Revealing the diversity of commensal corynebacteria from a single human skin site

Reyme Herman[1], Sean Meaden[1], Michelle Rudden[1,2], Robert Cornmell[4], Holly N. Wilkinson[2], Matthew J. Hardman[2], Anthony J. Wilkinson[3], Barry Murphy[4], Gavin H. Thomas[1*]

[1] Department of Biology, University of York, Wentworth Way, York, YO10 5DD, UK

[2] School of Life Sciences, University of Hull, Hull, HU6 7RX, UK

[3] York Structural Biology Laboratory, Department of Chemistry, University of York, Wentworth Way, York, YO10 5DD, UK

[4] Unilever Research & Development, Port Sunlight Laboratory, Quarry Road East, Bebington, Wirral, Merseyside, CH63 3JW, UK

[*] For correspondence: gavin.thomas@york.ac.uk

Abstract

Our understanding of the skin microbiome has dramatically improved since the pioneering studies and the improvements in sequencing technologies. Species of the genus *Corynebacterium* are known to form a major part of the human skin microbiome but most detailed studies have focussed on other similarly prevalent genera like *Staphylococcus* and *Cutibacteria*. Prior to this study, there were few complete genomes for skin commensals of the genus *Corynebacterium,* with only 9 complete genomes available for the most commonly identified species *C. tuberculostearicum*. In this study we explored the genus *Corynebacterium* from a single body site by swabbing the axilla/underarm of 4 individuals and using a selective media to enrich for corynebacteria. We then generated whole genome sequencing data of these corynebacteria enriched isolates using long-read sequencing and subsequent bioinformatics analysis to reveal an unparalleled diversity of this genus from a single skin site. Through this approach, we obtained the closed genomes of 215 isolates, 154 derived from a single individual. With this genetic information, we were able to identify 7 different species including species previously not associated to the skin and two novel species provisionally named *C. axilliensis* and *C. jamesii*. We used pangenome analysis on 30 genetically distinct isolates spanning the 7 species to identify putative metabolic differences, antimicrobial resistance profiles, novel biosynthetic gene clusters (BGCs), prophages and phage defence systems. Our culture-based Nanopore sequencing approach has dramatically improved our overall knowledge of skin corynebacteria, and uniquely here also providing in-depth analysis from a single skin site, revealing a multitude of differences between the isolates. Here we not only improved our knowledge of axillary corynebacteria but also greatly expanded the publicly available number of cutaneous corynebacterial genomes complementing recent studies seeking to understand the diversity of skin corynebacteria.

Introduction

The exploration of the human skin microbiome allows deeper understanding of how commensal microbes and the host interact with each other to maintain skin health. Human skin, generally a nutrient limited environment, hosts a variety of commensal microbes

spanning multiple genera. Skin sites can be characterised as dry, moist or oily (sebaceous), depending on the architecture of the skin and abundance of the skin associated glands, namely the eccrine, sebaceous or apocrine glands. These different environmental conditions can play a major role in determining microbial diversity [1]. While *Cutibacterium* and *Staphylococcus* are the two major commensal genera identified, other skin associated genera like *Corynebacterium* are also highly abundant [2, 3]. In a study by Grice *et al*., out of over 112,000 16S rRNA sequences derived from 20 different skin sites, corynebacteria comprises of 22.8% of the sequences second only to cutibacteria (23%) [4]. Some corynebacterial species have been found to be lipophilic and hence require lipids derived from the sebaceous gland or the stratum corneum [5, 6]. Commonly identified commensal skin corynebacteria like *C. jeikeium* [7], *C. amycolatum* [8] and *C. kroppenstedtii* [9] have been suggested to be opportunistic pathogens. *C. accolens* [10] and *C. tuberculostearicum* [11] have been shown to interact with human immune system causing inflammation. *C. accolens* has been shown to also inhibit *Streptococcus pneumoniae* [12, 13] while *C. striatum* is proposed to be able to shift *Staphylococcus aureus* from the virulent to a commensal state [14], thus demonstrating potential protection capabilities. Together, these studies highlight the importance of understanding the healthy human skin microbiome in our efforts to combat pathogenic species.

Human malodour is microbially generated from a metabolite secreted from the apocrine gland. Although a monophyletic group of staphylococci was attributed to this phenomenon [15], some studies have proposed a role for axillary *Corynebacterium* in the human malodour formation due to the correlation of abundance and odour profiles [16] and the presence of enzymatic components for the biotransformation of the human derived precursor [17–19]. Further, a genus level study analysing the diversity of axillary commensals of 9 individuals found that 59.5% of sequences were binned to the genus *Corynebacterium* [20]. Our understanding of the role of cutaneous corynebacteria in human malodour remains unclear. Attaining genomic information of these axillary corynebacteria could reveal if they do possess the relevant genes for malodour production.

Despite the abundance of corynebacteria on the human skin, the amount of genetic information on cutaneous corynebacterial species is surprisingly lacking which implies a lack of understanding of its true diversity within the collective skin microbiome. In the case of *C. tuberculostearicum*, a species widely found on various skin sites, only 7 complete genomes have been made publicly available on GenBank prior to this study. In comparison, there are 221 available complete genomes of another skin commensal, *Staphylococcus epidermidis*. Two recent comparative genomics studies have sought to address the lack of genetic information through a culture-based approach. The first focusses on skin isolates of the *C. tuberculostearicum* species complex which is made up of multiple closely related species including *C. kefirresidentii* [21]. In addition, the study reports the identification of a novel species of corynebacteria that was predominantly found on the toe web and toenails of multiple individuals. A second comparative genomics study on human skin corynebacteria [22] revealed the diversity of corynebacteria on the human facial skin. This study too identified two novel species of corynebacteria.

Here, we sought to elucidate the diversity of corynebacteria within the human underarm (axilla). This is a moist skin site with metabolites derived from the three skin associated glands

and the skin itself, making it a relatively nutrient rich skin environment when compared with most of the other skin sites. In this study we took a culture-based approach followed by sequencing, consistent with recent comparative genomics studies. However, these studies adopted similar approaches by initially performing 16S rRNA (V1-V3 region) sequencing and identifying representatives within amplicon sequence variants (ASVs) to whole genome sequence. We instead adopted a direct-to long-read whole genome sequencing strategy by omitting the 16S rRNA gene sequencing which could introduce sampling bias. In this study, we generated complete genomes of over 200 corynebacteria isolates derived from axillary swabs of 4 individuals using the Oxford Nanopore sequencing technology. Our sequencing revealed 7 distinct species of axillary corynebacteria, two of which are proposed to be novel species of corynebacteria. We also further clarified the diversity within the *C. tuberculostearicum* species complex to complement recent studies. Downstream bioinformatics analysis revealed the genetic and functional diversity within our collection of axillary corynebacteria which would help reveal the roles of this genus of skin commensal bacteria.

Results

Assessing the diversity of axillary corynebacteria isolates

Axillary swabs from 4 human volunteers (A1, A3, A4 and A5) were collected and then enriched for corynebacteria on selective agar. To assess the ability of our approach to capture the diversity of corynebacteria within our isolates, we generated complete genome and plasmid sequences for 150 corynebacteria isolates from a single individual A1 and complemented the study with small subsets of isolates from the three other volunteers; 22 isolates from A3, 26 isolates from A4 and 13 isolates from A5 (**Fig. 1A**).

In order to initially assess the diversity of the corynebacterial isolates from all 4 volunteers, [23] we compared pairwise average nucleotide identities (ANI) between assembled genomes from each individual (**Fig. 1B-E**) and collectively (**Supplementary Fig. 1**). The pairwise percentage identity plots for the 150 Isolates from A1 revealed that the majority of isolates recovered by culture belonged to two discrete species (labelled as I and II) (**Fig. 1B**). However, the depth of sampling also revealed a third cluster (group III) and number of other smaller groups, some only containing one isolate, suggesting the approach had recovered significant diversity from this single body site from a single individual. For individuals A3 and A5 there was a similar pattern of a number of different clusters identified, despite there being a much smaller total number of genomes (**Fig. 1C** and **Fig. 1E**). Surprisingly in individual A4 we found that almost all of the 26 isolates were very similar to each other, suggesting that one strain/species dominates this community (**Fig. 1D**).

