Corynebacterium Comparative Genomics Reveals a Role for Cobamide Sharing in the Skin Microbiome

Mary Hannah Swaney¹ and Lindsay R Kalan^{1,2}

¹Department of Medical Microbiology and Immunology, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA

²Department of Medicine, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA

Address for Correspondence: Lindsay R. Kalan, 1550 Linden Dr, 6155 Microbial Sciences Building, Madison, WI, USA, 53706. lkalan@wisc.edu

ABSTRACT

The human skin microbiome is a key player in human health, with diverse functions ranging from defense against pathogens to education of the immune system. Recent studies have begun unraveling the complex interactions within skin microbial communities, shedding light on the invaluable role that skin microorganisms have in maintaining a healthy skin barrier. While the *Corynebacterium* genus is a dominant taxon of the skin microbiome, relatively little is known how skin-associated Corynebacteria contribute to microbe-microbe and microbe-host interactions on the skin. Here, we performed a comparative genomics analysis of 71 *Corynebacterium* species from diverse ecosystems, which revealed functional differences between host- and environment-associated species. In particular, host-associated species were enriched for *de novo* biosynthesis of cobamides, which are a class of cofactor essential for metabolism in organisms across the tree of life but are produced by a limited number of prokaryotes. Because cobamides have been hypothesized to mediate community dynamics within microbial communities, we analyzed skin metagenomes for *Corynebacterium* cobamide producers, which revealed a positive correlation between cobamide producer abundance and microbiome diversity, a trait associated with skin health. We also provide the first metagenome-based assessment of cobamide biosynthesis and utilization in the skin microbiome, showing that both dominant and low abundant skin taxa encode for the *de novo* biosynthesis pathway and that cobamide-dependent enzymes are encoded by phylogenetically diverse taxa across the major bacterial phyla on the skin. Taken together, our results support a role for cobamide sharing within skin microbial communities, which we hypothesize mediates community dynamics.

INTRODUCTION

The human skin supports a diverse and complex ecosystem of bacterial, fungal, viral, and microeukaryote species, termed the skin microbiome. Highly adapted to live on the skin, these microorganisms form distinct and specialized communities across the various microenvironments of the skin, which include sebaceous, moist, dry, and foot sites. Rather than existing as innocuous bystanders, the skin microbiome plays a significant role in human health through contributing to immune system education and homeostasis, protecting against pathogen colonization, and promoting barrier maintenance and repair [1–6].

Advances in sequencing technology have allowed us to not only identify the taxa that comprise the microbial communities of the skin, but also to ascertain the functional potential of the skin microbiome and the contribution of particular taxa to microbiome structure, stability, and function. For example, genomics-guided approaches have identified a *Cutibacterium acnes* biosynthetic gene cluster that produces the novel small-molecule cutimycin, contributing to niche competition within the hair follicle microbial community [7]. In addition, pangenomic analyses of dominant skin species have revealed the potential for functional niche saturation in the skin microbiome, where distinct combinations of strains within the same species contribute similar genetic potential [8]. Genomics studies have also elucidated the lineage-specific genetic differences between *C. acnes* strains that may explain the dual commensal and pathogenic phenotypes exhibited by this skin species [9]. These studies exemplify the power of sequence data analysis in generating hypotheses and gathering functional insight to skin microbial communities.

The transition from taxonomic characterization of the skin microbiome towards the elucidation of microbe-microbe and microbe-host interactions has shed light on the truly complex nature of skin microbial communities. Recent work has demonstrated that skin commensals not only take part in synergistic and competitive interactions within skin microbial communities [10–13], but also participate in microbe-host interactions that can dictate skin health and community structure [1,14–16]. While these studies have provided fundamental insight into the roles that certain skin commensals, particularly *Staphylococcus* species and *C. acnes*, play on the skin, little is known regarding the contribution of other major taxa residing within the skin.

Until recently, species of the *Corynebacterium* genus have been underappreciated as significant members of skin microbial communities, predominantly due to the difficulty of growing these species in the lab, which is a result of their nutritionally fastidious and slow-growing nature [17].

However, sequencing efforts have revealed that *Corynebacteria* are a dominant taxon within the microbiome, particularly in moist skin microenvironments [8,18,19]. Further, recent studies have exemplified the importance of certain *Corynebacterium* species in limiting pathogen colonization and virulence [20,21] and modulating the host response [22,23]. Considering the prevalence and diversity of commensal *Corynebacterium* species on the skin, their full potential in mediating interactions within the microbiome is likely much more vast.

To identify functions important for skin colonization by commensal *Corynebacterium* species, we performed a comparative genomics analysis of 71 species from the *Corynebacterium* genus, including species from a diverse host and environment niche range, and have generated the most up-to-date phylogeny of the *Corynebacterium* genus. Our analysis revealed that a subset of skin-associated *Corynebacterium* species encode for *de novo* biosynthesis of cobamides, a class of cofactor that is essential for metabolism in many organisms across life but only produced by a small fraction of prokaryotes. Microbiome diversity analysis of taxonomic data from 585 published skin metagenomes revealed that the abundance of cobamide-producing *Corynebacterium* species is associated with higher microbiome diversity, suggesting a role for cobamides in mediating skin microbiome community dynamics. We then analyzed 906 publicly available skin metagenomes to predict cobamide dependence and biosynthesis within the skin microbiome and found that phylogenetically diverse skin taxa are likely to use cobamides. Taken together, our results suggest that cobamide sharing occurs within the skin microbiome and hypothesize that this is an area of importance in studying community dynamics.

METHODS

Corynebacterium comparative genomics

71 *Corynebacterium* isolate genomes were acquired either from the National Center for Biotechnology Information (NCBI) as complete assemblies or from human skin isolates as draft assemblies. Supplementary Material S1 reports accession numbers and other information for each isolate genome, including ecosystem association, which was assigned using genome metadata and species-specific literature. The pangenomics workflow from anvi'o v6.2 (<http://merenlab.org/2016/11/08/pangenomics-v2/>) [24,25] was used for comparative genomics analysis. Briefly, genomes were annotated using 'anvi-run-ncbi-cogs', which assigns functions from the Clusters of Orthologous Groups (COGs) database. The *Corynebacterium* pangenome was computed using the program 'anvi-pan-genome' was used with the flags '--minbit 0.5', '--mcl-inflation 6', and '--enforce-hierarchical-clustering'. Average nucleotide identity between genomes

was calculated using pyani within the anvi'o environment (<https://github.com/widdowquinn/pyani>) [26]. The program 'anvi-get-enriched-functions-per-pan-group' was utilized to identify enriched COGs between host- and environment-associated genomes. Genome summary statistics are presented in Table S1.

***Corynebacterium* phylogenetic analysis**

The anvi'o phylogenomics workflow (<http://merenlab.org/2017/06/07/phylogenomics/>) was used to create a *Corynebacterium* phylogeny. Within the anvi'o environment, single copy genes (SCGs) from the curated anvi'o collection Bacteria_71 were identified within each genome using hmmscan from the HMMER package (<http://hmmer.org/>), and the SCG amino acid sequences were concatenated and aligned using MUSCLE [27]. A phylogenomic tree was constructed using FastTree [28] within anvi'o. To identify cobamide biosynthesis genes within the 71 *Corynebacterium* genomes, 46 KEGG orthology (KO) identifiers from KEGG map00860 (porphyrin and chlorophyll metabolism) were used to create a custom profile for KOfamsan, a functional annotation program based on KO and HMMs [29]. Each genome was queried against this profile, and hits to the KOs above the predefined KOfamsan threshold were considered for further analysis. Visualization of the phylogenomic tree and cobamide biosynthesis pathway completeness was performed in R with the ggtree package [30].

