Forensic human identification using targeted clade-specific markers from skin microbiomes with supervised learning classification

Sarah E. Schmedes [1,2], August E. Woerner [1], and Bruce Budowle [1,2]

[1]Graduate School of Biomedical Sciences, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA
[2]Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA
[3]Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

The human microbiome contains individual-specific signatures due to genetic and environmental factors and contributes more than 5 million additional genes to the human gene repository. Several recent studies have demonstrated the utility of microbiome profiling for forensic applications, such as attributing skin microbiome signatures from touched objects to respective donors. These studies relied on various unsupervised methods to show associations between individuals and their respective samples, however very few studies have placed the problem of forensic human identification from skin microbiomes as a formal problem in classification. Further, previous methods often target the 16S rRNA gene or perform shotgun metagenomic sequencing, which may have limited species resolution and susceptibility to stochastic effects, respectively. A novel approach is described using the nucleotide diversity of stable clade-specific markers within skin microbiomes and supervised learning classification methods to associate skin microbiome samples to their respective donors. Publically-available shotgun metagenomic datasets from 12 healthy individuals, across 17 skin body sites, sampled at 3 different time points over a period of almost 3 years, were analyzed to identify stable, universal features to differentiate skin microbiomes sampled from individuals. Regularized logistic regression and 1-nearest-neighbor classification were performed in a cross validation framework, using the nucleotide diversity of clade-specific markers, to classify microbiomes sampled from each individual. Skin microbiomes were classified with 100% accuracy for samples from the cheek, hip, and knee sites. Microbiomes from the face, ear, arm, hand, and torso sites were associated with their respective donor with accuracies ranging from 81% to 97%, and microbiome samples from the foot sites had less than 25% accuracy, likely due to lower abundance of shared clades for these sites. Attribute selection also was performed to identify the most differentiating genetic markers, and these markers have been developed into a novel targeted sequencing panel for skin microbiome profiling. Our study is the first method to identify features which provide individual classification accuracy approaching 100% for specific skin sites. Stable, differentiating features at the strain level constitute a preliminary panel to develop a robust and reproducible method using skin microbiomes for forensic human identification.