Forensic Science International: Genetics 32 (2018) 50–61

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Targeted sequencing of clade-specific markers from skin microbiomes for forensic human identification

Sarah E. Schmedes[a,b], August E. Woerner[b], Nicole M.M. Novroski[a,b], Frank R. Wendt[a,b], Jonathan L. King[b], Kathryn M. Stephens[c], Bruce Budowle[a,b,d,*]

[a] Graduate School of Biomedical Sciences, University of North Texas Health Science Center,3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA
[b] Center for Human Identification, University of North Texas Health Science Center,3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA
[c] Verogen, Inc, 11111 Flintkote Ave, San Diego, CA 92121, USA
[d] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Keywords:
Skin microbiome
Human identification
Forensic profiling
Targeted sequencing
Supervised learning

ABSTRACT

The human skin microbiome is comprised of diverse communities of bacterial, eukaryotic, and viral taxa and contributes millions of additional genes to the repertoire of human genes, affecting human metabolism and immune response. Numerous genetic and environmental factors influence the microbiome composition and as such contribute to individual-specific microbial signatures which may be exploited for forensic applications. Previous studies have demonstrated the potential to associate skin microbial profiles collected from touched items to their individual owner, mainly using unsupervised methods from samples collected over short time intervals. Those studies utilize either targeted 16S rRNA or shotgun metagenomic sequencing to characterize skin microbiomes; however, these approaches have limited species and strain resolution and susceptibility to stochastic effects, respectively. Clade-specific markers from the skin microbiome, using supervised learning, can predict individual identity using skin microbiomes from their respective donors with high accuracy. In this study the hidSkinPlex is presented, a novel targeted sequencing method using skin microbiome markers developed for human identification. The hidSkinPlex (comprised of 286 bacterial (and phage) family-, genus-, species-, and subspecies-level markers), initially was evaluated on three bacterial control samples represented in the panel (i.e., *Propionibacterium acnes*, *Propionibacterium granulosum*, and *Rothia dentocariosa*) to assess the performance of the multiplex. The hidSkinPlex was further evaluated for prediction purposes. The hidSkinPlex markers were used to attribute skin microbiomes collected from eight individuals from three body sites (i.e., foot (Fb), hand (Hp) and manubrium (Mb)) to their host donor. Supervised learning, specifically regularized multinomial logistic regression and 1-nearest-neighbor classification were used to classify skin microbiomes to their hosts with up to 92% (Fb), 96% (Mb), and 100% (Hp) accuracy. All samples (n = 72) regardless of body site origin were correctly classified with up to 94% accuracy, and body site origin could be predicted with up to 86% accuracy. Finally, human short tandem repeat and single-nucleotide polymorphism profiles were generated from skin swab extracts from a single subject to highlight the potential to use microbiome profiling in conjunction with low-biomass samples. The hidSkinPlex is a novel targeted enrichment approach to profile skin microbiomes for human forensic identification purposes and provides a method to further characterize the utility of skin microflora for human identification in future studies, such as the stability and diversity of the personal skin microbiome.

1. Introduction

Diverse microbial communities of bacterial, fungal, and viral species compose the human skin microbiome [1–3]. The skin microbiome can be influenced by several genetic and environmental factors, such as geography, health/disease states, and lifestyle (i.e., diet, hygiene, frequent contact with others, etc.) [4–8], affecting the composition of an individual's microflora. Although, a large number of skin flora are common to most individuals, overall skin microbial community profiles can vary substantially in abundance of specific microbial taxa and unique strain signatures [3,9]. Skin microbiome strain profiles can be stable over long periods of time (e.g., at least up to 3 years [3]) and thus make ideal candidates for genetically profiling microbiomes for forensic purposes.

* Corresponding author at: Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Blvd Fort Worth, TX 76107, USA.
E-mail address: Bruce.Budowle@unthsc.edu (B. Budowle).

http://dx.doi.org/10.1016/j.fsigen.2017.10.004
Received 22 August 2017; Received in revised form 12 October 2017; Accepted 16 October 2017
Available online 18 October 2017
1872-4973/

Current forensic human identification methods typically rely on targeting autosomal markers (e.g., short-tandem repeats (STRs)) to create genetic profiles to compare evidentiary items with profiles generated from a reference sample from an individual(s) [10–16]. In some cases when the evidentiary sample may be degraded or contain low amounts of DNA (i.e., low-copy number (LCN) DNA), high-copy number (HCN) markers (e.g., the mitochondrial genome [17] or hypervariable regions of the mitochondrial genome [18,19]) are targeted. Other HCN markers, such as skin microbiome genetic markers may provide additional identifying genetic information which can be used independently or potentially in conjunction with partial human forensic marker profiles. Microbial cells transfer from the skin to objects just as with human cells, and these microbial cells are likely greater in number than human cells, ~10,000 bacterial cells/cm^2 collected per skin swab [20]. The higher number of skin microbial cells than human cells and presence of individual-specific skin microbiome signatures may make skin microbiome profiling a viable approach for potential forensic applications. However, before skin microbiome profiling can be used for forensic human identification, a robust and reproducible method targeting stable, microbial polymorphic genetic markers must be established.

Previous studies have demonstrated the potential to use skin microbiome profiling for forensic applications, mainly targeting the 16S rRNA gene and using unsupervised methods to demonstrate that skin microbiome profiles from touched objects resemble their individual donors [21–23]. Supervised learning (i.e., classification) has been used in a limited capacity to classify skin microbiome samples from individuals collected at a single time point or over short time intervals [24,25]. Most studies characterizing skin microbiomes have relied on either targeted 16S rRNA sequencing or shotgun metagenomic sequencing; however, neither of these methods are ideal for forensic characterization of skin microbiomes due to limited species and strain resolution and susceptibility of stochastic effects, respectively. An alternative approach would be to use targeted sequencing of select sets of informative markers shown to provide individualizing resolution that are stable over time. A reliable method with the capability of strain-level resolution could be developed for forensic analyses and allow for sufficient coverage of informative sites, even from body sites with low-abundant taxa.

In a previous study, Schmedes et al. [26] mined a publically available dataset [3] comprised of shotgun metagenomic skin microbiomes collected from 12 individuals, 17 skin body sites, sampled at three time points over a time period of > 2.5 years to identify stable clade-specific markers. Markers were identified that provided individualizing resolution at each body site based on skin microbiome profiles generated using the nucleotide diversity (i.e., a measure of strain-level heterogeneity of the microbial population (See Methods and materials)) of each marker. Supervised learning, specifically regularized multinomial logistic regression (RMLR) and 1-nearest-neighbor classification (1NN), was used to attribute skin microbiome profiles to their individual host with high accuracy [26]. Subsets of clade-specific markers also were selected, which provide comparable classification accuracies to that of using all markers evaluated, as candidates to develop a targeted panel for skin microbiome characterization for human identification purposes [26]. Candidate markers were selected from 14/17 body sites, excluding three sites from the feet, which lacked sufficient coverage and stability for classification [26].

In this study, a novel targeted sequencing panel, the hidSkinPlex, was developed based on candidate markers from Schmedes et al. [26] for skin microbiome profiling for forensic human identification. The markers within the hidSkinPlex panel are contained in one multiplex amplification assay for targeted sequencing on the Illumina MiSeq system. Initially, the performance (i.e., sensitivity and specificity) of the hidSkinPlex was assessed using control bacterial genomic DNA from three bacterial species, *Propionibacterium acnes*, *Propionibacterium granulosum*, and *Rothia dentocariosa*. The hidSkinPlex was further evaluated using skin microbiome samples collected from three skin sites, the toe web/ball of the foot (Fb), the palm of the non-dominant hand (Hp) and the manubrium (Mb), in eight individuals. RMLR and 1NN classification were used to predict skin microbiomes originating from specific body sites with their respective donors. Attribute selection also was performed to identify subsets of hidSkinPlex markers that provide similar or greater predictive power than the entire hidSkinPlex panel for individual classification at each body site. Additionally, maximum likelihood phylogenies of *P. acnes* strains, using *P. acnes*-specific markers from the hidSkinPlex were constructed to characterize *P. acnes* strains across body sites and individuals to determine if *P. acnes* strains were more related at the level of the individual or the individual at each body site. Finally, hidSkinPlex profiles and human-specific STR and single-nucleotide polymorphism (SNP) profiles generated from the same skin samples were compared to provide a case study on the potential to use skin microbiome profiles in conjunction with human genetic profiles for forensic investigative purposes.

2. Material and methods

2.1. Sample collection

Skin microbiome samples were collected from eight individuals (four females, four males) sampled from the Mb, Hp, and Fb, according to a protocol approved by the University of North Texas Health Science Center (UNTHSC) Internal Review Board (IRB). Skin microbiome samples were collected using 4N6FLOQSwabs™: Genetics (COPAN, Brescia, Italy) pre-moistened with 30 µL sterile, molecular-grade water (Phenix, Candler, NC). All skin swabs were collected by swabbing a separate section of skin per replicate with firm pressure for 10 s on one side of the swab head, rotated 180°, and then swabbed another 10 s. Mb skin sites were collected ~5 cm beneath the junction of the clavicles. Hp samples were collected by swabbing separate sections of the palm starting at the base of a finger (excluding the thumb) and extending across the entire length of the palm. Fb samples were collected by swabbing between each toe web space and extending down the entire length of the ball of the foot. Three replicate samples were collected from each body site for a total of nine swabs collected per individual (n = 72). Each subject filled out a questionnaire to retrieve associated metadata related to the subject regarding bioancestry, hygiene, health/disease state, and recent travel. No subjects were eliminated from the study due to answers on the questionnaire. Swabs were either stored at −20 °C until DNA extraction or extracted directly.

2.2. DNA extraction and quantification

Total DNA was extracted from skin swabs collected from subjects S001-S004 using the MO BIO BiOstic® Bacteremia DNA Isolation Kit (MO BIO Laboratories, Inc. Carlsbad, CA) following the manufacturer's protocol with the following modifications. The CB1 buffer was added directly to the MicroBead tube followed by adding the swab to the buffer/bead solution and allowed to soak for 5 min with occasional rotation of the swab. Next, the swab head was snapped off, along the break point on the plastic applicator, and left in the tube proceeding to the 70 °C incubation step; the remainder of the manufacturer's protocol was followed as prescribed. A swab blank was included with each extraction. DNA extracts were stored at −20 °C. Total DNA was extracted from skin swabs collected from subjects S005-S008 using the QIAamp BiOstic Bacteremia DNA Kit (Qiagen, Hilden, Germany), the new version of the previous MO BIO kit. The same modified swab protocol was followed except the swab head could not fit in the new PowerBead tube. Instead, the swab was soaked with agitation in the MBL buffer (previously CB1) for at least 5 min followed by adding the supernatant from the swab tube directly to the PowerBead tube, proceeding to the 70 °C incubation step; the remainder of the manufacturer's protocol was followed as prescribed. Total DNA was quantified using the Qubit® 2.0

Fluorometer with the Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific, Eugene, OR).