Microbiome diversity analysis of *Corynebacterium* cobamide producers

Using the taxonomic abundance information from 585 skin metagenomes from Oh *et al.* [8], the total sum abundance of 5 *Corynebacterium* species (*C. amycolatum*, *C. kroppenstedtii*, *C. ulcerans*, *C. pseudotuberculosis*, *C. diphtheriae*) shown to encode for *de novo* cobamide biosynthesis was calculated for each metagenome. Richness was determined by calculating the total number of observed species within each sample. Alpha diversity was determined by calculating the Shannon index using the diversity() function from the phyloseq v1.26.1 package in R. The Spearman correlation between *Corynebacterium* cobamide producer abundance and alpha diversity was calculated and plotted with ggscatter() and stat_cor() from the ggpubr v0.4.0 package. For beta diversity analysis, relative abundances were square root transformed to give more weight to low abundance taxa, and the Bray-Curtis dissimilarity index was calculated for samples within each skin site using vegdist() from the vegan v2.5-6 package. The indices were ordinated using non-metric multidimensional scaling with the vegan metaMDS() program.

Choice of profile HMMs for skin metagenome survey of cobamide biosynthesis and use

Profile HMMs, retrieved from the TIGRfam and Pfam databases, were used to detect cobamide biosynthesis, cobamide transport, and cobamide-dependent genes within skin metagenomic sequencing data. A total of 12 cobamide biosynthesis genes were selected because of their broad distribution throughout both the aerobic and anaerobic biosynthesis pathways and their presence within taxonomically-diverse cobamide producer genomes [31–33]. CbiZ was included as a marker of cobamide remodeling, and BtuB was included to assess cobamide transport. 19 cobamide-dependent enzymes and proteins with B₁₂-binding domains were chosen to evaluate cobamide use. The single copy gene *rpoB* was used as a phylogenetic marker to assess microbial community structure within each metagenome and as a proxy for sequence depth. All cobamide-associated genes used in this analysis can be found in Supplementary Material S3.

Metagenomic sequence search using HMMER

Raw sequencing data from 906 skin metagenomes was retrieved from the Sequence Read Archive (SRA) and converted to FASTA format, retaining only forward read files for analysis (Supplemental Table S2). Metagenomes were translated to each of 6 frame translations using transeq from the emboss v6.6.0 package [34]. The program hmmsearch from HMMER v3.3.1 [35] was used with default parameters and an E-value cutoff of 1E-06 to scan the metagenomic sequencing reads for homology to each cobamide-related HMM. The resulting hits were taxonomically classified to the family, genus, and species levels using Kraken 2 [36] and Bracken [37]. The number of hits for each gene was normalized to HMM length when analyzing individual metagenomes and to both HMM length and sequencing depth when analyzing groups of metagenomes. Taxonomic frequency profiles were generated for each cobamide-related gene by dividing the normalized number of gene hits per taxon by the total normalized number of gene hits.

Metagenomic sequence search using INFERNAL

Covariance models (CMs) for 3 cobalamin riboswitches from the Rfam clan CL00101 were retrieved from the Rfam database [38] (Supplementary Material S4). The program cmsearch from INFERNAL v1.1.2 [39] was used with default parameters and an E-value cutoff of 1E-06 to scan the metagenomes for RNA homologs to cobalamin riboswitches. The methods following hit identification are the same as described above for HMM analysis, except that the number of hits for each riboswitch were not normalized by CM length because the read lengths and CM lengths were relatively similar.

Mapping of cobalamin riboswitch hits to genomes

Reads identified from INFERnal were aligned against the genomes of *Cutibacterium acnes* KPA171202, *Veillonella parvula* DSM 2008, *Corynebacterium* sp. ATCC 6931, *Pseudomonas putida* KT2440, and *Streptococcus sanguinis* SK36 using bowtie2 v2.3.5.1 and visualized in R with the ggbio v1.30.0 package. Genes upstream and downstream of the aligned reads within each genome were assigned functions based on NCBI RefSeq annotations and visualized using the gggenes v0.4.0 R package. Genes within genomic regions that encoded for a cobalamin riboswitch but had no genes currently known to be under cobalamin riboswitch control were assigned putative functions based on NCBI BLAST search results.

RESULTS

Ecosystem association drives *Corynebacterium* genetic diversity.

Species within the *Corynebacterium* genus occupy highly diverse habitats, including soil, cheese rinds, coral mucus, and human skin, with certain species causing serious disease in humans and animals. To explore the genomic diversity within the *Corynebacterium* genus, we performed a pangenomic analysis using anvi'o. These analyses included 50 host-associated (HA) and 21 environment-associated (EA) genomes (Supplemental Table S1), acquired as complete assemblies from NCBI (n=68) or as draft assemblies from human skin strains we isolated (n=3). Gene clusters (GCs), which represent one or more genes contributed by one or more genomes, were inferred using the anvi'o anvi-pan-genome function. Across all species, we identified 42,154 total GCs, 495 of which are core GCs present in all genomes. In a subset of genomes, 13,235 GCs are shared (dispensable) and 28,424 GCs are found in only one genome (species-specific) (Figure 1). Genome size ranged from 2.0 to 3.6 Mbp, with an average of 2.7 ± 0.3 Mbp, and the number of GCs per genome ranged from 1858 to 3170 GCs, with an average of 2365 ± 294 GCs (Supplemental Table S1). Further, HA species have a significantly reduced average number of GCs per genome compared to EA species (2242 vs. 2661, p-value<0.0001) and a significantly reduced genome length (2.6 Mbp vs 3.0 Mbp, p-value<0.0001) (Figure 2A, 2B).

Because we observed significant differences in gene cluster number and genome size between HA and EA genomes, we hypothesized that HA and EA *Corynebacterium* species would form phylogenetically distinct clades. To test this hypothesis, we generated a *Corynebacterium* phylogenetic tree based on 71 conserved single copy genes. We found that generally, EA and HA species were interspersed throughout the phylogeny, with a few clear HA- or EA-specific

subgroups (Figure 3). For example, the HA *C. amycolatum* subgroup forms the deepest subline within this genus, consistent with previous findings [40]. Species within this subgroup are unique compared to most other Corynebacteria in that they do not contain mycolic acids in their cell walls. In addition, the clade containing *C. diphtheriae* and *C. pseudotuberculosis* also consists uniquely of HA species, several of which are serious pathogens in humans and other animals. Furthermore, a distinct lineage of EA soil dwelling Corynebacteria is formed by several species including *C. glutamicum*, which is an industrial producer of amino acids and other diverse products. We also observed that genome length was reflected by the phylogeny, where species within the same clade possessed similar genome length and EA clades having notably larger genomes (Figure 3).

De novo cobamide biosynthesis is enriched in host-associated Corynebacteria.

To identify functional gene predictions that differ between HA and EA genomes, we performed a functional enrichment analysis in *anvi'o* using Clusters of Orthologous Groups (COG) annotations. Within the top significantly enriched functions in EA genomes, we observed functions putatively involved in amino acid transport, metabolism of various substrates, including aromatic compounds, tetrahydropterin cofactors, citrate/malate, and alcohols, and other uncharacterized functions ($q < 0.05$) (Figure 2D). Within HA genomes, we observed a significant enrichment of functions involved in the transport of various substrates, as well as 8 COG functions putatively involved in cobamide biosynthesis in HA genomes ($q < 0.05$) (Fig. 2C). Cobamides, which consist of the vitamin B₁₂ family of cofactors, are required for metabolism by organisms across all domains of life. Specifically, they function in the catalysis of diverse enzymatic reactions, ranging from primary and secondary metabolism, including methionine and natural product synthesis, to environmentally impactful processes, such as methanogenesis and mercury methylation.

