THE PERSISTENCE, DEGRADATION, AND APPLICATION OF HUMAN DNA RECOVERED FROM SOIL

A thesis presented to the faculty of the Graduate School of Western Carolina University in partial fulfillment of the requirements for the degree of Master of Science in Biology

By

Hannah Lynn Noel

Director: Dr. Frankie West
Assistant Professor of Forensic Science
Department of Chemistry and Physics

Committee Members: Brittania Bintz, Chemistry and Physics
Dr. Rebecca George, Anthropology
Maureen Hickman, Biology
Dr. Angela Dautartas, Anthropology

July 2023

ACKNOWLEDGEMENTS

I would like to thank my wonderful thesis advisor, Dr. Frankie West, for her support and encouragement to pursue my doctorate. I would also like to thank my thesis committee, Dr. Rebecca George, Britt Bintz, Maureen Hickman, and Dr. Angela Dautartas, for all their effort to help me succeed. Additionally, thank you to Dr. Tom Martin for his assistance with statistical analysis. I appreciate all the faculty who have helped me learn to be a better researcher and dedicated their time to this project. I also would like to extend my gratitude towards the forensic anthropology students that assisted in the placement, photography, and recovery of donors at the Forensic Osteology Research Station during the Summer of 2022.

I express my deepest gratitude to the donors and the families of the donors whose remains were critical for the success of my research. Their donation is a selfless act that has helped my project and others like it, and I am grateful for the privilege of working with them.

I would like to say thank you for the financial support from the Forensic Science Program at WCU, as well as WCU's Summer 2022 & 2023 Graduate Research Assistantship Grants. Without these resources, none of my research would have been possible.

Finally, I would like to thank my family, friends, and graduate school cohort for their words of encouragement and for being there to support me through this project when I needed them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADD: accumulated degree days
BSA: bovine serum albumin
CDI: cadaver decomposition island
CODIS: Combined DNA Index System
Ct: Cycle Threshold
Ct-SD: Cycle threshold standard deviation
DI: degradation index
DPMA: DNA primer mix A
DPMB: DNA primer mix B
FOREST: Forensic Osteology Research Station
gDNA: genomic DNA
IPC: internal positive control
iSNP: identifying single nucleotide polymorphism
MFCS: MiSeq FGx™ control software
mtDNA: mitochondrial DNA
mtGenome: mitochondrial genome
NGS: next-generation sequencing
NOAA: National Oceanic Atmospheric Administration
nuDNA: nuclear DNA
PMI: post-mortem interval
qPCR: quantitative PCR
rCRS: revised Cambridge Reference Sequence
SNP: single nucleotide polymorphism
STR: short tandem repeat
TBS: total body score
UAS: universal analysis software
WGS1: whole genome mix set 1
WGS2: whole genome mix set 2

LIST OF EQUATIONS

ABSTRACT

THE PERSISTENCE, DEGRADATION, AND APPLICATION OF HUMAN DNA RECOVERED FROM SOIL

Hannah Noel, M.S.

Western Carolina University (July 2023)

Director: Frankie West, Ph.D.

The dynamics of the persistence of human DNA in soil are poorly understood, though increasingly crucial to genomics. Human DNA recovered from soil and sediments can be used in identifying the presence of human burial and decomposition sites as well as in paleoanthropological contexts to support phylogenetic research. To date, studies on human DNA persistence in soil have focused on the recovery and quantity of DNA in soil rather than the taphonomic processes underlying its persistence. Given these gaps in knowledge on how environmental factors influence the persistence of human DNA in soil, this experiment aims to discover trends between extrinsic factors (e.g., temperature, precipitation, body moisture) and the quantity and quality of DNA recoverable from soil following the surface-level decomposition of three body donors for eleven weeks. Nuclear DNA (nuDNA) was quantified via Quantifiler™ Trio DNA Quantification Kit while mitochondrial DNA (mtDNA) was quantified using the Kavlick triplex qPCR protocol to determine the duration of DNA persistence in soil, determine quality through degradation indices, and to determine if environmental factors had an effect on the quantity of DNA within the soil. This experiment also explores the potential application of soil-derived in human identification work by using next-generation sequencing (NGS) to analyze concordance between sample and donor reference profiles. nuDNA was sequenced using the

ForenSeq™ DNA Signature Prep Kit and the mitochondrial genome was sequenced using the ForenSeq™ mtDNA Whole Genome Kit, which were both ran on the MiSeq FGx™ Sequencing System and compared to donor reference profiles. Current results suggest that human nuDNA can persist in small quantities within the soil for up to six weeks, while mtDNA is able to persist for the entire eleven-week period. mtDNA was also a more viable route for human identification via soil extracts, as haplogroup determinations coincide with donor reference haplogroups for at least three weeks into the decomposition period. Time into the decomposition period, temperature, and total body score all have a significant effect on the quantity of DNA within the soil, though some results suggest moisture may play a role in DNA persistence and degradation. However, further research needs to be done at multiple locations across various seasons to grasp the nuance behind the role of environmental factors in the mechanisms driving DNA persistence within soil. The results of this aim to encourage future works involving the persistence of DNA in soil and its utility in forensic and anthropological contexts.

CHAPTER ONE: INTRODUCTION

Human decomposition studies have been a fundamental part of forensic anthropological research since the 1980s.[1] Moreover, the study of the taphonomic processes involved in human decomposition holds relevance for a variety of forensic scientific applications, including forensic entomology, forensic botany, geology, and forensic DNA analysis. Various factors, including the surrounding environment, can influence the rate of decomposition, and taphonomic changes that human remains are subjected to. Understanding the effects of these complex biotic and abiotic interactions are essential to making accurate estimations of time since death and informed decisions on how to process remains for DNA analysis, including appropriate DNA extraction methods. Scientists have developed models to predict the post-mortem interval (PMI) via accumulated degree days (ADD) by combining climatic factors with defined phases of decomposition: fresh, bloat, active decay, advanced decay, and skeletonization.[2] Since decomposition relies on environmental factors, it is critical to understand the mechanisms underlying these interactions to develop future methods and models to be used in forensic contexts.

During the active decay stage of human decomposition, the body will purge decomposition fluids. These fluids are rich in biological macromolecules such as lipids, proteins, nucleic acids, as well as organic compounds, and human-associated microbes. These fluids, like the remains, are subject to interactions with the surrounding environment. In soil, a cadaver decomposition island (CDI) is formed from the pulse of water, carbon, and other nutrients from the decomposing cadaver. The soil within the CDI will undergo changes in chemistry and physiology.[3] These changes include shifts in the taxonomic abundance of soil microbial species,

pH reduction, and the quantity of available carbon and nitrogen.[4-6] Microbial species will be introduced to the soil, such as human gut-associated *Bacteroides*, which can survive in the soil for up to seven months.[4,6-7] Shifts in the microbial community structure will perpetuate the changes in soil pH and nutrient availability for extended periods. Soil chemistry, along with the relative moisture level and temperature in an environment, are all known to contribute to the rate of decomposition and taxonomic change seen in remains.[3] Therefore, these factors will likely influence the persistence of biological macromolecules that infiltrate the soil.

Complex chemistry occurs in the soil surrounding the burial environment and surface-level decomposition. During active decay, autolysis of cells and subsequent degradation of their constituents occurs primarily through the activity of hydrolytic enzymes.[8] For instance, nuclease enzymes called DNases will cleave double-stranded nuclear DNA (nuDNA) into smaller duplexes until the strand is thoroughly degraded.[9] These DNase enzymes are found in the purged decomposition fluids and are produced by some soil microbes. Outside the body and away from DNA repair systems, extracellular DNA in the soil is also vulnerable to oxidation. Oxidation of DNA causes alterations of DNA base pairs, particularly in the imidazole ring of guanine, due to its low redox potential, causing single or double-stranded breaks.[10] Both of these mechanisms result in the degradation of DNA over time and, thus, a loss in genetic information.

However, in some circumstances, DNA may persist in sediments for thousands of years. For instance, ancient DNA from Neanderthals and Denisovans has been extracted and sequenced using Next Generation Sequencing (NGS) techniques from various cave environments.[11-14] Though heavily degraded, this ancient DNA still yielded analyzable mitochondrial DNA (mtDNA) sequences. Up to 73% of these sequences can be assigned to the Neanderthal genome.[12] This is not a stand-alone case, as the mitochondrial genome of a Middle Pleistocene

cave bear (*Ursus deningeri*) has been sequenced from fossils and sediment at Sima de los Huesos in the Sierra de Atapuerca, Spain. These fragments were all under 50bp regardless of historical climatic data revealing that this DNA had been preserved outside permafrost, suggesting mtDNA's resilience in sediments.[15] MtDNA is more resistant to degradation than nuDNA due to its circular structure which limits nuclease activity. Additionally, mtDNA has high copy number per cell, making it more abundant and more likely to persist than nuDNA.

Different qualities of the soil and sediments may also factor into how well DNA is preserved. Since DNA can be oxidized, resulting in single or double-stranded breaks,[10] soil porosity could be an essential factor in DNA persistence. Some research has shown that light-texture soils with fine particle size, such as clay, can stabilize organic compounds due to their small pore size minimizing oxidation and high surface area allowing for increased cationic binding to the soil matrix.[16] Similarly, soil type has also been cited as an essential factor in determining the effects of the decomposition of pig remains in burial contexts. Decomposition of pig carcasses was shown to occur at faster rates in loamy and organic soils than in clayey or sandy soils. These differences in decomposition rates could be due to the relationship between matric potential to retain moisture and microbial activity within soil types, along with soil loamy and organic soils displaying differences in insect genera present.[17] This is further supported by the ancient hominin DNA recovered from Denisovan Cave sediments, as the microstratigraphic layers of sediments were primarily comprised of clay and fine silts[11], which may have provided a suitable environment for DNA preservation.

Soil temperature and moisture also play a role in autolytic processes that can break down DNA, as they can contribute to variance in microbial activity. Putrefaction of human remains carried out by anaerobic bacteria in the burial environment occurs optimally between

temperatures of 21°C and 38°C and is reduced at temperatures below 10°C or above 40°C.[8] Similarly, studies of surface-level decomposition of pig carcasses have revealed that there is greater bacterial and eukaryotic diversity within carcass-enriched soils in the warmer summer months.[18] Therefore, there appears to be a trend of increased microbial activity, and potentially nuclease activity, within warmer soils. Wetter soil also tends to show greater rates of decomposition and microbial activity. In a study on the decomposition of juvenile rat cadavers in three different soil types, they found that cadavers in soils with lower matric potentials and more moisture present had greater rates of decomposition than those in drier soil. However, the optimal matric potential for cadaver decomposition was exceeded in soil with higher clay content, as the wettest soil type slowed the decomposition process.[19] Thus, moist soils may contribute to higher microbial degradation of human remains and extracellular DNA so long as optimal growth conditions for decomposition microorganisms are not exceeded.

Chemical interactions between molecules are another mechanism that allows for the persistence of extracellular DNA within soil. Since DNA has a negatively charged sugar-phosphate backbone, it can form ionic bonds with positively charged cations in the soil.[9] These bonds between the cations and adsorbed extracellular DNA can persist for extended periods. Physical characteristics of the soil particles can also aid in the formation of micro-aggregates and organomineral complexes that can slow the oxidation of extracellular DNA.[20] Furthermore, cation bridges can form between humic acids within the soil and DNA, deterring microbial degradation.[9]

Microbial activity is a primary contributor to the degradation of DNA in soil, as many microbes can produce DNase enzymes for metabolic purposes, which shear DNA.[21] However, as previously stated, the formation of cation bridges between humic acid and DNA can prevent

degradation. One study found that extracellular plasmid DNA from the soil bacteria *Bacillus subtilis* was rapidly adsorbed in clay and sandysoils high in humic acid content while maintaining genetic integrity. Furthermore, the extracellular *B. subtilis* plasmids bound to humic acid still maintained the ability to be transformed into competent cells *in vitro.*[22] This suggests that soils with high humic acid concentrations may be better at preserving human mtDNA from degradation, since mtDNA is similar in structure to plasmid DNA. The soil microbiome also plays a role in determining the physiology of the soil. The thanatomicrobiome, or postmortem microbial community of the human body associated with decomposition, follows general trends in microbial succession that alter the soil microbiome for extended periods.[8] Enzyme activities related to the decomposition of organic materials are shown to be strongly affected by soil pH and moisture, as they can lead to temporal shifts in bacterial abundance and diversity.[23] Therefore, DNA's persistence in soil is a nuanced mechanism that relies on microbial, chemical, physical, and climatic interactions.

A study by Emmons et al. tracked the short-term persistence of nuDNA and mtDNA during surface-level decomposition from fresh, bloat, active decay, advanced decay, and dry/remains.[24] Quantitative data of DNA extracted from the soil elucidated that mtDNA was generally more recoverable than nuDNA and that both were most abundant during active and advanced stages of decomposition when the body was purging. It was postulated that the acidic clay loam soil on-site may have influenced the ability of DNA to bind to the soil matrix.[23] This aligns with the idea that soil micro-aggregates and porosity may influence the oxidative decomposition of DNA and DNA persistence in soil. However, no concordance data between donor and soil samples were provided, nor did the study account for potential climatic factors that may affect the persistence of DNA in the soil.

This project aims to expand upon Emmons et al. by monitoring the persistence of DNA in soil following surface-level decomposition in conjunction with climatic data to understand general trends in DNA degradation and potential factors affecting its persistence. Since many potential environmental factors are involved in DNA preservation in soil, this study will focus primarily on temperature, humidity, rainfall, and body moisture levels. Temperature and moisture are also significant factors that contribute to increased microbial activity, which may lead to further nuclease activity in the soil. Additionally, donor profiles will be compared to all DNA recovered from the soil for concordance data to determine whether DNA recovered from soil can produce identifying DNA profiles.

NGS methods will be used to produce DNA profiles from soil samples due to their high sensitivity and ability to enable sequencing of heavily degraded DNA versus Sanger sequencing. Unlike Sanger sequencing which only sequences a single DNA fragment at a time, NGS is massively parallel for the sequencing of millions of fragments simultaneously. Furthermore, NGS methods can be suitable for samples leading to low input of DNA template. In a study using the ForenSeq™ DNA Signature Prep Kit, which targets over 200 markers, a comparison of 223 samples with DNA concentrations varying from 1 ng/µl – 125 pg/µl showed complete single nucleotide polymorphism (SNP) genotypes were produced in 87% of samples, with 99.6–99.9% genotype concordance to reference.[25] Though lower quantities of input DNA resulted in the loss of SNPs, as more than 50% of calls were missing if samples contained concentrations of 7.82 pg/µl nuDNA[25], this is still valuable information for developing a biological profile for positive human identification. Since the amplicons of the identity informative, ancestry, and phenotypic SNPs in ForenSeq™ DNA Signature Kit range from 63-261 bp in length, this allows for the detection of identifying loci in heavily fragmented nuDNA.[26] The ForenSeq™ DNA Signature

kit also targets 27 autosomal STRs, 24 Y-STRs, and 7 X-STRS with amplicon lengths ranging from 61-402 bp in length, providing sex estimation and identity informative information, including all 20 Combined DNA Index System (CODIS) STR loci.[26] Though these are longer fragments of DNA that may not be as easily detected in heavily degraded DNA samples, it is important to see how long these loci can be detected in human decomposition-enriched soils to determine the utility of soil-derived DNA for forensic casework contexts.

For sequencing mtDNA, the ForenSeq™ mtDNA Whole Genome Kit will be used, as it can sequence the entire 16,569 bp mtGenome.[27] To improve recovery of mtDNA from heavily degraded samples, the kit contains small amplicon sizes of less than 210 bp with a minimum overlap of 3 bp, allowing for sequencing of the entire mitochondrial genome (mtGenome) even when fragmented.[27] To further support this, the ForenSeq™ mtDNA Whole Genome Kit has been validated for sensitivity and reproducibility, achieving 97.9% coverage of the mitochondrial genome in dilutions of HL60 human positive control DNA containing 2 pg/µl genomic DNA (gDNA).[28] Due to the high sensitivity and ability to sequence short targets, NGS methods are appropriate for heavily degraded DNA samples derived from soils surrounding surface-level decomposition.

Understanding the dynamics underlying the persistence of DNA in soil has the potential to support forensic investigations and anthropological research. The ability to recover DNA from soil can be beneficial in identifying clandestine graves for forensic contexts. It could also be an alternative form of DNA sampling when few remains are present or destructive sampling of remains is not possible due to legal or ethical reasons. However, more research needs to be performed to determine the mechanisms behind DNA persistence in soil and its utility in forensic investigation so that it is generally accepted by the scientific community and admissible for

casework.[29] Furthermore, research on DNA preservation within sediments could benefit paleoanthropological research on understanding ancient hominin relationships. This project aims to address some of these knowledge gaps and act as a pilot study to encourage further research on the persistence of extracellular human DNA in soil.

CHAPTER TWO: MATERIALS AND METHODS

Donors and Sampling Location

Samples were collected from donated human cadavers and surrounding soil in Western Carolina University's Forensic Osteology Research Station (FOREST), an outdoor human decomposition facility located in Cullowhee, North Carolina. The FOREST has two separate enclosures; enclosure one is reserved for surface-level decomposition studies, while enclosure two is used for buried remains studies. For this research, all donors were placed in enclosure one for surface-level decomposition. Enclosure one is approximately 5,000 sq feet and sloped, with the top of the slope being northeast, with partial canopy coverage. Donors were refrigerated until delivered to the facility. Each individual received a unique identification number upon being received at the facility. The donors used for this study will be referred to as Donors 1, 2, and 3, respectively, to protect anonymity. All donors used in this study either personally or had next-of-kin provide additional consent for genetic studies to be performed. All donors were elderly, of self-identified European ancestry, and died of natural causes. The sex of Donor 1 was female, and Donors 2 and 3 were male. Donor 1 had a puncture from a feeding tube port on their lower left abdomen and an open sore on their right lower jawline. These antemortem wounds are of note as open wounds may be the first target for scavengers within a short period after placement.[30] Donor 2 had been used for cadaver dog training before being placed in their assigned plot. There were no open wounds or special uses for Donor 3 prior to their placement. Each donor was unclothed on the ground, in a supine position with their mouths open. The positioning of arms varied between donors. Donor 1 had arms crossed over their torso, Donor 2 had arms resting at their sides, and Donor 3's hands were resting on their hips. No scavenger

barriers were placed around the donor plots, aside from the wooden privacy fence and a chain-linked fence topped with razor wire surrounding the entire enclosure. The donors were placed in three different locations within the facility. Donor 1 was placed at the top of the slope in the facility with partial canopy cover, leaf litter, weedy vegetation, and red-clayey soil. Donor 2 was placed midway down the slope with some canopy cover, minimal leaf litter, dark clayey soil, and roughly a meter away from the mummified remains of another donor. Donor 3 was placed at the bottom of the slope to the right of the facility entrance with minimal canopy cover, weedy vegetation, and dark clayey soil. A trail camera was set at the foot of each body to monitor taphonomic scavenging activity.

Table 1. Donor sampling schedule for three human cadaveric subjects to study the persistence of DNA in soil following surface-level decomposition. Total Body Score (TBS) was determined using Megyesi et al.[2] scoring metrics. Highlighted samples denote that the body had been covered with a black plastic tarp to discourage the migration of remains through animal scavenging.

Donor ID	Sex	Date of Death	Sampling Event	Sampling Date	TBS
1	F	5/10/2022	0	5/12/2022 (Placement)	3
			1	5/18/2022	23
			2	5/25/2022	27
			3	6/1/2022	31
			4	6/8/2022	31
			5	6/15/2022	30
			6	6/22/2022	31
			7	6/29/2022	31
			8	7/6/2022	31
			9	7/13/2022	32
			10	7/20/2022	32
			11	1/27/2022	32
2	M	5/8/2022	0	5/13/2022 (Placement)	5
			1	5/18/2022	19
			2	5/25/2022	24
			3	6/1/2022	26
			4	6/8/2022	27
			5	6/15/2022	28
			6	6/22/2022	28
			7	6/29/2022	28
			8	7/6/2022	28
			9	7/13/2022	28
			10	7/20/2022	28
			11	7/27/2022	30
3	M	5/24/2022	0	5/27/2022 (Placement)	3
			1	6/1/2022	19
			2	6/8/2022	22
			3	6/15/2022	24
			4	6/22/2022	24
			5	6/29/2022	24
			6	7/6/2022	26
			7	7/13/2022	28
			8	7/20/2022	28
			9	7/27/2022	28
			10	8/3/2022	29
			11	8/10/2022	29

Sample Collection and Sampling Frequency

Samples consisted of 10mL of soil collected using Bel-Art Sterileware® Sampling Spatulas and 50mL conical tubes that had been sterilized via UV for 30 minutes before use. Organic matter and leaf litter were removed from the top layer of soil before being collected. Each tube was sealed using parafilm and stored at -80°C for later use. Four baseline soil samples were taken from each donor plot before placement, along with two general samples from the facility. One of the facility baseline soil samples was taken from the top of the slope, while the other was taken from the bottom, both in areas that had not been used as plots for several months to years. During decomposition, sampling occurred weekly from four regions around each donor: the left of the cranium, either side of the torso, and the right lower extremity. This was continued for eleven weeks for each donor. The positioning of Donor 1 was greatly affected by scavenging, causing the upper body and cranium to migrate below the lower extremities. To accommodate this positionality, samples from weeks four through eleven were taken from general quadrants of the plot: one towards the top where the cranium had been, one from each side of the center of the plot, and one towards the bottom of the plot where the right extremity had been.

Upon intake, two buccal swabs were collected with sterilized flocked swabs from each donor, along with hair samples with the roots intact. The swabs were returned to their sterile packaging, and the hair was placed in a paper envelope. Both were packaged in a plastic resealable bag for transport and then stored at -80°C. Donor reference profiles were obtained to compare sequence concordance with soil-derived DNA profiles.

Total Body Score, Accumulated Degree Days, and Taphonomic Change

Weekly assessments of decomposition were performed using TBS and accumulated degree days (ADD). TBS for each donor was determined using scoring metrics of the head,

torso, and limbs as outlined by Megyesi et al.[2], arranged in Table 2. ADD was calculated using Equation 1. Disarticulation and scavenging activity were documented via photography and trail camera footage of each donor.

Table 2. Weekly TBS and ADD of donor individuals across 11 weeks of surface-level decomposition.

