CUTIS-SEQ, a flexible bilocus sequence typing scheme that provides high resolution of Cutibacterium acnes strains across all subspecies

Anaerobe 79 (2023) 102671

Contents lists available at ScienceDirect

Anaerobe

journal homepage: www.elsevier.com/locate/anaerobe

CUTIS-SEQ, a flexible bilocus sequence typing scheme that provides high resolution of *Cutibacterium acnes* strains across all subspecies

Joseph McLaughlin [a], István Nagy [b, c], Georgios Miliotis [d], Andrew McDowell [a, e, *]

[a] Personalised Medicine Centre, School of Medicine, Ulster University, Londonderry, UK
[b] Institute of Biochemistry, Biological Research Centre, Szeged, Hungary
[c] Seqomics Biotechnology Ltd., Mórahalom, Hungary
[d] School of Medicine, National University of Ireland, Galway, Ireland
[e] Nutrition Innovation Centre for Food and Health (NICHE), School of Biomedical Sciences, Ulster University, Coleraine, UK

ARTICLE INFO

Article history:
Received 26 September 2022
Received in revised form
16 November 2022
Accepted 21 November 2022
Available online 29 November 2022

Keywords:
SLST
MLST
Discrimination
Cutibacterium acnes
Recombination
CUTIS-SEQ typing

ABSTRACT

Objectives: A 'high resolution' Single Locus Sequence Typing (SLST) scheme has been described for the anaerobic skin bacterium *Cutibacterium acnes* that seemingly discriminates sequence types (STs) to a level commensurate with previously described Multilocus Sequence Typing (MLST) methods ($MLST_4$; $MLST_8$; $MLST_9$). However, no quantifiable evaluation of SLST versus MLST for differentiation of *C. acnes* strains, especially in relation to the subspecies of the bacterium, known as *C. acnes* subsp. *acnes* (type I), *C. acnes* subsp. *defendens* (type II) and *C. acnes* subsp. *elongatum* (type III), has been performed which is vital given its increasing use. To address this, we examined the discriminatory power of SLST versus MLST with a large group of isolates representative of all subspecies.
Methods: Simpson's index of diversity (*D*) was used for quantitative comparison of the resolving power of the SLST and MLST schemes for 186 isolates of *C. acnes* covering all three subspecies.
Results: When strains were considered collectively, SLST and all three MLST approaches had similar *D* values > 90%. However, at the subspecies level there were significant differences between the methods, most strikingly a reduced discrimination of type II and type III strains (*D* <80%) by SLST versus $MLST_8$, and to a lesser extent $MLST_4$. The $MLST_9$ method also performed poorly for type II strains (*D* <70%), but did display the best results for type I (*D* = 90%). By combining the SLST locus with the *camp2* gene sequence to create a novel and flexible high-resolution Bilocus Sequence Typing (BLST) scheme, known as CUTIS-SEQ typing (**CUTI**bacterium acne**S** Bilocu**S** s**EQ**uence Typing), we achieved improved resolution at both species and, critically, subspp. levels.
Conclusions: CUTIS-SEQ provides an opportunity to improve differentiation of *C. acnes* isolates by SLST without significantly impacting laboratory workload, or compromising application to complex biological communities. A CUTIS-SEQ isolate database is now available as part of the *C. acnes* PubMLST database at https://pubmlst.org.

1. Introduction

Cutibacterium (previously *Propionibacterium*) *acnes* is a Gram-positive human anaerobe found predominantly on the skin as part of the normal microbiota. Specific lineages of *C. acnes* are strongly linked to the development of the skin disorders acne [1–3] and progressive macular hypomelanosis [4,5], as well as conditions beyond the skin, while others are associated with maintaining skin health [1,2,6–11]. A range of different DNA-based typing approaches for *C. acnes* have been described since 2005, including multiplex PCR, multiple locus variable number of tandem repeat analysis and sequencing of single (e.g., *recA* and *tly*) and multiple genes [12]. In the latter case, two distinct multilocus sequence typing (MLST) schemes based on the analysis of eight ($MLST_8$; 4253 bp) and nine ($MLST_9$; 4233 bp) protein-encoding loci [1,13], and their accompanying public databases, have been widely utilised; overall, these methods generate phylogenies that are broadly consistent with whole genome-based typing. Detailed analysis of the literature demonstrates that both schemes appear to have been

* Corresponding author. Nutrition Innovation Centre for Food and Health, School of Biomedical Sciences, Ulster Univetsity, Coleraine, UK.
E-mail address: a.mcdowell@ulster.ac.uk (A. McDowell).

https://doi.org/10.1016/j.anaerobe.2022.102671

used at similar rates, with their databases containing comparable numbers of resolved sequence types (STs). These schemes also provide similar clustering of strains into different clonal complexes (CC), but some differences do exist in the discrimination of individual STs due to the different gene targets analysed [13]. The typing schemes also provide much higher resolution of *C. acnes* versus a previously described ribotyping scheme (2, 13).

Although MLST has been the gold standard for global epidemiology and genetic population studies of bacteria, offering transferable results between laboratories, it is expensive and laborious. As a consequence, McDowell et al. [3] described an approach to reduce this burden for *C. acnes* by demonstrating that if only 4/8 alleles of the $MLST_8$ scheme (*aroE*, *guaA*, *camp2*, *tly*) are sequenced, this information can still be used to predict with high level accuracy the full eight-gene allelic profile via cross-referencing to previously determined STs deposited within the database available at PubMLST (https://pubmlst.org/) [3]. This approach, now known as $MLST_4$, is possible due to the clonal nature of *C. acnes* and the relatively restricted number of STs that occur. A similar approach was subsequently utilised by Lomholt et al. [14] to stratify isolates into phylogroups and CCs based on *recA* and *fba* allele combinations related to the $MLST_9$ scheme; this information was then used to select isolates for full $MLST_9$ analysis.

A 'high resolution' Single Locus Sequence Typing (SLST) scheme based on analysis of a 483–497 bp non-recombinant sequence present upstream of the CAMP factor 1 gene (KPA171202 genome coordinates, 1463044–1463527) was subsequently described for differentiation of *C. acnes* strains to a level matching that of MLST [15]. Although this SLST method does not facilitate CC analysis, and the clustering of certain phylogroups within the type I clade varies from that observed by conventional and core genome MLST analysis, its apparent capacity for high resolution while saving on time, labour and expense has rightfully promoted this scheme to the best approach for initial *C. acnes* strain typing. Furthermore, as it is a single locus it can also be used to profile multiple *C. acnes* STs in complex bacterial communities which is a major advantage [15]. Despite the increasing use of the SLST method, its index of diversity for differentiation of *C. acnes* strains, including within the three main subspecies, known as *C. acnes* subsp. *acnes* (type I), *C. acnes* subsp. *defendens* (type II) and *C. acnes* subsp. *elongatum* (type III) [16], has not been examined. Given that different strains, especially between the distinct subspecies, differ with respect to the production of putative virulence determinants, immunogenic, biochemical and morphological properties and, most critically, association with health and disease [17], confidence in high strain-level discrimination across these different groups is vital.

