AIMS Environmental Science, 4(3): 443-455.
DOI: 10.3934/environsci.2017.3.443
Received: 31 January 2017
Accepted: 08 May 2017
Published: 11 May 2017

http://www.aimspress.com/journal/environmental

Research article

Analysis of human mitochondrial DNA sequences from fecally polluted environmental waters as a tool to study population diversity

Vikram Kapoor[1,2,*], Michael Elk[3], Carlos Toledo-Hernandez[4] and Jorge W. Santo Domingo[2,*]

[1] Department of Civil and Environmental Engineering, University of Texas at San Antonio, San Antonio, TX 78249, USA
[2] U.S. Environmental Protection Agency, Office of Research and Development, Cincinnati, OH 45268, USA
[3] Pegasus Technical Services, Inc., Cincinnati, OH 45268, USA
[4] Department of Biology, University of Puerto Rico, Rio Piedras, PR 00930, USA

* **Correspondence:** Email: vikram.kapoor@utsa.edu, santodomingo.jorge@epa.gov;
Tel: +210-458-7198.

Abstract: Mitochondrial signature sequences have frequently been used to study human population diversity around the world. Traditionally, this requires obtaining samples directly from individuals which is cumbersome, time consuming and limited to the number of individuals that participated in these types of surveys. Here, we used environmental DNA extracts to determine the presence and sequence variability of human mitochondrial sequences as a means to study the diversity of populations inhabiting in areas nearby a tropical watershed impacted with human fecal pollution. We used high-throughput sequencing (Illumina) and barcoding to obtain thousands of sequences from the mitochondrial hypervariable region 2 (HVR2) and determined the different haplotypes present in 10 different water samples. Sequence analyses indicated a total of 19 distinct variants with frequency greater than 5%. The HVR2 sequences were associated with haplogroups of West Eurasian (57.6%), Sub-Saharan African (23.9%), and American Indian (11%) ancestry. This was in relative accordance with population census data from the watershed sites. The results from this study demonstrates the potential value of mitochondrial sequence data retrieved from fecally impacted environmental waters to study the population diversity of local municipalities. This environmental DNA approach may also have other public health implications such as tracking background levels of human mitochondrial genes associated with diseases. It may be possible to expand this approach to other animal species inhabiting or using natural water systems.

Keywords: Human mitochondrial DNA; fecal pollution; environmental waters; high-throughput sequencing; population diversity

1. Introduction

Human mitochondrial DNA (mtDNA) has proven to be a useful tool for a variety of anthropological investigations such as forensics genetics, human evolutionary history, migration patterns, and population studies [1-5]. Sequence diversity within the mitochondrial D-loop hypervariable regions (HVR1 and HVR2) has been applied for this purpose since the level of polymorphism in these regions is high enough to permit its use as an important tool in population diversity studies [6,7]. However, most of these studies are based on an analysis of a controlled cohort of individuals which are randomly selected to be representative of the population of the geographical region of interest [8,9]. In this study, we present an alternate approach for the analysis of population diversity by targeting human mtDNA directly from environmental waters impacted by human contamination.

DNA is naturally shed into the environment by virtually all animal species through feces, urine, exudates, or tissue residues [10,11]. There are numerous sources of human mtDNA in environmental waters. These include fecal waste from combined sewer overflows (CSO), sanitary sewer overflows, household sewage treatment systems, and agriculture/urban runoff [12,13]. Human fecal waste has been shown to have large amounts of exfoliated epithelial cells, each cell harboring thousands of mitochondrial copies making mtDNA an adequate molecular target in environmental studies. Recently, several studies have taken advantage of human-specific mtDNA signature sequences to implicate human feces as the primary source of contamination in fecally-contaminated effluents [14-16]. Consequently, human mtDNA sequences obtained from environmental waters are reliable, quantitative and real-time indicators of diversity of the contributing populations.

With the exception of one study, the aforementioned studies have focused on the detection of mtDNA using qPCR assays to detect fecal pollution sources. Recently, Kapoor et al. [13] demonstrated the use of mtDNA sequence analysis to both determine the importance of specific human fecal pollution sources in an urban watershed (Cincinnati, OH), as well as the relative abundance of population haplogroups associated with the contributing populations. We hypothesize that human mtDNA sequences in sewage are reliable, quantitative, and real-time indicators of population diversity in a community. To describe, characterize, and track human population diversity in a watershed region, we used high-throughput DNA sequencing technology to profile the HVR2 sequences in water samples taken from a tropical watershed (Río Grande de Arecibo (RGA), Puerto Rico) impacted by human sewage. Like previous controlled population studies [6,9], the single-nucleotide polymorphisms (SNPs) present in HVR2 was used to differentiate populations on the basis of their frequencies of occurrence. Furthermore, we extracted haplotypes and assigned mitochondrial haplogroups to identify the mtDNA biological ancestry of the populations impacting the watershed. We demonstrate the potential of these data for surveying the distribution of population diversity in this region and their intersection with orthogonal data like U.S. Census data. These data establish a regional-scale, baseline population profile, which represents a unique metagenomics tool for studying population diversity, regional migration, and other anthropological investigations.

