2020

DIET-BREADTH ANALYSIS IN THE AMERICAN SOUTHWEST: METABARCODING METHOD WITH COPROLITES

Paige Nicole Plattner
University of Montana, Missoula

DIET-BREADTH ANALYSIS IN THE AMERICAN SOUTHWEST:

METABARCODING METHOD WITH COPROLITES

By

PAIGE NICOLE PLATTNER

B.A. Anthropology, University of Montana, Missoula, MT, 2018

Thesis

presented in partial fulfillment of the requirements
for the degree of

Master of Arts
Anthropology

The University of Montana
Missoula, MT

May 2020

Approved by:

Scott Whittenburg,
Graduate School Dean

Prof. Meradeth Snow, Chair
Anthropology

Anna-Marie Prentiss, Co-Chair
Anthropology

David Xing, Co-Chair
UM Genomics Laboratory

Plattner, Paige, M.A., Spring 2020 *Anthropology*

Diet-Breadth Analysis in the American Southwest: Metabarcoding Method with Coprolites

Chairperson: Prof. Meradeth Snow

Abstract

Molecular analysis of coprolites is proposed as new means of studying bio-archaeological remains that aids in the ability to capture diet breadth from past populations. This research utilizes the targeted PCR method, known as metabarcoding, to study dietary elements preserved in the coprolites from three greater Southwest sites. The Illumina MiSeq technology was used to sequence the samples after multiple rounds of targeted PCR using two primers (16s and 12s) proven successful for taxonomic identification with animal DNA and one primer specific to plant DNA (Rias et al 2011).

The sequencing data was then run through two different bioinformatics pipelines that yielded varied results. The molecular analysis of coprolites did not yield any sequences of maize DNA; however, presence of maize in the control soil samples suggest an error in the amplification of processed and digested maize. The use of the 16s and 12s primers successfully added to the knowledge of our understanding of diet breadth for the three selected sites--and furthermore for the greater Southwest. The use of the plant specific primer was not as insightful as previously conducted microremains analysis; however, one of the two bioinformatics approaches yielded at least one sample with significant confidence of cacao DNA--potentially some of the earliest diet in the region (Crown and Hurst 2008).

Acknowledgements

I must first and foremost thank my academic advisor Prof. Meradeth Snow for her continued guidance, patience and unwavering support throughout my academic career. Without her, my enrollment in this Masters program and subsequent research would not have been possible.

I would additionally like to thank my committee members—Professor Anna-Marie Prentiss and Dr. David Xing—for their patience and encouragement throughout this process as well as their record-breaking email response times. Additionally, I must extend my gratitude to the University of Montana Genomics Core for the Core Voucher program, without which Illumina MiSeq sequencing would not have been possible.

Finally, I must thank my loving and supportive parents for encouraging a naive 11 years old who wanted to be an anthropologist.

TABLE OF CONTENTS

List of Figures

List of Tables

Introduction:

The research objective of this project is to better understand the diet breadth of individuals living in the early agricultural period of the American Southwest. Diet breadth studies are used to explore the energetic implication of food choice and apply an economic logic to past populations' dietary components (Gremillion 2004; Bettinger 1991; Kaplan and Hill 1992; Stephens and Krebs 1986). This research seeks to explore these past dietary components through a molecular analysis of paleofecal (coprolite) material. The DNA found within the coprolites are intended to provide snapshots of the individual's diet breadth by demonstrating what foods the individual was consuming. The coprolites used in this research were collected from the Boomerang Shelter, Marsh Pass, and Rio Zape archaeological sites. The Boomerang Shelter and Marsh Pass sites are associated with the Basketmaker II phase, a phase in the desert Southwest known for: coiled basketry, atlatls, slab-lined storage cists, bone tools, and cultivated maize (Matson 1991; Geib 2011; Roth 2016). The third site, Rio Zape, is located in the southern portion of what is considered the greater Southwest. While separated from the other sites by the modern US-Mexico border, scholars place this site with the American southwest based on similarities of their cultural complexes, including those listed above (Kirchhoff 1954).

Molecular analysis of coprolites is proposed as new means of studying diet breadth that aids in the ability to capture diet breadth from past populations with flawed or missing archaeological records. Several new genomic techniques have been developed to study human coprolites. A potential method to circumvent the preservation issue is PCR-based metabarcoding technique to reconstruct the past diets through analysis of genetic material preserved in selected coprolite material.

The metabarcoding technique is a targeted polymerase chain reaction (PCR) approach. Certain regions in the mitochondrial genome in animals as well as the chloroplast genome in plants are considered "conserved regions." These regions are subject to a lower mutation rate and therefore can be used to determine the taxonomy from which the DNA originated (Rias et al 2011). This technique has been used by biologists to noninvasively study the diets of targeted animal species (Rias et al 2011; Hugenholtz and Huber 2003; Ashelford 2006). In contrast, anthropologists have utilized the metabarcoding technique with ancient and historic human samples to test hypotheses regarding past human microbiomes (Jiménez et at 2013; Jervis-Bardy 2015; Wibowo et al. in prep). This research seeks to merge these two uses of the same genomic technique in reconstructing a past population's diet using human fecal remains.

The coprolite samples of this study were extracted in an ancient DNA laboratory with strict contamination avoidance procedures. Multiple PCR amplifications were performed for each sample using two sets of primers known to specifically target conserved regions of animal mitogenomes (12s and 16s) as well as a set of plant specific primers that targets the 6-loop region of chloroplast genomes (Riazz et al. 2011). Subsequent sequencing using the Illumina MiSeq technology created the molecular data needed for analysis. The samples were run through two different bioinformatics pipelines that filter and analyze the raw data in order to obtain diet specific data. The fist pipeline was an R-based primer-specific script. The second was a sample-specific cloud based pipeline in Galaxy. Both pipelines start with trimming primers and quality filtering. In the first script the Dada2 algorithm determines original sequence variants, then merges reads while removing chimeric sequences. Alternatively, the Galaxy pipeline masks and

assembles the processed files as a FASTA files output using the MEGAHIT tool (with minimum length of contigs to output set to 130bp). **Hypothesis 1)** The metabarcoding method, applied to coprolites, will expand the knowledge of diet breadth for the sites and respective region. **Hypothesis 2)** The primer-specific script, using R, will yield more diet specific species as it has processing steps that target only the DNA within the start and end of the primers. **Hypothesis 3)** The second pipeline, because it is sample specific, will facilitate individual comparisons between samples and sites in terms of what individuals had consumed in the meal(s) that are analyzed.

Test expectations: The first of the metabarcoding methods uses an R script designed to filter out bacterial and viral DNA, which were amplified non-specifically by the three primer sets, targeting on fauna mitogenome and the plant chloroplast genome. The cloud-based galaxy pipeline uses an all-in-one FASTQ pre-processor (Chen et al. 2018) that filters out low-quality reads and reads less than 120 base pairs in length but it does not have bacterial and viral filtering ability. If the first hypothesis is correct, the primer-specific R script method will result in more flora and fauna sequences, less bacterial and viral reads, and an easier assignment of dietary elements. The lack of a similar filter process in the galaxy pipeline will result in final files being heavily saturated with bacterial and viral reads; nonetheless, this method will still be valuable in its separate processing of samples. The samples can be compared individually against each other and against previously conducted microremains analysis of the samples, as well as grouped by sites and compared against each other and other dietary studies conducted in the region.

1.1 Site Backgrounds

The sites selected for analysis contained coprolites preserved in dry caves and alcoves. Despite the common perception, these caves and alcoves are neither dark nor deep. The rock shelters are the result of natural erosion in the cliff sides, only tens of meters wide at most. Many coprolites may have been moved, or exposed to natural weather or later human disruption. Any samples thought to have been affected by such events were excluded. Only the coprolites with exceptional preservation, believed to have remained dry and seemingly undisturbed since their deposit, were selected for the study.

The dry rock shelters from which the samples were collected are recorded archeological sites. The Boomerang Shelter archeological site is located in the southeast corner of the contemporary state of Utah in the Four Corners region of the United States. It is an early example of Basketmaker II in the Cedar Mesa region (Charles and Cole 2006, 181-182). The shelter was primarily occupied during the Basketmaker II times, with most remains dated to between 2500-1500 B.P., only a few pre-farmer artifacts dating as early as 8310 B.P. (ca. 7400 BCE). Two coprolite samples (UT30.3 and UT43.2) were carbon dated to 60AD and 10AD (see figure 1). Archeological evidence and previous studies suggest that at the time of occupation the populations were agriculturalists with corn/maize making up high proportions of their diets (Smiley et al. 1997; Roth 2016:147-155; Wibowo et al. in prep).

