2014 IEEE 14th International Conference on Bioinformatics and Bioengineering

Evaluation of Hand Bacteria as a Human Biometric Identifier

Amanda B. Holbert[1], Holly P. Whitelam[2], Letha J. Sooter[3], Jeremy M. Dawson[4], Lawrence A. Hornak[5]
West Virginia University
Morgantown, United States
[1]aholber1@mix.wvu.edu
[2] hpw1@pitt.edu
[3]lsooter@hsc.wvu.edu
[4];jeremy.dawson@mail.wvu.edu
[5]lawrence.hornak@mail.wvu.edu

Abstract— **Molecular biometrics is an advancing field that involves the analysis of a person's unique biological markers at a molecular level to ascertain identity. Bacteria communities found on the skin of the human hand have shown to be highly diverse and to have a low percentage of similarity between individuals. The goal of this research effort is to see if a person's demographics, primarily ethnicity, share a relationship with the bacteria communities that reside on their hand. A sample collection was carried out in which the left and right inner palms of 250 individuals were swabbed to obtain a total of 500 bacteria samples. Of these, 82 samples covering a range of age, gender, and ethnicity of participants were sequenced using 150 paired-end multiplex reads on an Illumina MiSeq to analyze the hypervariable V3 region of the 16S rRNA gene. Sequences were analyzed using a combination of commercial and custom bioinformatics tools. Results indicate that women that participated in the sample collection had a 9% higher diversity of bacteria at the genus level than men. Using a support vector machine with a 60% train and 40% test approach, ethnicities of individuals who provided samples could be classified with a range of 64-93% accuracy depending on the method used. Principal coordinate plots generated by using the unique fraction (UniFrac) algorithm devised by Lozupone et al at University of Colorado at Boulder showed that similar clustering appeared with people of Turkish, Asian Indian, and Middle Eastern descent and less clustering with people of Caucasian and African American descent. Although focused on a small subset of the human population with no temporal variance in bacterial diversity explored, these results provide a basis for performing identification based on human bacteria that can be expanded upon using time varying sampling and other regions of the 16S rRNA gene.**

Keywords— Ethnicity, identification of persons, molecular biometrics, next-generation sequencing, skin bacteria

I. INTRODUCTION

A biometric trait is a physical or behavioral characteristic that is unique to an individual. Unlike forensics, which applies to post event situations such as criminal events, biometrics deals with pre-event situations such as gaining access or verification. Fingerprints, facial, iris, and voice recognition are among the most commonly used biometric traits today. However, molecular biometrics is an emerging subset of biometrics that uses unique molecular and biological markers, such as, deoxyribonucleic acid (DNA) and odor, to perform human identification [1] [2] .

Body odor is caused by secretions of different skin glands and bacterial activity. Many microbial species colonize the human skin, and some of these colonies can be harmful to the body and cause infection; however, other colonies have the opposite effect and enhance the body's immune system [3] [4]. The density of bacterial communities on the human skin can be as high as 10^7 bacteria cells per square centimeter, and these communities can be more diverse than the bacteria that reside in the throat, stomach or gut [5]. The palm of the hand is thought to contain the most bacterial diversity due to its frequent exposure to different surfaces and environments. Even with exposure to varying conditions such as temperature fluctuations, soaps and detergents, lack of constant moisture, and ultraviolet radiation, an individual's hand bacterial communities have been shown to be highly resistant to these fluctuations and with composition containing some amount of consistency over time [6] [7] [8] .

In a study performed by the Fierer group at the University of Colorado at Boulder, is was observed that the diversity of skin bacterial communities found on the palm surfaces of individuals is quite large, with a total of 4,742 different phylotypes identified across 102 samples [6]. It was shown that women have a much higher level of bacterial diversity than that of men and certain bacteria groups were more abundant on one gender than the other. The similarities between two individuals' hand bacteria showed only a 13% similarity and only a 17% similarity between an individual's left and right hand. This uniqueness has enabled hand bacteria to be used as a forensic tool to link individuals to their personal belongings [9]. Bacteria were swabbed from individuals' fingertips, the individuals' keyboard keys, and additional keyboards from other individuals and random public keyboards. Comparisons between these samples showed that the bacteria communities on the individual's fingertips and the communities on their keyboard were far more similar to each other than the communities found on other individuals' fingertips and random keyboards. Since bacteria can reside on surfaces for at least two weeks, it is evident that people in the same household would have more shared bacteria than individuals outside of the household. Although observations have shown there are more shared bacteria with members in the same household, there is still a sufficient amount of uniqueness between the individuals [10].