2. Materials and Methods

2.1. Study area and sampling sites

The Río Grande de Arecibo (RGA) watershed is located along the western-central part of Puerto Rico and has a catchment area of approximately 769 km^2, with water flowing northward from the central mountain range into a coastal valley before discharging into the Atlantic Ocean. Multiple point sources, including leaking septic and sewer systems and discharge from wastewater treatment plants (WWTPs) contribute to human fecal pollution in the watershed, in addition to several nonpoint sources associated with recreational activities. Three secondary sewage treatment plants discharge disinfected secondary effluents into the watershed: two drain into Río Cidra and Río Caunillas, tributaries of the RGA, while the third drains directly into the RGA. The water quality of the RGA watershed is a major concern as it is an important drinking water reservoir and some sections are used in recreational activities. Thus, fecal contamination of the RGA is a significant public health concern and has a negative economic impact. Most of the population in the RGA watershed is located in the coastal alluvial plain near the municipality of Arecibo [17]. The upper watershed is mostly forested, undeveloped land.

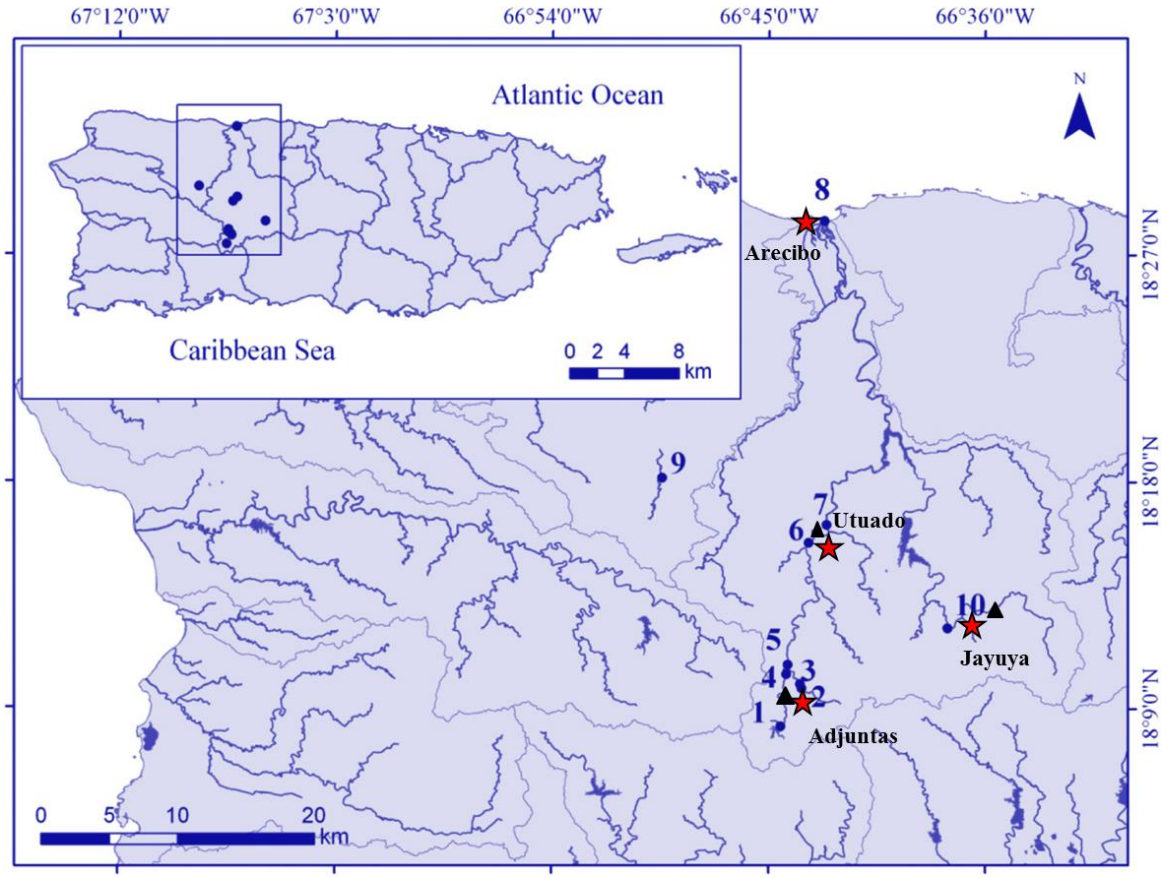

Figure 1. Location of sampling sites in Puerto Rico. Sites 4, 6, 7, 8 and 10 were used for sampling based on high levels of human fecal contamination. Wastewater treatment plants (WWTPs) are shown as black triangles and the major urban areas are highlighted by red stars.