We now describe a study comparing the diversity index of SLST versus MLST for differentiation of *C. acnes* STs, and propose a new Bilocus Sequence Typing (BLST) scheme, known as CUTIS-SEQ typing (**CUTI**bacterium acne**S** Bilocu**S** s**EQ**uence typing) that combines the SLST and *camp2* loci for flexible high resolution typing of the bacterium at the species and subspecies levels.

2. Methods

2.1. C. acnes isolates

A total of 168 publicly available draft and closed whole genome sequences of *C. acnes,* isolated from normal skin and a diverse range of clinical sources, were used for typing comparison and BLST development; this genomic data was retrieved from the National Centre for Biotechnology Information (NCBI) website. Isolates covered all three subspecies of the bacterium (116 type I; 47 type II; 5 type III). An additional 18 isolates in our bacterial collection, representing type II (n = 3) and type III (n = 15) strains, were also analysed as part of the investigation. On the basis of $MLST_8$ allelic profile, the evolutionary relationships between all the STs analysed is illustrated in Fig. S1. A very small number of draft *C. acnes* genomes deposited at NCBI were found to have unreliable sequence reads for the loci under investigation and were therefore not used. All isolates analysed and their CUTIS-SEQ genotypes are available to view as part of the *C. acnes* PubMLST database at https://pubmlst.org.

2.2. SLST analysis

SLST-PCR analysis was carried out following the method previously described [15]. Genomic DNA was prepared using a MasterPure gram-positive DNA Purification Kit (Lucigen, Cambio Ltd, UK) and PCR amplicons purified using a QIAquick PCR purification kit (Qiagen, UK). Sequencing reactions were performed using BigDye Terminator sequencing chemistry (ThermoFisher Scientific, UK), according to the manufacturer's instructions; SLST primers were used for forward and reverse sequencing. The samples were then analysed on an ABI Prism 3100 capillary electrophoresis system (ThermoFisher Scientific, UK). Sequences were trimmed and STs assigned via the SLST database (http://medbac.dk/slst/pacnes). STs for isolates where whole genome data was available were similarly determined after extraction of SLST sequences.

2.3. In silico typing

Allelic profiles and STs for the $MLST_8$ and $MLST_9$ schemes were determined after extraction of allelic sequences from whole genome data, followed by database analysis at http://pubmlst.org/pacnes/and http://pacnes.mlst.net/, respectively. $MLST_4$ profiles were determined as previously described [3]. All $MLST_9$ STs had been assigned from whole genome data prior to the start of this study. For CUTIS-SEQ development, extracted *tly* and *camp2* gene sequences were combined with SLST STs to create new ST designations for analysis.

2.4. Bioinformatic analysis

Phylogenetic analysis of SLST and BLST trees was carried out using the Minimum Evolution algorithm in MEGA v11.0 [18]. For genome phylogenies, strain assemblies were annotated using PROKKA v1.14.5 [19] and a core genome alignment created using MAFFT v 7.407 [20] before phylogenetic analysis using FastTree v2.1.10 [21]. Results were visualised using the Interactive Tree of Life (iTOL) v 5.0. [22]. A split decomposition tree was generated using SplitsTree version 4.1., and statistically significant recombination identified using the Phi test [23]. Formation of CCs and relationships between STs was determined by geoBURST analysis of allelic profiles [24].

2.5. Typing discrimination and statistical analysis

Simpson's index of diversity (D) for quantitative comparison of the resolving power of the different typing schemes was calculated as previously described [25] using the equation shown below, with a D threshold of $\geq$0.9 (90%) considered a minimum cutoff for confident interpretation of results [25].

$$D\,(\%) = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{s} n_j(n_j - 1)\ \text{x}\ 100$$

This index calculates the probability that two unrelated strains from a test population will be resolved into different groups. The

congruence between the typing schemes was determined using the adjusted RAND index. Confidence intervals (CI) were set at 95% and p values were calculated using the Jackknife pseudo-values resampling method with statistical significance set at $p < 0.05$.

3. Results

3.1. Comparison of SLST versus MLST schemes for species level discrimination

To examine differences in resolving power between SLST and all MLST schemes designed for *C. acnes*, we initially investigated available STs for 145 whole genomes (Table 1). This included a total of 116 type I, 27 type II and two type III strains. It was not possible to further expand the number of type II and III strains at this point of the analysis as the Aarhus $MLST_9$ database (http://www.mlst.net) was no longer available to facilitate ST assignment from newer whole genome data deposited; furthermore, the number of available type III whole genomes released was very small. At the species level, SLST resolved a smaller number of STs (n = 34) compared to the $MLST_4$ (n = 43), $MLST_8$ (n = 56) and $MLST_9$ (n = 37) methods, but the *D* value of 91.3% was above the ≥90% threshold and not statistically different from any of the other schemes despite being slightly lower in most cases (Table 1). We also compared all *D* values to a 'hybrid' 17-locus MLST scheme ($MLST_{17}$) that combined typing data from both the $MLST_8$ and $MLST_9$ methods. Given the greatly increased number of loci analysed, the $MLST_{17}$ approach showed extremely high resolution with a *D* of 96.6%, which was statistically very different from that generated with the other typing approaches ($p < 0.001$ in all cases), and therefore provided a very high discriminatory baseline for comparison to the other schemes. Despite having a propensity to generate a higher number of partitions, the $MLST_8$ scheme was not significantly different from $MLST_9$ in terms of overall discriminatory power ($p = 0.644$), with the latter having a marginally higher *D* value (Table 1). While no statistical difference in discrimination was observed between the $MLST_4$ approach versus $MLST_9$ ($p = 0.173$), a statistical difference between $MLST_4$ and $MLST_8$ was detected ($p = 0.009$) (Table 1) consistent with previous work [3].

With this dataset, we also used the adjusted Rand index to compare the overall congruence between SLST and the other typing methods. Overall, the congruence was limited and ranged from 52% (40.7-62.7%) for $MLST_9$, 50.7% (95% CI 39.4–62.4%) for $MLST_4$, 46.4% (95% CI 34.7–58.6%) for $MLST_8$ and 32% (20.4-43.2%) for the highly discriminatory $MLST_{17}$.

3.2. Comparison of SLST versus MLST schemes for type I and II subspecies level discrimination

We also examined differences across all typing schemes for the resolution of strains within *C. acnes* subsp. *acnes* (type I) and *C. acnes* subsp. *defendens* (type II); we were only able to examine type I and type II strains between all the schemes given the low number of *C. acnes* subsp. *elongatum* (type III) whole genomes available for any meaningful statistical analysis.