To validate the presence of the enriched cobamide biosynthesis genes and other genes of the approximately 25-enzyme cobamide biosynthesis pathway within the 71 *Corynebacterium* genomes, we scanned the genomes for these genes using KOfamScan [29]. We found that tetrapyrrole precursor synthesis, which is shared among the cobamide, heme, and chlorophyll biosynthesis pathways [31], was highly conserved throughout the genus (Figure 3). Corrin ring and nucleotide loop synthesis was predominantly intact and conserved within 5 distinct *Corynebacterium* lineages, including those of *C. diphtheriae*, *C. epidermidicans*, *C. argentoratense*, *C. kroppenstedtii*, and *C. amycolatum*. The species within these groups encode for all or nearly all of the genes required for cobamide biosynthesis, and notably, 21 out of 22 of

these predicted cobamide producers are host-associated (Figure 3). In addition, three EA species encode for the later portion of the pathway involved in nucleotide loop assembly, which can function in cobinamide salvaging. Taken together, these observations demonstrate a range of cobamide biosynthetic capabilities by *Corynebacteria*, including *de novo* producers, cobinamide salvagers, and non-producers.

Microbiome diversity is associated with the abundance of *Corynebacterium* cobamide producers. While 86% of bacteria have been found to encode at least one cobamide-dependent enzyme, only 37% of bacteria can produce the cofactor *de novo* [31]. Therefore, within microbial communities, cobamide sharing likely exists as a means to fulfill this nutritional requirement and is hypothesized to mediate community dynamics. Because several skin-associated *Corynebacteria* were identified to encode for *de novo* cobamide biosynthesis, we hypothesize that cobamide sharing occurs in the skin microbiome and mediates community dynamics. To assess changes at the community level associated with the presence of *Corynebacterium* cobamide producers, we explored the relationship between microbiome diversity and the abundance of *Corynebacterium* species producing cobamides. To do this, we analyzed published relative abundance data from 585 skin metagenomes [8]. Within each metagenome, we summed the abundance of five *Corynebacterium* species that encode for *de novo* cobamide biosynthesis (*C. amycolatum*, *C. kroppenstedtii*, *C. diphtheriae*, *C. ulcerans*, *C. tuberculosis*) and analyzed diversity within each community. We found that the abundance of these cobamide producers was positively correlated with alpha diversity (Shannon index) in 12/18 skin sites ($p < 0.05$) (Figure 4A), suggesting that increased abundance of cobamide-producing *Corynebacteria* (CPC) within the community may increase diversity. We also observed a positive correlation between CPC abundance and community richness, although only in 4/18 skin sites ($p < 0.05$) (Supplemental Figure 1). The observation that CPC abundance overall has a higher correlation with Shannon diversity, a metric taking into account total number of species and their overall proportion, compared to total species counts suggests that the increase in diversity we observe arises predominantly from an increase in species evenness within a given community. Supporting this argument, we found that communities with a low CPC abundance were usually dominated by single species including *Cutibacterium acnes* (formerly *Propionibacterium acnes*) and *Propionibacterium* phage, while communities with high CPC showed an expansion of other skin taxa and an overall more even species distribution within the community (Supplemental Figure 2).

To determine if CPC also shapes community composition, we calculated the Bray-Curtis dissimilarity index of square-root transformed abundance data of all samples within each skin site. A square root transformation was applied in order to give less weight to dominant species and more weight to low abundant species, which were prevalent in the dataset. The resulting matrices were ordinated by non-metric multidimensional scaling to visualize relatedness between individual communities. Within skin sites, particularly those that demonstrated a positive correlation between Shannon diversity and CPC abundance, samples clustered along gradients of both alpha diversity and CPC abundance (Figure 4B). This was especially evident for samples in the alar crease, back, manubrium, and inguinal crease, each of which had the strongest correlation coefficients between Shannon diversity and CPC abundance ($R=0.82$, $R=0.87$, $R=0.67$, $R=0.79$ respectively) (Figure 4A).

Cobamide biosynthesis, salvage, and remodeling genes are encoded by skin taxa.

Having identified that several skin-associated *Corynebacterium* species encode for *de novo* cobamide biosynthesis, we were interested in determining if other members within the skin microbiome also had this genomic capacity. We searched for cobamide biosynthesis genes within 906 skin metagenomes, encompassing samples from 21 distinct skin sites of 2 independent skin microbiome surveys [8,78] (Supplemental Table S2). We used 12 profile HMMs representing genes that demonstrated broad distribution within the *de novo* cobamide biosynthesis pathway and have been previously used to identify cobamide biosynthesis genes within metagenomic data [32,33]. Among site microenvironments, sebaceous sites were revealed to harbor the overall highest number of hits to cobamide biosynthesis genes ($n=139,128$ reads), followed by moist sites ($n=36,845$ reads), dry sites ($n=31,645$ reads), and foot sites ($n=6,411$ reads) (Supplemental Material S5). To assess the contribution of different taxa to cobamide biosynthesis, the metagenomic sequence classifier pipeline Kraken and Bracken was used to taxonomically classify the resulting gene hits. We identified the top taxa encoding for biosynthetic genes in descending order to be Propionibacteriaceae ($n=190,437$ reads), Pseudomonadaceae ($n=4,248$ reads), Corynebacteriaceae ($n=3,274$ reads), Veillonellaceae ($n=2036$ reads), and Streptococcaceae ($n=1,337$ reads) (Supplemental Material S6). Within individual metagenomes, we calculated the contribution of each taxa to cobamide biosynthesis gene hits by dividing the number of gene hits assigned to a given taxa by the total number of gene hits within the sample. We found that Propionibacteriaceae was the dominant contributor to cobamide biosynthesis gene hits, particularly in sebaceous sites (Figure 5). However, metagenomes from moist, dry, and foot

sites exhibited higher variability in Propionibacteriaceae contribution, which was also reflected in variable contribution by the other four taxa.

For a finer resolution of taxa contribution across individual biosynthesis genes, we determined taxa contribution to 12 cobamide biosynthesis marker genes for the top 40 abundant taxa within the dataset. This analysis revealed that the number of taxa encoding the full or nearly complete suite of cobamide biosynthesis markers varied according to the microenvironment (Figure 6). In particular, cobamide biosynthetic potential within the sebaceous microenvironment is uniquely restricted to Propionibacteriaceae, whereas numerous taxa within the moist, dry, and foot microenvironments encode for most cobamide biosynthesis markers. These include diverse taxa across the Firmicutes, Actinobacteria, and Proteobacteria phyla, such as Pseudomonadaceae, Veillonellaceae, and Rhodobacteraceae, all of which have previously been implicated in cobamide biosynthesis [41–43]. Corynebacteriaceae were also found to encode for nearly all cobamide biosynthesis markers, validating our comparative genomics analysis that identified specific *Corynebacterium* species as *de novo* producers. We also observed that a proportion of low abundant skin taxa encode for the full suite of cobamide biosynthesis markers, grouped into the “Other” category (<0.15% abundance based on rpoB hits), demonstrating that cobamide biosynthesis is likely encoded by low abundant or rare taxa as well. Overall, our analysis suggests the dominant cobamide producers on the skin are Propionibacteriaceae and Corynebacteriaceae, with other low abundant or rare taxa such as Veillonellaceae and Pseudomonadaceae contributing to biosynthesis in a site dependent manner [8].

