THF	Cofactor and vitamin metabolism	30
M00881	Lipoic acid biosynthesis, plants and bacteria, octanoyl-ACP => dihydrolipoyl-E2/H	Cofactor and vitamin metabolism	30

M00899	Thiamine salvage pathway, HMP/HET => TMP	Cofactor and vitamin metabolism	30
M00916	Pyridoxal-P biosynthesis, R5P + glyceraldehyde-3P + glutamine => pyridoxal-P	Cofactor and vitamin metabolism	30
M00926	Heme biosynthesis, bacteria, glutamyl-tRNA => coproporphyrin III => heme	Cofactor and vitamin metabolism	30
M00048	De novo purine biosynthesis, PRPP + glutamine => IMP	Purine metabolism	30
M00049	Adenine ribonucleotide biosynthesis, IMP => ADP,ATP	Purine metabolism	30
M00050	Guanine ribonucleotide biosynthesis, IMP => GDP,GTP	Purine metabolism	30
M00053	Deoxyribonucleotide biosynthesis, ADP/GDP/CDP/UDP => dATP/dGTP/dCTP/dUTP	Purine metabolism	30
M00938	Pyrimidine deoxyribonucleotide biosynthesis, UDP => dTTP	Pyrimidine metabolism	30
M00023	Tryptophan biosynthesis, chorismate => tryptophan	Aromatic amino acid metabolism	29
M00020	Serine biosynthesis, glycerate-3P => serine	Serine and threonine metabolism	29
M00140	C1-unit interconversion, prokaryotes	Cofactor and vitamin metabolism	29
M00307	Pyruvate oxidation, pyruvate => acetyl-CoA	Central carbohydrate metabolism	28
M00028	Ornithine biosynthesis, glutamate => ornithine	Arginine and proline metabolism	26
M00016	Lysine biosynthesis, succinyl-DAP pathway, aspartate => lysine	Lysine metabolism	26

553

554 **Table 2.** Microbiological, biochemical, and genomic characteristics of *Corynebacterium* strains

555 Strains: 1, *C. hallux* sp. nov. CTNIH22[T]; 2, *C. nasorum* sp. nov. KPL3804[T]; 3, *C. nasorum* sp. nov. MSK185; 4, *C. yonathiae* MSK136; 5, *C. yonathiae* KPL2619; 6,

556 *C. tuberculostearicum* ATCC 35692[T]; 7, *C. accolens* ATCC 49725[T]; 8, *C. macginleyi* ATCC 51787[T]. Data were generated in this study for all strains. +, positive;

557 w, weakly positive; -, negative; +/-, variable. BHI, brain-heart infusion; SBA, tryptic soy agar with 5% sheep blood; TSB, tryptic soy broth. Results in

558 parentheses indicate optimal values.

			1	2, 3	4, 5	6	7	8
Microbiological								
	Growth on solid media							
		SBA	+	+	+	w	+	(+)
		BHI	w	w	w	w	w	+
		BHI with 1% Tween 80	(+)	(+)	(+)	(+)	(+)	+
	Growth in liquid media							
		BHI	−	−	−	−	−	−
		BHI with 0.2% Tween 80	(+)	(+)	(+)	(+)	(+)	(+)
		TSB	−	−	−	−	−	−
		TSB with 0.2% Tween 80	+	+	+	+	+	+
	Temperature (°C) for growth		20, (30), 37, 42 (w)	30, (37), 42	20, (30), 37, 42 (w)	20 (w), 30, (37), 42 (w)	20, 30 (w), 37, 42	30, (37), 42 (w)
	Atmospheric conditions for growth							
		Aerobic	(+)	(+)	(+)	(+)	(+)	+
		Microaerophilic	+	+	+	+	+	(+)
		Anaerobic	−	w	w	−	w	w
	pH for growth		7 − 8 (7 − 8)	6 − 8 (7 − 8)	6 − 8 (7 − 8)	6 − 8 (7 − 8)	7 − 8 (7 − 8)	7 − 8 (8)
	Salinity for growth		≤14%	≤10%	≤10%	≤10%	≤14%	≤10%
Biochemical								
	Nitrate reduction		−	−	−	−	+	+
	Pyrazinamidase		−	w	−	−	−	−
	Pyrrolidonyl arylamidase		w	+/-	+	−	−	−
	Alkaline phosphatase		+	+	+	−	−	+
	β-Glucuronidase		−	−	−	−	−	−
	β-Galactosidase		−	−	−	−	−	−

α-Glucosidase	–	–	–	–	–	–
N-Acetyl-β-glucosaminidase	–	–	–	–	–	–
Esculin hydrolysis	–	–	–	–	–	–
Gelatin hydrolysis	–	–	–	–	–	–
Urea hydrolysis	–	–	–	–	–	–
Catalase	+	+	+	+	+	+
Oxidase	–	–	–	–	–	–
Motility	–	–	–	–	–	–
Carbohydrate fermentation						
D-glucose	–	–	–	–	+	+
D-ribose	–	–	–	–	+	+
D-xylose	–	–	–	–	–	–
D-mannitol	–	–	–	–	–	w
D-maltose	–	–	–	–	–	–
D-lactose	–	–	–	–	–	–
D-saccharose (sucrose)	–	–	–	–	–	+
Glycogen	–	–	–	–	–	–
Genomic						
GenBank accession number(s)	GCF_032821755.1	GCF_037908315.1, GCF_030229765.1	GCF_037908465.1, GCF_022288805.2	GCF_016728365.1	GCF_023520795.1	GCF_003688935.1
Genome length (Mbp)	2.53	2.45, 2.43	2.47, 2.35	2.45	2.47	2.43
GC content (mol%)	58.5	58.5, 58.7	58.5, 58.6	59.7	59.7	57.1