The ethnicity of an individual has also been able to be determined by specific bacterial communities found in the

978-1-4799-7502-0/14 $31.00 © 2014 IEEE
DOI 10.1109/BIBE.2014.36

mouth [11]. Analysis of oral bacterial communities collected from 192 individuals who were either non-Hispanic black, non-Hispanic white, Chinese or Latino proved to predict the individuals' ethnicity with a 62% accuracy. Observations showed that the ethnicity of an individual was best observed by the bacteria that were more influenced by the individual's heredity rather than the bacteria that were associated with the individual's food intake or hygiene.

Research studies exploring the uniqueness of hand bacteria distribution in the human population indicate the potential to use these distributions as a means of identification. If a subset of colonies can be identified as major 'markers' for identity, sensor technologies can be developed to look for those specific bacteria, similar in methodology to how short tandem repeats (STR) are used for forensic DNA profiling. Although it is known that bacteria collected from the gut have a high affiliation with that individual's genetic makeup [12] [13] , this area is not as easily accessible as bacteria collected on the skin. The aim of this study was to analyze the third hypervariable region (V3) of the 16S ribosomal RNA (rRNA) gene from bacteria collected from the palm of a hand, which can be used to distinguish a large population of bacteria down to the genus level [14], would produce similar results in uniqueness, and to determine if a person's demographics, primarily ethnicity, shared a relationship with the bacteria communities that exists on their hand.

II. METHODS

This section discusses how the hand bacteria samples were collected, what methods were used to extract and amplify the target region of interest, and what bioinformatics tools were used to analyze this region.

A. Sample Set

250 individuals participated in a sample collection held at WVU's Health Sciences Center over a three month period under IRB H-23693. Each participant that arrived during their scheduled appointment signed and dated a consent form describing in detail of the data collection procedure. Information about the individuals' demographics such as age, gender and ethnicity was also gathered after the consenting.

The participant's hands were swabbed by taking the end of a cotton swab that was dipped into a double distilled water solution containing 0.15 M NaCl and 0.1% Tween 20 [15]. The cotton swab had been sterilized prior to collection by being wrapped in aluminum foil and autoclaved. A staff member swabbed the entire inside of the participant's hand while rotating the cotton tip. After swabbing, the head of the tip was placed inside a 2ml bead solution tube from an Ultraclean Plant DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA), as has been reported previously [16], and then cut with clean scissors. A new cotton swab was then used for the other hand. When finished the same procedure was followed as with the first sample. The tubes were stored at -80° C until the DNA extraction procedure was performed.

B. Bacteria DNA Isolation and Amplification

DNA isolation was carried out on 102 samples of the sample set. A smaller sample subset was analyzed first to see what observations could be made. This subset contained a well-balanced variation of ethnicity and gender. Isolation was performed by following the Ultraclean Plant DNA Isolation Kit Protocol with the following modifications:

1) Prior to adding Solution P1 samples were placed in a 65°C water bath for 5 minutes
2) Step 9, 500 µl of supernatant was transferred to a clean 2 ml collection tube
3) Step 14, 600 µl of supernatant was transferred to a clean 2 ml collection tube
4) Step 23, Milli-Q H_2O was used for elution buffer instead of solution P5 and was left to sit for 5 minutes before centrifuging.

All other steps were followed as directed in the Ultraclean protocol. After isolation, the samples were stored at -20°C until amplification was carried out.