2.3. Development of the hidSkinPlex panel and multiplex primer design

Markers included in the hidSkinPlex panel were selected by Schmedes et al. [26]. Briefly, publically-available shotgun metagenomic sequence datasets generated by Oh et al. [3] were mined to identify universal clade-specific markers (i.e., markers unique to a particular microbial taxonomic clade) that were stable over the tested time interval, which could be used to differentiate individuals based on their individual-specific skin microbiome signatures. The Oh et al. [3] data were comprised of skin microbiomes from 12 healthy individuals, 17 skin body sites, and 3 time points (sampled over a period of > 2.5 years). The nucleotide diversities of clade-specific markers, from the MetaPhlAn2 [27] database, common to all individuals and time points at each body site, were calculated and used as features with RMLR and 1NN classification with and without attribute selection (e.g., correlation-based feature selection) to attribute skin microbiomes to their respective host donors. Attribute selected markers (i.e., a subset of markers with comparable predictive power as all shared markers) were included in the hidSkinPlex panel. Markers included in the hidSkinPlex panel identified in Schmedes et al. [26] were selected from samples which met the following criteria for sample inclusion: 50x maximum read depth at any shared marker site, 10x average read depth for all shared markers, and detected in all 3 time points for each individual per body site. Marker sites were included for analysis using a threshold of 5x read depth. Additional attribute selected markers, which were not selected by Schmedes et al. [26], were included in the hidSkinPlex panel to build in redundancy in the panel in case particular markers failed to amplify. Additional markers were selected using marker inclusion thresholds of 2x and 10x read depth and an additional sample set (30x maximum read depth at any shared marker site, 5x average read depth for all shared markers, and detected in all 3 time points for each individual per body site) with marker inclusion thresholds of 2x and 5x read depth at each marker site. The final hidSkinPlex panel contained 286 markers from 22 bacterial (and phage) clades (Table S1).

Custom primers (n = 572) for each hidSkinPlex marker (n = 286) were designed by Verogen, Inc. Primers were designed to produce amplicons with maximum coverage across each marker reference sequence with no overlapping primers. Primers for amplicons less than 200 bp also incorporated the Nextera Transposase sequence (Illumina, Inc., San Diego, CA) on the 5′ end of the primer to ensure transposition during library preparation.

2.4. Development and evaluation of the hidSkinPlex multiplex assay

The hidSkinPlex amplification assay was developed using the QIAGEN® Multiplex PCR Plus Kit (Qiagen) and three bacterial DNA controls (*P. acnes* Strain SK137, *P. granulosum* D-34, and *R. dentocariosa* Strain M567) (ATCC, Manassas, VA). The quantities of total bacterial genomic DNAs were determined using the Qubit® 2.0 Fluorometer with the Qubit® dsDNA BR Assay Kit (ThermoFisher Scientific). Each of the custom primers were at 100 μM final concentration (Integrated DNA Technologies, Coralville, IA) and pooled to make a working stock, 175 nM for each primer. Multiplex reaction conditions, following recommendations in the protocol for "Multiplex PCR fragments up to 1.5 kb in length" [28], were as follows for a 50 μL reaction: 25 μL Multiplex PCR Master Mix; 5 μL 10x primer mix; 5 μL 5x Q-Solution (with and without); 1 ng template DNA; molecular-grade water (volume varies according to volume of sample added). Separate PCRs with the following modified conditions were evaluated: 17.5 nM, 8.75 nM, and 4.375 nM final primer concentrations, with and without the addition of Q-Solution, 1 ng each control DNA and a 1:1:1 mixture including each bacterial control sample (1 ng total). PCR conditions were as follows: 95 °C for 5 min; 40 cycles (95 °C for 30 s; 55 °C, 57 °C or 59 °C for 3 min; 72 °C for 90 s); and 68 °C for 10 min. PCR product was purified using the MinElute® PCR Purification Kit (Qiagen) and quantified using the Qubit® 2.0 Fluorometer with the Qubit® dsDNA BR Assay Kit (ThermoFisher Scientific). Purified PCR product was visualized on the 2200 TapeStation system (Agilent Technologies, Santa Clara, CA) with the D1000 ScreenTape and reagents (Agilent Technologies) or with the High Sensitivity D1000 ScreenTape and reagents (Agilent Technologies) (using a 1:20 dilution of purified PCR product).

2.5. Library preparation and hidSkinPlex targeted sequencing

Targeted hidSkinPlex sequencing libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina) with the Nextera XT Index Kit v2 Set C (Illumina) and 1% spiked-in PhiX Control v3 (Illumina), following manufacturer's protocol, using 90 μL volume of Agencourt® AMPure® XP beads (Beckman Coulter, Inc., Brea, CA) during library cleanup. Libraries were quality controlled and visualized on the 2200 TapeStation system (Agilent Technologies) with the High Sensitivity D1000 ScreenTape and reagents. Pooled libraries were sequenced on the MiSeq (Illumina) using the MiSeq Reagent Kit v2 (300-cycles) (Illumina) with a 2 × 150 bp read length.

DNA extracts from S001, including a reference buccal swab, also were analyzed using the ForenSeq™ DNA Signature Prep Kit (Primer Mix A) (Illumina) and sequenced on the MiSeq FGx™ Forensic Genomics System (Illumina), following manufacturer's instructions. ForenSeq data were analyzed using STRait Razor v2 s [29].

2.6. Sequence quality control and data analysis

Sequence data were preprocessed using cutadapt [30] to trim bases with a quality score less than 20 and remove reads less than 50 bases in length. Adapters were previously removed on the MiSeq system before data analysis. MetaPhlAn2 [27] was used to align sequence reads to the MetaPhlAn2 reference database, which includes the markers in the hidSkinPlex panel. Samtools [31] programs view, sort, stats, index, bedcov and mpileup were used to retrieve alignment statistics and calculate read depth and variant calls for each aligned marker in the hidSkinPlex panel. To assess the performance of the hidSkinPlex, accuracy calls (i.e, true positive (TP), true negative (TN), false positive (FP), false negative (FN)) were designated by the following criteria: TP = expected marker, present; TN = expected absent marker, absent; FP = expected absent marker, present; FN = expected marker, absent. The sensitivity ($SN = \frac{TP}{TP+FN}$) and specificity ($SP = \frac{TN}{TN+FP}$) of the hidSkinPlex were calculated for accuracy calls using a threshold of 70 x read depth (i.e., > maximum read depth observed in the reagent blank).

Custom perl and R scripts were used to calculate the nucleotide diversity (π), $\pi = \frac{1}{n}\sum_{i}^{n} 2p_i(1-p_i)$, where p_i is the frequency of the reference base at the ith site in the nth base of the marker (as described in Nayfach et al. [32]) of each marker and construct feature vectors to use for statistical classification. Classification was performed to attribute skin microbiome profiles to their individual hosts using RMLR and 1NN in Weka [33] using n-fold cross validation where n is the sample size and the training set is of size $n - 1$ (i.e., "leave-one-out cross-validation" (LOOCV)). LOOCV helps provide precise estimates of prediction accuracy by testing each sample against a maximally-sized training set, minus the test sample, while mitigating the effects of overfitting. It should be noted that cross-validation approaches may still underestimate the overall (population) error rates. While this is especially true of approaches that do not perform within-fold optimization (note that hyperparameter optimization was not performed in this study), the error estimates presented in this study may still underestimate the overall population error rates as they are conducted using a convenience sample. Attribute selection, using the CfsSubsetEval in Weka [33], also was performed prior to each classification method to

select for a subset of markers which may have similar weight than using the full set. Subsets of markers were evaluated since hidSkinPlex markers were selected across 14 body sites at different read depth thresholds, potentially building in marker redundancy and markers performing best at particular body sites. Therefore, subsets of markers may be better suited for classification at specific body sites. Upper and lower 95% confidence intervals on the binomial probability of the classification accuracy estimates were calculated using the bionom.confint in the binom R library [34] using the asymptotic method. Fisher's Exact tests were performed in R using the fisher.text function. All figures were made in R using the ggplot2 [35] and cowplot [36] R libraries, unless otherwise stated.

Principal components analysis (PCA), using the nucleotide diversities of the hidSkinPlex markers, was performed using the prcomp function in R. Maximum likelihood phylogenies of hidSkinPlex *P. acnes* species-specific markers were constructed using MUSCLE [37] and RAxML [38] as implemented in StrainPhlAn [39] and the "strainphlan_ggtree.R" script from https://bitbucket.org/biobakery/breadcrumbs using the ggtree [40] and ggplot2 [35] R libraries.

2.7. Data and script accessibility

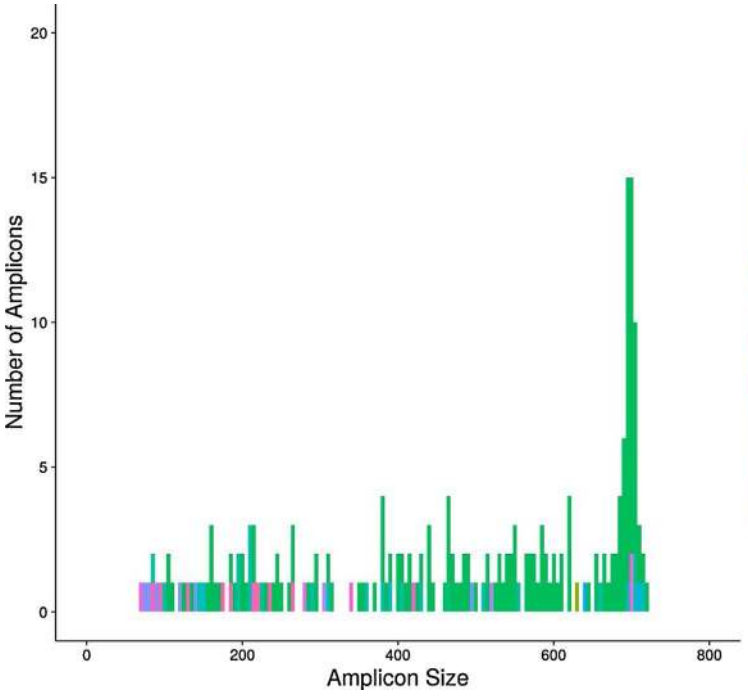
Sequence datasets can be found on the NCBI Sequence Read Archive under BioProject ID accession PRJNA398026. Custom perl and R scripts can be accessed at https://github.com/SESchmedes/hidSkinPlex.