Donor ID	Date Recorded	Head	Torso	Limbs	TBS	ADD
1	5/12/2022 (Intake)	1	1	1	3	67.30 ± 388.16
	5/18/2022	8	7	8	23	737.90 ± 388.16
	5/25/2022	11	8	8	27	1853.53 ± 388.16
	6/1/2022	12	10	9	31	5395.11 ± 388.16
	6/8/2022	12	10	9	31	5395.11 ± 388.16
	6/15/2022	12	9	9	30	4073.80 ± 388.16
	6/22/2022	12	10	9	31	5395.11 ± 388.16
	6/29/2022	12	10	9	31	5395.11 ± 388.16
	7/6/2022	12	10	9	31	5395.11 ± 388.16
	7/13/2022	12	11	9	32	7211.07 ± 388.16
	7/20/2022	12	11	9	32	7211.07 ± 388.16
	1/27/2022	12	11	9	32	7211.07 ± 388.16
2	5/13/2022 (Intake)	1	2	2	5	72.44 ± 388.16
	5/18/2022	8	4	7	19	340.41 ± 388.16
	5/25/2022	10	7	7	24	916.22 ± 388.16
	6/1/2022	10	8	8	26	1452.11 ± 388.16
	6/8/2022	10	9	8	27	1853.53 ± 388.16
	6/15/2022	11	9	8	28	2387.81 ± 388.16
	6/22/2022	11	9	8	28	2387.81 ± 388.16
	6/29/2022	11	9	8	28	2387.81 ± 388.16
	7/6/2022	11	9	8	28	2387.81 ± 388.16
	7/13/2022	11	9	8	28	2387.81 ± 388.16
	7/20/2022	11	9	8	28	2387.81 ± 388.16
	7/27/2022	11	10	9	30	4073.80 ± 388.16
3	5/27/2022 (Intake)	1	1	1	3	67.30 ± 388.16
	6/1/2022	7	6	9	19	340.41 ± 388.16
	6/8/2022	8	7	7	22	599.79 ± 388.16
	6/15/2022	9	8	7	24	916.22 ± 388.16
	6/22/2022	9	8	7	24	916.22 ± 388.16
	6/29/2022	9	8	7	24	916.22 ± 388.16
	7/6/2022	10	8	8	26	1452.11 ± 388.16
	7/13/2022	11	9	8	28	2387.81 ± 388.16
	7/20/2022	11	9	8	28	2387.81 ± 388.16
	7/27/2022	11	9	8	28	2387.81 ± 388.16
	8/3/2022	11	10	8	29	3104.56 ± 388.16
	8/10/2022	11	10	8	29	3104.56 ± 388.16

$$ADD = 10^{(0.002*TBS*TBS+1.81)} \pm 388.16$$

Equation 1. Formula for calculating ADD for donor individuals based on weekly TBS. This equation was developed by Megyesi et al.[2] to estimate PMI based on qualitative TBS scoring.

Weather and Environmental Data Collection

To assess the effects of temperature and relative humidity on the persistence of DNA in the soil, measurement of temperature and humidity were monitored using a National Oceanic and Atmospheric Administration (NOAA) weather station located at Jackson County Airport, Airport Road, Sylva, North Carolina 28779, which is located within a two-mile radius of the FOREST. The weather station continuously records temperature and relative humidity every 20 minutes, resulting in 72 records per day. These 72 records were averaged per day to calculate the daily mean temperature and humidity. Average rainfall was also measured using the Jackson County Airport weather station. Average precipitation was recorded every hour, for a total of 24 records per day, which were averaged for daily mean precipitation. The mean daily temperature, rainfall, and humidity calculations were then used to perform weekly averages to correspond to the weekly sampling from donors.

The effects of moisture retained by the cadaver on the persistence of DNA in soil were also assessed by recording weekly body moisture levels. A Delmhorst® RDM-3 moisture meter was used on the skin above the donor's left and right femora and the sternum to determine the moisture level of the flesh, as recorded in Table 3. The Delmhorst® RDM-3 can measure moisture content from 5%-60% using external electrodes that were placed against the flesh of the sampling areas on each donor. All measurements were taken using the default settings on the moisture meter. Donor 1 had no flesh remaining on the left femur from week 7 into the

decomposition period onward due to scavenging, so no readings were recorded for this extremity. By week 10, no flesh remained on any of the measurement regions for Donor 1 as the individual had entered advanced decay. Due to this, no moisture readings were taken from Donor 1 during the last two weeks of the decomposition period. For statistical analysis, moisture content for each of the three sampling sites (right femur, left femur, and sternum) per week per donor was averaged. If a reading was >60%, then the number was converted to 60 while averaging.

Table 3. Weekly moisture readings for each donor using the Delmhorst® RDM- moisture meter. Given the constraints of the instrument, any moisture reading above 60% could not be accurately assessed and is marked as >60%. An asterisk (*) is used to mark when no flesh remained, and samples could no longer be taken.

Donor ID	Date Recorded	Right Thigh (% moisture)	Left Thigh (% moisture)	Sternum (% moisture)
1	5/18/2022	54.7	>60	21.6
	5/25/2022	>60	>60	>60
	6/1/2022	27.7	25.2	48.4
	6/8/2022	24.5	42.2	46
	6/15/2022	30.1	>60	31.6
	6/22/2022	22.2	23.7	21.5
	6/29/2022	28.7	*	25.7
	7/6/2022	29.2	*	32
	7/13/2022	>60	*	>60
	7/20/2022	*	*	*
	1/27/2022	*	*	*
2	5/18/2022	16.2	20.8	21.7
	5/25/2022	>60	>60	>60
	6/1/2022	22.7	20.6	18.6
	6/8/2022	22.7	26.6	31.2
	6/15/2022	15.8	23.7	18.9
	6/22/2022	26.4	13.6	26.3
	6/29/2022	17.7	29.4	29.7
	7/6/2022	21.8	25.3	28.6
	7/13/2022	23.8	22.4	20.4
	7/20/2022	22.2	24.8	24.6
	7/27/2022	21.3	10.6	26.3
3	6/1/2022	18.5	55.7	17.6
	6/8/2022	36.3	30.7	24.5
	6/15/2022	18.4	21	18.6
	6/22/2022	18.2	16.8	17.8
	6/29/2022	16.7	19.3	19
	7/6/2022	21.7	23.1	22.4
	7/13/2022	17.1	17.4	16.3
	7/20/2022	20.6	19	23.4
	7/27/2022	21.3	17.9	20.9
	8/3/2022	32.1	23.8	26.2
	8/10/2022	39.6	24	29.3

Sample Nomenclature

Each sample was given a unique identifier that included the donor identification number, the location around the body where the sample was collected, and the week into the decomposition period. The location around the body was identified with the abbreviations C for the left side of the cranium, TR for the right side of the torso, TL for the left side of the torso, and LE for the right lower extremity. All extractions from reference buccal swabs were identified with the donor ID number and REF.

Sample Extraction

DNA Extraction from Soil

DNA was extracted from soil samples using the QIAGEN DNeasy® PowerSoil® Pro kit following the quick start protocol.[31] 250 mg of soil from each sample was placed in PowerBead Pro tubes with 800 µl of Solution CD1 and vortexed briefly to mix. Samples were attached to a horizontal vortex adapter and vortexed at high speed for 15 minutes. The solution was centrifuged at 12,100 x g for 1 minute, then the supernatant was moved to a fresh tube. 200 µl of Solution CD2 was added to the supernatant and vortexed for 5 seconds, then centrifuged at 12,100 x g for 1 minute. The supernatant was transferred and combined with 600 µL of Solution CD3 and vortexed for 5 seconds. 650 µl of lysate were loaded to an MB Spin Column and centrifuged at 12,100 x g for 1 minute. Flow-through was discarded and the process was repeated a second time to ensure all lysate had passed through the spin column. The spin column was placed in a new collection tube and 500 µl of Solution EA was added. The samples were centrifuged at 12,100 x g for 1 minute, and the flow-through was discarded. 500 µl of Solution CD5 was dispensed and columns were centrifuged at 12,100 x g for 1 minute. After discarding the flow-through, the spin columns were placed into a new collection tube and centrifuged at

12,100 x g for 2 minutes to remove excess solution. The spin columns were then loaded into elution tubes and 50 µl of Solution CD6 was added to the center of the filter membrane and centrifuged at 12,100 x g for 1 minute, resulting in a 50µl DNA elution. DNA extracts were stored in microcentrifuge tubes at -18°C. Each extraction run included a reagent blank to determine whether contamination had occurred during the extraction phase.

Reference DNA Extraction from Buccal Swabs

DNA from reference buccal swabs was extracted using the QIAGEN EZ1® DNA Investigator® Kit. Swabs were pretreated in 2 ml sample tubes with a Proteinase K digest of 10 µl Proteinase K to 290 µl of Buffer G2. The swabs were vortexed for 10 seconds, then incubated at 56°C with 900 rpm shaking for 20 minutes. The samples were loaded onto the QIAGEN EZ1® Advanced XL and processed with the automated "Tip Dance" protocol with TE elution buffer set to 50 µL.[32] Reference DNA extracts were stored at -18°C. A reagent blank was included to determine whether contamination occurred during extraction.

Sample DNA Quantitation

Nuclear

nuDNA from all experimental and reference samples was quantified using the Applied Biosystems Quantifiler™ Trio DNA Quantification Kit, hereafter referred to as Quantifiler™ Trio. Quantifiler™ Trio was used since it is a human-specific multiplex quantitative PCR (qPCR) kit that has three specific targets: a large autosomal target, a small autosomal target, and a Y-chromosomal target which is widely used in forensic DNA laboratories.[33] Having large and small autosomal targets allows for assessment of the amount of degradation seen in the DNA extracted by creating a Degradation Index (DI), as calculated by Equation 2. A DI for intact DNA will be ≤1, while a DI >1 indicates degradation. The Y-chromosomal target can provide

preemptive insight into the biological sex of the individual in question. The Quantifiler™ Trio kit quantifies DNA by using a 5-point standard curve, which is created from a stock of 100ng/µL Quantifiler™ THP DNA Standard serial diluted to 50.0 ng/µl, 5.0 ng/µl, 0.5 ng/µl, 0.05 ng/µl, and 0.005 ng/µl, respectively. Fresh master mix was made before each run using 8µl of Quantifiler™ Trio Primer Mix per reaction and 10µl of Quantifiler™ THP PCR Reaction Mix per reaction. The preparation of samples for quantification followed the general protocol outlined in the Quantifiler™ Trio DNA Quantification Kit user guide, in which 18µl of master mix and 2µl of sample DNA extract, DNA standard, or negative control was added to each well of the 96-well plate.[33] All DNA standards and a non-template control were loaded in duplicate on the 96-well plate, while DNA extracts and reagent blanks were loaded in triplicate wells to determine the average quantity of DNA per sample. Each plate was read on the Applied Biosystems™ 7500 Real-Time PCR system using the HID Real-Time PCR Analysis Software with the preset run method for the Quantifiler™ Trio Kit.

$$DI = \frac{Concentration\ of\ Small\ Autosomal\ Target\ DNA\ (ng/\mu l)}{Concentration\ of\ Large\ Autosomal\ Target\ \ DNA\ (ng/\mu l)}$$

Equation 2. Formula for calculating Degradation Index using the Quantifiler™ Trio kit. The large and small autosomal target DNA concentrations are based on the average of triplicate quants for each sample.

Mitochondrial DNA Quantitation

mtDNA in all experimental and reference DNA extracts was quantified using a triplex qPCR assay described by Kavlick.[34] This assay includes a short (105 bp) mtDNA target in the ND5 Gene, a long (316 bp) mtDNA target in the 16s rRNA Gene, and an Internal Positive Control (IPC).[34] Similar to the nuDNA quantitation, long and short targets of the Kavlick triplex assay can be used to calculate a DI, referred to as ΔΔCt (Equation 3). ΔΔCt provides insight into

the amount of degradation seen within the mtDNA in a sample. A ΔΔCt value of zero is interpreted to be undegraded, a ΔΔCt value of one is interpreted as slight degradation since there are half as many large target copies as the small target, and ΔΔCt values of four suggest 2^4 fewer large targets than small, indicating a higher state of sample degradation. The Kavlick assay uses a 7-point standard curve to quantify mtDNA copy number based on the 105 bp amplicon. This standard curve was created from a dsT8sig standard DNA stock that is diluted in a 10-fold serial dilution with TE buffer (10 mM Tris, pH 8, 0.1 mM EDTA) from 10^7 to 10^1 copies per 1 µl. Each reaction contains 2 µl of sample, standard, or control DNA and 18 µl of the qPCR master mix.[34] All DNA standards, a non-template control, and HL60 calibrator positive control DNA were loaded in duplicate on the 96-well plate. DNA extracts and reagent blanks were loaded in triplicate wells to determine the average quantity of DNA per sample. The plates were read on the Applied Biosystems™ 7500 Real-Time PCR system using the HID Real-Time PCR Analysis Software set for a custom assay. The following amplification settings were used for each run: 50°C for 2 minutes, 95°C for 20 seconds, and 40 cycles consisting of 95°C for 3 seconds and 60°C for 30 seconds.[34]

$$\Delta\Delta Ct = \Delta Ct_{Sample} - \Delta Ct_{Calibrator} = (Ct_{316} - Ct_{105})_{Sample} - (Ct_{316} - Ct_{105})_{Calibrator}$$

Equation 3. Formula for calculating ΔΔCt using the Kavlick mtDNA qPCR assay. Ct_{316} refers to the Ct value of the long mtDNA target, which is quantified relative to the short mtDNA target, referred to as Ct_{105} in the equation. This was done for both the sample DNA (ΔCt_{sample}) and the intact and undegraded HL60 calibrator DNA ($\Delta Ct_{calibrator}$) to normalize any differences in amplification efficiencies of the targets.

Verogen ForenSeq™ DNA Signature Prep Kit

The ForenSeq™ DNA Signature Prep Kit was used for sequencing nuDNA markers for all soil samples within the first six weeks of the decomposition period when nuDNA was still

quantifiable. Two samples with the highest quantity of small target nuDNA per week per donor were selected for sequencing. The kit includes two different primer mixes: DNA Primer Mix A (DPMA) and DNA Primer Mix B (DMPB). DMPB was used due to its wide coverage of 230 markers, including 27 autosomal short tandem repeats (STRs), 24 Y STRs, 7 X STRs, 94 identifying SNPs, 22 phenotypic SNPs, and 56 biogeographical ancestry SNPs. Sequencing was performed on the Illumina® MiSeq FGx™ Forensic Genomics System, hereafter referred to as the MiSeq FGx™. The MiSeq FGx™ was used as it is the first fully validated sequencing system for forensic genomics applications.[35]

Library Preparation

The Purified DNA protocol for the ForenSeq™ DNA Signature Prep Kit was followed for library preparation workflow.[26] Due to the low quantity of nuDNA in each sample, 5µl of undiluted sample DNA was added to each well of the ForenSeq Sample Plate during the amplify and tag targets step of the protocol. Thirty-two samples were pooled per library, which included 28 extracted DNA samples, two randomly selected reagent blanks from the first six weeks of extractions, a positive control, and a negative control. Libraries could then be stored for 30 days in -18°C storage.

Sequencing and Data Analysis

Libraries were denatured and diluted according to the ForenSeq™ DNA Signature Prep Kit reference guide[26] and loaded onto a prepared reagent cartridge. The sequencing run was set up in the MiSeq FGx™ Control Software (MFCS) v1.3 forensic genomics mode following the MiSeq FGx™ Sequencing System reference guide guidelines.[36] Summary of run metrics and individual sample sequencing results were analyzed in Verogen's Universal Analysis Software (UAS) v2.0.

Verogen ForenSeq™ mtDNA Whole Genome Kit

The ForenSeq™ mtDNA Whole Genome Kit was used to sequence the mtDNA from weekly soil samples for each donor plot during the first eight weeks of the decomposition period. The lack of sequencing for weeks nine through eleven of the decomposition period was due to sequencing reagents becoming depleted along with technical issues with the MiSeq FGx™ instrumentation. This prevented further sequencing of DNA extracts. One sample with the highest quantity of mtDNA per week per donor was selected for sequencing. The kit sequences the entire 16,569 bp human mtGenome, providing information on coding regions, non-coding regions, and haplogroups. This is done using two tiled primer mixes, Whole Genome Mix Set 1 (WGS1) and Whole Genome Mix Set 2 (WGS2), to establish total coverage of the mtGenome.[27] Sequencing was performed on the MiSeq FGx™ instrument by Verogen (San Diego, CA).

Library Preparation

Library preparation was performed following the protocol outlined in the ForenSeq™ mtDNA Whole Genome Kit Reference Guide.[27] A 96-well plate was divided into two sections, one for the WGS1 reaction and the other for the WGS2 reaction. Each sample was diluted to 8.33 pg/µl and 6µl was added to each half-reaction during the protocol's amplify and tag targets step. If a sample had a concentration of less than 8.33 pg/µl, undiluted sample was added. Sixteen samples were pooled per library, which included 13 extracted DNA samples, one randomly selected reagent blank from the sample extraction subset, one positive control, and one negative control.[27] This resulted in two separate sequencing runs. Each sample underwent two rounds of sample purification to account for low template input. All purified mtGenomic libraries from weeks one through three along with sample Donor 2 TR-4 were sent to Verogen (San Diego, CA) for normalization and sequencing. Weeks four through eight were sequenced in-

house with MiSeq FGx™ Control Software (MFCS) v2.0+ forensic genomics mode following the MiSeq FGx™ Sequencing System reference guide guidelines.[36] Summary of run metrics and individual sample sequencing results were analyzed in Verogen's Universal Analysis Software (UAS) v2.0.

Sequencing and Data Analysis

Data on the sequencing run were provided by Verogen for weeks one through three. These data were imported into a new run analysis in Verogen's UAS. Weeks four through eight were sequenced in house after instrument maintenance on the MiSeq FGx™. The summary of run metrics and individual sample sequencing results were also analyzed in the UAS. Haplogroup determinations for sequencing data were performed in Haplogrep 3 (version 3.2.1) using the PhyloTree 17 – Forensic Update 1.2 at the reference phylogenetic tree and the distance function set to Kulczynski.[37]

Statistical Analysis

DNA Persistence, Time, and Environmental Factors

Statistical analyses were performed in R Statistical Software v.4.1.3[38] and Microsoft Excel version 2304.[39] A PCA of environmental factors (TBS, body moisture content, weekly average rainfall, weekly average humidity, and weekly average temperature) and time into decomposition period was performed using the FactoMineR package.[40] The results of the PCA were visualized into PCA biplots with the factoextra package.[41] Linear models comparing the quantity of DNA to each dimension of the PCA were performed and analyzed with the anova() function in base R and visualized using the ggplot2 and extrafont packages.[38, 42-43] Spearman's rank correlation, or Spearman's ρ, was also performed in Excel to determine the degree of

association between environmental variables and sample quantity. The ranked average of each variable was calculated and Spearman's ρ was calculated using the CORREL() function.

CHAPTER THREE: RESULTS

The following chapter shows the results of all analyses performed on soil samples collected from the decomposition period. Quantitation of both nuDNA and mtDNA was performed until DNA was no longer detected by qPCR assays. NGS of the nuDNA covers the first six weeks of the decomposition period, while sequencing of the mtDNA was performed on extracts from the first eight weeks of decomposition. Sequencing was not possible for all eleven weeks of soil samples for nuDNA as DNA had become fully degraded after six weeks. As for mtDNA, sequencing only covers the first eight weeks of DNA extracts due to the depletion of sequencing reagents and technical issues discussed later within this chapter. Results of nuDNA quantitation and NGS are shown first, along with an analysis of the effects of environmental factors on nuDNA quantity in the soil. This is followed by quantitation and NGS results from mtDNA extracted from the soil, with statistical analysis of the effects of environmental factors on the quantity of mtDNA. Quantitation results for both nuclear and mitochondrial DNA will show the duration of DNA persistence in soil over the eleven-week decomposition period, while NGS data will determine the utility of using soil-derived DNA for producing human-identifying DNA profiles. Statistical analysis of environmental factors and their effect on DNA quantity within the soil will provide context for understanding the mechanisms behind DNA persistence and degradation within soil.

Extraction Troubleshooting

During the QIAGEN DNeasy® PowerSoil® Pro kit extraction from weekly soil samples, the samples Donor 1 C-2 and Donor 2 TR-1 experienced minimal elution after the final centrifugation step. Additional centrifugation was attempted with minimal effect. An additional

30 µl of Solution CD6 was added and the samples were centrifuged again, resulting in a preferred amount of elution.

In addition, due to complications in attempting alternative quantitation methods including droplet digital PCR, some extracts became depleted. As a result, these weekly soil samples were re-extracted for all downstream analyses. These samples include Donor 1 C-1, Donor 1 TR-1, Donor 1 TL-1, Donor 1 LE-1, Donor 2 C-1, Donor 2 TR-1, Donor 2 TL-1, Donor 2 LE-1, Donor 3 C-1, Donor 3 TR-1, Donor 3 TL-1, Donor 3 LE-1, Donor 1 LE-2, and Donor 1 TL-2.

For the eleventh week of decomposition, only one soil sample was selected per donor for extraction due to the number of DNA extraction minipreps being limited. These samples were selected as they would likely contain relatively high concentrations of DNA based on trends in quantitation results for prior weeks. The samples selected were Donor 1 TL-11, Donor 2 TL-11, and Donor 3 C-11.

Nuclear DNA

Quantification

Quantification of nuDNA from soil extracts was performed using the Quantifiler™ Trio DNA Quantification Kit.[33] The average quantity of large autosomal, small autosomal, and Y-chromosomal target DNA detected for each donor per week is present in Table 4, with DI calculated using Equation 2. These data are averages of the triplicate quantitation of each of the four weekly sampling locations per donor per week (Appendix A, Tables A1-A3). Due to the nature of qPCR being sensitive to spurious contamination and pipetting error, some replicate wells were removed from quantitation calculations. Outliers were removed based on the Ct standard deviation (Ct-SD) as outlined in Maussion et al.[44] If the Ct-SD was greater than 0.3, then the replicate furthest from the sample mean was removed from calculations. Highly variable

replicates were preserved if the absolute (mean-median)/median was less than 0.1, as no clear outlier can be determined. Graphical representations of Table 4 are found in Figures 1, 2, and 3. Data include error bars showing the standard error of the data calculated using the replicate quants of the four weekly samples per donor. Recovery of quantifiable nuDNA from soil extracts was only possible during the first six weeks of the decomposition period for all three donors.

Table 4. Quantification results for nuDNA from experimental samples including small autosomal target, large autosomal target, and Y-Chromosomal target DNA in nanograms per microliter. * - indicates when there was no large autosomal target DNA present to calculate DI.