The 16S rRNA gene was amplified using primer set E8F and E1541R, primer sequences were located just inside the target region [17] (TABLE I). Primers were synthesized by Eurofins Genomics (Huntsville, AL, USA). Using a 50 µl polymerase chain reaction (PCR) reaction mixture, each mixture consisted of 10 µl of DNA template, 10 µl of 5x HF Buffer (includes $MgCl_2$) (New England Biolabs, Ipswich, MA, USA), 1 µl of 20 µM of each reverse and forward primer, 2.5 µl of dimethyl sulfoxide (DMSO), 0.5 µl of Phusion *Taq* polymerase (New England Biolabs, Ipswich, MA, USA) and 0.5 µl deoxynucleotide triphosphate (dNTP). Settings for PCR using the MJ Mini Personal Thermal Cycler (Bio-Rad Laboratories, Hercules, CA, USA) included an initial 95°C for 5 min, 30 cycles of 95°C for 45 sec, 51.4°C for 1 min and 72°C for 30 sec and a final extension step of 72°C for 7 min. PCR products were purified using the PCR cleanup protocol from a Gel/PCR DNA Fragment Extraction Kit (IBI Scientific, Peosta, IA, USA) [18].

C. Library Generation and Sequencing

After purification, the V3 region was amplified again using forward primer 341F and modified reverse primer 518R which are located just outside the target region [19] [20] (synthesized by Eurofins Genomics, see TABLE I). The modified reverse primer contained a unique six base pair index e.g. CGTGAT, ACATCG, GCCTAA, which would allow for identification of each sample during multiplexing [20]. The same PCR reaction mix that was used prior for the 16S rRNA was used again except for the change of primers. Again, using the MJ Mini Personal Thermal Cycler, PCR settings were an initial 5 min denaturation set at 95°C, 20 cycles of 95°C for 1 min, 50°C for 1 min, and 72°C for 1 min and then completed with an extension step set at 72°C for 7 min.

TABLE I. BACTERIA SPECIFIC 16S RRNA AND V3 REGION PCR PRIMER AND SEQUENCE

Region	Code	Sequence 5′–3′
16S rRNA	E8F	AGAGTTTGATCCTGGCTCAG
	E1541R	AAGGAGGTGATCCANCCRCA
V3	341F	AATGATACGGCGACCACCGAG ATCTACACTCTTTCCCTACACG ACGCTCTTCCGATCTCCTACGG GAGGCAGCAG
	518R[a]	CAAGCAGAAGACGGCATACGA GATNNNNNNGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT ATTACCGCGGCTGCTGG

Each library consisted of samples could be identified by the unique index present in the reverse primer. Samples were again purified using the IBI Scientific kit and their absorbance was measured at 260nm using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The DNA concentration of each sample was determined using Beer's Law. The samples were then each diluted to 20ng/µl accordingly. 5µl of each sample was then pooled and sent to the Genomics Core Facility at West Virginia University for a 150bp paired-end multiplexed sequencing using the Illumina Miseq. Samples were loaded at a concentration of [8pM] as per guidelines provided by Illumina for the MiSeq V2 reagent chemistry. After project runs were complete the data was sent to BaseSpace [21], Illumina's next-generation sequencing cloud, for automatic analysis and storage. Index reads and base call quality scores for each sample were written to two FASTQ format files. One file for the forward run and the other for the reverse run. Due to some low quality scores with some of the samples, only 82 samples from the set were used for further evaluation.

D. Bioinformatics Method I

Using the FASTQ files, BaseSpace performed classification on the index reads using a Bayesian classifier. The classification process involved matching short subsequences of the reads (called words) to a set of 16S reference sequences from Greengenes database. The accumulated word matches for each read were used to assign reads to a particular taxonomic classification. Summary statistics provide the total number of classified clusters for each sample at each taxonomic levels: kingdom, phylum, class, order, family, and genus. Index reads were merged and aligned using Mothur Software [22] and sorted using custom Matlab scripts.

E. Bioinformatics Method II

The second method was provided by the Genomics Core Facility. This pipeline used the ERNE-filter for quality filtering, matched index reads using USEARCH [23], classified using QIIME [24], and parsed and merged indexes using custom scripts. Data was than analyzed using Vegan and R package [25].

[a] Insertion of six bases "NNNNNN" is for a unique index

III. RESULTS

Fig. 1 provides visualization of the classification breakdown of a sample sequenced. Sample results contained an average of 646,518 bacteria clusters that passed filter. Clusters that were able to be classified to genus level were used for further analysis. An average of 394,179 bacteria clusters were identified at genus level with an average of 231 identified operational taxonomic units (OTU's) per sample and a total of 777 different OTU's identified across the sample set. Women who participated in the collection were shown to have a 9% higher diversity of OTU's than men. A single rooted phylogenetic tree in Newick format was created in Mothur containing all OTU's at the genus level from each sample (including both left and right hand for majority of individuals). From the phylogenetic tree, the UniFrac algorithm [26] [27] [28] was used to create a pairwise dissimilarity matrix. The UniFrac algorithm measures the uniqueness between each sample by evaluating the branch lengths within the phylogenetic tree. A branch leading to an OTU from both samples is marked as shared, whereas a branch leading to an OTU that only appears in one sample is marked as unshared. If all branches are unique between two samples, the distance score for those two samples would be a 1.0. If the two given samples were identical, they would receive a score of 0.0.