3. Results

3.1. Development and evaluation of the hidSkinPlex targeted sequencing assay

The hidSkinPlex panel consists of 286 clade-specific markers from 22 bacterial (and phage) clades selected from the MetaPhlAn2 [27] reference database (Table S1), with > 65% of the markers from the dominant skin flora, *P. acnes*. Primers were designed to maximize coverage of each panel marker, without tiling, producing 286 amplicons (n = 572 primers) ranging in size from 72 bp to 721 bp (average 464 bp) (Fig. 1). The percentage of the marker reference sequences covered ranges from 32% to 100% (average 82%) with two amplicons designed with lengths greater than the reference genomic region. Nextera transposase sequences were incorporated into primers for amplicons < 200 bp to improve tagmentation efficiency during library preparation. Multiplex parameters including, the annealing temperature (i.e., 55 °C, 57 °C, and 59 °C), primer concentration (i.e., 17.5 nM, 8.75 nM, and 4.375 nM, each), and use of Q solution (QIAGEN) were evaluated to test the performance of the panel on 1 ng of bacterial control genomic DNA from *P. acnes* Strain SK137, *P. granulosum* D-34, and *R. dentocariosa* Strain M567 which include at least 200 markers from the hidSkinPlex panel. The hidSkinPlex also was assessed on a 1:1:1 mixture (1 ng total) of each bacterial control. Initially, the hidSkinPlex was evaluated using 17.5 nM primer concentration with an annealing temperature of 57 °C. However, primer-dimer concentrations were elevated (data not shown) and lower primer concentrations of 8.75 nM and 4.375 nM were used for a 3-stage temperature gradient (e.g., 55 °C, 57 °C, and 59 °C) assessment of the multiplex. Samples amplified using the following conditions were evaluated through the full sequencing workflow, based on amplification assessment on the Agilent 2200 TapeStation: 57 °C and 59 °C annealing temperatures; 8.75 nM primer concentration with and without Q solution (data not shown); and 4.375 nM primer concentration without the addition of Q solution (data not shown).

A total of 22.5 million raw sequencing reads (average ~0.94 million reads per sample) were generated with 14 million reads (average ~0.58 million reads per sample) remaining after quality trimming and filtering. A total of 242 markers amplified and were detected by sequencing; however, after implementing a threshold of ≥70x read depth, 200/200 expected markers were detected and were sequenced with average read depths (computed by total read depths at each base across the marker/length of amplicon) per marker ranging from 70x to > 49,000x read depth with an average of 1,278x ± 2,276 (SD) read depth (all reads in the reagent blank were < 70× and likely due to low-level bacterial contaminants from reagents [41,42]) (Fig. 2A). The performance of the panel was assessed by determining the proportion of true positives, false positives, true negatives, and false negatives based on expected marker presence/absence and calculating the sensitivity and specificity of the hidSkinPlex (Fig. 3A, Table S2). The proportion of true positives and true negatives ranged from < 30% to > 85% (Fig. 3A); however, after implementing a threshold of 70x read depth



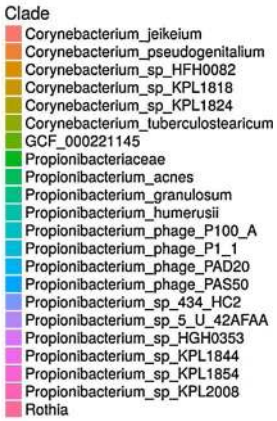


Fig. 1. A histogram of the amplicon sizes present in the hidSkinPlex panel. (Bin size = 5).

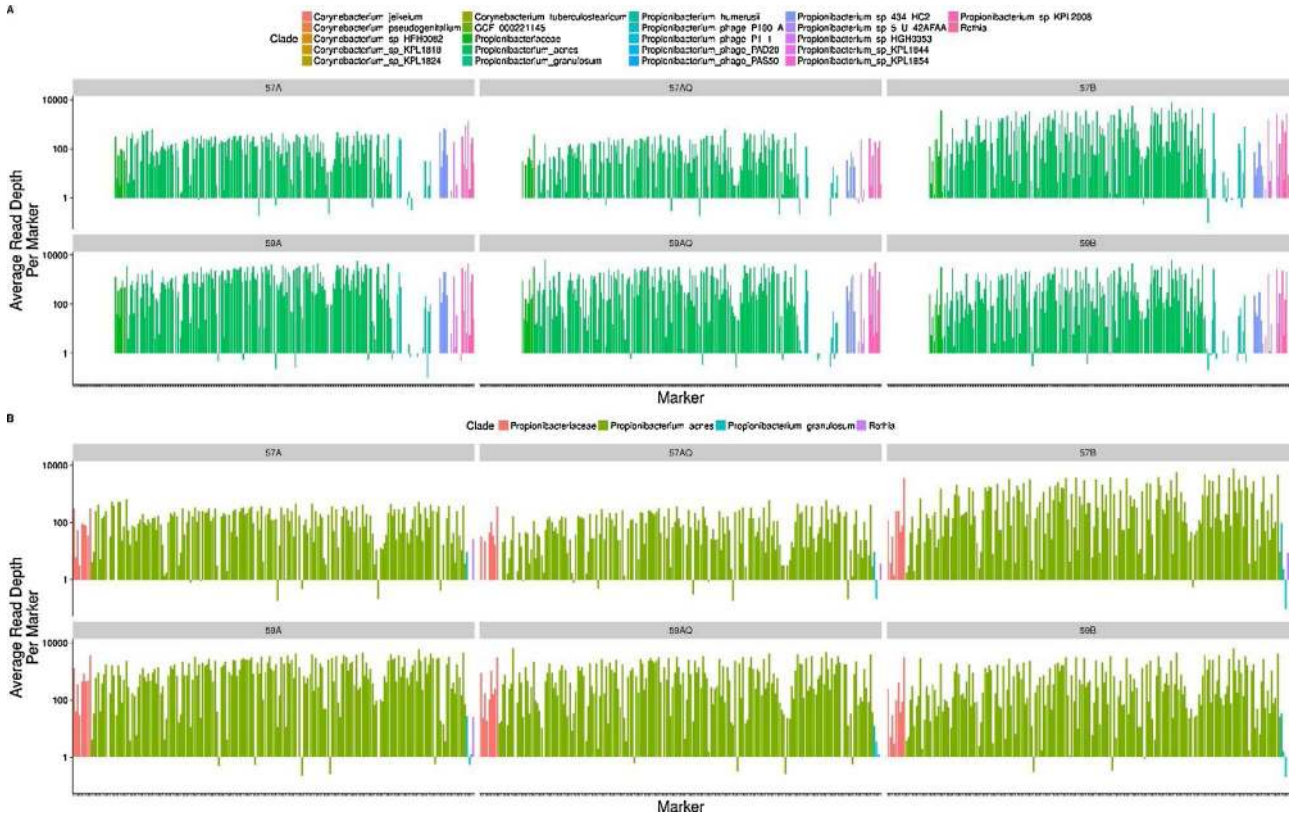


Fig. 2. The average read depth at each hidSkinPlex marker. A) Marker read depth at each marker in the hidSkinPlex (n = 286) for a synthetic bacterial mixture containing equal amounts of genomic DNA from *Propionibacterium acnes*, *Propionibacterium granulosum*, and *Rothia dentocariosa*. B) Marker read depth at each expected marker (i.e., "true positive", n = 200) for a synthetic bacterial mixture containing equal amounts of genomic DNA from *P. acnes*, *P. granulosum*, and *R. dentocariosa*. PCR parameters tested, include: 57 °C and 59 °C annealing temperatures; A = 8.75 nM final primer concentration; B = 4.375 nM final primer concentration; Q = addition of Q solution. (Markers ordered by clade then amplicon size for each PCR multiplex parameter, on a log scale).

the proportion of expected accuracy ranged from > 85% to 100% (Fig. 3B). The sensitivity of the hidSkinPlex panel, for 200/286 markers at 70x read depth, ranged from 85%–98% with a specificity range of 76%–90% (Table S2). Average read depth across each marker for true positives 70x read depth ranged from 70x to > 33,000x read depth (average of 1,123x ± 1,508 (SD) read depth) (Fig. 2B). PCR parameters of 59 °C and 8.75 nM primer concentration without Q solution (QIAGEN) were selected to assess the performance of the hidSkinPlex on skin microbiome samples since these parameters resulted in overall higher and more uniform read depth of expected markers evaluated on each control sample (Figs. 2 and 3, S1). More weight was given to the performance of *P. acnes* (n = 196) and the synthetic bacterial mixture (n = 200), since the majority of markers in the panel cover this species/sample as opposed to *P. granulosum* (n = 12) and *R. dentocariosa* (n = 1).

3.2. Skin microbiome profiling and classification using the hidSkinPlex

The hidSkinPlex was evaluated on skin microbiome samples to assess if enrichment of targeted clade-specific markers can be used to differentiate individuals based on microbiome profiles. Skin microbiome samples were collected from eight individuals, sampled in triplicate from Mb, Hp, and Fb (n = 72 samples). The Mb and Hp body sites were selected for this study due to their forensic relevance (i.e., Mb (shirt collar) and Hp (touch items)) and to overlap sites previously tested by Schmedes et al. [26], where the classification accuracies were generally higher. The foot was selected to determine if skin microbiomes from the foot can be used to differentiate individuals using targeted enrichment of informative hidSkinPlex markers. Previous attempts to use skin microbiome profiles using shotgun metagenomic data from the foot were not possible [26] due to low sequence read depth and/or coverage and high variability of markers at the foot body site [3].

DNA extracts (50 µL total volume) from the collected skin microbiome samples generated quantification results of total DNA ranging from < 0.5 to 934 pg/µL. A total of 1 ng of DNA template or up to 20 µL (maximum volume) of DNA template for each sample was amplified using the hidSkinPlex and sequenced generating 122 million raw sequencing reads (average of ~1.7 million reads per sample). Sequence reads were preprocessed to remove sequence adapters, trim bases with a quality score < 20 and remove reads < 50 bases in length resulting in 91.2 million total sequence reads (average ~1.3 million reads per sample) for downstream analysis. Sequence reads aligned to 282 out of the 286 total markers in the hidSkinPlex panel with read depths per marker ranging from 0.07x (less than 100% of the marker captured) to > 64,000x read depth (average of 2,117x ± 6,305 (SD) read depth) (Figs. 4, S2-4). A total of 183 markers, termed hereafter as universal markers, were common to all individuals and all body sites with a minimum of 2x read depth.