Donor	Week	Small Target (ng/µl)	Large Target (ng/µl)	Y Target (ng/µl)	Degradation Index
1	1	1.71E-03	3.27E-04	0	5.22
	2	1.09E-04	9.09E-06	0	12.00
	3	4.82E-04	0	0	*
	4	0	3.33333E-05	0	0.00
	5	5.00E-05	0	0	*
	6	1.17E-04	5.00E-05	0	2.33
2	1	7.07E-03	3.74E-03	6.91E-03	1.89
	2	7.45E-04	1.18E-04	3.27E-04	6.31
	3	4.17E-04	0	0	*
	4	3.33E-05	8.33E-05	0	0.40
	5	9.17E-05	0	0	*
	6	1.17E-04	1.67E-05	0	7.00
3	1	2.98E-03	9.00E-04	1.18E-03	3.31
	2	6.18E-04	2.36E-04	1.67E-05	2.62
	3	1.42E-04	2.50E-05	0	5.67
	4	0	0	0	*
	5	6.67E-05	0	0	*
	6	6.67E-05	8.33E-06	0	8.00

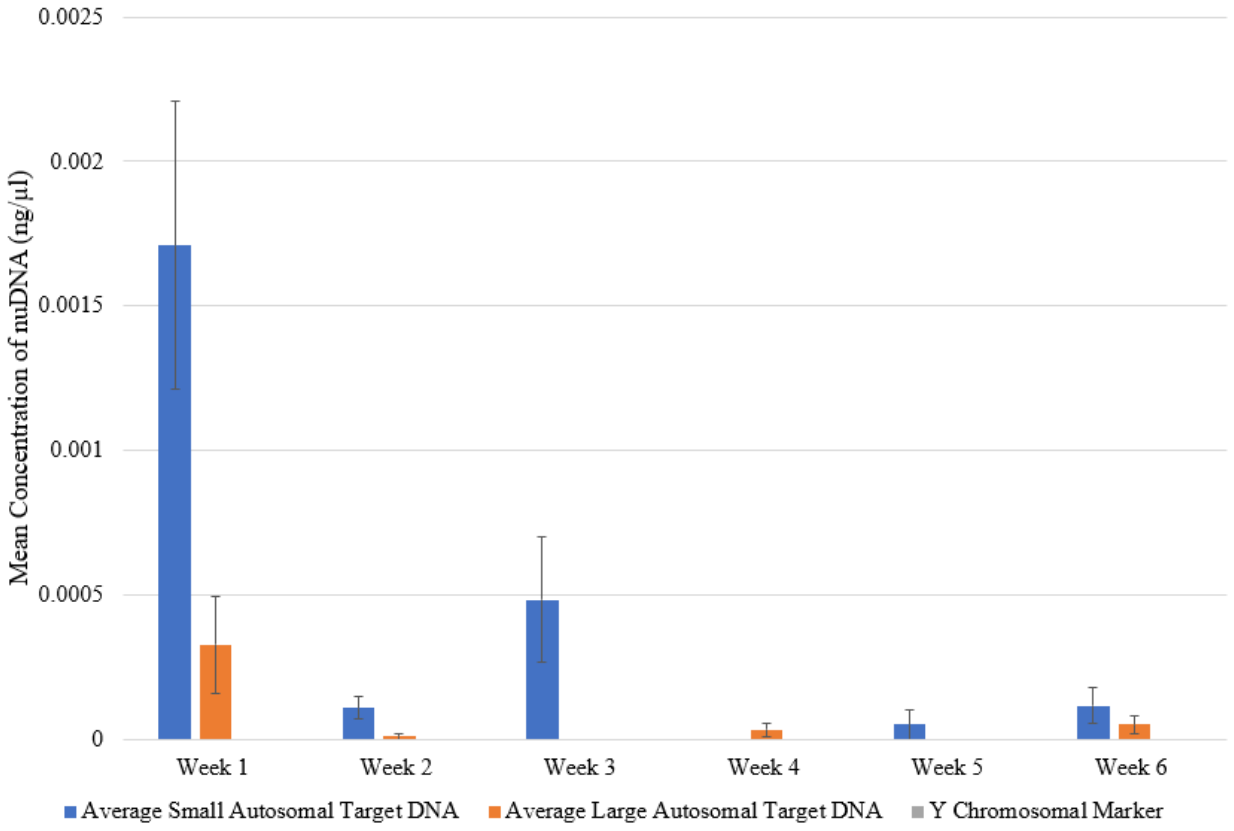

Figure 1. The average quantity of Large Autosomal, Small Autosomal Target DNA, and Average Y-Chromosomal Target DNA amplified from soil samples taken from the plot of **Donor 1** over six weeks of decomposition. Standard error was calculated using the replicate quantitations of each four weekly samples per donor.

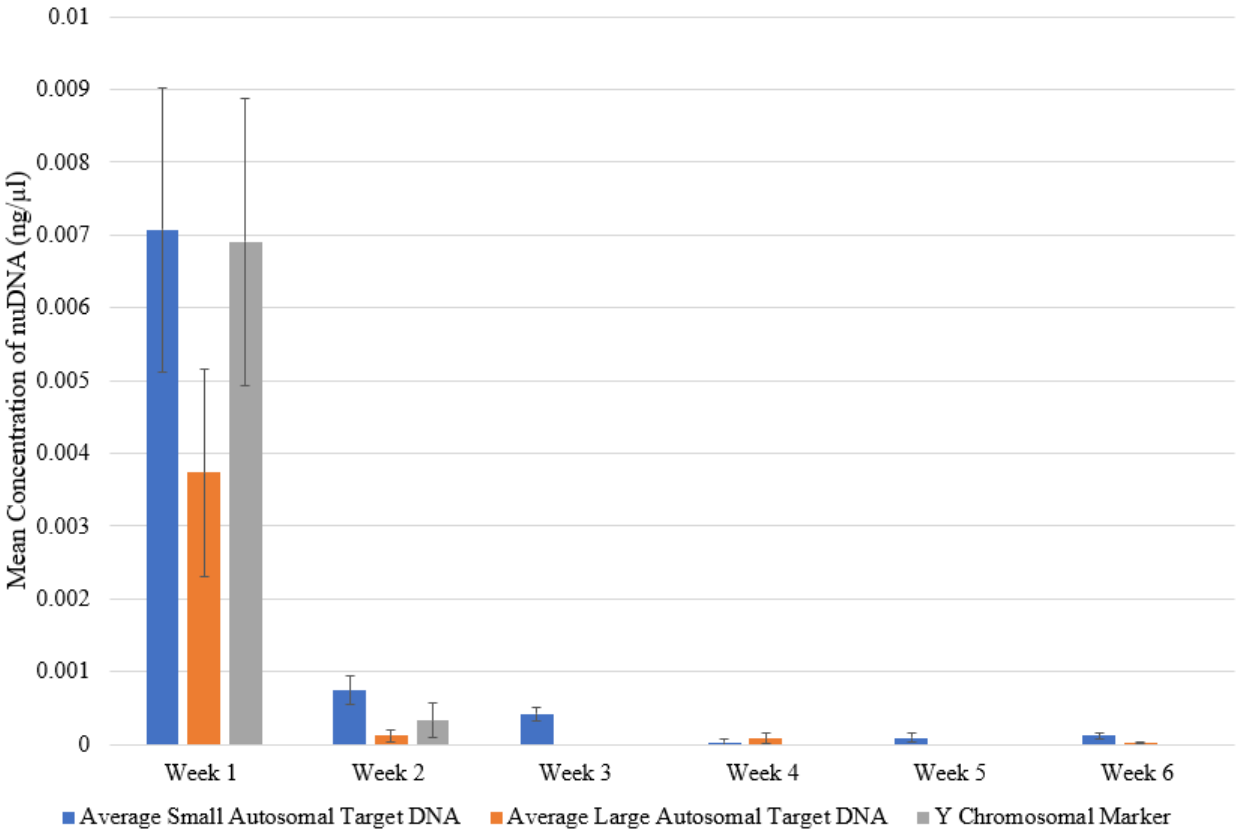

Figure 2. The average quantity of Large Autosomal, Small Autosomal Target DNA, and Average Y-Chromosomal Target DNA amplified from soil samples taken from the plot of **Donor 2** over six weeks of decomposition. Standard error was calculated using the replicate quantitations of each four weekly samples per donor.

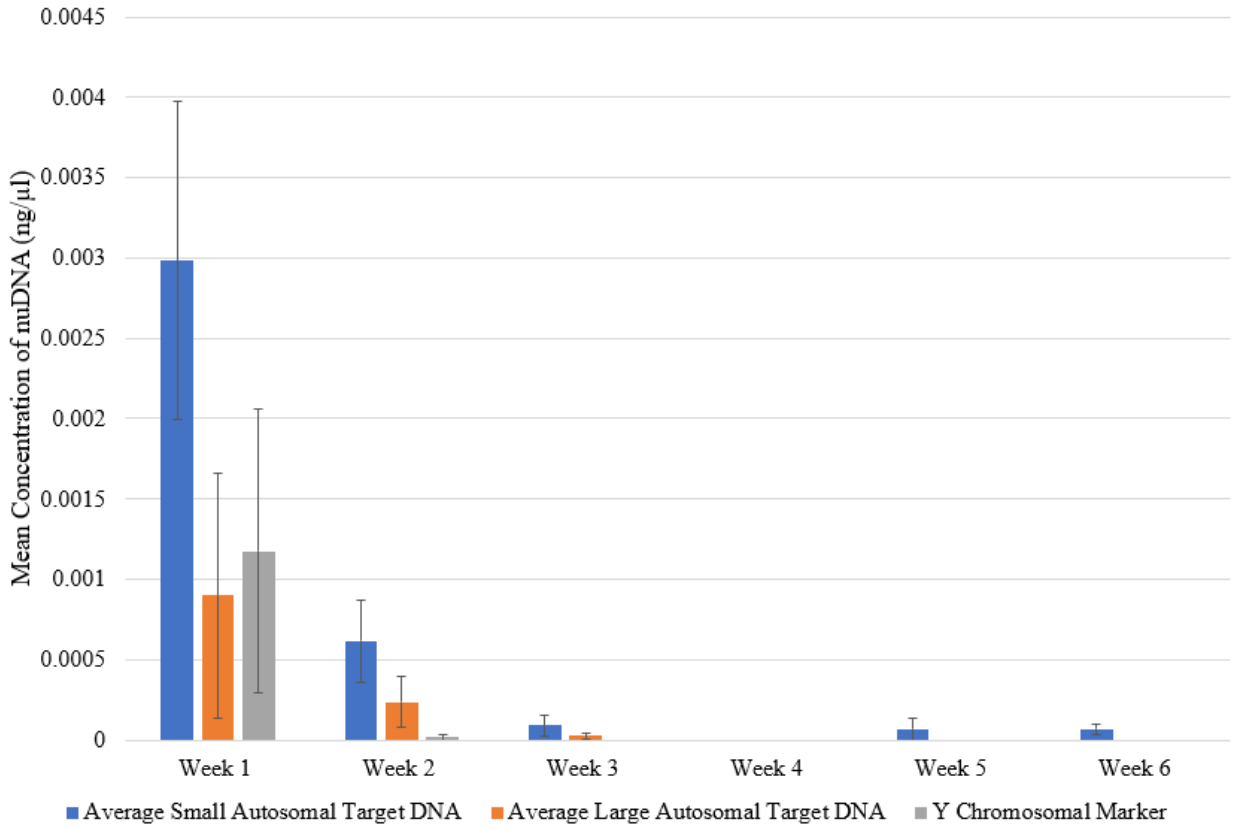

Figure 3. The average quantity of Large Autosomal, Small Autosomal Target DNA, and Average Y-Chromosomal Target DNA amplified from soil samples taken from the plot of **Donor 3** over six weeks of decomposition. Standard error was calculated using the replicate quantitations of each four weekly samples per donor.

For small target DNA recovery, concentrations ranged from undetectable to 7.07E-03 ng/µl. For large autosomal target DNA recovery, concentrations ranged from undetectable to 3.74E-03 ng/µl. Y-Chromosomal target DNA was only detected in weeks one and two for Donors 2 and 3. The quantity of nuDNA generally decreased over the six-week period while DI generally increased. However, nuDNA was more recoverable within the sixth week of the decomposition period than the fourth or fifth week. This will be further discussed in the following chapter. There was no indication of PCR inhibition for samples in which nuDNA was undetected.

Next-Generation Sequencing

Sequencing was performed with the pooled library on a standard MiSeq reagent Kit V2 flow cell. A custom run was created for the sequencing run with the ForenSeq™ DNA Signature Prep kit for nuDNA. Following sequencing on a standard MiSeq reagent Kit V2 flow cell, results were analyzed in Verogen's UAS. A summary of the number of loci detected is displayed in Table 5.

Table 5. Summary of the number of loci detected for each sample using the ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™.

Sample	Autosomal STRs	Y STRs	X STRs	Identifying SNPs (iSNPs)
Per Kit	27	24	7	94
Pos Control	26	23	7	94
Neg Control	0	0	0	0
Donor 2 LE-1	2	0	0	5
Donor 2 TR-2	1	0	0	1
Donor 2 C-2	0	1	1	1

NuDNA loci were only detected in three experimental samples from Donor 2, which were from one or two weeks into the decomposition period. No loci were detected in any other experimental samples. Despite the quality of the run, some loci were not detected in the positive control. Only 26/27 autosomal STRs and 23/24 Y STRs were detected within the positive control.

The allele calls and depth of reads for these detected loci are represented in Table 6. A read refers to a sequence of bases in a single DNA fragment, whereas a mapped read refers to a read from a sample sequence that directly aligns with a set of loci on a reference genome. The read depth is defined as the number of reads that overlap with the target loci region.[45] The

sample Donor 2 LE-1 was flagged on the TH01 locus for detecting more alleles than expected, exceeding the stutter threshold, and having a non-stutter allele between the analytical and interpretation threshold.

Table 6. Genotype and depth of coverage from soil samples sequenced with the ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™. * - indicates that the locus was flagged for issues in interpretation.

Sample	Target Type	Locus	Allele	Read Depth
Donor 2 LE-1	Autosomal STR	TPOX	8	49
Donor 2 LE-1	Autosomal STR	TH01	9, 9.3 *	75, 75 *
Donor 2 LE-1	iSNP	rs737681	C	55
Donor 2 LE-1	iSNP	rs1821380	G	34
Donor 2 LE-1	iSNP	rs1382387	T	52
Donor 2 LE-1	iSNP	rs576261	A	136
Donor 2 LE-1	iSNP	rs1028528	A	33
Donor 2 TR-2	Autosomal STR	D2S441	10	34
Donor 2 TR-2	iSNP	rs8037429	C	195
Donor 2 C-2	Y STR	DYS438	12	60
Donor 2 C-2	X STR	DXS7423	15	31
Donor 2 C-2	iSNP	rs8078417	C	32

Statistical Analysis of Environmental Factors Effect on nuDNA Persistence

All data for statistical analysis of environmental effects on nuDNA persistence can be found in Appendix B, Table B1. The two main components of the PCA for environmental factors and decomposition data, together, can explain 67.5% of the variation between samples (Figure 4). The correlation coefficient for each explanatory variable along each dimension can be found in Table 7. The variable with the highest positive correlation coefficient along dimension one is average weekly rainfall (0.871) followed by body moisture content (0.812). When looking at the vectors for average rainfall, body moisture content, and humidity on PCA biplot (Figure 4), their angle to horizontal axis and magnitude indicates that there is a moisture gradient responsible for explaining the variation in dimension 1. For dimension two, the variable with the highest positive

correlation coefficient was TBS (0.867) followed by average weekly temperature (0.675). Dimension three was also plotted against dimension 1 in a PCA biplot (Figure 5), which together account for 58.1% of the variation between samples. The variable with the highest positive correlation coefficient along dimension three was average humidity (0.695) while the variable with the highest negative correlation coefficient was TBS (-0.333).

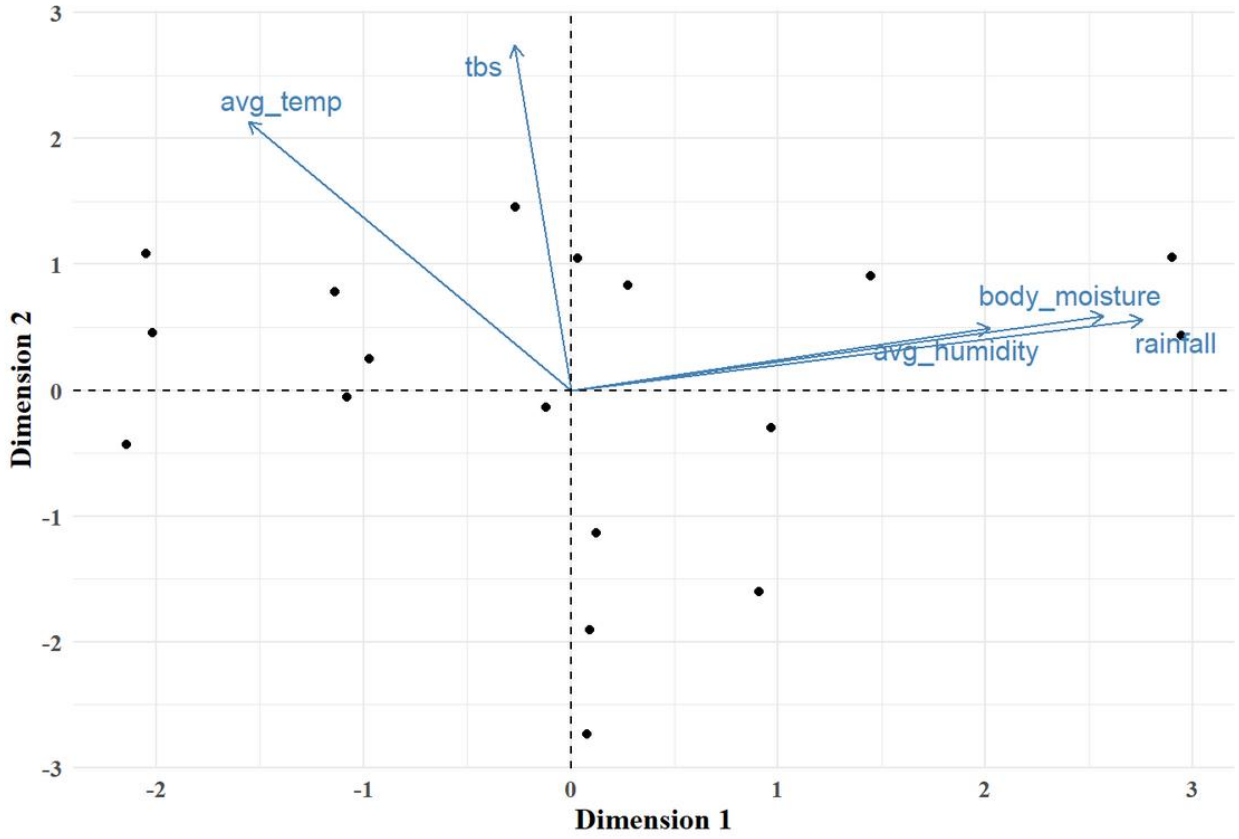

Figure 4. PCA biplot of dimensions one and two for the first six weeks of data collection from the human decomposition of three donors, including vectors indicating which environmental and decomposition factors were most important for differentiating samples.

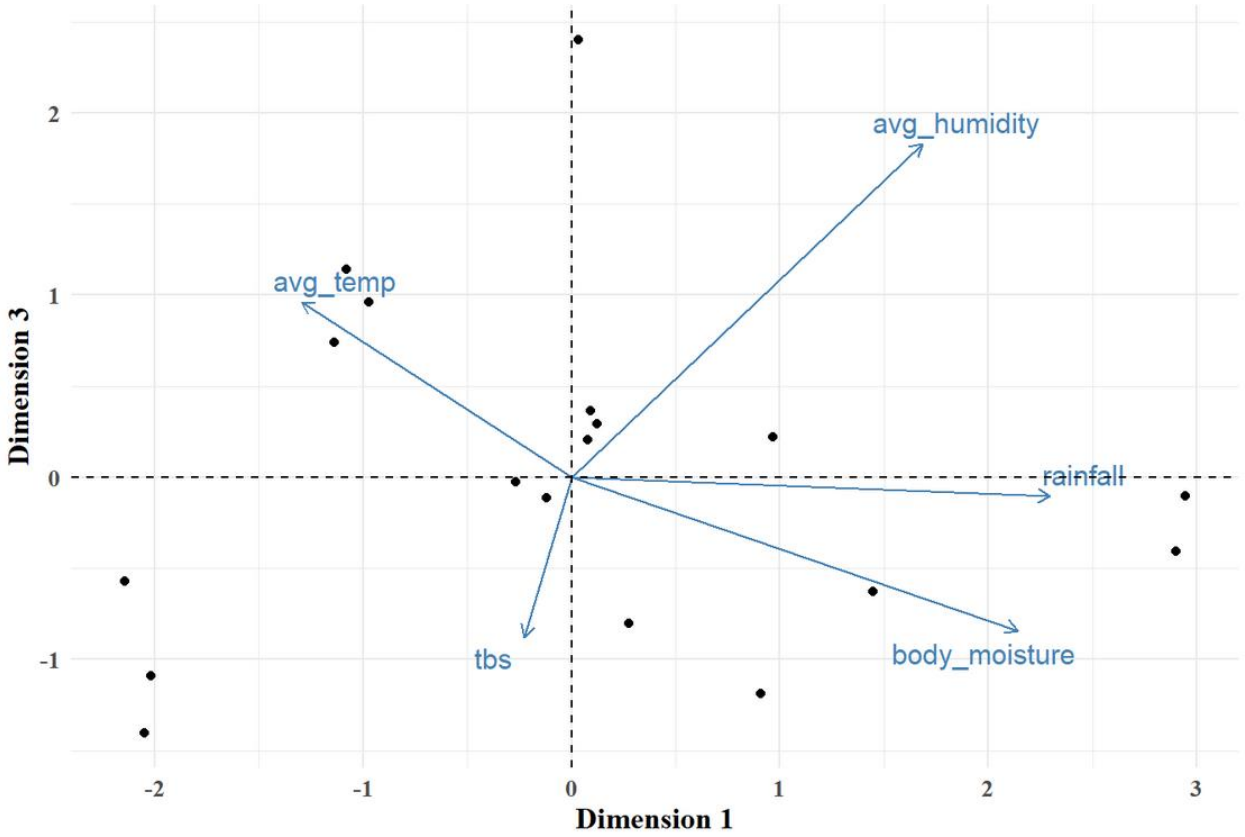

Figure 5. PCA biplot of dimensions one and three for the first six weeks of data collection from the human decomposition of three donor including vectors indicating which environmental and decomposition factors were most important for differentiating samples.

