


METHODOLOGY ARTICLE

Open Access



# Comparison of three amplicon sequencing approaches to determine staphylococcal populations on human skin

Charlotte Marie Ahle<sup>1,2</sup>, Kristian Stødkilde-Jørgensen<sup>3</sup>, Anja Poehlein<sup>4</sup>, Wolfgang R. Streit<sup>2</sup>, Jennifer Hüpeden<sup>1</sup> and Holger Brüggemann<sup>3\*</sup> 

## Abstract

**Background:** Staphylococci are important members of the human skin microbiome. Many staphylococcal species and strains are commensals of the healthy skin microbiota, while few play essential roles in skin diseases such as atopic dermatitis. To study the involvement of staphylococci in health and disease, it is essential to determine staphylococcal populations in skin samples beyond the genus and species level. Culture-independent approaches such as amplicon next-generation sequencing (NGS) are time- and cost-effective options. However, their suitability depends on the power of resolution.

**Results:** Here we compare three amplicon NGS schemes that rely on different targets within the genes *tuf* and *rpsK*, designated *tuf1*, *tuf2* and *rpsK* schemes. The schemes were tested on mock communities and on human skin samples. To obtain skin samples and build mock communities, skin swab samples of healthy volunteers were taken. In total, 254 staphylococcal strains were isolated and identified to the species level by MALDI-TOF mass spectrometry. A subset of ten strains belonging to different staphylococcal species were genome-sequenced. Two mock communities with nine and eighteen strains, respectively, as well as eight randomly selected skin samples were analysed with the three amplicon NGS methods. Our results imply that all three methods are suitable for species-level determination of staphylococcal populations. However, the novel *tuf2*-NGS scheme was superior in resolution power. It unambiguously allowed identification of *Staphylococcus saccharolyticus* and distinguish phylogenetically distinct clusters of *Staphylococcus epidermidis*.

**Conclusions:** Powerful amplicon NGS approaches for the detection and relative quantification of staphylococci in human samples exist that can resolve populations to the species and, to some extent, to the subspecies level. Our study highlights strengths, weaknesses and pitfalls of three currently available amplicon NGS approaches to determine staphylococcal populations. Applied to the analysis of healthy and diseased skin, these approaches can be useful to attribute host-beneficial and -detrimental roles to skin-resident staphylococcal species and subspecies.

## Background

Studying the skin microbiome is regarded increasingly important in understanding skin diseases as well as skin health. The genus *Staphylococcus* is one of the most abundant bacterial genera in the human skin

microbiome; it plays a central role on human skin and in health and disease [1–17]. While the skin colonization by *Staphylococcus aureus* is correlated with disease severity, as seen for example in atopic dermatitis [1], coagulase-negative staphylococci (CoNS) are regarded as having rather health-beneficial roles on human skin. Common CoNS species that can be found on human skin are *Staphylococcus epidermidis*, *Staphylococcus*

\* Correspondence: [brueggemann@biomed.au.dk](mailto:brueggemann@biomed.au.dk)

<sup>3</sup>Department of Biomedicine, Aarhus University, 8000 Aarhus, Denmark  
Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

*hominis*, *Staphylococcus capitis* and *Staphylococcus haemolyticus* and others [18, 19]. As an important host-beneficial mechanism of certain CoNS, colonization resistance can prevent the expansion of potential pathogens on the skin; this is achieved by different CoNS properties such as the production of bacteriocins and phenol-soluble modulins [2–5] and quorum-sensing interference [6, 7]. Other host beneficial mechanisms of CoNS include, for example, the training and fortification of skin immunity [8–11], supporting wound healing [12, 13], and, possibly, anti-cancer effects [14]. Such host-interacting functions of CoNS are often species-, subspecies-, phylotype- and even strain-specific [2, 11, 14, 20, 21].

It was shown that one individual is not only colonised by an array of different staphylococcal species, but also by different strains of each species, in particular of *S. epidermidis* [22, 23]. The population of *S. epidermidis* species consists of strains belonging to three main phylogenetically distinct clades (A, B and C) [24–26]. In addition, a myriad of individual strains within each clade can be distinguished that differ in the core genome by single nucleotide polymorphisms (SNPs) and in the flexible genome by strain-specific genomic islands and extrachromosomal plasmids [20, 24, 25, 27].

To comprehensively map staphylococci on human skin, specific methods are needed to determine populations beyond the genus and species level. Traditionally, studies concerning the determination of staphylococci employed cultivation approaches with solid agar-based media [28, 29]. This makes it possible to investigate the isolated strains regarding their geno- and phenotypes. Depending on the choice of media and growth conditions, cultivation methods can underrepresent slow-growing and fastidious microorganisms. In contrast, culture-independent methods employing next-generation sequencing (NGS) achieve a more comprehensive picture of the skin microbiome [30]. Previous culture-independent studies have often relied on the 16S rRNA gene. However, this gene is inadequate to sufficiently distinguish several different staphylococcal species, and does not discriminate populations beyond the species level [31, 32]. Alternative target genes for identification and differentiation of staphylococcal isolates were evaluated and proposed such as *kat* [33], *gap* [31, 34, 35], *hsp60* [36, 37], *rpoB* [38, 39] and *sodA* [40]. One of the most established gene targets is *tuf*, which codes for elongation factor Tu (EF-Tu) [41]. The *tuf* gene, or rather fragments thereof, is also used as a target for analysing mixed staphylococcal communities with NGS methods, hereafter referred to as amplicon NGS [42–45]. Furthermore, the staphylococcal *rpsK* gene that encodes the 30S ribosomal protein S11 was recently proposed for amplicon NGS [46].

Here, we first isolated staphylococcal strains obtained from skin swabs of healthy volunteers, in order to assemble staphylococcal mock communities. We then compared three amplicon NGS schemes for their suitability to determine the staphylococcal populations of these mock communities. Two tested amplicon NGS schemes target different *tuf* gene fragments; one was developed by Martineau et al. [41] and the other one by Ahle et al. [47], designated here *tuf1* and *tuf2* scheme, respectively. The third tested scheme, designated *rpsK* scheme, targets a fragment of the *rpsK* gene and was developed by Ederveen et al. [46]. Lastly, the three amplicon NGS schemes were tested on skin swab samples obtained from healthy volunteers.

## Results

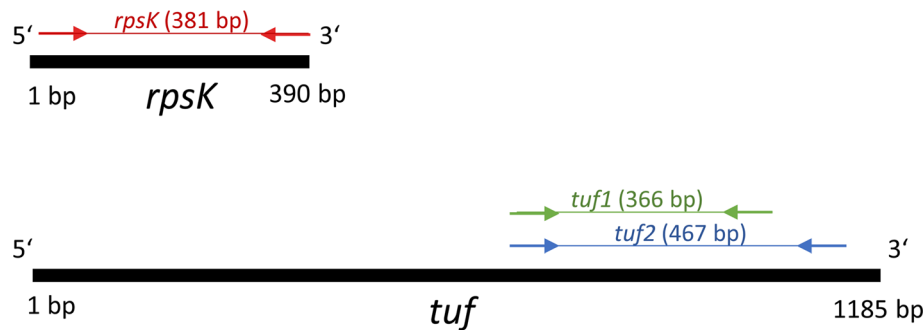
### Origin of targets for amplicon NGS

Here, we compared three different amplicon NGS schemes, all previously published and designed for determining staphylococcal populations in mixed communities: *tuf1* and *tuf2* schemes target the *tuf* gene and the *rpsK* scheme targets the *rpsK* gene. The *rpsK* and *tuf1* amplicon targets have a similar length with 381 and 366 bp, respectively, while the *tuf2* amplicon target is with 467 bp the longest among the three targets (Fig. 1). The amplicon targets *tuf1* and *tuf2* overlap to some extent (position 688 to 1053 bp and 685 to 1151 bp, respectively).