559

560 **Table 3.** Cellular fatty acid composition of *Corynebacterium* spp. strains by fatty acid methyl esters (FAME) analysis

561 Strains: 1, *C. hallux* sp. nov. CTNIH22[T]; 2, *C. nasorum* sp. nov. KPL3804[T]; 3, *C. nasorum* sp. nov. MSK185; 4, *C. yonathiae* MSK136; 5, *C. yonathiae* KPL2619; 6,
562 *C. tuberculostearicum* ATCC 35692[T]; 7, *C. accolens* ATCC 49725[T]; 8, *C. macginleyi* ATCC 51787[T]. Data were generated in this study for all strains. –, not
563 detected. Percentages may not sum to 100% due to rounding.

Fatty Acid	1	2	3	4	5	6	7	8
Saturated								
9:0	-	-	-	-	-	2.6	-	-
12:0	0.3	-	-	-	-	-	-	-
14:0	2.2	1.0	3.4	2.4	5.2	3.0	5.4	3.5
16:0	26.9	34.5	35.4	32.5	42.6	35.1	33.9	33.7
17:0	1.4	1.8	1.3	1.5	4.9	1.9	1.4	1.5
18:0	12.9	15.6	12.1	15.8	13.2	13.1	15.7	18.4
10Me-18:0	1.1	-	-	-	-	-	-	-
19:0	0.1	-	-	-	-	-	-	-
20:0	0.1	1.0	0.7	1.0		1.9	-	1.6
Unsaturated								
13:1 ω1c	0.6	-	-	-	-	2.2	-	18.0
14:1 ω5c	0.1	-	-	-	-	-	-	
15:1 ω5c	-	-	0.6	-	-	-	-	-
15:1 ω8c	0.2			-	-	-	-	-
17:1 ω8c	3.0	-	0.9	0.9	-	-	-	-
18:1 ω9c	32.5	21.4	17.5	22.8	19.3	20.3	16.6	-
20:1 ω9c	0.3	-	-	-	-	-	-	-
20:4 ω6,9,12,15c	1.7	3.3	3.4	2.8	-	2.6	3.7	1.9
Branched-chain								
15:0 iso	0.3	-	-	-	-	-	-	-
15:0 anteiso	0.4	-	0.6	-	-	-	-	-
16:0 iso	0.3	-	-	-	-	-	-	-
17:0 iso	0.5	-	0.6	-	-	-	-	-
17:0 anteiso	0.8	0.9	0.9	0.9	-	-	-	0.9
18:0 iso	0.3	-	-	-	-	-	-	-
19:0 iso	0.4	-	-	-	-	-	-	-

20:0 iso	-	-	0.4	-	-	-	-	-
Summed features*								
16:1 ω6c or 16:1 ω7c	1.4	2.1	2.2	1.8	-	1.8	2.6	1.7
18:1 ω6c or 18:1 ω7c	6.1	8.5	9.0	8.0	6.1	7.5	9.5	9.2
18:2 ω6,9c or 18:0 anteiso	5.9	10.1	11.1	9.5	8.7	8.0	11.2	9.9
19:1 ω9c or 19:1 ω11c	0.3	-	-	-	-	-	-	-

*groups of more than one fatty acid that could not be separated by gas chromatography

564

565 **Figure 1. Maximum-likelihood 16S rRNA gene phylogeny of new isolates and type strains of**

566 ***Corynebacterium* species.** A maximum likelihood phylogeny based on nearly full-length 16S rRNA

567 genes of *Corynebacterium* species with ultrafast bootstrap values below 95 at many nodes.