Table 7. Correlation coefficients for principal components analysis of environmental and decomposition factors influencing weekly soil samples following six weeks of surface level decomposition.

Variable	Dimension 1 Correlation Coefficient	Dimension 2 Correlation Coefficient	Dimension 3 Correlation Coefficient
TBS	-0.086	0.867	-0.333
Body Moisture Content	0.812	0.186	-0.319
Avg Weekly Rainfall	0.871	0.177	-0.039
Avg Weekly Humidity	0.639	0.157	0.695
Avg Weekly Temperature	-0.491	0.675	0.365

Linear regressions comparing dimensions one, two, and three of the PCA to the quantity of nuDNA in the soil were then performed and analyzed with an ANOVA, shown in Table 8. There appears to be a statistically significant relationship between dimension two of the PCA with the quantity of nuDNA in the soil (*p*-value = 9.696e-05), indicating that TBS and weekly average temperature have a significant effect on the persistence of nuDNA in soil. These data reflect a strong negative correlation between dimension two of the PCA and the quantity of nuDNA (Figure 6).

Table 8. ANOVA of linear regressions between each dimension of the PCA and concentration of nuDNA. ANOVAs with significant *p*-values (≤ 0.05) are in bold.

	DF	**Sum Sq**	**Mean Sq**	**F Value**	**Pr(>F)**
Dimension 1	1	7.5600e-07	7.5601e-07	0.2401	0.6308
Residuals	16	5.0382e-05	3.1488e-06		
Dimension 2	**1**	**3.1905e-05**	**3.1905e-05**	**26.542**	**9.696e-05**
Residuals	**16**	**1.9233e-05**	**1.2020e-06**		
Dimension 3	1	1.5000e-08	1.5000e-08	0.0047	0.9463
Residuals	16	5.1123e-05	3.1952e-06		

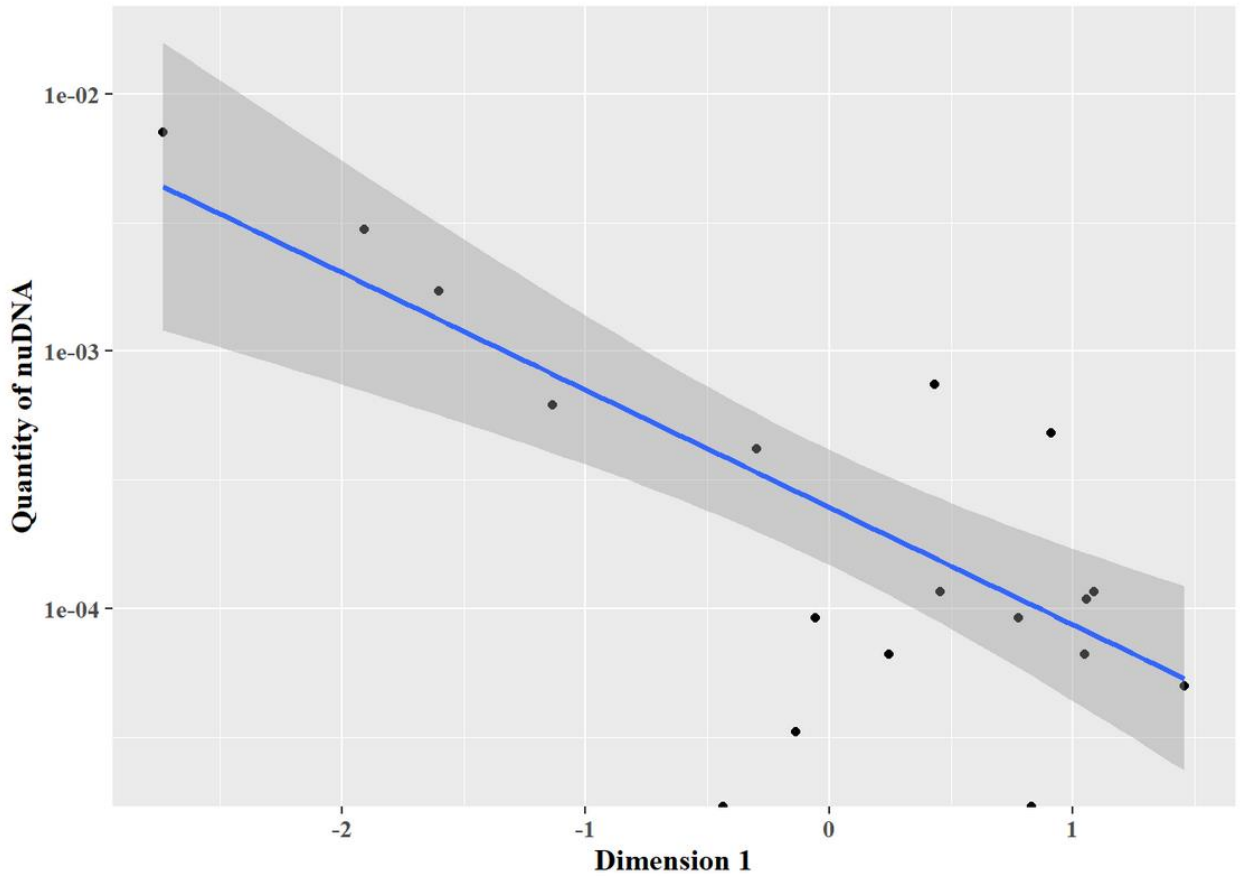

Figure 6. Linear regression of concentration of nuDNA in weekly soil samples compared to dimension two of the PCA. The y-axis has been converted to a logarithmic scale. The grey shading represents the 95% confidence interval and each soil sample per donor is represented by an individual black dot.

Spearman's rank correlation coefficients between environmental variables and the quantity of nuDNA are shown in Table 9. There are statistically significant negative correlations between the number of weeks the remains were decomposed and nuDNA concentration (Spearman's $\rho = -0.6594$, p-value = 0.0029), TBS of remains and nuDNA concentration (Spearman's $\rho = -0.4977$, p-value = 0.0356), and average weekly temperature and nuDNA concentration (Spearman's $\rho = -0.6184$, p-value = 0.0062). All other correlation calculations were insignificant.

Table 9. Spearman's rank correlation coefficients (Spearman's ρ) between weekly environmental factors and concentration of nuDNA derived from weekly soil samples. Correlations with significant *p*-values (≤ 0.05) are in bold.

Relationship	Spearman's ρ	T-Statistic	DF	*p*-value
Concentration nuDNA with Week into Decomposition Period	**-0.6594**	**3.5082**	**16**	**0.0029**
Concentration nuDNA with TBS	**-0.4977**	**2.2951**	**16**	**0.0356**
Concentration nuDNA with Weekly Body Moisture Content	0.2668	1.1074	16	0.2845
Concentration nuDNA with Total Weekly Rainfall	0.1729	0.7023	16	0.4926
Concentration nuDNA with Weekly Average Humidity	-0.0167	0.0667	16	0.9477
Concentration nuDNA with Weekly Average Temperature	**-0.6184**	**3.1479**	**16**	**0.0062**

Mitochondrial DNA

Quantification

Quantification of mtDNA from soil extracts was performed using Kavlick triplex qPCR assay.[34] The average quantity of short (105 bp) mtDNA target detected for each sample location per donor per week is present in Table 10. These quantities were calculated by averaging the triplicate quants of each weekly soil sample (Appendix A, Tables A3-A6). Outliers in each triplicate quantitation were removed using the Ct-SD, as described by Maussion et al.[43] Data from Table 10 can be found in graphical representations in Figures 7, 8, and 9. Data include error bars showing the standard deviation of the data. mtDNA was recoverable from soil extracts for all eleven weeks of the decomposition period.

Table 10. Quantification results for mtDNA from weekly experimental samples located left of the cranium, either side of the torso, and the right lower extremity in copy number per sample. NA indicates samples that were not extracted from and therefore, could not be quantified.

Donor ID	Week	Copy Number per Sample			
		Cranium	Right Torso	Left Torso	Lower Extremity
Donor 1	1	9533.589	1251.113	874.7076	641.8237
	2	609.1933	490.7316	3060.808	4621.655
	3	595.61	1252.579	15809.48	12297.62
	4	8772.39	798.2148	888.0397	668.9941
	5	442.0157	6391.292	571.7514	979.4836
	6	326.799	412.7456	6418.178	10050.42
	7	441.0747	585.9004	775.1834	692.9191
	8	492.4627	53.0694	64.9116	57.31903
	9	276.4725	366.0379	357.0675	278.0362
	10	204.7605	299.0701	226.1025	157.4057
	11	NA	NA	234.8125	NA
Donor 2	1	7949.136	13126.75	9378.515	21992.44
	2	7031.173	23122.97	27191.33	31869.53
	3	176163.1	123945	16863.4	287760.6
	4	833.5135	49105.97	8846.179	3681.66
	5	758.9095	2322.716	862.4975	886.317
	6	4506.705	4451.278	4737.329	4557.876
	7	962.1839	1200.856	1196.56	1229.645
	8	173.9251	121.0262	86.64017	156.4804
	9	483.9317	696.5407	1036.411	1033.236
	10	1613.794	1377.494	769.2741	4048.671
	11	NA	NA	1456.52	NA
Donor 3	1	3652.005	886.3371	18716.96	36282.58
	2	208.8542	266.9231	2637.38	16318.25
	3	437.6308	517.3273	542.8632	1717.409
	4	725.4743	240.8857	330.4717	1477.579
	5	662.3185	1269.03	924.012	1855.752
	6	4532.015	4657.145	5129.726	852.7269
	7	1291.822	1190.462	2166.3	1667.638
	8	107.0631	266.0392	3316.129	310.7740
	9	2348.609	1818.828	1537.603	141.7237
	10	3261.397	1109.265	729.6526	333.7573
	11	1983.269	NA	NA	NA

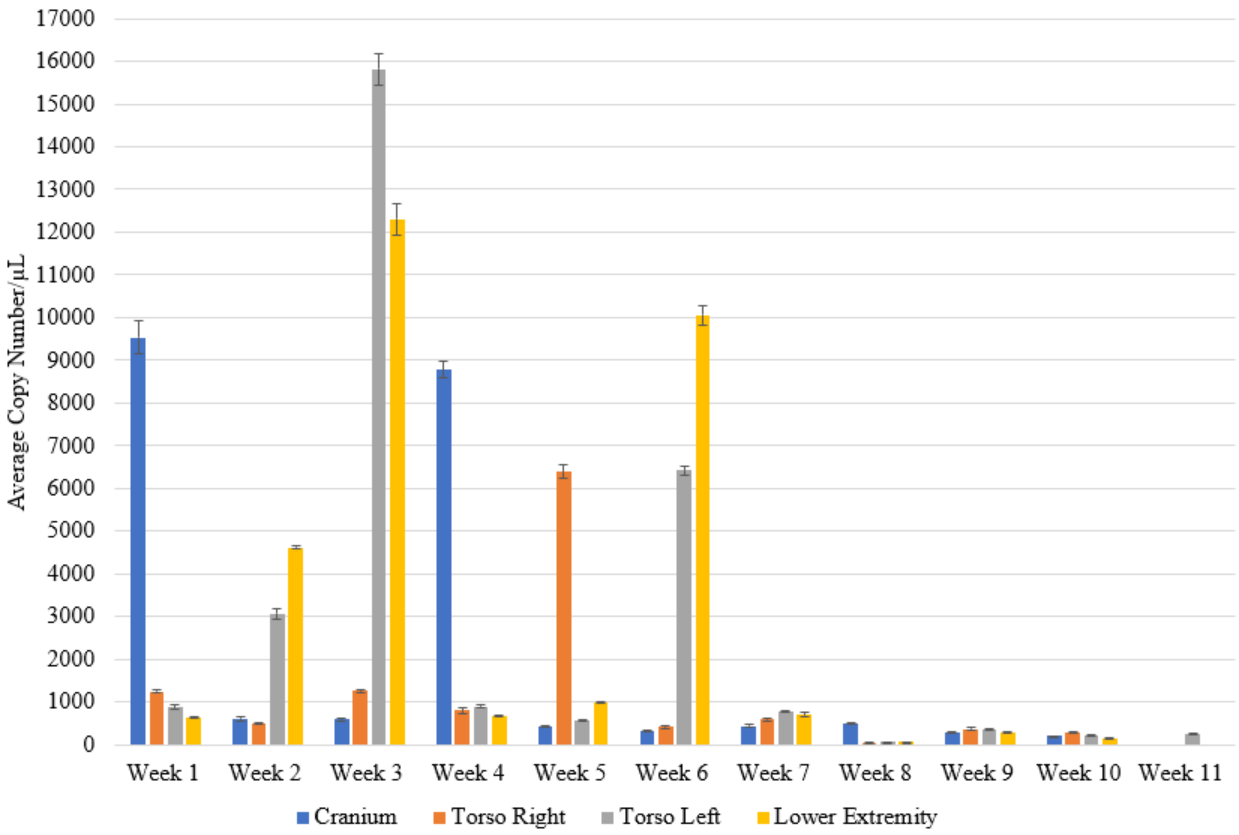

Figure 7. The average copy number per microliter of the short (105 bp) Target of the Kavlick mtDNA triplex assay amplified from soil samples taken from the plot of **Donor 1** over eleven weeks of decomposition. Standard error was calculated using the replicate quantitations of each four weekly samples per donor.

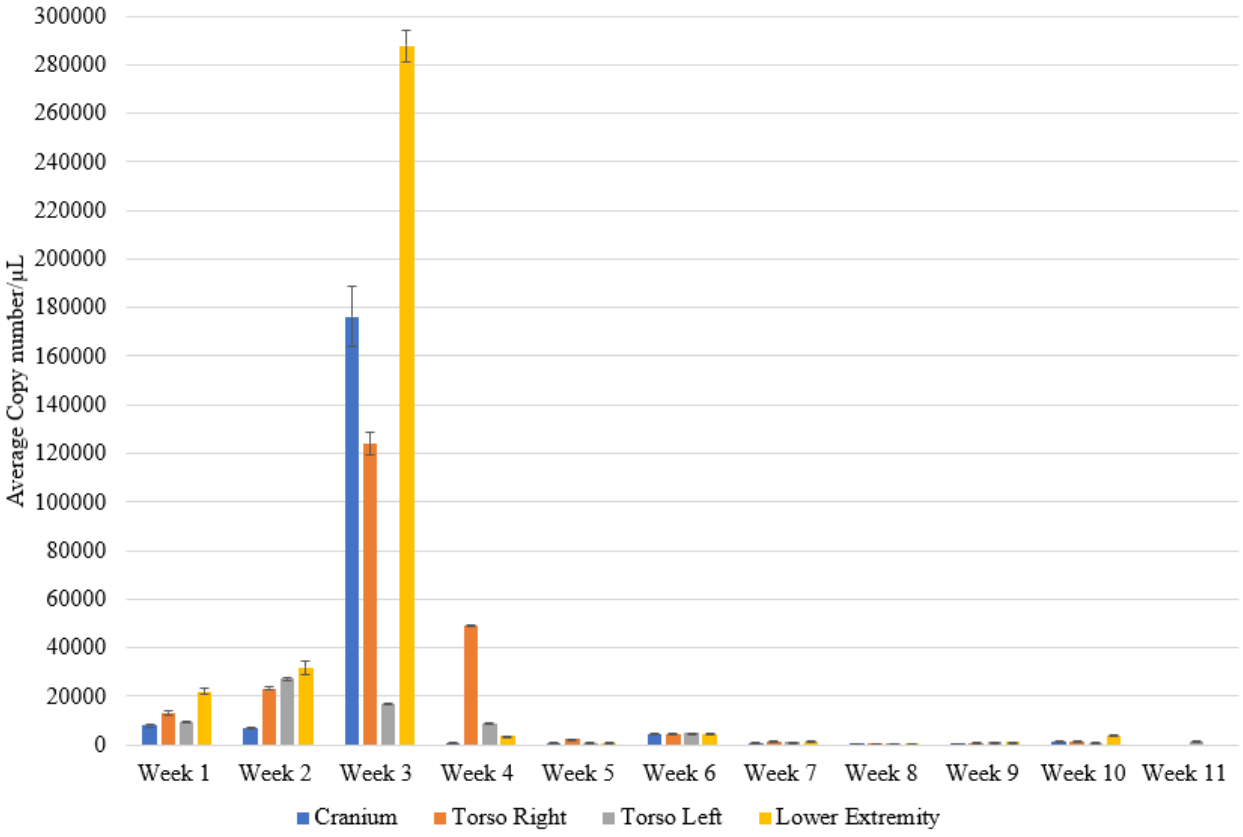

Figure 8. The average copy number per microliter of the short (105 bp) Target of the Kavlick mtDNA triplex assay amplified from soil samples taken from the plot of **Donor 2** over eleven weeks of decomposition. Standard error was calculated using the replicate quantitations of each four weekly samples per donor.

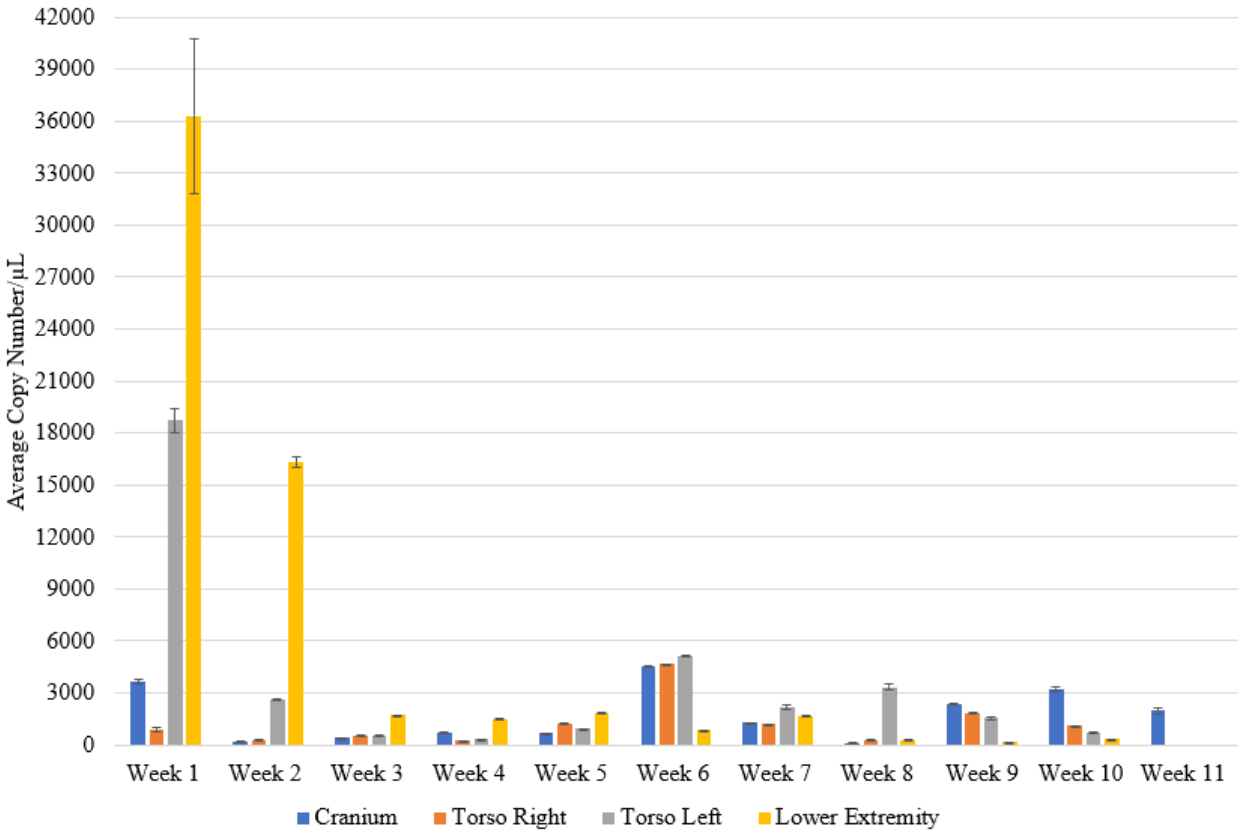

Figure 9. The average copy number per microliter of the short (105 bp) Target of the Kavlick mtDNA triplex assay amplified from soil samples taken from the plot of **Donor 3** over eleven weeks of decomposition. Standard error was calculated using the replicate quantitations of each four weekly samples per donor.

Copy number of the small 105bp target ranged from 53.0694 copies/µL found in soil from the right side of the torso of Donor 1 during the eighth week of decomposition to 28,7760.6 copies/µL found in soil right of the lower extremities of Donor 3 during the third week of decomposition. The quantity of mtDNA was highest during the first three weeks of the decomposition period and was detectable in minimal quantities through the remainder of the eleven weeks. Quantification of weekly samples from Donor 2 was generally an order of magnitude greater than those of Donors 1 and 3.

The degradation of mtDNA for each donor was assessed using Equation 3 to calculate ΔΔCt values, found in Table 11. The average Ct value for the small (105 bp) and large (316 bp) targets for each donor was calculated by taking the average of all three replicates of each of the four weekly samples per donor. Information on the Ct for each of these replicates can be found in Appendix A, Tables A7-12. The average Calibrator Ct values were calculated by taking the average of all replicates of the 20pg/µL HL60 positive control DNA for each corresponding run.

Table 11. Ct values and ΔΔCt calculations for soil samples from three donors across eleven weeks of surface-level decomposition using the Kavlick triplex mtDNA qPCR assay. * - indicates when there was no large target present to calculate ΔΔCt.