We identified representative genomes from groups of genetically similar isolates, with a 95% ANI used to distinguish between different species and an ANI of 99.5% to capture strain diversity (**Supplementary Fig. 2**). The isolates bin to 7 different species and the secondary clustering further identified 30 distinct isolates which we focussed further analysis on. To initially identify the species of these 30 isolates, their 16S rRNA (V1-V3 region) sequences were analysed using BLASTn and collectively aligned with the closest hits of each isolate (**Supplementary Fig. 3A**). Using this analysis, we observe 5 distinct clades with the largest clade containing 24 of our isolates which were found to be $\geq$ 99.8% identical (**Supplementary**

Fig. 3A, **red shading**). Further, the 16S rRNA (V1-V3 region) sequences of these 24 isolates were found to be ≥ 99.8% identical to three closely related but ultimately different species of corynebacteria which demonstrates the difficulty to speciate some corynebacterial isolates using this method. From here henceforth, the representative genomes will be referred to with the prefix YSMA (**Y**ork **S**kin **M**icrobiome: Study **A**)

While the 16S information alone could not accurate speciate our isolates, we were able to do this using the closed genomes using GTDB-tk [23, 24] (**Supplementary Fig. 3B**, **Supplementary Table 1**). Using this analysis, we observe further classification of the 24 isolates into three separate species, *C. kefirresidentii* (blue bar), *C. tuberculostearicum* (including *C. tuberculostearicum*_C, orange bar) and *C. aurimucosum*_E (black bar). These three species are commonly found on the skin and are thought to be part of the *C. tuberculostearicum* species complex which was previously reported [21, 25], confirming the importance of these organisms more generally at multiple human skin sites. ANI analysis of the GTDB representative for the species *C. aurimucosum*_E suggest that the species is misassigned and matches best with *C. marquesiae* [26]. From here henceforth, we classify our isolates that matched with *C. aurimucosum*_E on GTDB as *C. marquesiae*. Although two isolates (YSMAA1_1_B1 and YSMAA5_1_F11) matched closest with *C. tuberculostearicum* and *C. tuberculostearicum_C*, they were not classified to a specific species due to the predefined ANI reference radius of the analysis. As they share more than 95% identity to their closest matches, we classified both as *C. tuberculostearicum*.

YSMAA1_2_B5 and YSMAA1_2_C5 were classified as *C. amycolatum* which have been identified not only on the skin but also specifically in the human axilla [27, 28] (**Supplementary Table 1**). Interestingly, YSMAA1_1_F6 was identified to be *C. gottingense* which was first reported in 2017 [29] but until now, not being associated to the human skin. Finally, three isolates (YSMAA1_1_D6, YSMAA5_1_F9 and YSMAA1_1_F7) were matched with metagenome derived genomes of uncultured isolates suggesting that these are two new species. Isolates YSMAA1_1_D6 and YSMAA5_1_F9, tentatively named *C. axilliensis*, are most closely related to *C. singulare* with an average nucleotide identity of ~84% calculated using FastANI [30] (**Supplementary Fig. 4A**). *C. axilliensis* isolates were found on individuals A1 (~4% of corynebacteria genomes) and A5 (~23% of corynebacteria genomes). Isolate YSMAA1_1_F7, tentatively named *C. jamesii*, is most similar to *C. jeikeium* with an average nucleotide identity of ~87.5% calculated using FastANI [30] (**Supplementary Fig. 4B**). Overall, through our direct to whole genome sequencing approach, we not only revealed the species diversity of our axillary isolates from the four volunteers which would not be captured if a prior 16S rRNA screen was carried out but were also able to identify species not previously associated to healthy human skin including two novel species.

180 **Pangenome analysis of the 30 isolates**
181
182 We performed pangenome analysis of the 30 representative isolates to understand their
183 genetic differences. Firstly, we used Roary [31] to derive a core genome maximum likelihood
184 SNP tree using a BLASTp cutoff of 80% and seeded the tree with species representatives from
185 GenBank (**Supplementary Fig 5**). Consistent with the GTDB-tk analysis, we clearly see 7
186 different species including distinctions between the isolates from the species of the *C.*
187 *tuberculostearicum* species complex and obvious divergence between the two novel species
188 and their nearest related species. To further elucidate the genetic differences between our
189 isolates and their closest GenBank representatives, we performed a pangenome analysis
190 using anvi'o [32] (**Fig. 2**). Gene clusters of interest were then functionally predicted with
191 eggNOG-mapper[33, 34].
192
193 The two representative *C. axilliensis* isolates share a specific cluster of genes with the largest
194 proportion of genes with assigned functions found to be associated to transcription (K) or
195 replication, recombination and repair (L) (**Supplementary Fig. 6A**). Interestingly, analysis of
196 the gene clusters specific to either *C. axilliensis* YSMAA5_1_F9 or YSMAA1_1_D6 places the
197 largest proportion of genes with annotated functions in the cell wall/membrane/envelope
198 biogenesis (M) category suggesting differences in their cell envelope (**Supplementary Fig. 6B**
199 **and 6C**). Functionally annotated gene clusters specific to *C. jamesii* YSMAA1_1_F7 are mainly
200 found in the replication, recombination and repair (L) (**Supplementary Fig. 7**).
201
202 As expected, *C. kefirresidentii*, *C. marquesiae* and *C. tuberculostearicum* isolates share a larger
203 number of gene clusters than with other species. To clarify the main differences between the
204 three species, we generated the core genomes for each species from isolates within our
205 collection followed by a pangenome analysis (**Supplementary Fig. 8**) to extract the singletons
206 found within each species for functional analysis. Within the functionally predicted
207 singletons, genes classified to the inorganic ion transport and metabolism (P) category make
208 up the largest proportions in each of the three core genomes (**Supplementary Fig. 9A-C**)
209 consistent with a previous study on skin isolates of this species complex [21].
210
211 We further investigated the putative metabolic differences between these isolates by
212 identifying complete metabolic pathways present in each skin isolate using KEGG (**Fig. 3**). As
213 observed in the pangenome analysis, we generally observe KEGG modules clustering with
214 more similar species. Within the *C. tuberculostearicum* species complex isolates, we see
215 conservation of most KEGG modules but the largest differences seem to lie within the amino
216 acid metabolism pathways. Unlike other species within the complex, most *C. maquesiae*
217 isolates do not seem to possess the serine biosynthetic module (M00020) while *C.*
218 *tuberculostearicum* lack both lysine biosynthetic modules (M00016, M000526) and the
219 ornithine biosynthesis (M00028) modules and *C. kefirresidentii* lacks the tryptophan
220 biosynthetic module (M00023).
221
222 The tryptophan biosynthetic module (M00023) seems to also be absent in isolates all of *C.*
223 *gottingense*, *C. jamesii*, *C. amycolatum* and also *C. axilliensis* YSMAA1_1_D6. Interestingly, the
224 two isolates of *C. axilliensis* seem to possess different metabolic pathway profiles, not limited
225 to tryptophan biosynthesis. Another example is the fatty acid biosynthesis module (M00083),
226 found only in *C. axilliensis* YSMAA5_1_F9 and both *C. amycolatum* isolates.

227 Another category of metabolic pathways that we observe a large number of differences is the
228 metabolism of cofactors and vitamins. All but two isolates of *C. kefirresidentii* seem to be
229 biotin auxotrophs while *C. gottingense* and *C. amycolatum* isolates were suggested to be
230 pyridoxal phosphate auxotrophs. Finally, *C. amycolatum* YSMAA1_2_B5 seems to be the only
231 isolate that is able to synthesise pantothenate from valine or aspartate (M00119).
232
233 Overall, these differences including intra-species variation are captured in the 16
234 corynebacteria isolates from a single individual A1 (**Supplementary Fig 10**). Although we have
235 observed differences between isolates of the same species, we identified mainly species-
236 specific metabolic pathways which could be used to understand the potential modulate the
237 cutaneous corynebacterial community through metabolic means [37][38][39].
238
239 **Potential antimicrobial biosynthetic gene clusters and resistance cassettes**
240
241 Bacteria often produce complex small molecules that have various functions relating to
242 microbial competition such as antibiotics and siderophores, that are encoded within
243 recognisable biosynthetic gene clusters (BGCs). Resident bacteria of highly complex microbial
244 communities like the skin are known to produce antibiotics to suppress other species in their
245 niche, including preventing colonisation by pathogenic species [35] and conversely, allow for
246 the proliferation of others [36]. We used antiSMASH to identify putative BGCs within the
247 representative set of isolates [37]. For the purposes of this global corynebacteria BGC
248 analysis, we group similar predicted clusters together and determined the abundance of each
249 discrete cluster (**Fig 4**). The majority of the isolates share a predicted non-alpha poly-amino
250 acid (NAPAA), a type 1 polyketide synthase (T1PKS) and a terpene biosynthetic cluster. Other
251 predicted BGCs include non-ribosomal polyketide synthase (NRPS) and multiple terpene
252 clusters with different architectures (**Supplementary Fig. 11 and 12**). The *C.*
253 *tuberculostearicum* species complex isolates were generally found to possess similar BGCs
254 with *C. tuberculostearicum* isolates having the largest variation (**Supplementary Fig. 13**). As
255 expected, distantly related species like *C. amycolatum, C. gottingense* and *C. jamesii* were
256 predicted to have BGCs that are more distinct or of different classes (**Supplementary Fig. 14**).
257 Despite being outside the *C. tuberculostearicum* species complex, the analysed *C. axilliensis*
258 isolates seemed to also share similar predicted BGCs. With the exception of the cationic
259 peptide ε-Poly-L-lysine (NAPAA), we were not able to confidently predict the products of
260 these BGCs as the most similar known clusters for all of the predicted BGCs shared low
261 percentage similarities with their respective putative clusters. Nevertheless, our analysis
262 suggested different BGCs harboured by the corynebacterial isolates which could contribute
263 to the modulation of the skin microbiome through the production of metabolites like
264 siderophores (e.g. NI siderophore, aminopolycarboxylic-acid, NRPS) and even antimicrobials
265 (e.g. T1PKS, PKS-like + aminoglycoside).
266
267 The likely presence of antimicrobials on the human skin may promote the horizontal
268 acquisition of antimicrobial genes or develop beneficial mutants of existing cellular proteins.
269 ResFinder analysis of the isolates revealed 15 isolates with varying numbers and types of
270 acquired resistance genes but no predicted chromosomal mutations leading to resistance
271 (**Supplementary Table 2**). The most common resistance gene identified was *erm(X)* which is
272 proposed to confer resistance to erythromycin-related antibiotics and is widely found across
273 the genus [38–42]. *C. tuberculostearicum* YSMAA5_1_F11 and YSMAA5_1_G11 were also