### Amplicon NGS of bacterial mock communities

Two different bacterial mock communities (M1/M2) were prepared, containing DNA of nine (M1) and 18 (M2) different staphylococcal strains, respectively (Additional file 1). All utilized strains belonged to staphylococcal species commonly found on human skin. They originated either from publicly accessible collections or were isolated here from healthy skin. Strains isolated here, in total 254 strains, were first identified to the species level by MALDI-TOF mass spectrometry (Additional file 2). A subset of ten strains belonging to different staphylococcal species were subsequently genome sequenced to obtain a complete genome database for the two mock communities (Additional file 3).

To test whether the three NGS schemes can distinguish between staphylococci on species level, the mock community M1 was analysed. The mock community M1 contained one strain each of *S. aureus*, *S. capitis*, *S. epidermidis*, *S. haemolyticus*, *S. hominis*, *Staphylococcus saccharolyticus*, *Staphylococcus saprophyticus*, *Staphylococcus simulans* and *Staphylococcus warneri* (Additional file 1). DNA was pooled in equimolar ratios and the three amplicon NGS pipelines were applied. All three schemes were able to identify and distinguish each of the nine species. The *rpsK* and the *tuf1* schemes slightly



**Fig. 1** Location of amplicon targets for the three schemes to determine staphylococcal populations. The amplified region of the *rpsK* and *tuf* genes for the three schemes is shown (*rpsK* = red, *tuf1* = green, *tuf2* = blue)

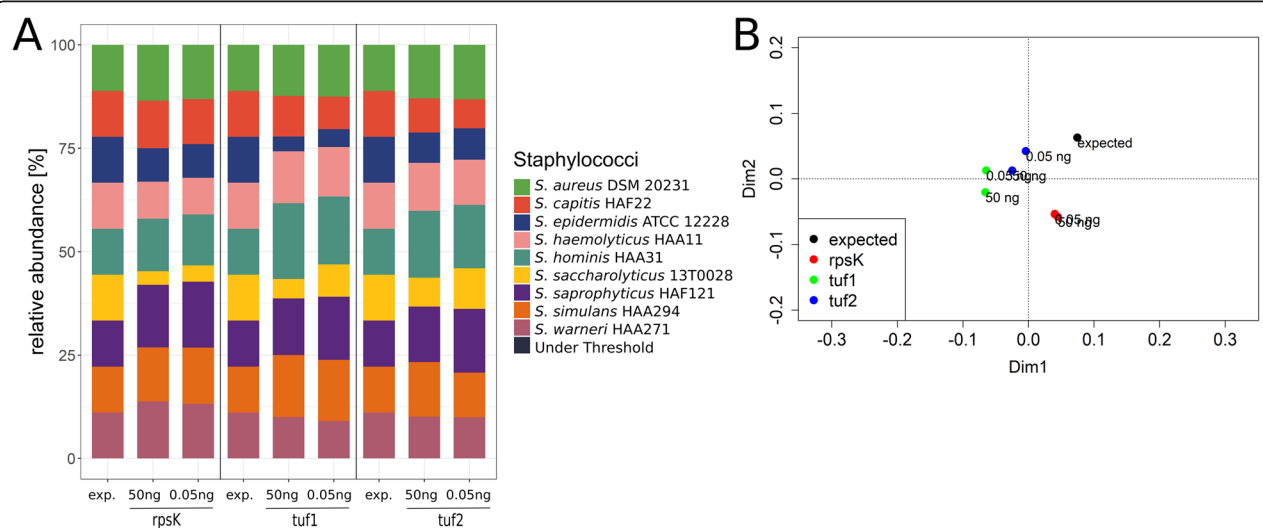
underrepresented *S. saccharolyticus* and *S. epidermidis*, respectively (Fig. 2A). A principal coordinate analysis (PCoA) plot of Bray Curtis dissimilarity was constructed to examine how accurate each scheme can represent the expected staphylococcal composition of mock community M1. The *tuf2* scheme represented the expected sample composition more accurately than the other two schemes. We repeated the experiment with different DNA input amounts, varying from 0.05 ng to 50 ng DNA per strain. The DNA input amount did only mildly influence the detected relative abundancies by the three schemes (Fig. 2 A and B).

The second mock community M2 was composed of 18 strains to investigate whether the schemes were able to resolve diversity of samples beyond the species level. The mock community M2 included genomic DNA of

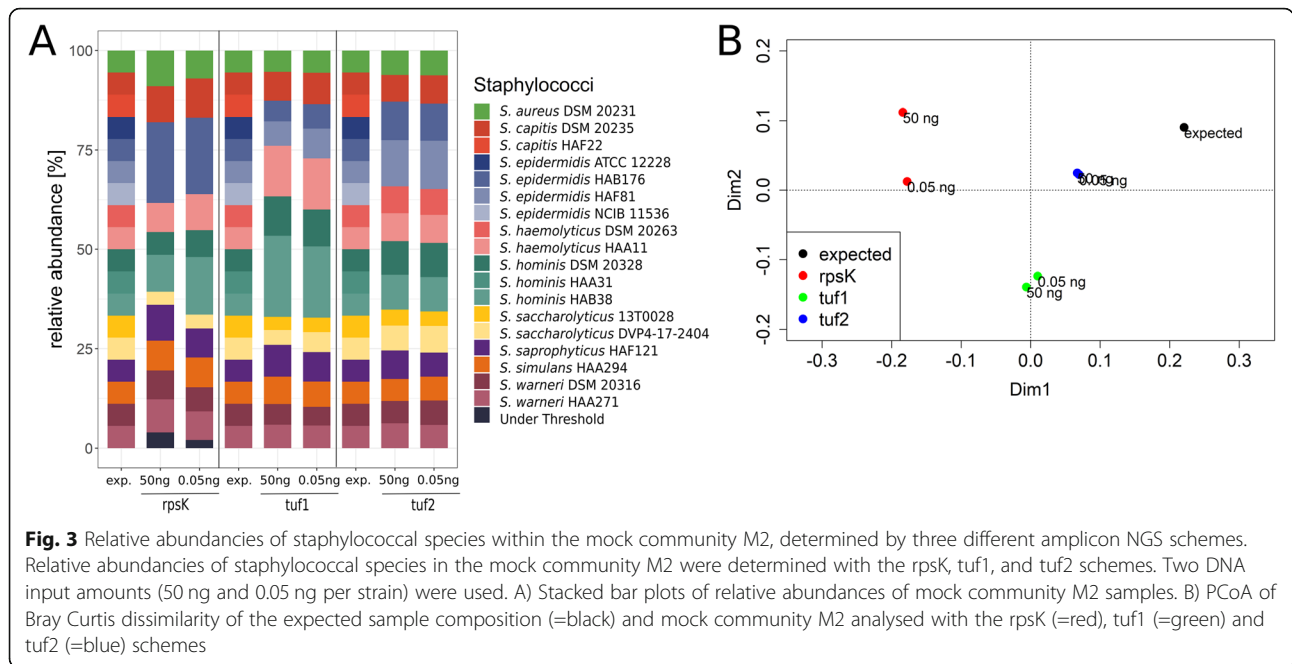
one *S. aureus* strain, two *S. capitis* strains, four *S. epidermidis* strains, two *S. haemolyticus* strains, three *S. hominis* strains, two *S. saccharolyticus* strains, one *S. saprophyticus* strain, one *S. simulans* strain and two *S. warneri* strains (Additional file 1).