568 *Mycobacterium tuberculosis* H37Rv[T] is the designated outgroup.
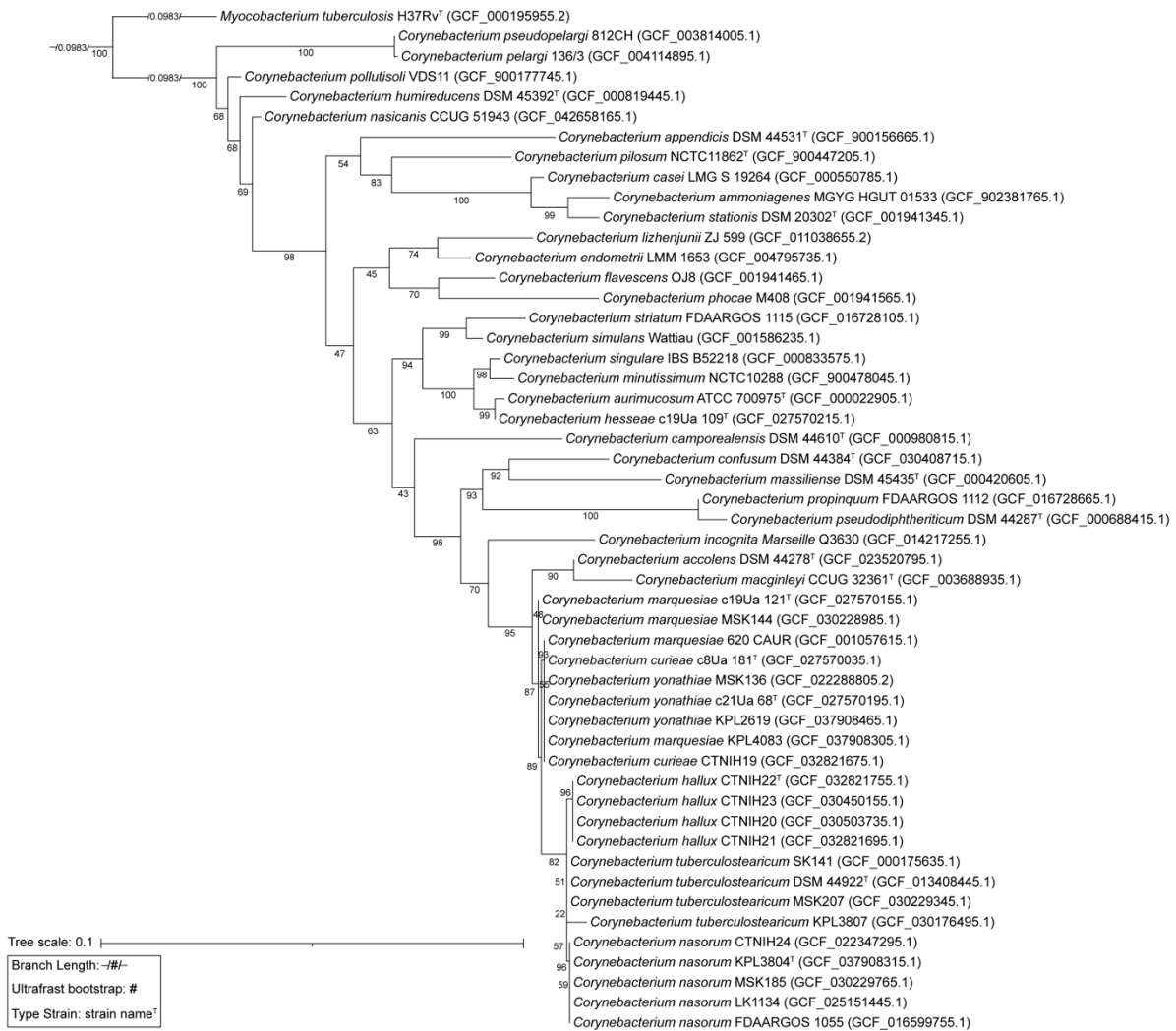
569

570

571 **Figure 2. Average nucleotide identity (ANI) among species closely related to *C. tuberculostearicum***

572 **revealed two novel species**. We used pyani v0.2.9 with ANIb BLAST+ to construct a whole-genome

573 identity ANI heat matrix (see Supplemental Methods) [22, 23]. Boxes highlight species boundaries

574 defined by an ANI threshold of 95% with purple for *C. nasorum* sp. nov. (Cna), pink for *C. curieae*

575 (Ccu), light green for *C. hallux* sp. nov. (Cha), dark green for *C. yonathiae* (Cyo), turquoise for *C.*

576 *marquesiae* (Cmq), and blue for *C. tuberculostearicum* (Ctu). For *C. tuberculostearicum,* all strains

577 reached a 95% ANI threshold compared to the type strain in at least one direction. The ANIb

578 comparisons indicate that the genome named C. aurimucosum_620_CAUR should be assigned to *C.*

579 *marquesiae*.