Donor	Week	Sample		Calibrator		ΔΔCt
		Average Ct (105 bp)	Average Ct (316 bp)	Average Ct (105 bp)	Average Ct (316 bp)	
Donor 1	1	27.9327	35.0961	24.6564	30.0982	1.7217
	2	31.0993	33.2415	25.8784	26.9797	1.0410
	3	27.3725	30.9891	24.1503	30.2703	-2.5035
	4	31.7507	33.5071	25.3173	26.5239	0.5498
	5	29.5636	34.5311	24.5537	26.5589	2.9623
	6	30.6174	33.2453	25.727	26.9782	1.3767
	7	30.3676	37.3093	24.3036	26.9267	4.3185
	8	29.8098	39.0249	21.0292	26.7955	3.4487
	9	32.2238	38.1639	25.0329	26.8787	4.0942
	10	33.9411	38.7213	26.3108	28.1167	2.9742
	11	33.8587	Undetermined	26.3108	28.1167	*
Donor 2	1	24.8219	30.8098	24.6564	30.0982	0.5461
	2	27.1202	28.6642	25.8784	26.9797	0.4427
	3	22.5518	26.5630	24.1503	30.2703	-2.1088
	4	29.8535	31.7044	25.3173	26.5239	0.6444
	5	29.6190	33.8933	24.5537	26.5589	2.2690
	6	29.1860	35.0682	25.727	26.9782	4.6320
	7	29.5044	35.6813	24.3036	26.9267	3.5538
	8	29.2901	37.3628	21.0292	26.7955	2.3063
	9	30.9803	35.1555	25.0329	26.8787	2.3295
	10	31.0233	34.3544	26.3108	28.1167	1.5252
	11	31.1827	37.6978	26.3108	28.1167	4.7091
Donor 3	1	25.7007	32.9339	24.6564	30.0982	1.7914
	2	31.1481	34.1331	25.8784	26.9797	1.8838
	3	30.1481	37.0190	24.1503	30.2703	0.7508
	4	33.2215	36.2404	25.3173	26.5239	1.8123
	5	29.6621	35.0380	24.5537	26.5589	3.3708
	6	29.7506	35.1514	25.727	26.9782	4.1496
	7	29.1306	36.7991	24.3036	26.9267	5.0454
	8	27.9293	36.8224	21.0292	26.7955	3.1267
	9	30.3522	34.4631	25.0329	26.8787	2.2651
	10	31.7804	35.1616	26.3108	28.1167	1.5753
	11	30.7379	35.4323	26.3108	28.1167	2.8884

There was degradation present in all weekly samples, except for samples during week three of decomposition for Donors 1 and 2. Donors 1 and 2 have ΔΔCt values of -2.5035 and

-2.1088 during the third week of decomposition, respectively. These negative values are interpreted as no degradation being present. Generally, there a greater ΔΔCt values further into the decomposition period indicating the degradation of mtDNA over time, but there is fluctuation within these values. There was no ΔΔCt calculation for Donor 1 during the eleventh week of decomposition, as the Ct value for the long target was undetermined, likely due to mtDNA degradation.

Next-Generation Sequencing

Sequencing of the entire 16,569 bp human mtGenome for weeks one through three in addition to sample Donor 2 TR-4 were performed by Verogen on a standard flow cell using the prepared purified library plate. Sequencing of weeks four through eight were performed in-house. The runs were analyzed in the Verogen UAS v2.0, which uses the revised Cambridge Reference Sequence (rCRS) as a reference comparison for each sample.[46] The analytical and interpretation thresholds for the run were both set at 0.06 with a minimum read count defined as 45 reads for an allele to be called. The positive HL60 control, negative control, and reagent blank all performed as expected. The summary of the run data for each sample is displayed in Table 12. All reference buccal swabs had over 95% coverage of the mtGenome besides buccal swabs from Donor 3, which only produced 81.8% coverage. Six samples: Donor 1 C-6, Donor 1 C-8, Donor 2 LE-7, Donor 2 LE-8, Donor 3 C-3, and Donor 3 C-7 did not have any base calls for the entirety of the mtGenome. Weeks four through eight which were sequenced in-house and failed on read two of the sequencing process and were thus sequenced a second time. On the second attempt, there was low cluster density of 344 k/mm^2. The results of the second run are shown in Tables 12 and 13.

Table 12. Summary metrics for experimental samples sequenced with the ForenSeq™ mtDNA Whole Genome Kit on the MiSeq FGx™. REF refers to sequences obtained from reference buccal swabs taken from donors upon intake.

Sample ID	mtGenome Coverage	Single Nucleotide Variants	Insertions	Deletions	No Call (bp)
Donor 1 REF	99.5%	37	1	2	89
Donor 1 C-1	5.3%	1	0	0	15,691
Donor 1 C-2	49.9%	19	0	0	8,308
Donor 1 LE-3	22.7%	10	0	0	12,811
Donor 1 TR-4	0.46%	0	0	0	16,492
Donor 1 TR-5	8.1%	4	0	0	15,220
Donor 1 C-6	0%	0	0	0	16,569
Donor 1 C-7	0.86%	0	0	0	16,424
Donor 1 C-8	0%	0	0	0	16,569
Donor 2 REF	96%	14	1	1	671
Donor 2 LE-1	64.8%	11	0	1	5,830
Donor 2 C-2	92.4%	16	0	1	1,265
Donor 2 C-3	85.4%	17	0	1	2,425
Donor 2 TR-4	98.8%	19	1	8	203
Donor 2 C-5	0.53%	0	0	0	16,482
Donor 2 C-6	1.1%	0	0	0	16,385
Donor 2 LE-7	0%	0	0	0	16,569
Donor 2 LE-8	0%	0	0	0	16,569
Donor 3 REF	81.8%	16	0	1	3,017
Donor 3 TL-1	53.3%	11	0	1	7,730
Donor 3 LE-2	39%	9	0	0	10,108
Donor 3 C-3	0%	0	0	0	16,569
Donor 3 LE-4	2.4%	2	0	0	16,176
Donor 3 TL-5	0%	0	0	0	16,569
Donor 3 LE-6	0.02%	0	0	0	16,565
Donor 3 C-7	1.7%	1	0	0	16,281
Donor 3 TL-8	12.2%	3	0	0	14,546

Information on the observed variants within the entirety of the mtGenome and haplogroup estimations are presented in Table 13. The majority of variants detected are crom the control region (CR) of the miGenome. The CR is roughly 1,122 bp in length and contains two hypervariable regions (HVI, HVII) that are highly polymorphic, allowing for potential human identification.[28] The haplogroups for the first three weekly samples from the plots of Donors 2

and 3 match reference buccal swabs. The haplogroups determined for Donor 3 C-7, and Donor 3 TL-8 fall within the same haplogroup cluster as the determinations from the first three weeks. The sample Donor 3 LE-4 was determined to be in haplogroup W1, however this sample was failed in Haplogrep 3 for containing minimal diagnostic loci. Donor 1 has variation in haplogroups, but all experimental samples profiles and the reference profile are within cluster K, aside for sample C-1 which fell within cluster U. As sample Donor 1 C-1 only had one polymorphism present, this may be the result of limited sequence coverage. Some samples were flagged if multiple unexpected polymorphisms or haplogroup quality was below the 0.9 quality threshold in Haplogrep 3. Some of these unexpected polymorphisms may be attributed to heteroplasmy, in which an mtDNA mutation arises from the presence of a mixture of mutant and wildtype mtDNA copies within a cell.[47] For instance, this may be present in Donor 2, where the 72Y mutation remains consistent across soil samples, suggesting that this donor has mtDNA genomes with either the base pair C or T at position 72. However, the high levels of mixed base calls indicate potential mixtures, as discussed later in this chapter. Samples Donor 1 LE-3, Donor 1 TR-5, Donor 2 TR-4, Donor 3 TL-1, and Donor 3 TL-8 are all marked as failed, as they missed the majority of expected polymorphisms, likely due to low sequence coverage set as 45 reads. The following samples had no variants detected due to lack of sequence coverage: Donor 1 TR-4, Donor 1 C-6, Donor 1 C-7, Donor 1 C-8, Donor 2 C-5, Donor 2 C-6, Donor 2 LE-7, Donor 2 LE-8, Donor 3 C-3, Donor 3 TL-5, and Donor LE-6.

Table 13. MtGenome variants and haplotype of mtDNA obtained from weekly soil samples using the ForenSeq™ mtDNA Whole Genome Kit on the MiSeq FGx™. REF refers to sequences obtained from reference buccal swabs taken from donors upon intake.

Sample ID	CR observed variants	Haplogroup
Donor 1 REF	73G, 146C, 152C, 263G, 315.1C, 498DEL, 543M, 750G, 1189C, 1438G, 1811G, 2706G, 3480G, 4769G, 7028T, 8860G, 9055A, 9093G, 9698C, 10398G, 10550G, 10938M, 11299C, 11377A, 11467G, 11719A, 12308G, 12372A, 13679M, 13683M, 13684M, 14167T, 14766T, 14798C, 15326G, 15900C, 16224C, 16311C, 16519C	K1c1b (Flagged)
Donor 1 C-1	11467G	U (Failed)
Donor 1 C-2	73G, 146C, 152C, 263G, 750G, 1438G, 1811G, 3480G, 8860G, 9093G, 10398G, 10550G, 11299C, 11377A, 11467G, 11719A, 15326G, 16311C, 16519C	K1c1 (Flagged)
Donor 1 LE-3	146C, 152C, 263G, 750G, 1811G, 10550G, 11467G, 15326G, 16311C, 16519C	K2a (Failed)
Donor 1 TR-5	146C, 152C, 750G, 11467G	K2a (Failed)
Donor 2 REF	263G, 750G, 1438G, 4769G, 5618C, 5899.1, 8723A, 8812G, 8860G, 13679M, 13684M, 15326G, 16258G, 16311C, 16519C	H104a
Donor 2 LE-1	72Y, 1750G, 4769G, 5618C, 8723R, 8812G, 8860G, 15326G, 6258G, 16311C, 16519C	H104a
Donor 2 C-2	72Y, 189R, 263G, 408W, 750G, 1438G, 4769G, 5618Y, 8723R, 8812R, 8860G, 11011R, 15326G, 16258G, 16311C, 16519C	H104a (Flagged)
Donor 2 C-3	72Y, 73R, 189R, 263G, 408W, 750G, 1438G, 4769G, 5618C, 8723A, 8812G, 8860G, 15326G, 16258G, 16311C, 16327Y, 16519C	H104a (Flagged)
Donor 2 TR-4	72Y, 189R, 263G, 297R, 310t, 311c, 312DEL, 313DEL, 314DEL, 315DEL, 316g, 319C, 408W, 750G, 1438G, 4769G, 5618C, 5899.1C, 7658R, 8723A, 8812G, 8857R, 8860G, 15326G, 16258G, 16311C, 16519C	H104a (Failed)
Donor 3 REF	152C, 263G, 709A, 750G, 1438G, 2259T, 2557T, 4745G, 4769G, 7337A, 8860G, 13326C, 13680T, 14831A, 14872T, 15326G	H13a1a1a*
Donor 3 TL-1	57K, 72Y, 73R, 152Y, 263G, 709R, 750G, 4745G, 13326C, 13680T, 15326G	H13a1a1a* (Failed)
Donor 3 LE-2	152C, 263G, 709A, 750G, 2259T, 4745G, 13326C, 13680T, 15326G	H13a1a1 (Flagged)
Donor 3 LE-4	7864Y, 13326C	W1 (Failed)
Donor 3 C-7	13326C	H1e1a2 (Failed)
Donor 3 TL-8	709A, 750G, 13326C	H13a1a1d (Failed)

To determine potential contamination within samples, mixed base calls were analyzed using the Verogen UAS v2.0 to look at read count, depth, and percent contribution of each allele call. If mixtures within a sample appeared to have a consistent major and minor contributor based on percent contribution of mixed base calls, the minor allele calls were analyzed in Haplogrep 3 to determine the haplogroup for the source of contamination. These results are presented in Table 14. The mixture 72Y which was consistently was excluded for samples Donor 2 C-2, Donor 2 C-3, and Donor 2 TR-4, as the C and T base calls were present in equal amounts with relatively high read depth. This could indicate that there was more than one source of contamination, as this variant was more pronounced than the minor alleles listed below. T in position 72 matches the rCRS, while C in position 72 is a variant that was diagnostic for the haplogroup H54* in Haplogrep 3.

Table 14. Minor alleles detected in mixtures within sequencing results of soil samples using the the ForenSeq™ mtDNA Whole Genome Kit on the MiSeq FGx™.

Sample ID	Minor Alleles in Mixture	Haplogroup
Donor 1 REF	543A, 10938A, 13679A, 13683A, 13684A	Flagged H2a2a1 (50%)
Donor 2 REF	13679A, 13684A	Flagged H2a21 (50%)
Donor 2 LE-1	8723G, 72C	Failed H54* (54%)
Donor 2 C-2	189G, 408A, 5618T, 8723G, 8812A, 11011G	Failed H2a2a1 (50%) H3v (38%)
Donor 2 C-3	73G, 189G, 408A, 16327T	Flagged H2a2a1 (50%) L3e1 (38%)
Donor 2 TR-4	189G, 297G, 408A, 7658G, 8857A	Flagged H2a2a1 (50%)
Donor 3 REF	57G, 72C, 73G, 152T, 709G	Failed H2a2a1 (50%) D1a1 (37%)
Donor 3 LE-4	7864Y	Flagged H2a2a1 (50%) W1 (30%)

By default, most determinations fell within the H2a2a1 haplogroup, as this is the haplogroup assigned to the rCRS. For some samples, such as Donor 1 REF and Donor 2 REF, none of the minor allele variants were known to the Haplogrep 3 database and thus could be the results of stochastic noise. Some samples have more than one determination if they had one or more allele variant present as a diagnostic polymorphism in a haplogroup other than H2a2a1. These determinations of minor contributors did not remain consistent across samples, suggesting that the mixtures may have already been present within soil samples prior to extraction, rather than occurring from one contributor during laboratory analysis.

Some other markers of note are 5899.1C and 16519C. The variant 5899.1C is an insertion found in both the reference sample and TR-4 from Donor 2 (Table 13) and is diagnostic for

haplogroup L3x1a1. As this is present in both reference and soil samples, this allele may have been introduced from lab personnel, as it does not fit the donor reference profile of H104a. The variant 16519C is a hotspot recognized within Haplogrep 3 that is present in all samples from Donor 2 and Donor 1 REF, C-2, and LE-3. As this variant occurs across multiple reference and experimental samples but does not fit within any sample determinations listed in Table 13, it is likely that this is also the result of contamination during laboratory analysis.

Statistical Analysis of Environmental Factors Effect on mtDNA Persistence

All data for statistical analysis on the effect of environmental factors on mtDNA persistence can be found in Appendix B, Table B2. The first two components of the PCA performed on environmental factors and decomposition data can explain 71.5% of the variation between samples (Figure 10). The correlation coefficients for each explanatory variable along each dimension are in Table 15. The two variables that explain the majority of the variation in dimension one are weekly average temperature (0.873) and TBS (0.671). For dimension two, most of the variation is explained by weekly total rainfall (0.717) and humidity (0.703), suggesting a moisture gradient. Dimension three was plotted against dimension one, shown in Figure 11, which together account for 53.9% of the variation between samples. The variables with the greatest correlation to dimension three are average humidity (-0.557) which is negatively correlated, and TBS (0.522) which is positively correlated.

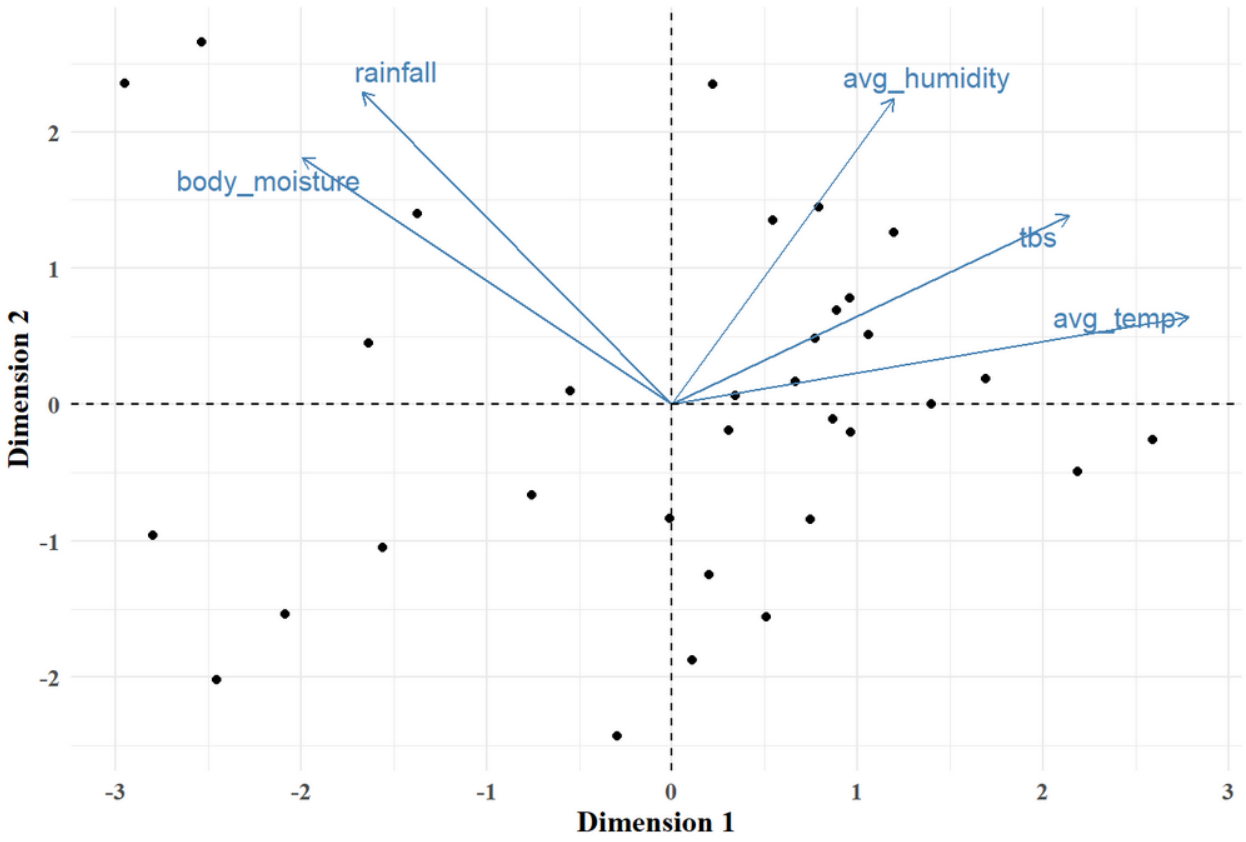

Figure 10. PCA biplot of dimensions one and two for eleven weeks of data collection from the human decomposition of three donors, including vectors indicating which environmental and decomposition factors were most important for differentiating samples.

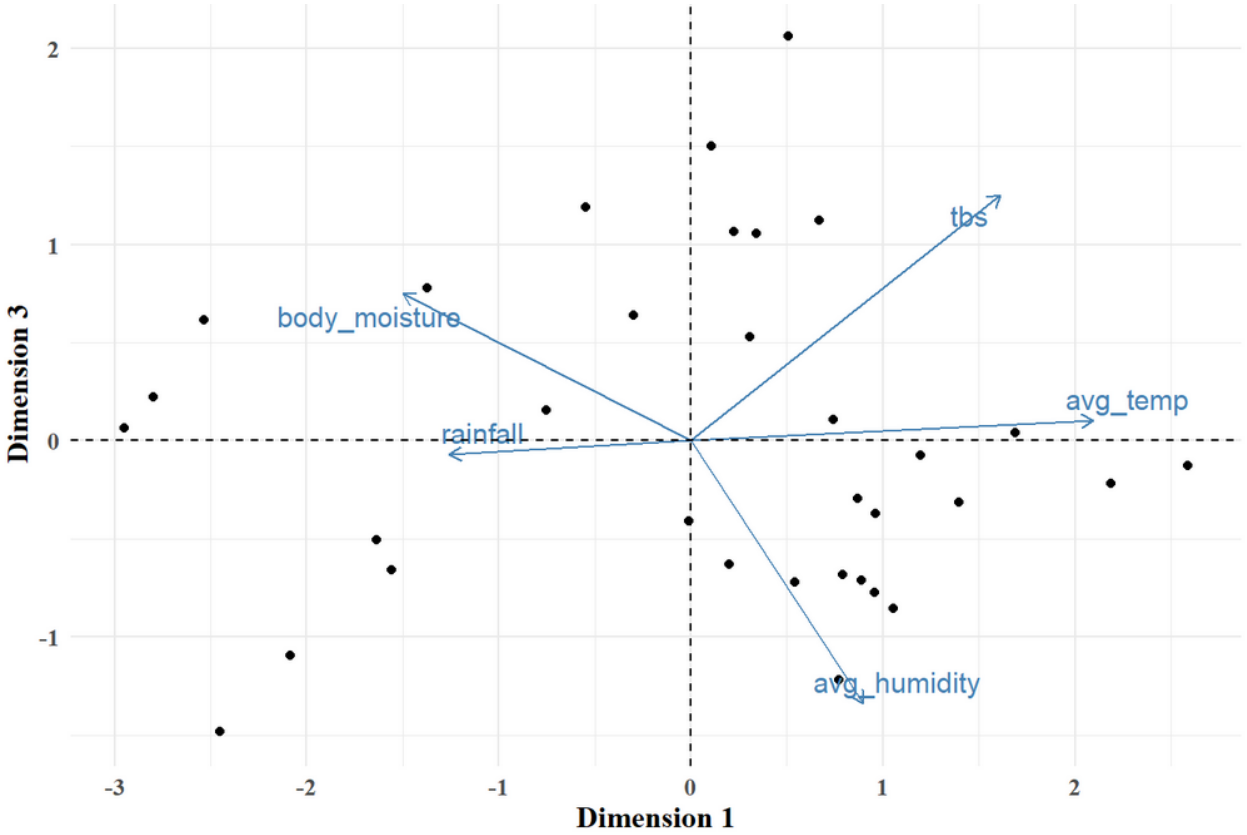

Figure 11. PCA biplot of dimensions one and three for eleven weeks of data collection from the human decomposition of three donors, including vectors indicating which environmental and decomposition factors were most important for differentiating samples.

Table 15. Correlation coefficients for principal components analysis of environmental and decomposition factors influencing weekly soil samples following eleven weeks of surface level decomposition.

Variable	Dimension 1 Correlation Coefficient	Dimension 2 Correlation Coefficient	Dimension 3 Correlation Coefficient
TBS	0.671	0.435	0.522
Body Moisture Content	-0.625	0.568	0.311
Avg Weekly Rainfall	-0.524	0.717	-0.029
Avg Weekly Humidity	0.374	0.703	-0.557
Avg Weekly Temperature	0.873	0.201	0.043

Linear regressions were performed to compare the first three dimensions of the PCA with the quantity of mtDNA in the soil per donor per week and analyzed with ANOVA, displayed in Table 16. There does not appear to be any statistically significant effect of any of the dimensions of the PCA on the quantity of mtDNA. However, dimension one has a *p*-value (*p*-value 0.06177) close to the alpha value of 0.05. This suggests that there may be a slight negative correlation between dimension one of the PCA and the quantity of mtDNA, as shown in Figure 12.

Table 16. ANOVA of linear regressions between each dimension of the PCA and concentration of mtDNA.

