Federal Agency: Office of Justice Programs, National Institute of Justice, Department of Justice

Federal Grant: Award Number 2015-NE-BX-K006

Project Title: Human Microbiome Species and Genes for Human Identification

PI: Bruce Budowle, University of North Texas Health Science Center at Fort Worth, Center for Human Identification, Fort Worth, TX 76107
tel: 817-735-2979; email: bruce.budowle@unthsc.edu

Submission Date: September 01, 2018

[redacted] [redacted]

[redacted] [redacted]

Recipient Organization: University of North Texas Health Science Center
3500 Camp Bowie Blvd., Fort Worth, TX 76107

Institutional Profile Number: 6108502

Project/Grant Period: 01-01-2016 - 09-30-2018

Reporting Period: 01-01-2018 - 06-30-2018

Report: **Final Summary Report**

Signature: Bruce Budowle

Bruce Budowle, Ph.D.

Summary

The Final Technical Report contains the details of the study and thus is not described in detail herein. Those interested in more explicit technical aspects should consult the Technical Report (dated 09-01-2018 and accepted by NIJ). Instead, a summary of the findings is provided to briefly capture the topic of the research.

When people touch items, they often transfer their DNA onto objects. Therefore, a genetic signature is left behind that can be exploited to determine the identity of an individual who may have handled an object, as well as to exclude individuals that did not come into contact with the item. Forensic DNA typing characterizes genetic signatures from human biological samples. The current DNA typing methodologies focus on markers in the human genome and are sensitive, highly discriminating and well-validated (for examples see Budowle and Eisenberg 2007, Budowle and van Daal 2008, and Honda et al 1999 and references within). However, the amount of human DNA deposited by touching an object often is very low, and most technologies are not designed to reliably type such low levels of DNA. To attempt to obtain results from such limited samples, modifications to methods are made that increase the sensitivity of detection of current DNA typing methods. Collectively, the suite of methods that increase the sensitivity of the well-validated human DNA typing protocols are known as low copy number (LCN) or low template typing (Gill et al 2000, Budowle et al 2001, Budowle et al 2009). Under LCN typing, limited template analyses suffer from displayed exaggerated stochastic effects and with increased sensitivity is a greater potential for contamination. This lack of reproducibility with LCN typing results has not deterred the interest in LCN typing and its potential use for developing investigative lead purposes. One way to partially overcome the limitations of LCN typing methodology is to employ an orthogonal approach. In fact, each technology, i.e., LCN typing and the orthogonal methodology, may not be individually robust, but when performed together, the information may be sufficiently corroborating to decrease uncertainty, enabling better use of trace biological sample analyses. One such orthogonal methodology that has a high sensitivity of detection is mitochondrial DNA sequencing, due to the large number of copies of the mitochondrial genome per cell. While robust and reliable, the discrimination power of this high copy marker is relatively low and likely will never approach that of the forensically-relevant nuclear markers. There is a need for another high copy marker system for human identity testing but with a much higher discrimination power.

Given the high number of bacterial cells on human skin (Grice et al 2009), it is conceivable that more bacterial cells and thus genomes are deposited on touched items than are human cells or DNA molecules. Some studies already support that bacterial signatures potentially may allow characterization to some degree of the human hosts who shed their microbes (i.e., have been deposited along with human DNA onto an item). These studies to characterize the microbiome primarily adopt one of two approaches: targeted 16S rRNA and whole-genome shotgun sequencing (WGS).

Most methods have focused on the single target 16S rRNA gene. However, it is likely that more genetic information is needed, than that of a single marker, to obtain a bacterial DNA profile sufficient for high resolution or high accuracy associations for human identity testing. Those studies using whole genome sequencing (WGS) metagenomics sequencing take a different approach than that of targeted 16S rRNA gene sequencing. WGS provides the possibility to sequence the entire genome of a single microorganism or an entire metagenome of many microorganisms in a given sample and thus should provide higher resolution. While more

comprehensive in coverage, with WGS the more area of any given genome(s) that is covered, the less read depth will be obtained for any particular site, increasing stochastic effects and potentially reducing the confidence of a base call from sequence data (both of which can impact accuracy).