Furthermore, certain taxa such as Moraxellaceae and Xanthomonadaceae only encode for cobamide biosynthesis genes cobP/cobU and cobS (Figure 6), which function in salvaging cobinamide from the environment [44]. This suggests that in addition to specific taxa producing cobamides *de novo*, there are other taxa that synthesize complete cobamides through salvaging cobamide precursors from the environment. In addition, certain taxa, such as Veillonellaceae, encode for adenosylcobinamide amidohydrolase CbiZ, which is involved in remodeling cobamides with a new lower ligand [44,45]. Most predicted cobamide producers identified in this analysis likely synthesize benzimidazole cobamides because they encode for bluB, the gene responsible for the aerobic synthesis of lower ligand 5,6-dimethylbenzimidazole (DMB) (Figure 6) [46–48]. Therefore, the identification of taxa that encode for cobamide remodeling enzymes suggests that structurally distinct cobamides are produced and used for microbial metabolism on the skin. Overall, these results demonstrate that certain skin taxa have the genetic potential to

produce complete cobamides through *de novo* biosynthesis, precursor salvage, or lower ligand remodeling.

Phylogenetically diverse skin taxa encode for cobamide-dependent enzymes and cobalamin riboswitches.

Our findings suggest that several taxa within the skin microbiome synthesize cobamides *de novo*. Therefore, cobamides and cobamide precursors are likely available in the community for uptake and use. To evaluate the potential utilization of cobamides by other taxa of the skin microbiome, we searched 906 skin metagenomes for the cobamide transport protein *btuB* and 19 other enzymes that carry out diverse cobamide dependent reactions (Supplemental Material S3). While the total number of cobamide-dependent gene hits varied by microenvironment (sebaceous=174,096, moist=70,994, dry=66,599, foot=45,412) (Supplemental Material S5), the relative proportion of taxa encoding the genes was similar across microenvironments (Figure 7A). Across sebaceous, moist, and dry microenvironments, *Propionibacteriaceae* was the dominant taxa encoding for cobamide-dependent enzymes D-ornithine aminomutase, methylmalonyl-CoA mutase, and ribonucleotide reductase class II, with this taxon contributing overall the most cobamide-dependent sequences within the dataset (Supplemental Table S7). In contrast, we found across the remaining cobamide-dependent enzymes, hits were assigned to phylogenetically diverse taxa across the four major phyla on the skin (Actinobacteria, Firmicutes, Proteobacteria, and Bacteroidetes) [17]. In particular, cobamide-dependent methionine synthase, epoxyqueosine reductase, ethanolamine lyase, and ribonucleotide reductase were the most common cobamide-dependent enzymes in the dataset, when considering the number of unique taxa that encode for these enzymes (Supplemental Figure 3A). Notably, more unique species within the dataset encode for at least one cobamide-dependent enzyme ($n=456$ species contributing at least five reads to any cobamide-dependent enzyme) than appear to appreciably contribute to *de novo* cobamide biosynthesis on the skin ($n=23$ species contributing at least five reads to at least 7/10 biosynthesis genes) (Supplemental Table S9). This supports a model for cobamide sharing, where a much larger number of skin taxa require cobamides than can produce the cofactor *de novo*.

Cobalamin riboswitches are elements found in bacterial mRNAs that regulate expression of cobamide-related transcripts through cobamide binding [49–51]. To further validate the use of cobamides within the skin microbiome, we searched the skin metagenomes for cobalamin

riboswitches from the Rfam clan CL00101 (Supplemental Material S4). Similar to the cobamide biosynthesis and cobamide-dependent gene analyses, phylogenetically diverse skin taxa encode for cobalamin riboswitches, with Propionibacteriaceae being the dominant taxa. (Figure 7B). At the species level, these hits were shown to be contributed predominantly by *Cutibacterium acnes* (Supplemental Material S8).

To further explore the observed role that cobamides play in *C. acnes*, we mapped the identified riboswitch metagenomic reads to the *C. acnes* KPA171202 reference genome and found that the reads mapped to several genomic regions, with nearby genes having functions involved in ABC transport, cobalt transport, cobamide biosynthesis, and cobamide-dependent and -independent reactions (Figure 8A). Regions 6, 7, and 8 did not encode for known cobamide-related genes, but all were found directly upstream of a small ~250 bp pseudogene and either a 1321 bp pseudogene (Region 6) or 1922 bp hypothetical protein (Regions 7 and 8). BLAST searches of the sequences revealed that the small pseudogenes are hypothetical adhesin protein fragments and the larger sequences that directly follow are thrombospondin type-3 repeat containing proteins (Supplemental Material S10). The role of cobalamin riboswitches in regulating these sequences is not clear. We also mapped the riboswitch-containing reads to the genomes of other skin taxa, finding that there were overall fewer cobalamin riboswitches within these genomes compared to *C. acnes*, but were similarly found nearby to genes with functions in cobamide biosynthesis, ABC transport, cobalt transport, and both cobamide-dependent and cobamide-independent isozymes (Figure 8B, C, D, E). Thus, within the skin microbiome, cobalamin riboswitches are likely to regulate diverse processes, including those that may not currently be characterized.

DISCUSSION

Corynebacterium species are well-equipped for growth on the skin due to their “lipid-loving” and halotolerant nature, allowing them to thrive in moist and sebaceous skin microenvironments [52]. However, many questions remain about the processes that govern skin colonization by this relatively understudied skin taxa and how these processes may impact or be impacted by microbe-microbe and microbe-host interactions on the skin.

Our *Corynebacterium* comparative genomics analysis revealed that host-associated Corynebacteria have significantly smaller genomes compared to environment-associated genomes and further, that distinct functions are enriched between the two groups (Figure 2). For

free-living microorganisms like EA *Corynebacteria*, selection likely drives the retention of metabolic functions, presumably due to inconsistent availability of certain nutrients and other compounds in the environment [53]. Whereas in microorganisms in association with a host, the host provides a relatively consistent supply of nutrients and metabolic intermediates, therefore eliminating the pressure to retain certain metabolic functions, leading to gene loss over time [54]. Therefore, we predict that the reduced genome size of HA *Corynebacterium* species arose from loss of metabolic functions, which is supported by the observation that EA genomes are enriched in functions associated with diverse substrate metabolism. A rather unexpected finding from our study was the enrichment of *de novo* cobamide biosynthesis within HA genomes. Retention of the energetically costly 25-enzyme cobamide biosynthesis pathway within HA *Corynebacterium* species, even with reduced genome size, suggests that synthesis of this cofactor is advantageous for host niche colonization.

While decades-old literature have implicated *Corynebacterium* species in *de novo* cobamide biosynthesis [55–57], only a few recent studies have explored this relatively rare function by members of the genus [31,58–60]. Here, we demonstrate that within the *Corynebacterium* genus, the *de novo* cobamide biosynthesis pathway is conserved among specific lineages, including those that contain skin-associated species (Figure 3). However, the existence of other lineages that do not encode for the complete suite of genes required for biosynthesis suggests that there has been evolutionary pressure for either retention or loss of the pathway. The presence of remnant cobamide biosynthesis genes within the *C. glutamicum* clade and the conserved nature of cobU/cobT and cobC throughout the genus, which are currently known to only function in cobamide biosynthesis [61–63], supports the likelihood of pathway loss.

