Metabolic capabilities are highly conserved among human nasal-associated *Corynebacterium* species in pangenomic analyses

Authors: Tommy H. Tran[a], Ari Q. Roberts[a], Isabel F. Escapa[a], Wei Gao[b,c,^], Julie A. Segre[d], Heidi H. Kong[e], Sean Conlan[d], Matthew S. Kelly[f], Katherine P. Lemon[a,g#]

Affiliations

[a]Alkek Center for Metagenomics & Microbiome Research, Department of Molecular Virology & Microbiology, Baylor College of Medicine, Houston, Texas, USA

[b]The Forsyth Institute (Microbiology), Cambridge, MA, USA

[c]Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Boston, MA, USA

[d]Microbial Genomics Section, Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

[e]Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD, USA

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

[g]Division of Infectious Diseases, Texas Children's Hospital, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA

Running Title: Nasal *Corynebacterium* spp. share metabolic abilities

[#]Address correspondence to Katherine P. Lemon, katherine.lemon@bcm.edu

Present affiliation:

^W.G. is now at Illumina Inc. (of San Diego, CA, USA) in Melbourne, Australia

Abstract word count: 248

Text word count: 7675

Keywords: nasal microbiota, phylogenetics, pangenomics, metabolism, *Corynebacterium accolens*, *Corynebacterium propinquum*, *Corynebacterium pseudodiphtheriticum*, *Corynebacterium tuberculostearicum, Corynebacterium*

ABSTRACT

Corynebacterium species are globally ubiquitous in human nasal microbiota across the lifespan. Moreover, nasal microbiota profiles typified by higher relative abundances of *Corynebacterium* are often positively associated with health. Among the most common human nasal *Corynebacterium* species are *C. propinquum*, *C. pseudodiphtheriticum, C. accolens*, and *C. tuberculostearicum*. Based on the prevalence of these species, at least two likely coexist in the nasal microbiota of 82% of adults. To gain insight into the functions of these four species, we identified genomic, phylogenomic, and pangenomic properties and estimated the functional protein repertoire and metabolic capabilities of 87 distinct human nasal *Corynebacterium* strain genomes: 31 from Botswana and 56 from the U.S. *C. pseudodiphtheriticum* had geographically distinct clades consistent with localized strain circulation, whereas some strains from the other species had wide geographic distribution across Africa and North America. All four species had similar genomic and pangenomic structures. Gene clusters assigned to all COG metabolic categories were overrepresented in the persistent (core) compared to the accessory genome of each species indicating limited strain-level variability in metabolic capacity. Moreover, core metabolic capabilities were highly conserved among the four species indicating limited species-level metabolic variation. Strikingly, strains in the U.S. clade of *C.*

pseudodiphtheriticum lacked genes for assimilatory sulfate reduction present in the Botswanan clade and in the other studied species, indicating a recent, geographically related loss of assimilatory sulfate reduction. Overall, the minimal species and strain variability in metabolic capacity implies coexisting strains might have limited ability to occupy distinct metabolic niches.

IMPORTANCE

Pangenomic analysis with estimation of functional capabilities facilitates our understanding of the full biologic diversity of bacterial species. We performed systematic genomic, phylogenomic, and pangenomic analyses with qualitative estimation of the metabolic capabilities of four common human nasal *Corynebacterium* species generating a foundational resource. The prevalence of each species in human nasal microbiota is consistent with the common coexistence of at least two species. We identified a notably high level of metabolic conservation within and among species indicating limited options for species to occupy distinct metabolic niches and pointing to the importance of investigating interactions among nasal *Corynebacterium* species. Comparing strains from two continents, *C. pseudodiphtheriticum* had restricted geographic strain distribution characterized by an evolutionarily recent loss of assimilatory sulfate reduction in North American strains. Our findings contribute to understanding the functions of *Corynebacterium* within human nasal microbiota and to evaluating their potential for future use as biotherapeutics.

INTRODUCTION

Nasal *Corynebacterium* species are frequently associated with health in compositional studies of human nasal microbiota. *Corynebacterium* are gram-positive bacteria in the phylum Actinobacteria (Actinomycetota). Based on studies from five continents, *Corynebacterium* species begin colonizing the human nasal passages before two months of age (1-13). *Corynebacterium* colonize both the skin-coated surface of the nasal vestibule (aka nostrils/nares) and the mucus-producing nasal respiratory epithelium coating the nasal passages posterior of the limen nasi through the nasopharynx (14-19). The bacterial microbiota of the human nasal passages from the nostrils through the nasopharynx is highly similar, and we refer to it herein as the human nasal microbiota.

Pediatric nasal microbiota profiles characterized by a high relative abundance of *Corynebacterium* (sometimes paired with a high relative abundance of *Dolosigranulum*) are often associated with health rather than a specific disease or disease-risk state in children (1, 2, 4, 7, 11-13, 20-30). In young children, the genus *Corynebacterium* (alone or with *Dolosigranulum*) is negatively associated with *Streptococcus pneumoniae* nasal colonization, which is important because pneumococcal colonization is a necessary precursor to invasive pneumococcal disease (13, 20, 22, 25, 28). For example, in young children in Botswana, the genus *Corynebacterium* is negatively associated with *S. pneumoniae* colonization both in a cross-sectional study of children younger than two years (25) and in a longitudinal study of infants followed from birth to one year (13). In contrast to these genus-level associations, little is known about species-level prevalence and relative abundance of nasal *Corynebacterium* in children. However, in a cultivation-based study, *C. pseudodiphtheriticum* is positively associated with ear and nasal health in young Indigenous Australian children (age 2-7 years), as was *D. pigrum* (29).

In adult nasal microbiota, the prevalence of *Corynebacterium* is as high as 98.6% based on 16S rRNA gene sequencing (Table S4A in (31)). Like children, some adults have nasal microbiota profiles characterized by a high relative abundance of *Corynebacterium* (18, 32). At least 23 validly published species of *Corynebacterium* can be cultivated from the adult nasal passages (17, 33). However, among these, *C. accolens* and *C. tuberculostearicum* (and/or other members of the *C. tuberculostearicum* species complex (34)) followed by *C. propinquum* and *C. pseudodiphtheriticum* are the most common in human nasal microbiota, in terms of both relative abundance and prevalence (15, 17, 31, 33). Indeed, the human nasal passages appear to be a primary habitat for *C. accolens, C. propinquum and C. pseudodiphtheriticum,* whereas *C. tuberculostearicum* is also common, often at high relative abundances, on other human skin sites (31, 34-36)].

Staphylococcus aureus nasal colonization increases the risk of invasive infection at distant body sites, and the infection isolate matches the colonizing isolate in ~80% of cases (37-40). In the absence of an effective vaccine, there is growing interest in identifying nasal microbionts that confer colonization resistance to *S. aureus*. Some studies of adult nasal microbiota report a negative association between *S. aureus* and either the genus *Corynebacterium* or specific species of *Corynebacterium* (14, 31, 32, 41-45), whereas others do not, which might reflect strain-level variation or differences in populations studied. In addition to studies in healthy adults, many studies on the nasal microbiota of adults focus on chronic rhinosinusitis and data on the relationship between *Corynebacterium* and chronic rhinosinusitis continues to evolve as larger cohorts are sampled (46).

Several small human studies support the potential use of *Corynebacterium* species to inhibit or eradicate pathobiont colonization of the human nasal passages. For example, repeated nasal application of *Corynebacterium* sp. Co304 eliminated *S. aureus* nasal colonization in 71% of 17 adults (41). Similarly, daily use of a nasal spray of *C. pseudodiphtheriticum* 090104 eliminated nasal *S. aureus* colonization in three of four adults (47). Studies in mouse models further support potential benefits of nasal *Corynebacterium.* Nasal priming of infant mice with *C. pseudodiphtheriticum* 090104 decreases severity of respiratory syncytial virus infection and prevents secondary pneumococcal infection (48). Similarly, short-term respiratory tract colonization of mice with *C. accolens* reduces lung inflammation and the number of *S. pneumoniae* in both the upper and lower respiratory tract after a challenge (49).

Inhibition of *S. pneumoniae* or *S. aureus in vitro* by nasal *Corynebacterium* species displays strain-level variation, highlighting the importance of sequencing the genomes of multiple strains per species. For example, *in vitro* antipneumococcal activity varies between strains of nasal *Corynebacterium* isolates (13, 28). Also, *C. accolens* displays strain-level (and/or assay-based) variation with respect to inhibition (50) versus enhancement (14) of *S. aureus* growth *in vitro.* Some of these inhibitory interactions are characterized, including identification of the mechanism. *C. accolens* strains secrete the triacylglycerol lipase LipS1 to hydrolyze host-surface triacylglycerols releasing nutritionally required free fatty acids that also inhibit *S. pneumoniae in vitro* (22, 49). In contrast to this single-species inhibition, *C. pseudodiphtheriticum* KPL1989 only inhibits *S. pneumoniae in vitro* in cocultivation with *D. pigrum,* via an unknown mechanism (51). Whereas *C. pseudodiphtheriticum* strains (in the absence of *D. pigrum*) display contact-

independent killing of *S. aureus* that requires *S. aureus* production of phenol-soluble modulins, and resistant *S. aureus* mutants are less virulent (52). There are also non-inhibitory interactions with *S. aureus*. Under *in vitro* conditions where both species grow, both human nasal- and skin-associated *Corynebacterium* species excrete an as-yet-unidentified compound that inhibits *S. aureus agr*-quorum-sensing autoinducing peptides, shifting *S. aureus* towards a colonization state (53).

Nasal *Corynebacterium* also interact with other common commensal/mutualistic nasal microbionts. For example, *C. accolens*, *C. pseudodiphtheriticum,* and *C. propinquum* enhance the growth yield of *D. pigrum in vitro* (51); this might be due to a positive metabolic interaction, such as cross-feeding. *C. propinquum* encodes a biosynthetic gene cluster (BGC) for the siderophore dehydroxynocardamine that is transcribed *in vivo* in human nasal passages, and dehydroxynocardamine iron sequestration inhibits the growth of coagulase-negative *Staphylococcus* species *in vitro* (54).

Overall, their ubiquity, frequent positive associations with health, and potential therapeutic use raise fundamental questions about the role of *Corynebacterium* species in human nasal microbiota. To increase genomic and metabolic knowledge of these, we performed systematic phylogenomic and pangenomic analyses of four common human nasal-associated *Corynebacterium* species. To increase the generalizability of our findings, we analyzed genomes of 87 nasal strains collected across two continents and from a broad age range of children and adults. Nasal strains of *C. pseudodiphtheriticum* overwhelmingly partitioned into clades by country of origin, consistent with geographically restricted strain circulation. Comparison of the core versus accessory genome of each of these four *Corynebacterium* species demonstrated that all COG categories associated

with metabolism were enriched in the core genome, indicating limited strain-level metabolic variation within each species. Furthermore, a qualitative analysis of KEGG modules revealed that these four species share the majority of KEGG modules with few species, or even clade, specific metabolic abilities. However, we did find that the clade of *C. pseudodiphtheriticum* dominated by strains from the U.S. lacked the module for assimilatory sulfate reduction, which is key for biosynthesis of sulfur-containing amino acids. To provide broader context, we compare the predicted metabolic abilities of nasal *Corynebacterium* species to two well-studied *Corynebacterium* species, *C. diphtheriae* and *C. glutamicum*, and to common nasal species from other bacterial genera.

RESULTS

***Corynebacterium pseudodiphtheriticum* displays geographically restricted strain circulation.** To compare the genomic content and phylogenomic relationships among and within four *Corynebacterium* species commonly found in human nasal microbiota, we isolated strains from the nasal vestibule (nostrils) of generally healthy children and adults in the U.S. and from nasopharyngeal swabs of mother-infant pairs in Botswana. After whole-genome sequencing of selected isolates using Illumina sequencing (**Table S1**), we compared our 87 distinct nasal strain genomes to publicly available genomes of the type strain plus several other reference strains of each species, for a total of 20 reference genomes (54-59).

We confidently assigned each new nasal isolate to a species in two steps. First, we generated a maximum-likelihood phylogenomic tree based on 632 shared single-copy core gene clusters (GCs) from 107 strain genomes (**Fig. S1A**) and identified which were

in a clade with the type strain of each species (**Fig. S1B**; type strains in bold). *C. macginleyi* is the closest relative of *C. accolens* and these two species are challenging to distinguish by partial 16S rRNA gene sequences. Inclusion of three *C. macginleyi* genomes in this phylogenomic analysis afforded confident assignment of candidate *C. accolens* strains to a species. This overall phylogenomic tree (**Fig. S1B**) confirmed that *C. propinquum* and *C. pseudodiphtheriticum* are more closely related to each other whereas *C. macginleyi*, *C. accolens*, and *C. tuberculostearicum* are more closely related to each other, with *C. macginleyi* closest to, yet distinct from *C. accolens.* Second, we confirmed each strain had a pairwise average nucleotide identity (ANI) of ≥ 95% for core GCs compared to the type strain of its assigned species (**Fig. S2A**). For each species, the pairwise ANIs for core GCs were very similar to those for all shared CoDing Sequences (CDS) (**Fig. S2B**).

To assess the evolutionary relationships between nasal isolates from both the U.S. and Botswana, we produced individual maximum-likelihood phylogenomic trees for each species (**Fig. 1**) based on its conservative core genome defined by the consensus of the three algorithms cluster of orthologs triangles (COGS) (60), Markov Cluster Algorithm OrthoMCL (OMCL) (61), and bidirectional best-hits (BDBH) (**Fig. S2C**). These species-specific phylogenies provided a refined view of the relationships between strains based on the larger number of shared single-copy core GCs within each species (ranging from 1345 to 1788). To better approximate the root of each species-specific tree (**Fig. 1**), we used the type strain of the most closely related species in the multispecies phylogenomic tree (**Fig. S1**) as the outgroup (**Fig. S2D**). With a relatively even representation of Botswanan and U.S. strains (40% vs. 58%), the phylogenomic tree for *C.*

pseudodiphtheriticum had two large, well supported clades dominated respectively by nasal strains from Botswana (15/15) or from the U.S. (20/22), indicating a restricted distribution of strains by country (**Fig. 1B**). Of MSK-named strains, 86% were collected in Botswana and 14% in the U.S., whereas all KPL-named strains were collected in the U.S. We avoided calculating geographic proportions within major clades for *C. propinquum* (**Fig. 1A**) and *C. accolens* (**Fig. 1C**) because of the disproportionately high representation of U.S. strains (80%) and for *C. tuberculostearicum* (**Fig. 1D**) because there were only 6 nasal strains with 5 from Botswana. Within these limitations, the phylogenomic analysis of these three species revealed some remarkably similar strains present in samples collected in the U.S. and Botswana based on their residing together in terminal clades. This raises the possibility of wide geographic distribution of a subset of similar strains for each of these three species across Africa and North America.

The sizes of the core genomes of four common nasal *Corynebacterium* species have leveled off. Based on rarefaction analysis, the core genomes of *C. propinquum, C. pseudodiphtheriticum, C. accolens,* and *C. tuberculostearicum* have reached a stable size and are unlikely to decrease much further with the sequencing of additional strains (**Fig. 2 Ai-Aiv**). Based on the respective Tettelin curves (red line), the *C. tuberculostearicum* core genome stabilized first at ~7 genomes; however, with the fewest genomes at 8 this might be an upper bound that will continue to decrease with additional strain genomes. In comparison, *C. pseudodiphtheriticum* had the largest number of strain genomes at 43, with pairwise ANIs of ≥ 96.2% (**Fig. S2Aii**), and its core genome stabilized last at ~37 genomes.

The proportion of an individual genome of each of these nasal *Corynebacterium* species devoted to conservative core GCs ranged from 72% for *C. pseudodiphtheriticum* (1517/2105) to 79% for *C. tuberculostearicum* (1788/2250). This is based on the average number of CDS per genome (**Table 1**). (We estimated CDS using Prokka:Prodigal (62) and then estimated the percentage of each genome occupied by conservative core GCs using GET_HOMOLOGUES (63).) The average / median genome size for each species ranged from ~2.33 / 2.33 Mb for *C. pseudodiphtheriticum* to ~2.51 / 2.52 Mb for *C. propinquum* with the average / median predicted CDS per genome ranging from 2105 / 2096 for *C. pseudodiphtheriticum* to 2265 / 2272 for *C. propinquum* (**Table 1**). These sizes and proportions are consistent with the reduced genome size and the GCs per genome of host-associated compared to environment-associated *Corynebacterium* species (64). For comparison, average and/or median reported genomes sizes for common human nasal microbionts from other genera are as follows: *Cutibacterium acnes* 2.51 Mb (65), *Staphylococcus epidermidis* 2.5 Mb (66), *Lawsonella clevelandensis* 1.79 Mb (https://www.ncbi.nlm.nih.gov/genome/?term=Lawsonella+clevelandensis on 05/03/23), *Haemophilus influenzae* 1.84 Mb (67), *D. pigrum* 1.93 / 1.91 Mb (68), and *Staphylococcus aureus* 2.83 Mb (https://www.ncbi.nlm.nih.gov/genome/?term=Staphylococcus%20aureus%5BOrganism%5D&cmd=DetailsSearch on 05/03/23).

The pangenomes of these four human nasal-associated *Corynebacterium* species remain open. With the number of strain genomes analyzed (**Table 1**), the pangenome of each of the four species continued to increase with each additional new genome, indicating that all are open (**Fig. 2 Bi-Biv**). Parameters used to generate a pangenome

via rarefaction yielded an overly conservative estimate of its size in GCs. Therefore, we used two other approaches to estimate the number of GCs in the pangenome for each species. These pangenome composition estimates are a lower bound for each species and will increase with sequencing of additional strain genomes. Starting with GET_HOMOLOGUES, we estimated pangenome size using the COG triangle and OMCL clustering algorithms. The pangenome size and its proportion contributed by core versus accessory GCs for each species ranged from 3232 GCs with 56% core and 44% accessory for *C. tuberculostearicum* to 4590 GCs with 33% core and 67% accessory for *C. pseudodiphtheriticum* (**Table 2; Fig. S3A**). The 56% core percentage for *C. tuberculostearicum* is likely an overestimate since this pangenome is based on only 8 genomes. This range of 33% to 56% for core genes per pangenome is similar to estimates for other human upper respiratory tract microbionts, such as *D. pigrum* (31%) (68), *Staphylococcus aureus* (36%), and *Streptococcus pyogenes* (37%) (69).

Next, we used anvi'o version 7.1.2 to estimate the core- and pangenomes (70). The number of GCs in the core genome of each species estimated with GET_HOMOLOGUES was within 6-23% of those estimated with anvi'o; however, the GET_HOMOLOGUES estimated pangenome sizes were 11-32% larger (**Fig. S3A, File S1**). Consistent with this, the estimated single-copy core as a proportion of the pangenome using anvi'o was higher for each species ranging from 41% to 64% (**Fig. S3B**).

We also used anvi'o to visualize the strain-level variation in gene presence and absence within the four human nasal-associated *Corynebacterium* species (**Fig. S3B**). Manually arraying the genomes in anvi'o to correspond with their species-specific phylogenomic tree (**Fig. 1**) showed that some blocks of gene presence/absence correlated with the core-

genome-based phylogenetic relationships among strains, but others did not (**Fig. S3B**). This is consistent with gene gain and loss playing a role in strain diversification with some of this due to mobile genetic elements and horizontal gene transfer (71, 72).

Gene clusters assigned to the COG categories associated with metabolism are highly enriched in the core genomes of common nasal *Corynebacterium* species. To predict and compare functions based on the pangenomes of each species, we assigned GCs to COG categories and used PPanGGOLiN to define the persistent versus the accessory genome (**Table S2**) (73). (Note that the PPanGGOLiN definition of persistent genome differs slightly from the GET_HOMOLOGUES definition of core genome.) As is common in bacteria, only about 65% of the GCs in the persistent genome and 26-36% of the GCs in the accessory genome of each species had an informative assignment to a definitive COG category (**Figs. 3Ai-iv & S4Ai-iv**). There was also variability in the size of the accessory genome among strains within each species (**Fig. S4A, S4B**). We next generated functional enrichment plots for COG categories in the persistent versus the accessory genome of each species (**Fig. 3Bi-vi**). GCs assigned to "mobile genetic elements" (MGEs; orange bar **Fig. 3B**) were overrepresented in the accessory genome of each species with the ratio of GCs in the accessory/persistent genome ranging from 4.2 (*C. tuberculostearicum*) to 36.1 (*C. pseudodiphtheriticum*). GCs assigned to "defense systems" (purple bar **Fig. 3B**), which protect bacteria from MGEs, were more evenly distributed with the ratio of GCs in the accessory/persistent genome ranging from 1 (*C. tuberculostearicum*) to 2.9 (*C. pseudodiphtheriticum*). These findings are consistent with pangenomic analyses of other bacterial species, including our prior analysis of the candidate beneficial nasal bacterium *D. pigrum* (68). Our COG-enrichment

analysis also showed that all the COG categories associated with metabolism, from "energy production and conversion" (pale orange) through "secondary metabolites" (pink) in **Fig. 3B,** were highly overrepresented in the persistent (or core) genome of each species with ratios of accessory/persistent ranging from 0.02 to 0.56 (median of 0.16). The exception was an accessory/persistent GC ratio of 1.2 for "secondary metabolites" in *C. pseudodiphtheriticum*. The overrepresentation of metabolism in the persistent genome of each species points to limited strain-level variation in metabolic capabilities, such as "carbohydrate or amino acid metabolism". This contrasts with our previous analysis of *D. pigrum* in which GCs assigned to the COG category "carbohydrate transport and metabolism" are enriched in the accessory genome (ratio 1.66) (68).

Common human nasal-associated *Corynebacterium* species have a largely shared metabolic capacity. In adult nasal microbiota, based on 16S rRNA V1-V3 sequences, the prevalence of the genus *Corynebacterium* is as high as 98.6%, with highly prevalent species including *C. accolens* (prevalence of 82%), *C. tuberculostearicum* (93%), *C. propinquum* (18%), and *C. pseudodiphtheriticum* (20%). In these data, 82% of the adult nostril samples contained ≥ 2 of these 4 *Corynebacterium* species, 30% contained ≥ 3, and 2.4% contained all 4 species (Tables S4A-B and S7 in (31). Thus, there is a high probability of coexistence of these *Corynebacterium* species in nasal microbial communities. This finding, combined with the enrichment of GCs assigned to metabolism COG categories in each persistent genome, led us to hypothesize that there would be much species-specific variation in core metabolic capabilities enabling the different nasal *Corynebacterium* species to occupy distinct metabolic niches within human nasal microbiota. To focus only on these 4 human nasal-associated species, we excluded the

3 *C. macginleyi* reference genomes and used the remaining 104 genomes for all subsequent analyses. To test our hypothesis, we assessed variation in metabolic capacity among the four species using anvi'o v7.1.2 to estimate the metabolic capabilities described by Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic annotations. This yielded KEGG Orthology family (KOfam) annotations for GCs as well as module completion results for each KEGG module that had at least one KOfam present (**Table S3A**). We examined the number of complete KEGG pathway modules, defined as "functional units of gene sets in metabolic pathways, including molecular complexes" per https://www.genome.jp/kegg/module.html. In contrast to our hypothesis, we learned that *C. propinquum*, *C. pseudodiphtheriticum*, *C. accolens,* and *C. tuberculostearicum* encode highly conserved core metabolic capabilities sharing 50 of 66 (76%) detected complete KEGG modules, with completeness defined as detection of ≥ 0.75 of the genes in a module (**Fig. 4A**). Various combinations of three of the four species shared an additional seven modules (11%). There were a few differences between the *C. propinquum-C. pseudodiphtheriticum* clade and the *C. accolens-C. tuberculostearicum* clade (**Fig. S1B**), with three and five clade-specific KEGG modules, respectively. Only *C. tuberculostearicum*, which is broadly distributed across human skin sites as well as in the nasal passages (31, 34-36) was predicted to encode two complete KEGG modules that were absent in the other three nasal species; these are discussed below.