Perhaps a combination of positive aspects of these two strategies – a single target gene and whole genome – may provide the sensitivity, robustness, and discrimination power necessary to support high resolution human identity (i.e., host) testing. A targeted panel of bacterial markers need to be identified for attribution purposes. These targets (or features) can be enriched (by, for example, PCR, as are current human identity markers) and subsequently sequenced (to obtain the highest genetic resolution) by massively parallel sequencing (MPS). The approach is essentially targeted genomics to survey the variety of a subset of microorganisms and selected markers present in a sample. It was proposed that this approach could provide a highly informative alternate human identity testing system.

To achieve human identification from the microbiome that is deposited on touch items this project first sought to identify selected microbial markers (with a primary focus on the most abundant species on human skin - *Propionibacterium acnes*, plus some additional taxa) from publicly available shotgun sequence data that were common to the skin of all individuals tested and to their body areas as well as being stable within an individual(s) for up to three years (Oh et al 2016). From these findings (Schmedes et al 2018), a targeted microbial marker multiplex, named hidSkinPlex, was created (Schmedes et al 2017). The approach to use this novel targeted panel is enrichment by PCR, accompanied with a sequencing method for skin microbiome profiling for forensic human identification. This approach provides greater depth of coverage than WGS and resolves bacteria of metagenomic samples down to species/strain level identification. The feasibility of classification of hidSkinPlex microbial profiles to their human hosts was evaluated using unsupervised and supervised machine learning approaches.

For the initial phase of mining possible features from publicly available shotgun sequence data, unsupervised learning techniques were evaluated to assess inter- versus intra-sample variation across host microbiomes sampled across 14 body sites (from Oh et al 2016). Two feature types capturing strain-level variation within shotgun metagenomes were compared using two supervised learning techniques. *Propionibacterium acnes* pangenome presence/absence features and the nucleotide diversities of clade-specific markers were used in conjunction with regularized multinomial logistic regression (RMLR) and 1-nearest-neighbor (1NN) classifiers to form predictions on host microbiomes based on samples from the same individuals up to three years apart. Feature selection was used to identify stable features which could be used to attribute skin microbiomes from multiple body sites to their respective hosts. This reduced set of markers was evaluated to determine if it could provide similar predictive power despite using much less information. The results from our classification algorithms were compared to evaluate if different body sites and different classification techniques significantly vary in their predictive capabilities.

The core skin microbial taxa comprised of all shared species within all tested individuals and stable over time included 10 bacterial species (*Corynebacterium aurimucosum*, *Corynebacterium jeikeium*, *Corynebacterium pseudogenitalium*, *Corynebacterium tuberculostearicum*, *Micrococcus luteus*, *Propionibacterium acnes*, *Propionibacterium granulosum*, *Pseudomonas* sp., unclassified, *Rothia mucilaginosa*, *Staphylococcus epidermidis*), 1 fungal species (*Malassezia globosa*), and 1 bacteriophage (*Propionibacterium* phage P101A). *Propionibacterium acnes* was the only species present in all samples at all body sites, ranging in average relative abundance from 35% to 89%.

Therefore, *P. acnes* is a suitable candidate for further characterization for skin microbiome profiling (although other species were assessed as well). RMLR accuracies for *P. acnes* absence/presence of features ranged from 66.67% at the ear (Ea) and interdigital web (Id) to 95.24% at the volar forearm (Vf). 1NN accuracies ranged from 58.33% at the inguinal crease (Ic) to 96.30% at the hypothenar palm (Hp). RMLR and 1NN classification also were evaluated on a reduced set of attribute selected markers (n=9 to 39). The attribute-selected loci had nearly identical classification accuracies as classification using all markers, indicating that the reduced set would suffice for association studies.

RMLR and 1NN classification were used to classify microbiome samples with respect to their individual donor in the same manner as the assessment of presence/absence markers but this time using nucleotide diversity. RMLR accuracies ranged from 66.67% at the inguinal crease (Ic) to 100% at the cheek (Ch). 1NN accuracies ranged from 56.67% at the alar crease (Al) to 100% at the inguinal crease (Ic) and popliteal fossa (Pc). RMLR and 1NN classification also were evaluated on a reduced set of attribute selected markers (n=14 to 47). The attribute-selected loci had nearly identical classification accuracies as classification using all markers collectively.