	DF	**Sum Sq**	**Mean Sq**	**F Value**	**Pr(>F)**
Dimension 1	1	2.3693e+09	2.3696e+09	3.756	0.06177
Residuals	31	1.9557e+10	6.3088e+08		
Dimension 2	1	7.0125e+07	7.0125e+07	0.0995	0.7546
Residuals	31	2.1857e+10	7.0506e+8		
Dimension 3	1	3.7518e+08	3.7518e+08	0.5397	0.4681
Residuals	31	2.1552e+10	6.9522e+08		

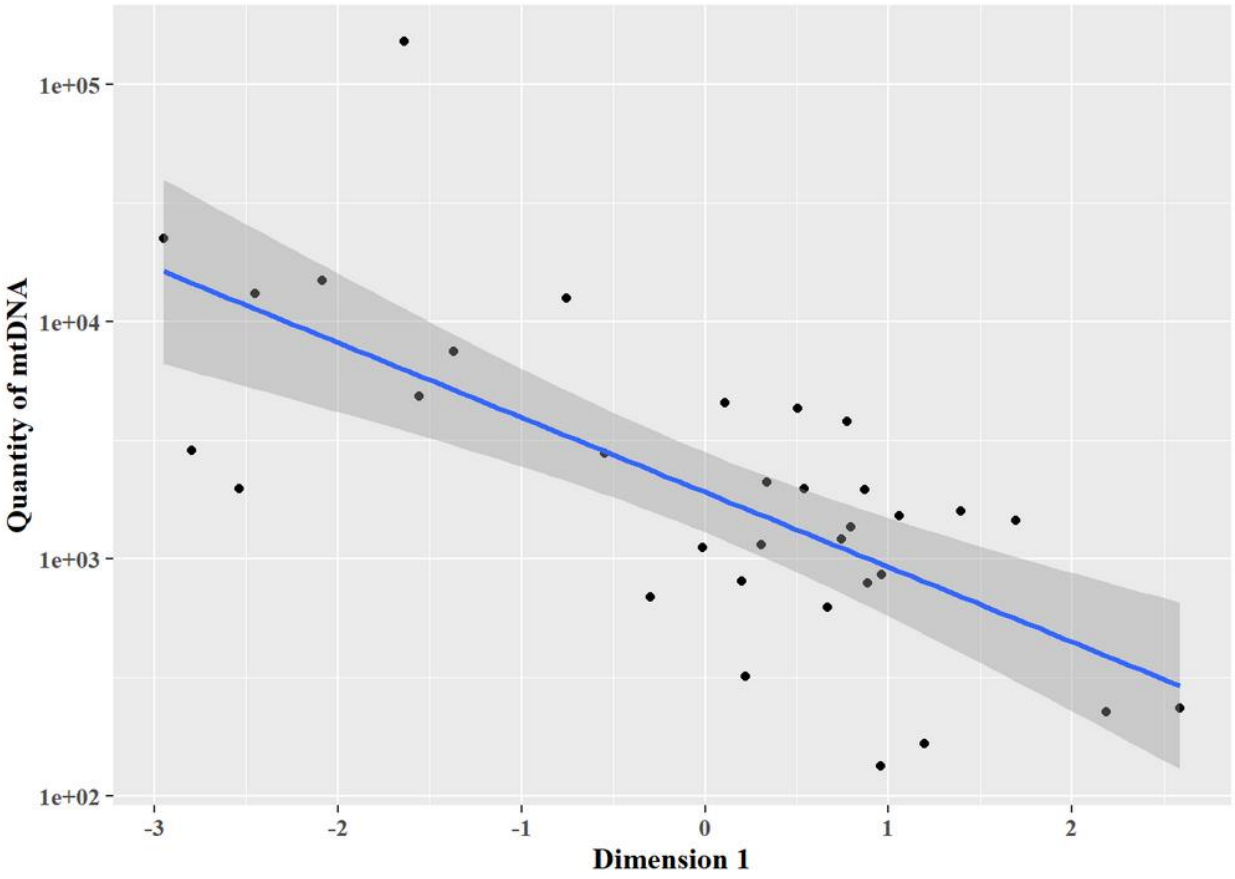

Figure 12. Linear regression of concentration of mtDNA in weekly soil samples compared to dimension one of the PCA. The y-axis has been converted to a logarithmic scale. The grey shading represents the 95% confidence interval and each soil sample per donor is represented by an individual black dot.

Spearman's rank correlations between environmental variables and the copy number of mtDNA are shown in Table 17. There are statistically negative correlations between the number of weeks into the decomposition period and mtDNA copy number (Spearman's ρ = -0.6219, *p*-value = 0.0001), TBS of remains and mtDNA copy number (Spearman's ρ = -0.4687, *p*-value = 0.0059), average weekly humidity and mtDNA copy number (Spearman's ρ = -0.3439, *p*-value = 0.0500), and average weekly temperature and mtDNA copy number (Spearman's ρ = -0.7200, *p*-value = 2.3245E-06). All other ranked correlation calculations were statistically insignificant.

Table 17. Spearman's rank correlation coefficients (Spearman's ρ) between weekly environmental factors and copy number of mtDNA derived from weekly soil samples. Correlations with significant *p*-values (≤ 0.05) are in bold.

Relationship	Spearman's ρ	T-Statistic	DF	*p*-value
Copy number mtDNA with Week into Decomposition Period	**-0.6219**	**4.422**	**31**	**0.0001**
Copy number mtDNA with TBS	**-0.4687**	**2.9539**	**31**	**0.0059**
Copy number mtDNA with Weekly Body Moisture Content	0.3274	1.9290	31	0.0629
Copy number mtDNA with Total Weekly Rainfall	0.1275	0.7160	31	0.4794
Copy number mtDNA with Weekly Average Humidity	**-0.3439**	**2.0391**	**31**	**0.0500**
Copy number mtDNA with Weekly Average Temperature	**-0.7200**	**5.7759**	**31**	**2.3245E-06**

CHAPTER FOUR: DISCUSSION AND CONCLUSION

This study investigated the persistence of human cadaveric DNA in soil following the surface-level decomposition of three donors across eleven weeks of decomposition. Human nuDNA was quantified using the Quantifiler™ Trio DNA Quantification Kit while mtDNA was quantified using the multiplex qPCR assay described by Kavlick[34] to assess the duration of DNA persistence in soil, the quantity present, and the degradation of DNA. Both nuDNA and mtDNA extracted from soil samples were sequenced using NGS methods on the MiSeq FGx™, using the ForenSeq™ DNA Signature Prep Kit for the nuDNA and the ForenSeq™ mtDNA Whole Genome Kit for mtDNA. This approach allowed for testing concordance between donor reference and experimental sample DNA profiles and assessment of the utility of using soil-derived DNA to produce profiles for human identification. Statistical analysis of environmental factors including time, TBS, body moisture content, weekly total rainfall, weekly average temperature, and weekly average humidity effect on the quantity of DNA recovered from the soil was also performed to provide context regarding factors impacting the persistence of DNA in soil.

Quantitation of DNA from Soil

Nuclear DNA Recovery

nuDNA was only recoverable from the soil surrounding surface level decomposition for six weeks into the decomposition period. Recovery of small autosomal target DNA ranged from undetectable in week four for both Donors 1 and 3, to 7.07E-03 ng/µl during week one for Donor 2 (Table 4). Generally, small autosomal target DNA was more recoverable than large autosomal target DNA, which was present in quantities ranging from undetectable to 3.74E-03

ng/µl during week one of the decomposition period for Donor 2 (Table 4). The amplicon length of the small autosomal target in the Quantifiler™ Trio DNA Quantification Kit is 80 bases in length, whereas the large autosomal target amplicon length is 214 bases.[33] Double-stranded breaks from nuclease activity and oxidation could have contributed to heavy fragmentation of the nuDNA in the soil, resulting in higher quantities of the small target DNA being present in soil samples as opposed to the large target. This is generally supported by the DI calculated in Table 4 increasing over the six weeks until the large autosomal target was no longer recoverable. However, for all three donors, higher quantities of small autosomal and large autosomal target DNA became recoverable during week six of sampling. Rainfall, high relative humidity, and elevated body moisture content observed during week six for all donors (Appendix B, Table B1) may be responsible for the resurgence of nuDNA into the soil resulting in higher recovery. During advanced decomposition, the remains dry, and purge fluids diminish.[2] However, it was observed that the remains in this study would absorb moisture after rainfalls, causing fluctuations in TBS as well as allowing additional fluids to seep into the soil after the initial purge. This could lead to the release of more cellular material (i.e., more recoverable genetic material) into the surrounding environment. This phenomenon has been observed in other temperate climates in South Africa during a decomposition study on *Sus scrofa*, in which pig remains would absorb water from rainfall which would rehydrate the flesh and cause further decomposition and insect activity.[48] Y-Chromosomal target DNA was also detectable in the first two weeks of soil samples for Donors 2 and 3, which matched donor biological profiles, providing positive sex estimation of soil-derived extracts. While Y-STR analysis was not conducted as part of this study, this finding supports the potential for Y-chromosomal testing in subsequent studies.

Mitochondrial DNA Recovery

mtDNA was recoverable from the surrounding soil for all eleven weeks of the decomposition period for all donors. This finding expands upon Emmons et al., demonstrating the recoverability of mtDNA for an additional three weeks beyond previously published data that showed recoverability for eight weeks. Recovery of the small 105bp target ranged from 53.0694 copies/µL from the soil surrounding the right side of the torso of Donor 1 during the eighth week of decomposition to 28,7760.6 copies/µL from the soil on the right side of the lower extremities of Donor 3 during the third week of decomposition (Table 10). Though detectable for all eleven weeks, mtDNA was found in the highest quantities for all donors during the first three weeks of the decomposition period and decreased over time, likely due to the degradation of mtDNA in the soil due to nuclease activity or oxidation. This is supported by the ΔΔCt values that tend to increase into the decomposition period (Table 11). Interestingly, Donors 1 and 2 have ΔΔCt values of -2.5035 and -2.1088 respectively, during the third week of decomposition, indicating no degradation being present (Table 11). This could be explained by the donors being in active decay and having recently purged decomposition fluids into the soil, therefore the mtDNA present in the soil has not been exposed to the extracellular environment for extended periods. Additionally, there was a noted slight increase in the quantity of mtDNA in the soil during week six, as seen in the nuDNA quantitation results (Table 10). Likewise, this could be due to extended periods of increased moisture around the remains, rehydrating them and causing a purge of decomposition byproducts into the soil (Appendix B, Table B2).

Overall Trends in DNA Recovery

Human nuclear and mitochondrial DNA is shown to be found in the highest quantities within the soil during the first three weeks of the decomposition period and decreases in

quantity over time. Furthermore, mtDNA appears to persist in the soil for longer periods and be largely more recoverable than nuDNA, similar to the findings in Emmons et al.[24] This is likely attributed to the robust circular structure of mtDNA causing it to be more resistant to nuclease activity, and high copy number per cell, as there are typically 1,000 mitochondria per cell with two to ten copies of the mtGenome per mitochondria, opposed to the two copies of nuDNA per human cell.[47]

Notably, soils surrounding Donor 2 consistently contained higher quantities of both nuclear and mitochondrial DNA than Donors 1 and 3. Compared to Donors 1 and 3, Donor 2 was greater in mass upon intake. Body mass is an intrinsic factor that can influence the rate of decomposition and has been observed to either increase the rate of decomposition due to the liquefaction of body fat, which has been demonstrated at the Anthropology Research Facility at The University of Tennessee, Knoxville[49] or slow the rate of decomposition due to a plateau phase as seen in a comparative study of pig carcasses at the Forensic Anthropology Body Farm at the University of Pretoria, South Africa.[48] Though there was an extended plateau in decomposition between weeks five and ten of the study for Donor 2, this was also observed in the other two donors (Appendix B, Table B2). Likewise, there did not appear to be differences in the rate of decomposition. However, the increased liquefaction of body fat could have provided increased amounts of purge fluids into the soil and deposition of more genetic material.

Next Generation Sequencing of DNA from Soil

ForenSeq™ DNA Signature Prep Kit

Sequencing with the ForenSeq™ DNA Signature Prep kit was performed on the first six weeks of soil samples. This kit targets 27 autosomal STRs, 24 Y STRs, 7 X STRs, and 94

iSNPs with amplicon lengths ranging from 61- 402 bases to aid in human identification.[26] Though the kit has a wide array of amplicon lengths to accommodate for heavily fragmented samples, nuDNA loci were only detected in three samples from Donor 2, all from the first two weeks of decomposition (Table 5). The loci that were detected in these samples were minimal, providing little genetic information on the donor. Of the samples with loci detected, the sample LE-1 from Donor 2 was flagged on the TH01 locus for detecting multiple alleles, exceeding the stutter threshold, and having a non-stutter allele between the analytical and interpretation threshold (Table 6). Furthermore, the positive control for the run only had 26/27 autosomal STRs and 23/24 Y STRs detected.

Though the nuDNA was heavily degraded, it was still expected that there would be better sequencing results given the sensitivity of the ForenSeq™ DNA Signature Prep. One potential explanation for the resulting data could be due to technical issues with the MiSeq FGx™ instrument used, as it was later discovered that the chiller component on the instrument was malfunctioning. This could have caused issues with temperature control, affecting the bridge PCR step of sequencing by causing issues with primer binding and extension of new DNA strands. However, reads were present for the majority of targets for the positive control of the sequencing one, indicating that there could have been potential inhibition within the DNA extracts. Though inhibition was not detected within the qPCR results from the soil extracts, the Quantifiler™ Trio kit is robust to common PCR inhibitors such as humic acid. In a developmental validation study performed by Holt et al. for the Quantifiler™ Trio kit, various concentrations of humic acid ranging from 200 ng/µl – 800 ng/µl were used to spike 0.1 ng/µl 007 human gDNA standards. They found that all targets were detected with concentrations of

humic acid as high as 300 ng/µl and inhibition being detected at quantities of 400 ng/µl – 800 ng/µl.[50]

The ForenSeq™ DNA Signature Prep on the other hand is not as robust to inhibition from humic acid. In a study on the effects of common inhibitors on sequencing results using the ForenSeq™ DNA Signature Prep Kit, it was shown that the total number of SNPs detected dropped to 9% with an average read depth of 55X when a humic acid concentration of 17 ng/µl was added to gDNA samples.[51] Similarly, a study by Sidstedt et al. found that humic acid concentrations as low as 5 ng/µl added to gDNA samples completely inhibited the ability to detect all autosomal STR targets while using the ForenSeq™ DNA Signature Prep Kit.[52] Though the QIAGEN DNeasy® PowerSoil® Pro kit contains wash steps to reduce the presence of common PCR inhibitors found within soil, the resulting extracts could still contain small quantities of inhibitors that may hinder sequencing capabilities. Alteration of PCR chemistries during the initial amplification phase of sequencing may help in troubleshooting low sequence coverage from soil-derived DNA extracts. Sidstedt et al. recommend the addition of 10µg of bovine serum albumin (BSA) to the master mix of PCR1, as it increased humic-acid tolerance and loci coverage in samples without hindering sequencing results of non-inhibited positive controls.[52]

ForenSeq™ mtDNA Whole Genome Kit

MtDNA was sequenced using the ForenSeq™ mtDNA Whole Genome Kit for the first three weeks of soil samples for all donors, including one sample from the fourth week of decomposition for Donor 2. Soil-derived samples had variable mtGenome coverage, ranging from no coverage in the third week of decomposition soil surrounding the left side of the cranium for Donor 3 to 98.8% coverage from the soil surrounding the right side of the torso of

Donor 2 during the fourth week of decomposition (Table 12). Overall, samples from Donor 2 had the highest quality of coverage. Since the maximum amount of input DNA for the ForenSeq™ mtDNA Whole Genome Kit is 6µl of 8.33pg/µl concentrated DNA,[27] experimental soil samples were undiluted. Since Donor 2 had the highest quantity of DNA in the surrounding soil overall, this could have led to better sequencing results as these quantities were closest to the manufacturer-recommended concentration of input DNA. Sensitivity studies have demonstrated the utility of the ForenSeq™ mtDNA Whole Genome Kit using dilutions of HL60 control DNA as low as 2 pg/µL with over 300,000 reads and the ability to detect at least 40% of variants when using two bead purification steps.[28] Though soil samples surrounding Donor 2 contained degraded DNA with concentrations less than 2pg/µl of DNA, the high sensitivity of the ForenSeq™ mtDNA Whole Genome Kit was still capable of retrieving ample diagnostic polymorphisms to determine haplogroup.

Polymorphisms within the control region of the mtGenome were analyzed using Haplogrep 3 software to determine mitochondrial haplogroup. The following haplogroups were determined for donor reference profiles: Donor 1 K1c1b, Donor 2 H104a, and Donor 3 H13a1a1a* (Table 13). Upon comparison to experimental samples, the haplogroups for Donors 2 and 3 match the reference profile within the first three weeks of the decomposition period. However, Donor 3 LE-4 and TL-8 vary from this determination as they are missing definitive polymorphisms, but fall within the same cluster as the reference haplogroup, likely due to the threshold of 45 reads and highly degraded nature of the input DNA. Donor 1 on the other hand had varying haplogroup estimations, but all determinations fell within cluster K. In some circumstances, samples were flagged because they contained multiple unexpected polymorphisms, or the sequencing information was below the 0.9 quality threshold in

Haplogrep 3. In some instances, these unexpected polymorphisms may be attributed to low levels of heteroplasmy present, where wildtype and mutated mtDNA coexist in an individual's mitochondria. Heteroplasmy can be intra-personal and inherited at low levels from the maternal parent.[46] However, the rates of these mixed allele calls were observed to be high, and many mixtures may be the result of contamination rather than heteroplasmy. In Table 14, the minor alleles are shown with potential haplogroup determinations. It should be noted that all samples from Donor 2 contained mixtures with varying haplogroup determinations for potential minor contributors. This may be due to the placement of Donor 2 at the FOREST being at a midway point on the slope next to a walkway. During rain, this walkway would often become a path for runoff, which could lead to mixtures being present within soil samples as decomposition fluids from other donors located uphill may enter the plot of Donor 2. Donor 1 was placed at the top of the slope, which likely prevented contamination from other donors. Donor 3 was at the bottom of the slope and some mixture was present in a soil sample from week four. Due to the small size and nature of the facility, many of these plots are used in rotation as new donors arrive, therefore, some contamination may also be the result of carcass enrichment in the soil prior to placement. Soil samples were taken from each plot prior to placement, which have not been extracted and analyzed for DNA quantitiation or sequencing, due to lack of supplies. These baseline samples should be analyzed and compared to current sequencing data to determine if DNA from previous donors was already present within the plots.

Furthermore, some contamination may be attributed to personnel present at the FOREST or during laboratory analysis. The variants 5899.1C and 16519C were found in both reference and experimental samples, and 16519C was also detected in both Donors 1 and 2, but were not diagnostic for haplogroup determinations found in Table 13. 5899.1C is diagnostic for

haplogroup L3x1a1 while 16519C is a common hotspot that occurs multiple times within the phylogenetic tree. This suggests that the source of contamination was introduced from a common source likely after the soil samples were collected. Reference buccal swabs have been taken of individuals that have contributed to this project at the FOREST and within the lab and will later be sequenced to determine the source of this contamination.

Five samples were marked as failed by Haplogrep 3, as they were missing diagnostic polymorphisms for haplogroup assignment. This could be due to low breadth coverage for sample LE-3 from Donor 1, as there was only 22.7% mtGenome coverage (Table 12), which may be due to DNA degradation. In Donor 2 TR-4, there are ambiguous calls at loci positions 310, 311, and 316, along with deletions from 312-315 (Table 13). This is likely due to a low depth of coverage below the Verogen UAS software threshold of 45 reads. Donor 3 TL-1 on the other hand seems to have low levels of spurious contamination, as there are three loci with mixtures present (57K, 72Y, and 73R) that do not appear in any other samples or donor reference (Table 13). Donor 1 TR-5 and Donor 3 TL-8 were both present on a sequencing run in which read two failed, resulting in a lack of sequencing results.

Issues Influencing Sequencing

As mentioned previously, there were technical issues with the MiSeq FGx™ used to sequence the nuDNA. The issues with the chiller not reaching proper cooling temperatures could have impacted primer binding and DNA extension, resulting in the minimal loci calls observed despite the sufficient quality score of the run. Furthermore, the presence of the humic acid within samples extracts would also have a significant impact on sequencing results, as it is a PCR inhibitor known to significantly impact the ability to detect loci and limit read depth.[51-52] The mtDNA samples from weeks one through three in addition to sample Donor 2 TR-4 were

sequenced on a different MiSeq FGx™ by Verogen (San Diego, CA), and resulted in higher quality sequencing data. However, this could also be due to the higher quantity of mtDNA present and lower levels of extracellular DNA degradation. The mtDNA run sequenced in-house from weeks four through eight had to be sequenced twice due to a failure in the first run of read two. Read two failing could have been the result of the fluidics in the instrument failing or the quality of the reagent cartridge.[36] A second run was performed on a new reagent cartridge, which resolved the failure on read two. However, cluster density was low, still resulting in a poor-quality sequencing run. The low cluster density on the second attempt to sequence could be the result of the loss of cluster density in the normalization plate over time, though it had not yet met its 30-day expiration date.[36] Due to prior instrument failure with the chiller component and read two failure, extracts should be sequenced externally to determine whether DNA concordance is possible after three weeks. Future research should also compare the results of various next-generation sequencing kits and platforms on soil-derived DNA to determine which commercially available kits and other non-commercial approaches have the highest sensitivity for sequencing heavily degraded samples.

Utility of Soil-Derived DNA Profiles

Currently, it appears that nuDNA derived from soil surrounding surface level decomposition is not a viable route for producing individually identifying DNA profiles. However, mtDNA provides better sequencing results and is in concordance with reference profiles for up to three weeks into the decomposition period. The ability to determine haplogroups from soil-derived mtDNA can also provide ancestry estimations of individuals along the maternal lineage. This might prove useful in forensic contexts to identify the presence

of clandestine graves if remains were later moved and the biological profile information of the individual that was once buried there was known.

Effects of Environmental Factors on DNA Persistence

Nuclear DNA

There is a strong statistically significant negative correlation between dimension two (correlated with TBS and average weekly temperature) of the PCA for the first six weeks of environmental factors and the quantity of nuDNA found in soil (*p*-value = 9.696e-05) as shown in Table 8. The variation in dimension two of the PCA is most closely correlated to the average weekly temperature and TBS (Table 7). Temperature is known to influence the rate of decomposition, as higher temperatures typically contribute to faster rates of decomposition[3] and will encourage microbial activity in the soil, contributing to potential nuclease activity. It is also expected that the further into advanced decay the remains are, the less likely DNA will be recoverable from the soil. This is further supported by Spearman's rank correlation results (Table 9) in which the relationship between the quantity of nuDNA found in the soil is significantly correlated with weekly average temperature (Spearman's ρ = -0.6184, *p*-value = 0.0062) and TBS (Spearman's ρ = -0.4977, *p*-value = 0.0356). Additionally, time in weeks into the decomposition period was significantly correlated with the quantity of nuDNA recovered from soil samples (Spearman's ρ = -0.6594, *p*-value = 0.0029). A similar study by Emmons et al. found that nuDNA was largely not recoverable from the soil across later stages of decomposition,[24] which was widely reflected within these results.