Nasal *Corynebacterium* species encode for central carbohydrate metabolism. To contextualize our findings within the genus *Corynebacterium*, we manually compared our results using genome annotation data from the KEGG database (not using anvi'o) to the genomes of the type strains of two well-studied *Corynebacterium* species: *C. glutamicum*

ATCC 13032 (*C. glutamicum*^{T}) (74) and *C. diphtheriae* NCTC11397 (*C. diphtheriae*^{T}). The soil bacterium, *C. glutamicum,* is used to produce amino acids and its metabolism is the best studied of *Corynebacterium* species (75). In contrast, *C. diphtheriae* colonizes the human pharynx and is extensively studied because toxigenic strains cause the human disease diphtheria (76). Along with *C. diphtheriae*^{T} and *C. glutamicum*^{T}**,** the four common human nasal *Corynebacterium* species all encoded complete modules for glycolysis, the pentose phosphate pathway, and the tricarboxylic acid (TCA) cycle (**Fig. 4B**). *C. propinquum* and *C. pseudodiphtheriticum*, along with *C. glutamicum*^{T} (which grows with acetate as a sole carbon source), also possessed a complete module for the glyoxylate cycle (M00012), a shortcut of the TCA cycle that allows microbes to utilize two-carbon compounds, such as acetate, as sources of carbon (77). This module is also present in skin-associated *Corynebacterium* species (78). Although the glyoxylate cycle might afford *C. propinquum* and *C. pseudodiphtheriticum* a distinct metabolic niche, two-carbon compounds are not detected in previous studies of the metabolites present in human nasal secretions (79, 80). Lastly, all four species encoded a complete module for UDP-N-acetyl-D-glucosamine biosynthesis (M00909), a precursor of cell wall peptidoglycan (81). Of note, we found 61 hits across 28 *D. pigrum* genomes and no hits in the *Corynebacterium* genomes for a sialidase (K01186), which can release sialic acid from sialylated glycans found in mucus providing bacteria with carbon and nitrogen (82).

Nasal *Corynebacterium* species encode for synthesis of key biosynthetic cofactors and electron transport chain components. The strain genomes of all four species contained complete modules for biosynthesis of cofactors required for synthesis of essential biomolecules and central metabolism (**Fig. 4B**; **Table 3**). These include

tetrahydrofolate (M00126), a coenzyme for synthesis of amino acids and nucleic acids; coenzyme A (M00120), required for the TCA cycle; and lipoic acid (M00881), an organosulfur cofactor required in central metabolism (83). Consistent with an intact TCA cycle, all four also had complete modules for the biosynthesis of key compounds involved in the electron transport chain, including menaquinone (M00116; this was incomplete in 3 strains: *C. accolens* ATCC49726, *C. propinquum* KPL3675, *C. pseudodiphtheriticum* 090104), heme/siroheme (M00121, M00868, M00926 / M00846), and riboflavin (M00125). We also detected the modules for coenzyme A, lipoic acid, and heme in both *C. glutamicum*^{T} and *C. diphtheriae*^{T}. Of the remaining modules detected in the nasal *Corynebacterium* species, more modules were shared with the human-associated bacterium *C. diphtheriae*^{T} (menaquinone, riboflavin, siroheme, molybdenum cofactor) than with the soil bacterium *C. glutamicum*^{T} (tetrahydrofolate). Of note, modules for the biosynthesis of cobalamin/Vitamin B12 (M00925, M00924, M00122) were incomplete or absent in all four nasal *Corynebacterium* species whereas both *C. glutamicum*^{T} and *C. diphtheriae*^{T} encoded a cobalamin biosynthesis module (M00122). Cobalamin is a cofactor for ribonucleotide reductases (K00525, K00526) (84), which we detected in all 104 *Corynebacterium* genomes.

Nasal *Corynebacterium* species share necessary modules for nucleotide synthesis and energy generation. All four of these common nasal *Corynebacterium* species had complete modules for synthesizing inosine monophosphate, adenine ribonucleotide, guanine ribonucleotide, uridine monophosphate, and pyrimidine deoxyribonucleotide (**Table 3**). All four species also had four complete modules for ATP synthesis (M00149, M00151, M00155, M00157) as well as a reductive pentose phosphate module, which

overlaps with reductive pentose phosphate modules mentioned earlier. Lastly, all four had complete modules for dTDP-L-rhamnose biosynthesis, a precursor to rhamnose cell wall polysaccharides. Rhamnose is part of the polysaccharide linker between peptidoglycan and arabinogalactan in members of *Mycobacteriales*, including *Corynebacterium* (81). Of the 11 KEGG modules related to nucleotide synthesis and energy generation shared by the four nasal *Corynebacterium* species, many are also present in other common nasal microbionts with 9/11 in *Cutibacterium acnes* KPA171202 (85), 9/11 in *Staphylococcus epidermidis* RP62A (86), 6/11 in *Streptococcus pneumoniae* TIGR4 (87), 5/11 in *S. aureus* USA300_FPR3757 (88), and 5/11 in a 27-strain *D. pigrum* pangenome (68), as well as 10/11 in *C. diphtheriae*^{T} and 9/11 in *C. glutamicum*^{T} (**Table 4**).

Human nasal *Corynebacterium* species have a broader metabolic capacity for biosynthesis of amino acids and cofactors/vitamins than *Dolosigranulum pigrum*. Many compositional studies of human nasal microbiota show a positive association at the genus level between *Corynebacterium* and *Dolosigranulum,* e.g., (1, 2, 4, 7, 24, 26, 51, 89). Presently, *D. pigrum* is the sole member of its genus. Nasal *Corynebacterium* species can enhance *in vitro* growth yields of *D. pigrum*, a lactic acid producing bacterium (51). Together with other prior analyses (90), these data indicate *D. pigrum* must access nutrients from its host and its microbial neighbors. We hypothesized that nasal *Corynebacterium* species with their larger genome sizes ranging from 2.3-to-2.6 Mb would have a greater number of complete KEGG modules than *D. pigrum* (1.9 Mb) in the pangenome for each species (**Fig. 5**). Using the enrichment analysis in anvi'o, we identified 30 complete modules shared by the majority (88%) of the 27 *D. pigrum* strain genomes (**Table S3B**), which is approximately half the number found in the majority of

the four *Corynebacterium* species' genomes (range 61-66). Only seven modules found in *D. pigrum* were undetectable in these four *Corynebacterium* species (**Table 5**), mostly related to carbohydrate metabolism. *D. pigrum* shared only about a third of the KEGG modules shared by all four *Corynebacterium* species (**Table 3**, highlighted in green). Specifically, the four *Corynebacterium* species shared 15 modules related to amino acid metabolism and 8 modules related to cofactor/vitamin metabolism in the majority of their genomes that were absent/incomplete in *D. pigrum* (**Table 3**; **Table S3B**). Overall, the four *Corynebacterium* species shared 50 complete KEGG modules with 15 for amino acids, 15 for carbohydrates, and 18 for cofactors and vitamins. In comparison, of the 30 complete KEGG modules identified in the majority of the *D. pigrum* genomes, there were 15 for carbohydrates and 4 for cofactors/vitamins. Previous genome-based estimations of *D. pigrum*'s metabolism predict auxotrophy for most (13/20) amino acids and synthesis of L-alanine, L-aspartate, L-asparagine, L-glutamine, glycine, L-serine, and L-tyrosine (90). In contrast, our current analysis, which includes additional strain genomes, found that 13-to-15 of these 27 genomes are predicted to synthesize 5 amino acids for which there are no KEGG modules: alanine, aspartate, asparagine, glutamine, and glycine. We did this by manually checking for KOfams encoding for enzymes that synthesize these amino acids. However, the KEGG modules for serine (M00020) (**Table S3B**) or tyrosine (M00025, M00040) biosynthesis were absent. *D. pigrum* also has predicted auxotrophies for polyamines (spermidine and putrescine); plus, many B vitamins, e.g., thiamine (vitamin B1), riboflavin (vitamin B2), niacin (vitamin B3), biotin (B7), and p-Aminobenzoate (PABA), a component of folate (vitamin B9) (51, 90). A potential limitation of this analysis is that, although representative members of *Lactobacillales* are well studied, it is possible

that current annotation software is more effective for *Corynebacterium* species than for *D. pigrum*.

***C. tuberculostearicum* is predicted to degrade glycogen unlike the other three nasal-associated *Corynebacterium* species.** Among the four nasal-associated species analyzed, *C. tuberculostearicum* was predicted to have two complete (≥ 0.75) KEGG modules absent from the other three (**Fig. 4A**): glycogen degradation (M00855) and UDP-glucose biosynthesis (M00549). To assess whether these two modules are present in other human skin-associated *Corynebacterium* species (78) or in non-*Corynebacterium* human nasal-associated bacterial species, we examined the KEGG database information for three common skin *Corynebacterium* species (*Corynebacterium simulans* PES1 (91), *Corynebacterium kroppenstedtii* DSM 44395 (92), and *Corynebacterium amycolatum* FDAARGOS_1108 (93)), and for the following common nasal species: *C. acnes* KPA171202 (85), *S. epidermidis* RP62A (86), *S. pneumoniae* TIGR4 (87), *S. aureus* USA300_FPR3757 (88), and *D. pigrum* (27-strain pangenome) (68) (**Table 6**). The glycogen degradation module (M00855) was detected in all 27 *D. pigrum* strain genomes, *S. epidermidis* RP62A, *C. acnes* KPA171202, and *S. pneumoniae* TIGR4, suggesting it is common in nasal microbiota, and in the skin-associated species *C. simulans* (and in *C. glutamicum*^{T}). However, *C. tuberculostearicum* and *D. pigrum* each lacked one of the four enzymes in the glycogen degradation module resulting in completion scores of 0.75. *C. tuberculostearicum* had the 4-alpha-glucanotransferase but lacked a glycogen debranching enzyme, whereas *D. pigrum* had a debranching enzyme but lacked the 4-alpha-glucanotransferase. Like *C. tuberculostearicum*, the KEGG database marks *C. glutamicum*^{T} as missing the gene for the glycogen debranching enzyme *glgX* (K02438,

455 EC:3.2.1.196). However, *C. glutamicum* both synthesizes and degrades glycogen when
456 grown in media with glucose and its GlgX performs glycogen debranching *in vitro* (94)
457 (**Fig. 4Ci**). Additionally, disruption of *glgX* in *C. glutamicum* has phenotypic effects related
458 to glycogen usage. The functional *C. glutamicum* GlgX has only 46% identity to the
459 functional *E. coli* GlgX (94). A blastP of the translated 836 amino acid sequence of *glgX*
460 from *C. glutamicum*^{T} against the non-redundant protein database returned hits in *C.*
461 *tuberculostearicum* at 73% identity and 85% query coverage, along with hits to multiple
462 other *Corynebacterium* species (**Table S3C**). This is consistent with published data from
463 Seibold and Eikmanns. A custom blastx of our 104 genomes against that same amino
464 acid sequence returned a single hit in each of the 8 *C. tuberculostearicum* genomes, at
465 72-73% identity along 86% of the length of GlgX (**Table S3D**). This predicts that, like *C.*
466 *glutamicum*, *C. tuberculostearicum* degrades glycogen, an ability shared with other skin
467 *Corynebacterium* and with some non-*Corynebacterium* nasal bacteria. The other 3 nasal
468 species analyzed lacked all steps in this module, including GlgX, indicating they either
469 use a currently unrecognized set of enzymes for glycogen degradation or lack this ability.
470 As with *cysH* (discussed below), the current KEGG annotations miss the *glgX* gene that
471 encodes the glycogen debranching enzyme in *C. glutamicum*^{T} and *C. tuberculostearicum*,
472 and likely other *Corynebacterium* species.

473 We next assessed the completion of the glycogen synthesis module because capacity for
474 both synthesis and degradation points to the use of glycogen for energy storage. The
475 glycogen synthesis module (M00854; **Fig. 4Cii**) was only 2/3 complete in all 8 *C.*
476 *tuberculostearicum* genomes, *C. diphtheriae*^{T}, and *C. glutamicum*^{T}. According to KEGG,
477 they are all missing a glycogen synthase (for which there are many orthologs listed in

KEGG including *glgA* EC 2.4.1.21). However, *C. glutamicum*^{T} is known to possess *glgA* (95). In contrast, this module was complete in 26/27 *D. pigrum* genomes, *S. epidermidis* RP62A, *C. acnes* KPA171202, and *S. pneumoniae* TIGR4. A custom blastx of our *Corynebacterium* genomes against the translated *glgA* amino acid sequence from *C. glutamicum*^{T} gave 8 hits, one in each of the *C. tuberculostearicum* genomes with 73-74% identity along its full length (390 aa) (**Table S3E**). Thus, it is likely that *C. tuberculostearicum* uses glycogen synthesis and degradation for energy storage similarly to *C. glutamicum*^{T}.

The UDP-glucose biosynthesis module (M00549) fully detected in *C. tuberculostearicum* (completion score = 1) was 2/3 complete in the other nasal *Corynebacterium* genomes and in the *D. pigrum* genomes, but present in the other species analyzed (**Table 6**). UDP-glucose is a key part of central metabolism. It is the activated form of glucose that serves as a precursor for other activated carbohydrates and is used by a vast majority of organisms for glucosyl transfers. Phosphoglucomutase (*pgm*) performs the second step in its three-step biosynthesis module (**Fig. 4Ciii)**. *C. tuberculostearicum* encoded *pgm*, which was absent in the other nasal *Corynebacterium* species. In contrast, the *D. pigrum* genomes lacked the third step, a UDP-sugar pyrophosphorylase or a UTP--glucose-1-phosphate uridylyltransferase. Upon blastx of the *pgm* amino acid sequence from *C. glutamicum*^{T} against the custom *Corynebacterium* genomes the only 8 hits were in each of the *C. tuberculostearicum* genomes with 78% identity along 537-553 of 554 amino acids (**Table S3F**). A blastx of the translated *pgm* amino acid sequence from *C. tuberculostearicum* KPL3807 (as identified from the Prokka annotation) against the custom *Corynebacterium* genomes only had hits among the *C. tuberculostearicum* with

96-97% identity along 536-538 of 538 amino acids (**Table S3G**). Since nucleotide sugars are so universal, either UDP-galactose is sufficient for the other three *Corynebacterium* species, or they encode an enzyme other than Pgm that performs the second step in this module. Our KEGG analysis supports the need for future experimental metabolic studies with these *Corynebacterium* species.

Nasal *Corynebacterium* species can synthesize most essential amino acids. This analysis estimated that all analyzed strains of these four common human nasal *Corynebacterium* species can synthesize 15 of the 20 standard amino acids. We detected complete biosynthetic modules for 11 amino acids in all 104 genomes, including the hydrophobic amino acids (isoleucine, leucine, methionine, valine, and tryptophan); the polar uncharged amino acids (threonine and serine); the charged amino acids (glutamate, arginine, histidine, lysine); and proline (**Table 3; Fig. 4B**). In these *Corynebacterium*, biosynthesis of methionine and arginine depends on aspartate as a precursor. For the five amino acids without KEGG modules, a manual search in KEGG and BioCyc databases for biosynthetic enzyme KOfams predicted that all four nasal *Corynebacterium* species plus *C. glutamicum*^{T} and *C. diphtheriae*^{T} can generate the following three additional amino acids: alanine, glutamine, and glycine (**File S1**).

For aspartate and asparagine, the remaining two without their own biosynthetic KEGG modules, this search identified asparagine synthase (K01953), which interconverts these amino acids, in all of the genomes, so if one is available exogenously the other could be made. Of the pathways on KEGG for generating aspartate (https://www.genome.jp/pathway/map00250), these nasal *Corynebacterium* species only had KOfam hits for this interconversion of aspartate and asparagine. This implies nasal

Corynebacterium species require an exogenous source of either aspartate or asparagine, but not both. KEGG lists similar annotations for interconversion in *C. glutamicum*^{T}.

Our analysis predicted that all four common human nasal *Corynebacterium* species are auxotrophic for phenylalanine (M00024, M00910) and tyrosine (M00025, M00040). The biosynthetic KEGG modules listed for phenylalanine and tyrosine were incomplete/absent in the 104 nasal *Corynebacterium* genomes examined and in *C. glutamicum*^{T} and *C. diphtheriae*^{T}. The pangenomes of the four nasal *Corynebacterium* species all lacked matches to enzymes needed for the last step in each module for synthesizing either phenylalanine (M00024, M00910) or tyrosine (M00024, M00040) from chorismate. The same two possible KOfam groups were missing: aromatic amino acid aminotransferase I / 2-aminoadipate transaminase (K00838) or aromatic-amino-acid transaminase (K00832).

Our analysis predicted all four species require exogenous phenylalanine, tyrosine, and either aspartate or asparagine. Of these four, mass spectrometry (MS) of aspirated human nasal secretions detects only phenylalanine (79); however, MS of nasal fluid collected with synthetic absorptive matrix strips detects aspartate, asparagine, phenylalanine, and tyrosine (80). Therefore, all four amino acids appear to be present in nasal secretions.

Predicted synthesis of methionine and cysteine varied by species. All genomes had a complete module to convert glycerate-3P to serine (M00020), to convert serine to cysteine (M00021), and to convert aspartate to methionine (M00017). However, all the genes in assimilatory sulfate reduction module M00176 were absent in 30 of the 43 *C.*

pseudodiphtheriticum genomes (all the U.S. strains and four strains from Botswana), and from *C. diphtheriae*^{T}. When performing an enrichment analysis within each of the four *Corynebacterium* species, only M00176 was significantly enriched in the Botswanan strains of 43 *C. pseudodiphtheriticum* (**Table S3H**). No other significant differences in KEGG module enrichment were detected in the other species' pangenome (**Table S3H-K**). Assimilatory sulfate reduction is key to the production of cysteine and methionine since it takes environmental sulfate and converts it to a usable form in the cell (96). The complete absence of all the components of this KEGG module implies this subset of *C. pseudodiphtheriticum* strains cannot synthesize methionine or cysteine. Of note, *C. glutamicum*^{T}, which lacks only the adenylyl sulfate kinase (*cysC*) component of this module, has proven assimilatory sulfate reduction capabilities via the *fpr2-cysIXHDNYZ* gene cluster (97). Moreover, all *C. accolens,* all *C. tuberculostearicum,* most *C. propinquum* (17/19), and 13 of 43 *C. pseudodiphtheriticum* genomes (13/17 strains from Botswana) had a completion score between 0.83 and 0.89 for this assimilatory sulfate reduction module (M00176). Like *C. glutamicum*^{T}, these other strains all lacked the adenylylsulfate kinase *cysC* (K00860) that converts adenosine phosphosulfate into phosphoadenosine phosphosulfate (**Fig. 4Civ**), but our KEGG analysis detected *cysN*, *cysD*, *cysH*, and *SIR* (K00956, K00957, K00390, K00392). *C. glutamicum cysH* encodes an adenosine 5'-phosphosulfate (APS) reductase that can complement an *E. coli cysC*-deficient mutant (97). However, KEGG annotations currently lack the option for *cysH* to substitute for *cysC*, which explains why there is experimental evidence that *C. glutamicum*^{T} performs assimilatory sulfate reduction but the KEGG module is marked as incomplete. Based on blastp results against the non-redundant protein sequences (nr)

database, orthologs of *C. glutamicum* CysH are overwhelmingly present in the Order *Mycobacteriales*, which includes *Corynebacterium* (**Table S3L**). When we performed a targeted blastx on our analyzed *Corynebacterium* genomes using the translated amino acid sequence of *C. glutamicum cysH*, the percent identity was 68-70% for *C. propinquum* and a subset of *C. pseudodiphtheriticum* strains and 58-61% for *C. accolens* and *C. tuberculostearicum* strains along the full length (228-231) of the CysH amino acid sequence (**Table S3M**). Thus, many human nasal *Corynebacterium* strains likely also perform assimilatory sulfate reduction using CysH rather than CysC (**Fig. 4Civ**). These findings also indicate a recent and geographically localized loss of the assimilatory sulfate reduction module M00176 within the *C. pseudodiphtheriticum* phylogeny.

Methionine might also be acquired from the environment. The KEGG database contains three KOfam entries for methionine transporter genes: *metN* (K02071), *metI* (K02072), and *metQ* (K02073), which encode for the methionine ABC-transporter (MetNI) and substrate binding protein (MetQ). We found hits for all 3 of these genes in all 104 *Corynebacterium* and 27 *D. pigrum* strains, often with multiple hits per gene, confirming our earlier results from Brugger et al. (51) and predicting these nasal *Corynebacterium* can transport methionine from their environment.

DISCUSSION

Here, we analyzed strain genomes of four common *Corynebacterium* species including those of 87 distinct human nasal isolates collected in Africa and North America across the human lifespan. Phylogenomic analysis showed *C. pseudodiphtheriticum* displays geographically restricted strain circulation and this corresponded with a recent

geographically restricted loss of the KEGG module for assimilatory sulfate reduction in strains isolated in the U.S, since this module was present in the other three species and most *C. pseudodiphtheriticum* strains from Botswana. Across the four species, genomic analysis revealed average genome sizes of 2.3 to 2.5 Mb, with the average CDS per genome ranging from 2105 to 2265 and with 72-79% of each individual genome encoding GCs of the shared conservative core genome of the respective species. For each species, the core genome size had leveled off while the pangenome remained open. An informative assignment to a definitive COG category was possible only for approximately 65% of the GCs in the persistent genome and 26-36% of the GCs in the accessory genome of each species, which points to the need for ongoing experimental research to identify the function of many bacterial GCs. GCs assigned to the COG categories for metabolism were enriched in the persistent genome of each species and all four species shared the majority (50 of 66) of complete KEGG modules identified, which implies limited strain- and species-level metabolic variation restricting the possibilities for strains to occupy distinct metabolic niches during the common occurrence of cocolonization of the human nasal passages. *Corynebacterium* species are often positively associated with *Dolosigranulum* in human nasal microbiota. We found human nasal *Corynebacterium* species have a broader metabolic capacity for biosynthesis of amino acids and cofactors/vitamins than *D. pigrum*. Our findings combined with data showing that the majority of adults likely host at least two *Corynebacterium* in their nasal passages points to the importance of future investigation into how *Corynebacterium* species interact with each other and with other microbionts in human nasal microbiota.

Our analysis predicts common human nasal *Corynebacterium* species all produce some compounds that *D. pigrum* cannot. This, combined with the positive association of these species in compositional studies, supports the possibility that *Corynebacterium* might cross-feed or serve as a source of nutrients for *D. pigrum,* and possibly other microbionts, in human nasal microbiota. These compounds include hydrophobic amino acids (isoleucine, leucine, methionine, valine and tryptophan); polar uncharged amino acids (threonine); charged amino acids (glutamate, arginine, histidine, lysine), and proline; along with the following cofactors/vitamins: biotin, heme, siroheme, lipoic acid, molybdenum cofactor, NAD, menaquinone, pantothenate, pyridoxal 5'-phosphate, riboflavin, and thiamine.

De novo cobamide biosynthesis is enriched in host-associated compared to environment-associated *Corynebacterium* species, with several human skin-associated *Corynebacterium* species encoding complete biosynthesis pathways (64, 78). Swaney, Sandstrom and Kalan used KOfamScan to identify the presence or absence of cobamide biosynthesis genes within genomes of diverse *Corynebacterium* species, exclusive of the four examined here (64). Checking for those same genes, we found only hits for genes encoding enzymes involved in biosynthesis of tetrapyrrole precursors (*hemA, hemL, hemB, hemC, hemD,* and *cobA*) and cobalamin nucleotide loop assembly (*cobC*) in our 104 *Corynebacterium* genomes. Moreover, none of the three strongly nasal-associated *Corynebacterium* species (*C. accolens, C. propinquum* or *C. pseudodiphtheriticum*) nor the common nasal- and skin-associated species, *C. tuberculostearicum*, encoded complete cobamide-production pathways (**Table S4**). Three incomplete modules for synthesis of cobalamin (vitamin B_{12}), a member of the cobamide family of compounds,

were detected in the 104 *Corynebacterium* genomes: 1) cobalamin biosynthesis (M00122) for cobyrinate a,c-diamide → cobalamin; 2) cobalamin biosynthesis, anaerobic (M00924) for uroporphyrinogen III → sirohydrochlorin → cobyrinate a,c-diamide; and 3) cobalamin biosynthesis, aerobic (M00925) for uroporphyrinogen III → precorrin 2 → cobyrinate a,c-diamide. Completeness scores for all three modules were very low, ranging between 0.083 and 0.286. In contrast, examining the four *Corynebacterium* species pangenomes for cobamide-dependent enzymes identified hits for ribonucleotide reductases (K00525, K00526) (84, 98) in all 104 *Corynebacterium* genomes. By analyzing the extent and distribution of cobamide production in 11,000 bacterial species, Shelton, Taga, and colleagues report 86% of bacteria in their dataset have at least 1 of the 15 cobamide-dependent enzyme families, but only 37% are predicted to synthesize their own cobamide (84). These data are consistent with human nasal *Corynebacterium* species requiring an exogenous source of cobamide compounds. *C. acnes*, which produces the cobamide vitamin B12 (cobalamin) (99) and is a highly prevalent and abundant member of human nasal microbiota, especially in adults (31), is a possible source.