The sampling sites (Figure 1) were identified and assessed for the presence of human fecal contamination through PCR-based detection of human fecal markers as described in a previous study [18]. These sites had a high human density based on previously recorded fecal pollution levels and potential impact from human fecal pollution via sewage overflow and watershed runoff [18]. Three sites (4, 7, and 10) were located downstream of a wastewater treatment plant (WWTP) for the municipalities of Adjuntas, Utuado, and Jayuya, respectively. Sites 6 and 7 represent sites before and after a WWTP. Site 6 is located approximately 1.62 km upstream from site 7, and site 7 is located 120 m downstream from the sewage treatment plant. Site 8 is located at the mouth of the watershed right before the RGA drains into the Atlantic Ocean and close to the town center of Arecibo.

2.2. Sample collection and DNA extraction

Ten samples (Table 1) were chosen from the water samples collected within the RGA watershed sites. The water samples collected within the RGA watershed represented different degrees of human contamination. Water sample collection and DNA extraction was performed as described earlier [18,19]. Briefly, all samples were collected using sterile bottles and transported on ice to the laboratory at the University of Puerto Rico—Río Piedras Campus where the samples (100 mL) were filtered through polycarbonate membranes (0.4-μm pore size, 47-mm diameter; GE Water and Process Technologies, Trevose, PA) and stored at −80 °C until DNA extraction. The membranes were shipped overnight on dry ice to the EPA laboratory (Cincinnati, OH) for DNA extraction. Total DNA was extracted from filters samples using the PowerSoil DNA isolation kit, following the manufacturer's instructions (Mo Bio Laboratories, Inc.). DNA extracts were stored at −20 °C until further processing.

Table 1. Description of samples collected in this study.

Sample	Site	Sampling Date	Location	Presumed human contamination source
1	7	6/10/2010	Downstream from Utuado WWTP	Sewage
2	7	10/28/2010	Downstream from Utuado WWTP	Sewage
3	7	5/27/2010	Downstream from Utuado WWTP	Sewage
4	8	11/12/2009	Mouth of Arecibo River	Urban runoff, recreation
5	8	9/23/2010	Mouth of Arecibo River	Urban runoff, recreation
6	8	10/24/2010	Mouth of Arecibo River	Urban runoff, recreation
7	4	11/23/2009	Downstream from Adjuntas WWTP, Cidra River	Sewage
8	4	5/27/2010	Downstream from Adjuntas WWTP, Cidra River	Sewage
9	6	10/28/2010	Upstream from Utuado WWTP	Septic tanks
10	10	11/12/2009	Downstream from Jayuya WWTP	Sewage

2.3. High throughput sequencing

The human mitochondrial hypervariable region II sequences were elucidated via Illumina sequencing of HVR2 libraries generated with DNA extracts and barcoded primers HVR2-F (5′-GGTCTATCACCCTATTAACCAC-3′) and HVR2-R (5′-CTGTTAAAAGTGCATACCGCC-3′) [13].

We generated PCR amplicon libraries for each water DNA extracts. PCR reactions were performed in 25 μL volumes using the Ex *Taq* kit (Takara) with 200 nM each of the forward and reverse primer and 2 μL of template DNA. Cycling conditions involved an initial 5 min denaturing step at 94 ℃, followed by 35 cycles of 45 s at 94 ℃, 60 s at 56 ℃, and 90 s at 72 ℃ and a final elongation step of 10 min at 72 ℃. Prior to multiplexed sequencing, PCR products were visualized on an agarose gel to confirm product sizes. Sequencing of the pooled library was performed on an Illumina Miseq benchtop sequencer using pair-end 250 bp kits at the Cincinnati Children's Hospital DNA Core facility. The HVRII sequence of the operator was also determined through Sanger sequencing and confirmed that it did not contribute to experimental data.

2.4. Bioinformatics analyses

All HVR2 sequences were sorted according to barcodes and grouped under their respective sampling event. The sequences were processed and cleaned using the software MOTHUR v1.25.1 [20]. Briefly, fastq files for forward and reverse reads were used to form contigs which were first screened for sequence length (no greater than 420 bp). To compensate for potential sequencing errors, sequences having an average quality under 20, having ambiguous bases (Ns), or being shorter than 300 bp were discarded. The quality-filtered sequences were then aligned to the revised Cambridge Reference Sequence (rCRS) [21] for human mitochondrial DNA (NC_012920.1| Homo sapiens mitochondrion, complete genome); and analyzed by using custom scripts to detect the SNPs present in the sequences. All SNPs with frequency greater than 5% were used for further analyses. Additionally, the sequences were exported to CLC Genomics Workbench Version 6.5 (CLC Bio, Cambridge, MA) and aligned to the rCRS, after which the Quality-based Variant Detection was called to detect insertions and deletions (indels) as well as SNPs with reference to the rCRS as described previously [13]. The mitochondrial genome databases, including MITOMAP [22], mtDB [23] and Phylotree [24] were referred to validate the occurrence of detected variants. Haplotypes were extracted and submitted to MITOMASTER version Beta 1 [25] to assign mitochondrial haplogroups based on variants present in HVR2.