To assess the ability of select subsets of hidSkinPlex markers to differentiate skin microbiomes from different individuals, skin microbiome profiles were constructed by calculating the nucleotide diversity for each marker (See Methods). Marker nucleotide diversity captures the level of heterozygosity of each marker and can capture strain level variation [26,32]. Nucleotide diversities were calculated for seven read depth thresholds (i.e., 2x, 10x, 25x, 50x, 100x, 150x, 200x) for samples at each body site and all sites combined. Classification was performed for all body site samples combined to test the prediction accuracy when the body site is unknown to the classifier, in contrast to previous studies [9,24–26] in which the body site was known (i.e., conditioning on the

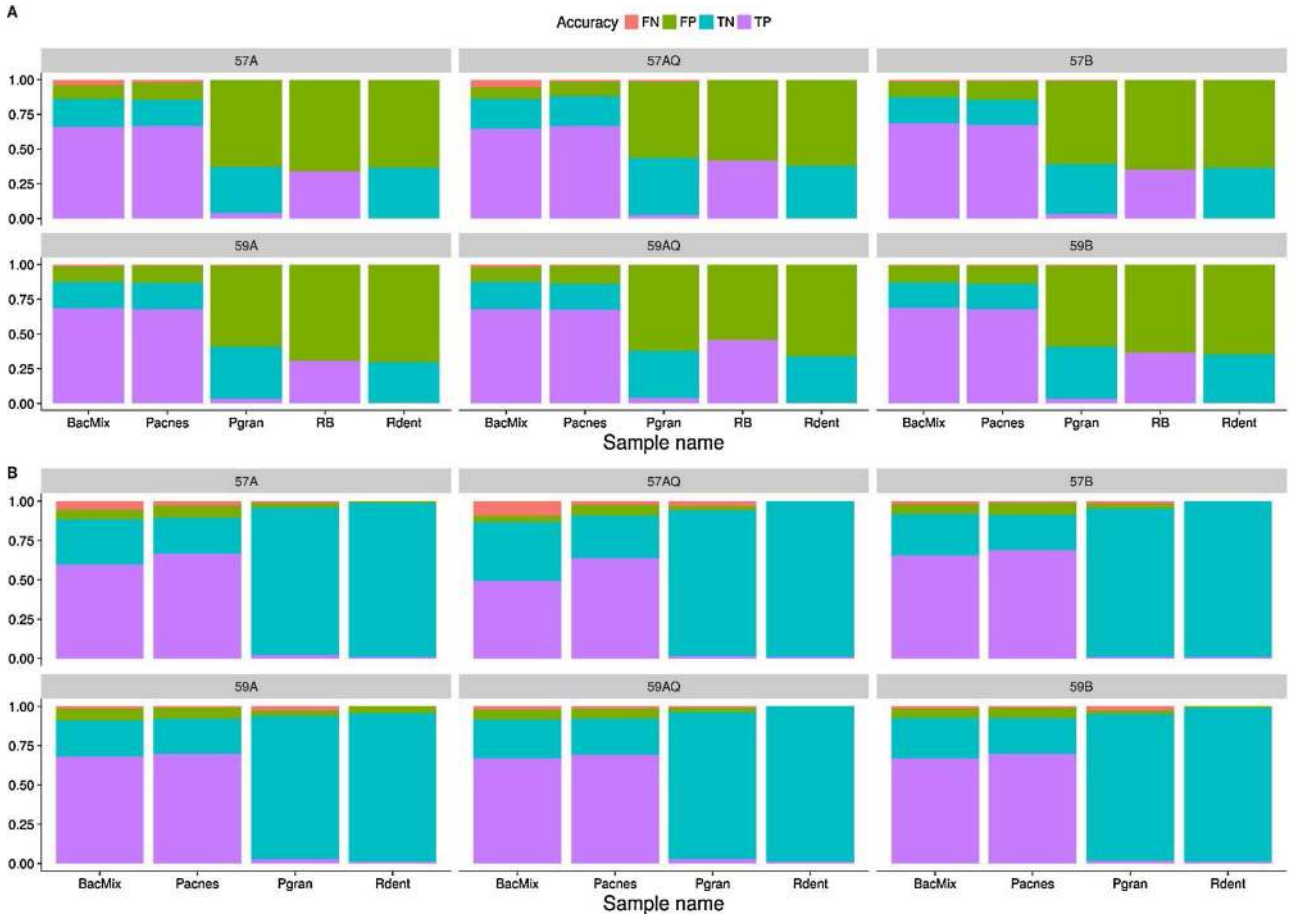


Fig. 3. Performance of the hidSkinPlex on bacterial controls *Propionibacterium acnes*, *Propionibacterium granulosum*, and *Rothia dentocariosa*. Accuracy calls (i.e, true positive (TP), true negative (TN), false positive (FP), false negative (FN)) were designated by the following criteria: TP = expected marker, present; TN = expected absent marker, absent; FP = expected absent marker, present; FN = expected marker, absent. A) Proportion of accuracy calls using a threshold of 1x read depth. B) Proportion of accuracy calls using a threshold of 70 x read depth (i.e., > maximum read depth observed in the reagent blank). Sample names: BacMix = synthetic bacterial mixture containing equal amounts of genomic DNA from *P. acnes*, *P. granulosum*, and *R. dentocariosa*; Pacnes = *P. acnes*; Pgran = *P. granulosum*; RB = reagent blank; Rdent = *R. dentocariosa*. PCR parameters tested, include: 57 °C and 59 °C annealing temperatures; A = 8.75 nM final primer concentration; B = 4.375 nM final primer concentration; Q = addition of Q solution.

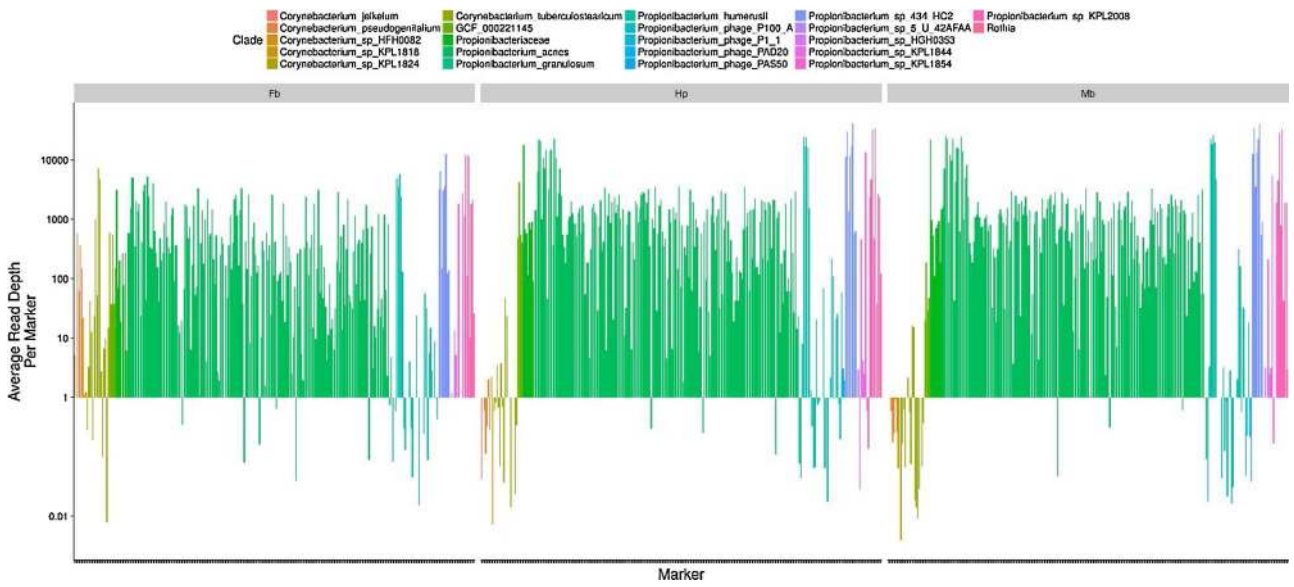


Fig. 4. The average read depth at each hidSkinPlex marker present in eight individuals from the toe web/ball of the foot (Fb), palm of the non-dominant hand (Hp) and manubrium (Mb). Markers are ordered by clade then amplicon size on a log scale.