Limitations of this study include the uneven representation of strains from the U.S. and Botswana for *C. propinquum* and *C. accolens*; the limited number of *C. tuberculostearicum* strains; the inherent limitations of KEGG analysis; and the predictive nature of genome-based metabolic estimation that requires future experimental validation. To our knowledge, this analysis includes the largest number of strain genomes for *C. pseudodiphtheriticum, C. propinquum*, and *C. accolens* to date, with the aforementioned smaller number for *C. tuberculostearicum*. However, compared to the

thousands of strain genomes that have been analyzed for nasal pathobionts, e.g., *S. aureus* (100), there are still a limited number of available genomes of nasal *Corynebacterium* species and a need to build large strain collections of human-associated *Corynebacterium* species to better assess the potential use of these strains for the promotion of human health. Similarly, although we included genomes for strains isolated from two continents and from a range of ages, the geographic sampling was limited compared to the distribution of human populations globally and there has yet to be a systematic large-scale sampling of nasal microbiota across the human lifespan.

Qualitatively, we isolated fewer *C. tuberculostearicum* from nasal swabs than expected based on its prevalence and relative abundance estimated in our earlier reanalysis of 16S rRNA gene V1-V3 sequences from human nasal samples (31). In contrast to using a single gene, here, we assigned isolates to *C. tuberculostearicum* based on WGS with an ANI of ≥ 95% to the type strain *C. tuberculostearicum* DSM 44922. Only a subset of our isolates from the US and Botswana with partial 16S rRNA gene Sanger sequences (approximately V1-V3) matching to *C. tuberculostearicum* met this criterion after WGS. This points to the existence of another common nasal *Corynebacterium* species that is closely related to *C. tuberculostearicum.* Recent human skin metagenomic analyses by Salamzade, Kalan and colleagues identify metagenome-assembled genomes and the strain genome LK1134 with ANI ≥ 95% to the genome called "*Corynebacterium kefirresidentii"* (101), which is not a validly published species, and show via phylogenomic analysis these are closely related to *C. tuberculostearicum* (34). Furthermore, using metagenomic analyses, they show sequences matching the "*C. kefirresidentii"* genome are more prevalent on nasal and nearby facial skin sites, whereas *C. tuberculostearicum*

is more prevalent and at higher relative abundance on foot-associated skin sites (34). Future work to validly name the species currently identified by the genome called "*C. kefirresidentii*" with designation and deposition of a type strain in publicly accessible stock centers is critical for experiments seeking to identify the function of this species in human nasal microbiota.

Isolation and whole genome sequencing of multiple strains for microbial species commonly detected in the human microbiome, such as this one, is an ongoing effort across multiple body sites. Collections of genome-sequenced strains from the microbiota are a critical resource for experimentally testing hypotheses generated from metagenomic and metatranscriptomic studies to identify the functions of human microbionts and mechanisms by which they persist in the microbiome. The Human Oral Microbiome Database (eHOMD) is an early and ongoing example of a body-site focused resource for the human microbiome based on a combination of culture-dependent and -independent data (102). Originally focused on the oral cavity, eHOMD now serves the full human respiratory tract (31). More recently, Saheb Kashaf, Finn, and colleagues established the Skin Microbial Genome Collection (SMGC), a combined cultivation- and metagenomic-based resource for the human skin microbiome (78). These well-curated, body-site focused databases serve a critical role in advancing microbiome research, including their importance in shedding light on so-called microbial and metagenomic "dark matter." The data we presented here serves as a foundational resource for future genomic, metagenomic, phenotypic, metabolic, functional, and mechanistic research on the role of nasal *Corynebacterium* species in human development and health.

MATERIALS AND METHODS

Collecting new nasal *Corynebacterium* sp. isolates. The U.S. *Corynebacterium* strains with KPL in their name were collected in Massachusetts, USA under a protocol approved by the Forsyth Institutional Review Board (FIRB #17-02) as described previously (68). In brief, adults and children participating in scientific outreach events in April 2017 and 2018 performed supervised self-sampling of their nostrils (nasal vestibule) with sterile swabs. They then inoculated their swab onto up to two types of agar medium: 1) brain heart infusion with 1% Tween80 (BHIT) and 25 microgram/ml fosfomycin (BHITF25) to enrich for *Corynebacterium* sp. and/or 2) BBL Columbia Colistin-Nalidixic acid agar with 5% sheep's blood (CNA BAP). These were incubated at 37°C for 48 hrs under either atmospheric (BHITF25) or 5% CO_2-enriched (CNA BAP) conditions. We selected colonies with a morphology typical of nasal *Corynebacterium* species and passed each two to three times for purification on BHIT with 100 ug/ml fosfomycin (BHITF100) at 37°C prior to storage in medium with 15-20% glycerol at -80°C. (Isolates from 2017 were picked from growth on BHITF100 at 37°C under atmospheric conditions that had been inoculated from sweeps of the original mixed growth on agar medium and stored at -80°C.)

The majority of the *Corynebacterium* strains with MSK in their name were cultured from nasopharyngeal swab samples collected from mothers and infants in a birth cohort study conducted in Botswana, as previously described (13), with a small number also collected from mid-turbinate nasal swab samples from patients cared for within the Duke University Health System (MSK074, MSK075, MSK076, MSK079, and MSK080). This work was reviewed and considered exempt by the Duke Health Institutional Review Board (Pro00102629). Bacteria were cultivated and isolated as previously described (13).

Selection of nasal *Corynebacterium* isolates for Illumina sequencing. For each KPL-named new isolate, Sanger sequencing (Macrogen, USA) was performed using primer 27F on a V1-V3 16S rRNA gene colony-PCR amplicon (GoTaq Green, Promega) of primers 27F and 519R. We assigned each initial isolate to a genus and a putative species based on blastn of each sequence against eHOMDv15.1 (31). We then selected a subset of these isolates for whole genome sequencing (WGS). For MSK-named new isolates, all isolates preliminarily assigned to *Corynebacterium* based on MALDI and/or Sanger sequencing of V1-V3 16S rRNA gene underwent WGS.

Genomic DNA extraction. We extracted genomic DNA (gDNA) from the KPL-named U.S. strains using the MasterPure Gram Positive Genomic DNA Extraction Kit with the following modifications to the manufacturer's protocol: 10 mg/mL lysozyme treatment at 37°C for 10 min and 2x 30 sec bead beat in Lysing Matrix B tubes (MP Biomedicals) at setting 6 on a FastPrep24 (MP Bio) with 1-minute interval on ice. To assess gDNA quality, we performed electrophoresis on 0.5% TBE agarose gel, used a NanoDrop spectrophotometer to quantify 260/280 and 260/230 ratios, and used a Qubit Fluorometric Quantification (Invitrogen) to measure concentration. We extracted gDNA from the MSK-named strains collected in Botswana and North Carolina using the Powersoil Pro extraction kit (Qiagen) following the manufacturer’s instructions. DNA concentrations were determined using Qubit dsDNA high-sensitivity assay kits (Thermo Fisher Scientific).

Whole genome sequencing and assembly. For the KPL-named U.S. isolates, Nextera XT (Illumina) libraries were generated from gDNA. Each isolate was sequenced using a paired-end 151-base dual index run on an Illumina Novaseq6000 at the NIH Intramural

Sequencing Center. The reads were subsampled to achieve 80x coverage and then assembled with SPAdes (version 3.13.0) (103) and polished using Pilon (version 1.22) (104). For the MSK-named isolates, which are mostly from Botswana, library preparation was performed using DNA Prep Kits (Illumina) and these libraries were sequenced on a NovaSeq 6000 instrument (Illumina) configured for 150 base pair paired-end reads. Adapter removal and read trimming were performed using Trimmomatic version 0.39 (105) to a Phred score of 30 across a 4-bp sliding window. Surviving reads shorter than 70 bp were discarded. The final quality of reads was assessed using FastQC version 0.11.9. Assembly was performed using SPAdes version 3.15.3 (106). The completeness of the genomes was evaluated with checkM version 1.1.3 (107) and all genomes with a completeness less than 95% were discarded. Genomic data are deposited under BioProjects PRJNA842433 for the KPL-named isolates (which are a subset of 94 *Corynebacterium* isolated in MA, USA) and PRJNA804245 for the MSK-named isolates (which are a subset of 71 genomes isolated from Botswana and the Duke University Health System). Table S1 includes GenBank IDs.

Selection of strain genomes for pangenomic analysis. To the 165 assemblies mentioned in the previous section, we added another 16 KPL-named *Corynebacterium* sp. nasal-isolate genomes originally sequenced as part of the HMP and deposited by the Broad at NCBI to consider for analysis (108). Furthermore, 31 reference assemblies for relevant *Corynebacterium* species, including the genome of the type strain of *C. propinquum*, *C. pseudodiphtheriticum*, *C. accolens,* and *C. tuberculostearicum* plus 3 strain genomes of *C. macginleyi,* were downloaded from NCBI using the PanACoTA v1.4.1 (109) `prepare -s` subcommand. We used default parameters such that genomes

with MASH distances to the type strain outside of the range 1e-4 to 0.06 were discarded to avoid redundant pairs or mislabeled assemblies and low-quality assemblies based on L90 ≤ 100 and number of contigs ≤ 999 were filtered out. The collected 212 assemblies were filtered using the `prepare --norefseq` subcommand as above to select higher quality assemblies (L90 ≤ 100 and number of contigs ≤ 999) and to eliminate redundant genomes defined by a MASH distance < 10^{-4} keeping the genome with the highest quality score from each redundant set. Finally, we confirmed the species-level assignment of our nasal isolates, and the nontype reference strains, based on an ANIb (nucleotide) of ≥ 95% for all shared CDS regions compared to the respective type strain of each species using GET_HOMOLOGUES (see below). For each species, this resulted in a set of distinct strain genomes (including the type strain) that we used for subsequent analyses, which totaled to 104 genomes: 19 *C. propinquum* genomes, 43 *C. pseudodiphtheriticum* genomes, 34 *C. accolens* genomes, and 8 *C. tuberculostearicum* genomes. Table S1 contains (A) a list of these 104 strain genomes selected for further analysis plus 3 *C. macginleyi* reference strains, and (B) the all-by-all MASH distance analysis result of the PanACoTA analysis for all 107 genomes.

Determination of the conservative core genome. We annotated all bacterial genomes with Prokka version 1.14.6 (62) with default parameters, including gene recognition and translation initiation site identification using Prodigal (110). Then, we used the `./get_homologues.pl` command from GET_HOMOLOGUES version 24082022 (63, 111) to determine a conservative core genome for each of the selected four *Corynebacterium* species based on the consensus of three algorithms: bidirectional best-hits (BDBH), cluster of orthologs triangles (COGS) v2.1 (60), and Markov Cluster Algorithm OrthoMCL

(OMCL) v2.4 (61) (**Fig. S1A, Fig. S2C**). Each of the three algorithms reported clustering at the protein level using blastp from NCBI BLAST v2.2 (112) with `-C 90` (min % coverage in BLAST pairwise alignments). The data output created from running the three different clustering algorithms was used to identify the intersection of the core GCs with the command `./compare_clusters.pl` with `-t # of genomes`. We ran this last command twice, with and without the -n flag to generate both nucleotide and protein outputs. Additional methods for genome annotations (https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_Prokka_Annotations.html) and for the determination of the conservative core genome are available online (https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_GET_HOMOLOGUES.html#conservative-core-and-phylogeny-of-all-107-corynebacterium-genomes).

Determination of core, soft core, shell, and cloud genomes. For this analysis to identify the core genomes, we dropped the strains Cps_090104 and Cac_ATCC_49726 from their respective species because, although these both passed CheckM analysis, their inclusion in the initial rarefaction analyses caused a splitting of data points towards the lower end resulting in an aberrant lower bound. We separately analyzed the genomes of each *Corynebacterium* species using GET_HOMOLOGUES with the command `./get_homologues.pl` and flag`-t 0` (to get all the possible clusters), first with `-M` (OrthoMCL) and a second time with `-G` (COGS). The COGS and OMCL results were then used with `./compare_clusters.pl` with `-m` (produce intersection pangenome matrices). Then the command `./parse_pangenome_matrix.pl` was used with `-s` (report

core, soft core, shell, and cloud clusters), and `-x` (produce matrix of intersection pangenome clusters) (**Fig. S3 Ai-Aiv**). In-depth code and descriptions to determine the core, soft core, shell, and cloud genomes are available online (https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_GET_HOMOLOGUES.html#create-a-pangenome-matrix-for-each-of-the-four-corynebacterium-species).

Calculating the conservative core genome for each species with the addition of an outgroup. We chose the type strain genome of the most closely related species from the full phylogenomic tree in **Fig. S1** to serve as the outgroup in each species-specific phylogeny with *C. pseudodiphtheriticum* DSM 44287 (GCF_000688415.1) for the *C. propinquum* phylogeny, *C. propinquum* DSM 44285 (GCF_000375525.1) for the *C. pseudodiphtheriticum* phylogeny*, C. macginleyi* CCUG 32361 (GCF_003688935.1) for the C. *accolens* phylogeny, and *C. accolens* ATCC 49725 (GCF_000159115.1) for the *C. tuberculostearicum* phylogeny. Addition of an outgroup resulted in four new datasets with n + 1 total genomes, where n denotes the number of strain genomes for each species. Each species-plus-outgroup dataset of genomes was analyzed as before using the commands `. /get_homologues.pl` (with BDBH, COG triangles, and OrthoMCL) and `./compare_clusters.pl`, but now with the outgroup genome included. (**Fig. S2D**). For construction of single-species trees (**Fig. 1**), we then combined this smaller conservative core of GCs shared between the species and its outgroup (shared intersection of each Venn diagram in Fig. S2D) with the subset of GCs belonging only to the conservative core of each species (which is the intersection of each Venn diagram in Fig. 2SC minus the intersection of those in Fig. 2SD). The code used to do this is available online

(https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_GET_HOMOLOGUES.html#generate-corynebacterium-species-specific-phylogenomic-trees-with-an-outgroup).

Construction of phylogenomic trees. We used GET_PHYLOMARKERS v.2.2.9.1 to concatenate and align the single copy core GCs evaluated for each species from GET_HOMOLOGUES (**Fig. S2C, S2D**) (113). The command `run_get_phylomarkers_pipeline.sh` was run with the flags: `-R 1 -t DNA -k 0.7 -m 0.7` on both the protein and nucleotide GET_HOMOLOGUES outputs. The flag `-R` was used to select optimal markers for phylogenomics; `-t` for whether the input is DNA or protein; `-k` for kde stringency; and `-m` for the minimum average support values for trees to be selected. The codon fasta alignments generated by GET_PHYLOMARKERS were analyzed with IQ-TREE v2.1.3 (114) with: `-p`, (uses edge linked partition model and ModelFinder functions (115-117)), `-alrt 1000` (perform replicate SH-like approximate likelihood ratio test) and `-B 1000` (number of ultrafast bootstrap replicates). The phylogenetic tool iTOL v6 (118) was used to visualize, scale, edit, annotate names, and root the tree at the midpoint for each phylogeny. Detailed code and methods to create these phylogenies are available online (https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_GET_HOMOLOGUES.html#concatenate-and-align-files-using-get_phylomarkers-for-each-of-the-four-corynebacterium-species).

Rarefaction analysis and Average Nucleotide Identity (ANI) across strain genomes. For each species, with GET_HOMOLOGUES, we used rarefaction analysis to estimate whether the core genome and pangenome were closed or open (**Fig. 2**). We modified

"$MIN_PERSEQID_HOM" and "$MIN_COVERAGE_HOM" values to equal 50 (inside "marfil_homology.pm") so only protein sequences with identity and coverage ≥ 50% will be called homologues.

Again, we made the decision to eliminate the two genomes, namely Cac_ATCC_49726 and Cps_090104 from this dataset. The reason for this action was that their core genome size resulted in the preliminary graph getting divided while conducting random sampling. This is because their core genome sizes were considerably lower, resulting in a misleading lower bound. The clustering was redone for each species using `./get_homologues.pl` and the following parameters: `-C 90`, `-c`, (genome composition analysis) and `-M` for OMCL. The rarefaction curve .tab files were produced from the `-c` flag. Rarefaction .tab files for each species were plotted into svg files using `/plot_pancore_matrix.pl` with flags `-f core_both` (displays both Tettelin and Willenbrock curves for core genome), and `-f pan` (curve for estimating pangenome size). Blastn instead of blastp was used to report GCs for ANI heat plots. To generate the core and the all shared CDS regions ANI .tab files, we used the command `./get_homologues.pl` with flags: `-d`, `-A`, `-t`, `-a`, `-M`. Furthermore, to plot the ANI heatmaps the `./plot_matrix_heatmap.sh` command was used.

Functional analysis of four *Corynebacterium* pangenomes using anvi'o and PPanGGOLiN. The pangenome for each *Corynebacterium* species was analyzed using anvi'o version 7.1.2 (70, 119). We used a workflow that allowed us to import Prokka annotated genomes into anvi'o (https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_Prokka_Annotations.html#prokka-annotation-for-anvio), followed by the addition of functional COG

annotations using the `anvi-run-ncbi-cogs` command with --sensitive flag (runs sensitive version of DIAMOND (120)) and the 2020 updated COG20 database (121, 122). Various annotations were also applied to each genome db file, such as KEGG/KOfam (123, 124), Pfam (125) and hmm-hits (126). The pangenome for each species and a combined pangenome for all the 104 strains was computed with the `anvi-pan-genome` command (flags: --mcl-inflation 10, and --use-ncbi-blast) using blastp search (127), muscle alignment (128) `minbit heuristic` (129) to filter weak hits, and the MCL algorithm (130). The functional and the geometric homogeneity index, and the rest of the layers shown within the anvi'o pangenome *Corynebacterium* figures were determined from the standard anvi'o pangenomic pipeline (https://merenlab.org/2016/11/08/pangenomics-v2/). The core (genes contained in all genomes), soft core (genes contained in 95% of the genomes), shell (genes contained in several genomes), and cloud (genes present in only a few genomes) assignments from GET_HOMOLOGUES were uploaded into the anvi'o pangenome for each species by creating bins in the anvi'o interactive interface. We also manually rearranged the order of strains in each species-specific analysis to match the order in our species-specific phylogenomic trees and imported our phylogenies into the anvi'o figures. This facilitated visualizing the strain genomes from an evolutionary standpoint and identifying GCs that might be unique to a strain or a group of strains in a clade. The code used to generate the anvi'o pangenomic analysis can be found at https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_Anvio.html. We then exported the summary files from the anvi'o pangenomic analyses to synchronize gene cluster identities with PPanGGOLiN v1.2.74 (**Table S2**) (for detailed code see https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_PPanGGOLi

N.html#ppanggolin-pangenome-analysis). GCs were defined as persistent or accessory by PPanGGOLiN and then we used an in-house R script (https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_COGS.html) to clean up and retrieve informative COG20 annotated GCs (122, 131) and to generate the functional enrichment plots shown in **Fig 3** and **Fig S4**.

Estimation of metabolic capabilities using anvi'o v7.1.2. We then used a workflow from anvi'o (https://merenlab.org/tutorials/infant-gut/#chapter-v-metabolism-prediction) that allowed us to estimate the completeness of KEGG modules in each of the four *Corynebacterium* pangenomes and in the combined pangenome for the 104 strains. To get a detailed KEGG metabolic output, we used the `anvi-estimate-metabolism` function using the flag `–kegg-output-modes modules, kofam_hits` to get details about the module completeness and KOFam hits in each pangenome. Additional detailed methods and code are available online (https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_Anvio.html#anvio-kegg-analysis-pipeline). We also estimated the metabolic capabilities of a pangenome of 27 *D. pigrum* strain genomes (68) and compared the predicted metabolic modules with those of the four *Corynebacterium* pangenomes. For the R code for generating **Fig. 5** in more detail, see https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_Anvio.html#make-the-distribution-plot.

Determining KEGG module enrichment. We used the `anvi-compute-metabolic-enrichment` function to identify which complete KEGG modules were enriched in some species within the combined pangenome (**Table S3A**) or in geographic sample groups

within each species **(Table S3H-K)**. Modules with an adjusted p-value < 0.05 were considered enriched in their associated group and modules that were not significantly enriched (adjusted p-value > 0.05) in comparisons in three out of the four species, were categorized as shared. We also performed a KEGG module enrichment analysis of 104 human nasal *Corynebacterium* genomes with 27 *D. pigrum* strain genomes (68) (**Table S3B**). Additional detailed methods and code are available online (https://klemonlab.github.io/CorPGA_Pangenomics/SupplementalMethods_Anvio.html#anvio-kegg-analysis-pipeline).

Comparisons with single representative strain genomes from KEGG. We manually compared the KEGG modules enriched in the four common nasal *Corynebacterium* species to those modules present in the strains listed in Table 7.

Blast analyses with specific genes. Using blastx, we queried each of the 104 human nasal *Corynebacterium* genomes against the translated amino acid sequences of the genes *glgX* (**Table S3D**), *glgA* (**Table S3E**), *pgm* (**Table S3F**) and *cysH* (**Table S3M**) from *C. glutamicum*^{T} (GCA_000011325.1) and *pgm* (**Table S3G**) from *C. tuberculostearicum* KPL3807. We performed blastx using ncbi-blast-2.13 with default settings. We filtered results by e-value < 0.05 for significant hits, hit length > 100 (since all sequences examined were ≥ 100 aa in length), and percent identity > 50% for similar sequences. We also performed a blastP on the NBCI site against the non-redundant protein sequences (nr) database (all non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects) using default settings (except for the number of results as the top 5000) of the

translated 836 amino acid sequence of *glgX* (**Table S3C**) and the translated 231 amino acid sequence of *cysH* (**Table S3L**) from *C. glutamicum*^{T}.

Data and code availability. All genomes are available from NCBI in Bioprojects PRJNA842433 and PRJNA804245. Table S1 lists the accession number for each *Corynebacterium* strain genome used in this study. Code used can be found at our GitHub repository https://github.com/KLemonLab/CorPGA_Pangenomics.

Acknowledgments. Thank you to all the individuals who donated nostril swab samples at a 2017 and 2018 science festival that were used to isolate KPL strains of *Corynebacterium* species. Their contributions continue to expand knowledge about human nasal microbiota. Thank you to lab members and colleagues who contributed to these outreach events, including Javier Fernandez Juarez, Kerry Maguire, Genevieve Holmes, Pallavi Murugkar, Pooja Balani, Sowmya Balasubramanian, Fan Zhu, Andrew Collins, Andy Kempczynski, Brian Klein, and Megan Lambert. For early planning and initial cultivation efforts we thank Silvio D. Brugger, Lindsey Bomar, Stephany Flores Ramos, and Sara M. Eslami for KPL strains; and Jhoanna N. Aquino and Christopher R. Polage for MSK strains. We also thank all of the individuals in the U.S. and Botswana who participated in the research studies that resulted in cultivation of the MSK strains. Thank you to Clay Deming and the NIH Intramural Sequencing Center for sequencing of the KPL strains. Thank you to Bruno Contreras-Moreira for advice on using GET_HOMOLOGUES; to A. Murat Eren for advice on using anvi'o; and to Ashlee M. Earl for early efforts to genome sequence nasal *Corynebacterium* isolates. Thank you to other members of our labs.

Authorship contributions: Conceptualization: KPL, WG, IFE. New Strain isolation: WG. Investigation: THT, AQR, WG, IFE, SC MSK. Bioinformatic analysis: THT, AQR, IFE, SC. Interpretation of data: THT, AQR, IFE, WG, JAS, HHK, SC, MSK, KPL. Visualization: THT, AQR, IFE. Wrote Original Draft: THT, AQR, KPL. Editing and review: KPL, MSK, SC, IFE, THT, AQR, WG, HHK, JAS. All authors read and approved the final manuscript. Supervision: KPL, MKS, HHK, JAS. Funding Acquisition: KPL, MSK, HHK, JAS.

Funding information: This work was supported by the National Institutes of Health through the National Institute of General Medical Sciences (grants R35 GM141806 and R01 GM117174 to K.P.L) and the National Institutes of Allergy and Infectious Diseases (grant K23 AI135090 to M.S.K.); and through the Intramural Research Programs of the National Human Genome Research Institute (J.A.S.) and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (H.H.K.). Additional funding was provided by the Forsyth Institute through a Pilot Grant (FPILOT45 to K.P.L.).

Table 1. Basic genomic information for four common human nasal-associated *Corynebacterium* species.