found to have an aminoglycoside resistance gene which could confer resistance to gentamycin and tobramycin. *C. marquesiae* YSMAA1_1_H4, *C. tuberculostearicum* YSMAA3_1_A4 and *C. axilliensis* YSMAA5_1_F9 seem to have a resistance gene to an unknown aminoglycoside initially found in *C. striatum*. Interestingly, *C. axilliensis* YSMAA5_1_F9 has genes conferring resistance to chloramphenicol (cmx), streptomycin (aph(6)-Id, aph(3")-Ib) and kanamycin (aph(3")-Ia). These were found in tandem in the same region of the genome surrounded by multiple insertion sequence (IS) elements and were also not found in *C. axilliensis* YSMAA1_1_D6, suggesting a recent acquisition. Finally, we identified a sole tetracycline resistance gene in *C. tuberculostearicum* YSMAA5_1_B12 which also resides near IS elements, suggesting another example of horizontally acquired resistance mechanisms. We attempted to experimentally verify what we identified using ResFinder using the EUCAST disc diffusion assay (**Supplementary Table 3**). Despite the common identification of erm(X) which should confer resistance to clindamycin, only 10 out of the 14 isolates predicted to be resistant were as such. *C. axilliensis* YSMAA5_1_F9 was experimentally resistant to chloramphenicol, as predicted by ResFinder. However, *C. tuberculostearicum* YSMAA5_1_B12 was thought to be doxycycline resistant but did not display the phenotype in our disc diffusion assay. The presence of certain resistance genes may not confer resistance in lab-based disc diffusion assays as we may not have captured the most appropriate conditions in which these isolates would be resistant. Interestingly, *C. jamesii* YSMAA1_1_F7 seem to be experimentally resistant to Penicillin G even though a β-lactamase was not predicted to be in the genome suggesting this isolate has some other innate mechanism of resistance.

Potential phage clusters and resistance mechanisms

Another type of commonly acquired mobile genetic element are prophages. Phages are commonly found in the microbiome and could play an important role in skin health by regulating the bacterial community. We identified 17 putative prophage sequences in our 30 representative genomes, with an average of 0.6 (± 1.4) prophages per genome [50] (**Table 1**). Putative prophages were identified using geNomad [43] followed by gene annotation using pharokka and phold[44]. Only putative prophages containing genes which were functionally annotated with multiple common features of a phage (e.g. tail proteins, head and packaging) were retained for further analysis. These putative prophages were then searched against the NCBI database using BLASTn to identify the closest related phage and bacteria species. Some putative prophages were either matched with phages but retaining poor coverage (A1_1_F7_phi1, A5_1_F9_phi1) or not matched with any (A1_2_B5_phi1). As expected, various regions of these putative prophages also matched with other corynebacteria and similar to prophage hits, we observe some bacteria hits with poor coverage (A1_1_F7_phi1, A3_1_H1_phi1). As some of these isolates were derived from the same skin site of the same individual, we sought to understand the diversity of these putative prophages as some prophages may be shared between isolates. We performed a multisequence alignment using MAFFT [45] to determine the genetic identities between these putative prophages (**Supplementary Figure 15**). Neither of these putative prophages seem to be shared between isolates, even within the isolates from the same individuals. Remarkably, none of the 9 representative *C. tuberculostearicum* genomes were found to possess any putative prophages.

321 In addition to the presence of prophages, we analysed the representative genomes with
322 PADLOC [46] to identify potential phage defence systems (**Fig 5**). Type I, II, IIG and IV
323 restriction modification (RM) systems are commonly found in all isolate genomes while the
324 Type III RM system is much less abundant. Other commonly identified systems include the
325 plasmid transformation protection system Wadjet Type I [47], the abortive phage infection
326 system AbiD [48]. We also observed the CRISPR-Cas Type I-E system in some isolates of *C.*
327 *amycolatum*, *C. tuberculostearicum* and *C. jamesii* and also the CRISPR-Cas Type I-G system
328 only in *C. gottingense.* We also predicted the presence of multiple low abundance phage
329 defence systems like the putative nucleotide-sensing Gabija defence [49], the pyrimidine
330 cyclase Pycsar [50] and other defence systems of unknown mechanisms like shedu and kiwa
331 [47]. Global PADLOC analysis of all 215 complete genomes of our axillary corynebacteria
332 isolates suggested typical conservation of these predicted phage defence systems within the
333 drep secondary clusters with some exceptions (**Supplementary Fig. 16 - 19**). Some of these
334 differences include select isolates from *C. marquesiae* cluster 4_5 and 4_8 and *C.*
335 *tuberculostearicum* cluster 3_2 which lack the RM type IV system. A minority of *C. marquesiae*
336 cluster 4_5 isolates were also predicted to have the putative phage RNA interacting Mokosh
337 Type II system [51]. Similarly, only two isolates of *C. kefirresidentii* cluster 2_5 lack the RM
338 Type IIG system and two other isolates of the same cluster were predicted to have the AbiD
339 system. Together, we have comprehensively covered the range of phage defence systems
340 predicted to be present in our axillary isolates and we expect these mechanisms to be present
341 in other skin associated corynebacteria.
342
343 <u>Discussion</u>
344
345 The rise in number of recent studies exploring the various roles of commensals in skin health,
346 be it antagonistic or mutualistic, accentuates the need to reveal the true constituents of the
347 skin microbiome. Within the population of skin commensal microbes, staphylococci and
348 cutibacteria have been widely studied since the pioneering skin microbiome studies [2–4, 52].
349 Even though a large proportion of the microbiomes of multiple types of skin sites are made
350 up of corynebacteria, commensals of this genus have been poorly studied highlighting the
351 importance of studies like this to not only uncover the roles of these commensals but to also
352 discover novel skin associated species. Others have also sought to address this by attempting
353 to clarify the diversity of cutaneous corynebacteria [21, 22, 25] but the genetic information
354 for this genus still lagged behind that of cutaneous staphylococci and cutibacteria. The recent
355 advances in long-read sequencing, particularly the R10.4.1 flow cell and Kit 14 chemistry by
356 Oxford Nanopore Technologies, have allowed for high quality sequencing to be more
357 accessible at lower costs to researchers. With this we developed and optimised a direct-to
358 sequencing method enabling us to contribute another 30-representative cutaneous
359 corynebacteria genomes with another 195 closed genomes which were filtered out during
360 dereplication but are available for analysis.
361
362 Unlike our study, prior efforts to understand the skin microbiome usually involves an initial
363 16S rRNA amplicon sequencing step which acts as a filter to select for specific isolates to
364 proceed with whole genome sequencing. The high throughput nature of 16S rRNA gene
365 amplicon sequencing makes it an appealing first step in determining the diversity of isolates.
366 A common way to compare the 16S rRNA gene sequences is by aligning their respective V1-
367 V3 fragment as it has been shown to be able to distinguish between species more reliably [53,

54]. However, this approach does not resolve the differences between isolates of closely related but ultimately different species. For example, the skin commensals *C. kefirresidentii* strain FDAARGOS_1055 and *C. tuberculostearicum* strain FDAARGOS_1198 share a 16S rRNA (V1-V3 fragment) sequence identity of 99.8% but only share a ~89% genomic sequence identity when calculated using FastANI. Clearly, the difference between the two species is only obvious when whole genome sequences are available. The difficulties in resolving the diversity of isolates were also previously reported in multiple other studies as microbial "dark matter", or isolates that cannot easily speciated using the current tools and information available [2, 55, 56]. Further, common mobile elements like prophages which known contributors to genetic variation would be missed without a whole genome sequencing approach. We were able to keep our whole genome sequencing cost low (~2.5x the cost of 16S rRNA gene sequencing) negating the need for an initial 16S rRNA gene filtering approach as the invaluable genetic information from whole genome sequencing outweighs the additional costs. While our method provides deeper insight into the species and strain diversity within single skin sites on single individuals, we are still likely undersampling all the diversity that is there, while obtaining a much deeper understanding of the recovered microbes and the power of this technique in assessing the complete axilla community will be discussed elsewhere.