The UniFrac distance matrix revealed uniqueness scores lower than observed scores in previous studies. The average uniqueness score for all of the samples was only 43%, whereas in [6], results showed a 13% to 17% similarity. This could be a result of only using the genus level and using a small target region. Principal coordinate analysis (PCoA) plots were created using the weighted UniFrac algorithm, which takes the abundance of each OTU into consideration. Fig. 2-4 displays PCoA plots with the 82 samples based off of ethnicity. Participants who identified themselves as Middle Eastern, Asian Indian, and Turkish descent revealed to have closer clustering, whereas participants who identified themselves as just Asian descent also showed clustering, while Caucasians and African Americans did not. Fig. 5 and Fig. 6 display plots based off of participants' age and gender. Neither plot shows any apparent clustering but as previously mentioned women differ from men by having a higher diversity of OTU's.

For both bioinformatics methods, the five most common OTU's revealed the highest potential for ethnicity classification. A supervised support vector machine (SVM) was used for training and cross validation. Using a third degree polynomial kernel function, 60% of the data set was used for training and the remaining 40% for testing. Training was randomly sampled for each ethnicity class. 400 iterations were competed and the average accuracy rate was reported. Table II shows the accuracy rate for each method. Method I used Sphingopyxis, Streptococcus, Corynebacterium, Straphylococcus, and Propionibacterium for the five features and Method II used Kaistobacter, Acinetobacter, Corynebacterium, Straphylococcus and Propionibacterium. Both methods showed similar results.

Asian and African American showed the lowest accuracy rate while Turkish and Hispanic showed the highest, but also had the least amount of samples, so less information was gathered. Fig. 7 and Fig. 8 display the average percentages of the five most commonly found bacteria for each ethnicity using Method I bioinformatics and Method II bioinformatics, respectively. Similar trends are again shown between both methods. Hispanic and Asian ethnicities displayed a higher percentage of Sphingopyxis (Method I) Kaistobacter (Method II) compared to other ethnicities. Middle Eastern and Asian Indian showed similar ratios of the 5 most common chosen OTU's. Other classification methods were performed, but the results were inconclusive due to the small size of the data set.

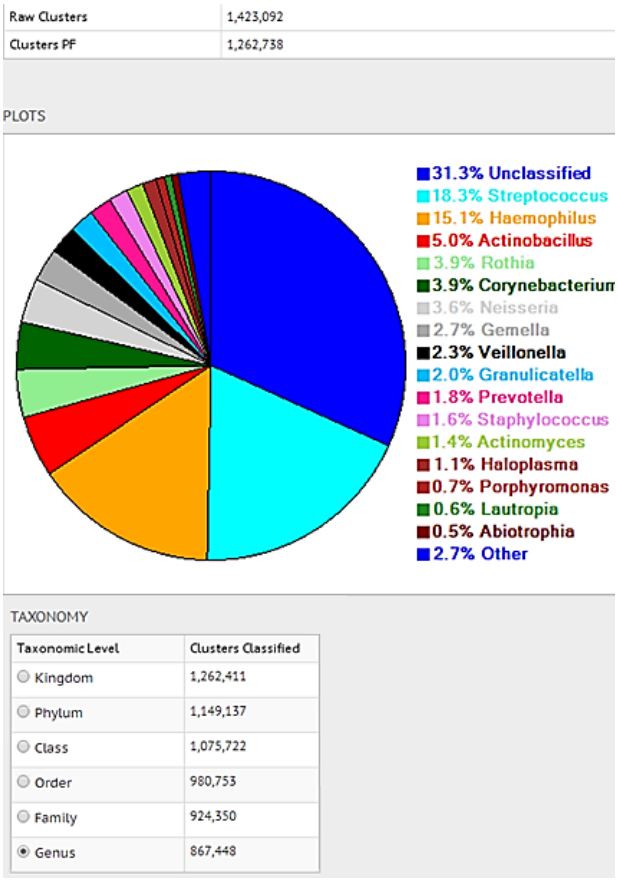

Figure 1. Screen shot from BaseSpace's metagenomics analysis showing visualization of clusters identified from a participant's hand sample.