Corynebacterium species	**# Strain Genomes (# Nasal isolates)**	**Average (median) genome size (Mb)**	**Average (median) CDS/genome**	**Average (median) G+C%**	**Conservative Core GCs/species[a]**	**Pangenome GCs/species[b]**
C. propinquum	19 (15)	2.51 (2.52)	2265 (2272)	56.47 (56.48)	1623	3777
C. pseudodiphtheriticum	43 (39)	2.33 (2.33)	2105 (2096)	55.29 (55.29)	1345	4590[c]
C. accolens	34 (32)	2.50 (2.49)	2304 (2294)	59.46 (59.44)	1561	4220[d]
C. tuberculostearicum	8 (6)	2.39 (2.39)	2250 (2253)	59.86 (59.88)	1788	3232

[a]GET_HOMOLOGUES conservative core GCs predicted from the consensus of BDBH, OMCL, and COGS algorithms.
[b]GET_HOMOLOGUES pangenome GCs predicted from the consensus of OMCL, and COGS algorithms.

[c]Cps_090104 removed from dataset for this analysis due to false % core lower bound.
[d]Cac_ATCC_49756 removed from dataset for this analysis due to false % core lower bound.

Table 2. Pangenomic estimation of human nasal-associated *Corynebacterium* species based on three different platforms.

Platform Species	Pangenome size (GCs)	% Core GCs/pangenome	% Accessory GCs/pangenome
GET_HOMOLOGUES[a]			
C. propinquum	3777	43%	57%
C. pseudodiphtheriticum[c]	4590	33%	67%
C. accolens[d]	4220	40%	60%
C. tuberculostearicum	3232	56%	44%
anvi'o			
C. propinquum	3108	59%	40%
C. pseudodiphtheriticum	3590	48%	51%
C. accolens	3427	57%	42%
C. tuberculostearicum	2907	66%	32%
PPanGGOLiN		% Persistent	
C. propinquum	[b]	63%	37%
C. pseudodiphtheriticum	[b]	49%	51%
C. accolens	[b]	59%	41%
C. tuberculostearicum	[b]	69%	31%

[a]GET_HOMOLOGUES pangenome size, % core, and % accessory are from the consensus of OMCL and COGS algorithms.
[b]Pangenome size was estimated in anvi'o then the GCs imported into PPanGGOLiN to estimate persistent vs. accessory genome percentages.
[c]Cps_090104 was removed from dataset only for this analysis due to an aberrant % core lower bound.
[d]Cac_ATCC_49756 was removed from dataset only for this analysis due to an aberrant % core lower bound.

Table 3. *D. pigrum* shares a subset (green) of the predicted 50 complete KEGG modules present in the genomes of all four *Corynebacterium* species.

KEGG module	**KEGG Module Name**	**KEGG Module Sub-Category**
M00015	Proline biosynthesis	Amino acid metabolism
M00016	Lysine biosynthesis	Amino acid metabolism
M00017	Methionine biosynthesis	Amino acid metabolism
M00018	Threonine biosynthesis	Amino acid metabolism
M00019	Valine/isoleucine biosynthesis	Amino acid metabolism
M00021	Cysteine biosynthesis	Amino acid metabolism

M00022	Shikimate pathway	Amino acid metabolism
M00023	Tryptophan biosynthesis	Amino acid metabolism
M00026	Histidine biosynthesis	Amino acid metabolism
M00045	Histidine degradation into glutamate	Amino acid metabolism
M00432	Leucine biosynthesis	Amino acid metabolism
M00526	Lysine biosynthesis	Amino acid metabolism
M00527	Lysine biosynthesis	Amino acid metabolism
M00570	Isoleucine biosynthesis	Amino acid metabolism
M00844	Arginine biosynthesis	Amino acid metabolism
M00149	Succinate dehydrogenase, prokaryotes	ATP synthesis
M00151	Cytochrome bc1 complex respiratory unit	ATP synthesis
M00155	Cytochrome c oxidase, prokaryotes	ATP synthesis
M00157	F-type ATPase, prokaryotes and chloroplasts	ATP synthesis
M00632	Galactose degradation, Leloir pathway	Carbohydrate metabolism
M00909	UDP-N-acetyl-D-glucosamine biosynthesis	Carbohydrate metabolism
M00167	Reductive pentose phosphate cycle	Carbon fixation
M00009	Citrate cycle (TCA cycle Krebs cycle)	Central carbohydrate metabolism
M00011	Citrate cycle second carbon oxidation	Central carbohydrate metabolism
M00001	Glycolysis (Embden-Meyerhof pathway)	Glycolysis
M00002	Glycolysis	Glycolysis
M00003	Gluconeogenesis	Glycolysis
M00086	beta-Oxidation, acyl-CoA synthesis	Lipid metabolism
M00116	Menaquinone biosynthesis	Metabolism of cofactors & vitamins
M00120	Coenzyme A biosynthesis	Metabolism of cofactors & vitamins
M00121	Heme biosynthesis plants and bacteria	Metabolism of cofactors & vitamins
M00125	Riboflavin biosynthesis plants and bacteria	Metabolism of cofactors & vitamins
M00126	Tetrahydrofolate biosynthesis	Metabolism of cofactors & vitamins
M00846	Siroheme biosynthesis	Metabolism of cofactors & vitamins
M00868	Heme biosynthesis	Metabolism of cofactors & vitamins
M00880	Molybdenum cofactor biosynthesis	Metabolism of cofactors & vitamins
M00881	Lipoic acid biosynthesis plants and bacteria	Metabolism of cofactors & vitamins

M00926	Heme biosynthesis bacteria	Metabolism of cofactors & vitamins
M00140	C1-unit interconversion, prokaryotes	Metabolism of cofactors and vitamins
M00048	Inosine monophosphate biosynthesis	Nucleotide metabolism
M00049	Adenine ribonucleotide biosynthesis	Nucleotide metabolism
M00050	Guanine ribonucleotide biosynthesis	Nucleotide metabolism
M00051	Uridine monophosphate biosynthesis	Nucleotide metabolism
M00053	Pyrimidine deoxyribonucleotide biosynthesis	Nucleotide metabolism
M00554	Nucleotide sugar biosynthesis (galactose)	Nucleotide metabolism
M00004	Pentose phosphate pathway	Pentose phosphate pathway
M00005	PRPP biosynthesis	Pentose phosphate pathway
M00006	Pentose phosphate pathway	Pentose phosphate pathway
M00007	Pentose phosphate pathway	Pentose phosphate pathway
M00793	dTDP-L-rhamnose biosynthesis	Polyketide sugar unit biosynthesis

Table 4. Comparison of biosynthetic nucleotide modules found in the four common human nasal *Corynebacterium* to representative strains of other nasal species[a].

KEGG modules	*Sau*	*Cacn*	*Spn*	*Dpi*	*Sep*	*Cgl*	*Cdi*
inosine monophosphate (M00048)	+	+	+	+	+	+	+
adenine ribonucleotide (M00049)	+	+	+	+	+	+	+
guanine ribonucleotide (M00050)	+	+	+	+	+	+	+
uridine monophosphate (M00051)	ND	ND	ND	+	ND	+	ND
pyrimidine deoxyribonucleotide (M00053)	+	+	+	ND	+	+	+
ATP synthesis (M00149	ND	+	ND	ND	+	ND	ND
M00151	ND	+	ND	ND	+	+	+
M00155	ND	+	ND	ND	+	+	+
M00157)	+	+	+	+	+	+	+

reductive pentose phosphate (M00167)	ND	ND	ND	ND	ND	ND	ND
dTDP-L-rhamnose (M00793)	ND	+	+	ND	+	+	+

+ = Detected, ND = Not Detected. [a]*Sau* = *Staphylococcus aureus* USA300_FPR3757, *Cacn* = *Cutibacterium acnes* KPA171202, *Spn* = *Streptococcus pneumoniae* TIGR4, *Dpi* = *D. pigrum* 27-strain pangenome, *Sep* = *Staphylococcus epidermidis* RP62A, *Cgl* = *C. glutamicum*[T] ATCC 13032, *Cdi* = *C. diphtheriae*[T] ATCC 700971.

Table 5. *D. pigrum* has eight complete KEGG modules that are absent from the four nasal *Corynebacterium* species analyzed.

KEGG module	KEGG Module Name	KEGG Module Sub-Category
M00159	V/A-type ATPase, prokaryotes	ATP synthesis
M00008	Entner-Doudoroff pathway, glucose-6P → glyceraldehyde-3P + pyruvate	Central carbohydrate metabolism
M00308	Semi-phosphorylative Entner-Doudoroff pathway, gluconate → glycerate-3P	Central carbohydrate metabolism
M00082	Fatty acid biosynthesis, initiation	Fatty acid metabolism
M00854	Glycogen biosynthesis, glucose-1P → glycogen/starch	Other carbohydrate metabolism
M00550	Ascorbate degradation, ascorbate → D-xylulose-5P	Other carbohydrate metabolism
M00061	D-Glucuronate degradation, D-glucuronate → pyruvate + D-glyceraldehyde 3P	Other carbohydrate metabolism

Table 6. Presence of the two KEGG modules detected in *C. tuberculostearicum* but not in *C. accolens*, *C. propinquum*, or *C. pseudodiphtheriticum* in other skin-associated *Corynebacterium* species and common nasal bacterial species.

species	Glycogen degradation (M00855)	UDP-glucose (M00549)
C. simulans PES1	+	+

C. kroppenstedtii DSM 44395	ND	+
C. amycolatum FDAARGOS_1108	ND	+
C. glutamicum[T] ATCC 13032	+	+
C. diphtheriae[T] NCTC11397	+	+
S. aureus USA300_FPR3757	ND	+
S. epidermidis RP62A	+	+
D. pigrum[a]	+	ND
C. acnes KPA171202	+	+
S. pneumoniae TIGR4	+	+

+ = Detected, ND = Not Detected. [a]Assessed 27-strain pangenome.

Table 7. Single representative strain genomes of other species for KEGG comparisons.

Genome	Reference	Link to KEGG Modules
C. glutamicum[T] ATCC 13032	(132)	https://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_modules&org=cgl
C. diphtheriae[T] NCTC11397	NA	https://www.kegg.jp/kegg-bin/show_organism?menu_type=pathway_modules&org=cdip
C. simulans PES1	(91)	https://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_modules&org=csp
C. kroppenstedtii DSM 44385	(92)	https://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_modules&org=ckp
C. amycolatum FDAARGOS_1108	(93)	https://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_modules&org=camc
C. acnes KPA171202	(85)	https://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_modules&org=pac
S. aureus subsp. USA300_FPR3757	(88)	https://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_modules&org=saa
S. epidermidis RP62A	(86)	https://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_modules&org=ser

S. pneumoniae TIGR4	(87)	https://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_modules&org=spn

SUPPLEMENTAL MATERIALS

File S1 (text). Supplemental methods and materials

Figure S1. Phylogenomic tree

Figure S2. Pairwise ANIs and conservative core GCs

Figure S3. Pangenome estimates

Figure S4. COG analysis of accessory genome by strain.

Table S1 (excel file). Strain genome details for the four human nasal-associated *Corynebacterium* species analyzed.

Table S2 (excel file). PPanGGOLiN assignment of GCs to the persistent or accessory genome used in the COG functional enrichment analysis.

Table S3 (excel file). Results from KEGG module enrichment analysis by anvi'o and blast analyses with specific genes.

Table S4 (excel file). Information on KEGG modules for cobamide biosynthesis in the nasal *Corynebacterium* species.

Supplemental File S1

Species assignment for *Corynebacterium* sp. KPL# previously submitted to NCBI.

The Broad Institute originally sequenced a set of *Corynebacterium* strains we had isolated from human nostril swabs, and we included in our current analyses any of these that were assigned to one of the four species of interest (108). The species assignment for each of

these based on phylogenomic analyses (**Fig. 1; Fig. S1**) and ANI (**Fig. S2A, S2B**) are as follows: *C. accolens* KPL1818; *C. accolens* KPL1824; *C. accolens* KPL1996 (*C. accolens* KPL1986, KPL1998, and KPL2004); and *C. pseudodiphtheriticum* KPL1989 (*C. pseudodiphtheriticum* KPL1995). Isolates in parentheses are genomes listed in NCBI that have a MASH distance ≤ 10^{-4} to the strain genome listed immediately before the parentheses.

Comparing core genomes and pangenomes predicted by GET_HOMOLOGUES and anvi'o. The difference in the outputs of gene clusters (GCs) from GET_HOMOLOGUES and anvi'o were measured using the percent difference formula $\frac{|V1-V2|}{\left[\frac{(V1+V2)}{2}\right]} \times 100 = \%\ difference$. Core and pangenomes sizes for both the pangenomic platforms are listed below for comparison, with their percentage differences.

Species	Core genome			Pangenome		
	GH	anvi'o	% diff.	GH	anvi'o	% diff.
C. propinquum	1637	1824	10.8%	3777	3108	19.4%
C. pseudodiphtheriticum	1360	1714	23.0%	4978	3590	32.4%
C. accolens	1579	1904	18.7%	4488	3427	26.8%
C. tuberculostearicum	1802	1915	6.1%	3232	3232	10.6%

Additional Information on Amino Acids. We manually detected amino acid synthesis capabilities for which there are no KEGG modules and found the following in the 104 nasal *Corynebacterium* genomes: alanine dehydrogenase (K00259) for pyruvate ←→ alanine; glutamine synthetase (K01915) for glutamate → glutamine; glycine hydroxy

methyltransferase (K00600) for serine ←→ glycine. We also detected ornithine biosynthesis (M00028) and ornithine is precursor for the biosynthesis of arginine (M00844). Both *C. glutamicum*^{T} and *C. diphtheriae*^{T} have a different enzyme for making alanine (alanine-synthesizing transaminase (K14260)), and the same capabilities for glycine and ornithine.

Other amino acid metabolism modules that were incomplete or absent in all 104 *Corynebacterium* genomes included the following: tyrosine degradation (M00044), homoprotocatechuate degradation (M00533), degradation of leucine (M00036), degradation of lysine (M00032) and hydroxyproline degradation (M00948). In these modules, the products can either directly enter the TCA cycle (fumarate, acetyl-CoA, 2-Oxoglutarate) or are intermediates (2-Oxohept-3-enedioate, acetoacetate) that can be converted into products (succinate, acetyl-CoA) that enter the TCA cycle. Thus, the absence/incompleteness of these four modules suggest that these four common nasal *Corynebacterium* cannot use these amino acids as a source of carbon/energy.

All 104 *Corynebacterium* genomes contained a complete module for histidine degradation into glutamate. However, none had a complete GABA (gamma-Aminobutyrate) shunt module (M00027), which converts glutamate into succinate to enter the TCA cycle. In contrast, *C. glutamicum*^{T} can degrade tyrosine (M00044), methionine (M00035), histidine (M0045), and leucine (M00036), whereas *C. diphtheriae*^{T} lacks these degradation modules.

All *Corynebacterium* had a complete module for methionine synthesis (M00017) starting from aspartate, whereas *D. pigrum* lacks a complete pathway for methionine

biosynthesis. Furthermore, Krismer et al. report that in their synthetic nasal medium (SNM3) *S. aureus* relies heavily on *de novo* methionine biosynthesis and that *S. aureus* from human noses has increased transcript levels of *metI*, which encodes a cysteine-gamma-synthase and is required for methionine biosynthesis (79).

Lastly, the four *Corynebacterium* species lacked the module encoding the degradation of methionine (M00035) into cystathionine. Cystathionine can be converted into cysteine in M00338 or M00609. Additionally, when looking at KEGG annotations in 27 *D. pigrum* genomes (the initial 11 from (51) plus 16 others (68)), the M00035 module for degrading methionine to cystathionine is incomplete in all. The methionine degradation module (M00035) is defined by the following enzymes: MetK, DMNT1 or DMNT3A or DMNT3B, AHCY, CBS or cysO, which does not contain any of the genes found by Brugger et al. Moreover, the methionine salvage module (M00034) is defined by a lot of enzymes, none of which include the ones we previously found in Brugger et al. Among other species commonly present in the human nasal passages, examination of representative genomes for the methionine degradation module (M00035) had mixed results; *S. aureus* USA300_FPR3757 (88) and *Streptococcus pneumoniae* TIGR4 (87) both lacked M00035, whereas *Cutibacterium acnes* KPA171202 (85) and *Staphylococcus epidermidis* RP62A (86) encoded M00035.

Polyamine synthesis. Modules for the biosynthesis of spermidine (M00133), GABA (M00136, M00135) and putrescine (M00134) were incomplete (M00133) or absent (M00134, M00136, M00135) in the human nasal *Corynebacterium* pangenomes, suggesting none can synthesize these polyamines. *D. pigrum* is also predicted to lack biosynthesis of spermidine and putrescine (51, 90).

Cofactors and Vitamins. All of the nasal *Corynebacterium* had complete module for biosynthesis of molybdenum cofactor (M00880), which is a cofactor for molybdenum enzymes in redox reactions (133). Thiamine is a cofactor to enzymes involved in carbohydrate and branched-chain amino acid metabolism (134). *C. accolens* and *C. tuberculostearicum* had a completion score of 0.8 for the thiamine biosynthesis for archaea module (M00896). However, both lacked high completion scores for the thiamine biosynthesis module for prokaryotes (M00895). Completion scores for M00895 were highest (0.72) in all *C. accolens* and 10 *C. pseudodiphtheriticum* genomes and lowest at 0.44 in all *C. tuberculostearicum* genomes, with the remaining genomes at 0.61. In contrast, thiamine salvage (M00899) is complete in all *C. accolens* and *C. tuberculostearicum* and incomplete in the rest. *C. accolens* and *C. tuberculostearicum* also encoded biosynthesis of pyridoxal 5'-phosphate (M00916), which is a coenzyme in many transamination reactions (135). This module was also found in *C. glutamicum*^{T} and *C. diphtheriae*^{T}. *C. propinquum,* and *C. pseudodiphtheriticum* (and *C. glutamicum*^{T}) had the module for pantothenate biosynthesis (M00119), which is a key precursor for the biosynthesis of coenzyme A and carrier proteins that have a phosphopantetheine prosthetic group (136). The ubiquinone module (M00117), an electron transport chain component, is incomplete/absent in all four *Corynebacterium* species.

REFERENCES

1. Biesbroek G, Bosch AA, Wang X, Keijser BJ, Veenhoven RH, Sanders EA, Bogaert D. 2014. The impact of breastfeeding on nasopharyngeal microbial communities in infants. Am J Respir Crit Care Med 190:298-308.

2. Biesbroek G, Tsivtsivadze E, Sanders EA, Montijn R, Veenhoven RH, Keijser BJ, Bogaert D. 2014. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. Am J Respir Crit Care Med 190:1283-92.
3. Mika M, Mack I, Korten I, Qi W, Aebi S, Frey U, Latzin P, Hilty M. 2015. Dynamics of the nasal microbiota in infancy: a prospective cohort study. J Allergy Clin Immunol 135:905-912 e11.
4. Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, Holt BJ, Hales BJ, Walker ML, Hollams E, Bochkov YA, Grindle K, Johnston SL, Gern JE, Sly PD, Holt PG, Holt KE, Inouye M. 2015. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. Cell Host Microbe 17:704-15.
5. Peterson SW, Knox NC, Golding GR, Tyler SD, Tyler AD, Mabon P, Embree JE, Fleming F, Fanella S, Van Domselaar G, Mulvey MR, Graham MR. 2016. A Study of the Infant Nasal Microbiome Development over the First Year of Life and in Relation to Their Primary Adult Caregivers Using cpn60 Universal Target (UT) as a Phylogenetic Marker. PLoS One 11:e0152493.
6. Bosch AATM, Levin E, Van Houten MA, Hasrat R, Kalkman G, Biesbroek G, De Steenhuijsen Piters WAA, De Groot P-KCM, Pernet P, Keijser BJF, Sanders EAM, Bogaert D. 2016. Development of Upper Respiratory Tract Microbiota in Infancy is Affected by Mode of Delivery. EBioMedicine 9:336-345.
7. Bosch A, de Steenhuijsen Piters WAA, van Houten MA, Chu M, Biesbroek G, Kool J, Pernet P, de Groot PCM, Eijkemans MJC, Keijser BJF, Sanders EAM, Bogaert D. 2017. Maturation of the Infant Respiratory Microbiota, Environmental Drivers, and Health Consequences. A Prospective Cohort Study. Am J Respir Crit Care Med 196:1582-1590.
8. Shilts MH, Rosas-Salazar C, Tovchigrechko A, Larkin EK, Torralba M, Akopov A, Halpin R, Peebles RS, Moore ML, Anderson LJ, Nelson KE, Hartert TV, Das SR. 2016. Minimally Invasive Sampling Method Identifies Differences in Taxonomic Richness of Nasal Microbiomes in Young Infants Associated with Mode of Delivery. Microb Ecol 71:233-42.
9. Chu DM, Ma J, Prince AL, Antony KM, Seferovic MD, Aagaard KM. 2017. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. Nat Med 23:314-326.
10. Salter SJ, Turner C, Watthanaworawit W, de Goffau MC, Wagner J, Parkhill J, Bentley SD, Goldblatt D, Nosten F, Turner P. 2017. A longitudinal study of the infant nasopharyngeal microbiota: The effects of age, illness and antibiotic use in a cohort of South East Asian children. PLoS Negl Trop Dis 11:e0005975.
11. Ta LDH, Yap GC, Tay CJX, Lim ASM, Huang CH, Chu CW, De Sessions PF, Shek LP, Goh A, Van Bever HPS, Teoh OH, Soh JY, Thomas B, Ramamurthy MB, Goh DYT, Lay C, Soh SE, Chan YH, Saw SM, Kwek K, Chong YS, Godfrey KM, Hibberd ML, Lee BW. 2018. Establishment of the nasal microbiota in the first 18 months of life: Correlation with early-onset rhinitis and wheezing. J Allergy Clin Immunol 142:86-95.

12. Toivonen L, Hasegawa K, Waris M, Ajami NJ, Petrosino JF, Camargo CA, Jr., Peltola V. 2019. Early nasal microbiota and acute respiratory infections during the first years of life. Thorax 74:592-599.
13. Kelly MS, Plunkett C, Yu Y, Aquino JN, Patel SM, Hurst JH, Young RR, Smieja M, Steenhoff AP, Arscott-Mills T, Feemster KA, Boiditswe S, Leburu T, Mazhani T, Patel MZ, Rawls JF, Jawahar J, Shah SS, Polage CR, Cunningham CK, Seed PC. 2022. Non-diphtheriae Corynebacterium species are associated with decreased risk of pneumococcal colonization during infancy. ISME J 16:655-665.
14. Yan M, Pamp SJ, Fukuyama J, Hwang PH, Cho DY, Holmes S, Relman DA. 2013. Nasal microenvironments and interspecific interactions influence nasal microbiota complexity and *S. aureus* carriage. Cell host & microbe 14:631-40.
15. Ramakrishnan VR, Feazel LM, Gitomer SA, Ir D, Robertson CE, Frank DN. 2013. The microbiome of the middle meatus in healthy adults. PLoS One 8:e85507.
16. Biswas K, Hoggard M, Jain R, Taylor MW, Douglas RG. 2015. The nasal microbiota in health and disease: variation within and between subjects. Front Microbiol 9:134.
17. Kaspar U, Kriegeskorte A, Schubert T, Peters G, Rudack C, Pieper DH, Wos-Oxley M, Becker K. 2016. The culturome of the human nose habitats reveals individual bacterial fingerprint patterns. Environ Microbiol 18:2130-42.
18. Wos-Oxley ML, Chaves-Moreno D, Jauregui R, Oxley AP, Kaspar U, Plumeier I, Kahl S, Rudack C, Becker K, Pieper DH. 2016. Exploring the bacterial assemblages along the human nasal passage. Environ Microbiol 18:2259-71.
19. De Boeck I, Wittouck S, Wuyts S, Oerlemans EFM, van den Broek MFL, Vandenheuvel D, Vanderveken O, Lebeer S. 2017. Comparing the Healthy Nose and Nasopharynx Microbiota Reveals Continuity As Well As Niche-Specificity. Front Microbiol 8:2372.
20. Laufer AS, Metlay JP, Gent JF, Fennie KP, Kong Y, Pettigrew MM. 2011. Microbial communities of the upper respiratory tract and otitis media in children. mBio 2:e00245-10.
21. Pettigrew MM, Laufer AS, Gent JF, Kong Y, Fennie KP, Metlay JP. 2012. Upper respiratory tract microbial communities, acute otitis media pathogens, and antibiotic use in healthy and sick children. Appl Environ Microbiol 78:6262-70.
22. Bomar L, Brugger SD, Yost BH, Davies SS, Lemon KP. 2016. *Corynebacterium accolens* Releases Antipneumococcal Free Fatty Acids from Human Nostril and Skin Surface Triacylglycerols. mBio 7:e01725-15.
23. Kelly MS, Surette MG, Smieja M, Pernica JM, Rossi L, Luinstra K, Steenhoff AP, Feemster KA, Goldfarb DM, Arscott-Mills T, Boiditswe S, Rulaganyang I, Muthoga C, Gaofiwe L, Mazhani T, Rawls JF, Cunningham CK, Shah SS, Seed PC. 2017. The Nasopharyngeal Microbiota of Children With Respiratory Infections in Botswana. Pediatr Infect Dis J 36:e211-e218.
24. Hasegawa K, Linnemann RW, Mansbach JM, Ajami NJ, Espinola JA, Petrosino JF, Piedra PA, Stevenson MD, Sullivan AF, Thompson AD, Camargo CA, Jr. 2017. Nasal Airway Microbiota Profile and Severe Bronchiolitis in Infants: A Case-control Study. Pediatr Infect Dis J 36:1044-1051.
25. Kelly MS, Surette MG, Smieja M, Rossi L, Luinstra K, Steenhoff AP, Goldfarb DM, Pernica JM, Arscott-Mills T, Boiditswe S, Mazhani T, Rawls JF, Cunningham CK,

Shah SS, Feemster KA, Seed PC. 2018. Pneumococcal Colonization and the Nasopharyngeal Microbiota of Children in Botswana. Pediatr Infect Dis J 37:1176-1183.