With type I strains (n = 116), the $MLST_9$ scheme resolved a total of 29 STs with a *D* of 90.4%, which was superior than the results generated with the $MLST_4$ (29 STs; 86%), $MLST_8$ (38 STs; 88%) and SLST (26 STs; 87.5%) methods (Table 1), all of which fell below the desirable ≥90% threshold value for *D*. While not statistically different to the $MLST_8$ ($p = 0.274$) or SLST ($p = 0.154$) schemes, it was when compared to $MLST_4$ ($p = 0.046$). As before, the $MLST_{17}$ scheme generated superior results with 55 resolved STs and a *D* value of 95.0% which was statistically very significant compared to all the other typing methods ($p > 0.001$ in all cases). The SLST method also had reduced ability to differentiate genotypes from acne-associated CC18 ($MLST_9$) ($p = 0.027$ versus $MLST_9$; 45 isolates examined), including the previously described epidemic clone ST18; this genotype was classified as A1 along with a number of other lineages ($p < 0.001$ versus $MLST_9$).

With type II strains (n = 27), the $MLST_8$ scheme resolved a total of 17 STs with a *D* of 94.3%, which was a striking difference compared to the results generated by both the $MLST_9$ (7 STs; 68.7%) and SLST (7 STs; 77.5%) approaches which demonstrated poor discrimination values that fell well below the ≥90% cut-off value (Table 1); these differences were statistically very significant ($p < 0.001$ and $p = 0.003$ versus $MLST_9$ and SLST, respectively). Given these surprising results, we examined the nature of these large differences at the phylogenetic level, especially for $MLST_9$, using a genetically heterogeneous group of 23 type II strains (from this dataset) which differentiated into three major clades or genetic groups based on a core genome phylogeny (Fig. 1). The $MLST_8$ method resolved these strains into 14 STs that were essentially consistent with their phylogenetic lineage, with the exception of ST6 in clade 3. In contrast, $MLST_9$ and SLST resolved only 6 and 7 STs, respectively. With $MLST_9$, clades 1 and 2 showed little differentiation, and within clade 3 all statistically significant subclades (100% bootstraps) were similarly not partitioned (Fig. 1). With SLST, better resolution was seen relative to $MLST_9$ consistent with the higher *D* value, but for some strains genotype K2 was observed between clades 1 and 2, while K1 was widely represented and shared between key subclades of group 3 (Fig. 1). Many of these latter typing issues were resolved with the BLST method (Fig. 1).

The $MLST_{17}$ scheme resolved 18 STs with a *D* of 94.6% which was almost identical to that obtained with the $MLST_8$ method, but very different to that for SLST and $MLST_9$ ($p = 0.003$ and < 0.001 versus SLST and $MLST_9$, respectively). The $MLST_4$ approach, while producing a lower *D* of 91.5% (13 STs), was not statistically different from either the $MLST_8$ ($p = 0.122$) or $MLST_{17}$ ($p = 0.118$) schemes for type II resolution, but was when compared to the $MLST_9$ ($p < 0.001$) and SLST ($p = 0.009$) methods.

Table 1
Number of STs and Simpson's index of diversity for 145 *C. acnes* genomes by MLST and SLST.

Typing	All (n = 145)		Type I (n = 116)		Type II (n = 27)		Type III (n = 2)	
	STs	*D* (95% CI)	STs	*D* (95% CI)	STs	*D* (95% CI)	STs	*D* (95% CI)
$MLST_{17}$	74	96.6% (95.1-98.1)	55	95.0% (92.8-97.2)	18	94.6% (89.0–1.00)	1	–
$MLST_9$	37	92.8% (90.8-94.8)	29	90.4% (87.3-93.5)	7	68.7% (57.6–79.7)	1	–
$MLST_8$	56	92.1% (89.2–95.1)	38	88.0% (83.8–92.2)	17	94.3% (88.8–99.8)	1	–
$MLST_4$	43	90.7% (87.7–93.8)	29	86.0% (81.6–90.3)	13	91.5% (86.1–96.8)	1	–
SLST	34	91.3% (88.5–94.0)	26	87.5% (83.5–91.6)	7	77.5% (68.0–87.0)	1	–

D = diversity index.
95% CI = 95% confidence interval.