The Marsh alcove of Arizona consists of the sites of White Dog Cave and Kinboko Caves I and II (Kidder and Guernsey 1919; Guernsey and Kidder 1921) and the Woodchuck Cave excavated in 1934 (Lockett and Hargrave 1953). They are also Basketmaker II sites and belong to the Black Mesa site group and the Marsh Pass alcove specifically has produced radiocarbon dates on maize that range from 500 B.C. to 100

B.C. (Matson 1991; Smiley 1993, 1994, 1997). The archeological remains have been associated with the Basketmaker II times through the Pueblo III times, ca 2500-800 B.P. The coprolites collected from this site were directly dated around the 600s AD. Sample AZ107 had a Carbon 14 corrected mean date of 595AD. Sample AZ108 is dated at 635AD, and the third sample, AZ110A, is dated at 620AD. These sites are approximately 500 or more years after the dates associated with the Boomerang Shelter. The important difference between these sites is that the individuals of the Marsh Pass site would have had corn as a staple of their diets for those additional 500 years.

The third and final site in the study is the La Cueva de los Muertos Chiquitos site (AD 660-1430). The site is located in the Rio Zape municipality of Guanaceví, Durango, Mexico. Despite the contemporary geo-political border that now separates this site from the contemporary American Southwest, the culture and traditions associated with this site lead it to be considered part of the greater prehistoric American Southwest (Kirchhoff 1954). The samples in the cave primarily date to the Gabriel San Loma Cultural Phase, more specifically from the 700s to the early 900s AD. Previous dietary analysis conducted on these samples found maize, agave, squash and cultivated beans to be dietary staples (Brooks et al. 1962; Reinhard et al. 2006; Hammer et al. 2015; de Araujo et al. 2019) with additional common dietary elements including; juniper, prickly pear cladodes, squash seeds, tomatillos, fish and rodents. “The site well represents the association of agave and maize documented previously” (Hammer et al. 2015). All coprolites samples selected for analysis yielded carbon 14 dates that correlate to the archeologically determined date ranges.

Sample ID	Site	Mean Date Range
TS924	Boomerang	60AD
TS904A	Boomerang	10AD
TS899	Rio Zape	920AD
TS900	Rio Zape	850AD
TS901	Rio Zape	725AD
TS1119A	Marsh Pass	595AD
TS1120A	Marsh Pass	635AD
TS1122A	Marsh Pass	620AD

Table 1: Site Date Ranges
Note: not all samples sequenced were dated.

Figure 1: Site Map. Sites noted are Boomerang (B), Marsh Pass (MP), and Rio Zape (RZ).

Archeological artifacts and visual examination of coprolites have provided some insights into the diet (Wibowo et al. in prep.; Matson and Chisholm 1991). While fibrous flora remains may have held up through the digestion process, the full breadth, particularly the fauna, remains a question. The sub-discipline of molecular anthropology

provides a pathway to study and potentially quantify the breadth of flora and fauna consumed by the past inhabitants of the sampled sites. Using both the archeological record and molecular anthropology allows for the study of the relationship between genetic changes from the domestication process and the technological innovations that had been implemented in the time since (Hünemeier et al. 2012).

1.2 Arrival of Maize in the Southwest

Early agriculture in the Southwest is of particular interest for anthropology because of the varied speed in which maize begins its transition into the primary diet staple of the region. Across the American Southwest it is known that maize was introduced from their agriculturalist neighbors to the south; however, the quantity grown and the extent to which the population was dependent on the crop is unclear across time and space (Matson 2006:152). The introduction of maize in this region has been associated with rock shelter occupation and small populations with some extent of residential mobility (Smiley 2002; Roth 2016). Maize had arrived in the Southwest's lowlands of the now United States around 4000 years ago (Swarts et al. 2017; Pennisi 2017). Much of the early archeological samples suggested maize arrived in the Southwest via a highland route but more recently this has been debated. Recent sampling shows morphological similarities of Southwest maize to the extant Mexican maize, thus suggesting a more lowland and Pacific coast route. Additionally, the temporal variation in morphology of maize cobs found in the Southwest suggest genetic changes were responsible for Southwestern environmental specific adaptations (da Fonseca et al. 2015), such as earlier flowering maize that fared better in the more temperate Southwest (Swarts et al. 2017).

2.0 Theoretical Framework: Diet Breadth Modeling

> *There is also, very clearly, no single and straight developmental pathway across this middle ground, no single evolutionary track leading to agriculture.*
>
> —SMITH 2001B:301
>
> *Smith 2001:300*

Diet breadth modeling has been central to discussions regarding forager-farmer transitions (Kennett and Winterhalder 2006; Prentiss 2019). The model was started in the field of environmental ecology with the early diet-choice theory seeking to model abundance and distribution, technology innovation and energy-extraction efficiency (Bettinger 1991; Kaplan and Hill 1992; Winterhalder and Goland 1997). By applying an economic mindset to early food production, costs can be calculated as investments and food handling times. The modeling favors buffering risk and thus a low trophic diet focused on domesticated foods, such as maize (Winterhalder and Goland 1997).

2.1 Significance of Dietary Analysis

The practice of dietary analysis that focuses specifically on agricultural transition periods serves to provide more understanding of the economic logic behind these decisions. Wile the transitions to a grain centered diet occurred separately across the world (Stanish, 2017 pg.170), the same question from archeologists and evolutionary anthropologists arises: did this transition occur slowly with small changes over time or was there a rapid depletion of the old ways of procuring food and dominance of the new ways? This very question is also asked in regard to the Basketmaker II phase and thus the variety of species consumed during this transition can serve as insight into the rationale of these individuals regarding their diet. Previous work can provide a starting framework to bring focus specifically to diet change in time periods that are known for the many

complex and collectively occurring (Stanish, 2017). Maize is often considered the most important crop in the Americas and the start of the Mesoamerican village lifestyle (Matsuoka et al., 2002; Jaenicke-Despre´s and Smith, 2006; Blake, 2006; Piperno, 2007; Ranere et al., 2009) but its importance in changing the desert Southwest's lifestyle is not as stark (Roth, 2016; Smiley et al. 1996). The inconsistency in which maize and other crops (squash and beans) spread across the region is all the more reason for a focused study in the diet of individuals living in the selected sites.

Originally maize was not highly productive, but its predictability is thought to have been added to the diet as a supplement to the variability in gathering pinon (pine nuts) (Minnis 1985, 1992; Roth 2016). A finding that is supported by repeated archeological findings of pinon and maize in the same settings (Smiley 2002). The Black Mesa early Basketmaker II period, from which the two US based sites belong to, is considered representative of a long period of adaptive stability following the initial introduction of maize (Smiley 2002). The strategy of storing pinon and maize and the integration of agriculture with foraging allowed the Basketmaker II groups to continue to successfully occupy the region (Roth 2016). Such a practice would be consistent with changing diet breadth as hunter-gatherers adopt a new food source (Gremillion 1996; Barlow 2002). If the samples are coming from the earlier part of an agricultural transition it will be expected that the diet breadth will be wider and more distinct species consumed; furthermore, it will be expected that the earlier Boomerang samples would have a larger diet breadth as compared to the Marsh Pass and Rio Zape samples that would have come from a society that has developed more dependence on maize agriculture.

3.0 Previous Studies: Microremains

Previous work completed with the same samples by Dr. Karl Reinhard (University of Nebraska) included a microremains analysis with the coprolites complemented by the recorded archeological remains from the sites and similar shelters. The analysis combined microscopic and pollen evidence of foods within the coprolites to gain understanding of the last meals consumed as well as some insight into parasites. As expected for the archaeological time period, the microremains suggested a diet of maize and other available remains from the region. While beans were not present in coprolites associated with the Boomerang shelter, they appeared to have been recently included in the various Marsh Pass sites of Northeastern Arizona and appeared to have been present longer and with more varieties in the Rio Zape cave sites (Wibowo et al. in prep; Pucu et al 2020). Other flora commonly recorded included: various grasses, pinyon pine nuts, cactus and agave, fruits and flowers. Fauna included: various rabbits, deer, small mammals, such as rodents, insects, larva, reptiles, and birds. Until the addition of beans most populations would have been dependent on meat consumption for protein (Wibowo et al 2019; Pucu et al 2020).