A key question that arises is why some *Corynebacterium* species have retained the *de novo* cobamide biosynthesis pathway, while others have not. From our analysis of skin metagenomes, we observed reads encoding for cobamide-dependent methionine synthase, methylmalonyl-CoA mutase, and ethanolamine ammonia lyase from the *Corynebacterium* genus, which is consistent with previous findings by Shelton *et al.* [31]. Therefore, cobamides are likely produced to fulfill metabolic requirements in methionine, propionate, and glycerophospholipid metabolism in *Corynebacteria*. Alternative cobamide-independent pathways exist for these functions, therefore cobamides may confer a distinct advantage for *Corynebacterium* cobamide-producing species. Indeed, metE, the cobamide-independent methionine synthase, is sensitive to oxidative stress and has reduced turnover compared to metH [64–66]. The skin in particular is subject to high

oxidative stress as a result of metabolic reactions, cosmetics, and UV irradiation exposure. [67–69]. Therefore, while the significance of employing cobamide-dependent vs -independent isozymes for bacterial metabolism on the skin is unknown, inherent features of the skin such as high oxidative stress may play a role.

While other studies of cobamides in microbial communities have demonstrated that *de novo* synthesis is carried out by a relatively small fraction of the community [33,70,71], our analysis of skin metagenomic data suggests that on the skin, cobamides are produced by taxa considered dominant within the microbiome, including *Cutibacterium* and *Corynebacterium* species. We identified that overall, *C. acnes* is the top cobamide producer and user within our dataset and is likely under tight regulation of cobamide biosynthesis and cobamide-dependent enzyme expression by more than 10 cobalamin riboswitches. Indeed, cobalamin riboswitch regulation has been demonstrated to decrease expression of *C. acnes* cobamide biosynthesis genes in the presence of exogenous vitamin B₁₂, both *in vitro* and on the skin of healthy individuals [72]. We also identified that three of the *C. acnes* cobalamin riboswitches are located upstream of proteins containing thrombospondin type 3 repeats (TT3Rs), which are efficient calcium ion binding motifs that have been well-studied in eukaryotes, but not prokaryotes [73]. TT3Rs that have been investigated in prokaryotes have been associated with the outer membrane of Gram-negative bacteria, therefore in the Gram-positive *C. acnes*, the role for TT3R-containing proteins is unknown and likely distinct. We predict that on the skin, cobalamin riboswitches could regulate functions not currently characterized to have a cobamide association.

Interestingly, we observed that other predicted cobamide producers on the skin, including *V. parvula* and *Corynebacterium* sp. ATCC 6931, only possess a few cobalamin riboswitches, and these riboswitches appear to regulate cobamide-dependent and -independent functions and ABC transport, as opposed to cobamide biosynthesis. This could suggest a more constitutive expression of cobamide biosynthesis by certain skin taxa, providing a possible explanation for the observed increase in microbiome diversity with cobamide-producing *Corynebacteria* as compared to the low diversity communities with high *C. acnes* abundance. Overall, cobamide production and riboswitch regulation are likely to act as mediators of microbe-microbe interactions on the skin.

In support of a key role for cobamides within the skin microbiome, we found that other phylogenetically diverse skin taxa, both dominant and low abundance, encode for metabolically

diverse cobamide-dependent enzymes, as well as proteins involved in cobamide transport, salvage, and remodeling (Figures 6 and 7). Because cobamide structure can dictate microbial growth and metabolism [74], microorganisms exhibit cobamide selectivity, employing numerous mechanisms for acquiring and using their preferred cobamide(s). These include selectivity in cobamide-dependent enzymes, differential cobamide import, and cobamide-specific gene regulation [74]. While cobamides are not likely to be scarce on the skin based on the observation that dominant taxa encode for *de novo* biosynthesis, the specific cobamides required for skin microbial metabolism is unknown. The presence of cobamide remodeling enzyme *cbiZ*, which is encoded by some skin taxa, may serve as an indicator that certain skin bacteria use cobamide-dependent enzymes with specific lower ligand preferences [45], thus implicating a role for cobamide salvaging and sharing within the skin microbiome in fulfilling metabolic requirements.

Within microbial communities, cobamides are hypothesized to mediate community dynamics because of the relative paucity of cobamide producers, yet evident requirement for this cofactor across the bacterial domain of life [31,74,75]. Our results suggest that on the skin, *Corynebacterium* cobamide-producing species may promote microbiome diversity (Figure 4) through an increase in community alpha diversity that arises from an overall more evenly distributed community. Interestingly, we observed that communities with low abundance of cobamide-producing *Corynebacteria* and thus low diversity were dominated by *Propionibacterium* species, predicted to be the dominant cobamide producer on the skin. Because we anticipate communities that produce more cobamides to have increased diversity, this result is confounding. However, because a spectrum of ecological niches exists on the skin, ranging from the anaerobic sebaceous follicle to the aerobic and desiccate skin surface, we hypothesize that cobamide interactions are dependent upon spatial structure of skin microbial communities. For example, *C. acnes* is an anaerobe that predominantly resides deep within the anaerobic sebaceous follicle [76], dominating between 60-90% of the follicle community [77]. As such, the opportunity for cobamide-mediated interactions is likely reduced as a result of the *C. acnes*-dominated sebaceous gland. Approaching the more oxygenated skin surface, the community becomes more diverse [18], thus increasing the incidence of cobamide interactions and subsequent effects on community dynamics. Furthermore, *C. acnes* and other related *Cutibacterium* and *Propionibacterium* species produce cobamides with DMB as the lower ligand, requiring the *bluB* gene. However, the *bluB* reaction must proceed under aerobic conditions. How *C. acnes* synthesizes complete cobamides within the anaerobic sebaceous follicle is unknown, but suggests a possible role for DMB sharing within the skin microbiome.

Overall, our findings reveal that skin-associated species within the *Corynebacterium* genus encode for the *de novo* cobamide biosynthesis pathway, which is a relatively rare metabolic function. Within skin microbial communities, abundance of these cobamide-producing *Corynebacteria* is correlated with increased microbiome diversity, suggesting that cobamides are important mediators of microbiome structure. We also demonstrate that cobamide use is widespread across phylogenetically diverse skin taxa, therefore we hypothesize that cobamides are likely to play a significant role in skin microbial community dynamics. Future studies to interrogate cobamide interactions between skin commensals will provide insight into the effects of cobamide biosynthesis and use within microbial communities.

REFERENCES

1. Scharschmidt TC, Vasquez KS, Pauli ML, Leitner EG, Chu K, Truong HA, et al. Commensal Microbes and Hair Follicle Morphogenesis Coordinately Drive Treg Migration into Neonatal Skin. *Cell Host Microbe*. 2017. doi:10.1016/j.chom.2017.03.001
2. Constantinides MG, Link VM, Tamoutounour S, Wong AC, Perez-Chaparro PJ, Han S-J, et al. MAIT cells are imprinted by the microbiota in early life and promote tissue repair. *Science*. 2019;366. doi:10.1126/science.aax6624
3. Linehan JL, Harrison OJ, Han S-J, Byrd AL, Vujkovic-Cvijin I, Villarino AV, et al. Non-classical Immunity Controls Microbiota Impact on Skin Immunity and Tissue Repair. *Cell*. 2018;172: 784–796.e18.
4. Belkaid Y, Harrison OJ. Homeostatic Immunity and the Microbiota. *Immunity*. 2017;46: 562–576.
5. Di Domizio J, Belkhodja C, Chenuet P, Fries A, Murray T, Mondéjar PM, et al. The commensal skin microbiota triggers type I IFN-dependent innate repair responses in injured skin. *Nat Immunol*. 2020;21: 1034–1045.
6. Wanke I, Steffen H, Christ C, Krismer B, Götz F, Peschel A, et al. Skin commensals amplify the innate immune response to pathogens by activation of distinct signaling pathways. *J Invest Dermatol*. 2011;131: 382–390.