3. Results and Discussion

3.1. Variant detection and frequency

In total, more than 100,000 sequence reads were retrieved with a mean output exceeding 20,000 per barcoded sample, which were then filtered and grouped according to their respective sampling events. HVR2 DNA from ten samples was sequenced and screened producing an average read length of approximately 423 bp. Of this, a 300 bp portion (i.e., from base position 50 to 350) was used for variant detection since SNPs in this region have been well documented [22]. A total of 19 distinct variants were detected with frequency greater than 5% of the total number of unique reads, all of which are present in MITOMAP—database of mtDNA Control Region Sequence Variants [22]. We observed some SNPs that were common to all samples with varying frequencies, while other SNPs were sample-specific, allowing each sample to have a unique human mtDNA signature in the form SNP allelic frequencies (Figure 2). The variation in SNP frequencies could be the result of several factors including limited sample size, population changes related to migration, changes in sampling time and storm runoff volumes during wet weather events, or a combination of them. Variants 73G and 263G were detected in all samples with high frequency (>90%), while variant 150T was detected

in all samples except the samples belonging to sites 4 and 8. Interestingly, variant 263G has been observed for mitochondrial genomes from European populations [26], and is compatible with the European ancestry that originated in the island over five centuries ago. All the variants detected at site 6 were also detected for samples belonging to site 7, except for 232G which was detected at site 6 with relatively low frequency. This is expected since sites 6 and 7 are located in close proximity to each other. Site 8 had two unique variants (67T, 81T) which may be attributed to the influx of water from several different tributaries, since site 8 is located right before the RGA drains into the Atlantic Ocean at sea level and it is the most downstream of all sampling sites.

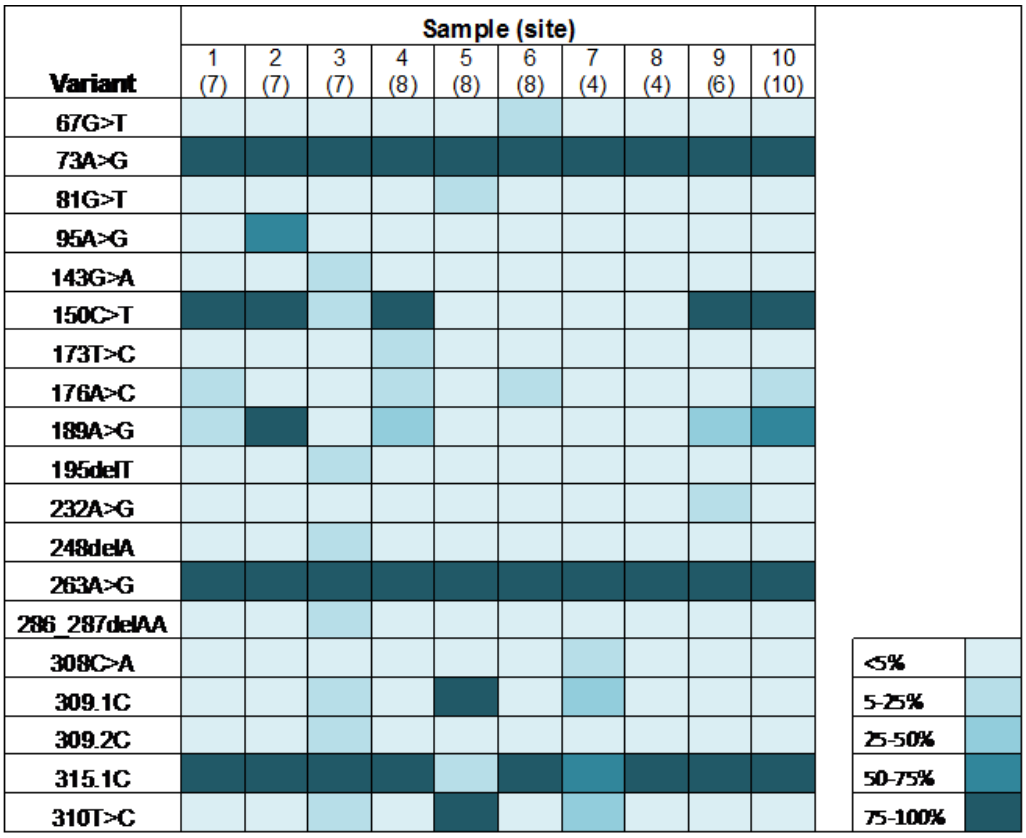

Figure 2. Heat map demonstrating the occurrence and frequency for variants detected in all samples (n = 10) in the human mitochondrial HVR2 region (position 50–350 bp relative to rCRS). Sampling sites are denoted within brackets next to the sample number. Variants are identified based on the revised Cambridge Reference Sequence (NC_012920.1| Homo sapiens mitochondrion, complete genome). Single nucleotide polymorphisms (SNPs) are denoted by ">" (73A>G means A is replaced by G at position 73). Insertions are denoted by "." followed by the number of nucleotides inserted at that position (309.1C means insertion of one C at position 309). Deletions are denoted by "del" followed by the nucleotides deleted (248delA means deletion of A from position 248).