26. Lappan R, Imbrogno K, Sikazwe C, Anderson D, Mok D, Coates H, Vijayasekaran S, Bumbak P, Blyth CC, Jamieson SE, Peacock CS. 2018. A microbiome case-control study of recurrent acute otitis media identified potentially protective bacterial genera. BMC Microbiol 18:13.
27. Man WH, Clerc M, de Steenhuijsen Piters WAA, van Houten MA, Chu M, Kool J, Keijser BJF, Sanders EAM, Bogaert D. 2019. Loss of Microbial Topography between Oral and Nasopharyngeal Microbiota and Development of Respiratory Infections Early in Life. Am J Respir Crit Care Med 15:760-770.
28. Xu L, Earl J, Pichichero ME. 2021. Nasopharyngeal microbiome composition associated with *Streptococcus pneumoniae* colonization suggests a protective role of *Corynebacterium* in young children. PLoS One 16:e0257207.
29. Coleman A, Bialasiewicz S, Marsh RL, Grahn Hakansson E, Cottrell K, Wood A, Jayasundara N, Ware RS, Zaugg J, Sidjabat HE, Adams J, Ferguson J, Brown M, Roos K, Cervin A. 2021. Upper Respiratory Microbiota in Relation to Ear and Nose Health Among Australian Aboriginal and Torres Strait Islander Children. J Pediatric Infect Dis Soc 10:468-476.
30. Coleman A, Zaugg J, Wood A, Cottrell K, Hakansson EG, Adams J, Brown M, Cervin A, Bialasiewicz S. 2021. Upper Respiratory Tract Microbiome of Australian Aboriginal and Torres Strait Islander Children in Ear and Nose Health and Disease. Microbiol Spectr 9:e0036721.
31. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. 2018. New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. mSystems 3:e00187-18.
32. Liu CM, Price LB, Hungate BA, Abraham AG, Larsen LA, Christensen K, Stegger M, Skov R, Andersen PS. 2015. *Staphylococcus aureus* and the ecology of the nasal microbiome. Sci Adv 1:e1400216.
33. Teutsch B, Berger A, Marosevic D, Schonberger K, Lam TT, Hubert K, Beer S, Wienert P, Ackermann N, Claus H, Drayss M, Thiel K, van der Linden M, Vogel U, Sing A. 2017. *Corynebacterium* species nasopharyngeal carriage in asymptomatic individuals aged >/= 65 years in Germany. Infection 45:607-611.
34. Salamzade R, Swaney MH, Kalan LR. 2022. Comparative Genomic and Metagenomic Investigations of the *Corynebacterium tuberculostearicum* Species Complex Reveals Potential Mechanisms Underlying Associations To Skin Health and Disease. Microbiol Spectr doi:10.1128/spectrum.03578-22:e0357822.
35. Oh J, Byrd AL, Park M, Program NCS, Kong HH, Segre JA. 2016. Temporal Stability of the Human Skin Microbiome. Cell 165:854-66.
36. Salamzade R, Cheong JZA, Sandstrom S, Swaney MH, Stubbendieck RM, Starr NL, Currie CR, Singh AM, Kalan LR. 2023. Evolutionary investigations of the biosynthetic diversity in the skin microbiome using lsaBGC. Microb Genom 9.
37. Kluytmans J, van Belkum A, Verbrugh H. 1997. Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. Clin Microbiol Rev 10:505-20.

38. von Eiff C, Becker K, Machka K, Stammer H, Peters G, Group FtS. 2001. Nasal carriage as a source of *Staphylococcus aureus* bacteremia. N Engl J Med 344:11-6.
39. Wertheim HF, Vos MC, Ott A, van Belkum A, Voss A, Kluytmans JA, van Keulen PH, Vandenbroucke-Grauls CM, Meester MH, Verbrugh HA. 2004. Risk and outcome of nosocomial *Staphylococcus aureus* bacteraemia in nasal carriers versus non-carriers. Lancet 364:703-5.
40. Young BC, Wu CH, Gordon NC, Cole K, Price JR, Liu E, Sheppard AE, Perera S, Charlesworth J, Golubchik T, Iqbal Z, Bowden R, Massey RC, Paul J, Crook DW, Peto TE, Walker AS, Llewelyn MJ, Wyllie DH, Wilson DJ. 2017. Severe infections emerge from commensal bacteria by adaptive evolution. Elife 6.
41. Uehara Y, Nakama H, Agematsu K, Uchida M, Kawakami Y, Abdul Fattah AS, Maruchi N. 2000. Bacterial interference among nasal inhabitants: eradication of *Staphylococcus aureus* from nasal cavities by artificial implantation of *Corynebacterium* sp. J Hosp Infect 44:127-33.
42. Lina G, Boutite F, Tristan A, Bes M, Etienne J, Vandenesch F. 2003. Bacterial competition for human nasal cavity colonization: role of Staphylococcal agr alleles. Appl Environ Microbiol 69:18-23.
43. Wos-Oxley ML, Plumeier I, von Eiff C, Taudien S, Platzer M, Vilchez-Vargas R, Becker K, Pieper DH. 2010. A poke into the diversity and associations within human anterior nare microbial communities. ISME J 4:839-51.
44. Johnson RC, Ellis MW, Lanier JB, Schlett CD, Cui T, Merrell DS. 2015. Correlation between nasal microbiome composition and remote purulent skin and soft tissue infections. Infection and immunity 83:802-11.
45. Blum FC, Whitmire JM, Bennett JW, Carey PM, Ellis MW, English CE, Law NN, Tribble DR, Millar EV, Merrell DS. 2022. Nasal microbiota evolution within the congregate setting imposed by military training. Sci Rep 12:11492.
46. Psaltis AJ, Mackenzie BW, Cope EK, Ramakrishnan VR. 2022. Unraveling the role of the microbiome in chronic rhinosinusitis. J Allergy Clin Immunol 149:1513-1521.
47. Kiryukhina NV, Melnikov VG, Suvorov AV, Morozova YA, Ilyin VK. 2013. Use of *Corynebacterium pseudodiphtheriticum* for elimination of *Staphylococcus aureus* from the nasal cavity in volunteers exposed to abnormal microclimate and altered gaseous environment. Probiotics and antimicrobial proteins 5:233-8.
48. Kanmani P, Clua P, Vizoso-Pinto MG, Rodriguez C, Alvarez S, Melnikov V, Takahashi H, Kitazawa H, Villena J. 2017. Respiratory Commensal Bacteria *Corynebacterium pseudodiphtheriticum* Improves Resistance of Infant Mice to Respiratory Syncytial Virus and *Streptococcus pneumoniae* Superinfection. Front Microbiol 8:1613.
49. Horn KJ, Jaberi Vivar AC, Arenas V, Andani S, Janoff EN, Clark SE. 2021. *Corynebacterium* Species Inhibit *Streptococcus pneumoniae* Colonization and Infection of the Mouse Airway. Front Microbiol 12:804935.
50. Menberu MA, Liu S, Cooksley C, Hayes AJ, Psaltis AJ, Wormald PJ, Vreugde S. 2021. *Corynebacterium accolens* Has Antimicrobial Activity against *Staphylococcus aureus* and Methicillin-Resistant *S. aureus* Pathogens Isolated from the Sinonasal Niche of Chronic Rhinosinusitis Patients. Pathogens 10.

51. Brugger SD, Eslami SM, Pettigrew MM, Escapa IF, Henke MT, Kong Y, Lemon KP. 2020. *Dolosigranulum pigrum* Cooperation and Competition in Human Nasal Microbiota. mSphere 5:e00852-20.
52. Hardy BL, Dickey SW, Plaut RD, Riggins DP, Stibitz S, Otto M, Merrell DS. 2019. *Corynebacterium pseudodiphtheriticum* Exploits *Staphylococcus aureus* Virulence Components in a Novel Polymicrobial Defense Strategy. mBio 10.
53. Ramsey MM, Freire MO, Gabrilska RA, Rumbaugh KP, Lemon KP. 2016. *Staphylococcus aureus* Shifts toward Commensalism in Response to *Corynebacterium* Species. Frontiers in Microbiology 7:1230.
54. Stubbendieck RM, May DS, Chevrette MG, Temkin MI, Wendt-Pienkowski E, Cagnazzo J, Carlson CM, Gern JE, Currie CR. 2019. Competition among Nasal Bacteria Suggests a Role for Siderophore-Mediated Interactions in Shaping the Human Nasal Microbiota. Appl Environ Microbiol 85.
55. Neubauer M, Šourek J, Rýc M, Boháček J, Mára M, Mňuková J. 1991. *Corynebacterium accolens* sp. nov., a Gram-Positive Rod Exhibiting Satellitism, from Clinical Material. Systematic and Applied Microbiology 14:46-51.
56. Riegel P, de Briel D, Prevost G, Jehl F, Monteil H. 1993. Proposal of *Corynebacterium propinquum* sp. nov. for *Corynebacterium* group ANF-3 strains. FEMS Microbiol Lett 113:229-234.
57. Karlyshev AV, Melnikov VG. 2013. Draft Genome Sequence of *Corynebacterium pseudodiphtheriticum* Strain 090104 "Sokolov". Genome Announc 1.
58. Bernier AM, Bernard K. 2018. Whole-Genome Sequences of *Corynebacterium macginleyi* CCUG 32361(T) and Clinical Isolates NML 080212 and NML 120205. Microbiol Resour Announc 7.
59. Zhang C, Song W, Ma HR, Peng X, Anderson DJ, Fowler VG, Jr., Thaden JT, Xiao M, You L. 2020. Temporal encoding of bacterial identity and traits in growth dynamics. Proc Natl Acad Sci U S A 117:20202-20210.
60. Kristensen DM, Kannan L, Coleman MK, Wolf YI, Sorokin A, Koonin EV, Mushegian A. 2010. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. Bioinformatics 26:1481-1487.
61. Li L. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Research 13:2178-2189.
62. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.
63. Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol 79:7696-701.
64. Swaney MH, Sandstrom S, Kalan LR. 2022. Cobamide Sharing Is Predicted in the Human Skin Microbiome. mSystems 7:e0067722.
65. Cobian N, Garlet A, Hidalgo-Cantabrana C, Barrangou R. 2021. Comparative Genomic Analyses and CRISPR-Cas Characterization of *Cutibacterium acnes* Provide Insights Into Genetic Diversity and Typing Applications. Front Microbiol 12:758749.
66. Conlan S, Mijares LA, Program NCS, Becker J, Blakesley RW, Bouffard GG, Brooks S, Coleman H, Gupta J, Gurson N, Park M, Schmidt B, Thomas PJ, Otto

M, Kong HH, Murray PR, Segre JA. 2012. *Staphylococcus epidermidis* pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. Genome Biol 13:R64.

67. Pettigrew MM, Ahearn CP, Gent JF, Kong Y, Gallo MC, Munro JB, D'Mello A, Sethi S, Tettelin H, Murphy TF. 2018. *Haemophilus influenzae* genome evolution during persistence in the human airways in chronic obstructive pulmonary disease. Proc Natl Acad Sci U S A 115:E3256-E3265.
68. Flores Ramos S, Brugger SD, Escapa IF, Skeete CA, Cotton SL, Eslami SM, Gao W, Bomar L, Tran TH, Jones DS, Minot S, Roberts RJ, Johnston CD, Lemon KP. 2021. Genomic Stability and Genetic Defense Systems in *Dolosigranulum pigrum*, a Candidate Beneficial Bacterium from the Human Microbiome. mSystems 6:e0042521.
69. McInerney JO, McNally A, O'Connell MJ. 2017. Why prokaryotes have pangenomes. Nat Microbiol 2:17040.
70. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, Trigodet F, Watson AR, Esen OC, Moore RM, Clayssen Q, Lee MD, Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjodin A, Scott JJ, Vazquez-Campos X, McKay LJ, McDaniel EA, Stevens SLR, Anderson RE, Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6:3-6.
71. Oliveira PH, Touchon M, Cury J, Rocha EPC. 2017. The chromosomal organization of horizontal gene transfer in bacteria. Nat Commun 8:841.
72. Iranzo J, Wolf YI, Koonin EV, Sela I. 2019. Gene gain and loss push prokaryotes beyond the homologous recombination barrier and accelerate genome sequence divergence. Nat Commun 10:5376.
73. Gautreau G, Bazin A, Gachet M, Planel R, Burlot L, Dubois M, Perrin A, Medigue C, Calteau A, Cruveiller S, Matias C, Ambroise C, Rocha EPC, Vallenet D. 2020. PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph. PLoS Comput Biol 16:e1007732.
74. Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, Dusch N, Eggeling L, Eikmanns BJ, Gaigalat L, Goesmann A, Hartmann M, Huthmacher K, Kramer R, Linke B, McHardy AC, Meyer F, Mockel B, Pfefferle W, Puhler A, Rey DA, Ruckert C, Rupp O, Sahm H, Wendisch VF, Wiegrabe I, Tauch A. 2003. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. J Biotechnol 104:5-25.
75. Wolf S, Becker J, Tsuge Y, Kawaguchi H, Kondo A, Marienhagen J, Bott M, Wendisch VF, Wittmann C. 2021. Advances in metabolic engineering of Corynebacterium glutamicum to produce high-value active ingredients for food, feed, human health, and well-being. Essays Biochem 65:197-212.
76. Oliveira A, Oliveira LC, Aburjaile F, Benevides L, Tiwari S, Jamal SB, Silva A, Figueiredo HCP, Ghosh P, Portela RW, De Carvalho Azevedo VA, Wattam AR. 2017. Insight of Genus *Corynebacterium*: Ascertaining the Role of Pathogenic and Non-pathogenic Species. Front Microbiol 8:1937.

77. Yang P, Liu W, Chen Y, Gong AD. 2022. Engineering the glyoxylate cycle for chemical bioproduction. Front Bioeng Biotechnol 10:1066651.
78. Saheb Kashaf S, Proctor DM, Deming C, Saary P, Holzer M, Program NCS, Taylor ME, Kong HH, Segre JA, Almeida A, Finn RD. 2022. Integrating cultivation and metagenomics for a multi-kingdom view of skin microbiome diversity and functions. Nat Microbiol 7:169-179.
79. Krismer B, Liebeke M, Janek D, Nega M, Rautenberg M, Hornig G, Unger C, Weidenmaier C, Lalk M, Peschel A. 2014. Nutrient limitation governs *Staphylococcus aureus* metabolism and niche adaptation in the human nose. PLoS Pathog 10:e1003862.
80. Farne H, Groves HT, Gill SK, Stokes I, McCulloch S, Karoly E, Trujillo-Torralbo MB, Johnston SL, Mallia P, Tregoning JS. 2018. Comparative Metabolomic Sampling of Upper and Lower Airways by Four Different Methods to Identify Biochemicals That May Support Bacterial Growth. Front Cell Infect Microbiol 8:432.
81. Burkovski A. 2013. Cell envelope of corynebacteria: structure and influence on pathogenicity. ISRN Microbiol 2013:935736.
82. Adolf LA, Heilbronner S. 2022. Nutritional Interactions between Bacterial Species Colonising the Human Nasal Cavity: Current Knowledge and Future Prospects. Metabolites 12.
83. Cronan JE. 2016. Assembly of Lipoic Acid on Its Cognate Enzymes: an Extraordinary and Essential Biosynthetic Pathway. Microbiol Mol Biol Rev 80:429-50.
84. Shelton AN, Seth EC, Mok KC, Han AW, Jackson SN, Haft DR, Taga ME. 2019. Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics. ISME J 13:789-804.
85. Bruggemann H, Henne A, Hoster F, Liesegang H, Wiezer A, Strittmatter A, Hujer S, Durre P, Gottschalk G. 2004. The complete genome sequence of *Propionibacterium acnes*, a commensal of human skin. Science 305:671-3.
86. Gill SR, Fouts DE, Archer GL, Mongodin EF, Deboy RT, Ravel J, Paulsen IT, Kolonay JF, Brinkac L, Beanan M, Dodson RJ, Daugherty SC, Madupu R, Angiuoli SV, Durkin AS, Haft DH, Vamathevan J, Khouri H, Utterback T, Lee C, Dimitrov G, Jiang L, Qin H, Weidman J, Tran K, Kang K, Hance IR, Nelson KE, Fraser CM. 2005. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. J Bacteriol 187:2426-38.
87. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, Heidelberg J, DeBoy RT, Haft DH, Dodson RJ, Durkin AS, Gwinn M, Kolonay JF, Nelson WC, Peterson JD, Umayam LA, White O, Salzberg SL, Lewis MR, Radune D, Holtzapple E, Khouri H, Wolf AM, Utterback TR, Hansen CL, McDonald LA, Feldblyum TV, Angiuoli S, Dickinson T, Hickey EK, Holt IE, Loftus BJ, Yang F, Smith HO, Venter JC, Dougherty BA, Morrison DA, Hollingshead SK, Fraser CM. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science 293:498-506.

88. Diep BA, Gill SR, Chang RF, Phan TH, Chen JH, Davidson MG, Lin F, Lin J, Carleton HA, Mongodin EF, Sensabaugh GF, Perdreau-Remington F. 2006. Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. Lancet 367:731-9.
89. Copeland E, Leonard K, Carney R, Kong J, Forer M, Naidoo Y, Oliver BGG, Seymour JR, Woodcock S, Burke CM, Stow NW. 2018. Chronic Rhinosinusitis: Potential Role of Microbial Dysbiosis and Recommendations for Sampling Sites. Front Cell Infect Microbiol 8:57.
90. Renz A, Widerspick L, Drager A. 2021. First Genome-Scale Metabolic Model of Dolosigranulum pigrum Confirms Multiple Auxotrophies. Metabolites 11.
91. Tsai YC, Conlan S, Deming C, Program NCS, Segre JA, Kong HH, Korlach J, Oh J. 2016. Resolving the Complexity of Human Skin Metagenomes Using Single-Molecule Sequencing. mBio 7:e01948-15.
92. Tauch A, Schneider J, Szczepanowski R, Tilker A, Viehoever P, Gartemann KH, Arnold W, Blom J, Brinkrolf K, Brune I, Gotker S, Weisshaar B, Goesmann A, Droge M, Puhler A. 2008. Ultrafast pyrosequencing of *Corynebacterium kroppenstedtii* DSM44385 revealed insights into the physiology of a lipophilic corynebacterium that lacks mycolic acids. J Biotechnol 136:22-30.
93. Sichtig H, Minogue T, Yan Y, Stefan C, Hall A, Tallon L, Sadzewicz L, Nadendla S, Klimke W, Hatcher E, Shumway M, Aldea DL, Allen J, Koehler J, Slezak T, Lovell S, Schoepp R, Scherf U. 2019. FDA-ARGOS is a database with public quality-controlled reference genomes for diagnostic use and regulatory science. Nat Commun 10:3313.
94. Seibold GM, Eikmanns BJ. 2007. The *glgX* gene product of *Corynebacterium glutamicum* is required for glycogen degradation and for fast adaptation to hyperosmotic stress. Microbiology (Reading) 153:2212-2220.
95. Seibold G, Dempf S, Schreiner J, Eikmanns BJ. 2007. Glycogen formation in Corynebacterium glutamicum and role of ADP-glucose pyrophosphorylase. Microbiology (Reading) 153:1275-1285.
96. Lensmire JM, Hammer ND. 2019. Nutrient sulfur acquisition strategies employed by bacterial pathogens. Curr Opin Microbiol 47:52-58.
97. Ruckert C, Koch DJ, Rey DA, Albersmeier A, Mormann S, Puhler A, Kalinowski J. 2005. Functional genomics and expression analysis of the *Corynebacterium glutamicum fpr2-cysIXHDNYZ* gene cluster involved in assimilatory sulphate reduction. BMC Genomics 6:121.
98. Sokolovskaya OM, Shelton AN, Taga ME. 2020. Sharing vitamins: Cobamides unveil microbial interactions. Science 369.
99. Jang J, Forbes VE, Sadowsky MJ. 2022. Probable role of *Cutibacterium acnes* in the gut of the polychaete *Capitella teleta*. Sci Total Environ 809:151127.
100. Howden BP, Giulieri SG, Wong Fok Lung T, Baines SL, Sharkey LK, Lee JYH, Hachani A, Monk IR, Stinear TP. 2023. *Staphylococcus aureus* host interactions and adaptation. Nat Rev Microbiol doi:10.1038/s41579-023-00852-y:1-16.
101. Blasche S, Kim Y, Patil KR. 2017. Draft Genome Sequence of *Corynebacterium kefirresidentii* SB, Isolated from Kefir. Genome Announc 5.
102. Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. 2010. The Human Oral Microbiome Database: a web accessible resource for investigating

oral microbe taxonomic and genomic information. Database (Oxford) 2010:baq013.

103. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol 20:714-37.

104. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963.

105. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.

106. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology 19:455-477.

107. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043-55.

108. Human Microbiome Jumpstart Reference Strains C, Nelson KE, Weinstock GM, Highlander SK, Worley KC, Creasy HH, Wortman JR, Rusch DB, Mitreva M, Sodergren E, Chinwalla AT, Feldgarden M, Gevers D, Haas BJ, Madupu R, Ward DV, Birren BW, Gibbs RA, Methe B, Petrosino JF, Strausberg RL, Sutton GG, White OR, Wilson RK, Durkin S, Giglio MG, Gujja S, Howarth C, Kodira CD, Kyrpides N, Mehta T, Muzny DM, Pearson M, Pepin K, Pati A, Qin X, Yandava C, Zeng Q, Zhang L, Berlin AM, Chen L, Hepburn TA, Johnson J, McCorrison J, Miller J, Minx P, Nusbaum C, Russ C, Sykes SM, Tomlinson CM, et al. 2010. A catalog of reference genomes from the human microbiome. Science 328:994-9.

109. Perrin A, Rocha EPC. 2021. PanACoTA: a modular tool for massive microbial comparative genomics. NAR Genom Bioinform 3:lqaa106.

110. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

111. Vinuesa P, Contreras-Moreira B. 2015. Robust identification of orthologues and paralogues for microbial pan-genomics using GET_HOMOLOGUES: a case study of pIncA/C plasmids. Methods Mol Biol 1231:203-32.

112. Altschul S. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389-3402.

113. Vinuesa P, Ochoa-Sanchez LE, Contreras-Moreira B. 2018. GET_PHYLOMARKERS, a Software Package to Select Optimal Orthologous Clusters for Phylogenomics and Inferring Pan-Genome Phylogenies, Used for a Critical Geno-Taxonomic Revision of the Genus Stenotrophomonas. Front Microbiol 9:771.

