HUMAN SKIN MICROBIOME: SELECTING INFORMATIVE SNPs FOR HUMAN IDENTIFICATION

Allison J Sherier M.S., Sarah E. Schmedes M.S., Ph.D., August E. Woerner Ph.D., Nicole M.M. Novroski M.S., Frank R. Wendt B.S., Jonathan L. King M.S., Bruce Budowle Ph.D.
University of North Texas Health Science Center, Center for Human Identification

The human microbiome is considered the second human genome, contributing to a large portion of the genetic content of the human body. The microbiome of the skin can add additional genetic information that can be used to supplement traditional DNA evidence. Past research has indicated that the skin microbiome is generally stable over time. In particular, in a limited sample of individuals (n=12), a shotgun metagenomics sequencing study showed that the same or similar taxa are found within a given body site, even when the same body-site was assessed up to three years later. Based on this past research, a targeted clade-specific multiplex containing 282 bacterial and 4 phage markers from 22 family-, genus-, species-, and subspecies-level clades, called hidSkinPlex, was developed and established the proof-of-concept work necessary for using the skin microbiome for human identification (HID). The hidSkinPlex was developed under the hypothesis that genes from stable, universal microbial species can differentiate skin microbiomes of individuals and be applied towards forensic HID purposes. Further testing of hidSkinPlex has produced extensive sequence data from the microbiomes of 51 unrelated individuals. Instead of focusing on microbial taxonomic similarities or differences as in previous research studies, we propose using genetic distance. Wright's fixation index (F_{ST}) was used as a novel genetic approach to select single nucleotide polymorphisms (SNPs) of interest in conjunction with supervised machine learning techniques (e.g., nearest neighbor and logistic regression) to improve upon previous classification accuracies. Selecting SNPs that differentiate individuals based on their microbial populations will allow for a novel approach to HID using the skin microbiome, similar to past approaches for selecting SNPs that differentiate individuals based on their ancestry. Using genetic distance improved performance compared to using taxonomic difference by defining informative SNPs and determining their diversity in different individuals' skin microbiome This work supports the hypothesis that the second human genome may be a viable source for HID and potentially improve analysis of samples currently yielding low level human DNA, such as the majority of touch samples.