3.1Boomerang Shelter Dietary Overview

In general, the Boomerang Shelter coprolites are considered atypical for the region and time period. Four samples were selected for molecular analysis and the following is the microremains analysis of the same samples. The first of the Boomerang shelter samples is TS904B. This sample primarily consisted of rough and rarely eaten foods such as: woody stems of a shrub and male flowers with significant pollen. In addition to these odd finds there were also some more typical dietary remains such as:

nuts, prickly pear pads, Chenopodium seeds, Sarcobatus pollen grains, and maize smut fungus otherwise known as huitlacoche in Mexico. The second sample, TS903B, contained some typical items such as grasshopper or cricket exoskeleton and fruit exocarp. More interestingly, the sample had a significant amount of ricegrass fibers and evidence of harvested maize pollen. Samples TS924 was significantly composed of cactus remains associated with prickly pear cladodes but there were also some traces of maize, seeds and bone. The last of the Boomerang samples, TS929A, predominantly consisted of maize, both milled and kernels. The alteration of the maize starch could be the result of fermentation. In addition to the maize the sample also included pigweed and goosefoot pollen, as well as an undetermined bean source (Fabaceae pollen).

3.2 Marsh Pass Dietary Overview

Four of eight coprolites from Northeastern Arizona sites were selected in the microremains analysis--the same four samples selected for the molecular analysis. Sample, TS1125A, only contained maize, specifically maize milled to 1-2mm. TS1122A, contained crushed bone, rabbit hair and more milled maize. TS1119A, consists of maize starch, non-cultivated grass caryopses, chaff and stems. Lastly, TS1120A is unique in that this is the only sample without maize starch. The sample instead consists of insects, more non-cultivated grass caryopses, chaff and stems. The only Marsh Pass sample without microremains information available is sample TS1126A.

3.3 Rio Zape Dietary Overview

The smallest group of samples come from the La Cueva de los Muertos Chiquitos. All three coprolites selected for molecular dietary analysis have recorded microremains

analysis. The first sample, TS899, consisted of mostly agave fiber and maize pollen but there were also traces of milled maize, fractured nuts and goosefoot seeds. The microscopic analysis of the maize pollen within the sample indicated that the pollinating tops (anthers and tassels) of the maize stock were also being processed as food. Sample TS900 varied in its composition; while there was a maize cob end, traces of ground maize, milled maize, and Huitlacoche (maize mushroom), the sample primarily consisted of succulent leaves and a fine spongy fiber with some traces of crushed nuts. Significantly, with 105,832,910 *Ustilago maydis* spores per gram of sample, TS900 sample is the most ancient evidence of Huitlacoche (maize mushroom) in Mexico and Greater Southwest region (Wibowo et al. in prep; Pucu et al 2020). The final sample, TS901, was also consistent with the diet anticipated for this region and time. The sample consisted of milled goosefoot seed, dropseed, maize, insects and pollen. The first pollen, *Amaranthaceae* pollen, most likely came from pigweed and its abundance in the sample suggests the source was ingested greens including buds. The second pollen, a squash pollen, suggests the individual ingested squash blossoms. While there were only three Rio Zape samples selected, their microremains suggested a broad diet consistent with indigenous foods for the region and time period (Wibowo et al. in prep; Pucu et al 2020). Their complexity complements the Boomerang and Marsh Pass samples and ultimately are an important addition to the analysis.

Analysis of the diet of other nearby sites in the region (Bostwick and Adams 2016; Swarts et al 2017), such as the inhabitants of the nearby site of Dyck cave in the Verde valley, have demonstrated a diverse diet of flora. The known plant dietary elements include: cultivated corn, squash, gourds, beans, and possibly cotton seeds, plus

wild resources such as acacia beans, agave, yucca, walnuts, piñon nuts, acorns, mesquite pods, hackberry seeds, juniper seeds, wild grapes, prickly pear, and cholla (Bostwick and Adams 2016:chp. 5). Additionally, a series of caves in Montezuma Castle National Monument, downstream from the Dyck cave site, included a plant assemblage similar to the neighboring Dyck cave. The soft tissues found included banana yucca fruit, cholla buds, and prickly pear cactus pads. Unlike their neighbors, the Montezuma Castle samples included three additional varieties of beans, an additional variety of both squash and yucca, Ephedra, jimson weed, and wild grape. The coprolites of the Boomerang, Marsh Pass and Rio Zape sites were anticipated to depict a similar breadth of flora as well as a previously unseen breadth of fauna.

4.0 Methods

All samples were received by the University of Montana (UM) Molecular Anthropology Laboratory in conical tubes, and after the outside had been wiped down with a bleach solution, a small portion was scraped from the center of the sample into a UV irradiated (for a minimum of 15 minutes) 15mL sterile tube. The samples were extracted in the UM laboratory--a controlled access facility where standard procedures require researchers wear Tyvek clean suits, foot coverings, hair nets, face masks, arm coverings, and gloves to enter. Everyday before and after any procedure all work surfaces are bleached using a 50% household bleach:water solution. At the end of each day an overhead UV light runs for an hour. The ancient DNA laboratory is positively pressurized and movement between the post-PCR laboratory and ancient DNA laboratory is not allowed.

4.1 DNA Extraction

Once all of the samples were received by the Snow laboratory, they were prepped for extraction. Approximately a gram of each sample was submerged in 5 mL of EDTA (0.5M, pH 8) and incubated for 2 days (48 hours) at room temperature in UV'ed 15mL Falcon tubes. After the initial incubation period, 20 µL of 1 mg/mL Proteinase K was added to samples and followed by another incubation at 52C with slow rotation (4rpm) for four hours to lyse the mtDNA within the coprolites' cells. The extraction protocol was based on the Dabney et al. (2011) protocol standard in the Snow ancient DNA laboratory. Once the cells within the coprolites are lysed, the samples are spun in a centrifuge at 1500 rpm, 1.5 mL of the EDTA supernatant is then transferred into a sterile, UV treated

15 mL polypropylene tube and inversion mixed with 13 mL of PB Buffer (Qiagen), which contains guanidinium thiocyanate. The resulting solution is then spun through Qiagen MinElute filters utilizing 50 mL polypropylene tubes and nested conical reservoirs (Zymo) with attached filters. These filters are then washed twice with PE Buffer (Qiagen), which is primarily ethanol, before the final elution step using two rounds of 50 µL DNAse free H2O. The samples elute into sterile, low-bind 2 mL tubes, where they are stored in a sterile freezer kept at -20C. Blank negatives (controls where water was used in place of sample) were used and checked throughout the extraction process to aid in determining if contamination had been introduced through the extraction process. There were no instances of contamination in subsequent DNA quantifications or analyses detected.

4.2 Metabarcoding PCR

After the extraction process, each sample was PCR amplified at least twice for each of the three primer sets. This research focuses on the economical metabarcoding (PCR-based NGS sequencing) sequencing method and uses known and validated PCR primers: 12V5, 16Svert, and Plant1(trnL-P6-loop) from Riaz et al. (2011). The first primer targets the 12S region of mitochondrial DNA, this is a conserved region of the mitogenome that leads itself well to distinguishing mammals, birds, amphibians, fishes, and some invertebrates (Kocher 1989). The second PCR primer is also targeting a conserved region of the mitogenome, known as the 16S region. The 16S region has been used by biologists in species identification since the mid-1990s (Muyzer and Uitterlinden 1993), and utilized ever since for its broad success in identifying vertebrates and utilizing known mutational differences to differentiate between them (Palumbi 1996). The third

PRC primer is chloroplast specific and amplifies the P6 loop. This loop is targeted because it has successfully amplified highly degraded DNA from processed foods and has been specifically recommended for use in "diet analyses based on feces and in ancient DNA studies," due to their success with degraded DNA(Taberlet et al. 2007). The PCR product was then indexed using the Meyer Kircher protocol (2010). The PCR-products were pooled first based on sets of primers, then indexed by sample, and pooled again based on DNA concentration, which was established using the Qubit technology. They were then sent to the University of Montana Genomics core to be sequenced using the Illumina MiSeq technology. The MiSeq technology utilizes a double-sided, single-lane flow cell, as the bases attach they emit a fluorescent dye which is recorded as an image repeatedly for each base creating a synthesis of DNA strands (Quail et al. 2012; Jervis-Bardy 2015). Data analysis was performed using two different bioinformatics techniques and compared for continuity.