Next stability over time was assessed with the publicly available shotgun sequence data. 1NN classification accuracies, with and without attribute selection, were compared between the shortest (sampling collection time points 2 vs. 3 (5-10 weeks)) and longest (sampling collection time points 1 vs. 3 (>2.5 years)) time intervals at each body site. Microbiome samples collected 5-10 weeks apart could be attributed to their host individual with higher accuracy than microbiomes samples collected >10-30 months apart. Long time interval accuracies ranged from 30% at the alar crease (Al) to 100% at the popliteal fossa (Pc) and inguinal crease (Ic). Short time interval accuracies ranged from 50% at the ear (Ea) to 100% at the forehead (Fh), inguinal crease (Ic), popliteal fossa (Pc), and volar forearm (Vf).

Of the body sites assessed, those that are likely of the greatest forensic relevance—the Mb body site (shirt) and the Hp body site (palm)— yielded highly accurate rates of classification (93%/96%, respectively, using 1NN classification on nucleotide diversity). Surprisingly, the hand demonstrated rather high accuracy.

Based on these *in silico* results, the hidSkinPlex was developed. The hidSkinPlex is comprised of 282 bacterial and 4 phage markers from 22 family-, genus-, species-, and subspecies-level clades, which were selected and described in the *in silico* work of this project. The markers are contained in one multiplex amplification assay. Amplification allows for enrichment of informative features which in turn will increase sensitivity of detection and reduce stochastic effects compared with the shotgun sequence data that were used to select the markers. The hidSkinPlex was 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), initially in eight individuals. Note the foot was chosen because it lacked microbial features common to other skin body sites (and testing with an enrichment assay may determine whether the species are present but at lower, sometimes undetectable, levels by shotgun sequencing). RMLR and 1NN classification were used to predict skin microbiomes originating from specific body sites with their respective donors. Attribute selection 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.

After some preliminary analytical conditions evaluations the performance of the hidSkinPlex was tested for prediction purposes on skin swab samples collected initially from eight and subsequently

from up to 51 individuals and from three body sites (in triplicate). Skin microbiome profiles were constructed by calculating the nucleotide diversity for each marker using subsets of universal (i.e., markers common to all initial eight individuals and body sites, including all replicates) and non-universal markers (i.e., all markers present across all samples, including common and unique markers). Classification accuracies were highest for the hand, ranging from 95.8-100% using universal markers. Classification accuracies for the manubrium ranged from 70.8-95.8%. The hidSkinPlex enrichment successfully amplified common markers shared by all individuals on the foot. The foot results are substantially different from using shotgun sequencing data, where only 2-5 markers were common to individuals (Schmedes et al 2018). Classification accuracies for the Fb ranged from 54.2-83.3%.

While highly successful and a clear demonstration of the potential of our targeted approach for human identity testing, there were cases in which the algorithms misclassified a sample and further review was necessary. As an example, one of the manubrium replicate swabbings from an individual was misclassified and associated with a different individual. Two explanations for this mis-association are: 1) the two different individuals' microbiomes share a recent common ancestry, to the exclusion of recent common ancestry between adjacent swabbings from the same individual; and 2) recent common ancestry is apparent between two individuals, as well as within replicates swabbings of the one individual, but it is lost when considering a singular, summary phylogeny. To address this misclassification issue a measure of genetic distance (F_{ST}) between populations was computed on variable sites both between replicate swabbings of the first individual and with that of its misclassified nearest neighbor the second (different) individual. The majority of F_{ST} values are consistent with the hypothesis that there is a greater distance between two swabbings from the manubrium of the first individual than between the swabbing of the first individual and that of the second individual. However, there are more sites with high F_{ST} between individuals than within an individual. High F_{ST} markers between sample results is consistent with the assumption that adjacent body sites share recent common ancestry. The two swabbing replicates (that misclassified) fail to have as many high F_{ST} markers. These results suggest that a classifier strategy that includes the selection of high F_{ST} markers, possibly based on the number and strength of these markers, coupled with the markers from which they reside, may contribute to the refinement of a more robust, high accuracy panel for identifying the host of a touch microbiome sample. Indeed, if there are high F_{ST} markers common to human skin microbiome communities, a highly accurate classification system could be developed.

Lastly, one yet to be explored facet of the human microbiome for identity testing is the relationship of the host microbiome with the human genomic signatures (i.e., forensically-relevant identity testing markers). While there are many aspects (requiring more data analyses than presented herein), from a forensics perspective this issue can be considered initially in terms of the quantity and quality of the DNA obtained from an individual. For example, it is well established that some people shed excess quantities of DNA (the good shedders) and others are poor shedders (and likely some people are in between these two states), regarding human DNA and typing success of touch samples. A study was performed to determine if the amount of recovery of human DNA (and typing success) was correlated with that of microbial DNA. The amount of DNA recovery is directly related to the number of reads (and hence the number of markers) observed.