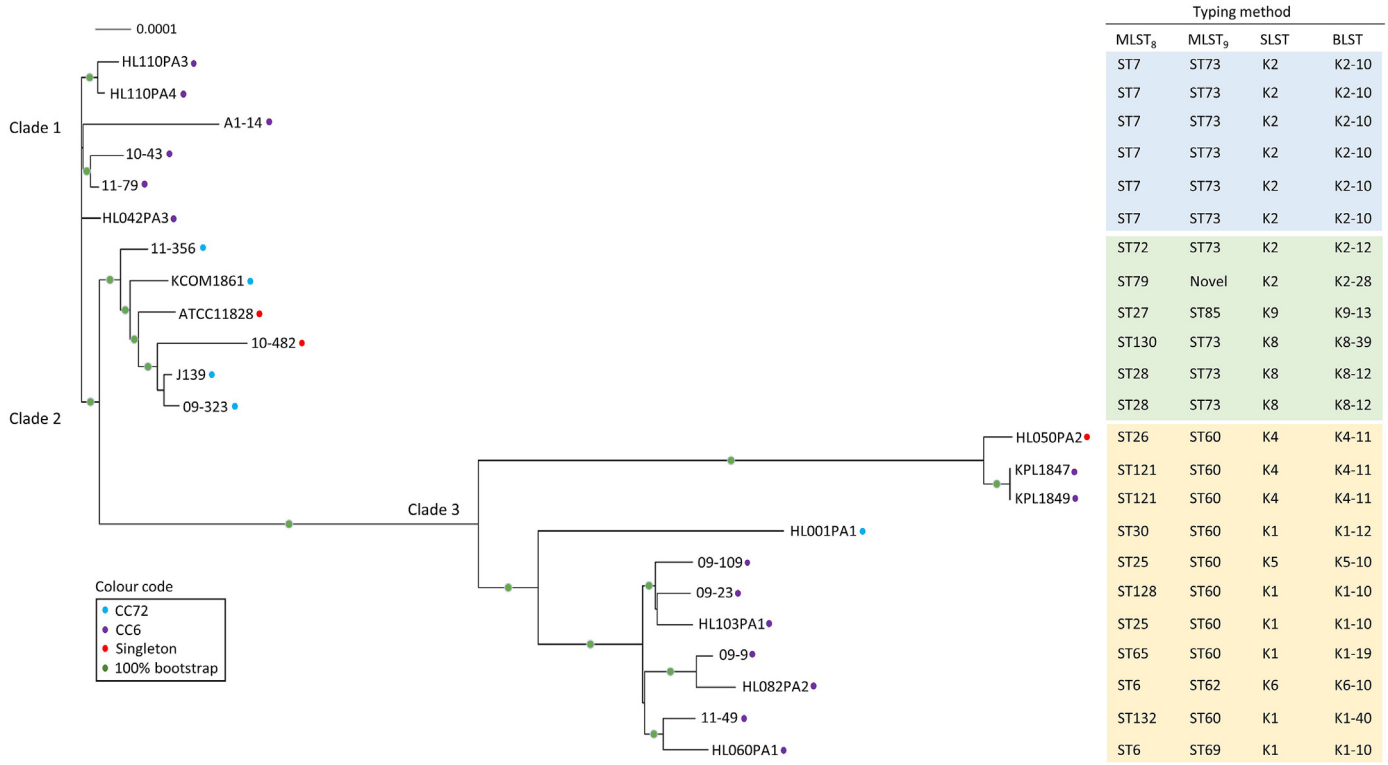

Fig. 1. Core genome phylogenies of 23 type II strains. Strain assemblies were annotated using PROKKA v1.14.5 (19) and a core genome alignment created using MAFFT v 7.407 (20) before phylogenetic analysis using FastTree v2.1.10 (21). Results were visualised using Interactive Tree of Life (iTOL) v 5.0 (23). MLST, SLST and BLST (SLST-*camp2*) STs are shown for each corresponding strain of the tree, with CCs indicated using a colour code.

3.3. Expanded analysis of SLST versus $MLST_8$ for discrimination of *C. acnes subsp. defendens* (type II)

Given the results obtained upon initial analysis of type II strains across all typing schemes, we investigated the robustness of the data with a larger cohort of type II isolates (n = 50). To do this, we compared the SLST results with $MLST_4$ and $MLST_8$ data for a further 20 type II isolates that comprised more newly deposited genome data, as well as three isolates in our collection which we analysed by SLST; MLST data was already available for these isolates. Analysis by $MLST_9$ could not be performed due to the unavailability of a database for ST assignment. With this much larger cohort of isolates, the $MLST_8$ scheme resolved a total of 22 STs with a *D* of 93.5% versus 9 STs and a *D* of 79.3% for the SLST method, and this was statistically very significant consistent with the previous results ($p > 0.001$) (Table 2). The $MLST_4$ approach identified 17 STs with a *D* of 90.6% which was better than SLST ($p = 0.009$), but poorer than $MLST_8$ with this expanded number of isolates ($p = 0.016$).

Table 2
Expanded analysis of SLST versus $MLST_4$ and $MLST_8$ for discrimination of *C. acnes* subsp. *acnes* (type II) and *C. acnes* subsp. *elongatum* (type III).

Method	Type II (n = 50)		Type III (n = 20)	
	STs	*D* (95% CI)	STs	*D* (95% CI)
$MLST_4$	17	90.6% (87.1–94.2)	7	82.6% (73.2–92.0)
$MLST_8$	22	93.5% (90.4–96.6)	10	91.1% (84.6–97.5)
SLST	9	79.3% (72.7–85.9)	6	76.8% (62.1–91.6)

D = diversity index.
95% CI = 95% confidence interval.

3.4. SLST versus $MLST_8$ for discrimination of *C. acnes subsp. elongatum* (type III)

To gain some insight into potential differences between typing schemes for the resolution of strains within *C. acnes* subsp. *elongatum* (type III), we compared SLST and $MLST_8$ for the discrimination of a total of 20 type III strains. These included the original two isolates and a further three for which openly available whole genome data was just available for $MLST_8$ analysis, alongside a further 15 isolates in our collection which we analysed by SLST PCR and sequencing; $MLST_8$ data was already available for these strains. As before, $MLST_9$ could not be performed due to unavailability of a database for ST assignment. The $MLST_8$ scheme resolved a total of 10 STs with a *D* of 91.1% versus six STs and a considerably reduced *D* of 76.8% by the SLST method which again fell well below the ≥90% cut-off value (Table 2). The $MLST_4$ method had a *D* of 82.6%, and this was not statistically different from either $MLST_8$ ($p = 0.062$) or SLST ($p = 0.601$).

3.5. Comparison of SLST versus BLST

To enhance the discrimination of strains across all subspecies, especially types II and III, we investigated the resolving power of a BLST scheme that comprised the SLST locus combined with either *tly* (777 bp; KAP171202 genome coordinates 1514497–1515273) or *camp2* (804–807 bp; KAP171202 genome coordinates 756446–757249) genes which comprise part of the $MLST_8$ method. With data from all the isolates combined (n = 186), SLST resolved a total of 41 STs with a *D* of 93.4% (CI 91.7–95.2%) versus 94.5% (CI 92.9–96.6%) and 93.7% (CI 91.7–95.6%) for $MLST_8$ and $MLST_4$, respectively. With the addition of *tly* and *camp2* sequence data to

the SLST locus, this increased significantly to 94.6% (51 STs; $p = 0.003$) and 94.9% (58 STs; $p < 0.001$), respectively (Table 3); individually, the *tly* and *camp2* genes generated *D* values of 82.4% (19 STs) and 87.3% (26 STs), respectively (Table 3).

At the subspecies level, the best improvements from SLST were observed with the SLST-*camp2* combination, especially in relation to partitioning of type II and type III strains where large statistically significant increases in *D* values were observed to >90% (Table 3; Figs. 1 and 2); $p < 0.001$ and 0.048 for type II and type III differences, respectively. Interestingly, the results obtained for discrimination of type II strains with the *camp2* gene alone where not significantly different to that obtained using the SLST approach, further demonstrating the latter's weaker resolving power for this subspecies (*D* values of 75.4 vs 79.3%, respectively; $p = 0.519$) (Table 3). Furthermore, the BLST method helped to resolve some of the discrepancies observed with SLST resolution of type II strains present in different clades upon core genome analysis (Fig. 1). With type I strains, where the baseline *D* value was much higher than that observed with type II and III strains, a more modest increase in *D* to a final value 88.6% was observed with the addition of *camp2* (Table 3; Fig. 2), but this did not reach significance despite the low p value of 0.054. Within the type I clade, the BLST scheme with *camp2* did help to discriminate a number of isolates that formed distinct lineages on previously described whole genome trees of *C. acnes*, but yet had the same SLST genotypes [15,26]. As example, type IA_2 organisms HL030PA2 and HL063PA2 form a clade distinct from that formed by P. acn31 and HL037PA1 (15,26). While the latter strain has the unique ST F6, the strains HL030PA2, HL063PA2 and P. acn31 all share ST F4. On BLST, this discrepancy is resolved as P. acn31 becomes ST F4-2, while HL030PA2 and HL063PA2 are STs F4-7 and F4-8, respectively. While the SLST-*tly* combination actually proved marginally better than *camp2* at discriminating type I strains with a *D* value of 89.4% ($p = 0.034$), the latter was still selected for use with the SLST locus given its better performance overall, especially in relation to type II and III strains.