3.2. Haplotypes and haplogroup classification

Mitochondrial haplogroups have arisen from mutation and migration during human evolution and largely correspond to the geographic regions of their origin [23,24]. These mitochondrial

haplogroups can be used to define ancestry based on the frequency of observation as a means of investigating population diversity [4,27]. For instance, there is broad correspondence between the L haplogroups and African ethnicity assignments, while the H haplogroups are most common among the Europeans. Consequently, we sought to use our human mtDNA sequences to extract haplotypes and classify them into haplogroups by comparing them to the Phylotree database [24]. We observed abundant diversity of haplotypes from HVR2 amplicons for all samples, which is consistent with the clear indication of human-associated pollution in the watershed [18]. The major haplotypes obtained from each sample are presented in Table 2.

Table 2. Major haplotypes detected in the samples.

Sample	Haplotype
1	73G, 150T, 263G, 315.1C; 73G, 150T, 189G, 263G, 315.1C; 73G, 150T, 176C, 263G, 315.1C; 73G, 150T, 263G; 73G, 150T, 176C, 189G, 263G, 315.1C; 73G, 150T, 315.1C
2	73G, 150T, 189G, 263G, 315.1C; 73G, 95G, 150T, 189G, 263G, 315.1C; 73G, 95G, 150T, 189G, 263G; 73G, 150T, 189G, 263G
3	73G, 263G, 315.1C; 73G, 150T, 263G, 315.1C; 73G, 263G, 309.1C, 310C; 73G, 263G, 310C 73G, 143A, 195delT, 248delA, 263G, 286delAA, 309.2C, 310C
4	73G, 150T, 263G, 315.1C; 73G, 150T, 189G, 263G, 315.1C; 73G, 150T, 173C, 263G, 315.1C; 73G, 150T, 176C, 263G, 315.1C; 73G, 150T, 176C, 189G, 263G, 315.1C; 73G, 150T, 263G; 73G, 150T, 189G, 263G
5	73G, 263G, 309.1C, 310C; 73G, 263G, 310C; 73G, 263G, 315.1C; 73G, 81T, 263G, 309.1C, 310C; 73G, 263G, 309.1C; 73G, 263G; 73G, 309.1C, 310C
6	73G, 263G, 315.1C; 73G, 263G, 388G; 67T, 73G, 263G, 315.1C; 73G, 176C, 263G, 315.1C; 73G, 263G
7	73G, 263G, 315.1C; 73G, 263G, 309.1C, 310C; 73G, 263G, 308A, 315.1C; 73G, 263G, 310C; 73G, 263G
8	73G, 263G, 315.1C; 73G, 263G
9	73G, 150T, 263G, 315.1C; 73G, 150T, 189G, 263G, 315.1C; 73G, 150T, 232G, 263G, 315.1C; 73G, 150T, 263G; 73G, 150T, 189G, 263G
10	73G, 150T, 263G, 315.1C; 73G, 150T, 189G, 263G, 315.1C; 73G, 150T, 176C, 263G, 315.1C; 73G, 150T, 176C, 189G, 263G, 315.1C; 73G, 150T, 263G; 73G, 150T, 189G, 263G

The mitochondrial sequences were compared and assigned to haplogroups based on the differences in HVR2 sequence mutations with respect to the rCRS. Since most accurate haplogroup

prediction is based on full mtDNA sequences, sequences were assigned to the closest haplogroup for which the HVR2 sequence contained all mutations that define the haplogroup. The most salient features of the haplogroup distribution (Figure 3) in the clustered sequences were the relatively high frequencies of haplogroup H (32%). This haplogroup is very common in Europe [28,29] and its presence in the mtDNA sequences from our study is in agreement with the presence of European population on the island. Other dominant haplogroups were T (25%), L (24%) and B (11%). As an additional verification step, HVR2 PCR products were cloned (TOPO TA Cloning Kit for Sequencing, Invitrogen, Carlsbad, CA) and 90 colonies were randomly picked and sent for Sanger sequencing. Nucleotide sequences were assembled and edited by using Sequencher 4.7 software (Gene Codes, Ann Arbor, MI) and analyzed for haplogroup prediction using MITOMASTER. Most of the sequences belonged to the haplogroup H (40%), followed by T (20%), L (12%) and B (10%). The results obtained with Sanger sequencing corresponded well with Illumina high-throughput sequencing supporting the reproducibility of the results by alternative sequencing methods.

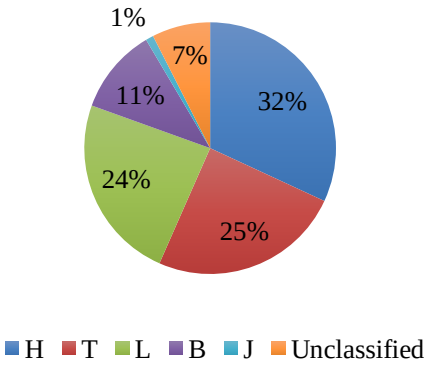

Figure 3. Pie chart showing the haplogroup distribution derived from the sequencing of HVR2 amplicons obtained from the water samples (n = 10) collected in the Río Grande de Arecibo Watershed in Puerto Rico.