One of the hand replicate swabs from each of the same 51 individuals was typed for human identity markers. To formally test whether these trends are evident a robust linear regression was conducted, modeling microbial coverage as a function of human coverage. There was only marginal

significance. Thus, given this sampling of 51 individuals, it appears that the amount of shed human DNA may have some slight correlation with the amount of microbial DNA. Therefore, at this time, there is little indication that a poor shedder (from a human DNA perspective) will be more successfully typed with the microbiome panel (or vice versa as well as for good shedders). However, there were some individual examples where human DNA reads were relatively low and microbiome reads were high. Given the much higher copy number of the most abundant species (*P. acnes*), it may be that greater sensitivity of detection can be achieved more consistently with the microbiome markers. These results are preliminary and only limited analyses were performed in this study. Future work will include quantifying the human component with a human DNA specific quantitation system, comparing information content between assays, and assessing whether specific features show departures from independence. In addition, larger sample sizes of subsets of the data may provide more insight into possible relationships.

In conclusion, the studies described herein were highly successful and met the goals of the project to develop an initial targeted multiplex skin microbiome panel that potentially could be used for human identification. This multiplex is the first targeted panel, i.e., the hidSkinPlex, designed to generate individual skin microbiome genetic profiles for forensic human identification. The hidSkinPlex improves upon the limitations of 16S rRNA and WGS by capturing informative sites, which can maximize coverage and read depth, thus reducing stochastic effects. Targeted enrichment provides the capability to identify individuals using samples from a body site affording reduced stochastic effects (e.g., the foot) and to identify individuals using samples across the body, regardless of body site as well as predict body site origin. Moreover, these studies attributed skin microbiome samples collected from the hand to their respective host with up to 100% accuracy. The hand is one of the most forensically-relevant sites, regarding touch DNA samples, and as such this finding is significant for the potential use of skin microbiome profiling independent of or in conjunction with traditional human forensic profiles to assist in criminal investigations, such as robberies, homicides, and sexual assaults.

Further work is needed to bring this system to forensic fruition. Future areas of refinement to consider include increasing classification accuracy (such as evaluating sequence variation within amplicons and seeking high Fst marker solutions), improving on panel design by reducing the number of markers for robust amplification and typing without compromising accuracy, developing relevant population data, and testing the panel on forensically-relevant samples. Additional analysis methods and supervised learning algorithms also need to be assessed to develop a standardized bioinformatics pipeline and interpretation guidelines for using the hidSkinPlex (or a more informative derivative) in the forensic setting. In addition to RMLR and 1NN other algorithms should be assessed, such as support vector machines, including linear and nonlinear classifiers, other forms of logistic regression, including lasso and elastic net, K-nearest-neighbors, including methods that learn a metric distance, random forest classification and neural networks. Lastly, the hidSkinPlex allows for a suitable set of markers to begin to assess necessary components for forensic applications in a specific and sensitive manner. Application testing includes: relevant population studies (to include relatedness with co-habitants), transfer studies (i.e., utility studies), stability studies and effects of related microbiomes on association inferences.

Account of the Activities

The primary goals of this research were to: This project can be divided into three sub studies or outcomes: 1) Identification of stable clade-specific markers within shotgun skin microbiome datasets, sampled over time, and identify patterns of marker presence/absence and nucleotide diversity that are most differentiating among individuals; 2) Development of a targeted sequencing panel (hidSkinPlex) comprised of clade-specific markers for skin microbiome forensic profiling and evaluation of the panel coupled with massively parallel sequencing and deep machine learning tested for performance parameters with skin swab samples collected from 51 unrelated individuals; and 3) Compare general performance of the hidSkinPlex and human DNA markers from samples from skin swabs taken from the hand.

Accomplishments

The project was highly successful in all activity areas.