3.6. Analysis of recombination

No evidence of statistically significant recombination was evident when the concatenated SLST-*camp2* sequences were examined by split decomposition analysis (Phi test $p = 0.280$) (Fig. 3). A clear tree-like structure where the major sub-species (types I, II and III) were resolved into distinct clusters with 100% bootstrap support was evident, with minimal network structure.

For comparison, Phi test analysis of sequences currently comprising the individual SLST and *camp2* databases (the latter is part of the $MLST_8$ database) also demonstrated non-statistically relevant levels of recombination with p values of 0.336 and 0.695, respectively. Inspection of Fig. 2 also demonstrated that no SLST or *camp2* alleles are shared between the three subspp. divisions of the bacterium.

4. Discussion

Our data demonstrates that while congruence or agreement between SLST and other published MLST schemes for typing of *C. acnes* appears limited based on adjusted Rand index, the discriminatory power is comparable when considering all strains at the species level. However, at the subspecies level, differences in resolution were observed, most notably a strikingly reduced differentiation of *C. acnes* subsp. *defendens* (type II) by SLST and $MLST_9$ compared to $MLST_8$, and to a lesser extent $MLST_4$. The SLST method also had a slightly lower discriminatory ability versus $MLST_9$ for resolution of *C. acnes* subsp. *acnes* strains, but this was not statistically significant; the $MLST_9$ method had the best *D* value for all the typing schemes examined in relation to type I strains (90%) (excluding $MLST_{17}$), although statistical significance was only reached versus $MLST_4$. Of particular note, SLST had a reduced ability to differentiate the acne-associated type IA_1 CC18, and its founding epidemic clone ST18 strongly linked with moderate-to-severe forms of the condition (1). Although *C. acnes* subsp. *elongatum* (type III) strains form a much tighter phylogenetic structure when compared to type I and II genetic groups, the SLST scheme still generated a much lower *D* value compared to $MLST_8$; this analysis was based on a smaller number of type III strains available for study, so further investigation with a larger collection of isolates when available will be required to confirm this result. Unfortunately, as the $MLST_9$ database was no longer available during this study, it was not possible to compare this scheme to the other typing methods for discrimination within the type III group.

The $MLST_8$ method is a modified and expanded version of the original University of Warwick $MLST_7$ scheme (3135 bp) for *C. acnes*; the latter resolved different STs within types IA_1, IA_2, IB, IC, II and III (13). The Warwick scheme was independently developed at the same time as the University of Bath $MLST_9$ method, which was subsequently modified and adopted by researchers at Aarhus University [1]. The $MLST_7$ typing approach was later amended by researchers at Queen's University, Belfast, with removal of the *recA* locus and addition of the genes for the extracellular/cell-surface-associated putative virulence determinants *tly* (haemolysin/cytotoxin and/or RNA-binding FtsJ-like methyltransferase) and *camp2* (co-haemolysin) (13). The use of such genes, which often demonstrate strong diversifying or directional selection, can be a good compromise for enhanced discrimination of clonal organisms, such as *C. acnes*, versus the sequencing of larger numbers of housekeeping loci that normally show less genetic variation; these genes can be used as 'genetic magnifying glasses' to help subtype clones of particular interest [27]. Furthermore, the addition of more rapidly evolving genes does not appear to significantly affect assignment to specific clonal lineages [27]. Indeed, assignment of STs to specific CCs by $MLST_8$ essentially matches that achieved by $MLST_9$, although a greater number of type II CCs are identified by $MLST_8$ due to its enhanced discriminatory power for this

Table 3
Discriminatory power of the BLST scheme based on a combination of the SLST locus and *tly* or *camp2* genes.

Typing method	All (n = 186)		Type I (n = 116)		Type II (n = 50)		Type III (n = 20)	
	STs	*D* (95% CI)	STs	*D* (95% CI)	STs	*D* (95% CI)	STs	*D* (95% CI)
tly	19	82.4% (79.9–84.8)	12	70.0% (65.7–74.4)	4	22.4% (7.10–37.8)	3	56.8% (40.7–72.9)
camp2	26	87.3% (84.9–89.7)	12	73.0% (68.7–77.4)	10	75.4% (66.9–84.0)	4	56.3% (33.7–78.9)
SLST[a]	41	93.4% (91.7–95.2)	26	87.5% (83.5–91.6)	9	79.3% (72.7–85.9)	6	76.8% (62.1–91.6)
SLST-*tly*	51	94.6% (93.0–96.1)	32	89.4% (85.7–93.1)	11	83.8% (78.5–89.1)	8	88.4% (82.3–94.5)
SLST-*camp2*	58	94.9% (93.2–96.7)	32	88.6% (84.6–92.7)	17	92.2% (89.7–94.8)	9	91.4% (84.4–96.6)

D = diversity index.
95% CI = 95% confidence intervals.
[a] SLST data combined with information available in Tables 1 and 2