4.3 R-script Pipeline

The first bioinformatics approach is based on version 1.8 of the dada2 pipeline (Callahan et al. 2016; https://benjjneb.github.io/dada2/tutorial_1_8.html) The pipeline uses paired-end fastQ files derived from Illumina sequencing as input. The fastQ files must be split/"demultiplexed" by samples with removal of the barcodes. The DNA strands of different known species were isolated from background bacterial sequences and the pipeline then produces the tables as seen in Tables 1-3 for every PCR-specific run. The sequences in the tables were compared in the NCBI (National Center for Biotechnology Information) database using the nucleotide BLAST feature. All confident

BLAST results, with a score of 97% confidence of higher, of known consumable flora and fauna species in the region were recorded as dietary elements.

4.4 Cloud-based Pipeline (Galaxy)

The second bioinformatics technique incorporated the online cloud based Galaxy program that outputs FASTA files, SAM files, and data for visualization using visual aids known as Circos plots. The all-in-one FASTQ pre-processor (Chen et al. 2018) was used to filter low quality and short read, cut low quality reads, trim and front and tail reads, cut adapters, correct mismatched base pairs and removes any polyX tailing. The preprocessed output of FASTQ files are then masked (Blankenberg et al. 2010) and assembled as a FASTA file output using MEGAHIT (Li et al. 2015). The file feature of the pipeline creates the visual aid of a Circos plot (Krzywinski et at. 2009) from the FASTA file. Each FASTA file is sample specific, individual sequences were compared in the NCBI (National Center for Biotechnology Information) database using the nucleotide BLAST feature and coordinated to Circos plot section, as seen in Figures 2-12 The Circos plots are a software package (Krzywinski et al. 2009) used for visualizing data and information in a circular layout, with the data in this case being the sequence quantity in each sample.

4.5 Metabarcoding's Limitations and Alternatives

The metabarcoding method uses identified barcode markers known to correspond to well-studied regions of mitogenomes and chloroplast genomes, thus facilitating comparisons within the public NCIB database. While the targeted-PCR method does not produce the same number of reads as more recently developed sequencing methods, such

as deep shotgun sequencing, it allows a means of economically affordably looking for exactly the DNA that is the focus of your research and no more. The use of these three primers builds the groundwork for future inquiries regarding the agricultural integration process in past populations. By targeting faunal species with the 12s and 16s primers, the breadth of animals consumed that may have otherwise been unknown, is observable. Additionally, the comparison of microremains' observations with the plant DNA found in the molecular analysis reveals plants otherwise unseen as well as amplification errors within the coprolites themselves.

The targeted-PCR metabarcoding method is known to produce more sequence data with the issues of missing species and data due to biased and unequal amplification of species'16S rRNA genes. Most anomalies in 16s rRNA are chimeras—"artificial sequences generated from two or more phylogenetically different DNA templates during PCR amplification" (Ashelford et al. 2005:1). The result of these anomalies are chimeric sequences, "representing phylogenetically novel non-existent organisms" (Hugenholtz and Huber 2003:1), routinely being overlooked despite knowledge of PCR-generated artefacts amongst researchers ("jumping-PCR"). The continual addition of 16s rRNA sequences into public repositories has serious implication for correct taxonomic identifications, in that if the databases themselves are incorrect the comparison to the database will be flawed.

The alternative shotgun metagenomic sequencing method sequences without this known bias. The shotgun sequencing method involves sequencing small DNA fragments that have overlapping sequences. These fragments are then layered over each other to reconstruct the genome (Sharpton 2014). Unfortunately, the certainty with which these

layers overlap–otherwise referred to as how deep they are sequenced--is sometimes not enough to detect the 16S rRNA genes of rare bacterial communities (Shah et al. 2011). However, if deep enough, shotgun metagenomic sequencing has the potential to sequence all of the genetic information in a sample, as it is not limited to merely the targeted region and can allow for identifying species by sequencing other identifiable DNA fragments found in the rest of the genome. That reconstructed genome can be used for additional analyses beyond taphonomy assignment (Laudadio et al. 2019; Wood and Salzberg 2014).

Deeper sequencing of the samples, such as shotgun sequencing, would provide a broader understanding of the diet of these individuals and the population's approach to agriculture. Metagenomic shotgun sequencing would provide more than just the taphonomy information. With strong enough reads, a metagenome could provide more information on metabolic function profiling, antibiotic resistance, and gene profiling. Furthermore, any human mitogenome picked up by the shotgun sequencing could be checked for mitochondrial haplotypes and aid in anthropological understanding of migration and marriage patterns in the region.

Despite the known problems with metabarcoding, this targeted approach leads to only sequencing regions known to distinguish exactly the species the analysis is focused on--in this case flora and fauna. For this research the results of metabarcoding were limited to plants and vertebrate animals consumed, using the PCR-primers that specifically targeted these species. While many non-dietary species, such as bacteria, also have the 12s and 16s regions, the large public depositories--despite their bacterial error rates (Ashelford et al. 2005)--are still efficient in estimating the likelihood of taxonomic

assignment (Quail 2012) as it is at least possible to estimate the genus/phylum a match may come from and exclude it should it be bacterial in origin. The unfortunate issue is that it is impossible to avoid amplifying these sequences with these PCR methods, and the resulting data are heavily biased toward the non-dietary elements that must be carefully sorted through. While it was hypothesized that the metabarcoding technique would be subject to the problematic loss of data, it was ultimately an appropriate means for studying the diet breadth of the individuals who deposited the samples. The two data analysis techniques did yield different results, they are expanded on in the following text.

5.0 Results

The R pipeline allowed for all samples to be run collectively but required a separate run for each of the three PCR types. The pipeline targeted all sequences that started and ended with the known primer index sequences. This prevented non-target reads from being processed and recorded in the datasheets. The pipeline yielded a matrix/table for every primer-specific run. The tables list a count of how many strands of each sequence was in each of the 15 samples (see tables 2-4). The sequences were then compared in the NCBI Genbank database using the BLAST feature to assign species.

Table 2
Samples and Affiliated Sites

Marsh Pass Samples					Rio Zape Samples			Boomarang (coprolite) Samples				Boomerang- soil Samples		
1119A	1120A	1122A	1125A	1126A	899	900	901	903B	904A	929A	TS924	1026	1043	3567

5.1 Results of R-Script Pipeline

Table 3
R-script:12S Primer

BLAST Results/Specie	Sample Numbers														
	1026	1043	1119A	1120A	1122A	1125A	1126A	3567	899	900	901	903B	904A	929A	TS924
canis lupis	5	14	0	0	23403	2292	596	10	7	5230	4	0	0	1	23
human	2764	264	3333	705	8	17	4	305	5280	4060	298	3327	915	4	510
Ovis aries (sheep)	0	311	0	0	0	0	0	0	0	1	0	0	0	5246	0
human	64	6	0	0	6	0	0	0	0	557	0	0	0	0	0
human (or chimp)	0	0	0	0	0	0	0	0	0	275	0	0	0	0	0
human (or chimp)	0	0	0	0	0	0	0	0	0	251	0	0	0	0	0
human	23	0	0	0	0	0	0	0	0	209	0	0	0	0	4
human	0	0	0	0	0	0	0	0	0	218	0	0	0	0	0
misc. primates	33	0	0	0	0	0	0	0	0	137	0	0	0	0	0
human	0	0	0	0	0	0	0	0	0	0	0	130	0	0	0
human	35	0	0	0	0	0	0	1	0	83	0	0	0	0	0
Sylvilagus floridanus/nuttalliiae(cottontail)	0	0	0	0	0	0	0	0	0	111	0	0	0	0	0
canis lupis	0	0	0	0	96	0	5	0	0	0	0	0	0	0	0