580

581
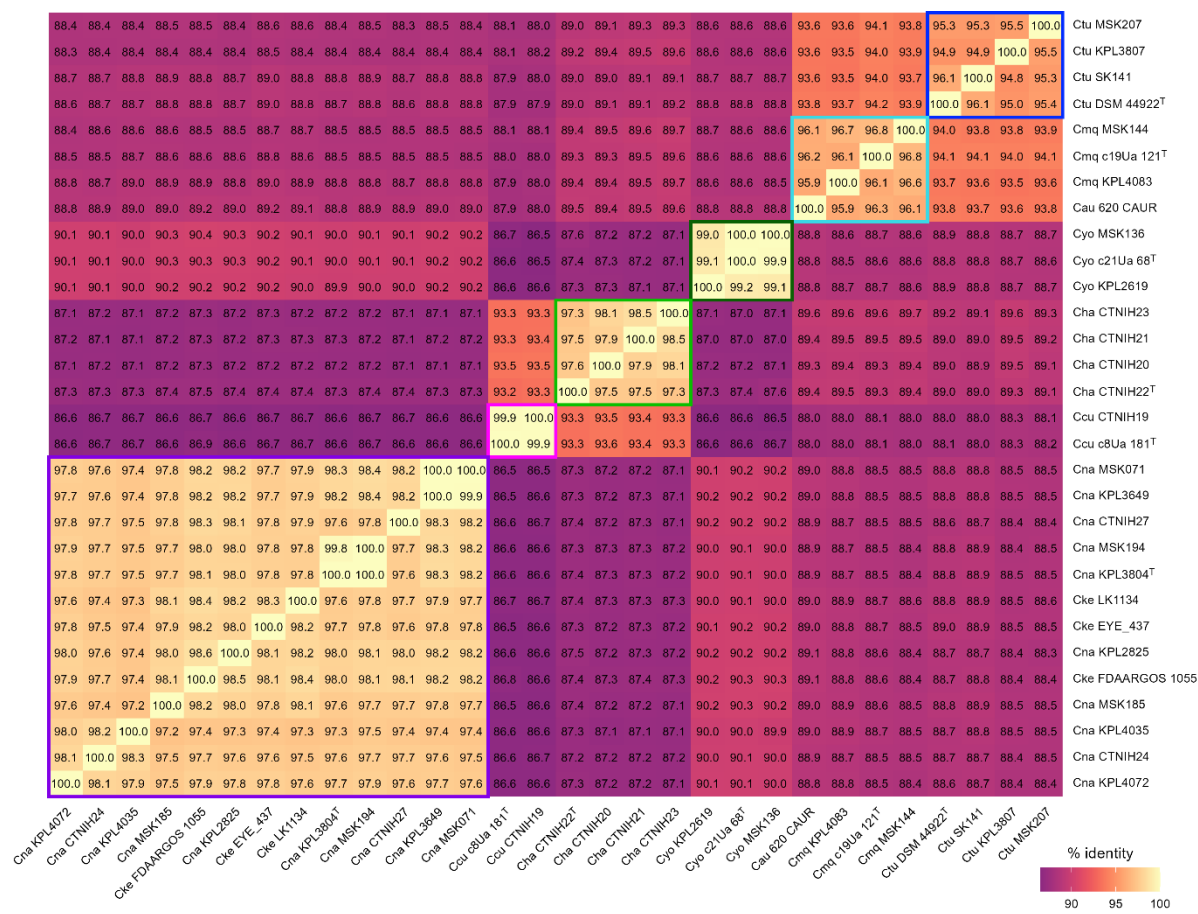
582 **Figure 3. Maximum-likelihood phylogenomic tree of new isolates and type strains of**

583 *Corynebacterium* **species.** A maximum likelihood phylogenomic tree was constructed using 305

584 shared conservative core gene clusters and *Mycobacterium tuberculosis* H37Rv^T as the designated

585 outgroup. This monophyletic tree shows robust separation of *Corynebacterium* species based on