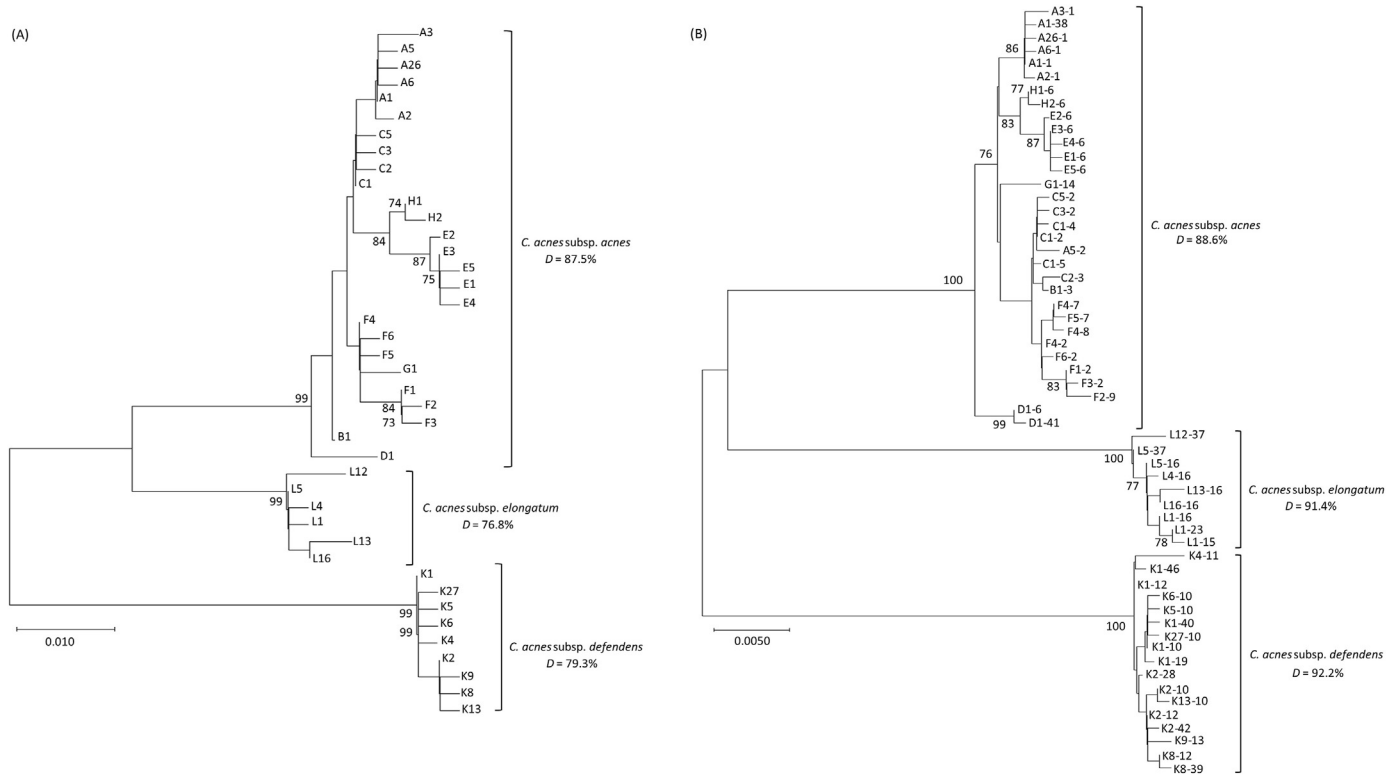

Fig. 2. Minimum evolution phylogenetic trees of STs resolved using SLST (A) and BLST (B) schemes. In the latter case, analysis was performed on concatenated sequences. Bootstrapping statistics were performed using 500 datasets, and only bootstrap values ≥ 70% are shown. The evolutionary distances were computed using the Tajima-Nei method. The subspecies status of all resolved isolates is indicated. The *D* values obtained for each subspecies are also shown for the SLST and BLST schemes.

subspecies. Interestingly, subsequent analysis of *tly* and *camp2* allele sequences has shown weak-to-no evidence of diversifying selection based on a range of selection tests, and these genes appear to have levels of genetic diversity akin to that for housekeeping loci with which they appear to have co-evolved [1,3]. This has led to the suggestion that the co-evolution of these genes with the core genome might reflect a role in commensal existence constraining their diversity and preventing loss from the *C. acnes* population (3). The absence of hypervariability in these loci is an especially noteworthy observation, particularly in relation to *camp2* which has been a target for the development of an acne vaccine [28].

As the *tly* and *camp2* genes comprise the $MLST_8$ scheme, their affiliation to a large number of isolates is already known. They also comprise a greater number of alleles within the MLST database due to the longer sequence length used versus other housekeeping loci; this affords more opportunity to identify novel SNPs. On this basis, we investigated whether the addition of one of these genes to the SLST locus could help create a novel high-resolution BLST scheme with an improved *D* value for strains from all *C. acnes* subspecies. Overall, combining the SLST locus with *camp2* generated the best results compared to *tly*, with increases in *D* value for type II and III strains to >90% being observed; the *tly* gene was also much closer to the SLST locus on the *C. acnes* chromosome versus *camp2* which was situated in a different section. While more modest increases in *D* value at the species and *C. acnes* subsp. *acnes* (type I) levels were obtained, these were still valuable and phylogenetically informative; in the latter case, BLST helped differentiate a number of type IA_2 isolates that formed distinct whole genome clades, but type IA_1 ST18 strains remained undifferentiated. Also, as *camp2* shows no evidence of statistically significant levels of recombination or diversifying selection, combining it with the SLST locus does not significantly impinge on the latter's non-recombinant nature, or affect major taxonomic inferences. In the latter case, strains sharing the same SLST letter designation reflective of subspecies and type I phylogroup relationships (15) remain clustered together on BLST analysis, with the exception of A5 and B1 which form part of the C clade. In the latter cases, this reflects $MLST_8$ classification of these strains as members of CC3 (SLST C designation).

As the *tly* and *camp2* loci have their own sequence databases, which in the case of *camp2* is embedded within the PubMLST database for *C. acnes*, it will be straightforward to assign existing or novel allele designations using these sites before combining to give an overall BLST ST, as in this study. Indeed, the presence of these open databases was a key driver in our decision to examine both *tly* and *camp2* for BLST. This method will also give researchers flexibility as the SLST approach can continue to be utilised as a standalone method, or it can be enhanced by combining with *camp2* to provide increased discrimination where appropriate. This 'opt in-opt out' approach could be at the species level, or confined to one or more subspecies of the bacterium depending on need; practically, the latter would focus on the type II and III strains given the much lower *D* values for these organisms. Furthermore, because we have used a hyphenated ST nomenclature comprising previous SLST and *camp2* allele type designations, tracking of BLST allele sequences to previous publications where SLST and $MLST_8$ have been used is facilitated which is another key feature of the method. Primer sequences for *camp2* analysis have also been previously described and widely used [29]. However, to help facilitate tracking and surveillance of CUTIS-SEQ genotypes, we have provided an isolate database as part of the PubMLST *C. acnes* database (https://pubmlst.org).

One valuable aspect of the SLST method has been its capacity to identify multiple *C. acnes* STs in complex biological samples via next