Mitochondrial DNA

When examining the correlations between the dimensions of the PCA of eleven weeks of environmental factors and the quantity of mtDNA, there are no statistically significant

relationships. However, the correlation between dimension one (correlated with eleven weeks of average weekly temperature and TBS data) and the quantity of mtDNA is close to the alpha of $p \leq 0.05$, with a *p*-value of 0.06177 (Table 16). The variation in dimension one of the PCA is most closely correlated with average weekly temperature and TBS. Though not statistically significant, these data are the result of a small sample size, and these factors may still be important driving forces in the degradation of mtDNA within the soil. This was similar to the results of the correlations between PCA dimensions and concentration of nuDNA in the soil, in which dimension two had a strong statistically significant correlation with nuDNA quantity, and was explained by temperature and TBS.

Since PCA can be sensitive to small sample size and outliers, a non-parametric Spearman's rank correlation test was also conducted. Correlations using Spearman's rank test between the copy number per microliter of mtDNA in the soil and different environmental factors show significant relationships between mtDNA copy number and time into the decomposition period (Spearman's $\rho = -0.6219$, *p*-value = 0.0001), TBS (Spearman's $\rho = -0.4687$, *p*-value = 0.0059), average weekly humidity (Spearman's $\rho = -0.3439$, *p*-value = 0.0500), and average weekly temperature (Spearman's $\rho = -0.7200$, *p*-value = 2.3245E-06). This partially supports the results of the correlation between dimensions of the PCA and quantity of mtDNA recovered, as temperature and TBS have a statistically significant correlation with mtDNA copy number per microliter. The correlation between weekly average humidity and copy number per microliter of mtDNA could be explained by moisture's role in the decomposition. Typically, more moisture present within soils can increase decomposition rates due to increased microbial activity[17], however, once the soil matric potential (SMP) is exceeded, decomposition may be hindered.[19] This suggests humidity may influence the quantity

of recoverable mtDNA in soil because it is closely correlated with microbial presence, and therefore, nuclease activity.

General Trends Between Quantity of DNA and Environmental Factors

The results of the statistical analysis of environmental factors on the concentration of DNA in the soil were similar between nuDNA and mtDNA. These results suggest that temperature, TBS, and time exposed to the external environment all have a significant correlation to the concentration of human DNA within the soil. This is supported by the literature, as high temperatures are shown to increase decomposition rates and microbial activity within the soil,[3,8,18] and advanced stages of decay further into the decomposition period are associated with the inability to recover nuDNA.[24] Humidity was shown to be significantly correlated with the quantity of mtDNA in the soil, but not nuDNA. Moisture likely plays a role in the persistence and degradation of DNA within the soil as high moisture levels typically have a positive impact on the rate of decomposition[3,19], promote microbial activity[17], and will influence the enzymatic potential for nuclease enzymes. This effect may be more evident in a larger sample size with more data points, as donors and time were limited for this study.

Applications: Forensic Investigation and Paleoanthropological Work

Downstream applications of soil-derived DNA extracts are limited by quantity, quality, and type of DNA recovered along with the amount of time extracellular DNA has been present within the soil. Since extracellular DNA within the soil is prone to degradation and heavy fragmentation caused by previously discussed variables, it limits the ability to sequence and retrieve genetic information from these extracts. NGS approaches are preferable due to the variety of high-sensitivity sequencing kits with a suite of short amplicons to target heavily fragmented DNA. Results from this and the Emmons et al. study do not indicated that nuDNA

derived from soil surrounding decomposing human remains can be a viable route for positive human identification as it is largely unrecoverable after early stages of decomposition and does not provide quality sequencing data for human identification.[24] Rather, mtDNA recovered from soil was able to be recovered up to eleven weeks into the decomposition period and was able to provide haplogroup data within concordance with donor reference profiles up to three weeks into the decomposition period. This could be useful in forensic contexts in conjunction with thanatomicrobiome analysis of soils to identify the presence of clandestine graves if remains have been moved by determining the presence of human decomposition from the recovery of human mtDNA in the soil and microbial succession signatures.[7] Further research are needed to determine best practices for soil sampling, extraction, and sequencing.

Furthermore, this type of research is beneficial to understand the complex mechanisms underlying the persistence of mtDNA within soil and sediments. As previously stated, given the proper environment, ancient hominin mtDNA can persist for thousands of years and still provide genetic information under unique environmental conditions.[11-14] Understanding how this DNA has been preserved for so long can help aid in discovering the best approaches to recovering ancient mtDNA and further paleoanthropological phylogenetic research. In the contemporary context of this study, it appears that some of the main factors influencing the persistence of mtDNA in soil are time, stage of decomposition, and temperature as they are likely influencing microbial and enzymatic activity leading to DNA degradation. The role of moisture on the persistence of DNA in soil is not fully understood, though it likely plays a nuanced role as it affects the rate of human decomposition as well as in DNA degradation. Further research needs to be performed on these factors to better understand their complex interactions and their effects on the recoverability of human DNA from soil.

Future Directions

Due to instrument complications and unexpected low-quality sequencing results of nuDNA, additional sequencing runs are suggested. Similarly, the author recommends comparative studies on different NGS protocols (commercial kits and other approaches) for sequencing soil-derived DNA extracts. Similar NGS platforms, such as the Ion Chef™ Instrument, could be used in comparative studies alongside the MiSeq FGx™ to determine which method has the highest sensitivity and reproducibility when sequencing soil-derived DNA profiles. The Ion S5™ Precision ID Identity Panel has contains 34 Y-clade SNPs and 90 autosomal SNPs with an average amplicon size of 138bp, making it ideal for heavily fragmented nuDNA.[53] The Precision ID Identity Panel has been shown to produce greater than 1500 X coverage of autosomal SNPs and greater than 780 X coverage of Y-clade SNPs for nuDNA concentrations as low as 0.2ng/µl with 100% congruence amongst genotype calls.[54] For mtDNA, the Ion S5™ Precision ID mtDNA Whole Genome Panel can provide whole mtGenome coverage with as little as 2pg/µl of input DNA.[55] In a study looking at the utility of the Precision ID mtDNA Whole Genome Panel on sequencing heavily degraded DNA from 52 skeletal remains aged at approximately 50 years, Ta et al. found that they could obtain 50 full and two partial control region-based haplotypes with a mean coverage of 2494 X in samples, even when nuDNA was unrecoverable.[56] Given the similarities between these kits with the ForenSeq™ DNA Signature Prep Kit and the ForenSeq™ mtDNA Whole Genome Kit for the MiSeq FGx™, they would be good candidates for future research in determining best practices for sequencing soil-derived DNA.

Future research also needs to be performed on the effects of environmental factors on the persistence of DNA in soil. Though this study focused primarily on climatic data, soil

edaphic factors such as pH, soil moisture, electrical conductivity, and cation concentrations would be essential to understanding how DNA could potentially bind to the soil matrix. This research would need to be performed at multiple human decomposition facilities across various seasons to understand the nuanced effects of seasonality, regional climate, and different soil types on the persistence of human DNA within the soil. With a larger sample size, across multiple facilities, these data could help build predictive modeling to determine whether DNA would be able to persist under certain environmental conditions and provide a better understanding of the mechanisms driving the persistence and degradation of human DNA within soil.

REFERENCES

1. Christensen, AM, Passalacqua, NV, Bartelink, EJ. Forensic taphonomy. In: Forensic Anthropology: Current methods and practice. 2nd ed. Elsevier Inc.; 2019. p. 145-182.

2. Megyesi, MS, Nawrocki, SP, Haskell, NH. Using accumulated degree-days to estimate the postmortem interval from decomposed human remains. *J. Forensic Sci*. 2005;50(3):618–626.

3. Carter, DO, Tibbett M. Cadaver decomposition and soil: Processes. In: Carter, DO, Tibbett M, editors. Soil analysis in forensic taphonomy: Chemical and biological effects of buried human remains. Boca Raton, Florida (FL): CRC Press, Taylor & Francis Group, LLC; 2008. p. 29-45.

4. Cobaugh KL, Schaeffer SM, DeBruyn JM. Functional and Structural Succession of Soil Microbial Communities below Decomposing Human Cadavers. *PLoS One*. 2015;10(6):e0130201. doi:http://dx.doi.org/10.1371/journal.pone.0130201

5. Finley SJ, Pechal JL, Benbow ME, Robertson BK, Javan GT. Microbial Signatures of Cadaver Gravesoil During Decomposition. *Microbial Ecol*. 2016;71(3):524–529. doi:10.1007/s00248-015-0725-1

6. Keenan SW, Emmons AL, Taylor LS, Phillips G, Mason AR, Mundorff AZ, Bernard EC, Davoren J, DeBruyn JM. Spatial impacts of a multi-individual grave on microbial and microfaunal communities and soil biogeochemistry. *PLoS One*. 2018;13(12):e0208845. doi:10.1371/journal.pone.0208845

7. Metcalf JL. Estimating the postmortem interval using microbes: Knowledge gaps and a path to technology adoption. *Forensic Sci. Int. Genet*. 2019;38:211–218. doi:10.1016/j.fsigen.2018.11.004

8. Forbes, SL. Decomposition chemistry in a burial environment. In: Carter, DO, Tibbett M, editors. Soil analysis in forensic taphonomy: Chemical and biological effects of buried human remains. Boca Raton, Florida (FL): CRC Press, Taylor & Francis Group, LLC; 2008. p. 203-217.

9. Levy-Booth DJ, Campbell RG, Gulden RH, Hart MM, Powell JR, Klironomos JN, Pauls KP, Swanton CJ, Trevors JT, Dunfield KE. Cycling of extracellular DNA in the soil environment. *Soil Biol. Biochem*. 2007;39(12):2977–2991. doi:10.1016/j.soilbio.2007.06.020

10. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature*. 1993;362:709-715.

11. Morley MW, Link to external site this link will open in a new window, Goldberg P, Uliyanov VA, Kozlikin MB, Shunkov MV, Derevianko AP, Jacobs Z, Roberts RG, Link to external site this link will open in a new window. Hominin and animal activities in the microstratigraphic record from Denisova Cave (Altai Mountains, Russia). *Sci. Rep-UK*. 2019;9(1).

12. Massilani D, Morley MW, Mentzer SM, Aldeias V, Vernot B, Miller C, Stahlschmidt M, Kozlikin MB, Shunkov MV, Derevianko AP, et al. Microstratigraphic preservation of ancient faunal and hominin DNA in Pleistocene cave sediments. *Proc. Natl. Acad. Sci*. 2022;119(1):e2113666118. doi:10.1073/pnas.2113666118

13. Slon V, Hopfe C, Weiß CL, Mafessoni F, de la Rasilla M, Lalueza-Fox C, Rosas A, Soressi M, Knul MV, Miller R, et al. Neandertal and Denisovan DNA from Pleistocene sediments. *Science*. 2017;356(6338):605–608. doi:10.1126/science.aam9695

14. Vernot B, Zavala EI, Gómez-Olivencia A, Jacobs Z, Slon V, Mafessoni F, Romagné F, Pearson A, Petr M, Sala N, et al. Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. *Science*. 2021;372(6542):eabf1667. doi:10.1126/science.abf1667

15. Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga J-L, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci*. 2013;110(39):15758–15763. doi:10.1073/pnas.1314445110

16. Ladd, JN, Foster, RC, Nannipieri, P, Oades, JM. Soil structure and biological activity. *Soil Biochem*. 1996;9:23-78.

17. Tumer AR, Karacaoglu E, Namli A, Keten A, Farasat S, Akcan R, Sert O, Odabaşi AB. Effects of different types of soil on decomposition: An experimental study. *Legal Med*. 2013;15(3):149–156. doi:10.1016/j.legalmed.2012.11.003

18. Carter DO, Metcalf JL, Bibat A, Knight R. Seasonal variation of postmortem microbial communities. *Forensic Sci. Med. Pathol*. 2015;11(2):202–207. doi:10.1007/s12024-015-9667-7

19. Carter DO, Yellowlees D, Tibbett M. Moisture can be the dominant environmental parameter governing cadaver decomposition in soil. *Forensic Sci. Int.* 2010;200(1):60–66. doi:10.1016/j.forsciint.2010.03.031

20. Pathan SI, Arfaioli P, Ceccherini MT, Ascher-Jenull J, Nannipieri P, Pietramellara G, D'Acqui LP. Physical protection of extracellular and intracellular DNA in soil aggregates against simulated natural oxidative processes. *Appl. Soil Ecol*. 2021;165:104002. doi:10.1016/j.apsoil.2021.104002

21. Nielsen KM, Johnsen PJ, Bensasson D, Daffonchio D. Release and persistence of extracellular DNA in the environment. *Environ. Biosaf. Res*. 2007;6(1–2):37–53. doi:http://dx.doi.org/10.1051/ebr:2007031

22. Crecchio C, Stotzky G. Binding of DNA on humic acids: Effect on transformation of Bacillus subtilis and resistance to DNase. *Soil Biol. Biochem*. 1998;30(8):1061–1067. doi:10.1016/S0038-0717(97)00248-4

23. Douterelo I, Goulder R, Lillie M. Enzyme activities and compositional shifts in the community structure of bacterial groups in English wetland soils associated with preservation of organic remains in archaeological sites. *Int. Biodeter. Biodegr.* 2011;65(3):435–443. doi:10.1016/j.ibiod.2010.11.017

24. Emmons, A, Debruyn, J, Mundorff, A, Cobaugh, K, Cabana, G. The persistence of human DNA in soil following surface decomposition. *Sci. Justice.* 2017 57:341–348.

25. Ballard D, Winkler-Galicki J, Wesoły J. Massive parallel sequencing in forensics: advantages, issues, technicalities, and prospects. *Int. J. Legal Med*. 2020;134(4):1291–1303. doi:10.1007/s00414-020-02294-0

26. Verogen. ForenSeq™ DNA Signature Prep Kit Reference Guide; 2022. (Publication No. VD2018005 Rev. D).

27. Verogen. ForenSeq™ mtDNA Whole Genome Kit Reference Guide; 2020. (Publication No. VD2020006 Rev. B).

28. Holt CL, Stephens KM, Walichiewicz P, Fleming KD, Forouzmand E, Wu S-F. Human Mitochondrial Control Region and mtGenome: Design and Forensic Validation of NGS Multiplexes, Sequencing and Analytical Software. *Genes*. 2021;12(4):599. doi:10.3390/genes12040599

29. Daubert v. Merrell Dow Pharmaceuticals, Inc. (n.d.). Oyez. Retrieved December 18, 2022, from https://www.oyez.org/cases/1992/92-102

30. Erkol, Z, Hösükler, E. Postmortem animal attacks on human corpses. In: Dogan, KH editor. Post mortem examination and autopsy - Current issues from death to laboratory analysis. IntechOpen; 2018. doi:10.5772/intechopen.72929

31. Qiagen. DNeasy® PowerSoil® Pro Kit Quick-start Protocol; 2019. (Publication No. HB-2494-003).

32. Qiagen. EZ1&2® DNA Investigator® Kit Handbook; 2022. (Publication No. HB-0122-007).
33. Applied Biosystems. User Guide: Quantifiler HP and Trio DNA Quantification Kits; 2018. (Publication No. 4485354 Rev. H).

34. Kavlick MF. Development of a triplex mtDNA qPCR assay to assess quantification, degradation, inhibition, and amplification target copy numbers. *Mitochondrion*. 2019;46:41–50. doi:10.1016/j.mito.2018.09.007

35. Illumina. MiSeq FGx™ Forensic Genomics System Specification Sheet: Forensic Genomics; 2016. (Publication No. 1470-2014-004).

36. Verogen. MiSeq FGx™ Sequencing System Reference Guide; 2021. (Publication No. VD2018006 Rev. F).

37. Institute of Genetic Epidemiology. *Haplogrep 3 [Computer software]*. Innsbruck, Austria: University in Innsbruck; 2023. https://haplogrep.i-med.ac.at/

38. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2021. https://www.R-project.org/.

39. Microsoft Corporation. *Microsoft Excel [Computer software]*; 2019. https://office.microsoft.com/excel

40. Le S, Josse J, Husson F. **FactoMineR**: An *R* Package for Multivariate Analysis. *J. Stat Softw*. 2008;25(1), 1-18. doi:10.18637/jss.v025.i01

41. Kassambara A, Mundt F. *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.7.; 2020. https://CRAN.R-project.org/package=factoextra

42. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2016.

43. Chang W. *extrafont: Tools for Using Fonts.* R package version 0.19.; 2023. https://CRAN.R-project.org/package=extrafont

44. Maussion G, Thomas RA, Demirova I, Gu G, Cai E, Chen CX-Q, Abdian N, Strauss TJP, Kelaï S, Nauleau-Javaudin A, et al. Auto-qPCR; a python-based web app for automated and reproducible analysis of qPCR data. *Sci. Rep.* 2021;11(1):21293. doi:10.1038/s41598-021-99727-6

45. Bruijns B, Tiggelaar R, Gardeniers H. Massively parallel sequencing techniques for forensics: A review. *Electrophoresis*. 2018;39(21):2642–2654. doi:10.1002/elps.201800082

46. Verogen. Universal Analysis Software v2.0 Reference Guide; 2021. (Publication No. VD2019002 Rev. E).

47. Guo Y, Li C-I, Sheng Q, Winther JF, Cai Q, Boice JD, Shyr Y. Very low-level heteroplasmy mtDNA variations are inherited in humans. *J. Genet. Genomics*. 2013;40(12):607–615. doi:10.1016/j.jgg.2013.10.003

48. Sutherland A, Myburgh J, Steyn M, Becker PJ. The effect of body size on the rate of decomposition in a temperate region of South Africa. *Forensic Sci. Int.* 2013;231(1):257–62. doi:10.1016/j.forsciint.2013.05.035

49. Mann RW, Bass WM, Meadows L. Time since death and decomposition of the human body: Variables and observations in case and experimental field studies. *J. Forensic Sci.* 1990;31(1):103-111.

50. Holt A, Wootton SC, Mulero JJ, Brzoska PM, Langit E, Green RL. Developmental validation of the Quantifiler® HP and Trio Kits for human DNA quantification in forensic samples. *Forensic Sci. Int. Genet*. 2016;21:145–157. doi:10.1016/j.fsigen.2015.12.007

51. Elwick K, Zeng X, King J, Budowle B, Hughes-Stamm S. Comparative tolerance of two massively parallel sequencing systems to common PCR inhibitors. *Int. J. Legal Med.* 2018;132(4):983–995. doi:10.1007/s00414-017-1693-4

52. Sidstedt M, Steffen CR, Kiesler KM, Vallone PM, Rådström P, Hedman J. The impact of common PCR inhibitors on forensic MPS analysis. *Forensic Sci. Int. Genet*. 2019;40:182–191. doi:10.1016/j.fsigen.2019.03.001

53. Applied Biosystems. Precision ID SNP Panels with the HID Ion S5™/HID Ion GeneStudio™ S5 System application guide; 2019. (Publication No. MAN0017767 Rev. C.0).

54. Meiklejohn KA, Robertson JM. Evaluation of the Precision ID Identity Panel for the Ion Torrent™ PGM™ Sequencer. *Forensic Sci. Int. Genet.* 2017;31. doi:10.1016/j.fsigen.2017.08.009

55. Applied Biosystems. Precision ID mtDNA Panels with the HID Ion S5™/HID Ion GeneStudio™ S5 System application guide; 2021. (Publication No. MAN0017770 Rev. C.0).

56. Ta MTA, Nguyen NN, Tran DM, Nguyen TH, Vu TA, Le DT, Le PT, Hong DTT, Ha H, Chu HH. Massively parallel sequencing of human skeletal remains in Vietnam using the precision ID mtDNA control region panel on the Ion S5™ system. International Journal of Legal Medicine. 2021;135(6):2285–2294. doi:10.1007/s00414-021-02649-1

APPENDIX A: DNA QUANTITATION

Table A1: Quantifiler™ Trio qPCR replicate results on the quantity of small autosomal, large autosomal, and Y-chromosomal target nuDNA present in soil samples from six weeks of surface level decomposition in ng/µl from **Donor 1**. Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities.

	Small Autosomal (ng/µl)			Large Autosomal (ng/ul)			Y-Chromosomal (ng/µl)		
Sample	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
C-1	0.0035	0.005*	0.0033	0.0011	0*	0.0017	0	0	0*
TR-1	0.0005	0	0.0001	0	0	0	0	0	0
TL-1	0.0004	0.0004	0.0003	0	0.0001	0	0	0	0
LE-1	0.0031	0.0036	0.0036	0.0001	0.0004	0.0002	0	0	0
C-2	0.0003	0.0003	0.0011*	0	0	0.0001*	0	0	0*
TR-2	0	0	0	0	0	0	0	0	0
TL-2	0.0002	0	0.0002	0	0	0.0001	0	0	0
LE-2	0.0002	0	0	0	0	0	0	0	0
C-3	0	0	0	0	0	0	0	0	0
TR-3	0	0	0	0	0	0	0	0	0
TL-3	0.0006	0.0006	0.0004	0	0	0	0	0	0
LE-3	0.0026*	0.0019	0.0018	0.0003*	0	0	0*	0	0
C-4	0	0	0	0	0.0002	0	0	0	0
TR-4	0	0	0	0	0.0002	0	0	0	0
TL-4	0	0	0	0	0	0	0	0	0
LE-4	0	0	0	0	0	0	0	0	0
C-5	0	0	0	0	0	0	0	0	0
TR-5	0.0006	0	0	0	0	0	0	0	0
TL-5	0	0	0	0	0	0	0	0	0
LE-5	0	0	0	0	0	0	0	0	0
C-6	0.0006	0	0	0	0	0.0001	0	0	0
TR-6	0.0004	0.0004	0	0	0	0	0	0	0
TL-6	0	0	0	0.0002	0	0	0	0	0
LE-6	0	0	0	0.0003	0	0	0	0	0

Table A2: Quantifiler™ Trio qPCR replicate results on the quantity of small autosomal, large autosomal, and Y-chromosomal target nuDNA present in soil samples from six weeks of surface level decomposition in ng/µl from **Donor 2**. Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities.