First, we calculated the theoretical resolution power of each scheme regarding the M2 community. To do so, we extracted the three target alleles of each of the 18 genomes present in M2 and built phylogenetic trees. The trees showed that the *rpsK*, *tuf1* and *tuf2* schemes should distinguish 12, 14 and 15 strains, respectively (Additional file 4). Thus, in silico, the *tuf2* scheme is superior in resolution power.

Next, we applied the three schemes to analyse the DNA cocktail of the mock community M2. The *rpsK* scheme detected 11 strains, while the *tuf1* and *tuf2*

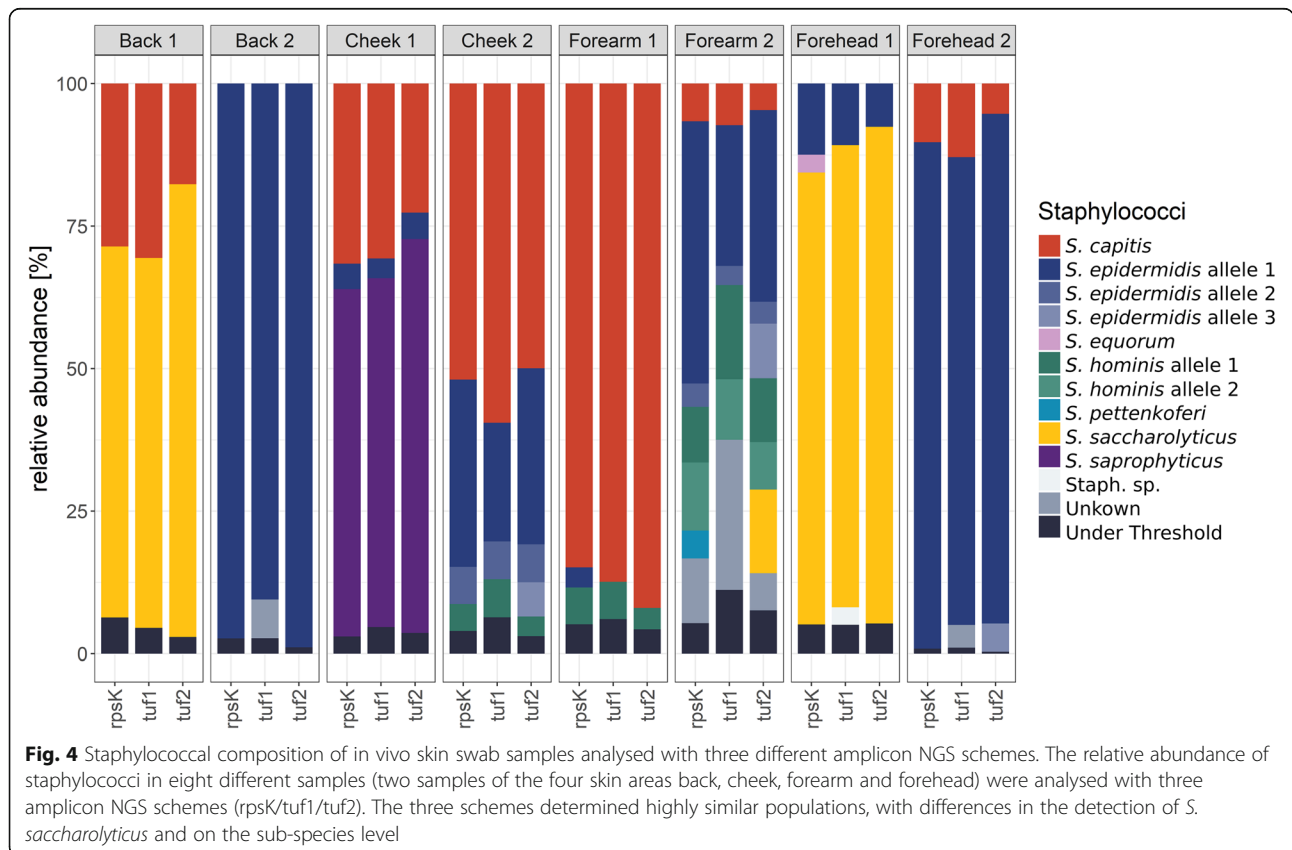


**Fig. 2** Relative abundancies of staphylococcal species within the mock community M1, determined by three different amplicon NGS schemes. Relative abundancies of staphylococcal species in the mock community M1 were determined with the *rpsK*, *tuf1*, and *tuf2* schemes. Two DNA input amounts (50 ng and 0.05 ng per strain) were used. A) Stacked bar plots of relative abundancies of mock community M1 samples. B) PCoA of Bray Curtis dissimilarity of the expected sample composition (=black) and mock community M1 analysed with the *rpsK* (=red), *tuf1* (=green) and *tuf2* (=blue) schemes



schemes detected 13 and 14 strains, respectively (Fig. 3A). The PCoA plot from Bray-Curtis dissimilarity showed that the expected sample composition was best reflected by the data generated with the tuf2 scheme.

Furthermore, compared to the two tuf schemes, data generated with the rpsK scheme showed a higher divergence when different DNA input amounts were used (Fig. 3 A and 3B).



### Amplicon NGS of human skin swab samples

Next, we applied the three schemes for the determination of staphylococcal populations in vivo. Eight skin swab samples from eight different volunteers were randomly selected. These eight samples included two samples from each of the four skin sites investigated (forehead, cheek, forearm, back).

Overall, all three schemes detected a similar staphylococcal species composition in the analysed skin swab samples (Fig. 4). The most prominent differences were seen in one of the forearm samples ("forearm 2"); here, the *tuf2* scheme detected *S. saccharolyticus*, whereas the *rpsK* and *tuf1* schemes did not. Moreover, small differences were noted: first, the *rpsK* scheme detected *S. epidermidis* and *Staphylococcus pettenkoferi* in forearm samples, and *Staphylococcus equorum* in a forehead sample ("forehead 1"), all of which with low relative abundances; in contrast, these three species were not detected by the two *tuf* schemes. Second, the *tuf2* scheme detected three different *S. epidermidis* alleles in one cheek ("cheek 2") and one forearm sample ("forearm 2"). In addition, in one forehead sample ("forehead 2"), the *tuf2* scheme found two different *S. epidermidis* alleles. In contrast, the other two schemes detected one *S. epidermidis* allele less in each sample.