- 564 7. Claesen J, Spagnolo JB, Ramos SF, Kurita KL, Byrd AL, Aksenov AA, et al. A
565 Cutibacterium acnes antibiotic modulates human skin microbiota composition in hair
566 follicles. Sci Transl Med. 2020;12. doi:10.1126/scitranslmed.aay5445
- 567 8. Oh J, Byrd AL, Park M, Kong HH, Segre JA. Temporal Stability of the Human Skin
568 Microbiome. Cell. 2016;165: 854–866.
- 569 9. Tomida S, Nguyen L, Chiu B-H, Liu J, Sodergren E, Weinstock GM, et al. Pan-genome
570 and comparative genome analyses of propionibacterium acnes reveal its genomic diversity
571 in the healthy and diseased human skin microbiome. MBio. 2013;4: e00003–13.
- 572 10. Wollenberg MS, Claesen J, Escapa IF, Aldridge KL, Fischbach MA, Lemon KP.
573 Propionibacterium-produced coproporphyrin III induces Staphylococcus aureus
574 aggregation and biofilm formation. MBio. 2014;5: e01286–14.
- 575 11. Christensen GJM, Scholz CFP, Enghild J, Rohde H, Kilian M, Thürmer A, et al.
576 Antagonism between Staphylococcus epidermidis and Propionibacterium acnes and its
577 genomic basis. BMC Genomics. 2016;17: 152.
- 578 12. O’Sullivan JN, Rea MC, O’Connor PM, Hill C, Ross RP. Human skin microbiota is a rich
579 source of bacteriocin-producing staphylococci that kill human pathogens. FEMS Microbiol
580 Ecol. 2019;95. doi:10.1093/femsec/fiy241
- 581 13. Nakatsuji T, Chen TH, Narala S, Chun KA, Two AM, Yun T, et al. Antimicrobials from
582 human skin commensal bacteria protect against Staphylococcus aureus and are deficient
583 in atopic dermatitis. Sci Transl Med. 2017;9. doi:10.1126/scitranslmed.aah4680
- 584 14. Naik S, Bouladoux N, Wilhelm C, Molloy MJ, Salcedo R, Kastenmuller W, et al.
585 Compartmentalized control of skin immunity by resident commensals. Science. 2012;337:
586 1115–1119.
- 587 15. Gallo RL, Nakatsuji T. Microbial symbiosis with the innate immune defense system of the
588 skin. J Invest Dermatol. 2011;131: 1974–1980.
- 589 16. Brandwein M, Bentwich Z, Steinberg D. Endogenous Antimicrobial Peptide Expression in
590 Response to Bacterial Epidermal Colonization. Front Immunol. 2017;8: 1637.
- 591 17. Grice EA, Segre JA. The skin microbiome. Nat Rev Microbiol. 2011;9: 244–253.

18. Oh J, Byrd AL, Deming C, Conlan S, NISC Comparative Sequencing Program, Kong HH, et al. Biogeography and individuality shape function in the human skin metagenome. *Nature*. 2014;514: 59–64.
19. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, et al. Topographical and temporal diversity of the human skin microbiome. *Science*. 2009;324: 1190–1192.
20. Bomar L, Brugger SD, Yost BH, Davies SS, Lemon KP. *Corynebacterium accolens* Releases Antipneumococcal Free Fatty Acids from Human Nostril and Skin Surface Triacylglycerols. *MBio*. 2016;7: e01725–15.
21. Ramsey MM, Freire MO, Gabriliska RA, Rumbaugh KP, Lemon KP. *Staphylococcus aureus* Shifts toward commensalism in response to *corynebacterium* species. *Front Microbiol*. 2016. doi:10.3389/fmicb.2016.01230
22. Altonsy MO, Kurwa HA, Lauzon GJ, Amrein M, Gerber AN, Almishri W, et al. *Corynebacterium tuberculostearicum*, a human skin colonizer, induces the canonical nuclear factor- κ B inflammatory signaling pathway in human skin cells. *Immun Inflamm Dis*. 2020;8: 62–79.
23. Ridaura VK, Bouladoux N, Claesen J, Chen YE, Byrd AL, Constantinides MG, et al. Contextual control of skin immunity and inflammation by *Corynebacterium*. *J Exp Med*. 2018;215: 785–799.
24. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ*. 2015;2015: e1319.
25. Delmont TO, Murat Eren A. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ*. 2018. p. e4320. doi:10.7717/peerj.4320
26. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods*. 2015;8: 12–24.
27. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32: 1792–1797.

- 619 28. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for
620 large alignments. PLoS One. 2010;5: e9490.
- 621 29. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al.
622 KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score
623 threshold. Bioinformatics. 2020;36: 2251–2252.
- 624 30. Yu G, Smith DK, Zhu H, Guan Y, Lam TT. ggtree : an r package for visualization and
625 annotation of phylogenetic trees with their covariates and other associated data. McInerney
626 G, editor. Methods Ecol Evol. 2017;8: 28–36.
- 627 31. Shelton AN, Seth EC, Mok KC, Han AW, Jackson SN, Haft DR, et al. Uneven distribution
628 of cobamide biosynthesis and dependence in bacteria predicted by comparative
629 genomics. ISME J. 2019;13: 789–804.
- 630 32. Doxey AC, Kurtz DA, Lynch MDJ, Sauder LA, Neufeld JD. Aquatic metagenomes
631 implicate Thaumarchaeota in global cobalamin production. ISME J. 2015;9: 461–471.
- 632 33. Lu X, Heal KR, Ingalls AE, Doxey AC, Neufeld JD. Metagenomic and chemical
633 characterization of soil cobalamin production. ISME J. 2020;14: 53–66.
- 634 34. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software
635 Suite. Trends Genet. 2000;16: 276–277.
- 636 35. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7: 1002195.
- 637 36. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome
638 Biol. 2019;20: 257.
- 639 37. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: Estimating species abundance in
640 metagenomics data. PeerJ Computer Science. 2017;2017: e104.
- 641 38. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam
642 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids
643 Res. 2018;46: D335–D342.
- 644 39. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches.
645 BIOINFORMATICS APPLICATIONS. 2013;29: 2933–2935.