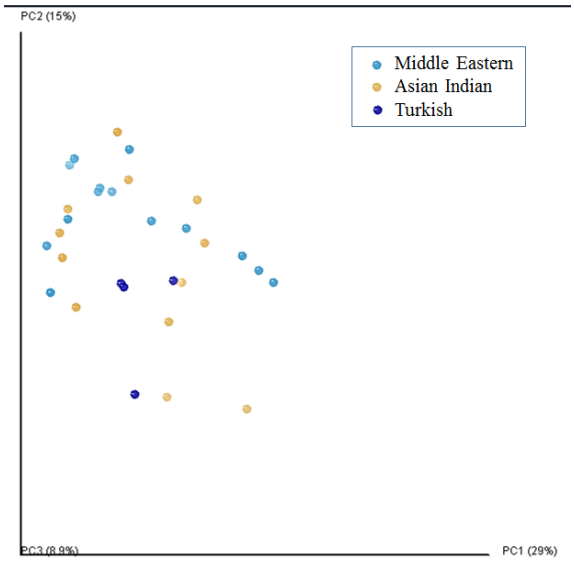

Figure 2. Clustering of Middle Eastern, Asian Indian and Turkish ethnicities (more clustering on left hand side).

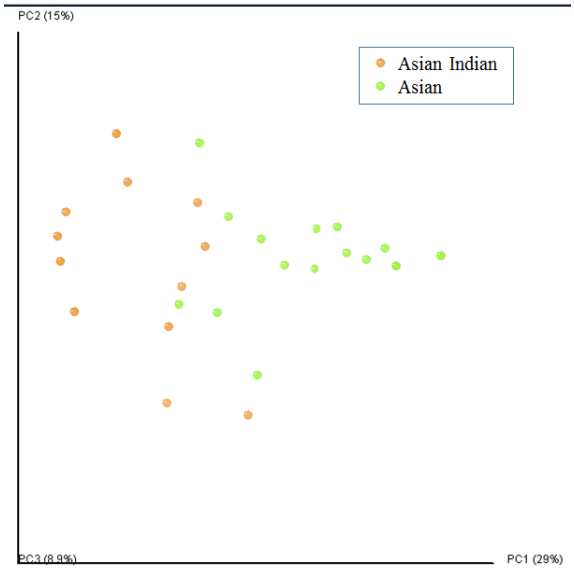

Figure 3. Dissimilarity between Asian and Asian Indian ethnicities

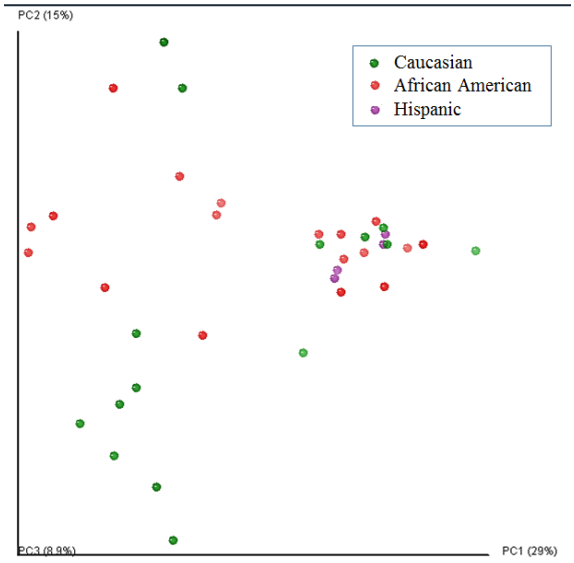

Figure 4. No clustering apparent with African American or Caucasian ethnicities. Slight clustering with two participants (four samples) identified as Hispanic.