### In silico comparison of *S. epidermidis* amplicon targets

We observed that three different alleles of *S. epidermidis* were detected with the *tuf2* scheme (Fig. 4). Since *S. epidermidis* is the most dominating staphylococcal species on human skin and its phylogenetic diversification into three clades (designated A, B and C) is well reported, we performed a detailed analysis of the resolution power of the three amplicon NGS schemes regarding *S. epidermidis*. A phylogenetic tree based on the core genome of 308 publicly available *S. epidermidis* genomes (Additional file 5) was built and analysed regarding the question whether the amplicon schemes can mirror the population structure. The data showed that the *rpsK* scheme could unambiguously identify B clade strains of *S. epidermidis*, but it could not distinguish strains of the A and C clades (Fig. 5). The *tuf1* and *tuf2* schemes could both unambiguously identify most strains regarding their assignment to the phylogenetic clades A, B and C. However, the *tuf1* scheme could distinguish less A and B clade strains compared to the *tuf2* scheme. Overall, the *tuf2* scheme was superior in resolution power.

### Discussion

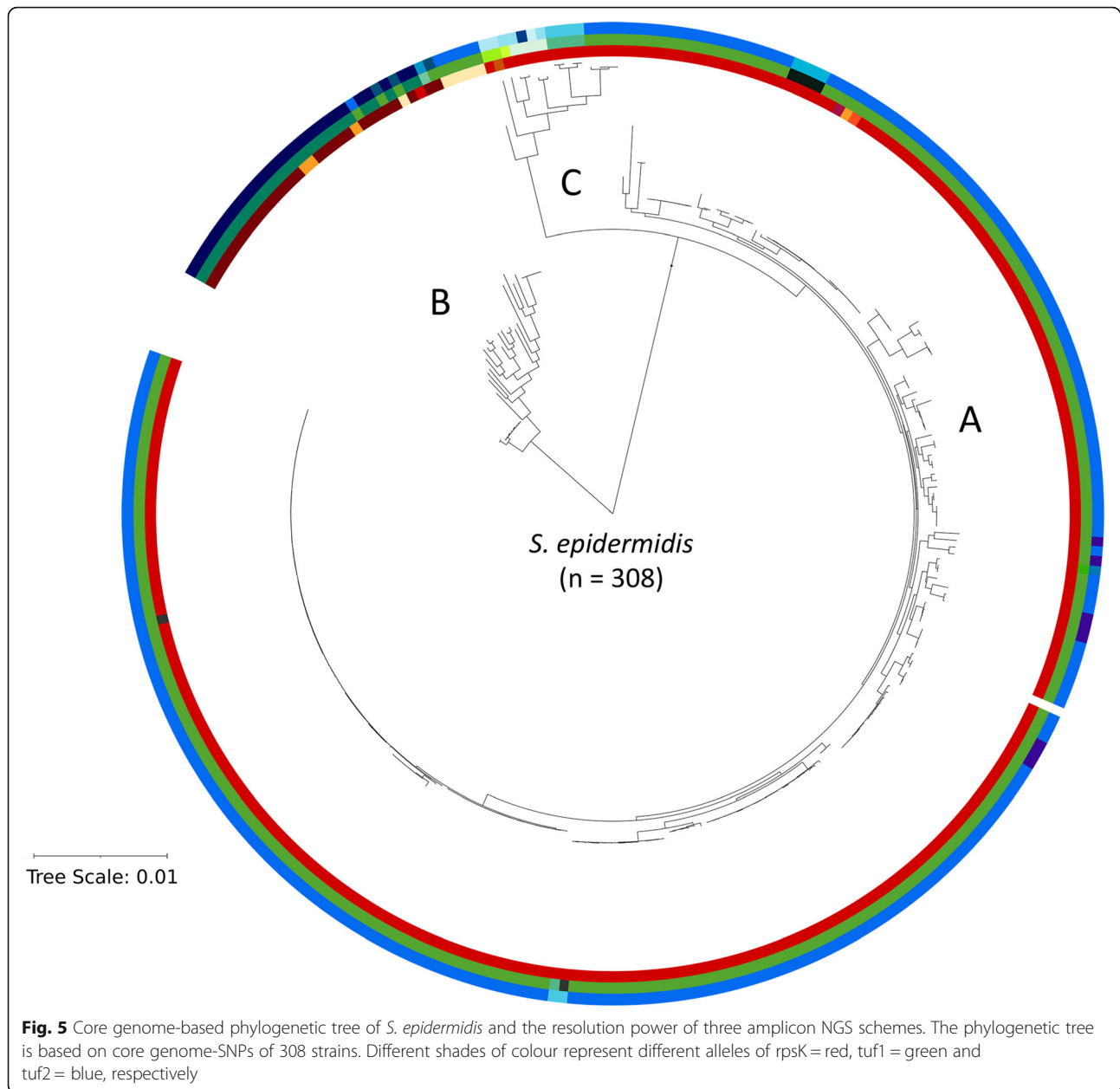
Staphylococcal populations are an important part of the human skin microbiome and play a central role in health and disease, in a species- and often also in a strain-dependent manner [1–14]. There is a need for efficient tools to determine and discriminate staphylococcal

populations on human skin, ideally beyond the species level. Here, we compared three culture-independent amplicon NGS schemes to investigate their suitability and accuracy for analysing staphylococcal populations in human skin samples.

The *tuf1* gene fragment was first used for the identification of staphylococcal isolates to the species level [41]. Later, this *tuf1* target sequence was applied in an amplicon NGS approach for determining staphylococcal populations of pig skin and pig noses [42]. A modified scheme, relying on a different *tuf* gene fragment (*tuf2*) was recently developed [47]. In the latter study, samples from healthy skin were analysed, and a surprising finding was the identification of the species *S. saccharolyticus* in relatively high quantities, which was not seen before in other amplicon NGS studies or in culture-dependent studies. In the third scheme analysed here, a *rpsK* gene fragment was used, derived from a prediction with a bioinformatics pipeline [46]. The scheme was used for determining staphylococcal populations of atopic dermatitis affected skin versus healthy skin.

Overall, we could show that all three amplicon NGS schemes accurately identified staphylococcal populations of mock communities as well as of in vivo skin swab samples. All three methods performed comparably well, regardless of DNA input amounts. However, a few differences were detected regarding the detection of *S. saccharolyticus* and different alleles of *S. epidermidis*. First, on the species level, two schemes, the *rpsK* and *tuf1* scheme, were unable to detect *S. saccharolyticus* in all samples, while the *tuf2* scheme was able to detect this species. This could be due to a mismatch in the primers to amplify *rpsK* and *tuf1*: both reverse primers have one mismatch with the corresponding regions in the genome sequence of *S. saccharolyticus* DVP4-17-2404 and 13 T0028 (data not shown). Second, the *tuf1* scheme had problems to accurately detect *S. epidermidis* in both mock communities. A reason could be a mismatch of the reverse *tuf1* primer with the corresponding region in the genome sequences of the *S. epidermidis* strains included in mock community M2. Such primer mismatches can lead to an amplification bias and thus an underestimation of the corresponding target [48, 49]. Another reason for the higher resolution power of the *tuf2* scheme can also be the amplicon size: it is longer (467 bp), compared to the other two schemes (*rpsK* = 381 bp; *tuf1* = 366 bp). This leads to a better resolution due to a higher number of SNPs in the amplicon, compared to the *tuf1* and *rpsK* amplicons. This is likely the reason why the *tuf2* scheme could distinguish more strains in mock community M2 and was able to differentiate between more *S. epidermidis* alleles in in vivo skin samples.