Using our method, we were able to strongly confirm the widespread presence of the species *C. tuberculostearicum*, *C. kefirresidentii* and *C. marquesiae* (formerly *C. aurimucosum*_E) (**Supplementary Fig. 20**), which together form the *C. tuberculostearicum* species complex [26]. In addition, we found more rarely *C. amycolatum* strains in the axilla and unexpectedly, we also identified isolates of *C. gottingense*, a species previously not associated to the skin. This species has been isolated from blood of patients with bacteraemia, the cerebrospinal fluid (CSF) and urine [29, 57], suggesting that it possibly originates from the skin. In addition to known species, we also identified two novel species tentatively named *C. jamesii* and *C. axilliensis*. Isolates of *C. axilliensis* were found on individuals A1 and A5 and makes up a significant 23% of sequenced corynebacterial isolates from the latter suggesting this novel species could be a major previously unrecognised components of the axillary microbiome of some individuals.

Although the fundamental roles of cutaneous commensal corynebacteria are still poorly understood, there have been increasing studies demonstrating how bacteria from this genus can contribute to colonisation protection including against other human pathogens. For example, *C. striatum* has been demonstrated to shift *S. aureus* towards commensalism through the suppression of the *agr* quorum sensing system [14]. In the presence of *C. striatum*, *S. aureus* exhibited increased human epithelial cell adhesion, decreased haemolysin activity and decreased fitness during infection indicating a decrease in virulence as compared to a monoculture. Rosenstein *et al*. [36] have proposed a cooperative role of corynebacteria against *S. aureus* in the nasal cavity with some species of commensal corynebacteria aiding in the proliferation of commensal *Staphylococcus lugdunensis* which in turn produces the antibiotic lugdunin that is active against *S. aureus*. These studies have highlighted the importance in understanding the true diversity of cutaneous corynebacteria species as we may yet uncover more roles that could be exploited to improve tools and treatments for skin health.

415 Studies advancing our understanding of the skin microbiome including reports focussing on
416 cutaneous corynebacteria seem to only be scratching the surface of the skin, the biggest
417 organ of the human body. Our knowledge of commensal corynebacteria lags behind other
418 equally prevalent genera like *Staphylococcus* and *Cutibacterium*. Here, we sought to further
419 understand the genetic diversity of the genus *Corynebacterium* using a low cost direct-to
420 whole genome sequencing approach on isolates derived from a single skin site, the axilla.
421 Using the invaluable genetic information, we not only identified species never before seen on
422 the skin including two novel species but we also identified a multitude of genetic differences
423 within all 215 closed genomes including putative metabolic differences, the presence of
424 biosynthetic gene clusters and prophages providing multiple new avenues of exploration.
425 These publicly available genomes can be further mined by other researchers to build on our
426 analysis to further elucidate the roles of this poorly understudied genus on the skin. Further,
427 these and other matched isolates of the York Skin Microbiome (YSM) library are made
428 available for multiple other studies to facilitate our efforts to understand the complex
429 environment that is the skin microbiome.
430
431 Materials and Methods
432
433 **Sampling procedures**
434 Four adult male volunteers (Age range: 32 to 78) were recruited for axillary swabbing at
435 Unilever R&D, Bebington, UK. The volunteers were in good general health, did not use any
436 body products (e.g. deodorants) for 24 hours before the study, were not on antibiotics, with
437 no active skin conditions on any part of their bodies, not suffered from eczema in the last 5
438 years nor have they ever had psoriasis. Prior to swabbing, the volunteers were "sniff tested"
439 by a panel of testers and scored according to odour levels (High malodour: A1 and A3, Low
440 malodour: A4 and A5). Axillary swabs were then collected by swabbing in a linear motion 20
441 times using eSwabs (Copan) on both underarms and then placed into Amies transport
442 medium.
443
444 **Bacterial culturing**
445 Swabs were plated onto ACP solid media (39.5 g L^{-1} Blood Agar base no. 2, 3 g L^{-1} yeast extract,
446 2 g L^{-1} glucose, 5 mL L^{-1} Tween 80, 50 mL L^{-1} defibrinated horse blood, 100 mg L^{-1} fosfomycin)
447 to enrich for corynebacteria and grown for 2 days at 37°C. Colonies were picked and grown
448 overnight in brain-heart infusion broth + 1 % Tween-80 (BHIT) at 37°C, shaking at 200 rpm.
449 Cultures were stocked in BHIT supplemented with 10 % glycerol in 2 mL 96 well plates.
450
451 **DNA extraction**
452 For genomic and plasmid DNA extraction, 0.5 mL of BHIT was inoculated with each isolate in
453 2 mL 96 well plates and grown for 20 hours at 37°C, shaking at 200 rpm. Cultures were then
454 pelleted by centrifugation before resuspension in 100 μL enzymatic lysis buffer (50 mM Tris-
455 HCl, pH 8, 50 mM EDTA, pH 8, 0.5% Tween 20, 0.5% Triton-X100, 25 mg/mL lysozyme). The
456 mixtures were incubated at 37°C overnight with gentle shaking (<50 rpm). 500 μL of lysis
457 buffer (0.5 M NaCl, 100 mM Tris-HCl, pH 8, 50mM EDTA, pH 8, 1.5% SDS) was then mixed into
458 each well and incubated at 80°C for 30 minutes in a thermomixer. The mixture was allowed
459 to cool before the addition of 4μl of 4mg/mL RNase A Solution (Promega) before incubation
460 at 37°C for a further hour. Cellular proteins were then degraded with 20 μl of 20 mg/mL
461 Proteinase K Solution (Promega) and incubation at 56°C for 15 minutes. Remaining proteins

were precipitated with 200 µl of protein precipitation buffer (5 M Potassium acetate). The lysates were then clarified by centrifugation at 16,000 xg for 5 minutes. DNA was precipitated from the supernatant using 600 µl isopropanol and incubated overnight at 4°C. The precipitated DNA were pelleted by centrifugation at 16,000xg for 5 minutes. The DNA pellets were washed with 600 µl 70% ethanol and then air dried for 15 minutes. Finally, the DNA pellets were rehydrated with 50 µl nuclease-free water and incubation overnight at room-temperature. The DNA concentrations and QC measurements were performed using Qubit (Thermo Fisher) and Nanodrop (Thermo Fisher).

DNA library preparation, whole genome sequencing and assembly

The Oxford Nanopore Native Barcoding Kit 96 V14 was used to prepare and barcode the prepared genomic and plasmid DNA. The recommended protocol was followed with the following changes: 1. Half the recommended volume of the DNA repair and end-prep reaction was prepared. 2. Half the recommended volume of the native barcode ligation reaction was prepared. The prepared library was loaded into a PromethION R10.4.1 flow cell and sequencing was allowed to take place over 72 hours and base-called using the singleplex high-accuracy model, 400 bps on MinKNOW 23.07.5. The reads were assembled with the EPI2ME Labs wf-bacterial-genomes isolate workflow (Flye [58], Medaka (Oxford Nanopore Technologies), Prokka [59], ResFinder [60, 61]).

Bioinformatics tools

Pairwise genome comparisons of all 215 assembled genomes were performed using the ANIm [62] analysis on pyANI v0.2.12 [63]. The genomes were dereplicated with drep v3.4.2 [64] using Mash [65] and MUMmer 3.0 [66], employing a primary clustering ANI cutoff of 95% to distinguish between different species and a secondary clustering ANI cutoff of 99.5% to capture strain diversity to derive a representative set of genomes. The core genomes were determined using Roary[31] with an 80% BLASTp cutoff. PhyML 3.3 [67] trees 16S rRNA V1-V3 and core genome alignments of the representative genomes were constructed using the HKY85 substitution model with approximate likelihood ratio tests (aLRT). Trees were visualised on iTOL v6.9 [68]. Speciation using GTDB-tk v2.3.2 [23] was performed on KBase [24]. The genome alignments of *C. axilliensis* and *C. jamesii* were prepared using BLASTn and visualised using BRIG [69].

Pangenome analysis was performed using the anvi-pan-genome package and visualised on anvi'o v8 [32]. The representative genomes were also compared using pyANI and visualised with the pangenome on anvi'o. Relevant gene clusters were extracted and functionally analysed with eggNOG-mapper v2 [70, 71] to determine the respective COG categories. Metabolic pathways were identified using blastKOALA and KEGG Reconstruct [72, 73]. Biosynthetic gene cluster analysis was performed using the default settings on antiSMASH v7 [37]. Putative prophages were predicted using geNomad v1.8 [43] and then annotated using pharokka 1.7.2 and phold [44]. Phage defence systems were predicted using PADLOC v4.3 [74].