First, publicly available shotgun sequence data were mined from microbiome samples of human skin body areas across 14 body sites from 12 healthy individuals (from Oh et al 2016). Two feature types capturing strain-level variation within shotgun metagenomes were compared using two supervised learning techniques. *Propionibacterium acnes* pangenome presence/absence features and the nucleotide diversities of clade-specific markers were used in conjunction with regularized multinomial logistic regression (RMLR) and 1-nearest-neighbor (1NN) classifiers to form predictions on host microbiomes based on samples from the same individuals up to three years apart. Feature selection was used to identify stable microbial features which could be used to attribute skin microbiomes from multiple body sites to their respective hosts. This reduced set of markers was evaluated to determine if it could provide similar predictive power despite using much less information. The results provided a subset of taxa that could be used for microbiome association to respective hosts. This study was published – see Schmedes, S.E., Woerner, A.E., and Budowle, B.: Forensic human identification using skin microbiomes. Applied Environ. Microbiol. (in press).

Second, the hidSkinPlex was developed based on candidate markers from the *in silico* data described in the previous section. The hidSkinPlex panel consists of 286 clade-specific markers from 22 bacterial (and phage) clades selected from the MetaPhlAn2 (Truong et al 2015) reference database, with > 65% of the markers from the abundant 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). The markers are contained in one multiplex amplification assay. Amplification allows for enrichment of informative features which in turn will increase sensitivity of detection and reduce stochastic effects compared with the shotgun sequence data used to build the panel. The hidSkinPlex was 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), initially in eight individuals. 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)). The foot was selected to determine if skin microbiomes from this body sight can be used to differentiate individuals using targeted enrichment of informative hidSkinPlex markers. RMLR and 1NN classification were used to predict skin microbiomes originating from specific body sites with their respective donors. Attribute selection 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. Classification accuracies were highest for Hp, ranging from 95.8-100%. Classification accuracies for Mb. The hidSkinPlex enrichment successfully amplified common markers shared by all individuals on Fb. The Fb results are substantially different from using shotgun sequencing data, where only 2-5 markers were common to individuals (Schmedes et al 2018). Classification accuracy was superior using nucleotide diversity compared with presence/absence of features. This study was published – see Schmedes, S.E., Woerner, A.E., Novroski, N.M.M., Wendt, F., King, J.L., and Budowle, B.: Targeted sequencing of clade-specific markers from skin microbiomes for forensic human identification. Forens. Sci. Int. Genet. 32:50-61, 2017.

Third, the number of individuals typed with the hidSkinPlex was increased from eight to 51 so that a larger data set could be assessed. In this study two strategies were assessed: 1) phylogenetic distance to predict the host individual, operating under the premise that microbes within individuals are more closely related than microbes between/among individuals; and 2) population genetic measures of diversity at clade-specific markers, serving as a fine-grained assessment of microbial composition and quantification. NN and reverse nearest neighbor (rNN) classifiers were constructed based on the pooled data, and yielded 71% and 78% accuracy, respectively, when diversity was considered. In contrast, they performed significantly lower than a phylogenetic distance approach (54% and 63% accuracy, respectively). However, empirical estimates of classification accuracy were 100% when conditioned on a maximum nearest neighbor distance when diversity was used, while identification based on a phylogenetic distance failed to reach saturation. These findings suggest that microbial strain composition is more individualizing than that of a phylogeny, indicating that microbial composition may be more individualizing than recent common ancestry. Lastly, the data herein show that Fst may be useful for reducing misclassifications, improving the likelihood of accuracy regarding association of microbiome sample and host. This study was published – see Woerner, A.E., Novroski, N., Wendt, F.R., Ambers, A., Wiley, R., Schmedes, S.E., and Budowle, B.: Forensic human identification with targeted microbiome markers using nearest neighbor classification. Forens. Sci. Int. Genet. (in press).

Fourth, one of the hand replicate swabs from each of the same 51 individuals described above was typed for human identity markers. The intent was to perform a preliminary comparison of efficiency of the two systems – a commercially available massively parallel sequencing (MPS) kit (ForenSeq™ DNA Signature Prep Kit, Verogen, Inc, San Diego, CA) designed for typing human identity markers, i.e., STRs and SNPs, and our in-house hidSkinPlex panel designed for selected microbial targets. The latter panel has not been optimized. Results are preliminary as a human specific quantitation assay has yet to be performed on these samples and the human samples were not assessed for mixtures (future work). Only number of reads and total features/markers recovered by swabbing were assessed. The amount of recovery is directly related to the number of reads (and hence the number of markers) observed. It is well established that some people shed excess quantities of DNA (so-called good shedders) and others are poor shedders (and likely some people are in between these two states), with the DNA in question referring to human DNA. There is substantial variance in the amount of recovered material, in terms of the human DNA and perhaps even more so with respect to the microbial DNA. These effects are not restricted to any one taxonomic unit, though sampling from the hand leads to enrichments of *P. acnes* over other skin sites. A robust linear regression was conducted and it is only marginally significant (Coefficient: 0.11016, SE=0.06106, p=0.0775),

although there is a positive correlation (in the rank coverage). To test for nonlinear relationship, a spearman rank correlation test between these variables was conducted (R function cor.test), and this test was significant (S=15144, p=0.0249). Thus, given this sampling of 51 individuals, it appears that the amount of shed human DNA may have some slight correlation with the amount of microbial DNA. Larger sample sizes of subsets of the data may provide more insight into possible relationships. These data are being analyzed further for a manuscript in preparation.