On the other hand, the longer *tuf2* amplicon length can create a potential problem, since it leads to fewer usable paired-end reads after quality processing, due to a low sequence quality of reads at the 3' ends with the applied Illumina MiSeq sequencing approach. The analysis of the sequence data obtained with the *rpsK* and *tuf1* schemes showed that a similar portion of paired-end reads passed the quality processing, in average 21.7 and 21.0% respectively (Additional file 6). The proportion of passed reads for the *tuf2* scheme was lower compared to *rpsK* and *tuf1* schemes, with an average of 11.6%. This did not seem to have an impact on the quality of results generated in this study, since the mock community

populations were best determined by the *tuf2* scheme, but could lead to problems when the input DNA concentration of in vivo samples is extremely low.

We further investigated whether the three amplicon NGS schemes can not only detect overall staphylococcal populations on species level, but also whether they can differentiate subspecies and phylogenetic clades of CoNS. Exemplarily, we focused on *S. epidermidis*, because of its abundance on skin and the extensive knowledge about its population structure [24–26]. Previous studies have shown that the population of *S. epidermidis* can be divided in three main clades (A, B and C) [24–26] and that each individual and each skin site is

colonised by multiple founder lineages of *S. epidermidis* of different phylogenetic clades [50]. In Espadinha et al. [25] it was shown that *S. epidermidis* strains from the A and C clades were more often associated with hospital-infections, while the B clade was mainly associated with commensal *S. epidermidis* strains. We analysed in silico if the schemes can give an accurate picture of the genetic diversity regarding the three phylogenetic clades of *S. epidermidis*. The *tuf2* scheme was superior in differentiating the three main clades of *S. epidermidis*, probably due to a longer amplicon sequence and thus higher resolution power. Thus, the *tuf2* scheme is able to analyse the presence of each *S. epidermidis* clade and could show which clade is mainly present e.g. in skin samples. However, all schemes including the *tuf2* schemes have only limited powers to resolve the population structure of *S. epidermidis*. A species-specific scheme would need to be employed, such as the duplex-amplicon NGS scheme developed by Rendboe et al. [51] to more accurately resolve the diversity of *S. epidermidis*.

Besides the three amplicon NGS schemes analysed here, other studies have used different *tuf* gene fragments for the analysis of staphylococcal populations [43–45]. For example, the *tuf* gene fragment used by McMurray et al. [43] was predicted to distinguish fourteen out of eighteen strains of mock community M2 (data not shown). Two additional *tuf* schemes were developed in the last year [44, 45]; these schemes were predicted to distinguish both 13 strains of mock community M2 (data not shown). Thus, these additional three schemes based on *tuf* gene fragments were predicted to distinguish fewer strains of the mock community M2 than the *tuf2* scheme.

The overall limitation of the amplicon NGS schemes analysed here is that they cannot resolve staphylococcal populations to the strain level. This could be achieved by shotgun metagenomic sequencing; however, it relies on a sufficiently high DNA input amount and high sequencing depth and is thus still expensive.

## Conclusion

All three schemes included in this study performed well when analysing staphylococcal populations in mock communities as well as in skin swab samples. However, the *tuf2* amplicon NGS scheme determined the expected sample composition best; it could distinguish between more *S. epidermidis* alleles in in vivo samples and detected *S. saccharolyticus* most reliably.

## Methods

### Participants and skin swabbing

Skin swab samples with moistened cotton tips were taken from 13 volunteers (female,  $n = 5$ ; male,  $n = 8$ ) with an age range of 22–43 years from forehead, cheek,

back and forearm, as described previously [47]. None of the volunteers had a history of skin disease; none had undergone treatment with topical medicine or antibiotics during the last 6 months. Written informed consent was obtained from all volunteers and the study was approved by the International Medical & Dental Ethics Commission GmbH (IMDEC), Freiburg (Study no. 67885).

### Cultivation of swab sample and species identification

Skin swab samples obtained were diluted in 0.9% NaCl solution and plated out on Columbia agar with 5% sheep blood and cultivated at 37 °C for 24 h. Up to five colonies that resembled staphylococci based on colony size and colour were randomly picked of each plate and pure cultures were obtained by sub-cultivation on the same agar, cultivated at 37 °C for 24 h. Each isolate (254 in total) was assigned to species level by MALDI-TOF mass spectrometry (Additional file 2).

### DNA extraction from skin swab samples

Eight skin swab samples were randomly selected for DNA extraction. Prior to DNA extraction, skin swab samples were centrifuged, and the supernatant was discarded. The pellets were lysed by using lysostaphin (0.05 mg/mL, Sigma) and lysozyme (9.5 mg/mL, Sigma). DNA was extracted using the DNeasy PowerSoil Kit (QIAGEN), following the manufacturer's instructions. DNA concentrations were measured with the Qubit dsDNA HS Assay (ThermoFisher Scientific) at a Qubit fluorometer.

### Whole genome sequencing

Bacterial isolates were grown on Columbia agar with 5% sheep blood for 24 h at 37 °C. Bacteria were harvested and lysed with lysostaphin (0.05 mg/mL, Sigma). DNA extraction was performed using DNeasy UltraClean Microbial Kit by following manufacturer's instructions. DNA concentration and purity were measured by Nanodrop. DNA integrity was examined with Genomic DNA ScreenTape (Agilent) at the 4200 TapeStation System. The extracted bacterial DNA was used to generate Illumina shotgun libraries; they were prepared using the Nextera XT DNA Sample Preparation Kit and subsequently sequenced on a MiSeq system using the v3 reagent kit with 600 cycles (Illumina, San Diego, CA, USA) as recommended by the manufacturer. Quality filtering was done with version 0.36 of Trimmomatic [52]. Assembly was performed with version 3.13.0 of the SPAdes genome assembler software [53]. Version 2.2.1 of Quali-map [54] was used to validate the assembly and determine the sequence coverage. Additional file 3 contains information regarding the sequencing and genome

**Table 1** Primer sets used in this study (without adapter sequences)

Primer pair name	Amplicon position <sup>a</sup>	Target gene		5' --> 3'	Amplicon length [bp]	reference
rpsK	2–382	rpsK	fw	TGGCACGTAACAAGTATC	381	[46]
			rev	GACGACGTTTTGGTGGAC		
tuf1	688–1053	tuf	fw	GGCCGTGTTGAACGTGGTCAAATC	366	[41, 42]
			rev	TIACCATTTCAGTACCTTCTGGTAA		
tuf2	685–1151	tuf	fw	ACAGGCCGTGTTGAACGTG	467	[47]
			rev	ACAGTACGTCCACCTTCACG		

<sup>a</sup> Amplicon position in genes (*tuf/rpsK*) of *S. epidermidis* ATCC 12228 (GCA\_000007645.1)

statistics, i.e. coverage, contig number, N50 and GenBank accession numbers.

### Bacterial mock communities

Genomic DNA of 18 strains was used to build two mock communities. The DNA was combined in equimolar ratios, with 0.05 ng or 50 ng DNA of each strain. The following bacterial strains were used for two mock communities M1 and M2 (Additional file 1).