Antibiotic susceptibility testing

508 Antibiotic susceptibility testing was performed using plate-based disc-diffusion assays using
509 the recommended manufacturers protocol and EUCAST instructions. Overnight cultures were
510 prepared BHIT and grown at 37°C. Cultures were then diluted to a 0.5 McFarland standard
511 (Remel, Thermo Scientific) before plating onto BHIT-agar. Six antibiotic susceptibility discs
512 were placed on the agar (Doxycycline-30μg, Chloramphenicol-30μg, Fosfomycin-200μg,
513 Clindamycin-10μg, Vancomycin-30μg, Penicillin G-1μg) (Oxoid). Plates were allowed to grow
514 for 24 hours at 37°C before measurements of any zones of inhibition.
515
516 Ethics statement
517
518 This study was conducted in accordance with the ethical principles of Good Clinical Practice
519 and the Declaration of Helsinki. The local ethics committees of Unilever R&D (Port Sunlight,
520 UK) approved the protocols before commencement of the studies and all subjects gave
521 written informed consent.
522
523 Data availability
524
525 The 24 dereplicated genomes and associated plasmids were deposited to GenBank within
526 BioProject PRJNA1191350.
527
528 Bibliography
529
530 1. Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. Nature Reviews Microbiology
531 2018 16:3. 2018;16:143–55.
532 2. Oh J, Byrd AL, Park M, Kong HH, Segre JA. Temporal Stability of the Human Skin
533 Microbiome. Cell. 2016;165:854–66.
534 3. Oh J, Byrd AL, Deming C, Conlan S, Kong HH, Segre JA, et al. Biogeography and
535 individuality shape function in the human skin metagenome. Nature 2014 514:7520.
536 2014;514:59–64.
537 4. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, et al. Topographical and
538 temporal diversity of the human skin microbiome. Science (1979). 2009;324:1190–2.
539 5. POLLOCK MR, WAINWRIGHT SD, MANSON EE. The presence of oleic acid-requiring
540 diphtheroids on human skin. J Pathol Bacteriol. 1949;61:274–6.
541 6. McGinley KJ, Labows JN, Zechman JM, Nordstrom KM, Webster GF, Leyden JJ. Analysis of
542 Cellular Components, Biochemical Reactions, and Habitat of Human Cutaneous Lipophilic
543 Diphtheroids. Journal of Investigative Dermatology. 1985;85:374–7.
544 7. Tauch A, Kaiser O, Hain T, Goesmann A, Weisshaar B, Albersmeier A, et al. Complete
545 genome sequence and analysis of the multiresistant nosocomial pathogen Corynebacterium
546 jeikeium K411, a lipid-requiring bacterium of the human skin flora. J Bacteriol.
547 2005;187:4671–82.
548 8. Esteban J, Nieto E, Calvo R, Fernández-Roblas R, Valero-Guillén PL, Soriano F.
549 Microbiological characterization and clinical significance of Corynebacterium amycolatum
550 strains. European Journal of Clinical Microbiology and Infectious Diseases. 1999;18:518–21.
551 9. Paviour S, Musaad S, Roberts S, Taylor G, Taylor S, Shore K, et al. Corynebacterium species
552 isolated from patients with mastitis. Clinical Infectious Diseases. 2002;35:1434–40.

10. Ridaura VK, Bouladoux N, Claesen J, Erin Chen Y, Byrd AL, Constantinides MG, et al. Contextual control of skin immunity and inflammation by Corynebacterium. Journal of Experimental Medicine. 2018;215:785–99.
11. 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-κB inflammatory signaling pathway in human skin cells. Immun Inflamm Dis. 2020;8:62–79.
12. 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.
13. Horn KJ, Jaberi Vivar AC, Arenas V, Andani S, Janoff EN, Clark SE. Corynebacterium Species Inhibit Streptococcus pneumoniae Colonization and Infection of the Mouse Airway. Front Microbiol. 2022;12:804935.
14. Ramsey MM, Freire MO, Gabrilska RA, Rumbaugh KP, Lemon KP. Staphylococcus aureus Shifts toward commensalism in response to corynebacterium species. Front Microbiol. 2016;7 AUG:212804.
15. Rudden M, Herman R, Rose M, Bawdon D, Cox DS, Dodson E, et al. The molecular basis of thioalcohol production in human body odour. Sci Rep. 2020;10:12500.
16. Troccaz M, Gaïa N, Beccucci S, Schrenzel J, Cayeux I, Starkenmann C, et al. Mapping axillary microbiota responsible for body odours using a culture-independent approach. Microbiome. 2015;3:1–15.
17. Natsch A, Emter R. The specific biochemistry of human axilla odour formation viewed in an evolutionary context. Philosophical Transactions of the Royal Society B. 2020;375.
18. Emter R, Natsch A. The sequential action of a dipeptidase and a beta-lyase is required for the release of the human body odorant 3-methyl-3-sulfanylhexan-1-ol from a secreted Cys-Gly-(S) conjugate by Corynebacteria. J Biol Chem. 2008;283:20645–52.
19. Natsch A, Schmid J, Flachsmann F. Identification of Odoriferous Sulfanylalkanols in Human Axilla Secretions and Their Formation through Cleavage of Cysteine Precursors by a C-S Lyase Isolated from Axilla bacteria. Chem Biodivers. 2004;1:1058–72.
20. Callewaert C, Kerckhof FM, Granitsiotis MS, van Gele M, van de Wiele T, Boon N. Characterization of Staphylococcus and Corynebacterium Clusters in the Human Axillary Region. PLoS One. 2013;8:e70538.
21. Ahmed N, Joglekar P, Deming C, Lemon KP, Kong HH, Segre JA, et al. Genomic characterization of the C. tuberculostearicum species complex, a prominent member of the human skin microbiome . mSystems. 2023;8.
22. Jensen MG, Svraka L, Baez E, Lund M, Poehlein A, Brüggemann H. Species- and strain-level diversity of Corynebacteria isolated from human facial skin. BMC Microbiol. 2023;23:1–14.
23. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics. 2022;38:5315–6.
24. Chivian D, Jungbluth SP, Dehal PS, Wood-Charlson EM, Canon RS, Allen BH, et al. Metagenome-assembled genome extraction and analysis from microbiomes using KBase. Nature Protocols 2022 18:1. 2022;18:208–38.
25. Salamzade R, Swaney MH, Kalan LR. Comparative Genomic and Metagenomic Investigations of the Corynebacterium tuberculostearicum Species Complex Reveals Potential Mechanisms Underlying Associations To Skin Health and Disease. Microbiol Spectr. 2023;11.

600 26. Cappelli EA, Ksiezarek M, Wolf J, Neumann-Schaal M, Ribeiro TG, Peixe L. Expanding the
601 Bacterial Diversity of the Female Urinary Microbiome: Description of Eight New
602 Corynebacterium Species. Microorganisms. 2023;11.
603 27. Taylor D, Daulby A, Grimshaw S, James G, Mercer J, Vaziri S. Characterization of the
604 microflora of the human axilla. Int J Cosmet Sci. 2003;25:137–45.
605 28. Bawdon D, Cox DS, Ashford D, James AG, Thomas GH. Identification of axillary
606 Staphylococcus sp. involved in the production of the malodorous thioalcohol 3-methyl-3-
607 sufanylhexan-1-ol. FEMS Microbiol Lett. 2015;362:1–10.
608 29. Atasayar E, Zimmermann O, Spröer C, Schumann P, Groß U. Corynebacterium
609 gottingense sp. Nov., isolated from a clinical patient. Int J Syst Evol Microbiol.
610 2017;67:4494–9.
611 30. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI
612 analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature
613 Communications 2018 9:1. 2018;9:1–8.
614 31. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid
615 large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3.
616 32. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, et al. Community-led,
617 integrated, reproducible multi-omics with anvi'o. Nature Microbiology 2020 6:1. 2020;6:3–
618 6.
619 33. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al.
620 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource
621 based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309–14.
622 34. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper
623 v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the
624 Metagenomic Scale. Mol Biol Evol. 2021;38:5825–9.
625 35. Zipperer A, Konnerth MC, Laux C, Berscheid A, Janek D, Weidenmaier C, et al. Human
626 commensals producing a novel antibiotic impair pathogen colonization. Nature 2016
627 535:7613. 2016;535:511–6.
628 36. Rosenstein R, Salazar BOT, Sauer C, Heilbronner S, Krismer B, Peschel A. Siderophore
629 piracy enables the nasal commensal Staphylococcus lugdunensis to antagonize the
630 pathogen Staphylococcus aureus. bioRxiv. 2024;:2024.02.29.582731.
631 37. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 7.0: new
632 and improved predictions for detection, regulation, chemical structures and visualisation.
633 Nucleic Acids Res. 2023;51:W46–50.
634 38. Olender A, Niemcewicz M. Macrolide, lincosamide, and streptogramin B-constitutive-
635 type resistance in Corynebacterium pseudodiphtheriticum isolated from upper respiratory
636 tract specimens. Microb Drug Resist. 2010;16:119–22.
637 39. Ortiz-Pérez A, Martín-De-Hijas NZ, Esteban J, Fernández-Natal MI, García-Cía JI,
638 Fernández-Roblas R. High frequency of macrolide resistance mechanisms in clinical isolates
639 of Corynebacterium species. Microb Drug Resist. 2010;16:273–7.
640 40. TAUCH A, KASSING F, KALINOWSKI J, PÜHLER A. The Corynebacterium xerosis composite
641 transposon Tn5432 consists of two identical insertion sequences, designated IS1249,
642 flanking the erythromycin resistance gene ermCX. Plasmid. 1995;34:119–31.
643 41. Hodgson ALM, Krywult J, Radford A. Nucleotide sequence of the erythromycin resistance
644 gene from the Corynebacterium plasmid pNG2. Nucleic Acids Res. 1990;18:1891.
645 42. Rosato AE, Lee BS, Nash KA. Inducible macrolide resistance in Corynebacterium
646 jeikeium. Antimicrob Agents Chemother. 2001;45:1982–9.