586 ultrafast bootstrap values.
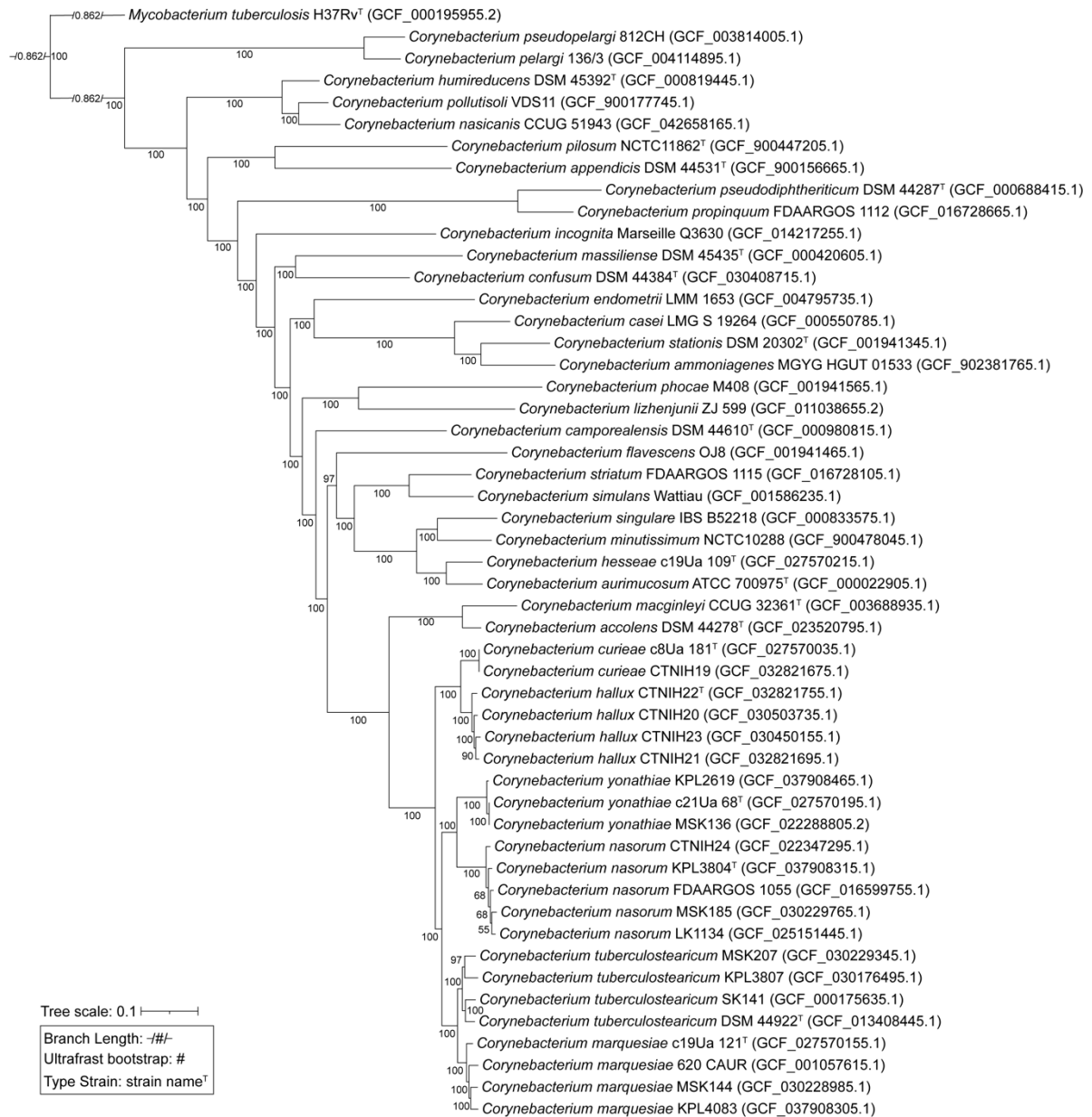
587

588

589 **Figure 4. Species closely related to *C. tuberculostearicum* are estimated to largely share a common**

590 **set of metabolic capabilities**. The heatmap represents average estimated module stepwise

591 completion scores by KEGG subcategories for each of the 30 genomes from Figure 2 covering six

592 species that clade closely with *C. tuberculostearicum* (**Figures 3 and S2A**). Average stepwise

593 completion scores were calculated including only modules detected in at least one of the analyzed

594 genomes. (P) represents pathway modules; (S) represents signature modules. Cna, *C. nasorum* sp.

595 nov.; Ccu, *C. curieae*; Cha, *C. hallux* sp. nov.; Cyo, *C. yonathiae*; Cmq, *C. marquesiae*; Ctu, *C.*

596 *tuberculostearicum*.

597

598

599 **Figure 5**. Scanning electron microscopy images (12,000x) of strains *C. hallux* sp. nov. CTNIH22[T] (**A**), *C. nasorum* sp. nov. KPL3804[T] (**B**), and *C. yonathiae*
600 KPL2619 (**C**).

601 **A.**

602

603 **C.**

604
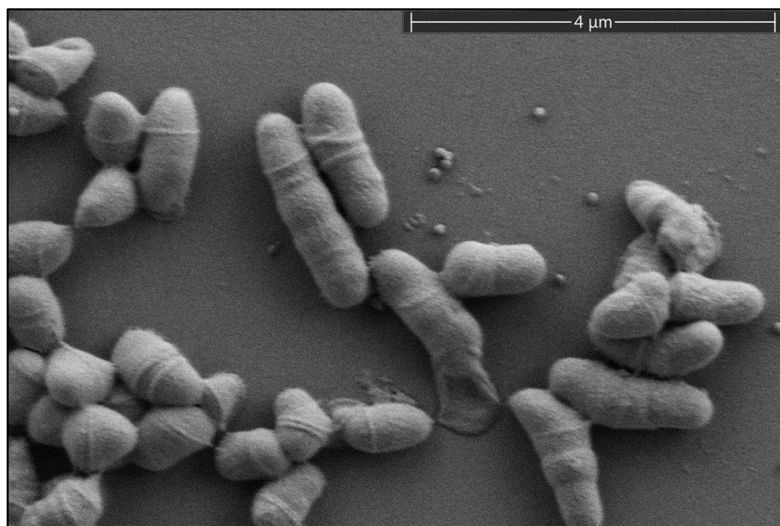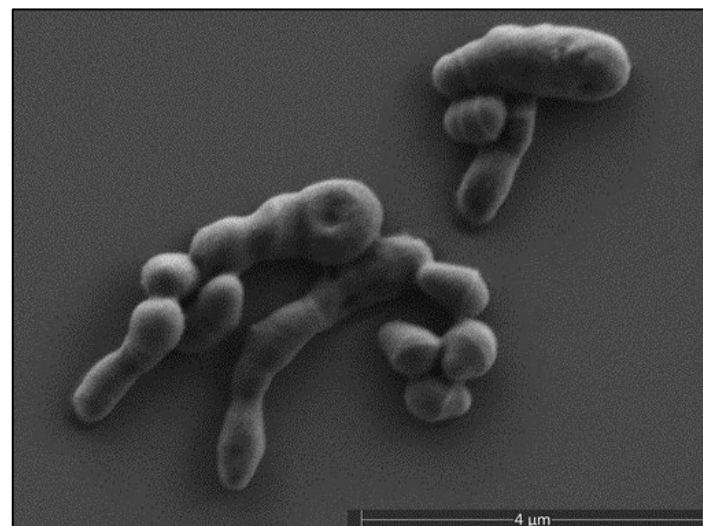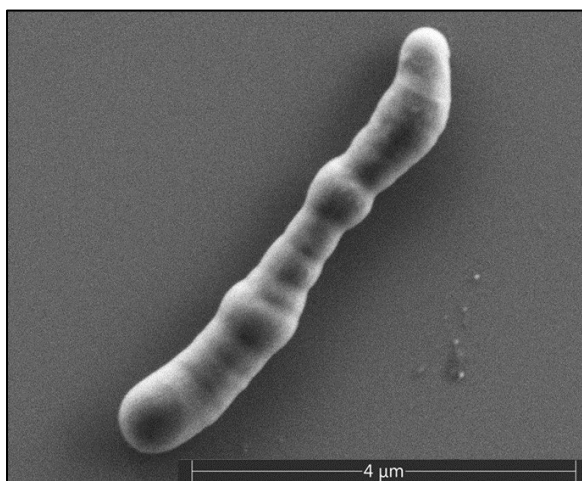
Supplementary Figures