The nine-strain-community (M1) contained: *S. epidermidis* ATCC 12228, *S. hominis* HAA31, *S. capitis* HAF22, *S. aureus* DSM 20231, *S. warneri* HAA271, *S. haemolyticus* HAA11, *S. saprophyticus* HAF121, *S. simulans* HAA294 and *S. saccharolyticus* 13 T0028. The 18-strain mixture (M2) contained the same strains as above plus the additional strains: *S. epidermidis* NCIB 11536, *S. epidermidis* HAF81, *S. epidermidis* HAB176, *S. hominis* DSM 20328, *S. hominis* HAB38, *S. capitis* DSM 20325, *S. warneri* DSM 20316, *S. haemolyticus* DSM 20263 and *S. saccharolyticus* DVP4-17-2404. GenBank accession numbers of all genomes of the listed strains are given in Additional file 1.

### Amplicon polymerase chain reaction (PCR) and sequencing

The target fragments, designated *tuf1* [41], *tuf2* [47], and *rpsK* [46], were amplified using specific primer sets (Table 1). PCR reaction mixtures were made in a total volume of 25 µl and comprised 5 µl of DNA sample, 2.5 µl AccuPrime PCR Buffer II (Invitrogen, Waltham, MA, USA), 1.5 µl of each primer (10 µM) (DNA Technology, Risskov, Denmark), 0.15 µl AccuPrime Taq DNA Polymerase High Fidelity (Invitrogen, Waltham, MA, USA), and 14.35 µl of PCR grade water. The PCR reaction was performed using the following cycle conditions: an initial denaturation at 94 °C for 2 min, followed by 35 cycles of denaturation at 94 °C for 20 s, annealing at 55 °C for 30 s, elongation at 68 °C for 1 min, and a final elongation step at 72 °C for 5 min. PCR products were verified on an agarose gel and purified using the Qiagen Genscreen™ Size Selection kit (Qiagen, Hilden, Germany). The concentration of the purified PCR products was

measured with a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). Amplicon NGS was performed as described previously [47].

### Amplicon NGS data analysis and visualization

FASTQ sequences were processed using QIIME2 (v. 2019.7) [55] as described previously [47]. A cut-off of 99% identity against *tuf* and *rpsK* gene databases was used. The database was built based on all closed staphylococcal genomes available in GenBank (status 02/01/2021). Mock community samples were analysed with a database which contained the *rpsK/tuf1/tuf2* allele for each strain. Data was normalized, low abundant features were filtered with a threshold of 2.5%, and figures were prepared in R (v. 4.0.1) with the packages ggplot2 [56] and gplots [57]. Bray-Curtis dissimilarity of mock community sample data was calculated in the vegan package [58] and ordinated in a principal coordinate analysis (PCoA).

### Phylogenomic amplicon target analysis

For phylogenomic analyses, all closed and scaffold genome sequence data of *S. epidermidis* was obtained from NCBI RefSeq (status 07.01.2021). GenBank accession numbers of all used genomes are given in Additional file 5. Genomes were aligned and clustered based on SNPs in their core genome using Parsnp (v 1.0) [59]. The different amplicon alleles were identified for each strain using Blast+ (v 2.11.0). Visualization of the tree was done with iTOL (v 5.7).

### Abbreviations

CoNS: coagulase-negative staphylococci; EF-Tu: elongation factor Tu; MALDI-TOF: Matrix-Assisted Laser Desorption – Ionization - Time of Flight; NGS: Next-generation sequencing; PCoA: principal coordinate analysis; PCR: polymerase chain reaction; SNP: single nucleotide polymorphism; SRA: Sequence Read Archive

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12866-021-02284-1>.

**Additional file 1.** Table S1: Composition of two staphylococcal mock communities.



**Additional file 2.** Table S2: Cultivation results of skin swabs from 13 volunteers.

**Additional file 3.** Table S3: Genome statistics of sequenced staphylococcal genomes.

**Additional file 4.** Fig. S1: Phylogenetic trees of the alleles of three amplicon targets (*tuf1*, *tuf2* and *rpsK*), extracted from 18 staphylococcal genomes present in the M2 mock community.

**Additional file 5** Table S4: Accession numbers of *S. epidermidis* genomes included in the phylogenomic analysis.

**Additional file 6.** Table S5: Amplicon sequencing statistics of the mock communities.

### Acknowledgements

We thank Lise Hald Schultz for excellent technical assistance. Data processing was performed on the GenomeDK cluster; we would like to thank GenomeDK and Aarhus University for providing computational resources that contributed to these research results.

### Authors' contributions

Conceptualization, CA, JH, WRS, HB; methodology and experiments, CA, AP; data analysis, CA, KS, AP, HB; writing—original draft preparation, CA, HB; writing—review and editing, KS, AP, WRS, JH; supervision, JH, WRS, HB; project administration, JH, WRS, HB; funding acquisition, HB. All authors have read and agreed to the published version of the manuscript.

### Funding

This work was supported by grants from the NovoNordisk Foundation (grant no. NNF18OC0053172).

### Availability of data and materials

Whole genome sequences generated for this project are available in the NCBI BioProject database under the project ID PRJNA702288, and NCBI GenBank accession numbers are listed in Additional file 3. Whole genome sequences of strains used in the mock communities were obtained from NCBI GenBank, and all NCBI GenBank accession numbers are listed in Additional file 1. Amplicon sequencing data are available in the NCBI Sequence Read Archive (SRA) under the project ID PRJNA702649. Whole genome sequences of *S. epidermidis* used for the phylogenomic analysis were obtained from NCBI RefSeq, and NCBI RefSeq accession numbers are listed in Additional file 5.

### Declarations

#### Ethics approval and consent to participate

Written informed consent was obtained from all volunteers and the study was approved by the International Medical & Dental Ethics Commission GmbH (IMDEC), Freiburg (Study no. 67885).

#### Consent for publication

Not applicable.

#### Competing interests

Conflicts of Interest: CA and JH are employees at Beiersdorf AG. The other authors declare no conflict of interest.

#### Author details

<sup>1</sup>Beiersdorf AG, Research & Development, Front End Innovation, 20245 Hamburg, Germany. <sup>2</sup>Department of Microbiology and Biotechnology, University of Hamburg, 22609 Hamburg, Germany. <sup>3</sup>Department of Biomedicine, Aarhus University, 8000 Aarhus, Denmark. <sup>4</sup>Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, University of Göttingen, 37073 Göttingen, Germany.

Received: 25 March 2021 Accepted: 21 July 2021

Published online: 28 July 2021

### References

- Bunikowski R, Mielke ME, Skarabis H, Worm M, Anagnostopoulos I, Kolde G, et al. Evidence for a disease-promoting effect of *Staphylococcus aureus*-