647 43. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of mobile
648 genetic elements with geNomad. Nature Biotechnology 2023. 2023;:1–10.
649 44. Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald PJ, Vreugde S. Pharokka: a fast
650 scalable bacteriophage annotation tool. Bioinformatics. 2023;39.
651 45. Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple
652 sequence alignment based on fast Fourier transform.
653 46. Payne LJ, Todeschini TC, Wu Y, Perry BJ, Ronson CW, Fineran PC, et al. Identification and
654 classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new
655 system types. Nucleic Acids Res. 2021;49:10868–78.
656 47. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery
657 of antiphage defense systems in the microbial pangenome. Science (1979). 2018;359.
658 48. McLandsborough LA, Kolaetis KM, Requena T, McKay LL. Cloning and characterization of
659 the abortive infection genetic determinant abiD isolated from pBF61 of Lactococcus lactis
660 subsp. lactis KR5. Appl Environ Microbiol. 1995;61:2023–6.
661 49. Cheng R, Huang F, Wu H, Lu X, Yan Y, Yu B, et al. A nucleotide-sensing endonuclease
662 from the Gabija bacterial defense system. Nucleic Acids Res. 2021;49:5216–29.
663 50. Tal N, Morehouse BR, Millman A, Stokar-Avihail A, Avraham C, Fedorenko T, et al. Cyclic
664 CMP and cyclic UMP mediate bacterial immunity against phages. Cell. 2021;184:5728-
665 5739.e16.
666 51. Millman A, Melamed S, Leavitt A, Doron S, Bernheim A, Hör J, et al. An expanded arsenal
667 of immune systems that protect bacteria from phages. Cell Host Microbe. 2022;30:1556-
668 1569.e5.
669 52. Grice EA, Segre JA. The skin microbiome. Nature Reviews Microbiology 2011 9:4.
670 2011;9:244–53.
671 53. Oh J, Freeman AF, Comparative N, Program S, Park M, Sokolic R, et al. The altered
672 landscape of the human skin microbiome in patients with primary immunodeficiencies.
673 https://doi.org/10.1101/gr.159467.113.
674 54. Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, et al. Topographic diversity of
675 fungal and bacterial communities in human skin. Nature 2013 498:7454. 2013;498:367–70.
676 55. Kashaf SS, Proctor DM, Deming C, Saary P, Hölzer M, Comparative N, et al. Integrating
677 cultivation and metagenomics for a multi-kingdom view of skin microbiome diversity and
678 functions. NATuRE MICRobIology |. 2022;7:169–79.
679 56. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, et al. Insights into
680 the phylogeny and coding potential of microbial dark matter. Nature. 2013;499:431–7.
681 57. Bernard KA, Pacheco AL, Burdz T, Wiebe D, Bernier AM. Corynebacterium godavarianum
682 jani et al. 2018 and corynebacterium hadale wei et al. 2018 are both later heterotypic
683 synonyms of corynebacterium gottingense atasayar et al. 2017, proposal of an emended
684 description of corynebacterium gottingense atasayar et al. 2017. Int J Syst Evol Microbiol.
685 2020;70:3534–40.
686 58. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using
687 repeat graphs. Nature Biotechnology 2019 37:5. 2019;37:540–6.
688 59. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics.
689 2014;30:2068–9.
690 60. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+:
691 Architecture and applications. BMC Bioinformatics. 2009;10:1–9.
692 61. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for
693 predictions of phenotypes from genotypes. J Antimicrob Chemother. 2020;75:3491–500.

62. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 2009;106:19126–31.
63. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical Methods. 2015;8:12–24.
64. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11:2864–8.
65. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: Fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17:1–14.
66. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:1–9.
67. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol. 2010;59:307–21.
68. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2013;1:13–4.
69. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. BMC Genomics. 2011;12:1–10.
70. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38:5825–9.
71. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309–14.
72. Kanehisa M, Sato Y, Kawashima M. KEGG mapping tools for uncovering hidden features in biological data. Protein Sci. 2022;31:47–53.
73. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci. 2020;29:28–35.
74. Payne LJ, Meaden S, Mestre MR, Palmer C, Toro N, Fineran PC, et al. PADLOC: a web server for the identification of antiviral defence systems in microbial genomes. Nucleic Acids Res. 2022;50:W541–50.

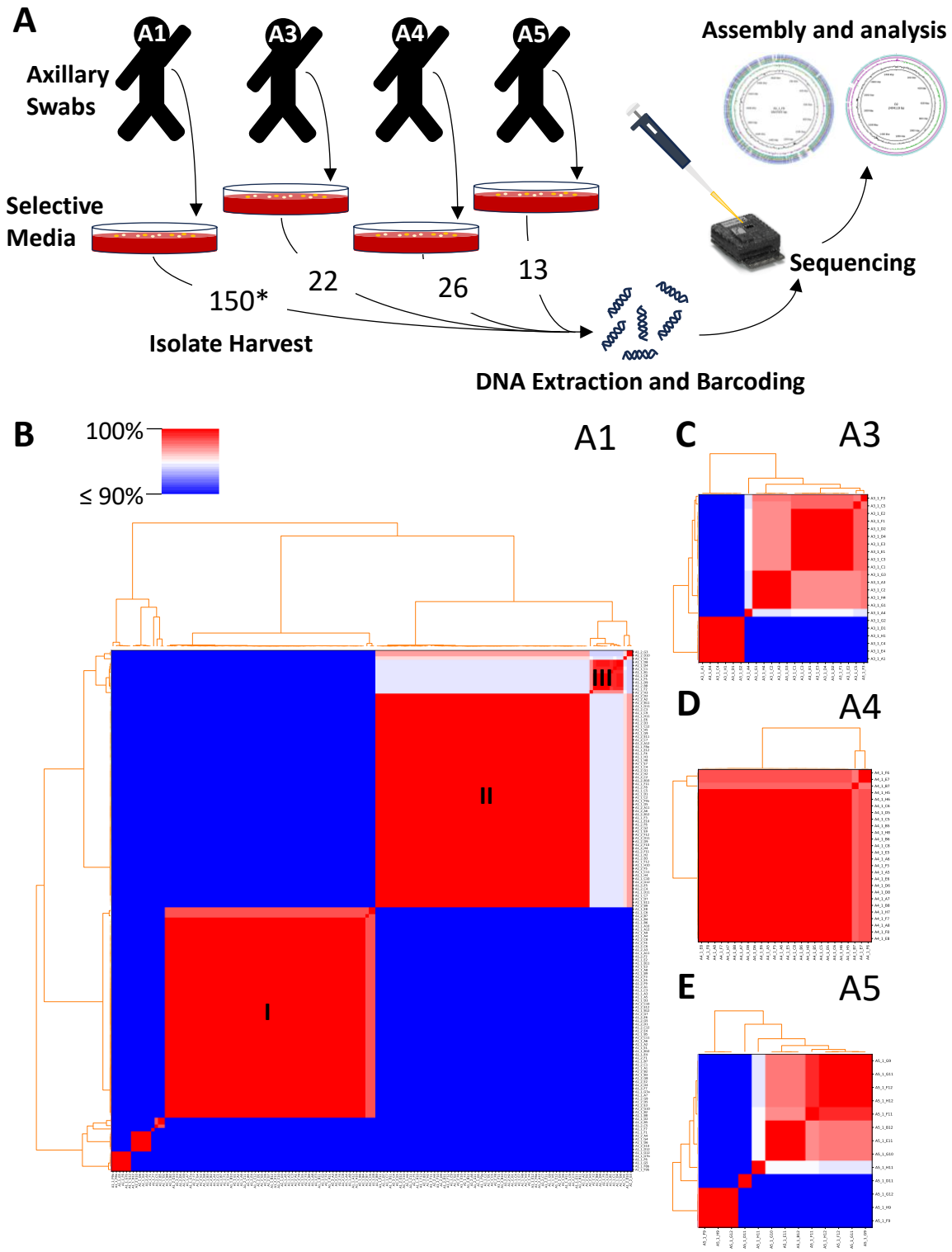

Figure 1. Capturing the diversity of axillary corynebacteria. (A) Axillary *Corynebacterium* isolates were harvested from swabs from 4 volunteers. Long-read whole genome sequencing and downstream analysis was performed on a 150 isolates of volunteer A1, 22 from volunteer A3, 26 from volunteer A4 and 13 from volunteer A5. (B-E) Pairwise average nucleotide identities (ANI) were calculated with pyANI and used to assess the genetic diversity of the isolates. ANI values were placed onto a coloured pairwise matrix according to the colour scale depicted in (B) for isolates derived from each volunteer. Three major groups of isolates were identified with I, II and III from the isolates from A1.