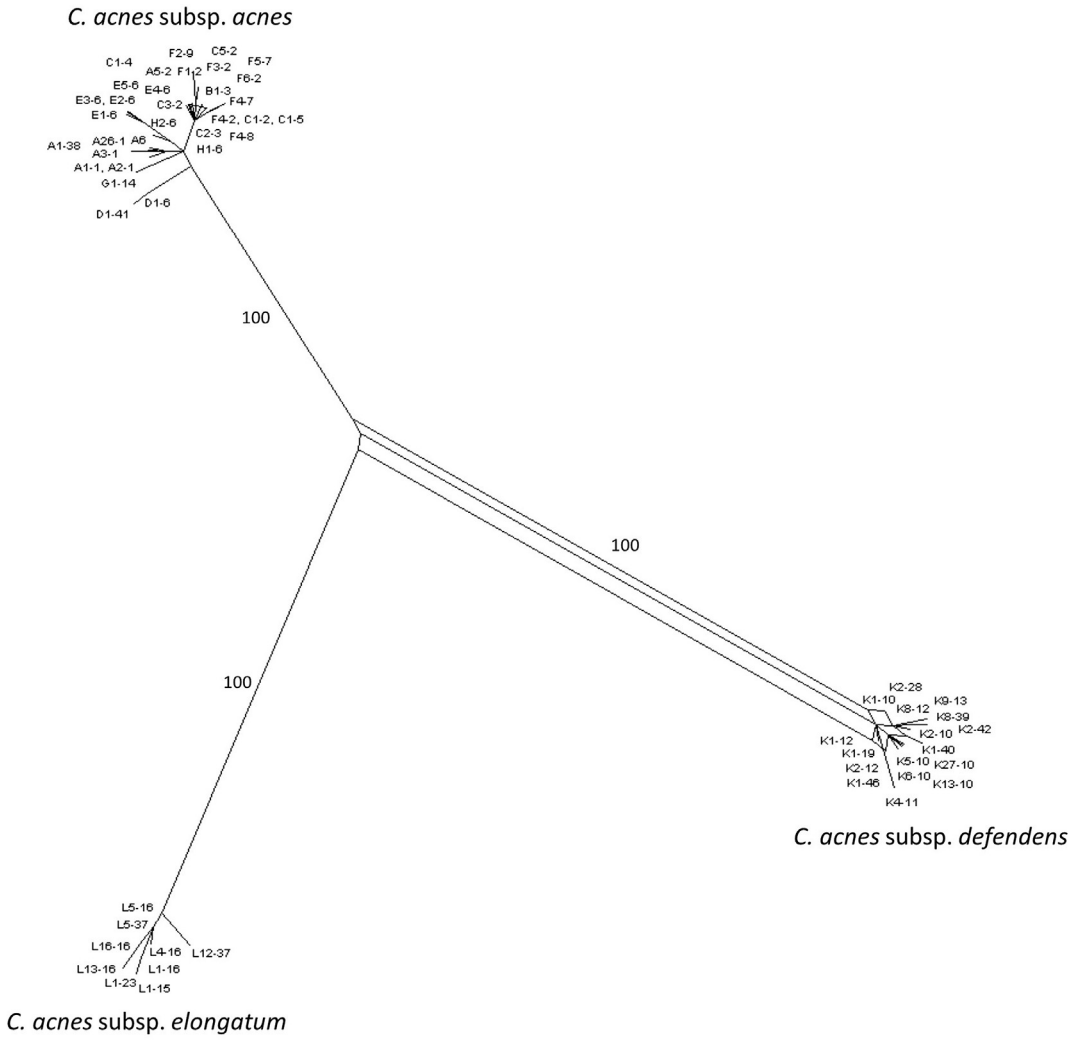

Fig. 3. Split decomposition analysis of concatenated BLST sequences using uncorrected p distances and bootstrapping statistics with 1000 datasets. A tree-like structure is apparent with all three subspecies resolved (100% bootstrap support). Phi test analysis revealed no statistically significant levels of recombination ($p = 0.280$), although some minimal parallelogram or network structure is observed.

generation amplicon sequencing (15). With multiple loci, however, linking sequenced alleles to different strains within a bacterial community is problematic. While the introduction of the software tool MetaMLST has enabled reconstruction of the most abundant ST of a species in complex metagenomic samples, including skin samples, it does not provide the relative abundances of the STs identified [30]. More recently, a new approach, known as metagenomic multi-locus sequence typing (MG-MLST), has been described for the identification of ST composition within a microbial community using the ancestry prediction algorithm STRUCTURE [31]; this method can be used for both shotgun and amplicon data. This method not only identifies the most dominant ST present, but less abundant strains alongside their relative proportions. The method has also been validated on clinical skin samples which included *camp2* as part of the gene set analysed. As a consequence, BLST analysis could also be applied directly to complex microbial communities for quantitative determination of *C. acnes* ST composition using the MG-MLST approach which we are now doing. This would not only help to reduce any bias introduced by analyzing only cultivable strains, but will facilitate high resolution typing of microbiomes for detailed strain-level differences in *C. acnes* bacteria associated with health and disease, as well as ecosystems of the human body.

5. Conclusion

We have found that SLST and MLST schemes display similar levels of resolution when strains are considered at the species level, but at the subspecies level significant differences exist between the methods. Of particular note, SLST demonstrates weaker resolution of type II and III strains with *D* values below the $\geq$90% threshold (<80%) which is the minimum value considered for reliable typing data. The type II clade represents a large group of strains which are believed to play an important role in maintaining skin health, but may also have a propensity to cause infections related to soft tissues, implants and the prostate gland [3,32,33]. In all these cases, further studies are required to better understand the role specific type II STs play in these conditions. Combining the SLST locus with the *camp2* gene will enhance population genetic typing of *C. acnes*, and create a novel and highly flexible high-resolution BLST scheme (~1250 bp) that we now call CUTIS-SEQ typing. Most importantly, the analysis of two loci should not greatly compromise sample

preparation time and labour, and the method should still be directly applicable to the analysis of *C. acnes* populations in clinical samples based on MG-MLST. If further population genetic data is required in relation to CC membership, this can be achieved using the $MLST_4$ or $MLST_8$ methods.

CRediT authorship contribution statement

Joseph McLaughlin: Formal analysis, Investigation, Writing – original draft, Visualization. **István Nagy:** Investigation, Writing – review & editing. **Georgios Miliotis:** Formal analysis, Writing – review & editing. **Andrew McDowell:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Visualization, Writing – original draft, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Data availability

Data will be made available on request.

Acknowledgments

J.M. is currently funded by a British Skin Foundation studentship to A.M.D (Grant Project No. 025/s/16). The study also made use of the *C. acnes* PubMLST database and we also thank Keith Jolley for assistance in setting up the CUTIS-SEQ isolate database within this site.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.anaerobe.2022.102671.