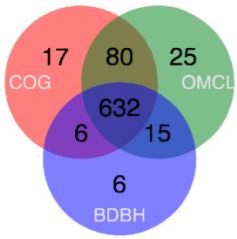
114. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37:1530-1534.
115. Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. Syst Biol 65:997-1008.
116. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35:518-522.
117. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14:587-589.
118. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293-W296.
119. Shaiber A, Willis AD, Delmont TO, Roux S, Chen LX, Schmid AC, Yousef M, Watson AR, Lolans K, Esen OC, Lee STM, Downey N, Morrison HG, Dewhirst FE, Mark Welch JL, Eren AM. 2020. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. Genome Biol 21:292.
120. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59-60.
121. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29:22-8.
122. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. 2021. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res 49:D274-D281.
123. Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27-30.
124. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457-62.
125. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: The protein families database in 2021. Nucleic Acids Res 49:D412-D419.
126. Eddy SR. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol 7:e1002195.
127. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-10.
128. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-7.
129. Benedict MN, Henriksen JR, Metcalf WW, Whitaker RJ, Price ND. 2014. ITEP: an integrated toolkit for exploration of microbial pan-genomes. BMC Genomics 15:8.
130. van Dongen S, Abreu-Goodger C. 2012. Using MCL to Extract Clusters from Networks. *In* van Helden J, Toussaint A, Thieffry D (ed), Bacterial Molecular Networks Methods in Molecular Biology (Methods and Protocols), vol 804. Springer, New York, NY.

131. Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. Science 278:631-7.
132. Ikeda M, Nakagawa S. 2003. The Corynebacterium glutamicum genome: features and impacts on biotechnological processes. Appl Microbiol Biotechnol 62:99-109.
133. Mendel RR. 2013. The molybdenum cofactor. J Biol Chem 288:13165-72.
134. Settembre E, Begley TP, Ealick SE. 2003. Structural biology of enzymes of the thiamin biosynthesis pathway. Curr Opin Struct Biol 13:739-47.
135. Tramonti A, Nardella C, di Salvo ML, Barile A, D'Alessio F, de Crecy-Lagard V, Contestabile R. 2021. Knowns and Unknowns of Vitamin B(6) Metabolism in Escherichia coli. EcoSal Plus 9.
136. Leonardi R, Jackowski S. 2007. Biosynthesis of Pantothenic Acid and Coenzyme A. EcoSal Plus 2.

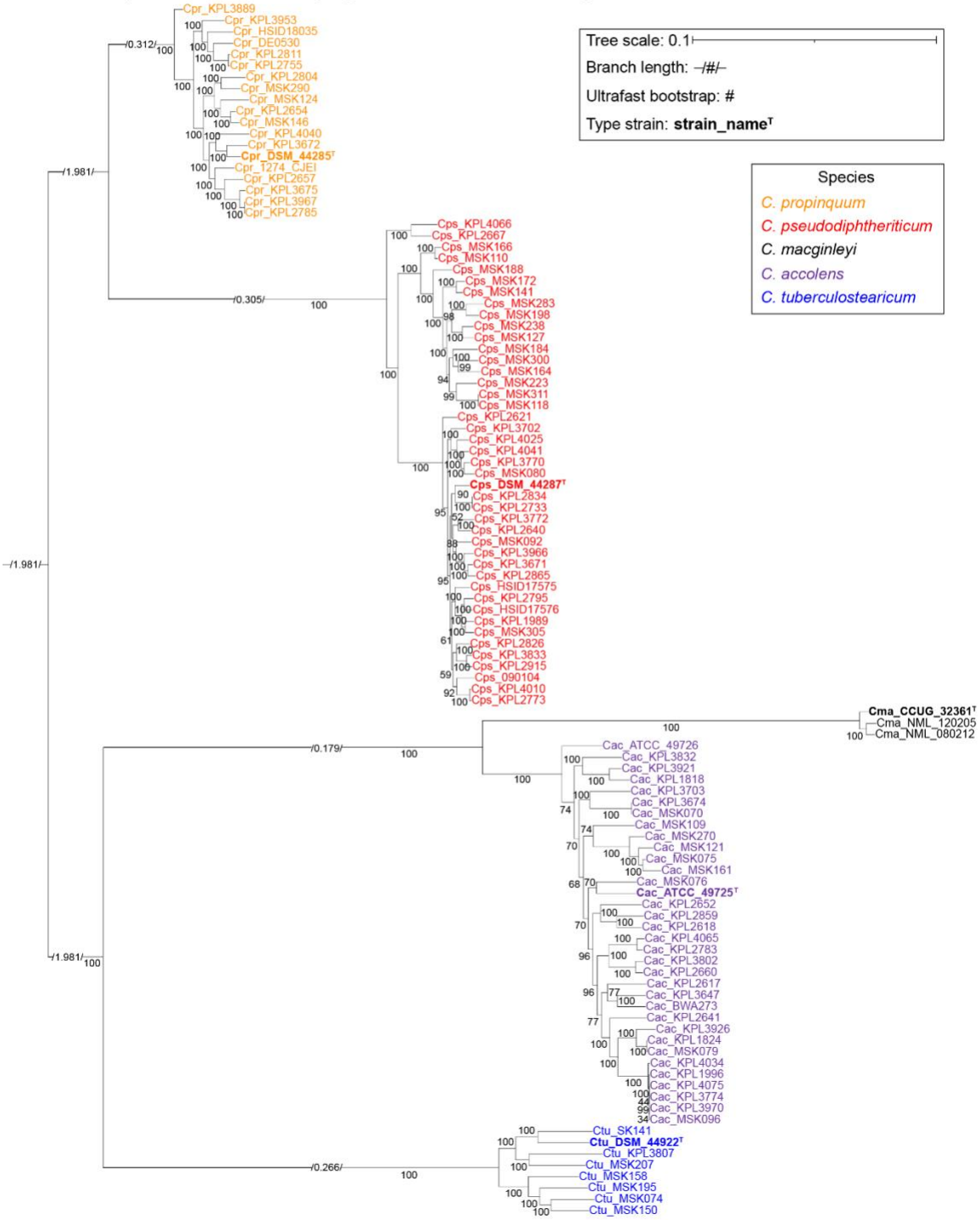
Main and Supplemental Figures (in order of appearance in text)

Figure S1. Phylogenomic tree based on the shared conservative core gene clusters from four human-nasal-associated *Corynebacterium* species. (**A**) Venn diagram showing 632 single-copy GCs are shared between 107 *Corynebacterium* strain genomes based on the consensus of BDBH, COG triangle, and OMCL algorithms using GET_HOMOLOGUES v24082022. (**B**) A maximum-likelihood phylogenomic tree constructed from 632 concatenated shared single-copy GCs from 107 *Corynebacterium* strain genomes. Species are indicated by name and color-code (see key) with type strains highlighted in bold. We used IQ-Tree v2.1.3 with model finder (BIC value 11679476.6345), edge-linked-proportional partition model, and 1,000 ultrafast rapid bootstraps (values on tree). Branch lengths are indicated (see key). The phylogeny contains 31 and 56 nasal-isolated strain genomes from Botswana (86% of the MSK-named strains) and the U.S. (14% of the MSK strains and all KPL-named strains). Many branches are highly supported with ultrafast bootstrap values ≥ 95.

A



B Midpoint Rooted ML Phylogenomic Tree of 107 *Corynebacterium* Strain Genomes



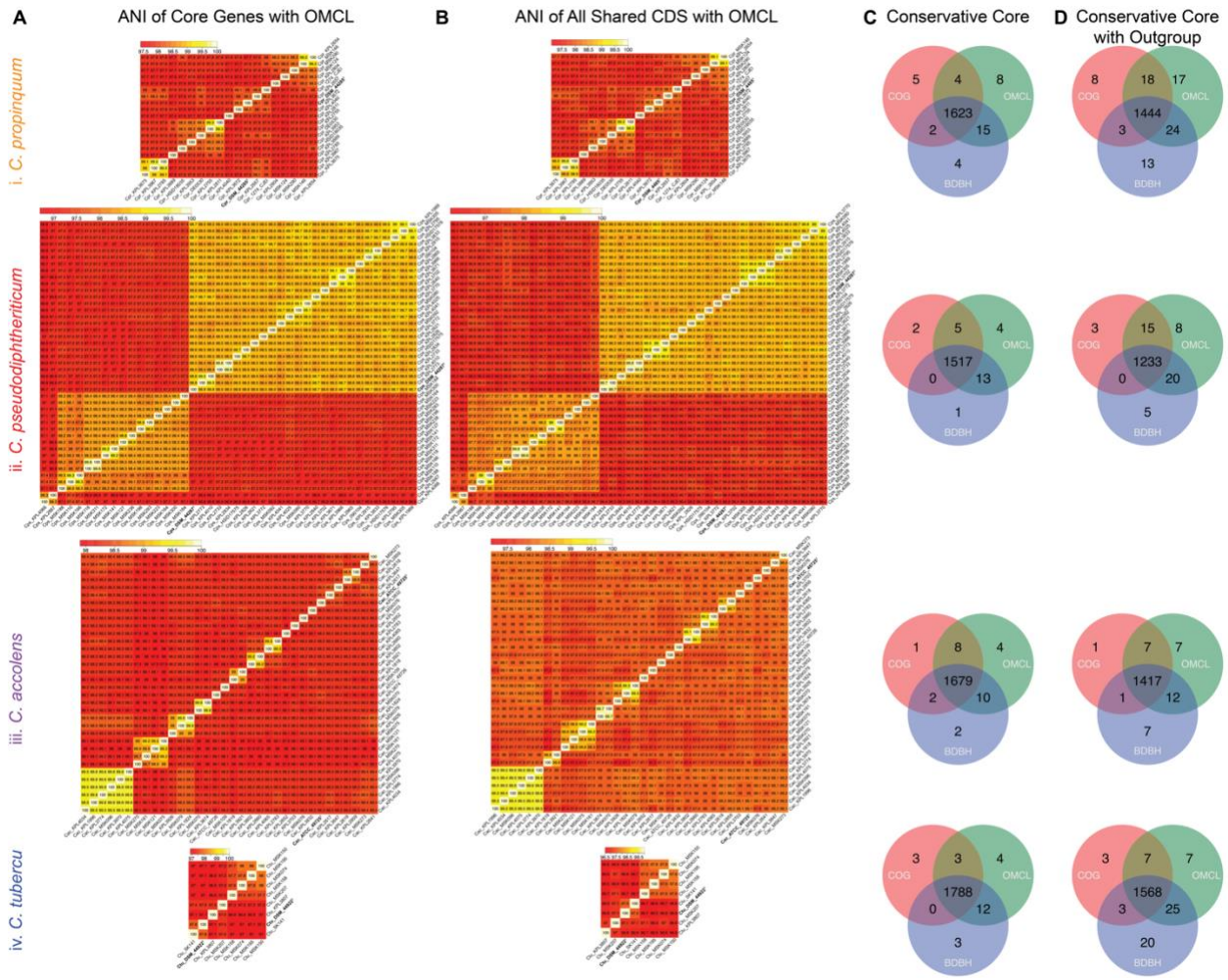


Figure S2. Species-level pairwise average nucleotide identities and conservative core GCs. (A) The shared core genes among all paired strain genomes within each species (**i-iv**) have ≥ 95% ANI (average nucleotide identity) to the type strain using the OMCL algorithm. (**B**) The ANI pairwise comparisons for all shared CDS regions are slightly less than core ANI values but still ≥ 95% among strain genomes within each species. The order of strains in A and B differ due to changes in pairwise identities because more SNPs were included in B. (**C**) The shared conservative core single-copy GCs determined by the consensus of BDBH, COG triangle, and OMCL for each of the *Corynebacterium* species. (**D**) The conservative core CGs for each species plus the

type strain of its closest relative in the overall phylogenomic tree (Fig. S1B). From 112-240 conservative core GCs were absent in the outgroup of each species, with the fewest absent in the outgroup for *C. pseudodiphtheriticum* and the most absent in the outgroup for *C. tuberculostearicum*.

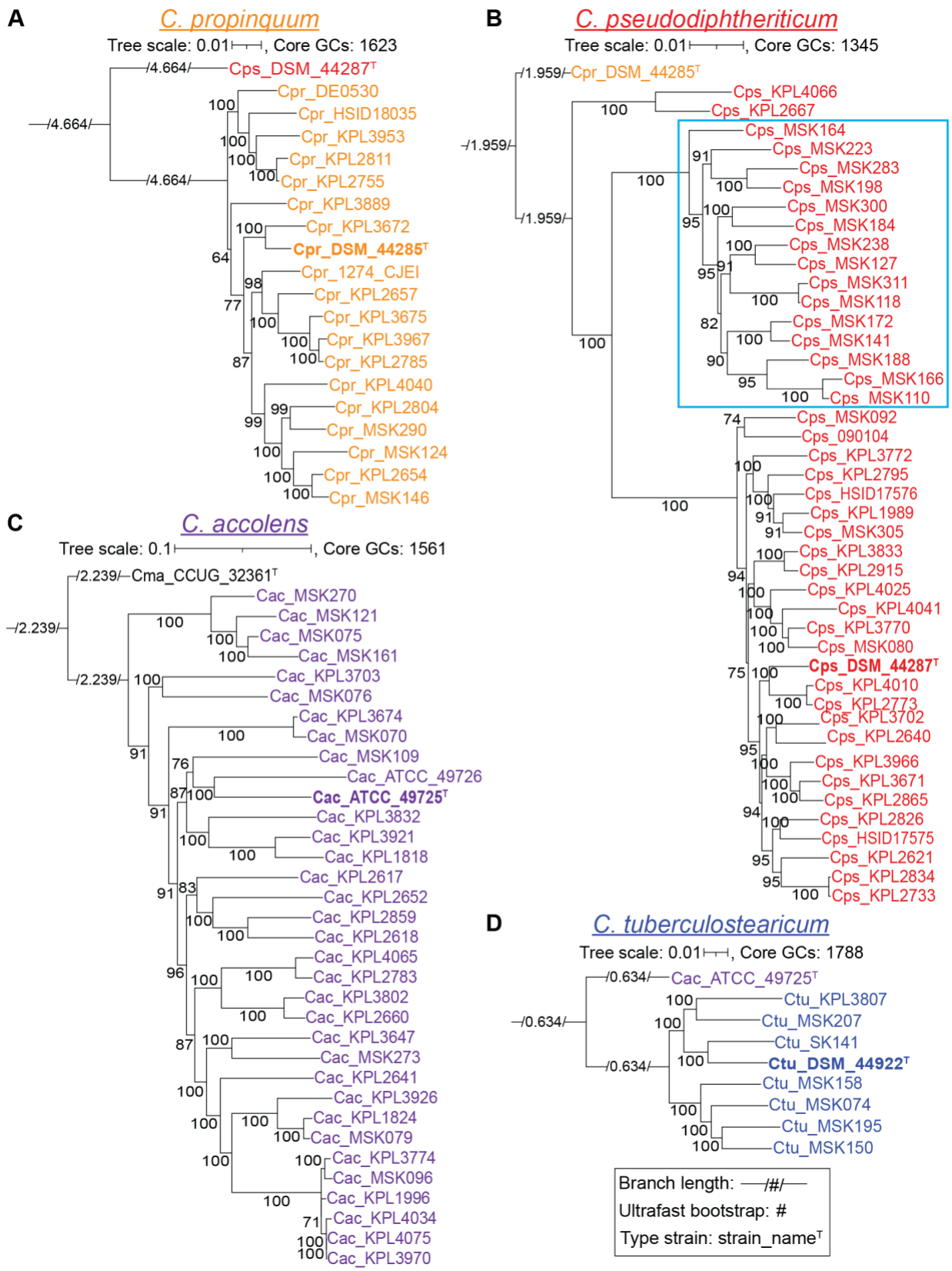

A
C. propinquum
Tree scale: 0.01, Core GCs: 1623
Cps_DSM_44287T
Cpr_DE0530
Cpr_HSID18035
Cpr_KPL3953
Cpr_KPL2811
Cpr_KPL2755
Cpr_KPL3889
Cpr_KPL3672
Cpr_DSM_44285T
Cpr_1274_CJEI
Cpr_KPL2657
Cpr_KPL3675
Cpr_KPL3967
Cpr_KPL2785
Cpr_KPL4040
Cpr_KPL2804
Cpr_MSK290
Cpr_MSK124
Cpr_KPL2654
Cpr_MSK146
B
C. pseudodiphtheriticum
Tree scale: 0.01, Core GCs: 1345
Cpr_DSM_44285T
Cps_KPL4066
Cps_KPL2667
Cps_MSK164
Cps_MSK223
Cps_MSK283
Cps_MSK198
Cps_MSK300
Cps_MSK184
Cps_MSK238
Cps_MSK127
Cps_MSK311
Cps_MSK118
Cps_MSK172
Cps_MSK141
Cps_MSK188
Cps_MSK166
Cps_MSK110
Cps_MSK092
Cps_090104
Cps_KPL3772
Cps_KPL2795
Cps_HSID17576
Cps_KPL1989
Cps_MSK305
Cps_KPL3833
Cps_KPL2915
Cps_KPL4025
Cps_KPL4041
Cps_KPL3770
Cps_MSK080
Cps_DSM_44287T
Cps_KPL4010
Cps_KPL2773
Cps_KPL3702
Cps_KPL2640
Cps_KPL3966
Cps_KPL3671
Cps_KPL2865
Cps_KPL2826
Cps_HSID17575
Cps_KPL2621
Cps_KPL2834
Cps_KPL2733
C
C. accolens
Tree scale: 0.1, Core GCs: 1561
Cma_CCUG_32361T
Cac_MSK270
Cac_MSK121
Cac_MSK075
Cac_MSK161
Cac_KPL3703
Cac_MSK076
Cac_KPL3674
Cac_MSK070
Cac_MSK109
Cac_ATCC_49726
Cac_ATCC_49725T
Cac_KPL3832
Cac_KPL3921
Cac_KPL1818
Cac_KPL2617
Cac_KPL2652
Cac_KPL2859
Cac_KPL2618
Cac_KPL4065
Cac_KPL2783
Cac_KPL3802
Cac_KPL2660
Cac_KPL3647
Cac_MSK273
Cac_KPL2641
Cac_KPL3926
Cac_KPL1824
Cac_MSK079
Cac_KPL3774
Cac_MSK096
Cac_KPL1996
Cac_KPL4034
Cac_KPL4075
Cac_KPL3970
D
C. tuberculostearicum
Tree scale: 0.01, Core GCs: 1788
Cac_ATCC_49725T
Ctu_KPL3807
Ctu_MSK207
Ctu_SK141
Ctu_DSM_44922T
Ctu_MSK158
Ctu_MSK074
Ctu_MSK195
Ctu_MSK150
Branch length: /#/
Ultrafast bootstrap: #
Type strain: strain_nameT

Figure 1. Species-specific phylogenomic trees show a geographic pattern of clades for *C. pseudodiphtheriticum*. Each panel shows a core-genome-based maximum likelihood species-specific phylogeny. The majority (86%) of the MSK-named strains are from Botswana, whereas all KPL-named strains are from the U.S. (**A**) Phylogeny of 19 *C. propinquum* strains based on 1,623 core GCs shows two major clades (BIC value 9762417.2123). (**B**) Phylogeny of 43 *C. pseudodiphtheriticum* strains based on 1,345 core GCs shows three major clades, one of which is entirely composed of MSK strains from Botswana (15/15, outlined in light blue), whereas the other two have a majority of U.S. nasal strains (KPL), with 2/2 and 20/22, respectively (BIC value 10177769.6675). The branching pattern separating the Botswanan and U.S. clades was well supported with ultrafast bootstrap values ≥ 95% (61). (**C**) Phylogeny of 34 *C. accolens* strains based on 1,561 core GCs with the majority collected in the U.S. shows most MSK strains dispersed throughout (BIC value 10700765.2332). (**D**) Phylogeny of eight *C. tuberculostearicum* strains based on 1,788 core GCs with 6 nasal isolates from Botswana and the U.S. (BIC value 10452720.3067). For each species-specific phylogeny, we chose the type strain from the most closely related species (Fig. S1B) as the outgroup. Each phylogeny was made from all shared conservative core GCs for a given species (**Fig. S2C**), including the subset of GCs that were absent in the corresponding outgroup (**Fig. S2D**) to keep the phylogenies at the highest resolution possible for each species. These species level phylogenies were created to increase the resolution between *Corynebacterium* strain genomes from Figure S1, which is based on only the 632 GCs shared by all strains. The increase in shared GCs for species specific trees ranged from 601-936 GCs. Each phylogeny was generated with

IQ-Tree v2.1.3 using model finder, edge-linked-proportional partition model, and 1,000 ultrafast rapid bootstraps. A large majority of the branches have highly supported ultrafast bootstrap values with the lowest at 64 on an ancestral branch in the *C. propinquum* phylogeny. Type strains are indicated with a superscript T. Ancestral branch lengths are indicated numerically within a visually short branch to fit on the page.

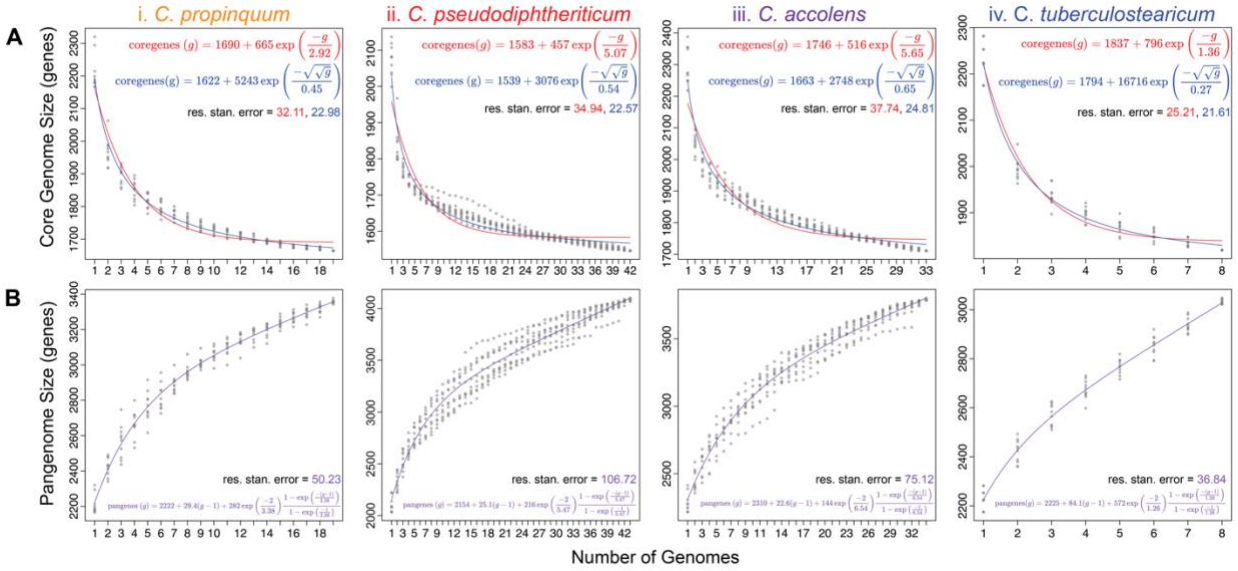


Figure 2. *C. propinquum*, *C. pseudodiphtheriticum*, *C. accolens*, and *C. tuberculostearicum* have core genomes that have leveled off and pangenomes that remain open. (**A**) All four *Corynebacterium* species have a core genome that has leveled off using a Tettelin curve fit model. (**Ai**) The *C. propinquum* core genome (n = 19) leveled off at ~12 genomes. (**Aii**) The *C. pseudodiphtheriticum* core genome (n = 42) leveled off at ~19 genomes. (**Aiii**) The *C. accolens* core genome (n = 33) leveled off at ~ 21 genomes. (**Aiv**) The *C. tuberculostearicum* core genome (n = 8) leveled off at ~7 genomes. Two best fit curve line models are shown for the core genome: Tettelin (red) and Willenbrock (blue). (**B**) The pangenomes for the four *Corynebacterium* species (i-vi) remain open as indicated by the continuous steep slope of the best fit line shown in purple. Core and pangenome size estimations were calculated from 10 random genome samplings (represented by gray dots) using the OMCL algorithm predicted GCs with GET_HOMOLOGUES v24082022.