- derived exotoxins in atopic dermatitis. *J Allergy Clin Immunol.* 2000;105(4): 814–9. <https://doi.org/10.1067/mai.2000.105528>.
- Nakatsuji T, Chen TH, Narala S, Chun KA, Two AM, Yun T, et al. Antimicrobials from human skin commensal bacteria protect against *Staphylococcus aureus* and are deficient in atopic dermatitis. *Sci Transl Med.* 2017;9(378):eaah4680. <https://doi.org/10.1126/scitranslmed.aah4680>.
- Schnell N, Entian KD, Schneider U, Gotz F, Zahner H, Kellner R, et al. Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. *Nature.* 1988;333(6170):276–8. <https://doi.org/10.1038/333276a0>.
- Sandiford S, Upton M. Identification, characterization, and recombinant expression of epidermin NI01, a novel unmodified bacteriocin produced by *Staphylococcus epidermidis* that displays potent activity against staphylococci. *Antimicrob Agents Chemother.* 2012;56(3):1539–47. <https://doi.org/10.1128/AAC.05397-11>.
- O'Neill AM, Nakatsuji T, Hayachi A, Williams MR, Mills RH, Gonzalez DJ, et al. Identification of a human skin commensal bacterium that selectively kills *Cutibacterium acnes*. *J Invest Dermatol.* 2020.
- Brown MM, Kwiecinski JM, Cruz LM, Shahbandi A, Todd DA, Cech NB, et al. Novel peptide from commensal *Staphylococcus simulans* blocks methicillin-resistant *Staphylococcus aureus* quorum sensing and protects host skin from damage. *Antimicrob Agents Chemother.* 2020;64(6). <https://doi.org/10.1128/AAC.00172-20>.
- Peng P, Baldry M, Gless BH, Bojer MS, Espinosa-Gongora C, Baig SJ, et al. Effect of Co-inhabiting Coagulase Negative Staphylococci on *S. aureus* agr Quorum Sensing, Host Factor Binding, and Biofilm Formation. *Front Microbiol.* 2019;10:2212. <https://doi.org/10.3389/fmicb.2019.02212>.
- Scharschmidt TC, Vasquez KS, Truong HA, Gearty SV, Pauli ML, Nosbaum A, et al. A wave of regulatory T cells into neonatal skin mediates tolerance to commensal microbes. *Immunity.* 2015;43(5):1011–21. <https://doi.org/10.1016/j.immuni.2015.10.016>.
- Naik S, Bouladoux N, Linehan JL, Han SJ, Harrison OJ, Wilhelm C, et al. Commensal-dendritic-cell interaction specifies a unique protective skin immune signature. *Nature.* 2015;520(7545):104–8. <https://doi.org/10.1038/nature14052>.
- Stacy A, Belkaid Y. Microbial guardians of skin health. *Science.* 2019; 363(6424):227–8. <https://doi.org/10.1126/science.aat4326>.
- Parlet CP, Brown MM, Horswill AR. Commensal staphylococci influence *Staphylococcus aureus* skin colonization and disease. *Trends Microbiol.* 2019;27(6):497–507. <https://doi.org/10.1016/j.tim.2019.01.008>.
- Leonel C, Sena IFG, Silva WN, Prazeres P, Fernandes GR, Mancha Agresti P, et al. *Staphylococcus epidermidis* role in the skin microenvironment. *J Cell Mol Med.* 2019;23(9):5949–55. <https://doi.org/10.1111/jcmm.14415>.
- Lai Y, Di Nardo A, Nakatsuji T, Leichtle A, Yang Y, Cogen AL, et al. Commensal bacteria regulate toll-like receptor 3-dependent inflammation after skin injury. *Nat Med.* 2009;15(12):1377–82. <https://doi.org/10.1038/nm.2062>.
- Nakatsuji T, Chen TH, Butcher AM, Trzoss LL, Nam SJ, Shirakawa KT, et al. A commensal strain of *Staphylococcus epidermidis* protects against skin neoplasia. *Sci Adv.* 2018;4:eaao4502.
- Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, et al. Topographical and temporal diversity of the human skin microbiome. *Science.* 2009;324(5931):1190–2. <https://doi.org/10.1126/science.1171700>.
- Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. *Nat Rev Microbiol.* 2018;16(3):143–55. <https://doi.org/10.1038/nrmicro.2017.157>.
- Chen YE, Fischbach MA, Belkaid Y. Skin microbiota-host interactions. *Nature.* 2018;553(7689):427–36. <https://doi.org/10.1038/nature25177>.
- Schleiferi KHK, W. E. Isolation and characterization of staphylococci from human skin I. amended descriptions of *Staphylococcus epidermidis* and *Staphylococcus saprophyticus* and descriptions of three new species: *Staphylococcus cohnii*, *Staphylococcus haemolyticus*, and *Staphylococcus xylosus*. *Int J Syst Evol Microbiol.* 1975;25:50–61.
- Kloss WES, K. H. Isolation and characterization of staphylococci from human skin II. Descriptions of four new species: *Staphylococcus warneri*, *Staphylococcus capitis*, *Staphylococcus hominis*, and *Staphylococcus simulans*. *Int J Syst Bacteriol.* 1975;25(1):62–79. <https://doi.org/10.1099/00207713-25-1-62>.
- Meric G, Mageiros L, Pensar J, Laabei M, Yahara K, Pascoe B, et al. Disease-associated genotypes of the commensal skin bacterium *Staphylococcus epidermidis*. *Nat Commun.* 2018;9(1):5034. <https://doi.org/10.1038/s41467-018-07368-7>.
- Both A, Huang J, Qi M, Lausmann C, Weisselberg S, Buttner H, et al. Distinct clonal lineages and within-host diversification shape invasive *Staphylococcus epidermidis* populations. *PLoS Pathog.* 2021;17(2):e1009304. <https://doi.org/10.1371/journal.ppat.1009304>.