3.3. Population diversity

To further explore the applicability of our HVR2-derived haplogroup data to local population diversity, several mitochondrial databases and studies were consulted to assign haplogroups to the general population groups found in Puerto Rico. We assigned haplogroups H, T and J to "West Eurasians"; haplogroup L to "Sub-Saharan African"; and haplogroup B to "American Indian" according to Martínez-Cruzado et al. [30]. Based on the average distribution of HVR2-derived population groups, most mtDNA haplogroups were identified as of West Eurasian ancestry (57.6%), followed by those of African (23.9%) and American Indian (11%) ancestries (Figure 4). According to U.S. census data for 2010 [31], populations belonging to these groups live in and around the study area. Figure 5 presents the comparative analysis of the population data obtained through the two strategies— census data for population (by race) viz-a-viz the HVR2-derived population groups for three different locations in the watershed. There was a strong correlation between the federal census data and the mitochondrial haplogroups as an indicator of population composition (Pearson product-moment correlation coefficient, $r = 0.9$) demonstrating the suitability of human mitochondrial sequences to infer the population structure of the neighborhoods impacting the watershed.

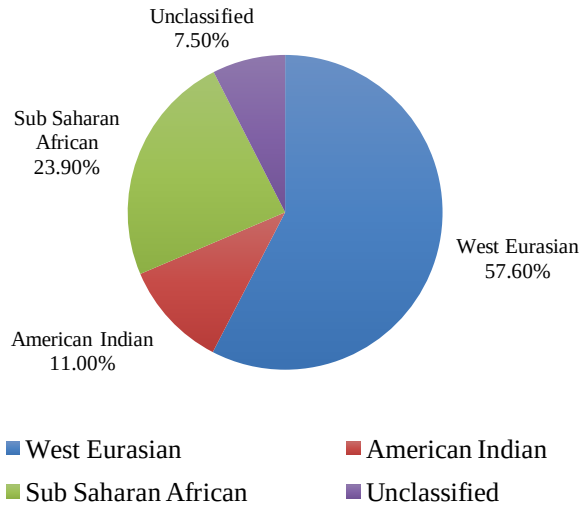

Figure 4. Pie chart showing the average population diversity of the sampling region (Río Grande de Arecibo Watershed in Puerto Rico) determined through the HVR2 derived haplogroups (n = 10).

While their relative abundance is different, the average census abundance patterns (White > African American > American Indian) are similar to our findings suggesting that the results correspond with the census data for population (by race). The mtDNA sequencing analysis suggests that American Indian ancestry is more prevalent than that the census data reports. Similarly, results from studies using HVR1 and other mtDNA-restriction profiles have also suggested that the presence of American Indian signals in Puerto Rico is more prevalent than previously considered [30], which is in agreement with our findings. The HVR2 motifs that are characteristic of the 'American Indian' haplogroups detected in this study are 73G, 143A and 263G (32) whereas Martínez-Cruzado (30) used predetermined restriction motifs as defining markers, along with HVR1 sequences to resolve inconclusive results. The latter approach to define haplogroups is more exact since it is based on haplogroup-defining markers for the entire mtDNA and not only just HVR2. However, classification of haplogroups based on analysis of small mtDNA regions with maximal discriminative power is useful for environmental studies due to concerns related to DNA damage in the environment. This approach has proven useful in past anthropological studies involving analysis of Neanderthal-type specimen to sequence small regions (300–350 bp) of Neanderthal mtDNA [33,34]. Deducing population diversity from mtDNA sequences retrieved from waste streams may be more accurate than census data since these are limited to people who respond to surveys and are subject to misclassification of self-declared racial/ethnic background while waste streams are impacted by everyone connected to the public sewer system. However, it is also possible that certain groups are overrepresented using the current approach either because they disproportionately use the water resources, are not connected to the sewer pipelines (e.g., use of septic tanks) and/or live in close proximity to sampling sites than others. Signature sequences from areas impacted by leaky septic tanks and combined sewer overflows will also be reflected in these types of molecular surveys. While further studies are needed to better understand how all these different sources may impact haplogroup distribution, we suggest that the use of these methods could provide complementary information in epidemiological studies.