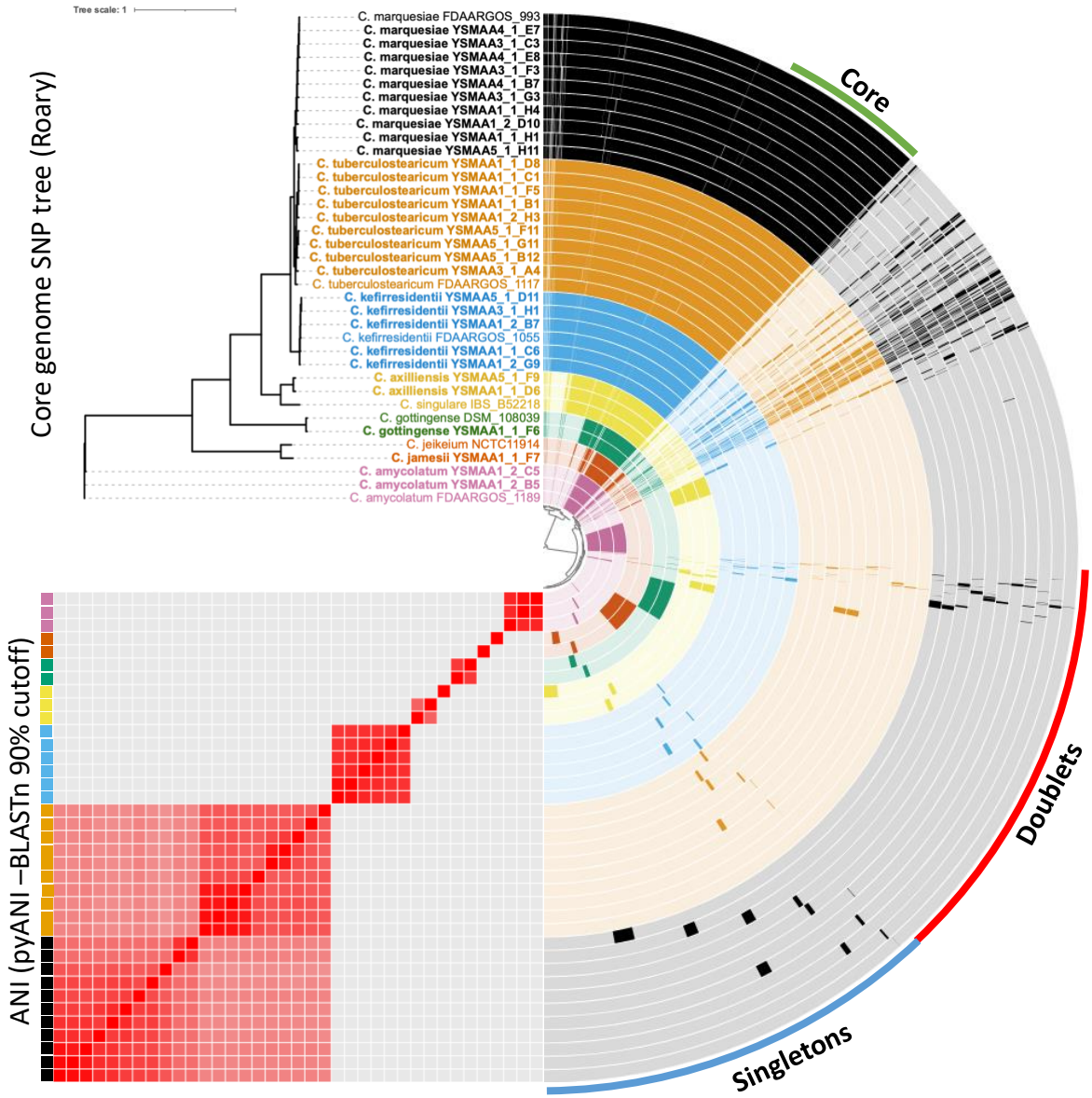

Figure 2. The pangenome of the axillary corynebacteria isolates (bold) with the closest RefSeq representatives for each species calculated and visualised using anvi'o. Gene clusters found in either one (Singletons) or two (Doublets) were identified. A core genome was calculated using Roary with an 80% BLASTp cutoff and visualised on a tree generated using PhyML and visualised on iTOL. An ANI comparison using BLASTn on pyANI with a 90% cutoff was generated and visualised on anvi'o. Isolates of the 7 clusters of species were colour coded.

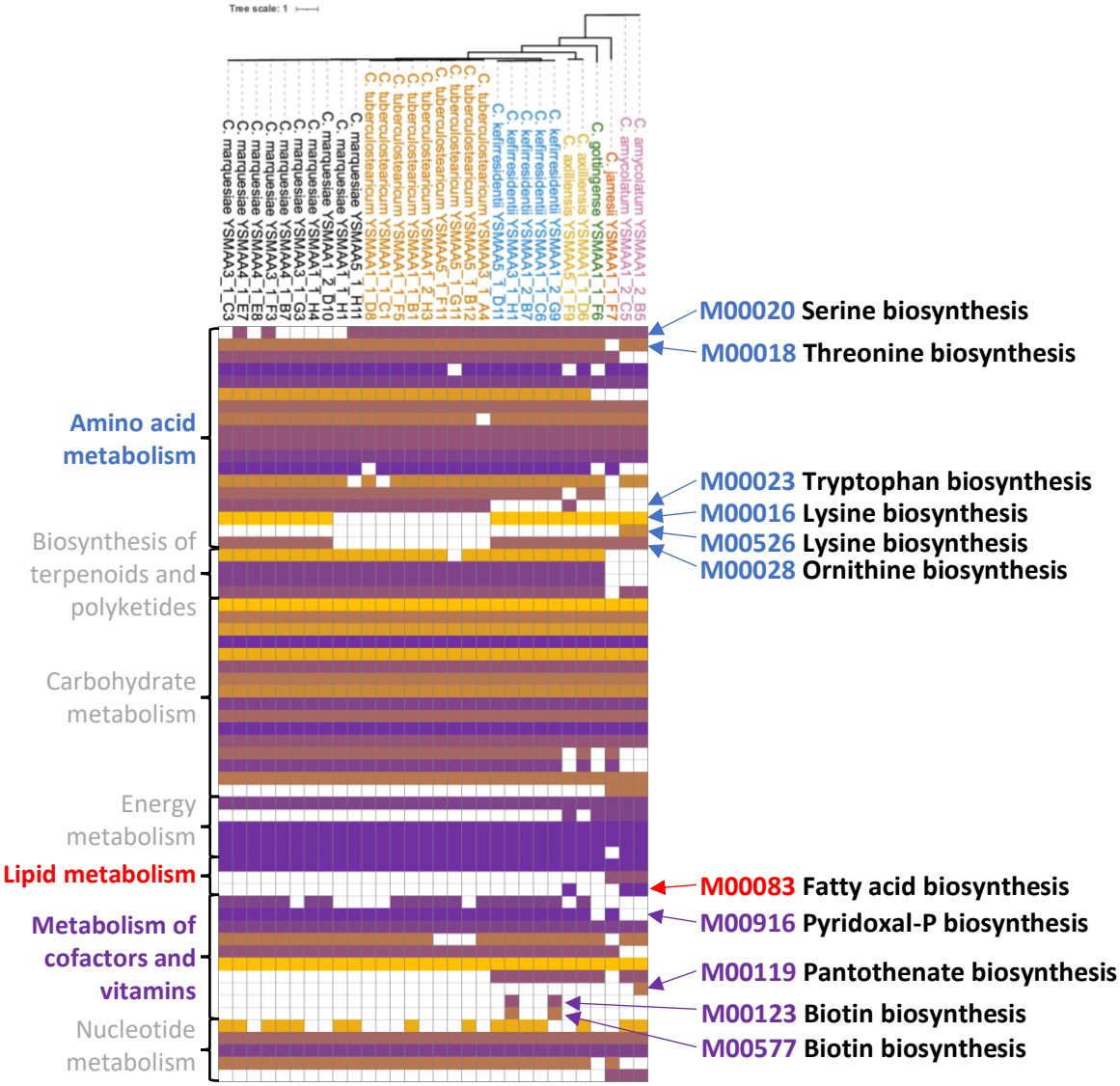

Figure 3. Metabolic pathways in axillary corynebacteria reveals species specific general profiles and some variation within species. Each KEGG module for each isolate were placed on a heatmap where the modules with the largest number of reactions are in yellow and lowest in purple with absent modules in white. Modules of interest were described on the right of the heatmap. KEGG modules were grouped according to the categories on the left of the heatmap. Isolates were grouped by species with a core genome tree depicting the variation.