References

[1] H.B. Lomholt, M. Kilian, Population genetic analysis of Propionibacterium acnes identifies a subpopulation and epidemic clones associated with acne, PLoS One 5 (2010), e12277.
[2] S. Fitz-Gibbon, S. Tomida, B.-H. Chiu, L. Nguyen, C. Du, M. Liu, D. Elashoff, M.C. Erfe, A. Loncaric, J. Kim, R.L. Modlin, J.F. Miller, E. Sodergren, N. Craft, G.M. Weinstock, H. Li, *Propionibacterium acnes* strain populations in the human skin microbiome associated with acne, J. Invest. Dermatol. 133 (2013) 2152–2160.
[3] A. McDowell, I. Nagy, M. Magyari, E. Barnard, S. Patrick, The opportunistic pathogen Propionibacterium acnes: insights into typing, human disease, clonal diversification and CAMP factor evolution, PLoS One 8 (2013), e70897.
[4] E. Barnard, J. Liu, E. Yankova, S.M. Cavalcanti, M. Magalhães, H. Li, S. Patrick, A. McDowell, Strains of the *Propionibacterium acnes* type III lineage are associated with the skin condition progressive macular hypomelanosis, Sci. Rep. 6 (2016) 3196.
[5] R.L. Petersen, C.F. Scholz, A. Jensen, H. Brüggemann, H.B. Lomholt, *Propionibacterium acnes* phylogenetic type III is associated with progressive macular hypomelanosis, Eur. J. Microbiol. Immunol. (Bp) 7 (2017) 37–45.
[6] K.E. Piper, M.J. Jacobson, R.H. Cofield, J.W. Sperling, J. Sanchez-Sotelo, D.R. Osmon, A. McDowell, S. Patrick, J.M. Steckelberg, J.N. Mandrekar, M.F. Fernandez Sampdero, R. Patel, Microbiologic diagnosis of prosthetic shoulder infection by use of implant sonication, J. Clin. Microbiol. 47 (2009) 1878–1884.
[7] R.J. Cohen, B.A. Shannon, J.E. McNeal, T. Shannon, K.L. Garrett, *Propionibacterium acnes* associated with inflammation in radical prostatectomy specimens: a possible link to cancer evolution? J. Urol. 173 (2005) 1969–1974.
[8] S.A. Niazi, D. Clarke, T. Do, S.C. Gilbert, F. Mannocci, D. Beighton, *Propionibacterium acnes* and *Staphylococcus epidermidis* isolated from refractory endodontic lesions are opportunistic pathogens, J. Clin. Microbiol. 48 (2010) 3859–3869.
[9] L. Zaffiri, R. Abdulmassih, S. Boyaji, I. Bagh, A.R. Campbell, M.E. Loehrke, Brain abscess induced by *Propionibacterium acnes* in a patient with severe chronic sinusitis, New Microbiol. 36 (2013) 325–329.
[10] Y. Eishi, Etiologic link between sarcoidosis and *Propionibacterium acnes*, Respir. Invest. 51 (2013) 56–68.
[11] H.B. Albert, P. Lambert, J. Rollason, J.S. Sorensen, T. Worthington, M.B. Pedersen, H.S. Nørgaard, A. Vernallis, F. Busch, C. Manniche, T. Elliott, Does nuclear tissue infected with bacteria following disc herniations lead to Modic changes in the adjacent vertebrae? Eur. Spine J. 22 (2013) 690–696.
[12] A. McDowell, Over a Decade of *recA* and *tly* gene sequence typing of the skin bacterium *Propionibacterium acnes*: what have we learnt? Microorganisms 6 (2017) 1.
[13] A. McDowell, E. Barnard, I. Nagy, A. Gao, S. Tomida, H. Li, A. Eady, J. Cove, C.E. Nord, S. Patrick, An expanded multilocus sequence typing scheme for Propionibacterium acnes: investigation of "pathogenic", "commensal" and antibiotic resistant strains, PLoS One 7 (2012), e41480.
[14] H.B. Lomholt, C.F.P. Scholz, H. Brüggemann, H. Tettelin, M. Kilian, A comparative study of *Cutibacterium (Propionibacterium) acnes* clones from acne patients and healthy controls, Anaerobe 47 (2017) 57–63.
[15] C.F.P. Scholz, A. Jensen, H.B. Lomholt, H. Brüggemann, M. Kilian, A novel high-resolution single locus sequence typing scheme for mixed populations of *Propionibacterium acnes in vivo*, PLoS One 9 (2014), e104199.
[16] I. Dekio, A. McDowell, M. Sakamoto, S. Tomida, M. Ohkuma, Proposal of new combination, *Cutibacterium acnes* subsp. *elongatum* comb. nov., and emended descriptions of the genus *Cutibacterium, Cutibacterium acnes* subsp. *acnes* and *Cutibacterium acnes* subsp. *defendens*, Int. J. Syst. Evol. Microbiol. 69 (2019) 1087–1092.
[17] J. McLaughlin, S. Watterson, A.M. Layton, A.J. Bjourson, E. Barnard, A. McDowell, *Propionibacterium acnes* and acne vulgaris: new insights from the integration of population genetic, multi-omic, biochemical and host-microbe studies, Microorganisms 7 (2019) 128.
[18] K. Tamura, G. Stecher, S. Kumar, MEGA11: molecular evolutionary genetics analysis version 11, Mol. Biol. Evol. 38 (2021) 3022–3027.
[19] T. Seemann, Prokka: rapid prokaryotic genome annotation, Bioinformatics 30 (2014) 2068–2069.
[20] K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, Nucleic Acids Res. 30 (2002) 3059–3066.
[21] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree 2 - approximately maximum-likelihood trees for large alignments, PLoS One 5 (2010), e9490.
[22] I. Letunic, P. Bork, Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation, Bioinformatics 23 (2007) 127–128.
[23] D.H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies, Mol. Biol. Evol. 23 (2006) 254–267.
[24] A.P. Francisco, M. Bugalho, M. Ramirez, J.A. Carrico, Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach, BMC Bioinf. 10 (2009) 152.
[25] P.R. Hunter, M.A. Gaston, Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity, J. Clin. Microbiol. 26 (1988) 2465–2466, 1988.
[26] S. Tomida, L. Nguyen, B.H. Chiu, J. Liu, E. Sodergren, G.M. Weinstock, H. Li, Pan-genome and comparative genome analyses of *Propionibacterium acnes* reveal its genomic diversity in the healthy and diseased human skin microbiome, mBio 4 (2013) e00003–e00013.
[27] J.E. Cooper, E.J. Feil, Multilocus sequence typing – what is resolved? Trends Microbiol. 12 (2004) 373–377.
[28] S. Keshari, M. Kumar, A. Balasubramaniam, T.-W. Chang, Y. Tong, C.-M. Huang, Prospects of acne vaccines targeting secreted virulence factors of Cutibacterium acnes, Expert Rev. Vaccines 5 (2019) 433–437.
[29] S. Valanne, A. McDowell, G. Ramage, M.M. Tunney, G.G. Einarsson, S. O'Hagan, G.B. Wisdom, D. Fairley, A. Bhatia, J.F. Maisonneuve, M. Lodes, D.H. Persing, S. Patrick, CAMP factor homologues in *Propionibacterium acnes*: a new protein family differentially expressed by types I and II, Microbiology (Read.) 151 (2005) 1369–1379.
[30] M. Zolfo, A. Tett, O. Jousson, C. Donati, N. Segata, MetaMLST: multi-locus strain-level bacterial typing from metagenomic samples, Nucleic Acids Res. 45 (2017) e7.
[31] N.J. Bangayan, B. Shi, J. Trinh, E. Barnard, G. Kasimatis, E. Curd, H. Li, MG-MLST: characterizing the microbiome at the strain level in metagenomic data, Microorganisms 8 (2020) 684.
[32] H. Brüggemann, L. Salar-Vidal, H.P.M. Gollnick, R. Lood, A Janus-faced bacterium: host-beneficial and -detrimental roles of *Cutibacterium acnes*, Front. Microbiol. 12 (2021), 673845.
[33] K.J. Spittaels, R. Ongena, C.C. Zouboulis, A. Crabbé, T. Coenye, *Cutibacterium acnes* phylotype I and II strains interact differently with human skin cells, Front. Cell. Infect. Microbiol. 10 (2020), 575164.