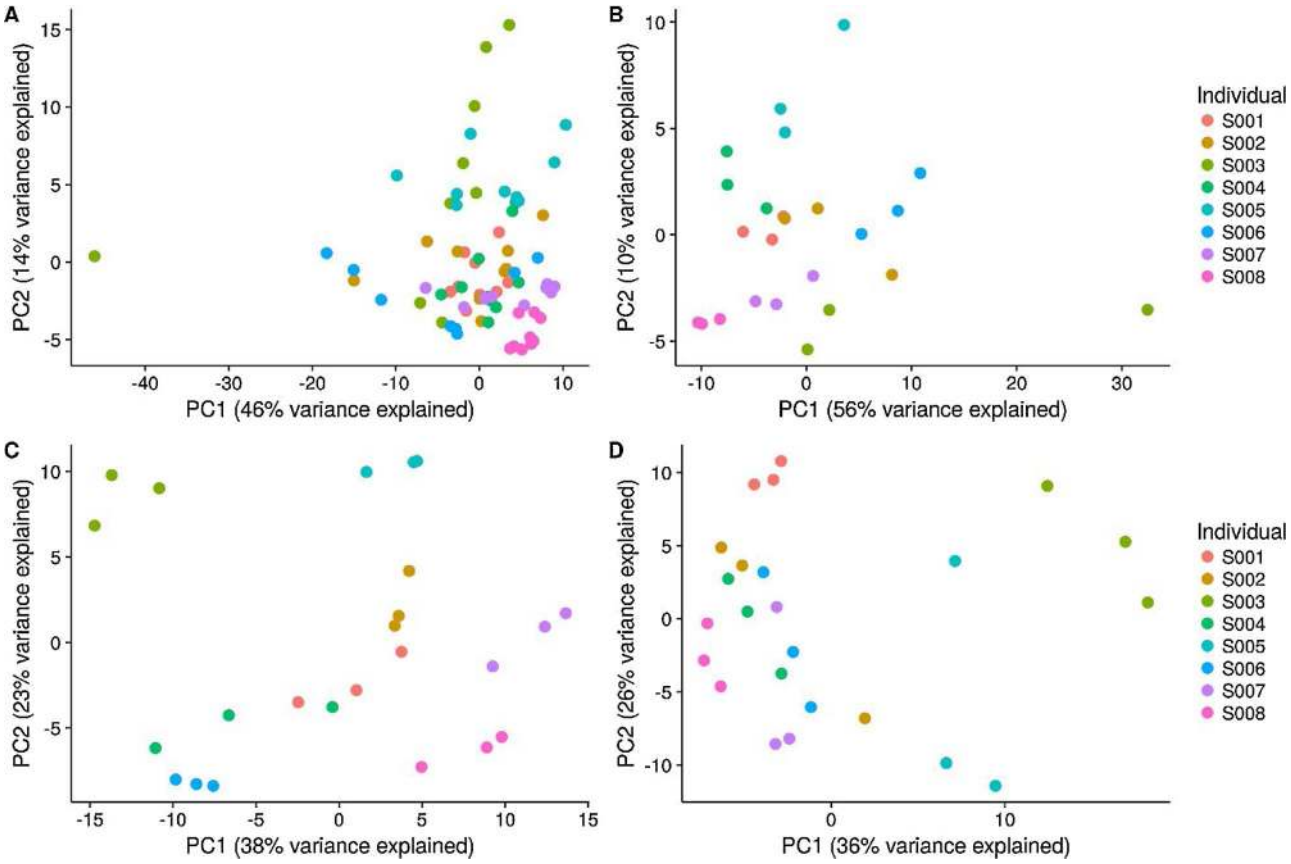


Fig. 5. Principal components analysis of the nucleotide diversity of universal hidSkinPlex markers for each body site (threshold, ≥10x read depth). A) All samples regardless of body site. B) Toe web/ball of the foot (Fb). C) Palm of the non-dominant hand (Hp). D) Manubrium (Mb).

body site). Skin microbiome profiles were assessed using subsets of universal (i.e., markers common to all individuals and body sites, including all replicates) and non-universal markers (i.e., all markers present across all samples, including common and unique markers). PCA of skin microbiomes profiles using universal markers depicted samples from the same individual at Fb, Hp, and Mb body sites clustering more closely than samples from different individuals, with few exceptions (Fig. 5). This cluster pattern was less apparent when considering all samples together, regardless of body site; however, some clustering was still observed (Fig. 5). While unsupervised learning, such as PCA, can facilitate data visualization, supervised methods are necessary to calculate predictive accuracies for sample classification.

RMLR and 1NN classification were used to attribute skin microbiome samples to their respective individual donors using LOOCV. RMLR and 1NN were performed using skin microbiome profiles comprised of universal and non-universal markers at each read depth threshold (i.e., 2x, 10x, 25x, 50x, 100x, 150x, 200x) (Fig. 6). Classification accuracies (i.e., the percentage of samples classified correctly) were highest for Hp, ranging from 95.8 to 100% (average 97.9% ± 2.1 (SD)) using 98–207 (threshold 200 x and 2×, respectively) universal markers (Table S3). Classification accuracies for Mb (threshold 200x and 25x/50x/150x, respectively) ranged from 70.8-95.8% (average 86.3% ± 6.9 (SD)). Classification accuracies calculated using enriched hidSkinPlex markers from the Hp and Mb were comparable and not significantly different from classification accuracies calculated using shotgun data [26] ($p = 1$ for Hp and Mb; Fisher's Exact Test). The hidSkinPlex enrichment successfully amplified common markers shared by all individuals on Fb, 37–188 markers (threshold 200x and 2x, respectively). The Fb results are substantially different from using shotgun sequencing data, where only 2–5 markers were common to individuals [26]. Classification accuracies for the Fb ranged from 54.2–83.3% (average 73.2% ± 7.5 (SD)). Another notable difference using targeted enrichment of common markers across body sites was the ability to classify microbiomes to their respective donor using all samples, when the body site was unknown to the classifier, in contrast to previous studies when the body site was known/assumed [9,24–26]. Classification accuracies for all samples ranged from 68.06 to 97.2% (average 87.6% ± 7.7 (SD)) using 17–183 markers (threshold 200 x and 2x, respectively). RMLR and 1NN also were performed using non-universal markers at each threshold; however, average classification accuracies were lower for all body sites (Table S4). The only improvement using non-universal markers was an increase in classification accuracy up to 91.7% (threshold 10x) using 254 markers on Fb. To compare classification accuracies using targeted markers and shotgun data from the Fb, RMLR and 1NN were performed using shotgun data from the plantar heel (Ph), toenail (Tn), and toe web space (Tw) (body sites excluded from Schmedes et al. [26]) and were found to be significantly lower than classification accuracies calculated using enriched hidSkinPlex markers ($p < 0.00001$; Fisher's Exact Test). The highest classification accuracy from the foot using shotgun data was 23% at the Tw body site.

Attribute selection (see Methods) was performed using LOOCV with RMLR and 1NN classification to determine if reduced subsets of hidSkinPlex markers produce comparable or increased classification accuracies (Fig. 6). Additionally, attribute selection may allow for the selection of the most differentiating markers which may be better suited for microbiome profiling of particular body sites. hidSkinPlex marker subsets ranged in size from 8 to 20 (all), 15–31 (Fb), 38–64 (Hp), and 13–43 markers (Mb) (Table S3). Classification accuracies using attribute selected markers were similar to accuracies using full sets of markers, a finding previously reported by Schmedes et al. [26] with shotgun metagenomic data. This finding also was observed when using

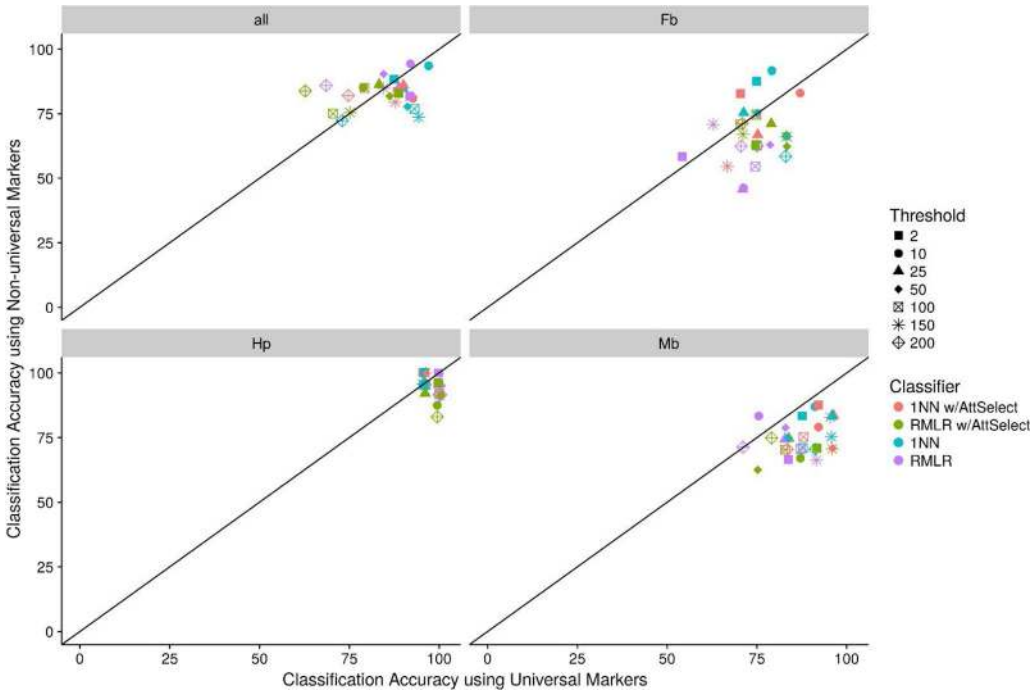


Fig. 6. Comparison of skin microbiome classification accuracies using universal and non-universal hidSkinPlex markers. RMLR and 1NN, with and without attribute selection, were performed to attribute skin microbiomes to their respective individual host at each body site (i.e., all samples (all), toe web/ball of the foot (Fb), palm of non-dominant hand (Hp), and manubrium (Mb)).

non-universal markers (Table S4).

3.3. *Propionibacterium acnes* strain characterization

P. acnes has been shown to be a dominant skin flora with single-nucleotide variant (SNV) profiles [3], clade-specific marker phylogenies and pangenome gene presence/absence profiles that are stable over time [26]. To determine if *P. acnes* strains are more closely related at the individual level (i.e., regardless of body site) or more closely related at a particular body site for each individual, maximum-likelihood phylogenies were constructed using *P. acnes*-specific hidSkinPlex markers (> 65% of the hidSkinPlex panel) enriched in each body site to evaluate *P. acnes* strain-level variation across all body sites and individuals (Figs. 7, S5–S7). If *P. acnes* strains are more closely related at the individual level, all nine samples from a particular individual would be more closely related and branch out from a common node, a pattern not observed in Fig. 7. Only all nine samples for one individual form an individual-specific clade in the tree; however, samples from Mb and Hp from two additional individuals do form unique clades specific to those particular individuals. Instead, *P. acnes* strains tend to be more closely related if originating from the same individual and same body site, although some exceptions are evident (Fig. 7). Unique individual-specific clades of *P. acnes* strains were most evident for Hp and Mb as compared to Fb (Figs. S5–7) and less diversity was observed between strains from different individuals in samples from Hp (Fig. S6).

3.4. Body site classification

The hidSkinPlex panel was developed with clade-specific markers selected for their ability to differentiate skin microbiome samples from different individuals. While the main purpose of the hidSkinPlex is for individual identification, body site identification was evaluated to determine if hidSkinPlex markers could serve a dual classification purpose. PCA of nucleotide diversities of non-universal hidSkinPlex markers showed clustering of skin microbiome samples from samples collected from all three body sites, with greater variance observed across Fb than Hp and Mb, thus resolving Fb more so from Hp and Mb (Fig. 8). RMLR and 1NN classification were performed, with and without attribute selection, as previously described, using skin microbiome profiles comprised of nucleotide diversities of non-universal hidSkinPlex markers to predict body site classification (Table S5). Body site classification was predicted with 69.4-86.1% accuracy (average 78.5% ± 4.2 (SD)) using 232–275 non-universal hidSkinPlex markers, respectively. Classification accuracies using 15–23 attribute selected markers were nearly identical (Table S5).