Figure S1. A phylogenomic tree of 68 members of the *Corynebacterium* genus illustrates the distinct species closely related to *Corynebacterium tuberculostearicum*. In contrast, phylogenetic trees of the same 68 *Corynebacterium species* based on either the (B) 16S rRNA gene or (C) *rpoB* gene showed the limitations of single-gene phylogenies for this genus. (**A**) A maximum-likelihood phylogenetic tree of 68 *Corynebacterium* species based on 16S rRNA gene sequences from 72 *Corynebacterium* strain genomes (**Table S1**) representing species across the breadth of the phylogeny of this genus has a number of poorly supported branches. (**B**) Although better than the full-length 16S rRNA gene phylogeny, a maximum-likelihood phylogeny based on full-length *rpoB* gene sequences of the same 72 strains also illustrates the limitation of single-gene phylogenies for resolving closely related *Corynebacterium* species. For example, the strains *C. yonathiae* KPL2619 (Cyo_KPL2619) and *C. marquesiae* c19Ua_121 have an average nucleotide identity based on BLAST+ (ANIb) below 95% (**Figure 2**), indicating they are distinct species, yet these incorrectly clade together here. Similarly, several strains assigned to the proposed species *C. nasorum* sp. nov. based on ANIb values are incorrectly in clades of other closely related species. (**C**) A maximum-likelihood phylogenomic tree of the same 68 *Corynebacterium* species was constructed using 193 concatenated and aligned shared single-copy core gene clusters from 72 *Corynebacterium* strain genomes (**Table S1**). The majority of branches in this phylogeny exhibit strong support with ultrafast bootstrap values of 95 or higher. See Supplemental Methods for description of the construction of these phylogenies.

626
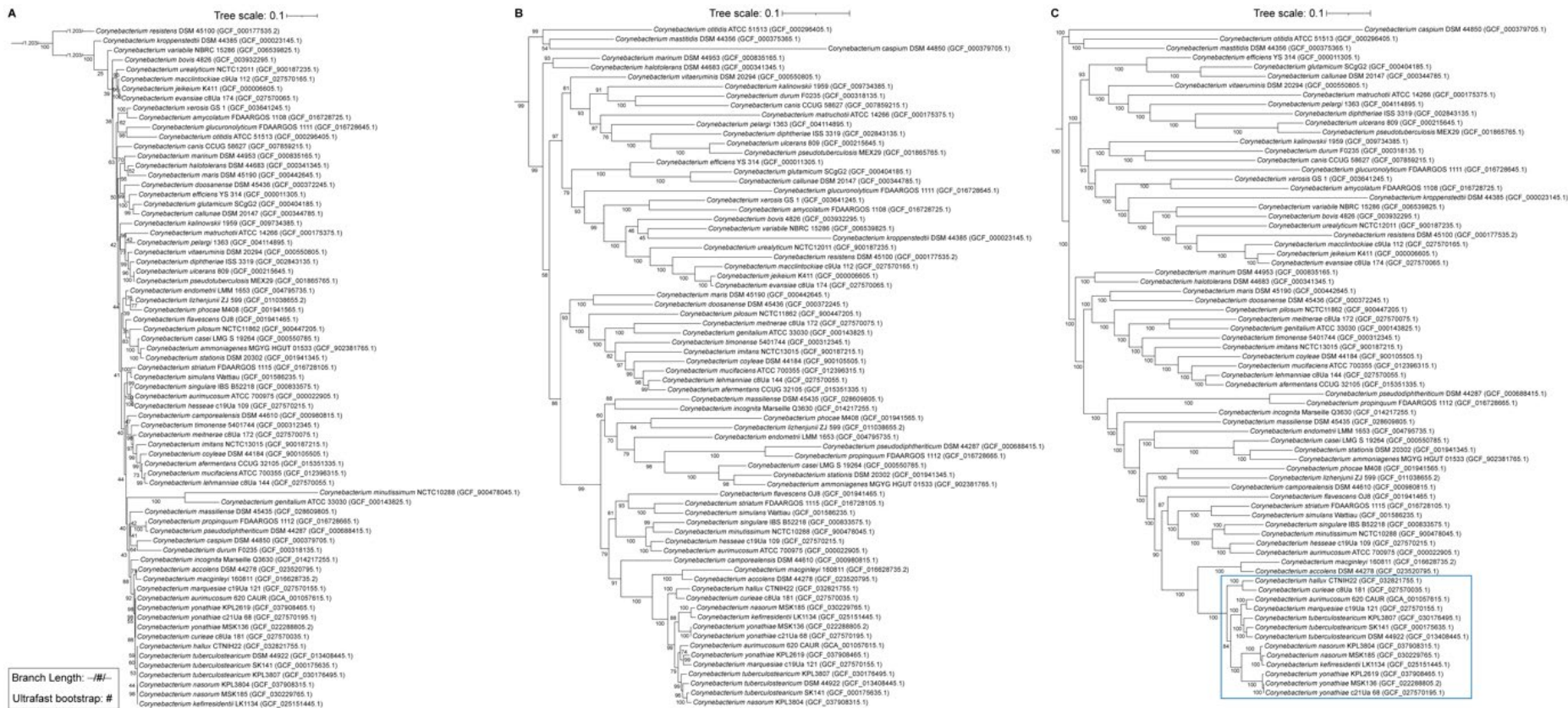
627 **Figure S2. A phylogenomic tree provided superior resolution among the species most closely**