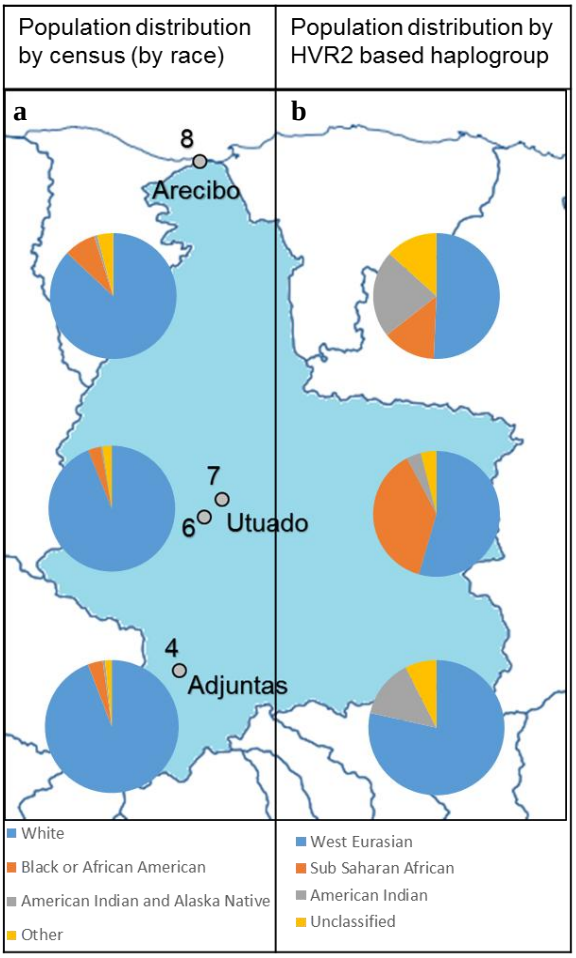

Figure 5. Pie charts demonstrating the population racial diversity in three different municipalities in the Río Grande de Arecibo Watershed obtained through (a) 2010 population census data (by race), and (b) annotation of HVRII sequences (obtained during 2009-10) into haplogroups. Site 8 is located in Arecibo; sites 6 and 7 are at Utuado; and site 4 is at Adjuntas.

The overall bioinformatics strategy in this study included the following steps: (i) trim/clean sequencing reads and group them according to sites, (ii) map sample specific reads to the rCRS, (iii) annotate the mapped sequences to detect variants in HVR2 region, and (iv) extract haplotypes from individual reads and assign haplogroups based on HVR2 sequence motifs. As reported here, next-generation sequencing technology of the mitochondrial hypervariable sequences enabled the identification of a great number of mtDNA variants and at varied allele frequencies. The methods used in our study for haplogrouping uses only HVR2 sequences which may result in coarse haplogroup assignments. Since most accurate haplogroup prediction is based on full mtDNA sequences, sequences were assigned to the closest haplogroup for which the HVR2 sequences contain all SNPs that define the haplogroup. We believe that future global sequencing efforts associated with distinct populations will provide improved phylogenetic resolution of the human mtDNA hypervariable regions as a tool for defining genetic ancestry. It has not escaped our attention that extending our technique to include other mtDNA regions and/or assembling full mitochondrial genomes through metagenomics approaches on a massively parallel scale would allow for tracking humans through

public waste streams, thus ethical concerns remain an important consideration in future work.

Mitochondrial DNA analysis has been applied in several biomedical investigations of human evolution, for example, studies tracing the origin of modern humans or of certain human populations. In addition, mtDNA analysis is extremely effective in a forensic setting for the identification of criminals and victims of crimes or accidents. Although our study was confined to analysis of HVRII region of human mtDNA for samples collected in a limited number of geographic locations, it can be inferred that by targeting specific regions of mtDNA, we can estimate cancer rates, occurrence of diseases, and population diversity in watershed regions impacted by human contamination. Moreover, we envision that a similar approach could be used to study the population diversity of different animal species in natural settings, such as local versus migratory birds.

4. Conclusions

We investigated the occurrence of HVR2 allelic frequencies of human mtDNA derived from water samples taken within a fecally impacted tropical watershed. The SNPs within the human HVR2 sequences represented a unique molecular signature for evaluating anthropogenic site-specific inputs. We observed several HVR2 haplotypes linked to these samples, and used this haplogroup data to derive human population diversity within the different sites of the watershed. There was a strong correspondence between the demographic census data and the population composition based on mitochondrial haplogroups, demonstrating the suitability of human mitochondrial sequences to infer the population structure of the neighborhoods impacting the watershed. As the levels of human mtDNA is significantly high in point and non-point sources of fecal pollution, detecting mtDNA allelic signatures in environmental waters provides a unique approach for simultaneously studying fecal waste source tracking, human population diversity and other many anthropological investigations.

Acknowledgements

We would like to thank Mehdi Keddache for help in data analysis, and David Wendell for providing access to the CLC Genomics program. VK was supported by U. S. Environmental Protection Agency (EPA) via a post-doctoral appointment administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. EPA. The manuscript has been subjected to the EPA's peer review and has been approved as an EPA publication. Mention of trade names or commercial products does not constitute endorsement or recommendation by the EPA for use. The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. EPA.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. Byrne EM, McRae AF, Zhao ZZ, et al. (2008) The use of common mitochondrial variants to detect and characterise population structure in the Australian population: implications for genome-wide association studies. *Eur J Hum Genet* 16: 1396-1403.