Lastly, the studies described herein were highly successful and met the goals of the project to develop an initial targeted multiplex skin microbiome panel that potentially could be used for human identification. This multiplex is the first targeted panel, i.e., the hidSkinPlex, designed to generate individual skin microbiome genetic profiles for forensic human identification. The hidSkinPlex improves upon the limitations of 16S rRNA and WGS by capturing informative sites, which can maximize coverage and read depth, thus reducing stochastic effects. Targeted enrichment provides the capability to identify individuals using samples from a body site affording reduced stochastic effects (e.g., the foot) and to identify individuals using samples across the body, regardless of body site as well as predict body site origin. Moreover, these studies attributed skin microbiome samples collected from the hand to their respective host with up to 100% accuracy. The hand is one of the most forensically-relevant sites, regarding touch DNA samples, and as such this finding is significant for the potential use of skin microbiome profiling independent of or in conjunction with traditional human forensic profiles to assist in criminal investigations, such as robberies, homicides, and sexual assaults.

References

Budowle B, Hobson DL, Smerick JB, Smith JAL. (2001) Low Copy Number - Consideration and Caution. In: Twelfth International Symposium on Human Identification 2001, Promega Corporation, Madison, Wisconsin, 2001. Available: http://www.promega.com/ussymp12proc/default.htm.

Budowle B, Eisenberg AJ. (2007). Forensic Genetics, in Rimoin, DL, Connor JM, Pyeritz RE, Korf BR (eds), Emery and Rimoin's Principles and Practice of Medical Genetics, fifth edition, Vol. 1, Elsevier, Philadelphia, pp 501-517.

Budowle B, van Daal A. (2008) Forensically relevant SNP classes. Biotechniques 44:603-610.

Budowle B, Eisenberg AJ, van Daal A. (2009). Validity of low copy number typing and applications to forensic science. Croatian Med J 50(3):207-217.

Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. Forensic Sci Int 24:112(1):17-40.

Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC et al. (2009) Topographical and temporal diversity of the human skin microbiome. Science 324:1190– 1192.

Honda K, Roewer L, de Knijff P. (1999) Male DNA typing from 25-year-old vaginal swabs using Y chromosomal STR polymorphisms in a retrial request case. J Forensic Sci 44(4):868-872.

Oh J, Byrd AL, Park M, Kong HH, Segre JA. (2016) Temporal Stability of the Human Skin Microbiome. Cell 165:854–866.

Schmedes SE, Woerner AE, Budowle B. (2018) Forensic human identification using skin microbiomes. Appl Environ Microbiol (in press).

Schmedes SE, Woerner AE, Novroski NMM, Wendt F, King JL, Budowle B. (2017) Targeted sequencing of clade-specific markers from skin microbiomes for forensic human identification. Forensic Science International: Genetics 32:50-61.

Products Produced

In addition to the research results obtained, numerous presentations and a peer-review published paper have been produced that document the work.

Presentations at National and International Meetings that were supported by this work

Budowle B. Microbial forensics for microbial and human identification in criminal and civil investigations, 10th International Society of Applied Biological Sciences Conference, Dubrovnik, Croatia, 2017.

Schmedes SE, Woerner AE, Budowle B. Forensic human identification using targeted clade-specific markers from skin microbiomes with supervised learning classification, 28th International Symposium on Human Identification, Seattle, WA, 2017.

Budowle B, Schmedes SE, Woerner AE.: Microbial forensics and human identification, 9th Asian Forensic Sciences Network Annual Meeting and Symposium, Singapore, 2017.

Publications

Schmedes SE, Woerner AE, Budowle B. (2018) Forensic human identification using skin microbiomes. Appl Environ Microbiol (in press).