628 **related to *Corynebacterium tuberculostearicum* compared to single-gene phylogenies. (A)** The

629 maximum likelihood phylogenomic tree constructed based on 1447 conservative core gene clusters

630 provided higher resolution of the distinct clades within the *Corynebacterium tuberculostearicum*

631 species complex. The use of a large pool of single copy gene clusters shared across 30

632 *Corynebacterium* genomes within the *C. tuberculostearicum* species complex enhanced species

633 delineation, with robust ultrafast bootstrap values supporting the distinct clades. Ccu (pink

634 branches) is *C. curieae*; Cha (light green branches) is *C. hallux* sp. nov.; Ctu (blue branches) is *C.*

635 *tuberculostearicum*; Cmq (turquoise branches) is *C. marquesiae*; Cyo (dark green branches) is *C.*

636 *yonathiae*; and Cna (purple branches) is *C. nasorum* sp. nov. **(B)** In contrast, a full-length 16S rRNA

637 gene maximum likelihood phylogeny had strains of different species sometimes intermingled in a

638 single clade and was poorly supported based on ultrafast bootstrap values. This is consistent with

639 the known limitations of using 16S rRNA gene phylogenies within this genus and highlights the

640 limitations of using the 16S rRNA gene for resolving evolutionary relationships within the *C.*

641 *tuberculostearicum* species complex. **(C)** The full-length *rpoB* gene maximum likelihood phylogeny

642 had better support than the 16S rRNA gene for the species within the *C. tuberculostearicum* species

643 complex. However, it still had two clades with intermingled species and was inferior to the

644 phylogenomic tree using conservative core gene clusters shown in S2A.

645
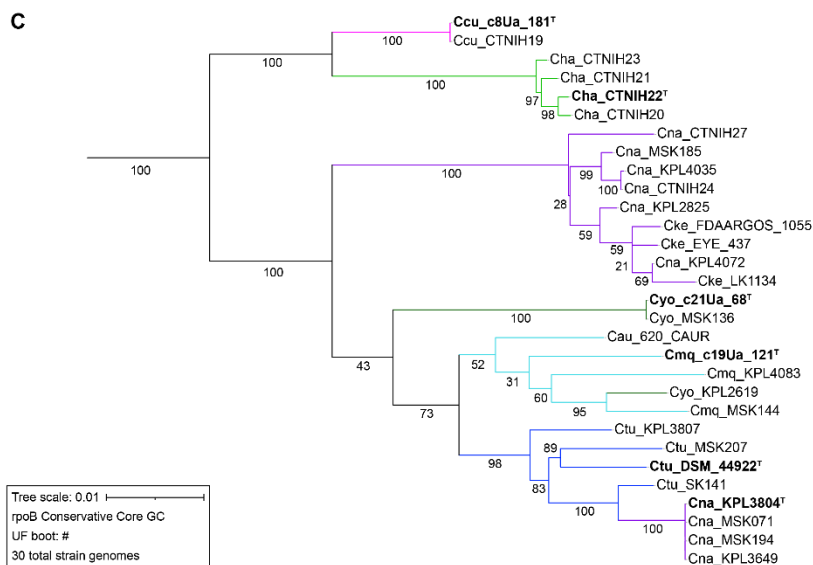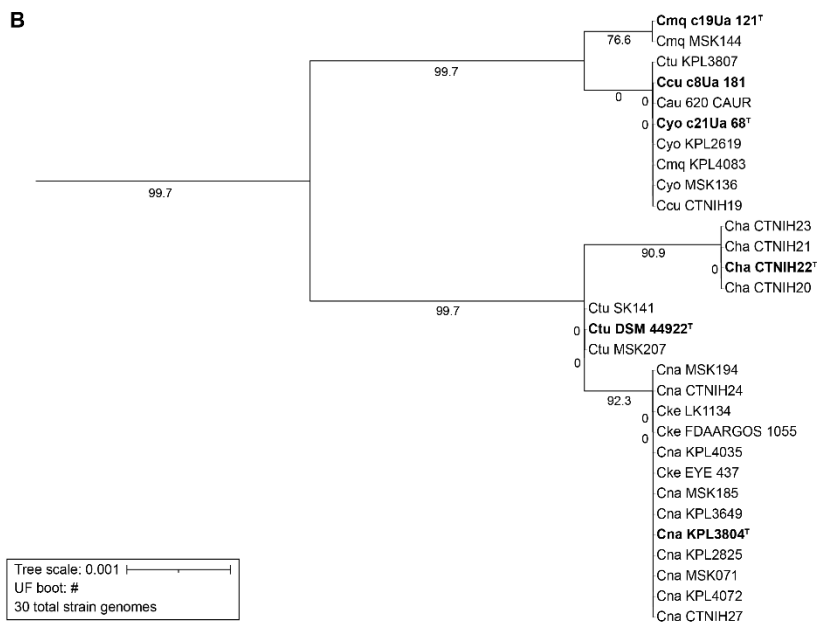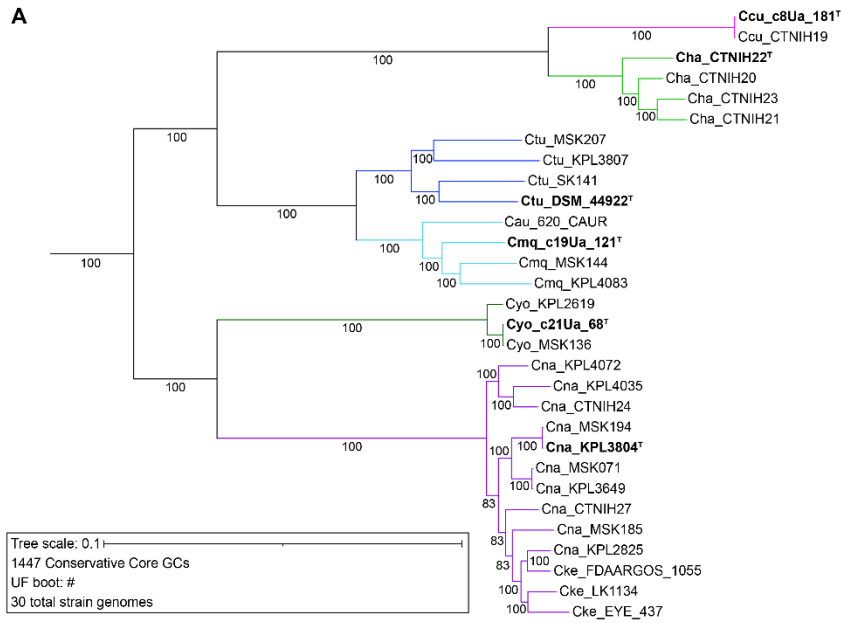
Supplementary Methods