Table 4
R-script:16S Primer

BLAST results/ Species	Sample Numbers													
	1026	1043	1120A	1122A	1125A	1126A	3567	899	900	901	903B	904A	929A	TS924
human	277	38	0	0	5	2	24	14405	160	15	0	1185	0	961
canis lupis: wolf	0	0	0	0	2641	734	0	2	0	0	0	0	0	1
Sm. Mammal: likely squirrel or marmot	0	0	0	0	0	0	0	0	5	0	0	0	0	22
Eimeria mitis(parasite),Cyprinus carpio(carp),or Tyto alba(barn owl)	0	0	0	0	0	0	0	0	17	0	0	0	0	0
Cyprinus carpio(carp)	0	0	9	0	0	0	0	0	0	0	0	0	0	0
Dipodomys (kangaroo rat)	0	0	0	0	0	0	0	0	0	0	0	0	0	6
Cyprinus carpio(carp), Eimeria mitis(parasite) or Tyto alba(barn owl)	4	0	0	0	0	0	0	0	0	0	0	0	0	0
Tyto alba(barn owl), Cyprinus carpio(carp), or Eimeria mitis(parasite)	0	0	0	0	0	0	0	0	0	0	4	0	0	0
Cyprinus carpio(carp), Eimeria mitis(parasite) or Tyto alba(barn owl)	0	0	0	2	0	0	0	0	0	0	0	0	0	0

Table 5

R-script: Plant (trnL-P6-loop) Primer

Blast Results/ Species	Sample Numbers														
	1026	1043	1119A	1120A	1122A	1125A	1126A	3567	899	900	901	903B	904A	929A	TS924
Lasthenia californica	0	0	0	0	0	0	0	0	0	216	0	0	0	0	0
Lasthenia californica or Dioon seed plant	0	0	0	0	0	0	0	0	0	0	0	185	0	0	0
No significant match	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0
Dioon seed plant	0	0	0	0	0	0	0	0	0	0	0	0	0	46	0
Lasthenia californica, Dioon seed plant or Theobroma cacao	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0
Dioon seed plant	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0
Lasthenia californica	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0
Theobroma cacao	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0

5.2 Results of Cloud-based Pipeline

The Galaxy pipeline treated the samples as the whole and were not PCR specific, nor did it have an adapter specific filtering step. FASTA files and Circos plot for each of the samples were created. The FASTA files were added into NCBI, spliced into sequence strands, and were then assigned a most likely species using the BLAST feature. Due to the lack of a filtration step, many of the strands were difficult to confidently assign a species that would have been a dietary element.

Figure 2: Sample 900
Domesticated barley: Hordeum vulgare OR Mites: Culicoides sonorensis
Mostly viral OR Mites: Culicoides sonorensis
Domesticated barley: Hordeum vulgare OR Mites: Culicoides sonorensis
Domesticated barley or viral

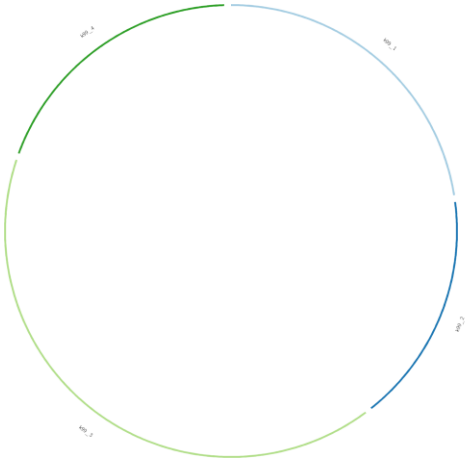
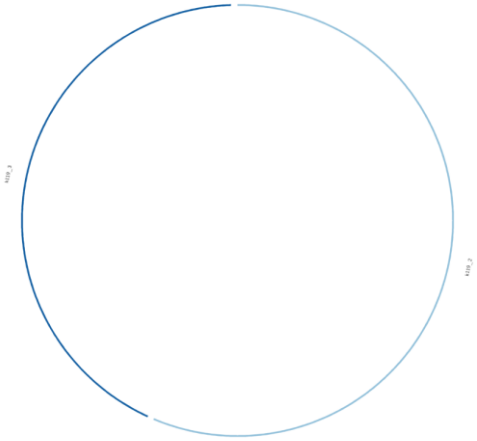


Figure 3: Sample 901
Domesticated (Dom.) Barely or viral
Synthetic OR Dom. Barely OR Mites: Culicoides sonorensis



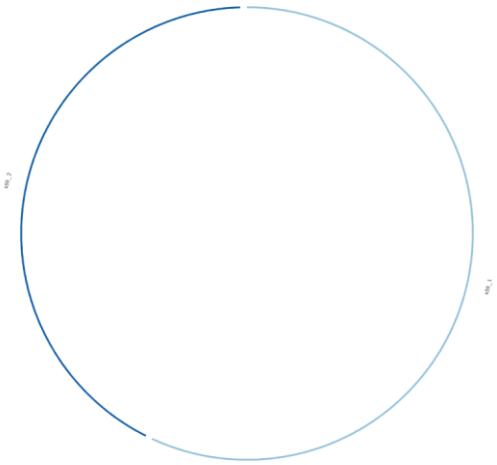


Figure 4: Sample 903B
Synthetic OR Dom. Barely OR Mites: Culicoides sonorensis
Sea anemone: Nematostella vectensis OR carp: Cyprinus carpio OR Barn owl-Tyto alba OR Flower: Lasthenia californica

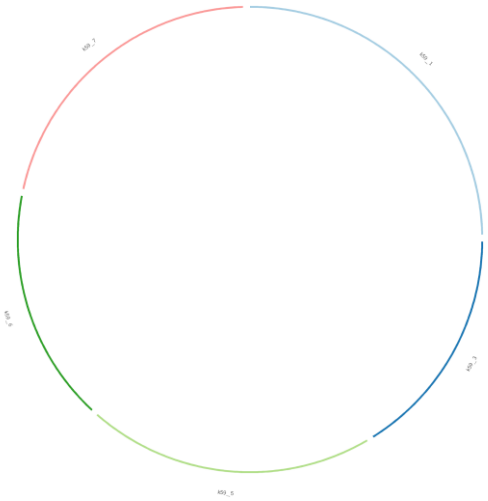


Figure 5: Sample 1125A
Mites: Culicoides sonorensis,
Dom. Barely or Mites: Culicoides sonorensis
Dom. Barely or Mites: Culicoides sonorensis
Dom. Barely or Mites: Culicoides sonorensis
Dom. Barely or Mites: Culicoides sonorensis

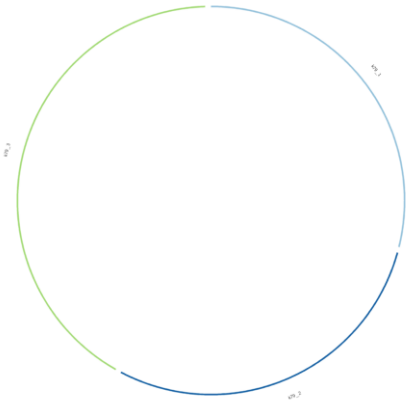


Figure 6: Sample 1122A
Carp: Cyprinus carpio OR Sea anemone-Nematostella vectensis
Mites: Culicoides sonorensis (or viral)
Mites: Culicoides sonorensis (or viral)

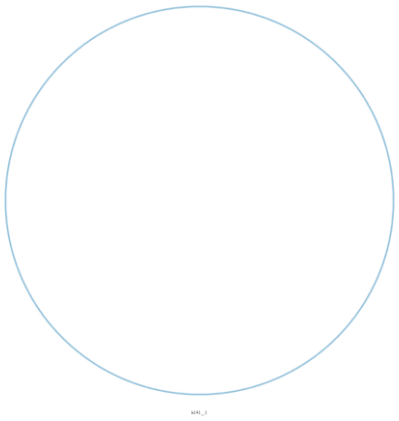


Figure 7: Sample 1120A
Carp: Cyprinus carpio or Sea anemone: Nematostella vectensis or sea snail: Conus episcopatus

Figure 8: Sample 1126A

Mites: Culicoides sonorensis (or viral)
Mites: Culicoides sonorensis (or viral)
Dom. Barely or Mites: Culicoides sonorensis
No Significant match