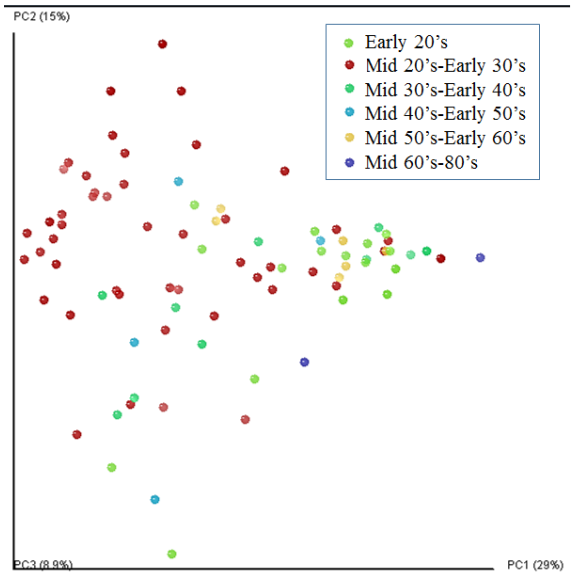

Figure 5. No clustering apparent with participants based on their age.

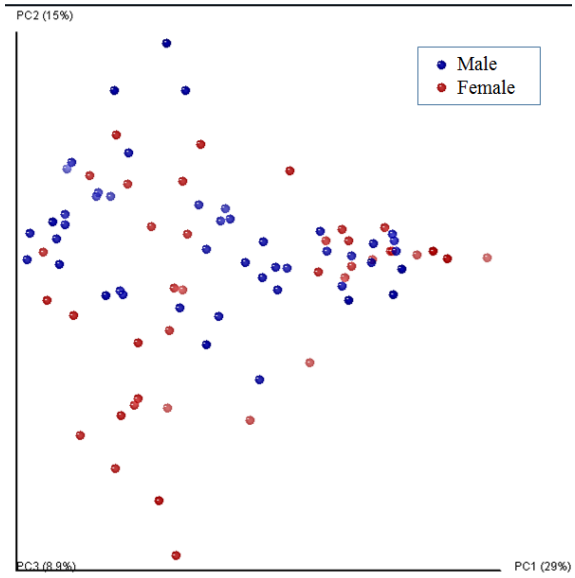

Figure 6. No clustering apparent with participants based on gender.

TABLE II. ACCURACY RATE OF SUPPORT VECTOR MACHINE CLASSIFICATION METHOD

Ethnicity	Number of Samples	Method I (%)	Method II (%)
Caucasian	15	81	77
Hispanic	4	85	88
African American	18	66	70
Turkish	4	92	93
Middle Eastern	14	76	79
Asian	15	64	67
Asian Indian	12	76	77

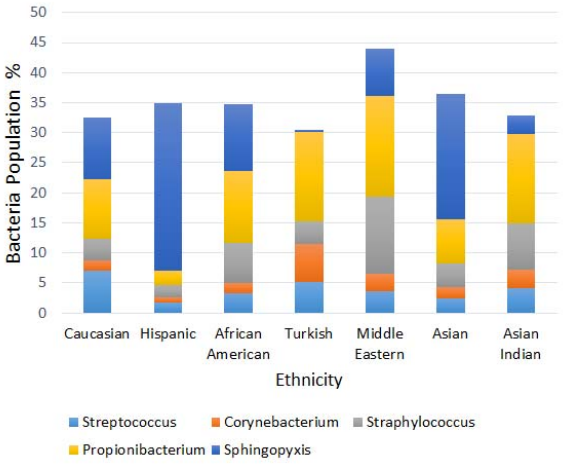

Figure 7. Percentage of the five most common bacteria found using Method I bioinformatics.

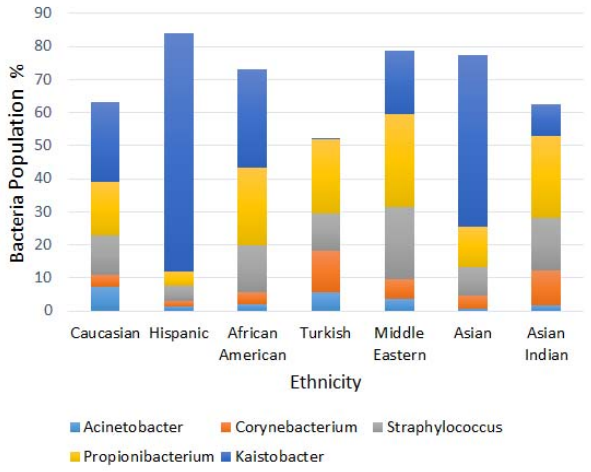

Figure 8. Percentage of the five most common bacteria found using Method II bioinformatics.