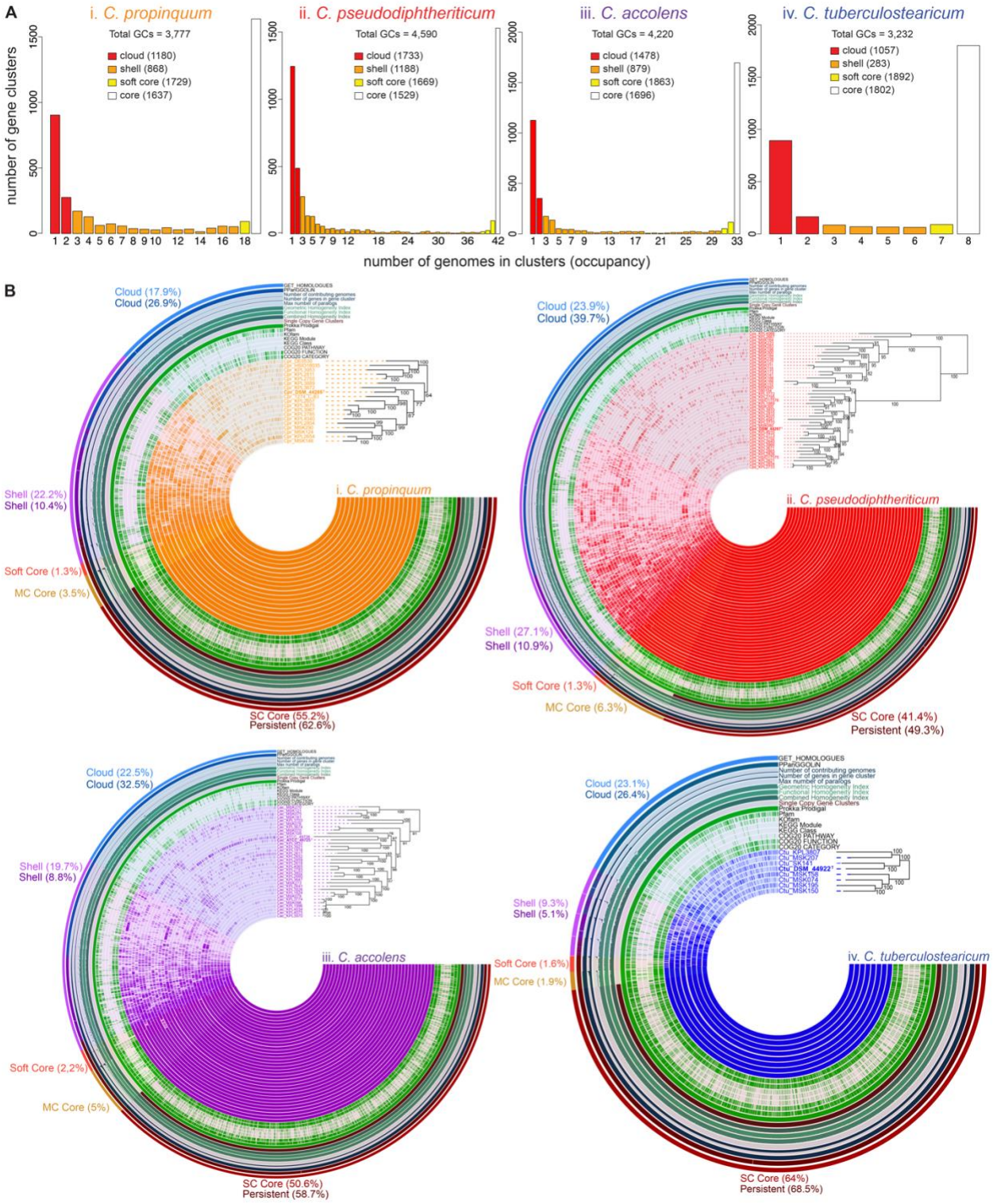


Figure S3. Three different analysis platforms estimated a similar pangenome for each individual *Corynebacterium* species. (A) Representation of the pangenome

estimated by GET_HOMOLOGUES with COG triangle and OMCL of (**i**) *C. propinquum*, 19 genomes; (**ii)** *C. pseudodiphtheriticum*, 42 genomes; (**iii**) *C. accolens*, 33 genomes; and (**iv**) *C. tuberculostearicum,* 8 genomes. (**B**) An anvi'o v7.1.2 concentric pangenome dendrogram for each species (**i-vi**) manually reordered to match the species-specific phylogenies in **Fig. 1**. The GCs present in the core/persistent genome consistently accounted for over half of the pangenome in each of the nasal *Corynebacterium* species. The outermost concentric ring indicates GCs (with percentages) in the single copy core (dark red), multicopy core (beige), soft core (red), shell (light purple), and cloud (light blue) based on GET_HOMOLOGUES with anvi'o further dividing the core into single copy, multicopy, and soft core. The next to outermost ring indicates GCs (with percentages) in the persistent (dark maroon), shell (purple), and cloud (blue) as assigned by PPanGGOLiN v1.2.74. The average core+soft core or persistent, shell, and cloud sizes across the four species from GET_HOMOLOGUES were 58.58%, 19.58%, and 21.85% respectively and from PPanGGOLiN, 56.63%, 10.13%, and 31.38%. The differences between the two algorithms were 1.95% for core, 9.45% for shell, and 9.53% for cloud. GET_HOMOLOGUES defines the pangenome compartments as GCs shared across a certain number of genomes, whereas PPanGGOLiN uses both presence and analysis for genomic neighborhoods to infer GCs in the persistent genome, which likely explains these differences. The anvi'o estimates for single-copy core GCs were (55.2% (1,715/3,108) for *C. propinquum*, 41.4% (1,488/3,590) for *C. pseudodiphtheriticum*, 50.6% (1,734/3,427) for *C. accolens*, and 64% (1,860/2,907) for *C. tuberculostearicum*).

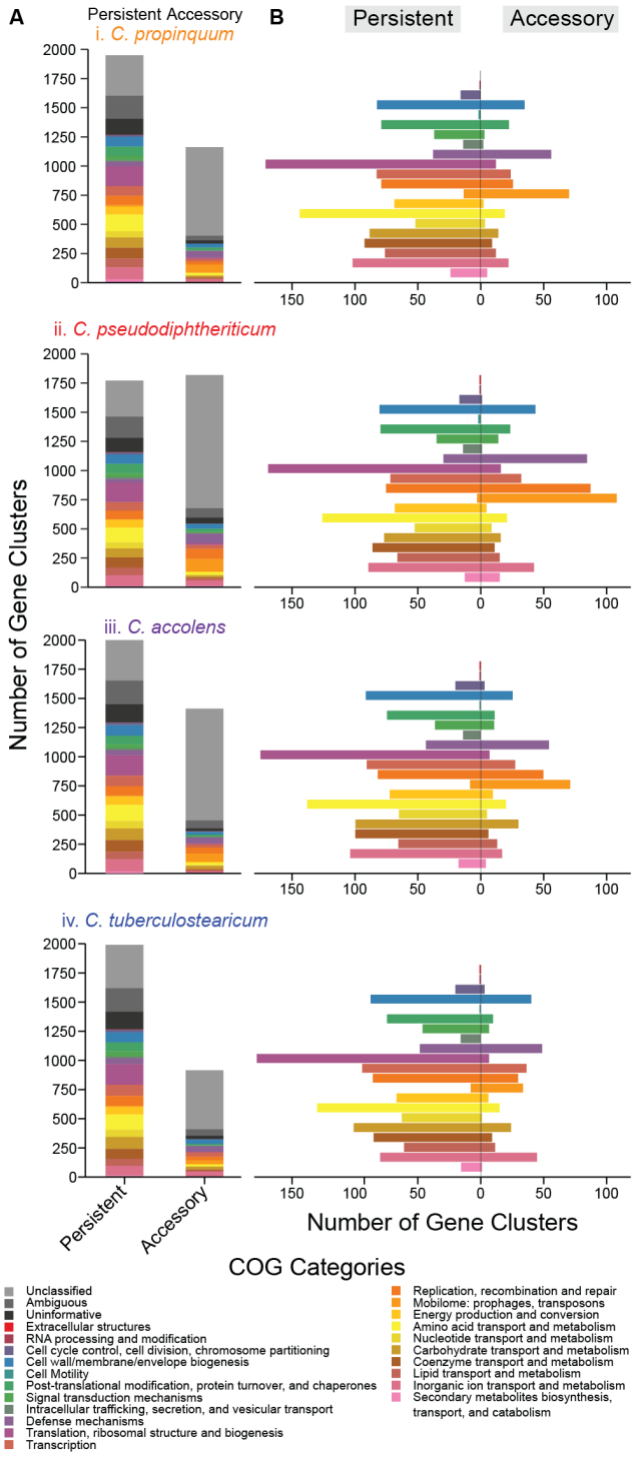


Figure 3. GCs assigned to COG metabolism categories are overrepresented in the persistent compared to the accessory genomes of each species indicating limited strain-specific metabolism. We identified the COG functional annotations for GCs

using anvi'o and then used PPanGGOLiN to assign GCs to the persistent vs. accessory genome. (**A**) Over one-third of the GCs in each species (**i-iv**) were assigned as uninformative (black), ambiguous (dark gray), or unclassified (gray) across both the persistent and accessory genome. The combined percentage of each of these categories out of all the genes per species was 38.1% *Cpr*, 37.9% *Cps*, 37.1% *Cac*, and 38.3% *Ctu*. For each species, the percentage of GCs with an informative COG assignment was higher in the persistent genome, 64.9% *Cpr* (1262), 65.3% *Cps* (1156), 64.7% *Cac* (1300), and 63.5% *Ctu* (1264), than in the accessory genome, with 28.9% *Cpr* (336), 29.9% *Cps* (543), 25.7% *Cac* (363), and 35.6% Ctu (326). (**B**) Functional enrichment of GCs in the persistent vs. the accessory genome for the different COG categories. Metabolic COG categories, e.g., those involved in energy production (pale orange), or in amino acid (yellow), nucleotide (gold), carbohydrate (khaki), and lipid metabolism (dark salmon), were enriched in the persistent genome of each species. In contrast, mobilome (bright orange) and to a lesser extent defense mechanisms were enriched in the accessory genomes. Each *Corynebacterium* species shared similar COG functional enrichment ratios of GCs in its persistent vs. its accessory genome.

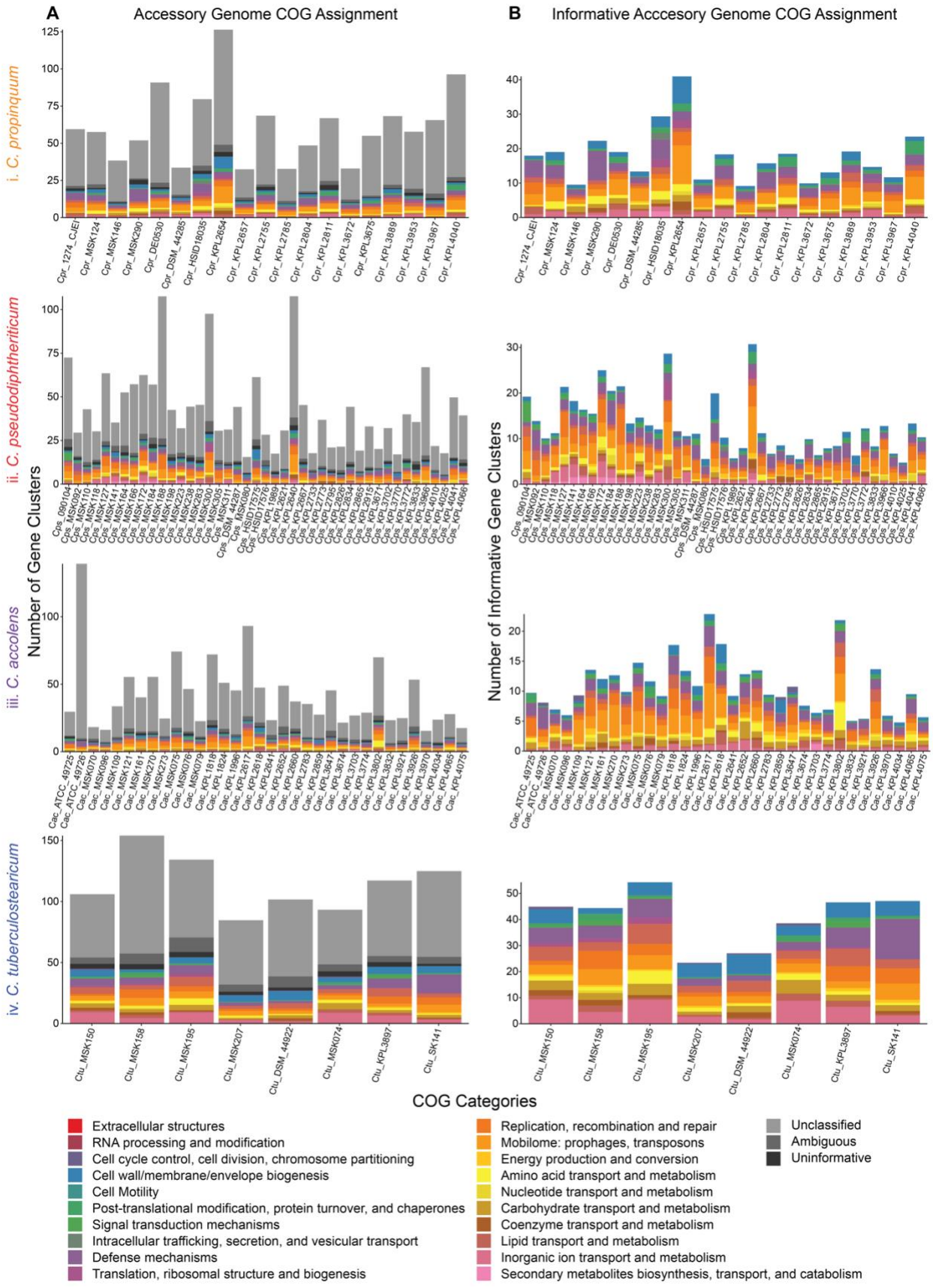

A
Accessory Genome COG Assignment
B
Informative Acccesory Genome COG Assignment
i. C. propinquum
ii. C. pseudodiphtheriticum
iii. C. accolens
iv. C. tuberculostearicum
Number of Gene Clusters
Number of Informative Gene Clusters
COG Categories
Extracellular structures
RNA processing and modification
Cell cycle control, cell division, chromosome partitioning
Cell wall/membrane/envelope biogenesis
Cell Motility
Post-translational modification, protein turnover, and chaperones
Signal transduction mechanisms
Intracellular trafficking, secretion, and vesicular transport
Defense mechanisms
Translation, ribosomal structure and biogenesis
Replication, recombination and repair
Mobilome: prophages, transposons
Energy production and conversion
Amino acid transport and metabolism
Nucleotide transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis, transport, and catabolism
Unclassified
Ambiguous
Uninformative

Figure S4. The size of the accessory genome varies between strains within each nasal *Corynebacterium* species. We used anvi'o to identify COG functional annotations and PPanGGOLiN to assign GCs to the persistent vs. accessory genome. (**A**) For each species (**i-vi**), most of the GCs in each strain's accessory genome had either no annotation ("unclassified"), an ambiguous COG categorization ("ambiguous"), or belonged to an uninformative S or R COG category ("uninformative"). The calculated average percentages for unclassified, ambiguous, and uninformative categories per species were (**i**) 65.2%, 3.5%, and 2.4% for *C. propinquum*; (**ii**) 62.8%, 4.4%, and 2.9% for *C. pseudodiphtheriticum*; (**iii**) 67.6%, 4.7%, and 2.0% for *C. accolens*; and (**iv**) 55.1%, 6.2%, and 3.1% for *C. tuberculostearicum.* The accessory genome of *C. accolens* had the highest percentage of COGS belonging to these three categories and *C. tuberculostearicum* had the lowest. Averages of the informative assigned COG annotations (colored categories) in the accessory genome were calculated for each species as follows: (**i**) 28.9% for *C*. *propinquum*, (**ii**) 29.9% for *C. pseudodiphtheriticum*, (**iii**) 25.7% for *C. accolens*, and (**iv**) 35.6% for *C. tuberculostearicum*. (**B**) There was a relatively proportional distribution of accessory informative functional GCs among each set of genomes. The size of the accessory genome varied between strains from each of the four species. (**Ai**) For *C. propinquum,* KPL2654 and KPL4040 had the largest accessory sizes, whereas KPL2785 and KPL3672 had the smallest. (**Aii**) For *C. pseudodiphtheriticum,* KPL2640 and BWA188 had the largest accessory sizes, whereas DU080 and KPL3070 had the smallest. (**Aiii**) For *C. accolens,* ATCC_49726 and KPL2617 had the largest accessory sizes, whereas BWA096 and KPL3970 had the

smallest. (**Aiv**) For *C. tuberculostearicum* BWA158 and BWA195 had the largest accessory sizes, whereas BWA207 and DU074 had the smallest.

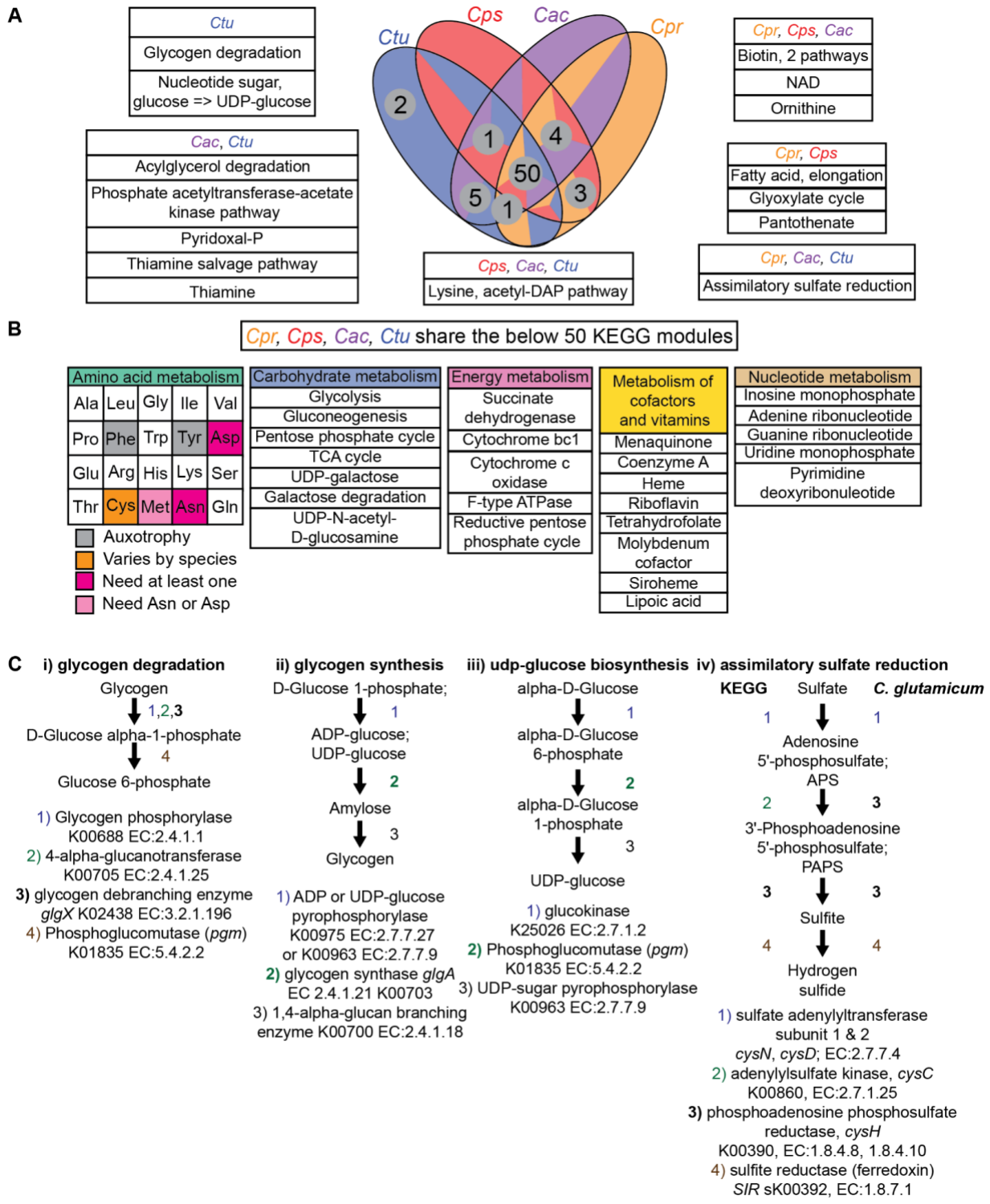


Figure 4. These four common human-nasal-associated *Corynebacterium* species have a largely shared metabolic capacity. All genomes were analyzed using KEGG annotations of genes in each anvi'o contigs database with KEGG Orthology (KO)

numbers from the KEGG KOfam database. Completion scores were calculated for each module in each genome ranging from 0-1 representing the fraction of enzymes present out of total enzymes that make up each module. To identify the complete KEGG modules shared by the four common nasal *Corynebacterium* species, we ran a module enrichment analysis in which we assigned each genome a species group and compared whether that complete module was more likely to be found in one group or multiple groups. Modules with an adjusted p-value <0.05 were considered enriched in their associated species and modules that were not significantly enriched (adjusted p-value >0.05) in comparisons in three out of the four species, were categorized as shared. (**A**) Venn diagram summarizing complete (completion score ≥ 0.75) KEGG modules shared between the four *Corynebacterium* species: *C. propinquum* (*Cpr*)*, C. pseudodiphtheriticum (Cps), C. accolens (Cac),* and *C. tuberculostearicum (Ctu*). Specific modules are grouped by the subgroup of ≤ 3 species encoding these and shown in boxes surrounding the Venn diagram with species labels. (**B**) Detailed list of the functions encompassed by the 50 KEGG modules shared by all four *Corynebacterium* species grouped by module category. In amino acid metabolism, gray indicates amino acids lacking complete biosynthesis modules; orange indicates that biosynthetic capabilities that vary by species (as influenced by predicted assimilatory sulfate reduction capabilities); deep pink indicates a predicted requirement for an exogenous source of either Asp or Asn; and light pink indicates the ability to make Met depends on either exogenous Asp or Asn. (**C**) Manual assessment for *Corynebacterium* orthologs predicted full biosynthetic pathway in several instances where KEGG annotations gave completion scores < 1. We performed blastX with enzymes with

names in bold. (**Ci**) glycogen degradation and synthesis. (**Cii**) synthesis of UDP-glucose. (**Ciii**) Assimilatory sulfate reduction. Noted on the right side is the pathway defined in the KEGG module and on the left is the pathway identified in *C. glutamicum* by Ruckert, Kalinowski, and colleagues (97).