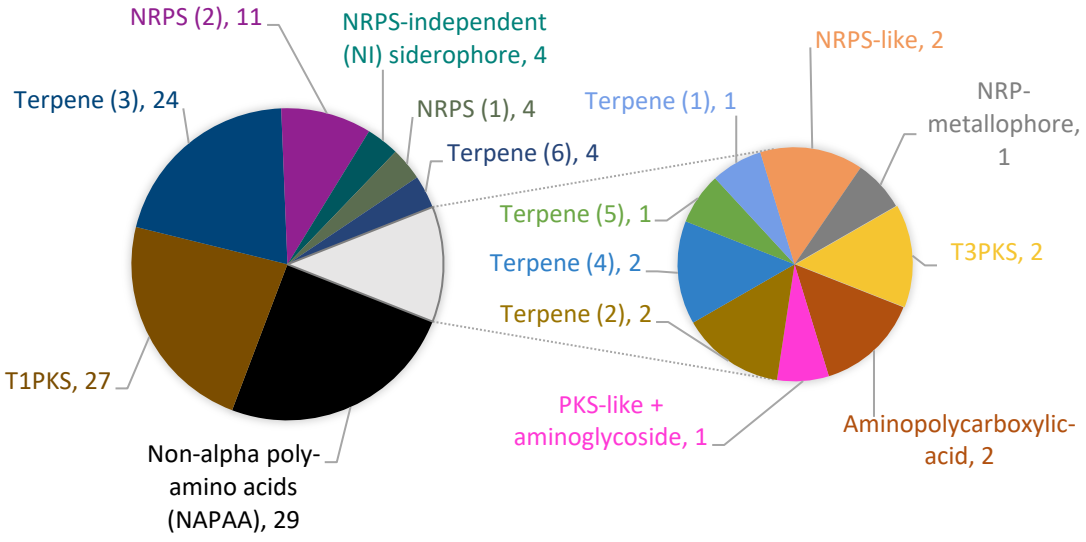

Figure 4. Multiple putative biosynthetic gene cluster (BGC) types were identified in the representative isolate genomes using antiSMASH 7. Names of the different BGC types were included with different cluster architectures in numbered parentheses. The abundance of each BGC type were also included.

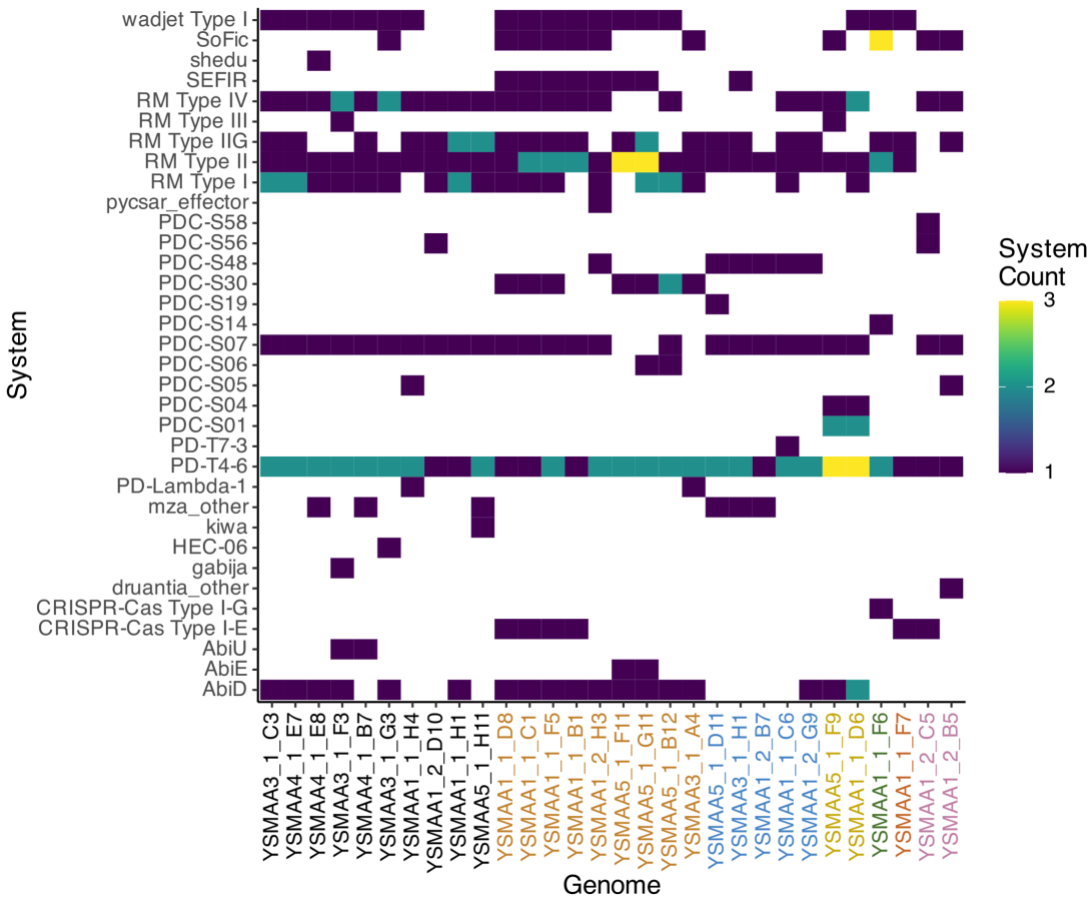

Figure 5. Various phage defence systems were predicted in the representative genomes using PADLOC. The heatmap depicts the abundance of each phage defence system within each isolate according to the scale on the right with absent systems in white. The genomes were arranged according to the different corynebacteria species and coloured accordingly.

Putative prophage name	Putative prophage length	Source species	Closest phage hit (% identity, % query cover)	Closest bacteria hit (% identity, % query cover)
A1_1_C6_phi1_2	52250	*C. kefirresidentii*	ctNWy2 (92.03%, 69%)	*C. tuberculostearicum* CT (91.51%, 76%)
A1_1_D6_phi1	42557	*C. axilliensis*	ctIBL2 (83.98%, 29%)	*C. aurimucosum* DSM 44532 (87.61%, 26%)
A1_1_F7_phi1	56384	*C. jamesii*	ctwl93 (77.01%, <1%)	*C. jeikeium* DSM 7171 (91.59%, 7%)
A1_1_H1_phi1	49192	*C. marquesiae*	ctfR81 (90.01%, 67%)	*C. aurimucosum* FDAARGOS_1110 (84.54%, 49%)
A1_2_B5_phi1	38625	*C. amycolatum*	-	*C. jeikeium* FDAARGOS_574 (92.93%, 37%)
A1_2_D10_phi1	42401	*C. marquesiae*	ctKpD5 (81.46%, 43%)	*C. accolens* CA1 (80.24%, 49%)
A1_2_G9_phi1	26374	*C. kefirresidentii*	ctOSk1 (95.99%, 64%)	*C. tuberculostearicum* CT (91.72%, 65%)
A3_1_C3_phi1	38823	*C. marquesiae*	ctSSL1 (93.44%, 45%)	*C. imitans* NCTC13015 (81.58%, 28%)
A3_1_F3_phi1	54302	*C. marquesiae*	ctcbl1 (92.73%, 70%)	*C. accolens* CA9 (75.43%, 38%)
A3_1_G3_phi1	47617	*C. marquesiae*	ct1BD1 (92.15%, 54%)	*C. tuberculostearicum* CTNIH13 (89.64%, 42%)
A3_1_H1_phi1	47347	*C. kefirresidentii*	ctLCM1 (95.25%, 44%)	*C. tuberculostearicum* CTNIH23 (97.62%, 9%)
A4_1_E7_phi1	47738	*C. marquesiae*	ct1BD1 (86.18%, 54%)	*C. tuberculostearicum* CTNIH13 (89.94%, 40%)
A5_1_D11_phi5_2_1	49670	*C. kefirresidentii*	ctNWy2 (95.17%, 85%)	*C. tuberculostearicum* CT (99.84%, 100%)
A5_1_D11_phi3_4	50578	*C. kefirresidentii*	ctNWy2 (96.26%, 64%)	*C. tuberculostearicum* CT (96.35%, 81%)
A5_1_F9_phi1	43162	*C. axilliensis*	ctIBL2 (84.39%, 19%)	*C. aurimucosum* DSM 44532 (88.11%, 56%)
A5_1_H11_phi1	50212	*C. marquesiae*	ctXFF1 (97.39%, 85%)	*C. accolens* CA9 (84.87%, 51%)
A5_1_H11_phi2	49906	*C. marquesiae*	ct1BD1 (92.69%, 46%)	*C. tuberculostearicum* CTNIH13 (91.20%, 75%)

Table 1. The closest phages and bacteria genomes were identified using a BLASTn analysis on the putative prophages identified by geNomad. The percentage sequence identity and percentage coverage of each hit were included.