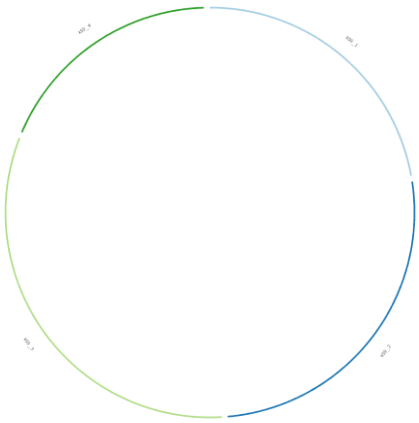
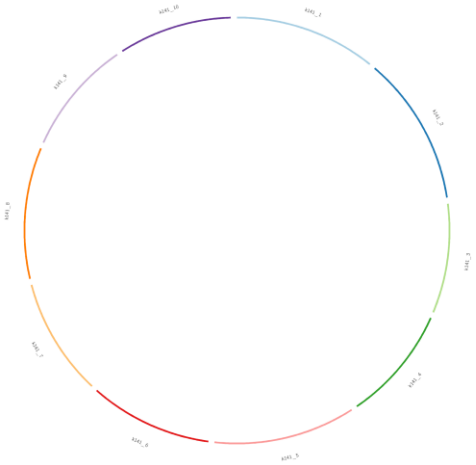


Figure 9: Sample 899

Dom. Barely or Mites: Culicoides sonorensis,
Dom. Barely or Mites: Culicoides sonorensis,
Dom. Barely or Mites: Culicoides sonorensis,
Barn owl-Tyto alba OR carp OR dom. Barley,
Dom. Barely or Mites: Culicoides sonorensis,
Dom. Barely or Mites: Culicoides sonorensis,
Dom. Barely or Mites: Culicoides sonorensis,
Dom. Barely or Mites: Culicoides sonorensis(mostly viral)
Barn owl-Tyto alba OR carp
Barn owl-Tyto alba OR carp OR Flower-Lasthenia californica



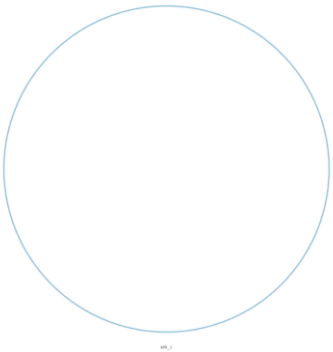


Figure 10: Sample 1026.1.4 soil
Human (contamination)

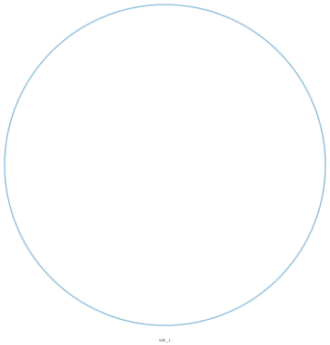


Figure 11: Sample 1043.4.1 soil
Corn

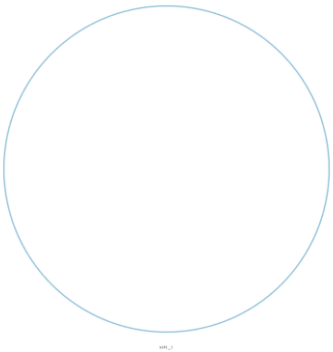


Figure 12: Sample 3567.1.1- soil
Corn

5.3 Results of Molecular Analysis with Microremains

Table 6: Molecular and Microremains results

Samples	**R-script pipeline**	**Galaxy pipeline**	**Microremains**
TS903A (Boomerang)	Human, mites (or owl or fish), Flower (*Lasthenia californica*) or Dioon seed	Grain, mites (*Culicoides sonorensis*), or sea anemone (*Nematostella vectensis*, or fish (*Cyprinus carpio*), or barn owl (*Tyto alba*), or Flower (*Lasthenia californica*)	Grasshopper or cricket exoskeleton, fruit, ricegrass fibers and maize pollen.
TS904A (Boomerang)	Human and cacao	Poor DNA	Woody shrubs, nuts, prickly pear pads, Chenopodium seeds, Sarcobatus pollen grains, and huitlacoche
TS924 (Boomerang)	Canine, human, sheep, small mammal and kangaroo rat	Poor DNA	Cactus remains (prickly pear), maize, seeds and bone
TS929A (Boomerang)	Canine, human, dioon seed	Poor DNA	Maize (both milled and kernels),

			pigweed and goosefoot pollen, and undetermined bean
TS1119A (Marsh Pass)	Human	Poor DNA	Maize, non-cultivated grass caryopses, chaff and stems
TS1120A (Marsh Pass)	Human, fish (*Cyprinus carpio*), dioon seed	Fish (*Cyprinus carpio*) or Sea anemone (*Nematostella vectensis*) or sea snail (*Conus episcopatus*)	Insects, non-cultivated grass caryopses, chaff and stems
TS1122A (Marsh Pass)	Canine, human and mites (or owl or carp)	Fish (*Cyprinus carpio*), or Sea anemone (*Nematostella vectensis*), Mites (*Culicoides sonorensis*)	Bone, rabbit hair and maize
TS1125A (Marsh Pass)	Canine and human	Grain or mites (*Culicoides sonorensis*)	Maize
TS1126A (Marsh Pass)	Canine and human	Grain or mites (Culicoides sonorensis)	NA
TS899 (Rio Zape)	Canine, human, flower (*Lasthenia californica*), and flower or cacao	Grain, mites (*Culicoides sonorensis*), or fish (*Cyprinus carpio*), or barn owl (*Tyto alba*), or Flower (*Lasthenia californica*)	Agave fiber, maize, nuts and goosefoot seeds
TS900 (Rio Zape)	Canine, human, sheep (*Ovis aries)*, rabbit (*Sylvilagus floridanus/nuttalliiae*), flower (*Lasthenia californica*) and mites (or fish or owl)	Grain or mites (*Culicoides sonorensis*)	Maize (cob end, ground and milled), Huitlacoche, succulent leaves, fine spongy fiber and nuts.
TS901 (Rio Zape)	Canine and human	Grain or mites (*Culicoides sonorensis*)	Goosefoot seed, dropseed, maize, insects, pigweed

			pollen and squash blossom pollen
TS1026 (Boomerang-soil)	Canine and human	Human DNA	NA
TS1043 (Boomerang-soil)	Canine, human and sheep (*Ovis aries)*	Maize	NA
TS3567 (Boomerang-soil)	Canine and human	Maize	NA

The diet breadth model would predict that the older sites that have had maize agriculture for a shorter period of time would be more reliant on maize and have a narrower diet breadth (Gremillion 1996; Barlow 2002). This observation was confirmed by the two Basketmaker II sites. The observation of 14 dietary elements among the 4 samples, creates an average of 3.5 dietary elements sequenced per sample (using both pipelines) for the Boomerang site, the oldest of the sites. The Marsh Pass site's inhabitants had maize agriculture for 500 or more years and in confirmation with the diet breadth hypothesis had a lower average of 2.6 dietary elements sequenced per sample (13 elements among the 5 samples). Interestingly, the Rio Zape site is both the youngest of the three sites as well as the site with the highest average dietary elements sequences per sample at 5 elements. The calculated average, rather that the recorded elements present, should be the bases of comparison among the sites as the coprolites chosen are not representative of occupational density at the sites nor relative to the complete coprolite collection from each site. In whole, the calculated averages contrast the expected diet breadth but can be explain through preservation factors of the different environment. In previous analyses the Rio Zape sites have been noted as extremely well preserved (Jiménez et al 2012; Wibowo et al. in prep; Pucu 2020). The drier climate of the Rio Zape site may have facilitated better preservation of the DNA within the coprolites than

the more temperate Basketmaker II sites. A subsequent study of more sites with known dates should be conducted to further analysis the findings in this research.

6.0 Discussion

The results of this study lead to several interesting implications both to the knowledge of the greater Southwest, as well as molecular approaches toward archeological coprolites. The diet analysis was successful in expanding our existing knowledge of the diet of individuals living at these sites. While the floral remains were better suited for microremains analysis, the faunal components of the inhabitants' diet may have otherwise remained speculative, and the cacao unseen. Follow-up studies with primers better suited for degraded plant DNA would be a valuable addition to the information gathered in this thesis. Additionally, a hybrid bioinformatics pipeline that both treats the samples individually and screens for the specific primers used would be a valuable approach for similar studies.