3.5. hidSkinPlex profile coupled with human-specific STR and SNP profile

Skin microbiome profiling provides potential to generate additional identifying genetic data than human genetic profiles alone for human identification purposes. Given the higher copy number of microbial cells to human cells, skin microbiome profiles may be used individually but also in conjunction with partial human DNA profiles for investigative purposes, especially from touched evidentiary items. To demonstrate this proof-of-concept, DNA extracts from female subject S001 from Fb, Hp, and Mb (n = 9), in addition to a buccal reference sample, were sequenced on the Illumina MiSeq FGx™ Forensic Genomics System using the Illumina ForenSeq™ panel (Primer Mix A). The recommended DNA input for the ForenSeq assay is 1 ng (5 µL maximum input volume of template); however, DNA extracts from S001 (50 µL total volume) were low bio-mass samples with total DNA concentrations ranging from < 0.5 pg/µL for samples from the hand Hp and Mb and 56–86 pg/µL for samples from Fb. Thus all samples were below the ForenSeq optimum input recommendation. Only one of the nine samples yielded a full profile (Mb, replicate #3), while 8/9 samples yielded partial profiles ranging from 32% (Hp, replicate #1) to 99% (Mb, replicate #2) alleles detected (Fig. 9). The lowest numbers of alleles detected were from samples collected from the hand (Hp), the same samples which were classified with 100% accuracy using the hidSkinPlex profile (Fig. 6). Considering these skin samples were swabbed directly from the skin of the subject, similar trends (if not lower amounts) would likely be recovered from touch items in a forensic setting. The potentially more robust microbial profiles might be able to increase the strength of an association of a sample with a donor.

Additional trace human alleles were detected from all nine skin swab samples from female subject S001. This observation in and of itself was not surprising, considering human skin comes into contact with touched objects and other people on a daily basis; however, in some

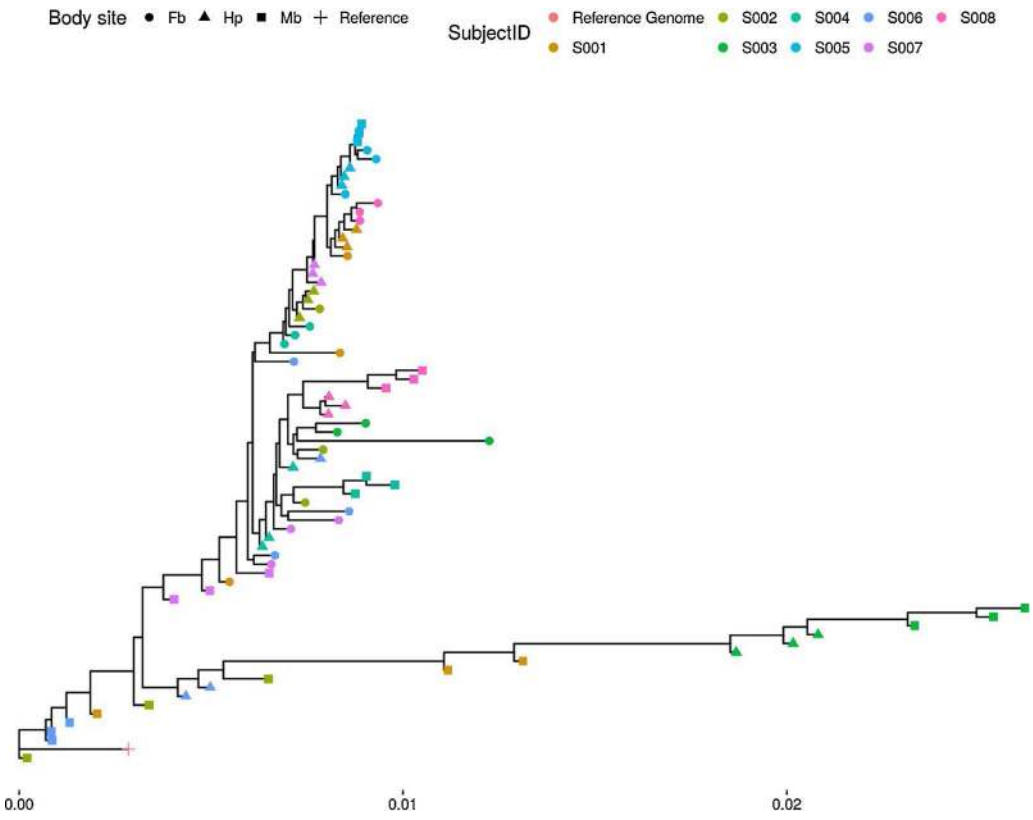


Fig. 7. Maximum likelihood phylogeny of *Propionibacterium acnes* strains present in skin microbiomes from three skin body sites and eight individuals. The *P. acnes* phylogeny was constructed using all *P. acnes*-specific markers in the hidSkinPlex panel (n = 187).

cases the trace alleles were the major contributor (File S1). Several of these trace alleles were Y-chromosome STR alleles, potentially from 1 to 2 male donors (File S1). The majority of Y-STR alleles (n = 14 loci) were detected from skin samples collected from Fb, although alleles from the suspected male donors could be detected across all three body sites. One could presume these likely come from cohabitating male family member(s) such as a partner/spouse or other relatives. Future studies would need to be conducted to collect samples from cohabiting individuals to test the hypothesis that trace levels of human DNA, as well as shared microbial DNA, are prevalent on the surface of the skin

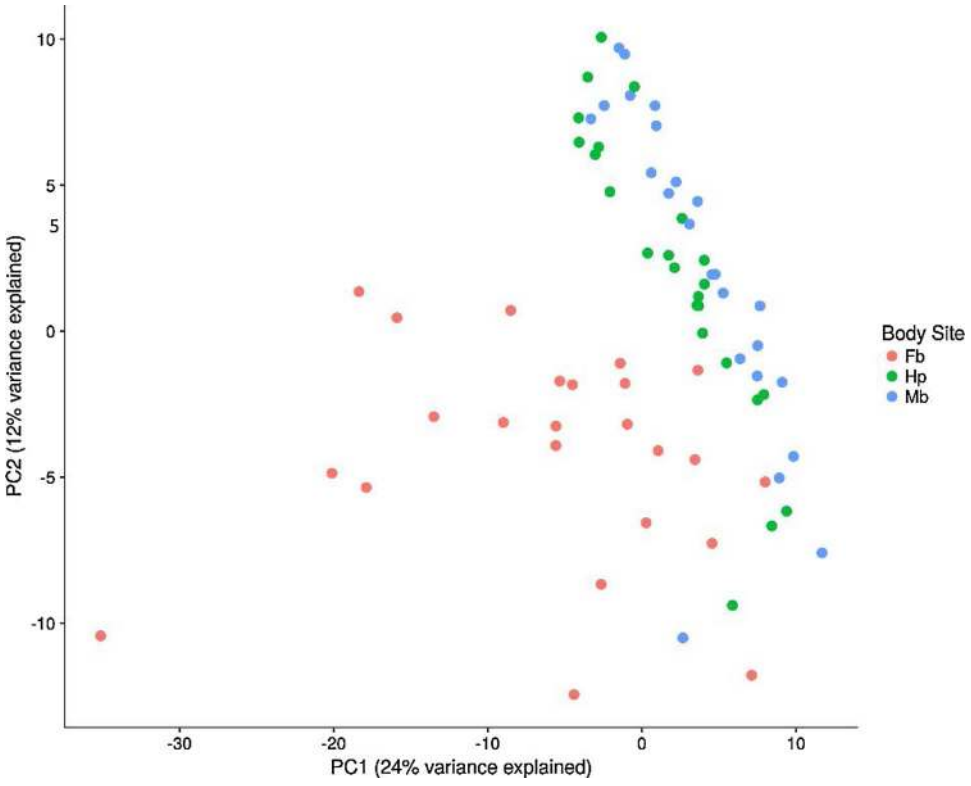


Fig. 8. Principal components analysis of the nucleotide diversity of 261 non-universal hidSkinPlex markers for all body sites (threshold, ≥10 x read depth).

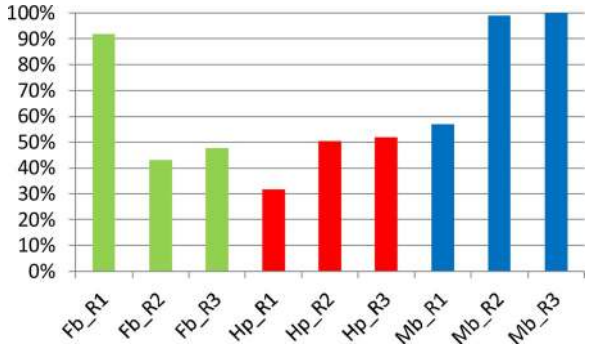


Fig. 9. Percentage of ForenSeq STR/SNP alleles detected from skin swabs collected from subject S001 from the toe web/ball of the foot (Fb = green), palm of the non-dominant hand (Hp = red), and manubrium (Mb = blue). ForenSeq profiles were generated from skin swab samples collected from female subject S001 and compared to a buccal reference swab to determine of the percentage of ForenSeq STR and SNP alleles (n = 195) called from low-biomass skin swab samples.

for periods of time. Trace human and microbial DNA profiles might be able to determine close contact and frequency of contact between individuals, potentially assisting sexual assault investigations.

4. Discussion

In this study, the hidSkinPlex, a novel targeted sequencing panel for skin microbiome profiling for forensic human identification, is described. The hidSkinPlex was developed to create a targeted enrichment solution to maximize marker detection and read depth for skin microbiome profiling. The hidSkinPlex is comprised of 286 bacterial (and phage) family-, genus-, species-, and subspecies-level markers previously selected by Schmedes et al. [26] by mining shotgun metagenomic datasets from skin microbiomes, sampled from 12 individuals over a 3 year period. These markers were deemed likely to be informative to differentiate microbiomes from individuals with a high degree of accuracy. The hidSkinPlex was designed to be coupled with Nextera XT library preparation to sequence on the Illumina MiSeq system. Three bacterial controls (i.e., *P. acnes* Strain SK137, *P. granulosum* D-34, and *R. dentocariosa* Strain M567) were used to evaluate the performance of the hidSkinPlex and yielded > 85% − 100% amplification of expected markers (Fig. 3). The hidSkinPlex was evaluated on skin swab samples collected from eight individuals and three body sites (i.e., Fb, Hp, and Mb). Amplification of hidSkinPlex markers was successful for all samples (n = 72), with amplification of 282/286 markers across all individuals and body sites (average 2,117x sequencing read depth), and 183 markers were common to all samples. Four markers from *Propionibacterium* phage P100 A, *Propionibacterium* phage P1 1, *Propionibacterium* phage PAD20, and *Propionibacterium* phage PAS50 were not detected in collected skin microbiomes samples, as well as the bacterial controls. Possible explanations for not observing the phages are they were not present, they failed to co-extract, or amplification failed due to primer design or PCR conditions. The last explanation may be likely given *Propionibacterium* phages are prevalent in the skin microbiome [2,3]. Further evaluation will be needed to determine the cause of the absence of phage markers.

