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