6.1 Maize and Lack thereof

It was anticipated, based on archeological data of the region and time period, as well as through physical inspections of the coprolites, that maize would be a component of many of the samples; however, maize sequences were only found in the soil samples using the non-specific second bioinformatics pipeline. This would suggest that the primers were capable of amplifying corn DNA but the primers either were not well suited for the corn's DNA in the coprolites or the coprolites' corn DNA was not well suited for PCR amplification.

Studying corn genetics is complicated by the thousands of years of human interference with its genome. A strong example of this can be seen by the recent finding that "5,310 years ago, maize in the Tehuacan Valley was on the whole genetically closer to modern maize than to its wild counterpart" (Ramos-Madrigal et al 2016). The maize

genome is slightly smaller than our own 3 billion bases; however, the DNA content within the genome has been shown to differ by up to a billion bases of what are seemingly identical strains on a hillside (Pennisi, 2017). More particularly in the American Southwest, the adaptation for surviving the arid conditions has been observed as a 60% diversity reduction surrounding the sugary1 locus--the locus responsible for sweetcorn's mutation (Larson, 2015).

While the complications with the corn genome could explain the lack of corn DNA in the first pipeline's results, the BLAST results of corn in the second pipeline's soil sample analysis suggests there is more likely a problem with the corn DNA within the coprolites. It is likely that the processing of the maize ingested by past individuals is inhibiting the DNA from being PCR amplified. The processing of the maize both in food preparation and human digestion may have degraded the DNA to the extent that the PCR was not able to amplify any strands within the coprolites. Such a conclusion would also be supported by the lack of any sequences for the soil samples in the plant-primer specific table (table 3). Problems with the digested corn DNA would explain the lack of presence of maize in the coprolite samples and contrasting significant quantity in the soil samples seen in the galaxy pipeline results (samples 1043 and 3567, Figures 11 and 12).

The results of the molecular analysis, without comparison to the previously conducted microremains analysis, may have suggested that maize was not a significant component of the diet of the individuals at the three sites selected. Having the results of the previous study on the same coprolites, as well as archeological remains of dietary elements of similar sites all suggest problems with the plant primer used for this molecular analysis. This conclusion should serve as a cautious reminder to future studies

working with aDNA and coprolites that the lack of a species in the final results file does not necessarily correlate to a lacking presence of that species, but rather that DNA was damaged and/or lost in the many steps between harvest and bioinformatics results analysis.

6.2 Finding Cacao and Research Implications

Despite its flaws, the plant-primer did suggest the presence of cacao at the Boomerang shelter and possibly at the Rio Zape site. Boomerang shelter is located in contemporary Southern Utah; therefore, the presence of cacao at this site would have significant implications for cacao use in the Prehispanic American Southwest. Further studies with either improved plant primers and/or newer sequencing methods ought to be conducted to confirm these results and expand upon the existing knowledge of cacao in the desert southwest. It is foreseeable that many archeologists and molecular anthropologists will begin looking to newer methods such as shotgun sequencing for analyzing larger portions of their data. While this is a goal to strive for there is still the current need for better primers suited to the economical metabarcoding technique. In the meantime, more primers targeting specific species can be tested with coprolite remains to better establish the presence or absence of a species in the diet of past individuals.

While the presence of maize was not observable with the molecular analysis, the diet breadth was still observable. Through this research it was shown that the diet breadth in the American Southwest consisted of extensive dietary elements, particularly faunal, and while maize was a dietary staple in the region (Larson 2015) many insights can still be gained by the non-agricultural elements. Other plant and animal remains were found in the microremains and found again through molecular analysis- particularly small

mammals, rabbit, fish, owls, Dioon seeding plants and pollinating flowers (*Lasthenia californica*). These elements suggest the inhabitants of the sites were sustained by a wide diet.

6.3 Future Research

The molecular analysis of the remains was both complemented and complicated by the separately conducted microremains analysis. The microremains analysis was compiled of plant-based results, as it is hard to identify other dietary sources after the digestion process. The molecular approach provided more insight into the variety of fauna consumed with results such as sheep, squirrels, and other small mammals. Unfortunately, the molecular approach did not yield as much insight into the flora aspects of the diet, particularly in comparison to the previous microremains analysis. The most likely explanations for this would be that there was a problem with the plant specific primer or lack of amplifiable plant DNA (due to damage or processing) in the coprolites. Both options are expanded upon in the discussion.

7.0 Conclusion

The metabarcoding method, with the use of coprolites, did expand the knowledge of diet breadth for the sites and respective region. In line with the test expectations, it was easier to assign dietary elements using the primer-specific R script method because it increased the ratio of flora and fauna sequences to bacterial and viral reads. These results support hypothesis one. The test expectations regarding hypothesis two predicted the samples would be easier to compare individually against each other, against previously conducted microremains analysis of the samples, and as site-based groups against each other and other dietary studies conducted in the region. The results of the second bioinformatics method resulted in many non-specific sequences that yielded a variety of possible species using the BLAST feature, all with similar confidence scores. As a result, the species assignment process was difficult, and often left to a human judgement call. The individual samples could be compared against the microremains analysis of previously conducted studies but the species identification was too uncertain to compare samples against each other, or based on site. While hypothesis two was not fully supported by the results, the sample specific results did show that corn/maize DNA was only amplified in the soil samples. The lack of corn DNA in coprolites with microremains evidence for maize is explored further in the discussion section.

While neither pipeline method yielded as broad of an array of flora results as the microremains analysis, the R script table from the plant primer-specific run did find that at least one sample had sequences that were significantly likely to be cacao. Sample TS904A is a Boomerang shelter coprolite with 14 strands of sequenced DNA that, when compared in the NCBI database using the BLAST feature, appear to be no other plant

species other than *Theobroma cacao*. In addition to this finding, the sample TS899 had 28 sequences that, when compared in the NCBI database using the BLAST feature, share similar likelihood of being either the daisy species *Lasthenia californica*, some variety of a Dioon seed plant, or *Theobroma cacao*. This same sample has another sequence with 14 strands associated with the *Lasthenia californica* plant. This same plant species is a common BLAST result for both pipeline methods.

8.0 References:

Ashelford, K., Chuzhanova, N., Fry, J., Jones A., and Weightman, A. 2005. "New Screening Software Shows that Most Recent Large 16S rRNA Gene Clone Libraries Contain Chimeras"Applied and Environmental Microbiology 71, 7724.

Boddy, Jessica. 2016. "Catching Ancient Maize Domestication in the Act." Science. American Association for the Advancement of Science. https://doi.org/10.1126/science.354.6315.953.

Barlow, R. W. 2002. "Predicting maize agriculture among the Fremont: An economic comparison of farming and foraging in the American Southwest. American Antiquity, 67, 65-88.

Blankenberg, D. and Gordon, A. and Von Kuster, G. and Coraor, N. and Taylor, J. and Nekrutenko, A. 2010. Manipulation of FASTQ data with Galaxy. In Bioinformatics, 26 (14), pp. 1783â□"1785. [doi:10.1093/bioinformatics/btq281]

Bostwick, Todd W, and Karen R Adams. 2016. Plant Remains.

Brooks, R. H., Kaplan, L., Cutler, H. C. & Whitaker, T. W. 1962. "Plant Material from a Cave on the Rio Zape, Durango, Mexico." Am. Antiq. 27, 356–369.

Callahan, Benjamin J, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson, and Susan P Holmes. 2016. "DADA2: High-Resolution Sample Inference from Illumina Amplicon Data." Nature Methods 13 (7): 581–83. https://doi.org/10.1038/nmeth.3869.

Charles, Mona C, and Sally J Cole. 2006. "Chronology and Cultural Variation in Basketmaker II." Source: Kiva. Vol. 72. Winter.

Charles, Mona C, Leslie M Sesler, and Timothy D Hovezak. 2006. "Understanding Eastern Basketmaker II Chronology and Migrations." Source: Kiva. Vol. 72. Winter.