646

647

648 **Construction of phylogenic trees**. To generate the maximum likelihood 16S rRNA gene phylogenies

649 shown in Figures 1, S1A, and S2B, we performed the following steps. First, to identify the 16S rRNA

650 genes present in each genome, we ran barrnap v0.9 (https://github.com/tseemann/barrnap) with

651 default parameters on the fasta files containing the genome assemblies for each phylogeny. We

652 then used seqkit (v2.6.0) grep -r -n -p '16S_rRNA' to select the 16S rRNA gene sequences from each

653 genome's total rRNA sequences [33]. For genomes that had multiple copies of the 16S rRNA gene,

654 we manually inspected the sequences and removed copies that were less than 50% of the expected

655 16S rRNA gene sequence length using AliView v1.28 [34], aligned the remaining copies using

656 MUSCLE v3.8.1551 with default parameters, and generated a consensus 16S rRNA gene sequence

657 using the EMBOSS cons command [35, 36]. We concatenated and aligned the single and consensus

658 16S sequences with the linux `cat` command and MUSCLE. The resulting 16S rRNA gene alignment

659 was used as input for IQ-TREE2 v2.1.3 [37] and we set the parameters -alrt to 1000 and -B to 1000.

660

661 To generate the maximum-likelihood *rpoB* phylogenies shown in Figures S1B and S2C, we identified

662 the single copy *rpoB* gene cluster from the conservative core determined with GET_HOMOLOGUES

663 [38] (see below) and then aligned and concatenated the *rpoB* gene from all the genomes with

664 GET_PHYLOMARKERS [39]. Then we used IQ-TREE2 with the same parameters as the 16S rRNA tree

665 [37].

666

667 To generate the maximum likelihood phylogenomic trees in Figures 3, S1C and S2A, we used Prokka

668 v1.14.6 [40] with default settings to annotate each bacterial genome, based on the prediction of

669 coding sequences with Prodigal [41]. For detailed methods on the annotation of genomic assemblies

670   , please see https://klemonlab.github.io/NovCor_Manuscript/Methods_Prokka_Annotations.html.

671   We then used GET_HOMOLOGUES (version 13062023) [38] to separately identify the core gene

672   clusters (GCs) shared by the set of strains used for each individual tree. The consensus of the single

673   copy core GCs from three clustering algorithms; bidirectional best-hits, cluster of orthologs triangles

674   (COGS) v2.1 [42], and Markov Cluster Algorithm OrthoMCL (OMCL) v2.4 [43], defined the

675   conservative shared core genome for each group using ./get_homologues.pl. Subsequently, we

676   employed GET_PHYLOMARKERS v2.2.9.1 [39] to align and concatenate the shared single copy core

677   gene clusters. These were then analyzed using IQ-TREE2 v2.1.3 [37] with the following parameters: -

678   p (edge-linked partition model and ModelFinder functions) [44, 45], -alrt 1000 (replicate SH-like

679   approximate likelihood ratio test) [46], and -B 1000 (number of ultrafast bootstrap replicates) [21].

680

681   To visualize, scale, edit, annotate names, and root trees at the midpoint for each phylogeny, we used

682   the phylogenetic tool iTOL version 6 [47]. For detailed methods on the construction of all

683   phylogenetic trees, please see

684   https://klemonlab.github.io/NovCor_Manuscript/Methods_Phylogenies.html.

685

686   **Average Nucleotide Identity.** We used pyani (version 0.2.9) [22] with ANIb BLAST+ [23, 48] to

687   construct a WGS identity ANI heat matrix. We used the data in the pyani output file

688   ANIb_percentage_identity.tab to create Figure 2 in R, with genome order based on the

689   corresponding .svg file. For detailed methods, see

690   https://klemonlab.github.io/NovCor_Manuscript/Methods_ANIs.html.