40. Pascual C, Lawson PA, Farrow JA, Gimenez MN, Collins MD. Phylogenetic analysis of the genus *Corynebacterium* based on 16S rRNA gene sequences. *Int J Syst Bacteriol*. 1995;45: 724–728.
41. Pollich M, Klug G. Identification and sequence analysis of genes involved in late steps in cobalamin (vitamin B12) synthesis in *Rhodobacter capsulatus*. *J Bacteriol*. 1995;177: 4481–4487.
42. Fang H, Kang J, Zhang D. Microbial production of vitamin B12: a review and future perspectives. *Microb Cell Fact*. 2017;16: 15.
43. Hazra AB, Tran JLA, Crofts TS, Taga ME. Analysis of substrate specificity in CobT homologs reveals widespread preference for DMB, the lower axial ligand of vitamin B(12). *Chem Biol*. 2013;20: 1275–1285.
44. Gray MJ, Tavares NK, Escalante-Semerena JC. The genome of *Rhodobacter sphaeroides* strain 2.4.1 encodes functional cobinamide salvaging systems of archaeal and bacterial origins. *Mol Microbiol*. 2008;70: 824–836.
45. Gray MJ, Escalante-Semerena JC. The cobinamide amidohydrolase (cobyric acid-forming) CbiZ enzyme: a critical activity of the cobamide remodelling system of *Rhodobacter sphaeroides*. *Mol Microbiol*. 2009;74: 1198–1210.
46. Taga ME, Larsen NA, Howard-Jones AR, Walsh CT, Walker GC. BluB cannibalizes flavin to form the lower ligand of vitamin B12. *Nature*. 2007;446: 449–453.
47. Campbell GRO, Taga ME, Mistry K, Lloret J, Anderson PJ, Roth JR, et al. *Sinorhizobium meliloti* bluB is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of B12. *Proc Natl Acad Sci U S A*. 2006;103: 4634–4639.
48. Gray MJ, Escalante-Semerena JC. Single-enzyme conversion of FMNH₂ to 5,6-dimethylbenzimidazole, the lower ligand of B12. *Proc Natl Acad Sci U S A*. 2007;104: 2921–2926.
49. Garst AD, Edwards AL, Batey RT. Riboswitches: Structures and mechanisms. Cold Spring Harbor Perspectives in Biology. Cold Spring Harbor Laboratory Press; 2011. pp. 1–13. doi:10.1101/cshperspect.a003533

- 674 50. Nahvi A, Barrick JE, Breaker RR. Coenzyme B12 riboswitches are widespread genetic
675 control elements in prokaryotes. *Nucleic Acids Res.* 2004;32: 143–150.
- 676 51. Polaski JT, Webster SM, Johnson JE, Batey RT. Cobalamin riboswitches exhibit a broad
677 range of ability to discriminate between methylcobalamin and adenosylcobalamin. *J Biol*
678 *Chem.* 2017;292: 11650–11658.
- 679 52. Scharschmidt TC, Fischbach MA. What Lives On Our Skin: Ecology, Genomics and
680 Therapeutic Opportunities Of the Skin Microbiome. *Drug Discov Today Dis Mech.*
681 2013;10. doi:10.1016/j.ddmec.2012.12.003
- 682 53. Wernegreen JJ. Endosymbiont evolution: predictions from theory and surprises from
683 genomes. *Ann N Y Acad Sci.* 2015;1360: 16–35.
- 684 54. Moran NA. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr Opin*
685 *Microbiol.* 2003;6: 512–518.
- 686 55. Perlman D. Microbial Synthesis of Cobamides¹¹ Presented in part in lectures before
687 School of Pharmacy, University of Wisconsin, March, 1958. In: Umbreit WW, editor.
688 *Advances in Applied Microbiology.* Academic Press; 1959. pp. 87–122.
- 689 56. Andrianthosa M, Laget M, Cremieux A, Dumenil G. Constant fed-batch culture of methanol-
690 utilizing corynebacterium producing vitamin B 12. *Biotechnol Lett.* 1984;6: 783–788.
- 691 57. Fujii K, Takeuchi H, Shimizu S, Fukui S. Vitamin B12-dependent Methionine Biosynthesis
692 and Its Metabolic Role in *Corynebacterium simplex* ATCC 6946, a Vitamin B12-producing
693 and Hydrocarbon-utilizable Bacterium. *Agric Biol Chem.* 1972;36: 2323–2334.
- 694 58. dos S. Peinado R, Olivier DS, Eberle RJ, de Moraes FR, Amaral MS, Arni RK, et al.
695 Binding studies of a putative C . pseudotuberculosis target protein from Vitamin B 12
696 Metabolism. *Sci Rep.* 2019;9: 1–13.
- 697 59. Al-Dilaimi A, Albersmeier A, Kalinowski J, Rückert C. Complete genome sequence of
698 *Corynebacterium vitaeruminis* DSM 20294T, isolated from the cow rumen as a vitamin B
699 producer. *J Biotechnol.* 2014;189: 70–71.

60. Antonov IV. Two cobalt chelatase subunits can be generated from a single chlD gene via programmed frameshifting. *Mol Biol Evol.* 2020. Available: <https://academic.oup.com/mbe/article-abstract/doi/10.1093/molbev/msaa081/5811573>
61. Zayas CL, Escalante-Semerena JC. Reassessment of the late steps of coenzyme B₁₂ synthesis in *Salmonella enterica*: evidence that dephosphorylation of adenosylcobalamin-5'-phosphate by the CobC phosphatase is the last step of the pathway. *J Bacteriol.* 2007;189: 2210–2218.
62. O'Toole GA, Trzebiatowski JR, Escalante-Semerena JC. The cobC gene of *Salmonella typhimurium* codes for a novel phosphatase involved in the assembly of the nucleotide loop of cobalamin. *J Biol Chem.* 1994;269: 26503–26511.
63. Cheong C-G, Escalante-Semerena JC, Rayment I. The three-dimensional structures of nicotinate mononucleotide: 5, 6-dimethylbenzimidazole phosphoribosyltransferase (CobT) from *Salmonella typhimurium* complexed with 5, 6-dimethylbenzimidazole and its reaction products determined to 1.9 Å resolution. *Biochemistry.* 1999;38: 16125–16135.
64. González JC, Banerjee RV, Huang S, Sumner JS, Matthews RG. Comparison of cobalamin-independent and cobalamin-dependent methionine synthases from *Escherichia coli*: two solutions to the same chemical problem. *Biochemistry.* 1992;31: 6045–6056.
65. Hondorp ER, Matthews RG. Oxidative stress inactivates cobalamin-independent methionine synthase (MetE) in *Escherichia coli*. *PLoS Biol.* 2004;2: e336.
66. Leichert LI, Jakob U. Protein thiol modifications visualized in vivo. *PLoS Biol.* 2004;2: e333.
67. Andersson T, Ertürk Bergdahl G, Saleh K, Magnúsdóttir H, Stødkilde K, Andersen CBF, et al. Common skin bacteria protect their host from oxidative stress through secreted antioxidant RoxP. *Sci Rep.* 2019;9: 3596.
68. Kawashima S, Funakoshi T, Sato Y, Saito N, Ohsawa H, Kurita K, et al. Protective effect of pre- and post-vitamin C treatments on UVB-irradiation-induced skin damage. *Sci Rep.* 2018;8: 16199.