Bacterial contamination was observed in both the reagents blanks for the control sequencing portion of the study as well as in the swab blanks sequenced along with subject skin microbiome samples. In the control sequencing run all reads in the reagent blank were 70x read depth; however, an average 71× ± 183 (SD) read depth was observed for the swab blanks sequenced with subject samples. These observations were expected and likely unavoidable due to the prevalence of bacterial contamination in laboratory reagents and consumables. Microbial contamination within DNA extraction kits and laboratory water has been observed and highlighted as cause for caution for microbiome and other low-abundance microbial studies [41,42]. Two of the dominant genera in the hidSkinPlex panel, *Propionibacterium* and *Corynebacterium*, have been previously reported as common contaminants in reagents [41]. Negative controls should be used to establish a baseline for such studies. The performance of the hidSkinPlex was assessed using a threshold of 70x read depth, to subtract reads from the reagent blank, to calculate true positives and negatives. However, since it was unknown what markers to expect or observe in each of the skin microbiome samples, a threshold was not used to remove reads. Given the high classification accuracies observed in this study (e.g., up to 92% (Fb) − 100% (Hp)), contamination likely did not significantly interfere with classification. In future studies, deeper analysis of these contaminant reads could be used to bioinformatically remove known contaminant reads from subject samples. Since bacterial contamination in laboratory reagents is a common issue, reagent and swab blanks should always be processed through the entire workflow and sequenced to identify any contaminants which may be present in reagents and consumables.

Classification accuracies using enriched clade-specific markers with 1NN classification for the Hp (up to 100%) and Mb (up to 96%) (Fig. 6, Table S3) were both comparable and not significantly different ($p = 1$ for Hp and Mb; Fisher's Exact Test) than accuracies observed from shotgun metagenomic data, using clade-specific markers with 1NN [26]. Additionally, hidSkinPlex markers from Fb were successfully amplified and yielded classification accuracies up to 92% using non-universal markers (Table S4). Individual classification accuracies using skin microbiomes from Fb were significantly higher ($p < 0.00001$; Fisher's Exact Test) using enriched hidSkinPlex markers, as opposed to markers from shotgun data which only yielded up to 23% classification accuracy for the toe web space (Tw) foot site [26]. The ability to classify skin microbiomes from the foot using a targeted enrichment method is significant since the foot harbors highly variable and low-abundant microbial communities [3], hindering classification using shotgun metagenomic data [26]. The foot is a forensically relevant skin site, and Goga [22] attempted to associate skin microbiome samples collected from shoe insoles with the correct owners' of the shoes using unsupervised methods. The hidSkinPlex with RMLR and 1NN offers a supervised approach to identify skin microbiomes sampled from the foot.

Enrichment of hidSkinPlex markers provides the capability to identify skin microbiomes from individuals when the body site is not known to the classifier with up to 97% accuracy using markers shared across Fb, Hp, and Mb (Fig. 6, Table S3-S4) and provides the ability to identify the body site of origin of the skin microbiome sample with up to 86% accuracy (Table S5). Thus, the hidSkinPlex can serve a dual purpose, providing a method to not only identify individuals but also predict the body site origin of skin microbiome evidentiary samples. While the hidSkinPlex was not originally designed for body site classification, the addition of body site specific markers would likely yield higher body site classification accuracies. Further analyses of body site specific markers from shotgun metagenomic data would need to be performed to assess the utility of additional marker inclusion to the hidSkinPlex panel for body site identification capabilities.

P. acnes is a highly informative, forensically relevant target due its high abundance on all skin surfaces and stability of individual-specific strain-level profiles [3,26]. Oh et al. [3] previously reported that *P. acnes* strain and SNV profiles are individual-specific and are similar across body sites. Schmedes et al. [26] described the stability of individual-specific *P. acnes* pangenome gene presence/absence profiles and *P. acnes* clade-specific phylogenies at individual body sites. Since > 65% of hidSkinPlex markers are from *P. acnes*, *P. acnes* strain diversity was assessed across all body sites to determine if strains are more closely related to individuals regardless of body site or more closely related to individuals at a specific body site. With few exceptions, *P. acnes* strains tend to be more closely related by individual and body site, in contrast to findings from Oh et al. [3] (Fig. 7). However, additional analysis of *P. acnes* strain-specific SNPs identified in Oh et al.

[3], outside the hidSkinPlex markers, would need to be performed to make a more appropriate comparison. The fact that these samples are associated within individuals across body sites, in some cases, may indicate samples collected from other body sites may be sufficient to identify an individual, even if the forensic sample is from an un-tested body site. To partially test this, classification was performed using all hidSkinPlex markers without conditioning on the body site and accuracies remained high (Fig. 6, Table S3-S4). Due to the influence of *P. acnes* markers on individual classification at Hp and Mb (Figs. S6 and 7), additional *P. acnes*-strain specific SNP loci may be informative additions to the hidSkinPlex panel for improved individual identification capability, especially across multiple body sites.

Skin microbiome profiling serves as a potential tool to use in conjunction with low-biomass or degraded samples which fail to yield full human STR/SNP profiles of touched evidentiary items. In a small case study, skin microbiome profiles (hidSkinPlex) and human forensic profiles (Illumina ForenSeq panel A) were generated from the same DNA extracts sampled for one female study subject to assess each profile type generated from the same low-biomass samples. Few samples yielded complete or nearly-complete (92–99%) STR/SNP profiles from Fb (n = 1) and Mb (n = 2). Only partial profiles, 32–52% complete, were generated for all samples from the hand; however, for these same samples using hidSkinPlex, profiles were able to be classified to their respective individual host with 100% accuracy, highlighting the potential microbiome profiles can provide, especially when used in conjunction with partial human STR/SNP profiles. These samples were collected directly from the skin, and not touched evidentiary items; touched samples would likely yield lower profile completeness. Multiple trace alleles were detected on all skin surfaces sampled from subject S001, including Y-STR alleles, with some alleles comprising the major contributor to the profile (File S1). Although, spurious alleles would be expected at low levels, likely from coming into contact with daily objects and surfaces touched by other individuals, detection of alleles common in multiple samples, and in some cases the major contributor, are likely to be due to frequent contact of subject S001, such as a spouse or family member(s). In fact, previous studies have demonstrated that microflora are more commonly shared among cohabitating family members and couples than with individuals from different households [6,43]. Indeed, Ross et al. [43] reported that microflora from the foot were more similar among couples than other body sites. Interestingly, the majority of Y-STR alleles, potentially from 1 male donor, were detected on the foot from S001. Future studies will address the level of trace human DNA shared by cohabiting couples and family members, as well as microbial DNA using the hidSkinPlex, to determine the potential of using foreign human and microbial DNA from persons as trace evidence.

5. Conclusion

In this study, the initial development and evaluation of the hidSkinPlex, a targeted sequencing panel for skin microbiome profiling for forensic human identification, are presented. Skin microbiome profiles generated using the hidSkinPlex from the foot, hand, and manubrium were attributed to their respective individual host with up to 92% (Fb) – 100% (Hp) accuracy. Additionally, body site origin could be predicted with up to 86% accuracy. Future studies will assess the stability of skin microbiomes collected over varying time intervals, skin microbiome identification from touch samples coupled with human genetic profiles, and the degree of shared microbiome signatures between cohabitating couples and family members. Additional markers for the foot body site, likely from *Corynebacterium* spp. (a common genus colonizing the foot) and body site specific markers will be evaluated for inclusion into the hidSkinPlex. Further development of the hidSkinPlex will remove redundant markers (i.e., keep attribute selected markers) and identify the most differentiating regions within each marker in order to reduce amplicon size of these regions. Since the hidSkinPlex is not yet optimized, primer redesign and concentrations will be further evaluated to provide more uniform coverage and read depth across markers. Finally, additional analysis and statistical methods will be explored to develop analysis and interpretation guidelines for use of skin microbiome profiling in the forensic setting.

Data availability

Sequence datasets can be found on the NCBI Sequence Read Archive under BioProject ID accession PRJNA398026. Custom perl and R scripts can be accessed at https://github.com/SESchmedes/hidSkinPlex.

Competing interests

Kathryn Stephens is employed by Verogen, Inc.

Funding

This project was supported by the National Institute of Justice, Award Number 2015-NE-BX-K006, the Texas Branch of the American Society for Microbiology, 2014 Eugene and Millicent Goldschmidt Graduate Student Award, and the Department of Defense, Award Number W911NF-16-0085.

Acknowledgements

We would like to acknowledge and thank Kameran Wong and Cydne Holt from Illumina, Inc. for their support and technical assistance with primer and multiplex design. We also would like to thank the subject volunteers for contributing skin swab samples for this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2017.10.004.

References

[1] E. A. Grice, H.H. Kong, S. Conlan, C.B. Deming, J. Davis, A.C. Young, G.G. Bouffard, R.W. Blakesley, P.R. Murray, E.D. Green, M.L. Turner, J. A. Segre, Topographical and temporal diversity of the human skin microbiome, Science 324 (2009) 1190–1192, http://dx.doi.org/10.1126/science.1171700.

[2] G.D. Hannigan, J.S. Meisel, A.S. Tyldsley, Q. Zheng, B.P. Hodkinson, A.J. Sanmiguel, S. Minot, F.D. Bushman, E.A. Grice, A. Grice, The human skin double-stranded DNA virome: topographical and temporal diversity, genetic enrichment, and dynamic associations with the host microbiome, MBio 6 (2015), http://dx.doi.org/10.1128/mBio.01578-15. Editor (e01578-15).

[3] J. Oh, A.L. Byrd, M. Park, H.H. Kong, J.A. Segre, Temporal stability of the human skin microbiome, Cell 165 (2016) 854–866, http://dx.doi.org/10.1016/j.cell.2016.04.008.