22. Zhou W, Spoto M, Hardy R, Guan C, Fleming E, Larson PJ, et al. Host-specific evolutionary and transmission dynamics shape the functional diversification of *Staphylococcus epidermidis* in human skin. *Cell*. 2020;180(3):454–70 e418. <https://doi.org/10.1016/j.cell.2020.01.006>.
23. Oh J, Byrd AL, Deming C, Conlan S, Program NCS, Kong HH, et al. Biogeography and individuality shape function in the human skin metagenome. *Nature*. 2014;514(7520):59–64. <https://doi.org/10.1038/nature13786>.
24. Conlan S, Mijares LA, Program NCS, Becker J, Blakesley RW, Bouffard GG, et al. *Staphylococcus epidermidis* pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. *Genome Biol*. 2012;13(7):R64. <https://doi.org/10.1186/gb-2012-13-7-r64>.
25. Espadinha D, Sobral RG, Mendes CI, Meric G, Sheppard SK, Carrico JA, et al. Distinct phenotypic and genomic signatures underlie contrasting pathogenic potential of *Staphylococcus epidermidis* clonal lineages. *Front Microbiol*. 2019;10:1971. <https://doi.org/10.3389/fmicb.2019.01971>.
26. Meric G, Miragaia M, de Been M, Yahara K, Pascoe B, Mageiros L, et al. Ecological overlap and horizontal gene transfer in *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Genome Biol Evol*. 2015;7(5):1313–28. <https://doi.org/10.1093/gbe/ew066>.
27. Miragaia M, Thomas JC, Couto I, Enright MC, de Lencastre H. Inferring a population structure for *Staphylococcus epidermidis* from multilocus sequence typing data. *J Bacteriol*. 2007;189(6):2540–52. <https://doi.org/10.1128/JB.01484-06>.
28. Kloos WE, Musselwhite MS. Distribution and persistence of *Staphylococcus* and *Micrococcus* species and other aerobic bacteria on human skin. *Appl Microbiol*. 1975;30(3):381–5. <https://doi.org/10.1128/am.30.3.381-395.1975>.
29. Kloos WE, Schleifer KH. Isolation and characterization of staphylococci from human skin. *Int J Syst Bacteriol*. 1975;25(1):62–79. <https://doi.org/10.1099/00207713-25-1-62>.
30. Ward DM, Weller L, Bateson MM. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature*. 1990;345(6270):63–5. <https://doi.org/10.1038/345063a0>.
31. Ghebremedhin B, Layer F, König W, König B. Genetic classification and distinguishing of *Staphylococcus* species based on different partial gap, 16S rRNA, hsp60, rpoB, sodA, and tuf gene sequences. *J Clin Microbiol*. 2008;46(3):1019–25. <https://doi.org/10.1128/JCM.02058-07>.
32. Meisel JS, Hannigan GD, Tyldsley AS, SanMiguel AJ, Hodkinson BP, Zheng Q, et al. Skin microbiome surveys are strongly influenced by experimental design. *J Invest Dermatol*. 2016;136(5):947–56. <https://doi.org/10.1016/j.jid.2016.01.016>.
33. Blaiotta G, Fusco V, Ercolini D, Pepe O, Coppola S. Diversity of *Staphylococcus* species strains based on partial kat (catalase) gene sequences and design of a PCR-restriction fragment length polymorphism assay for identification and differentiation of coagulase-positive species (*S. aureus*, *S. delphini*, *S. hyicus*, *S. intermedius*, *S. pseudintermedius*, and *S. schleiferi* subsp. *coagulans*). *J Clin Microbiol*. 2010;48(1):192–201. <https://doi.org/10.1128/JCM.00542-09>.
34. Yugueros J, Temprano A, Berzal B, Sanchez M, Hernanz C, Luengo JM, et al. Glycerinaldehyde-3-phosphate dehydrogenase-encoding gene as a useful taxonomic tool for *Staphylococcus* spp. *J Clin Microbiol*. 2000;38(12):4351–5. <https://doi.org/10.1128/JCM.38.12.4351-4355.2000>.
35. Yugueros J, Temprano A, Sanchez M, Luengo JM, Naharro G. Identification of *Staphylococcus* spp. by PCR-restriction fragment length polymorphism of gap gene. *J Clin Microbiol*. 2001;39(10):3693–5. <https://doi.org/10.1128/JCM.39.10.3693-3695.2001>.
36. Goh SH, Potter S, Wood JO, Hemmingsen SM, Reynolds RP, Chow AW. HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. *J Clin Microbiol*. 1996;34(4):818–23. <https://doi.org/10.1128/jcm.34.4.818-823.1996>.
37. Goh SH, Santucci Z, Kloos WE, Falynn M, George CG, Driedger D, et al. Identification of *Staphylococcus* species and subspecies by the chaperonin 60 gene identification method and reverse checkerboard hybridization. *J Clin Microbiol*. 1997;35(12):3116–21. <https://doi.org/10.1128/jcm.35.12.3116-3121.1997>.
38. Mollet C, Drancourt M, Raoult D. rpoB sequence analysis as a novel basis for bacterial identification. *Mol Microbiol*. 1997;26(5):1005–11. <https://doi.org/10.1046/j.1365-2958.1997.6382009.x>.
39. Drancourt M, Raoult D. rpoB gene sequence-based identification of *Staphylococcus* species. *J Clin Microbiol*. 2002;40(4):1333–8. <https://doi.org/10.1128/JCM.40.4.1333-1338.2002>.
40. Poyart C, Quesne G, Boumaila C, Trieu-Cuot P. Rapid and accurate species-level identification of coagulase-negative staphylococci by using the sodA gene as a target. *J Clin Microbiol*. 2001;39(12):4296–301. <https://doi.org/10.1128/JCM.39.12.4296-4301.2001>.
41. Martineau F, Picard FJ, Ke D, Paradis S, Roy PH, Ouellette M, et al. Development of a PCR assay for identification of staphylococci at genus and species levels. *J Clin Microbiol*. 2001;39(7):2541–7. <https://doi.org/10.1128/JCM.39.7.2541-2547.2001>.
42. Strube ML, Hansen JE, Rasmussen S, Pedersen K. A detailed investigation of the porcine skin and nose microbiome using universal and *Staphylococcus* specific primers. *Sci Rep*. 2018;8(1):12751. <https://doi.org/10.1038/s41598-018-30689-y>.
43. McMurray CL, Hardy KJ, Calus ST, Loman NJ, Hawkey PM. *Staphylococcal* species heterogeneity in the nasal microbiome following antibiotic prophylaxis revealed by tuf gene deep sequencing. *Microbiome*. 2016;4(1):63. <https://doi.org/10.1186/s40168-016-0210-1>.
44. Van Reckem E, De Vuyst L, Leroy F, Weckx S. Amplicon-based high-throughput sequencing method capable of species-level identification of coagulase-negative staphylococci in diverse communities. *Microorganisms*. 2020;8.
45. Iversen S, Johannesen TB, Ingham AC, Edslev SM, Tevell S, Mansson E, et al. Alteration of bacterial communities in anterior nares and skin sites of patients undergoing arthroplasty surgery: analysis by 16S rRNA and staphylococcal-specific tuf gene sequencing. *Microorganisms*. 2020;8(12). <https://doi.org/10.3390/microorganisms8121977>.
46. Ederveen THA, Smits JPH, Hajo K, van Schalkwijk S, Kouwenhoven TA, Lukovac S, et al. A generic workflow for single locus sequence typing (SLST) design and subspecies characterization of microbiota. *Sci Rep*. 2019;9(1):19834. <https://doi.org/10.1038/s41598-019-56065-y>.
47. Ahle CM, Stodkilde K, Afshar M, Poehlein A, Ogilvie LA, Soderquist B, et al. *Staphylococcus saccharolyticus*: an overlooked human skin colonizer. *Microorganisms*. 2020;8(8). <https://doi.org/10.3390/microorganisms8081105>.
48. Stadhouders R, Pas SD, Anber J, Voermans J, Mes TH, Schutten M. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay. *J Mol Diagn*. 2010;12(1):109–17. <https://doi.org/10.2353/jmoldx.2010.090035>.
49. Sipoš R, Szekely AJ, Palatinszky M, Revesz S, Marialigeti K, Nikolausz M. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiol Ecol*. 2007;60(2):341–50. <https://doi.org/10.1111/j.1574-6941.2007.00283.x>.
50. Zhou H, Shi L, Ren Y, Tan X, Liu W, Liu Z. Applications of human skin microbiota in the cutaneous disorders for ecology-based therapy. *Front Cell Infect Microbiol*. 2020;10:570261. <https://doi.org/10.3389/fcimb.2020.570261>.
51. Rendboe AK, Johannesen TB, Ingham AC, Mansson E, Iversen S, Baig S, et al. The Epidome - a species-specific approach to assess the population structure and heterogeneity of *Staphylococcus epidermidis* colonization and infection. *BMC Microbiol*. 2020;20(1):362. <https://doi.org/10.1186/s12866-020-02041-w>.
52. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
53. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455–77. <https://doi.org/10.1089/cmb.2012.0021>.
54. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, et al. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics*. 2012;28(20):2678–9. <https://doi.org/10.1093/bioinformatics/bts503>.
55. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37(8):852–7. <https://doi.org/10.1038/s41587-019-0209-9>.
56. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2009.
57. Warnes GRB, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, et al. gplots: Various R Programming Tools for Plotting Data; 2020.
58. Oksanen J, Friendly M, Kindt R, Legendre P, McGinn D, Minchin PR, et al. vegan: Community Ecology Package. R package version 2.5–6; 2019.
59. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol*. 2014;15(1):524. <https://doi.org/10.1186/s13059-014-0524-x>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.