69. Hakoziaki T, Date A, Yoshii T, Toyokuni S, Yasui H, Sakurai H. Visualization and characterization of UVB-induced reactive oxygen species in a human skin equivalent model. *Arch Dermatol Res.* 2008;300 Suppl 1: S51–6.
70. Magnúsdóttir S, Ravcheev D, de Crécy-Lagard V, Thiele I. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Front Genet.* 2015;6: 148.
71. Romine MF, Rodionov DA, Maezato Y, Osterman AL, Nelson WC. Underlying mechanisms for syntrophic metabolism of essential enzyme cofactors in microbial communities. *ISME J.* 2017;11: 1434–1446.
72. Kang D, Shi B, Erfe MC, Craft N, Li H. Vitamin B 12 modulates the transcriptome of the skin microbiota in acne pathogenesis. *Sci Transl Med.* 2015;7: 293ra103–293ra103.
73. Dai S, Sun C, Tan K, Ye S, Zhang R. Structure of thrombospondin type 3 repeats in bacterial outer membrane protein A reveals its intra-repeat disulfide bond-dependent calcium-binding capability. *Cell Calcium.* 2017;66: 78–89.
74. Sokolovskaya OM, Shelton AN, Taga ME. Sharing vitamins: Cobamides unveil microbial interactions. *Science.* 2020;369: eaba0165.
75. Degnan Patrick H. Taga Michiko E. Goodman AL. Vitamin B12 as a modulator of gut microbial ecology. *Cell Metab.* 2014;20: 1–769–778.
76. Dréno B, Pécastaings S, Corvec S, Veraldi S, Khammari A, Roques C. *Cutibacterium acnes* (*Propionibacterium acnes*) and acne vulgaris: a brief look at the latest updates. *J Eur Acad Dermatol Venereol.* 2018;32: 5–14.
77. Hall JB, Cong Z, Imamura-Kawasawa Y, Kidd BA, Dudley JT, Thiboutot DM, et al. Isolation and Identification of the Follicular Microbiome: Implications for Acne Research. *J Invest Dermatol.* 2018;138: 2033–2040.
78. Hannigan GD, Meisel JS, Tyldsley AS, Zheng Q, Hodkinson BP, SanMiguel AJ, et al. The human skin double-stranded DNA virome: topographical and temporal diversity, genetic enrichment, and dynamic associations with the host microbiome. *MBio.* 2015;6: e01578–15.

FIGURE LEGENDS

Figure 1. *Corynebacterium* pangenome. Pangenome analysis generated with anvio. 42,154 gene clusters (combined core, dispensable, and singletons) were identified from 71 *Corynebacterium* genomes and are ordered by gene cluster frequency (opaque, present; transparent, absent). Each gene cluster contains one or more genes contributed by one or more genomes. Genomes are colored by ecosystem association and ordered by the phylogeny based on 71 single copy genes (unrooted). ANI scale (0.7-0.8). Singleton gene clusters (grey) are collapsed.

Figure 2. Host- and environment-associated *Corynebacterium* genomes have distinct genetic features. A) Genome length and B) number of gene clusters for 71 *Corynebacterium* genomes was determined using anvio. Significance was determined using Tukey's Test for post hoc analysis. C) Significantly enriched COG functions in C) host-associated or D) environment-associated genomes were identified with anvio. The top 20 significantly enriched COG functions ($q < 0.05$) are shown, ordered by ascending significance. Blue = host-associated, orange = environment-associated.

Figure 3. *De novo* cobamide biosynthesis is host-associated within the *Corynebacterium* genus. A *Corynebacterium* phylogenetic tree based on comparison of 71 conserved single copy genes was generated using FastTree within the anvio environment. The tree is rooted with *Tsukamurella paurometabola*. Species are colored by host (blue) or environment (orange) association, and by genome length (dark blue). KOfamScan was used to identify the presence (dark pink) or absence (light pink) of cobamide biosynthesis genes within each genome. Cobamide biosynthesis is separated into tetrapyrrole precursor, corrin ring, nucleotide loop, and lower ligand synthesis.

Figure 4. Microbiome diversity is associated with abundance of cobamide-producing *Corynebacterium* species. Within each metagenome, the total abundance of 5 cobamide-producing *Corynebacterium* (CPC) species was calculated. A) CPC abundance is plotted against the Shannon diversity indices for samples within each skin site. R =Spearman's correlation coefficient. B) NMDS plots based on Bray-Curtis indices for samples within each skin site are shown. Points are colored by log₁₀ *Corynebacterium* cobamide producer abundance and sized by alpha diversity (Shannon).

Figure 5. Taxonomic contribution of cobamide biosynthesis genes across skin sites is site-dependent. For each taxa shown, the number of cobamide biosynthesis gene hits (normalized by profile HMM coverages) assigned to the given taxa was divided by the overall total number of cobamide biosynthesis gene hits within the sample. Taxon contributions are shown for the top 5 taxa with the most cobamide biosynthesis gene hits, grouped by skin site. Color indicates microenvironment classification.

Figure 6. Many skin taxa encode for genes required for *de novo* cobamide biosynthesis. The top 40 abundant taxa within the dataset were determined by totaling the hits to single copy gene *rpoB* (normalized by profile HMM coverage). The remaining taxa were grouped into “Other”. Individual values in the heatmap represent the number of hits assigned to the taxon for a particular cobamide biosynthesis gene divided by the total number of hits to the gene. Gene hits were normalized by profile HMM coverage and sequencing depth prior to calculation.

Figure 7. Phylogenetically diverse skin bacteria encode cobamide-dependent enzymes, cobamide transporters, and cobalamin riboswitches A) The total normalized hits for 17 cobamide-dependent enzymes and cobamide transport protein *btuB* are shown (totals hits normalized to profile HMM coverage and sequence depth), with the taxonomic abundance of the hits expanded above. Hits to distinct B12-dependent radical SAM proteins are grouped together as “B12-dep radical SAM”. B) The taxonomic abundance of hits for cobalamin riboswitches (Rfam clan CL00101) are shown, with an expanded view of low abundance hits to the right. Total cobalamin riboswitch hits within each microenvironment are indicated.

Figure 8. Cobalamin riboswitches regulate diverse functions within skin bacteria. Cobalamin riboswitch-containing reads identified from INFERnal analysis were aligned to A) *Cutibacterium acnes* KPA171202, B) *Veillonella parvula* DSM 2008, C) *Pseudomonas putida* KT2440, D) *Corynebacterium* sp. ATCC 6931, and E) *Streptococcus sanguinis* SK36 genomes. Pink lines along the light grey genome track indicate the position of mapped INFERnal hits within the genome. Genes upstream and downstream of the riboswitches are colored by their general functional annotation. White (unrelated) indicates genes not currently known to be associated with cobamides. Grey (hypothetical) indicates a hypothetical protein that has no functional annotation. Genomic regions are not to scale.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. *Corynebacterium* cobamide producer abundance vs. Richness.

Within each metagenome, the total abundance of 5 cobamide-producing *Corynebacterium* (CPC) species was calculated. CPC abundance is plotted against richness (observed species) for samples within each skin site. R=Spearman's correlation coefficient.

Supplementary Figure 2. Relative abundance of metagenomes with low or high

***Corynebacterium* cobamide producer abundance.** The first (0.05%) and third (0.5%) quartiles of the cobamide-producing *Corynebacterium* (CPC) relative abundance across all samples were used to group samples below 0.05% or above 0.5% CPC abundance. Relative abundances for samples within each group are shown. Species less than 10% relative abundance are grouped into "Other".

Supplemental Figure 3. Number of unique species encoding for cobamide-dependent

enzymes, and cobalamin riboswitches. For A) cobamide-dependent enzymes and B) cobalamin riboswitches, the total number of unique species contributing at least 5 reads to each gene were calculated.

Figure 1

Corynebacterium Pangenome

bioRxiv preprint doi: <https://doi.org/10.1101/2020.12.02.407393>; this version posted December 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 2

bioRxiv preprint doi: <https://doi.org/10.1101/2020.12.02.407395>; this version posted December 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

A

B

Figure 8

Gene category

- Cobamide-independent ribonucleoside reductase
- Cobamide-independent methionine synthase
- Cobalt transport
- Cobamide biosynthesis
- Cobamide-dependent methylmalonyl-CoA mutase
- Cobamide-dependent ribonucleoside reductase
- ABC transporter
- Cobamide-dependent ethanolamine ammonia lyase
- Other function
- Hypothetical
- Pseudogene

Supplementary Figure 1

Supplementary Figure 2

Supplementary Figure 3

A

B