[4] R. Blekhman, J.K. Goodrich, K. Huang, Q. Sun, R. Bukowski, J.T. Bell, T.D. Spector, A. Keinan, R.E. Ley, D. Gevers, A.G. Clark, Host genetic variation impacts microbiome composition across human body sites, Genome Biol. 16 (2015) 191, http://dx.doi.org/10.1186/s13059-015-0759-1.

[5] T. Yatsunenko, F.E. Rey, M.J. Manary, I. Trehan, M.G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R.N. Baldassano, A.P. Anokhin, A.C. Heath, B. Warner, J. Reeder, J. Kuczynski, J.G. Caporaso, C.A. Lozupone, C. Lauber, J.C. Clemente, D. Knights, R. Knight, J.I. Gordon, Human gut microbiome viewed across age and geography, Nature 486 (2012) 222–227, http://dx.doi.org/10.1038/nature11053.

[6] S.J. Song, C. Lauber, E.K. Costello, C. a Lozupone, G. Humphrey, D. Berg-Lyons, J.G. Caporaso, D. Knights, J.C. Clemente, S. Nakielny, J.I. Gordon, N. Fierer, R. Knight, Cohabiting family members share microbiota with one another and with their dogs, Elife 2 (2013) e00458, http://dx.doi.org/10.7554/eLife.00458.

[7] H.H. Kong, J. Oh, C. Deming, S. Conlan, E.A. Grice, M.A. Beatson, E. Nomicos, E.C. Polley, H.D. Komarow, NISC Comparative Sequence Program, P.R., Murray, M.L., Turner, J.A., Segre, Temporal shifts in the skin microbiome associated with disease flare and treatment in children with atopic dermatitis, Genome Res. 22 (2012) 850–859, http://dx.doi.org/10.1101/gr.131029.111.

[8] N. Fierer, M. Hamady, C.L. Lauber, R. Knight, The influence of sex, handedness, and washing on the diversity of hand surface bacteria, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 17994–17999, http://dx.doi.org/10.1073/pnas.0807920105.

[9] E.A. Franzosa, K. Huang, J.F. Meadow, D. Gevers, K.P. Lemon, B.J.M. Bohannan, C. Huttenhower, Identifying personal microbiomes using metagenomic codes, Proc.

Natl. Acad. Sci. 112 (2015) E2930–E2938, http://dx.doi.org/10.1073/pnas.1423854112.
[10] D. Tautz, Hypervariabflity of simple sequences as a general source for polymorphic DNA markers, Nucleic Acids Res. 17 (1989) 6463–6471, http://dx.doi.org/10.1093/nar/17.16.6463.
[11] A. Edwards, A. Civitello, H.A. Hammond, C.T. Caskey, DNA typing and genetic mapping with trimeric and tetrameric tandem repeats, Am. J. Hum. Genet. 49 (1991) 746–756.
[12] A. Edwards, H.A. Hammond, L. Jin, C.T. Caskey, R. Chakraborty, Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups, Genomics 12 (1992) 241–253, http://dx.doi.org/10.1016/0888-7543(92)90371-X.
[13] P.J. Collins, L.K. Hennessy, C.S. Leibelt, R.K. Roby, D.J. Reeder, P.A. Foxall, Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR identifiler PCR amplificaiont kit, J. Forensic Sci. 49 (2004), http://dx.doi.org/10.1016/j.fsigen.2016.10.016 (JFS2002195).
[14] B.E. Krenke, A. Tereba, S.J. Anderson, E. Buel, S. Culhane, C.J. Finis, C.S. Tomsey, J.M. Zachetti, A. Masibay, D.R. Rabbach, E. a Amiott, C.J. Sprecher, Validation of a 16-locus fluorescent multiplex system, J. Forensic Sci. 47 (2002) 773–785 http://www.ncbi.nlm.nih.gov/pubmed/20457027.
[15] S. Flores, J. Sun, J. King, B. Budowle, Internal validation of the GlobalFiler??? Express PCR Amplification Kit for the direct amplification of reference DNA samples on a high-throughput automated workflow, Forensic Sci Int. Genet. 10 (2014) 33–39, http://dx.doi.org/10.1016/j.fsigen.2014.01.005.
[16] M.G. Ensenberger, K.A. Lenz, L.K. Matthies, G.M. Hadinoto, J.E. Schienman, A.J. Przech, M.W. Morganti, D.T. Renstrom, V.M. Baker, K.M. Gawrys, M. Hoogendoorn, C.R. Steffen, P. Martín, A. Alonso, H.R. Olson, C.J. Sprecher, D.R. Storts, Developmental validation of the PowerPlex® fusion 6C system, Forensic Sci Int. Genet. 21 (2016) 134–144, http://dx.doi.org/10.1016/j.fsigen.2015.12.011.
[17] J.L. King, B.L. LaRue, N.M. Novroski, M. Stoljarova, S.B. Seo, X. Zeng, D.H. Warshauer, C.P. Davis, W. Parson, A. Sajantila, B. Budowle, High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq, IForensic Sci. Int. Genet. 12C (2014) 128–135, http://dx.doi.org/10.1016/j.fsigen.2014.06.001.
[18] M.R. Wilson, J.A. DiZinno, D. Polanskey, J. Replogle, B. Budowle, Validation of mitochondrial DNA sequencing for forensic casework analysis, Int. J. Legal Med. 108 (1995) 68–74, http://dx.doi.org/10.1007/BF01369907.
[19] M.M. Holland, T.J. Parsons, Mitochondrial DNA sequence analysis – validation and use for forensic casework, Forensic Sci. Rev. 11 (1999) 21–50.
[20] E. a Grice, H.H. Kong, G. Renaud, A.C. Young, G.G. Bouffard, R.W. Blakesley, T.G. Wolfsberg, M.L. Turner, J. a Segre, A diversity profile of the human skin microbiota, Genome Res. 18 (2008) 1043–1050, http://dx.doi.org/10.1101/gr.075549.107.
[21] N. Fierer, C.L. Lauber, N. Zhou, D. McDonald, E.K. Costello, R. Knight, Forensic identification using skin bacterial communities, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 6477–6481, http://dx.doi.org/10.1073/pnas.1000162107.
[22] H. Goga, Comparison of bacterial DNA profiles of footwear insoles and soles of feet for the forensic discrimination of footwear owners, Int. J. Legal Med. 126 (2012) 815–823, http://dx.doi.org/10.1007/s00414-012-0733-3.
[23] J.F. Meadow, A.E. Altrichter, J.L. Green, Mobile phones carry the personal microbiome of their owners, Peer J. 2 (2014) e447, http://dx.doi.org/10.7717/peerj.447.
[24] S. Lax, J.T. Hampton-Marcell, S.M. Gibbons, G.B. Colares, D. Smith, J. a Eisen, J. a Gilbert, Forensic analysis of the microbiome of phones and shoes, Microbiome 3 (2015) 21, http://dx.doi.org/10.1186/s40168-015-0082-9.
[25] D.W. Williams, G. Gibson, Individualization of pubic hair bacterial communities and the effects of storage time and temperature, Forensic Sci. Int. Genet. 26 (2017) 12–20, http://dx.doi.org/10.1016/j.fsigen.2016.09.006.
[26] S.E. Schmedes, A.E. Woerner, B. Budowle, Forensic human identification using skin microbiomes, Appl. Environ. Microbiol. (2017), http://dx.doi.org/10.1128/AEM.01672-17 (in press).
[27] D.T. Truong, E.A. Franzosa, T.L. Tickle, M. Scholz, G. Weingart, E. Pasolli, A. Tett, C. Huttenhower, N. Segata, MetaPhlAn2 for enhanced metagenomic taxonomic profiling, Nat. Methods. 12 (2015) 902–903, http://dx.doi.org/10.1038/nmeth.3589.
[28] QIAGEN, QIAGEN Multiplex PCR Plus Handbook, (2016) https://www.qiagen.com/ https://www.qiagen.com/.
[29] J.L. King, F.R. Wendt, J. Sun, B. Budowle, STRait Razor v2s: advancing sequence-based STR allele reporting and beyond to other marker systems, Forensic Sci. Int. Genet. 29 (2017) 21–28, http://dx.doi.org/10.1016/j.fsigen.2017.03.013.
[30] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet. J. 17 (2011) 10, http://dx.doi.org/10.14806/ej.17.1.200.
[31] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079, http://dx.doi.org/10.1093/bioinformatics/btp352.
[32] S. Nayfach, K.S. Pollard, Population genetic analyses of metagenomes reveal extensive strain-level variation in prevalent human-associated bacteria, bioRxiv (2015), http://dx.doi.org/10.1101/031757.
[33] E. Frank, M.A. Hall, I.H. Witten, Online appendix for data mining: practical machine learning tools and techniques, in: M. Kauffmann (Ed.), WEKA Work, 4th ed., 2016.
[34] S. Dorai-Raj, Binom. Binomial Confidence Intervals for Several Parameterizations. R Package Version 1. 1-1, (2014) https://cran.r-project.org/package=binom.
[35] H. Wickham, Ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2009.
[36] C.O. Wilke, Cowplot: Streamlined Plot Theme and Plot Annotations for Ggplot2. R Package Version 0.7.0, (2016) https://cran.r-project.org/package=cowplot.
[37] R.C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797, http://dx.doi.org/10.1093/nar/gkh340.
[38] A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (2014) 1312–1313, http://dx.doi.org/10.1093/bioinformatics/btu033.
[39] D.T. Truong, A. Tett, E. Pasolli, C. Huttenhower, N. Segata, Microbial strain-level population structure & genetic diversity from metagenomes, Genome Res. 27 (2017) 626–638, http://dx.doi.org/10.1101/gr.216242.116.
[40] G. Yu, D.K. Smith, H. Zhu, Y. Guan, T.T.Y. Lam, Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data, Methods Ecol. Evol. 8 (2017) 28–36, http://dx.doi.org/10.1111/2041-210X.12628.
[41] S.J. Salter, M.J. Cox, E.M. Turek, S.T. Calus, W.O. Cookson, M.F. Moffatt, P. Turner, J. Parkhill, N.J. Loman, A.W. Walker, Reagent and laboratory contamination can critically impact sequence-based microbiome analyses, BMC Biol. 12 (2014) 87, http://dx.doi.org/10.1186/s12915-014-0087-z.
[42] M. Laurence, C. Hatzis, D.E. Brash, Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes, PLoS One. 9 (2014) e97876, http://dx.doi.org/10.1371/journal.pone.0097876.
[43] A.A. Ross, A.C. Doxey, J.D. Neufeld, The skin microbiome of cohabiting couples, mSystems (2017) e00017–e00043, http://dx.doi.org/10.1128/mSystems.