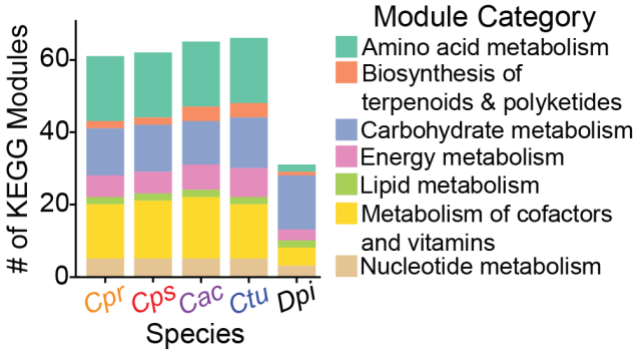
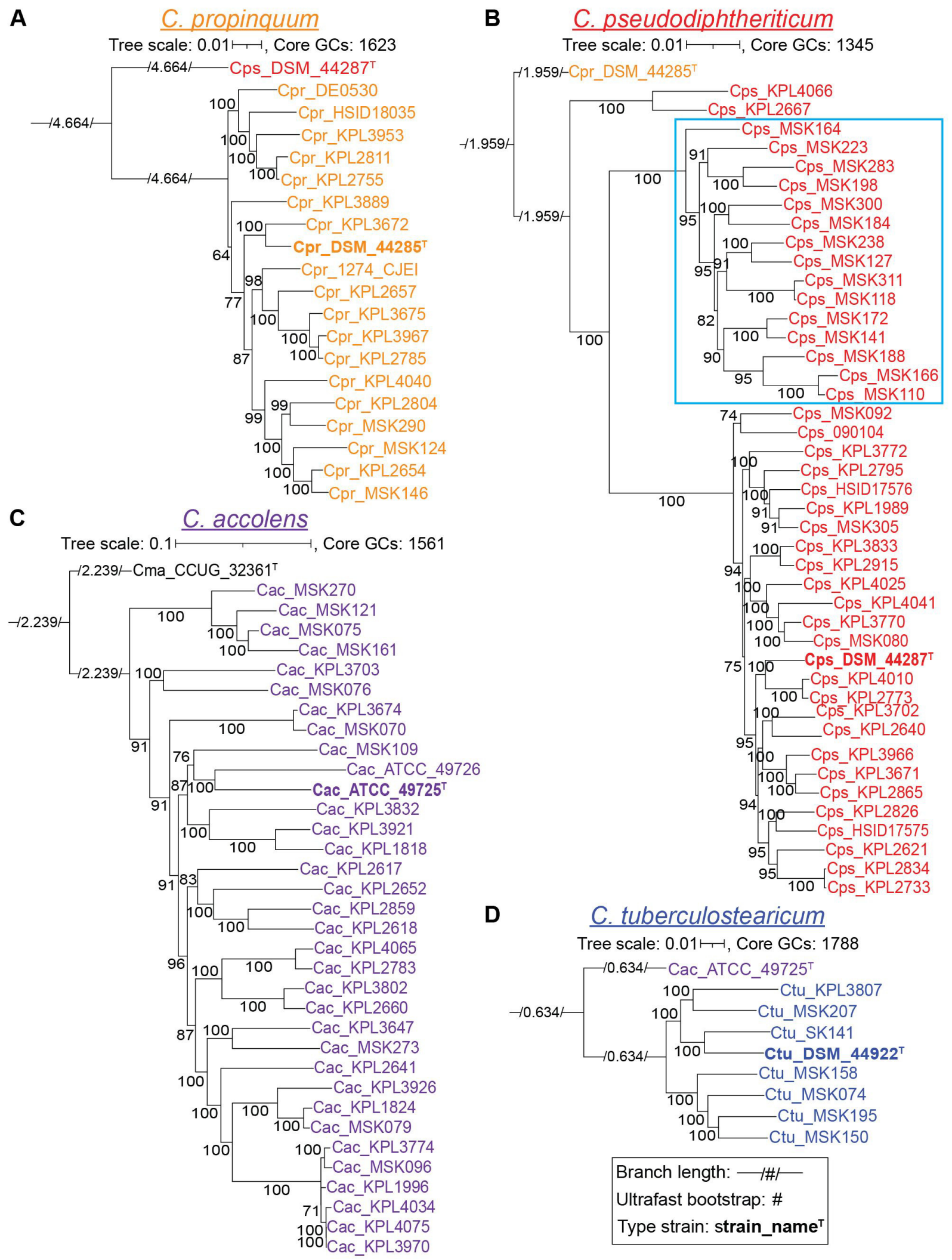
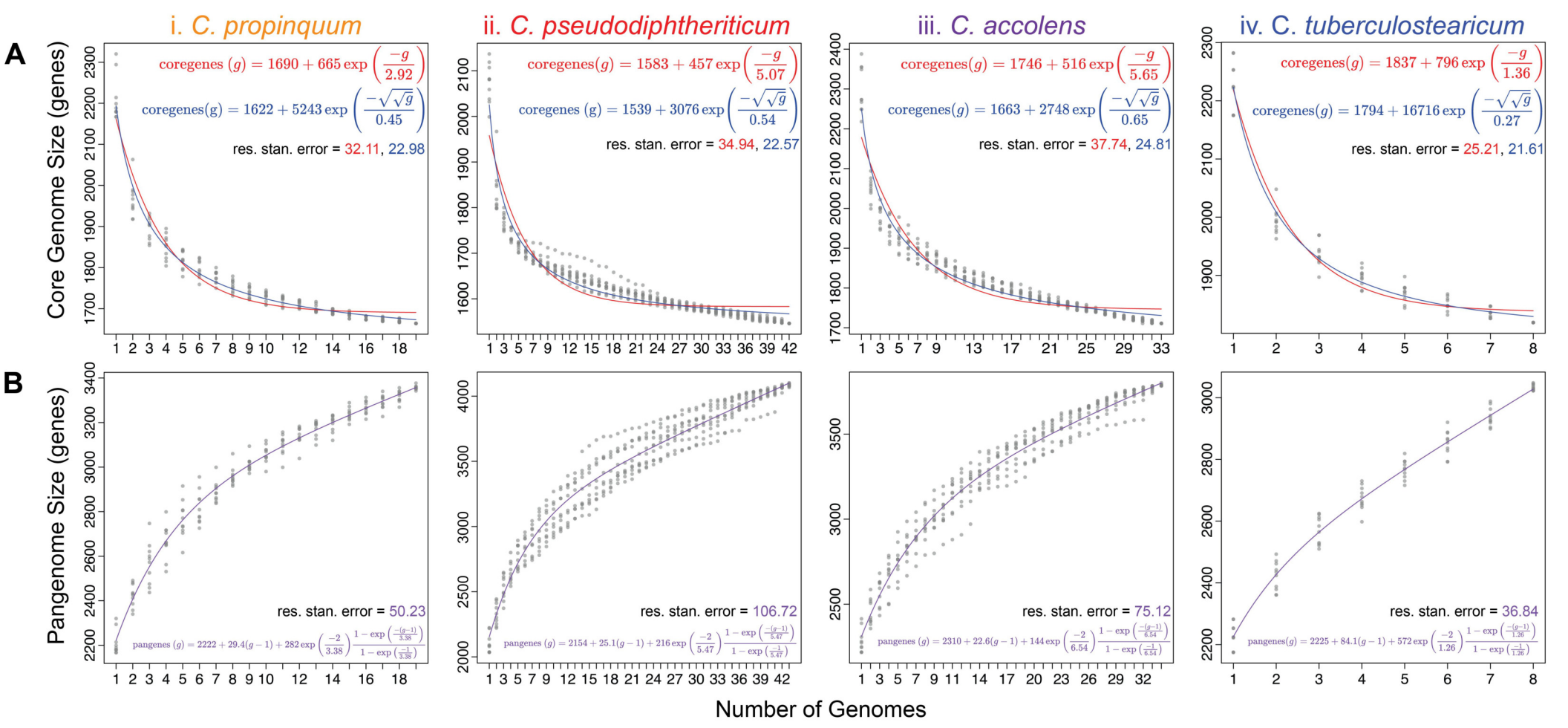
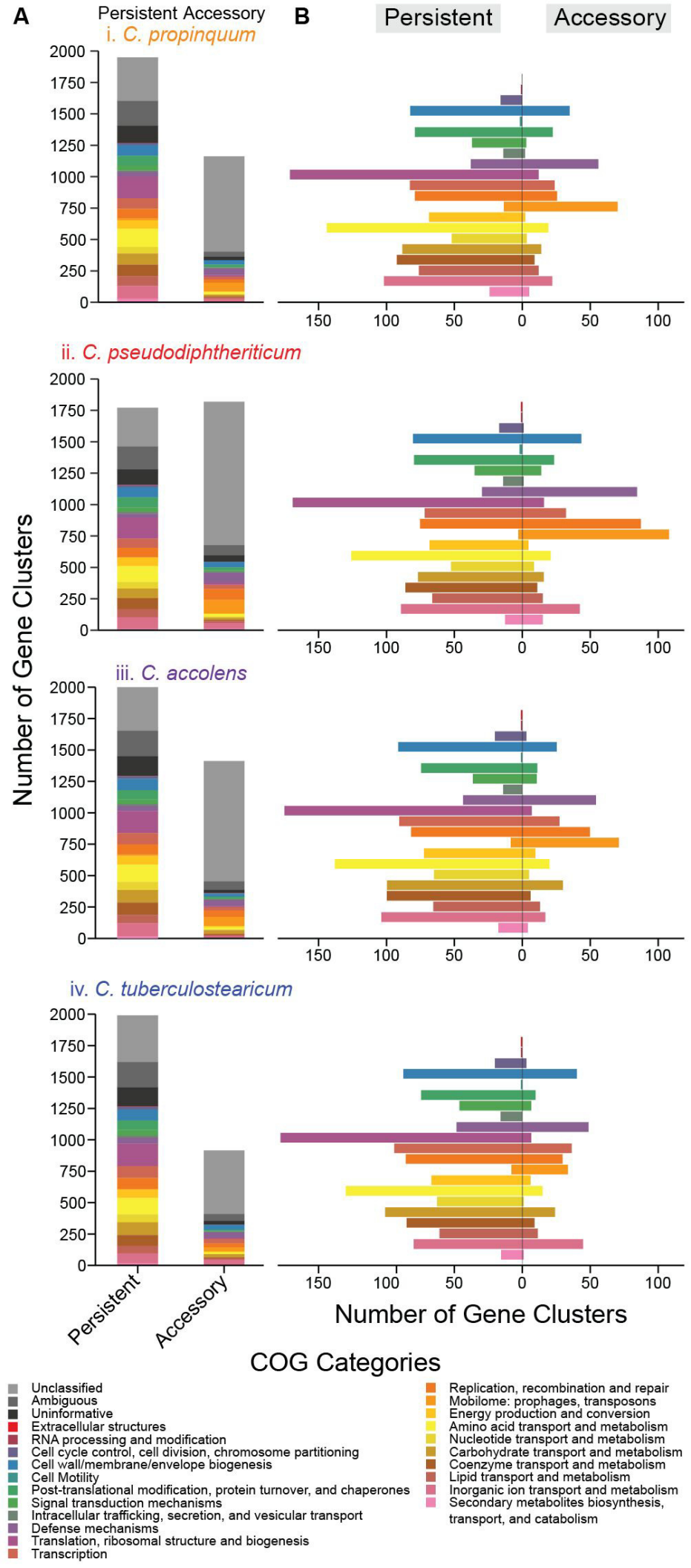


Figure 5. The pangenome of each of the four nasal *Corynebacterium* species have more amino acid and cofactor/vitamin metabolic capacity than does *D. pigrum*. Distribution of KEGG modules with a completion score ≥ 0.75 by module category in the pangenomes of four common human-nasal-associated species of *Corynebacterium* (avg. genome size 2.3 - 2.5 Mb) and in *D. pigrum* (avg. genome size 1.9 Mb), a common human-nasal-associated bacterium that is often positively associated with the genus *Corynebacterium* in compositional studies of human nasal microbiota.


A
C. propinquum
Tree scale: 0.01, Core GCs: 1623
Cps_DSM_44287T
Cpr_DE0530
Cpr_HSID18035
Cpr_KPL3953
Cpr_KPL2811
Cpr_KPL2755
Cpr_KPL3889
Cpr_KPL3672
Cpr_DSM_44285T
Cpr_1274_CJEI
Cpr_KPL2657
Cpr_KPL3675
Cpr_KPL3967
Cpr_KPL2785
Cpr_KPL4040
Cpr_KPL2804
Cpr_MSK290
Cpr_MSK124
Cpr_KPL2654
Cpr_MSK146
B
C. pseudodiphtheriticum
Tree scale: 0.01, Core GCs: 1345
Cpr_DSM_44285T
Cps_KPL4066
Cps_KPL2667
Cps_MSK164
Cps_MSK223
Cps_MSK283
Cps_MSK198
Cps_MSK300
Cps_MSK184
Cps_MSK238
Cps_MSK127
Cps_MSK311
Cps_MSK118
Cps_MSK172
Cps_MSK141
Cps_MSK188
Cps_MSK166
Cps_MSK110
Cps_MSK092
Cps_090104
Cps_KPL3772
Cps_KPL2795
Cps_HSID17576
Cps_KPL1989
Cps_MSK305
Cps_KPL3833
Cps_KPL2915
Cps_KPL4025
Cps_KPL4041
Cps_KPL3770
Cps_MSK080
Cps_DSM_44287T
Cps_KPL4010
Cps_KPL2773
Cps_KPL3702
Cps_KPL2640
Cps_KPL3966
Cps_KPL3671
Cps_KPL2865
Cps_KPL2826
Cps_HSID17575
Cps_KPL2621
Cps_KPL2834
Cps_KPL2733
C
C. accolens
Tree scale: 0.1, Core GCs: 1561
Cma_CCUG_32361T
Cac_MSK270
Cac_MSK121
Cac_MSK075
Cac_MSK161
Cac_KPL3703
Cac_MSK076
Cac_KPL3674
Cac_MSK070
Cac_MSK109
Cac_ATCC_49726
Cac_ATCC_49725T
Cac_KPL3832
Cac_KPL3921
Cac_KPL1818
Cac_KPL2617
Cac_KPL2652
Cac_KPL2859
Cac_KPL2618
Cac_KPL4065
Cac_KPL2783
Cac_KPL3802
Cac_KPL2660
Cac_KPL3647
Cac_MSK273
Cac_KPL2641
Cac_KPL3926
Cac_KPL1824
Cac_MSK079
Cac_KPL3774
Cac_MSK096
Cac_KPL1996
Cac_KPL4034
Cac_KPL4075
Cac_KPL3970
D
C. tuberculostearicum
Tree scale: 0.01, Core GCs: 1788
Cac_ATCC_49725T
Ctu_KPL3807
Ctu_MSK207
Ctu_SK141
Ctu_DSM_44922T
Ctu_MSK158
Ctu_MSK074
Ctu_MSK195
Ctu_MSK150
Branch length: /#/
Ultrafast bootstrap: #
Type strain: strain_nameT

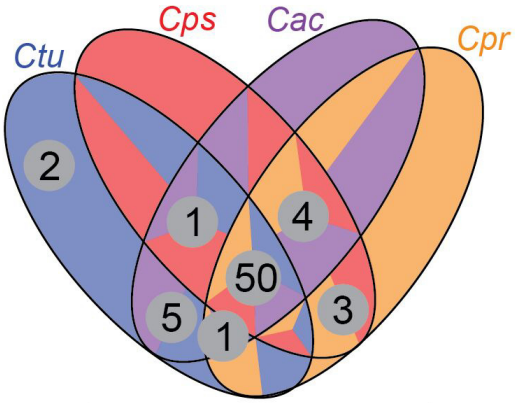

A
B
i. C. propinquum
ii. C. pseudodiphtheriticum
iii. C. accolens
iv. C. tuberculostearicum
Core Genome Size (genes)
Pangenome Size (genes)
Number of Genomes
$\text{coregenes}(g) = 1690 + 665 \exp\left(\frac{-g}{2.92}\right)$
$\text{coregenes}(g) = 1622 + 5243 \exp\left(\frac{-\sqrt{\sqrt{g}}}{0.45}\right)$
res. stan. error = 32.11, 22.98
$\text{coregenes}(g) = 1583 + 457 \exp\left(\frac{-g}{5.07}\right)$
$\text{coregenes}(g) = 1539 + 3076 \exp\left(\frac{-\sqrt{\sqrt{g}}}{0.54}\right)$
res. stan. error = 34.94, 22.57
$\text{coregenes}(g) = 1746 + 516 \exp\left(\frac{-g}{5.65}\right)$
$\text{coregenes}(g) = 1663 + 2748 \exp\left(\frac{-\sqrt{\sqrt{g}}}{0.65}\right)$
res. stan. error = 37.74, 24.81
$\text{coregenes}(g) = 1837 + 796 \exp\left(\frac{-g}{1.36}\right)$
$\text{coregenes}(g) = 1794 + 16716 \exp\left(\frac{-\sqrt{\sqrt{g}}}{0.27}\right)$
res. stan. error = 25.21, 21.61
res. stan. error = 50.23
$\text{pangenes}(g) = 2222 + 29.4(g-1) + 282 \exp\left(\frac{-2}{3.38}\right) \frac{1-\exp\left(\frac{-(g-1)}{3.38}\right)}{1-\exp\left(\frac{-1}{3.38}\right)}$
res. stan. error = 106.72
$\text{pangenes}(g) = 2154 + 25.1(g-1) + 216 \exp\left(\frac{-2}{5.47}\right) \frac{1-\exp\left(\frac{-(g-1)}{5.47}\right)}{1-\exp\left(\frac{-1}{5.47}\right)}$
res. stan. error = 75.12
$\text{pangenes}(g) = 2310 + 22.6(g-1) + 144 \exp\left(\frac{-2}{6.54}\right) \frac{1-\exp\left(\frac{-(g-1)}{6.54}\right)}{1-\exp\left(\frac{-1}{6.54}\right)}$
res. stan. error = 36.84
$\text{pangenes}(g) = 2225 + 84.1(g-1) + 572 \exp\left(\frac{-2}{1.26}\right) \frac{1-\exp\left(\frac{-(g-1)}{1.26}\right)}{1-\exp\left(\frac{-1}{1.26}\right)}$


A
Persistent Accessory
B
Persistent
Accessory
i. C. propinquum
ii. C. pseudodiphtheriticum
iii. C. accolens
iv. C. tuberculostearicum
Number of Gene Clusters
2000
1750
1500
1250
1000
750
500
250
0
150
100
50
0
50
100
Persistent
Accessory
Number of Gene Clusters
COG Categories
Unclassified
Ambiguous
Uninformative
Extracellular structures
RNA processing and modification
Cell cycle control, cell division, chromosome partitioning
Cell wall/membrane/envelope biogenesis
Cell Motility
Post-translational modification, protein turnover, and chaperones
Signal transduction mechanisms
Intracellular trafficking, secretion, and vesicular transport
Defense mechanisms
Translation, ribosomal structure and biogenesis
Transcription
Replication, recombination and repair
Mobilome: prophages, transposons
Energy production and conversion
Amino acid transport and metabolism
Nucleotide transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis, transport, and catabolism

A

Ctu
Glycogen degradation
Nucleotide sugar, glucose => UDP-glucose

Cac, *Ctu*
Acylglycerol degradation
Phosphate acetyltransferase-acetate kinase pathway
Pyridoxal-P
Thiamine salvage pathway
Thiamine



Cps, *Cac*, *Ctu*
Lysine, acetyl-DAP pathway

Cpr, *Cps*, *Cac*
Biotin, 2 pathways
NAD
Ornithine

Cpr, *Cps*
Fatty acid, elongation
Glyoxylate cycle
Pantothenate

Cpr, *Cac*, *Ctu*
Assimilatory sulfate reduction

B

Cpr, *Cps*, *Cac*, *Ctu* share the below 50 KEGG modules

Amino acid metabolism				
Ala	Leu	Gly	Ile	Val
Pro	Phe	Trp	Tyr	Asp
Glu	Arg	His	Lys	Ser
Thr	Cys	Met	Asn	Gln

- Auxotrophy
- Varies by species
- Need at least one
- Need Asn or Asp

Carbohydrate metabolism
Glycolysis
Gluconeogenesis
Pentose phosphate cycle
TCA cycle
UDP-galactose
Galactose degradation
UDP-N-acetyl-D-glucosamine

Energy metabolism
Succinate dehydrogenase
Cytochrome bc1
Cytochrome c oxidase
F-type ATPase
Reductive pentose phosphate cycle

Metabolism of cofactors and vitamins
Menaquinone
Coenzyme A
Heme
Riboflavin
Tetrahydrofolate
Molybdenum cofactor
Siroheme
Lipoic acid

Nucleotide metabolism
Inosine monophosphate
Adenine ribonucleotide
Guanine ribonucleotide
Uridine monophosphate
Pyrimidine deoxyribonuleotide

C

i) glycogen degradation

Glycogen
↓ 1,2,**3**
D-Glucose alpha-1-phosphate
↓ 4
Glucose 6-phosphate

1) Glycogen phosphorylase K00688 EC:2.4.1.1
2) 4-alpha-glucanotransferase K00705 EC:2.4.1.25
3) glycogen debranching enzyme *glgX* K02438 EC:3.2.1.196
4) Phosphoglucomutase (*pgm*) K01835 EC:5.4.2.2

ii) glycogen synthesis

D-Glucose 1-phosphate;
↓ 1
ADP-glucose; UDP-glucose
↓ **2**
Amylose
↓ 3
Glycogen

1) ADP or UDP-glucose pyrophosphorylase K00975 EC:2.7.7.27 or K00963 EC:2.7.7.9
2) glycogen synthase *glgA* EC 2.4.1.21 K00703
3) 1,4-alpha-glucan branching enzyme K00700 EC:2.4.1.18

iii) udp-glucose biosynthesis

alpha-D-Glucose
↓ 1
alpha-D-Glucose 6-phosphate
↓ **2**
alpha-D-Glucose 1-phosphate
↓ 3
UDP-glucose

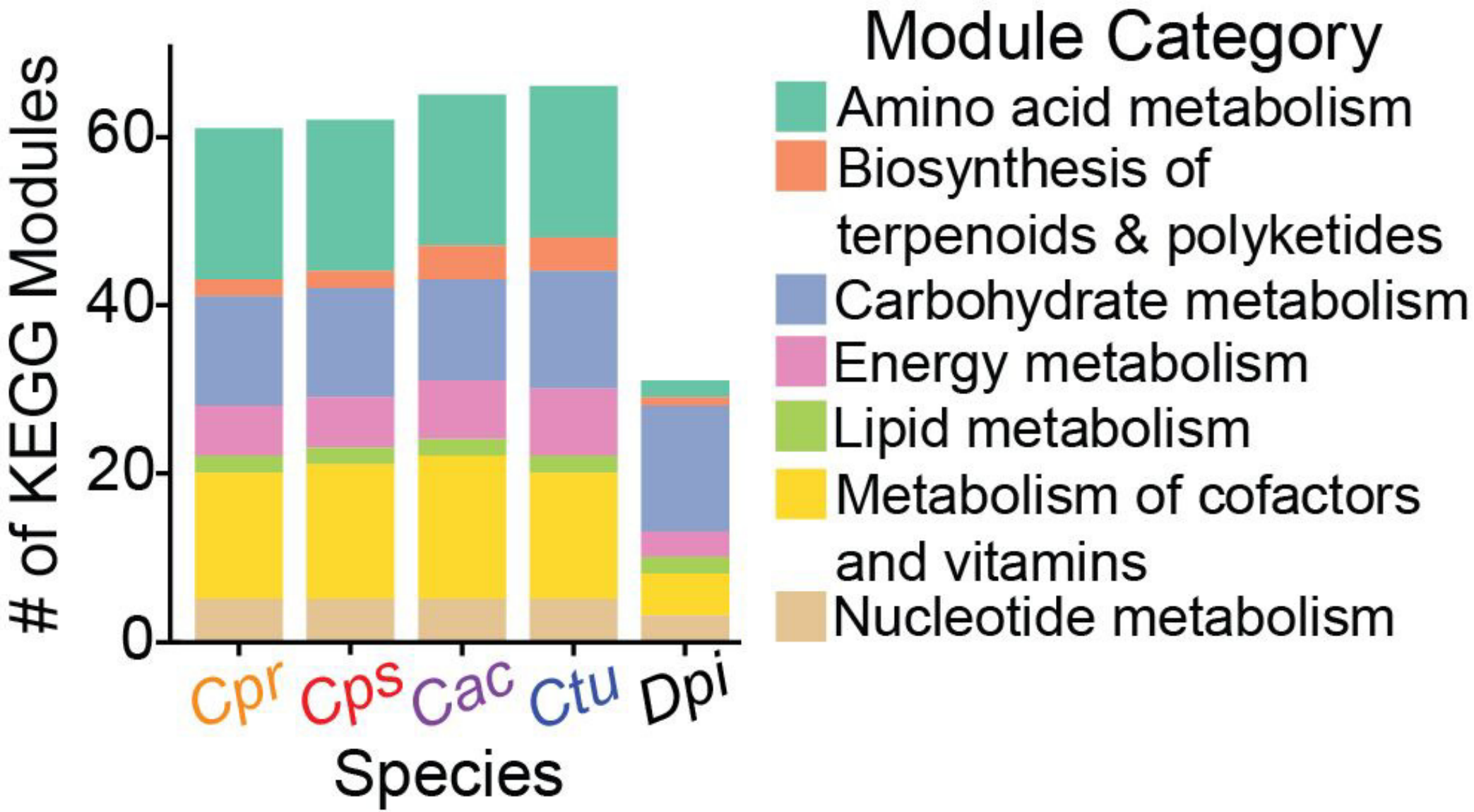
1) glucokinase K25026 EC:2.7.1.2
2) Phosphoglucomutase (*pgm*) K01835 EC:5.4.2.2
3) UDP-sugar pyrophosphorylase K00963 EC:2.7.7.9

iv) assimilatory sulfate reduction

KEGG Sulfate ***C. glutamicum***

1 ↓ 1
Adenosine 5'-phosphosulfate; APS
2 ↓ **3**
3'-Phosphoadenosine 5'-phosphosulfate; PAPS
3 ↓ **3**
Sulfite
4 ↓ 4
Hydrogen sulfide

1) sulfate adenylyltransferase subunit 1 & 2 *cysN*, *cysD*; EC:2.7.7.4
2) adenylylsulfate kinase, *cysC* K00860, EC:2.7.1.25
3) phosphoadenosine phosphosulfate reductase, *cysH* K00390, EC:1.8.4.8, 1.8.4.10
4) sulfite reductase (ferredoxin) *SIR* sK00392, EC:1.8.7.1


Module Category
Amino acid metabolism
Biosynthesis of terpenoids & polyketides
Carbohydrate metabolism
Energy metabolism
Lipid metabolism
Metabolism of cofactors and vitamins
Nucleotide metabolism
# of KEGG Modules
0
20
40
60
Cpr
Cps
Cac
Ctu
Dpi
Species