Schmedes SE, Woerner AE, Novroski NMM, Wendt F, King JL, Budowle B. (2017) Targeted sequencing of clade-specific markers from skin microbiomes for forensic human identification. Forensic Sci Int Genet 32:50-61.

Woerner, A.E., Novroski, N., Wendt, F.R., Ambers, A., Wiley, R., Schmedes, S.E., and Budowle, B. (2018) Forensic human identification with targeted microbiome markers using nearest neighbor classification. Forens. Sci. Int. Genet. (in press).

Invention Report

None

No patents were submitted related to this project

Participants and Other Collaborating Organizations

What individuals have worked on the project?

Name: Bruce Budowle Project role:
PI

Nearest person month worked: 1 year
Contribution to Project: Dr. Budowle has provided the overall direction and management of the project as well as participating in the technical research and data analysis.

Name: Sarah Schmedes (Note that Sarah Schmedes has graduated and is employed now at the CDC; she will no longer be working actively on the project)
Nearest person month worked: 1 year
Contribution to Project: Sarah Schmedes is a graduate student performing the bulk of the work identifying gene candidates, downloading genome data, curating data, performing laboratory work, assisting in the development of bioinformatics analysis methods, and data analysis.

Name: Frank Wendt
Nearest person month worked: 1 year
Contribution to Project: Frank Wendt is a graduate student performing sample collection and analysis with the the hidSkinPlex, performing laboratory work, assisting in the bioinformatics data analysis.

Name: Jonathan King
Nearest person month worked: 1 year
Contribution to Project: Mr. King is the laboratory manager performing daily analysis and evaluation of supporting data accumulation and will be assisting in laboratory work data analysis.

Name: August Woerner
Nearest person month worked: 1 year
Contribution to Project: Dr. Woerner is a post-doctoral research associate in the Center for Human Identification and is assisting in the development of bioinformatics analysis methods, data accumulation, and assisting in data analysis.

Name: Nicole Novroski (gratis)
Nearest person month worked: 1 year
Contribution to Project: Nicole Novroki is a graduate student performing sample collection and analysis with the the hidSkinPlex, and performing laboratory work.

Name: Rachel Wiley (gratis)
Nearest person month worked: 1 year
Contribution to Project: Rachel Wiley is a graduate student performing sample collection and analysis with the the hidSkinPlex, and performing laboratory work.

Name: Angela Ambers (gratis)
Nearest person month worked: 1 year
Contribution to Project: Angela Ambers is a postdoctoral fellow performing sample collection and analysis with the the hidSkinPlex, and performing laboratory work.

What other organizations have been involved as partners? None
Have other collaborators or contacts been involved? No

Research Impact on Criminal Justice System

What is the impact of the project on the criminal justice system?

The current impact of this project on the criminal justice system cannot be assessed, as further development is needed. The results are a fundamental leap in targets on which to focus, use of machine learning, and used of population parameters to reduce misclassification. Additional studies should consider looking at sequence data, high Fst markers, a reduced marker panel for greater sensitivity, and additional classification strategies. The potential impact (with continued support) will be substantial. Being able to have another genetic marker source for characterizing an individual who may have touched an object is an immensely beneficial tool. As more challenging samples are being presented for human DNA analyses, the limited amounts of DNA recovered are such that many samples yield inconclusive or limited partial results. The outcome of this effort will be improved capabilities to analyze touch samples. There will be development of a marker panel and an assay for characterizing the human microbiome that is deposited on objects. The outcome is that more biological evidence will be analyzed successfully, which in turn will result in more and better investigative leads to solve crimes.

How has it contributed to crime laboratories?

The current contribution cannot be assessed yet, but the anticipated outcome of this effort will be development of a panel of microbial target markers that will enable associating touch samples with their host(s). With this additional DNA typing method, samples that would otherwise provide limited results could be typable. Overall, an orthogonal typing method would improve confidence in analysis of low quantity low quality human DNA analyses of forensic biological evidence.

What is the impact on technology transfer? NA

Changes/Problems

We asked for a no cost extension until 09-30-2018. This change allowed for comparison of the microbiome marker system with that of a human marker system. We note that such a comparison has yet to be made by any group researching the microbiome for human host attribution and this gap can be addressed with this short extension.

We also asked for a GAN to ensure sufficient funds for sequencing. The funds will come from a reduction of salary support of Dr. Budowle. Dr. Budowle's effort to the project remained the same.

Proprietary Information

There was no proprietary information related to this work.