IV. Conclusion and Recommendations

New methods for molecular biometrics has become more desirable with the advancement in sequencing technology. The goal of this research effort was to identify bacteria using the V3 region of the 16S rRNA gene and to use these bacteria as 'markers' for identification of an individual. Findings revealed there are differences in bacteria diversity between men and women, and that no two samples were the same. Results also indicated that there is some clustering between Asian, Asian Indian, Middle Eastern, and Turkish ethnicities, which shows these ethnicities have a higher similarity between one another than to other ethnicities that didn't such as Caucasian and African American.

Although this study focuses on a small sample set with no additional samples or samples taken at different times for comparison and validation, these results do provide a basis for performing identification based on bacteria found on the human hand. Clustering was apparent with some of the ethnicities and particular bacteria did show more prominent on certain ethnicities than others. With more samples left that can be processed, the dataset can be enlarged, which may result in improved pattern recognition and decision making. Additionally, with the remaining samples it is possible to explore and combine other regions of the 16S rRNA gene to identify a higher level of uniqueness between individuals, thus resulting in a more reliable tool for human identification in the future.

Acknowledgments

This work was supported by DOJ contract no. 2010-DD-BX-0161 and the Center for Identification Technology Research (CITeR http:/www.citer.wvu.edu). We would like to note that Holly Whitelam participated in this project as part of the summer 2013 CITeR Research Experience for Undergraduates (REU) Program. We would like to acknowledge the use of the West Virginia University Genomics Core Facility for sequencing and assistance in bioinformatics methods.

References

[1] J. N. Lundstrom and M. J. Olsson, "Functional neuronal processing of human body odors," *Vitam Horm,* vol. 83, pp. 1-23, 2010.

[2] A. K. Jain, A. A. Ross and K. Nandakumar, "Chapter 1: Introduction," in *Introduction to Biometrics*, Springer, 2011, pp. 1-4.

[3] V. Ki and C. Rotstein, "Bacterial skin and soft tissue infections in adults: A review of their epidemiology, pathogenesis, diagnosis, treatment and site of care," *Can J Infect Dis Med Microbial,* vol. 19, no. 2, pp. 173-184, 2008.

[4] R. L. Gallo and T. Nakatsuji, "Microbial symbiosis with the innate immune defense System of the skin," *J Invest Dermatol,* vol. 131, no. 10, pp. 1974-1980, 2011.

[5] D. N. Fredricks, "Microbial ecology of human skin in health and disease," *Journal of Investigative Dermatology Symposium Proceedings,* vol. 6, pp. 167-169, 2001.

[6] N. Fierer, M. Hamady, C. L. Lauber and K. Rob, "The influence of sex, handedness, and washing on the diversity of hand surface bacteria," *Proc Natl Acad Sci USA,* vol. 105, no. 46, pp. 17994-17999, 2008.

[7] S. M. Smith, R. H. Eng and F. T. J. Padberg, "Survival of nosocomial pathogenic bacteria at ambient temperature," *Journal of Medicine,* vol. 27, no. 5-6, pp. 293-302, 1996.

[8] R. R. Roth and W. D. James, "Microbial ccology of the skin," *Annu Rev Microbiol,* vol. 42, pp. 441-464, 1988.

[9] N. Fierer, C. L. Lauber, N. Zhou, D. McDonald, E. K. Costello and R. Knight, "Forensic identification using skin bacterial communities," *Proc Natl Acad Sci USA,* vol. 107, no. 14, pp. 6477-6481, 2010.

[10] 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 and R. Knight, "Cohabiting family members share microbiota with one another and with their dogs," *eLife,* vol. 2, no. e00458, 2013.

[11] M. R. Mason, H. N. Nagaraja, T. Camerlengo, V. Joshi and P. S. Kumar, "Deep sequencing identifies ethnicity-specific bacterial signatures in the oral microbiome," *Plos One ,* vol. 8, no. e77287, 2013.