	Small Autosomal (ng/µl)			Large Autosomal (ng/ul)			Y-Chromosomal (ng/µl)		
Sample	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
C-1	0.0114	0.0092	0.0121	0	0.0001	0.0001	0	0.0004	0.001
TR-1	0.0003	0	0.0005	0.0003*	0	0	0.0001*	0.0006	0.0005
TL-1	0.0004*	0.0008	0.001	0.0026	0.003	0.003	0.0115	0.0083	0.0109
LE-1	0.0146	0.0128	0.0151	0.0126	0.0104	0.0093	0.0177	0.0138	0.0113
C-2	0.0008	0.0012	0.0023	0	0.0001	0*	0.0024	0	0.0009
TR-2	0.001	0.001	0.0017*	0	0.0002	0	0	0	0.0012
TL-2	0	0.0007	0	0	0	0	0	0	0
LE-2	0.0004	0.0008	0	0	0.001	0	0	0	0
C-3	0.0011	0.0006	0	0	0	0	0	0	0
TR-3	0.0006	0.0006	0	0	0	0	0	0	0
TL-3	0	0.0005	0	0	0	0	0	0	0
LE-3	0.0005	0.0006	0.0005	0	0	0	0	0	0
C-4	0	0	0	0	0	0	0	0	0
TR-4	0.0004	0	0	0.0002	0	0.0008	0	0	0
TL-4	0	0	0	0	0	0	0	0	0
LE-4	0	0	0	0	0	0	0	0	0
C-5	0	0.0007	0	0	0	0	0	0	0
TR-5	0.0004	0	0	0	0	0	0	0	0
TL-5	0	0	0	0	0	0	0	0	0
LE-5	0	0	0	0	0	0	0	0	0
C-6	0	0.0004	0.0004	0.0001	0	0.0001	0	0	0
TR-6	0	0.0003	0	0	0	0	0	0	0
TL-6	0	0	0	0	0	0	0	0	0
LE-6	0	0.0003	0	0	0	0	0	0	0

Table A3: Quantifiler™ Trio qPCR replicate results on the quantity of small autosomal, large autosomal, and Y-chromosomal target nuDNA present in soil samples from six weeks of surface level decomposition in ng/µl from **Donor 3**. Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities.

	Small Autosomal (ng/µl)			Large Autosomal (ng/ul)			Y-Chromosomal (ng/µl)		
Sample	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
C-1	0.0029	0.0021	0.0027	0	0.0002	0.0002	0.0005	0	0.0009
TR-1	0	0	0	0	0	0.0002	0	0	0
TL-1	0.0056	0.0056	0.0061	0.0003	0.0004	0.0001	0.0014	0.0005	0
LE-1	0.0029	0.0021	0.0027	0.0001	0	0.0093	0	0	0.0108
C-2	0	0	0	0	0	0	0	0	0
TR-2	0	0	0	0	0	0	0	0	0
TL-2	0.0013	0.0013	0.0007*	0	0	0*	0	0	0*
LE-2	0.002	0.002	0.0002	0.0005	0.0004	0.0017	0.0002	0	0
C-3	0.0007	0.0004	0	0	0	0	0	0	0
TR-3	0	0	0	0	0.0001	0	0	0	0
TL-3	0	0	0	0	0	0	0	0	0
LE-3	0	0	0	0	0.0002	0	0	0	0
C-4	0	0	0	0	0	0	0	0	0
TR-4	0	0	0	0	0	0	0	0	0
TL-4	0	0	0	0	0	0	0	0	0
LE-4	0	0	0	0	0	0	0	0	0
C-5	0	0	0	0	0	0	0	0	0
TR-5	0	0	0	0	0	0	0	0	0
TL-5	0	0	0.0008	0	0	0	0	0	0
LE-5	0	0	0	0	0	0	0	0	0
C-6	0.0002	0	0	0	0	0	0	0	0
TR-6	0	0.0003	0	0	0	0	0	0	0
TL-6	0	0	0	0	0	0	0	0	0
LE-6	0.0003	0	0	0	0.0001	0	0	0	0

Table A4: Kavlick mtDNA multiplex qPCR replicate results on the quantity of mtDNA present in soil samples from eleven weeks of surface-level decomposition in copy number/µl from **Donor 1**. Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities.

	Cranium			Torso Right			Torso Left			Lower Extremity		
Wk	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
1	9911.36	5769.32*	9155.82	1227.80	1350.16	1175.38	1383.72*	924.975	824.440	657.857	625.790	931.449
2	508.108	712.580	606.892	468.961	486.593	516.642	3283.999	2826.67	3071.76	4659.10	4584.21	2022.13*
3	645.032	539.914	601.884	1176.89	1276.91	1303.93	16236.5	16236.5	15104.6	13011.3	11797.7	12083.9
4	8992.67	8378.37	8946.13	684.120	817.187	893.348	959.453	835.744	868.923	655.910	664.181	686.892
5	458.660	453.447	413.940	6086.85	6437.10	6649.93	617.114	547.732	550.409	997.280	935.806	1005.36
6	345.818	335.929	298.650	363.331	413.897	461.009	6569.50	6463.98	6221.05	10303.2	9575.16	10272.9
7	505.482	431.422	386.320	669.205	575.704	512.792	713.190	803.055	809.305	666.274	616.906	795.578
8	494.189	472.673	510.526	57.0714	46.1621	55.9747	72.4776	53.0923	69.1649	64.0645	52.0731	55.8195
9	285.71	304.440	239.268	422.174	341.266	334.674	344.527	399.349	327.327	281.010	234.899	318.200
10	214.958	210.384	188.940	287.832	323.064	286.315	326.171*	205.931	246.275	141.920	172.892	215.082*
11	0	0	0	0	0	0	218.091	251.534	879.692	0	0	0

Table A5: Kavlick mtDNA multiplex qPCR replicate results on the quantity of mtDNA present in soil samples from eleven weeks of surface-level decomposition in copy number/µl from **Donor 2.** Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities.

	Cranium			Torso Right			Torso Left			Lower Extremity		
Wk	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
1	7154	8744.3	11372*	14917	12860	11604	9122.7	9634.4	14141*	20895	23090	14586
2	7185.7	6554.2	7353.7	24015	22153	23201	26574	26469	28531	28139	30278	37192
3	170657	157982	199851	115332	125792	130711	17081	16406	17102.7	297241	290614	275426
4	790.20	926.60	783.74	49295	31230*	48917	9134.4	9026.8	8377.3	3839.4	3561.4	3644.2
5	698.38	798.37	779.98	2353.6	2201.1	2413.4	844.61	831.44	911.44	868.45	862.87	927.63
6	4122.2	4627.8	4770.1	4677.3	4474.6	4201.9	4742.7	4793.4	4675.9	4294.8	4633.6	4745.3
7	844.99	974.20	1067.4	1137.3	1436.5	1028.8	1169.8	1093.5	1326.4	1021.1	1334.8	1333.1
8	162.02	158.92	200.84	125.50	566.84 *	116.55	88.331	83.296	88.293	162.18	275.97 *	150.78
9	439.91	498.32	513.57	732.31	658.92	698.39	1036.5	959.81	1112.97	601.76*	981.81	1084.7
10	1625.5	1539.9	1675.96	1427.2	1351.7	1353.5	705.85	764.54	837.44	3862.9	4074.4	4208.8
11	0	0	0	0	0	0	1568.1	1337.4	1464.01	0	0	0

Table A6: Kavlick mtDNA multiplex qPCR replicate results on the quantity of mtDNA present in soil samples from eleven weeks of surface-level decomposition in copy number/µl from **Donor 3**. Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities.

	Cranium			Torso Right			Torso Left			Lower Extremity		
Wk	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
1	3559.3	3744.8	5609.7*	776.86	1237.6*	995.81	19389	18045	27509*	19519*	31787	40778
2	193.04	227.18	206.35	261.91	248.41	290.46	2730.97	2573.9	2607.3	16847	16292	15816
3	424.88	443.82	444.19	547.66	495.48	508.85	514.41	510.58	603.60	1714.6	1714.7	1722.9
4	722.97	798.75	654.71	255.02	225.78	241.86	353.32	334.02	304.08	1471.0	1527.3	1434.4
5	648.85	721.72	616.39	1193.8	1322.0	1291.2	967.40	909.51	895.12	904.78*	1846.2	1865.3
6	4614.8	4460.8	4520.4	4635.3	4606.6	4729.6	5286.6	4996.5	5106.1	873.76	867.26	817.16
7	1308.5	1280.6	1286.3	1147.9	1219.7	1203.8	2055.1	2277.5	1471.1*	1642.6	1618.3	1742.0
8	118.21	95.914	187.19*	296.97	244.04	257.11	3487.7	6006.4*	3144.5	307.87	303.74	320.71
9	2368.6	2425.4	2251.9	1830.5	1778.8	1847.3	2195.2*	1619.8	1455.4	136.39	227.49*	147.06
10	3004.4	3400.4	3379.3	1102.7	1099.7	1125.4	817.23	641.09	730.64	332.94	358.51	309.83
11	2329.2	1773.7	1846.9	0	0	0	0	0	0	0	0	0

Table A7: Kavlick mtDNA multiplex qPCR Ct results for the small (**105bp**) target present in soil samples from eleven weeks of surface-level decomposition from **Donor 1**. Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities.

	Cranium			Torso Right			Torso Left			Lower Extremity		
Wk	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
1	25.134	25.954*	25.254	28.298	28.154	28.364	28.117*	28.727	28.901	29.243	29.319	28.716*
2	32.461	31.965	32.201	32.579	32.525	32.437	29.722	29.942	29.820	29.209	29.232	30.434*
3	29.774	30.028	29.873	28.915	28.799	28.769	25.167	25.180	25.270	25.483	25.623	25.589
4	28.961	29.068	28.969	32.862	32.593	32.458	32.350	32.559	32.5	32.926	32.907	32.856
5	30.797	30.813	30.939	27.227	27.150	27.105	30.387	30.552	30.545	29.725	29.812	29.713
6	32.961	33.003	33.175	32.888	32.698	32.540	28.651	28.675	28.731	27.992	28.099	27.996
7	30.621	30.840	30.993	30.234	30.442	30.602	30.146	29.982	29.971	30.240	30.346	29.995
8	27.996	28.060	27.949	30.433	30.732	30.460	30.096	30.535	30.162	30.270	30.562	30.464
9	32.365	32.273	32.623	31.797	32.107	32.135	32.093	31.878	32.167	32.389	32.650	32.209
10	33.985	34.016	34.174	33.557	33.387	33.564	33.373*	34.047	33.785	34.593	34.304	33.984*
11	0	0	0	0	0	0	33.963	33.754	31.919*	0	0	0

Table A8: Kavlick mtDNA multiplex qPCR Ct results for the large (**316bp**) target present in soil samples from eleven weeks of surface-level decomposition from **Donor 1**. Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities. Dashes (-) represent undetermined Ct.

	Cranium			Torso Right			Torso Left			Lower Extremity		
Wk	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
1	30.843	31.626*	30.731	35.438	36.032	35.065	35.586*	36.130	37.581	36.573	37.472	37.534*
2	34.915	34.022	34.308	35.257	35.192	35.249	31.574	31.652	31.562	30.863	31.063	32.348*
3	-	-	-	33.471	33.467	33.103	29.448	29.532	29.522	29.968	30.257	30.134
4	31.154	30.536	30.365	34.931	34.053	33.770	33.948	34.027	34.375	35.011	35.202	34.714
5	36.095	36.705	37.517	29.907	29.788	29.925	39.341	39.341	37.600	33.665	33.387	33.579
6	-	-	-	36.993	36.330	36.089	32.115	32.143	32.247	31.021	31.176	31.094
7	39.268	-	38.584	35.859	35.683	35.873	36.127	36.501	-	38.136	39.064	37.9996
8	-	-	-	-	-	-	38.6999	39.350	-	-	-	-
9	37.406	38.476	38.545	38.626	37.462	38.497	38.276	38.024	-	-	-	-
10	-	-	-	39.236	38.699	38.229	-	-	-	-	-	-
11	0	0	0	0	0	0	-	-	-	0	0	0

Table A9: Kavlick mtDNA multiplex qPCR Ct results for the small (**105bp**) target present in soil samples from eleven weeks of surface-level decomposition from **Donor 2**. Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities.

	Cranium			Torso Right			Torso Left			Lower Extremity		
Wk	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
1	25.628	25.324	24.926*	24.515	24.740	24.896	25.260	25.177	24.596*	24.005	23.853	24.549*
2	28.573	28.708	28.539	26.802	26.92	26.852	26.653	26.659	26.549	26.569	26.461	26.160
3	21.807	21.917	21.581	22.366	22.242	22.188	25.094	25.152	25.092	21.014	21.046	21.123
4	32.644	32.403	32.656	26.385	27.076*	26.396	28.938	28.955	29.069	30.25	30.364	30.329
5	30.216	30.032	30.064	28.539	28.631	28.504	29.954	29.976	29.849	29.916	29.924	29.824
6	29.333	29.164	29.119	29.148	29.213	29.305	29.128	29.112	29.149	29.273	29.162	29.127
7	29.911	29.715	29.589	29.501	29.179	29.640	29.462	29.556	29.289	29.650	29.28	29.282
8	28.96	28.987	28.657	29.321	29.321*	29.425	29.816	29.899	29.817	28.959	28.208*	29.061
9	31.737	31.556	31.512	30.995	31.149	31.064	30.490	30.602	30.386	31.281*	30.569	30.424
10	31.019	31.098	30.974	31.209	31.289	31.287	32.242	32.125	31.991	29.750	29.672	29.624
11	0	0	0	0	0	0	31.071	31.305	31.172	0	0	0

Table A10: Kavlick mtDNA multiplex qPCR Ct results for the large (**316bp**) target present in soil samples from eleven weeks of surface-level decomposition from **Donor 2**. Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities. Dashes (-) represent undetermined Ct.

	Cranium			Torso Right			Torso Left			Lower Extremity		
Wk	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
1	32.439	31.243	31.083*	30.421	30.325	30.753	31.161	30.995	30.770*	29.999	29.953	30.564
2	30.273	30.222	30.114	28.697	28.783	28.626	27.918	27.965	27.859	27.968	27.928	27.618
3	25.754	25.804	25.530	26.333	26.103	26.051	29.154	29.299	29.288	25.124	25.127	25.190
4	34.598	34.484	35.451	28.039	28.625*	27.930	30.695	30.788	31.042	31.917	31.927	31.877
5	34.770	34.846	34.892	31.419	31.409	31.279	34.626	34.455	34.420	35.033	35.219	35.219
6	35.3997	35.157	35.573	35.894	35.330	35.415	34.598	34.326	34.387	34.869	34.981	34.888
7	35.0666	35.437	35.197	35.447	35.493	35.769	34.763	34.891	34.966	36.972	37.169	36.974
8	36.979	36.917	36.385	37.092	36.932*	38.006	38.791	38.353	37.345	36.644	36.583*	37.117
9	36.145	36.346	35.831	34.755	35.029	34.802	33.568	33.527	33.747	35.723*	35.809	36.585
10	34.489	34.546	34.140	34.373	34.273	34.395	35.282	35.804	35.229	33.360	33.184	33.179
11	0	0	0	0	0	0	37.420	37.197	38.476	0	0	0

Table A11: Kavlick mtDNA multiplex qPCR Ct results for the small (**105bp**) target present in soil samples from eleven weeks of surface-level decomposition from **Donor 3**. Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities.

	Cranium			Torso Right			Torso Left			Lower Extremity		
Wk	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
1	26.686	26.609	25.997*	28.991	28.286*	28.615	24.118	24.227	23.588*	24.108*	23.369	22.992
2	33.882	33.643	33.784	33.434	33.512	33.282	29.993	30.080	30.061	27.322	27.371	27.415
3	30.371	30.308	30.307	30.008	30.151	30.113	30.098	30.108	29.869	28.378	28.378	28.371
4	32.778	32.628	32.929	34.356	34.541	34.437	33.863	33.948	34.090	31.703	31.646	31.741
5	30.318	30.171	30.389	29.476	29.335	29.368	29.767	29.852	29.874	29.859*	28.874	28.860
6	29.168	29.218	29.198	29.161	29.170	29.132	28.969	29.051	29.020	31.604	31.615	31.702
7	29.307	29.337	29.331	29.488	29.405	29.423	28.684	28.542	29.146*	28.993	29.014	28.912
8	29.405	29.7	28.756*	28.105	28.382	28.308	24.627	23.860*	24.774	28.677	28.697	28.618
9	29.287	29.252	29.360	29.662	29.704	29.649	29.397*	29.840	29.996	33.442	32.697*	33.332
10	30.118	29.937	29.946	31.588	31.592	31.558	32.027	32.383	32.191	33.343	33.235	33.449
11	30.491	30.891	30.832	0	0	0	0	0	0	0	0	0

Table A12: Kavlick mtDNA multiplex qPCR Ct results for the large (**316bp**) target present in soil samples from eleven weeks of surface-level decomposition from **Donor 3**. Values with an asterisk (*) were removed based on Ct-SD when calculating average quantities. Dashes (-) represent undetermined Ct.

	Cranium			Torso Right			Torso Left			Lower Extremity		
Wk	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
1	32.721	32.805	32.092*	39.195	38.204*	39.234	30.334	30.290	29.612*	30.171*	29.214	29.678
2	37.894	-	-	38.653	39.303	39.608	32.401	32.653	32.632	29.347	29.449	29.393
3	39.878	-	-	38.378	38.796	-	37.793	38.029	39.442	33.553	33.587	33.715
4	34.231	34.785	34.367	38.516	38.812	38.972	38.533	38.447	36.883	33.639	33.822	33.877
5	-	-	-	33.997	33.917	33.997	35.372	35.125	34.993	35.325*	36.754	36.150
6	37.399	36.335	37.628	36.489	35.719	36.169	31.42	31.063	31.076	35.874	36.574	36.073
7	35.631	35.84	35.528	38.165	37.614	38.930	37.825	38.717	38.212*	35.310	35.341	35.889
8	-	-	-	-	-	-	-	-	-	39.779	38.261	38.824
9	32.268	32.218	32.325	32.957	33.245	33.081	34.244*	33.957	34.118	38.369	38.518*	38.257
10	32.721	32.530	32.587	34.489	34.812	34.530	35.378	35.354	35.499	-	39.934	38.944
11	35.748	35.115	35.434	0	0	0	0	0	0	0	0	0

APPENDIX B: STATISTICAL ANALYSIS

Table B1: Data set used to calculate the effects of environmental factors on quantity of nuDNA (ng/µl). Environmental factors include time in weeks, TBS, body moisture concentration, weekly total rainfall (cm), average weekly humidity and average weekly temperature (Celsius).

Donor	Week	TBS	Body Moisture	Rainfall	Average Humidity	Average Temperature	Quantity nuDNA
2	1	19	19.57	1.2446	76.33333	17.28009	0.007073
2	2	24	60	7.3406	80.14286	20.11905	0.000745
2	3	26	20.63	5.588	78.85714	18.82937	0.000417
2	4	27	26.83	1.2954	77.14286	19.3254	3.33E-05
2	5	28	19.47	0.1016	78.14286	21.47156	9.17E-05
2	6	28	22.1	0.0762	67.42857	21.36905	0.000117
1	1	23	45.43	1.2446	74.28571	17.37103	0.001709
1	2	27	60	7.3406	80.14286	20.11905	0.000109
1	3	31	33.77	5.588	78.85714	18.82937	0.000482
1	4	31	37.56	1.2954	77.14286	19.3254	0
1	5	30	40.57	0.1016	78.14286	21.47156	0.00005
1	6	31	22.47	0.0762	67.42857	21.36905	0.000117
3	1	19	30.6	1.0922	76.16667	19.0625	0.002983
3	2	22	30.5	1.2954	77.14286	19.3254	0.000618
3	3	24	19.33	0.1016	78.14286	21.47156	9.17E-05
3	4	24	17.6	0.0762	67.42857	21.36905	0
3	5	24	18.33	2.1082	76	22.1131	6.67E-05
3	6	26	22.4	1.7018	86.57143	22.24206	6.67E-05

Table B2: Data set used to calculate the effects of environmental factors on quantity of mtDNA (copy number/µl). Environmental factors include time in weeks, TBS, body moisture concentration, weekly total rainfall (cm), average weekly humidity and average weekly temperature (Celsius).

Donor	Week	TBS	Body Moisture	Rainfall	Average Humidity	Average Temperature	Quantity mtDNA
2	1	19	19.57	1.2446	76.33333	17.28009	13113.38
2	2	24	60	7.3406	80.14286	20.11905	22303.75
2	3	26	20.63	5.588	78.85714	18.82937	151183
2	4	27	26.83	1.2954	77.14286	19.3254	12572.36
2	5	28	19.47	0.1016	78.14286	21.47156	1207.61
2	6	28	22.1	0.0762	67.42857	21.36905	4563.297
2	7	28	25.6	2.1082	76	22.1131	1147.311
2	8	28	25.23	1.7018	86.57143	22.24206	133.6709
2	9	28	22.2	2.1844	85.28571	22.21561	792.4656
2	10	28	23.87	0.4826	82.42857	21.81548	1952.309
2	11	30	19.4	0.9652	82	23.13492	1456.52
1	1	23	45.43	1.2446	74.28571	17.37103	2872.62
1	2	27	60	7.3406	80.14286	20.11905	1975.046
1	3	31	33.77	5.588	78.85714	18.82937	7488.824
1	4	31	37.56	1.2954	77.14286	19.3254	2781.91
1	5	30	40.57	0.1016	78.14286	21.47156	2096.136
1	6	31	22.47	0.0762	67.42857	21.36905	4302.036
1	7	31	27.2	2.1082	76	22.1131	623.7694
1	8	31	30.6	1.7018	86.57143	22.24206	166.9407
1	9	32	60	2.1844	85.28571	22.21561	319.4035
1	10	32	0	0.4826	82.42857	21.81548	227.8508
1	11	32	0	0.9652	82	23.13492	234.8125
3	1	19	30.6	1.0922	76.16667	19.0625	14884.47
3	2	22	30.5	1.2954	77.14286	19.3254	4857.852
3	3	24	19.33	0.1016	78.14286	21.47156	803.8077
3	4	24	17.6	0.0762	67.42857	21.36905	693.6028
3	5	24	18.33	2.1082	76	22.1131	1116.144
3	6	26	22.4	1.7018	86.57143	22.24206	3792.903
3	7	28	16.93	2.1844	85.28571	22.21561	1525.67
3	8	28	21	0.4826	82.42857	21.81548	857.6823
3	9	28	20.03	0.9652	82	23.13492	1586.097
3	10	29	27.37	3.0226	87.57143	22.18254	1358.518
3	11	29	30.97	2.2606	88.57143	21.33929	1983.269