2. Bonatto SL, Salzano FM (1997) A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci USA* 94: 1866-1871.
3. Wallace DC (1994) Mitochondrial DNA sequence variation in human evolution and disease. *Proc Natl Acad Sci USA* 91: 8739-8746.
4. Wallace DC, Brown MD, Lott MT (1999) Mitochondrial DNA variation in human evolution and disease. *Gene* 238: 211-230.
5. Wilson MR, DiZinno JA, Polanskey D, et al. (1995) Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int J Leg Med* 108: 68-74.
6. Salas A, Lareu V, Calafell F, et al. (2000) mtDNA hypervariable region II (HVII) sequences in human evolution studies. *Eur J Hum Genet* 8: 964-974.
7. Baasner A, Schäfer C, Junge A, et al. (1998) Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. *Forensic Sci Int* 98: 169-178.
8. Johnson DC, Shrestha S, Wiener HW, et al. (2015) Mitochondrial DNA diversity in the African American population. *Mitochondrial DNA* 26: 445-451.
9. Comas D, Reynolds R, Sajantila A (1999) Analysis of mtDNA HVRII in several human populations using an immobilised SSO probe hybridisation assay. *Eur J Hum Genet* 7: 459-468.
10. Shokralla S, Spall JL, Gibson JF, et al. (2012) Next-generation sequencing technologies for environmental DNA research. *Mol Ecol* 21: 1794-1805.
11. Ficetola GF, Miaud C, Pompanon F, et al. (2008) Species detection using environmental DNA from water samples. *Biol Lett* 4: 423-425.
12. Glassmeyer ST, Furlong ET, Kolpin DW, et al. (2005) Transport of chemical and microbial compounds from known wastewater discharges: potential for use as indicators of human fecal contamination. *Environ Sci Technol* 39: 5157-5169.
13. Kapoor V, DeBry RW, Boccelli DL, et al. (2014) Sequencing human mitochondrial hypervariable region II as a molecular fingerprint for environmental waters. *Environ Sci Technol* 48: 10648-10655.
14. Kapoor V, Smith C, Santo Domingo JW, et al. (2013) Correlative assessment of fecal indicators using human mitochondrial DNA as a direct marker. *Environ Sci Technol* 47: 10485-10493.
15. Martellini A, Payment P, Villemur R (2005) Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine and ovine sources in fecally contaminated surface water. *Water Res* 39: 541-548.
16. Caldwell JM, Raley ME, Levine JF (2007) Mitochondrial multiplex real-time PCR as a source tracking method in fecal-contaminated effluents. *Environ Sci Technol* 41: 3277-3283.
17. Martinuzzi S, Gould WA, González OMR (2007) Land development, land use, and urban sprawl in Puerto Rico integrating remote sensing and population census data. *Landscape Urban Plan* 79: 288-297.
18. Toledo-Hernandez C, Ryu H, Gonzalez-Nieves J, et al. (2013) Tracking the primary sources of fecal pollution in a tropical watershed in a one-year study. *Appl Environ Microbiol* 79: 1689-1696.
19. Kapoor V, Pitkänen T, Ryu H, et al. (2015) Distribution of human-specific bacteroidales and fecal indicator bacteria in an urban watershed impacted by sewage pollution, determined using RNA-and DNA-based quantitative PCR assays. *Appl Environ Microbiol* 81: 91-99.

20. Schloss PD, Westcott SL, Ryabin T, et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75: 7537-7541.
21. Andrews RM, Kubacka I, Chinnery PF, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23: 147-147.
22. Brandon MC, Lott MT, Nguyen KC, et al. (2005) MITOMAP: a human mitochondrial genome database—2004 update. *Nucleic Acids Res* 33: D611-D613.
23. Ingman M, Gyllensten U (2006) mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res* 34: D749-D751.
24. van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30: E386-E394.
25. Brandon MC, Ruiz-Pesini E, Mishmar D, et al. (2009) MITOMASTER: a bioinformatics tool for the analysis of mitochondrial DNA sequences. *Hum Mutat* 30: 1-6.
26. Röck AW, Dür A, van Oven M, et al. (2013) Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). *Forensic Sci Int Genet* 7: 601-609.
27. Lee C, Măndoiu II, Nelson CE (2011) Inferring ethnicity from mitochondrial DNA sequence. *BMC Proc* 5: S11.
28. Richards M, Macaulay V, Hickey E, et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67: 1251-1276.
29. Roostalu U, Kutuev I, Loogväli EL, et al. (2007) Origin and Expansion of Haplogroup H, the Dominant Human Mitochondrial DNA Lineage in West Eurasia: The Near Eastern and Caucasian Perspective. *Mol Biol Evol* 24: 436-448.
30. Martínez-Cruzado JC, Toro-Labrador G, Viera-Vera J, et al. (2005) Reconstructing the population history of Puerto Rico by means of mtDNA phylogeographic analysis. *Am J Phy Anthropol* 128: 131-155.
31. U.S. Census Bureau 2010 Census of Population and Housing, Summary Population and Housing Characteristics, CPH-1-53, Puerto Rico U.S. Government Printing Office, Washington, DC.
32. Guardado-Estrada M, Juarez-Torres E, Medina-Martinez I, et al. (2009) A great diversity of Amerindian mitochondrial DNA ancestry is present in the Mexican mestizo population. *J Human Genetics* 54: 695-705.
33. Krings M, Geisert H, Schmitz RW, et al. (1999) DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. *Proc Nat Acad Sci* 96: 5581-5585.
34. Ovchinnikov IV, Götherström A, Romanova GP, et al. (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404: 490-493.

AIMS Press