[12] Consortium, The Human Microbiome Project , "Structure, function and diversity of the healthy human microbiome," *Nature,* vol. 486, no. 7402, pp. 207-214, 2012.

[13] N. Marathe, S. Shetty, V. Lanjekar, D. Ranade and Y. Shouche, "Changes in human gut flora with age: an Indian familial study," *BMC Microbiology,* vol. 12, no. 1, p. 222, 2012.

[14] S. Chakravorty, D. Helb, M. Burday, N. Connell and D. Alland, "A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria," *J Microbiol Methods,* vol. 69, no. 2, pp. 330-339, 2007.

[15] L. C. Paulino, C.-H. Tseng, B. E. Strober and M. J. Blaser, "Molecular analysis of fungal microbiota in samples from healthy human skin and psoriatic lesions," *J Clin Micobiol,* vol. 44, no. 8, pp. 2933-2941, 2006.

[16] MO BIO Laboratories Inc, "UltraClean® Plant DNA Isolation Kit, Instruction Manual," MO BIO Laboratories, Inc, [Online]. Available: http://www.mobio.com/images/custom/file/protocol/13000.pdf.

[17] G. C. Baker, J. J. Smith and D. A. Cowan, "Review and re-analysis of domain-specific 16S primers," *Journal of Microbial Methods ,* vol. 55, no. 3, pp. 541-555, 2003.

[18] IBI Scientific, "Gel/PCR DNA Fragments Extraction Kit," IBI Scientific, 2014. [Online]. Available: https://www.ibisci.com/images/IB47010IB47020IB47030Protocol.pdf.

[19] G. Muyzer, E. C. De Waal and A. G. Uitterlinden, "Profiling of complex microbial populations by denaturing gradient gel Eelectrophoresis analysis of polymerase chain reaction-amplified

genes coding for 16S rRNA," *Applied and Environmental Microbiology,* vol. 59, no. 3, pp. 695-700, 1993.

[20] A. K. Bartram, M. D. Lynch, J. C. Stearns, G. Moreno-Hagelsieb and J. D. Neufeld, "Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads," *Appl Environ Microbial,* vol. 77, no. 15, p. 5569, 2011.

[21] "BaseSpace™: Getting Started," Illumina BaseSpace™ Developers, [Online]. Available: http://developer.basespace.illumina.com/docs/content/documentation/getting-started/overview.

[22] P. D. Schloss, S. L. Westcott, T. Ryabin, J. R. Hall, R. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley and D. H. Parks, "Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities," *Appl Environ Microbiol,* vol. 75, no. 23, pp. 7537-7541, 2009.

[23] R. C. Edge, "Search and clustering orders of magnitude faster than BLAST," *Bioinformatics,* vol. 26, no. 19, pp. 2460-2461, 2010.

[24] J. G. Caporaso, J. Kuczynski, J. Stombaugh and K. Bittinger, "QIIME allows analysis of high-throughput community sequencing data," *Nat Methods,* vol. 7, no. 5, p. 335–336, 2010.

[25] J. Oksanen, F. G. Blanchet, R. Kindt, P. Legendre, R. B. O'Hara and G. L. Simpson, "Package 'vegan'," 2011. [Online]. Available: http://CRAN.R-project.org/package=vegan.

[26] C. Lozupone, M. E. Lladser, D. Knights, J. Stombaugh and R. Knight, "UniFrac: an effective distance metric for microbial community comparison," *The ISME Journal,* vol. 5, no. 2, pp. 169-172, 2011.

[27] C. Lozupone and R. Knight, "UniFrac: a new phylogenetic method for comparing microbial communities," *Appl Environ Microbiol,* vol. 71, no. 12, p. 8228–8235, 2005.

[28] C. Lozupone, M. Hamady and R. Knight, "Unifrac - An online tool for comparing microbial community diversity in a phylogenetic context," *BMC Bioinformatics ,* vol. 7, no. 371, 2006.

[29] M. Hamady, C. Lozupone and R. Knight, "Fast Unifrac: faciliating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data," *The ISME Journal,* vol. 4, no. 1, pp. 17-27, 2009.

[30] T. Yatsunenko, F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras and M. Magris, "Human gut microbiome viewed across age and geography," *Nature,* vol. 486, no. 7402, pp. 222-227, 2012.