Chen, Shifu and Zhou, Yanqing and Chen, Yaru and Gu, Jia. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. [doi:10.1101/274100]

Crandall, J. J., Martin, D. L. & Thompson, J. L. 2012. "Evidence of Child Sacrifice at La Cueva de los Muertos Chiquitos (660-1430 AD)." in Landscapes of Violence vol. 2 No. 2, Article 12.

Crown, Patricia L, and W Jeffrey Hurst. 2008. "Evidence of Cacao Use in the Prehispanic American Southwest." Smithsonian Institution. www.pnas.orgcgidoi10.1073pnas.0812817106.

Dabney, Jesse, Michael Knapp, Isabelle Glocke, Marie-Theres Gansauge, Antje Weihmann, Birgit Nickel, Cristina Valdiosera, et al. 2013. “Complete Mitochondrial Genome Sequence of a Middle Pleistocene Cave Bear Reconstructed from Ultrashort DNA Fragments.” Proceedings of the National Academy of Sciences 110 (39): 15758. https://doi.org/10.1073/pnas.1314445110.

de Araujo E, P., Russ, J. & Reinhard, K. 2019. Gut Microbiome and Diet: What Can Coprolites Tell Us?Archaeological and Anthropological Sciences (accepted for publication).

Fonseca, Rute R. Da, Bruce D. Smith, Nathan Wales, Enrico Cappellini, Pontus Skoglund, Matteo Fumagalli, José Alfredo Samaniego, et al. 2015. “The Origin and Evolution of Maize in the Southwestern United States.” Nature Plants 1 (January). https://doi.org/10.1038/nplants.2014.3.

Geib, Phil. 2011. Foragers and Farmers of the Northern Kayenta Region: Excavations along the Navajo Mountain Road. University of Utah Press.

Gremillion, K.J. 1996. “Diffusion and Adoption of Crops in Evolutionary Perspective” Journal of Anthropological Archaeology, 15, 183-204.

Guernsey, S. J. & Kidder, A. V. 1921. Basket-Maker Caves of Northeastern Arizona: Report on the Explorations, 1916-17. vol. 8 No. 2. Cambridge, Mass.: Papers of the Peabody Musuem of American Archaeology and Ethnology.

1931. Explorations in Northeastern Arizona: Report on the Archaeological Fieldwork of 1920-1923. (The Museum).

Hugenholtz, Philip, and Thomas Huber. 2003. “Chimeric 16S RDNA Sequences of Diverse Origin Are Accumulating in the Public Databases.” International Journal of Systematic and Evolutionary Microbiology 53 (1): 289–93. https://doi.org/10.1099/ijs.0.02441-0.

Hünemeier, Tábita, Carlos Eduardo Guerra Amorim, Soledad Azevedo, Veronica Contini, Víctor Acuña-Alonzo, Francisco Rothhammer, Jean Michel Dugoujon, et al. 2012. “Evolutionary Responses to a Constructed Niche: Ancient Mesoamericans as a Model of Gene-Culture Coevolution.” PLoS ONE 7 (6). https://doi.org/10.1371/journal.pone.0038862.

Hammerl, E. E., Baier, M. A. & Reinhard, K. J. 2015. “Agave Chewing and Dental Wear: Evidence from Quids.” PLoS One 10, e0133710.

Jackson, Fatimah. 1996. “The Coevolutionary Relationship of Humans and Domesticated Plants.” YEARBOOK OF PHYSICAL ANTHROPOLOGY. Vol. 39.

Jervis-Bardy, J., Leong, L.E.X., Marri, S. et al. 2015."Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data." Microbiome 3, 19

Jiménez, F. Agustín, Scott L. Gardner, Adauto Araújo, Martín Fugassa, Richard H. Brooks, Elizabeth Racz, and Karl J. Reinhard. 2012. "Zoonotic and Human Parasites of Inhabitants of Cueva de Los Muertos Chiquitos, Rio Zape Valley, Durango, Mexico." Journal of Parasitology 98 (2): 304–9. https://doi.org/10.1645/ge-2915.1.

Kennett, Douglas J., and Bruce Winterhalder. 2006. Behavioral Ecology and the Transition to Agriculture. Origins of Human Behavior and Culture. Edited by and Bruce Winterhalder Douglas J. Kennett. Berkeley, Los Angeles, London: University of California Press. https://ebookcentral.proquest.com/lib/msoumt/reader.action?docID=254883&query=.

Kirchhoff, Paul. 1954. "Gatherers and Farmers in the Greater Southwest: A Problem in Classification." American Anthropologist: 56

Kidder, A. V. & Guernsey, S. J. 1919. Archaeological Explorations in North-eastern Arizona. Washington D. C.: Government Printing Office.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC. 1989. "Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers." Proc. Natl Acad. Sci. USA. 86:6196–6200.

Krzywinski, M. and Schein, J. and Birol, I. and Connors, J. and Gascoyne, R. and Horsman, D. and Jones, S. J. and Marra, M. A. 2009. Circos: An information aesthetic for comparative genomics. In Genome Research, 19 (9), pp. 1639â□"1645. [doi:10.1101/gr.092759.109]

Larson, Greger. 2015. "Crop Domestication: Corn in the USA." Nature Plants. Palgrave Macmillan Ltd. https://doi.org/10.1038/nplants.2014.9.

Laudadio I, Fulci V, Stronati L, Carissimi C. 2019. "Next-Generation Metagenomics: Methodological Challenges and Opportunities." Omics a Journal of Integrative Biology 23(7): 327-333.

Li, Dinghua and Liu, Chi-Man and Luo, Ruibang and Sadakane, Kunihiko and Lam, Tak-Wah. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. In Bioinformatics, 31 (10), pp. 1674â□"1676. [doi:10.1093/bioinformatics/btv033]

Meyer, Matthias & Kircher, Martin. 2010. "Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing." Cold Spring Harbor protocols.

Matson, R G. and Chisholm, Brian. 1991. “Basketmaker II Subsistence: Carbon Isotopes and Other Dietary Indicators from Cedar.” Source: American Antiquity. Vol. 56.

Matson, R G. 2006. “What Is Basketmaker II?” Source: Kiva. Vol. 72. Winter.

Morrow, J. J. & Reinhard, K. J. 2016. “Cryptosporidium parvum Among Coprolites from La Cueva de los Muertos Chiquitos (600-800 CE), Rio Zape Valley, Durango, Mexico.” J. Parasitol. 102, 429–435.

Mowrer, Kathy. 2006. “Basketmaker II Mortuary Practices: Social Differentiation and Regional Variation.” Source: Kiva. Vol. 72. Winter.

Muyzer, E. C. de Waal, G. and Uitterlinden, A. G. 1993. Applied Environmental Microbiology. 59, 695

Palumbi S. 1996. Nucleic acids II: the polymerase chain reaction. In: Hillis D, Moritz C, Mable B, editors. Molecular Systematics. 2nd edn. Sunderland, MA: Sinauer Assoc.; pp. 205–247.

Pennisi, Elizabeth. 2017. “Unlocking a Key to Maize’s Amazing Success.” Science. American Association for the Advancement of Science. https://doi.org/10.1126/science.357.6348.240.

Pucu, Elisa, Julia Russ, and Karl Reinhard. 2020. “Diet Analysis Reveals Pre-Historic Meals among the Loma San Gabriel at La Cueva de Los Muertos Chiquitos, Rio Zape, Mexico (600–800 CE).” Archaeological and Anthropological Sciences 12 (1). https://doi.org/10.1007/s12520-019-00950-0.

Quail, Michael A., Miriam Smith, Paul Coupland, Thomas D. Otto, Simon R. Harris, Thomas R. Connor, Anna Bertoni, Harold P. Swerdlow, and Yong Gu. 2012. “A Tale of Three next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq Sequencers.” BMC Genomics 13 (1). https://doi.org/10.1186/1471-2164-13-341.

Ramos-Madrigal, Jazmín, Bruce D. Smith, J. Víctor Moreno-Mayar, Shyam Gopalakrishnan, Jeffrey Ross-Ibarra, M. Thomas P. Gilbert, and Nathan Wales. 2016. “Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication.” Current Biology 26 (23): 3195–3201. https://doi.org/10.1016/j.cub.2016.09.036.

Reinhard, K. J., Edwards, S., Damon, T. R. & Meier, D. K. 2006. “Pollen concentration analysis of Ancestral Pueblo dietary variation.” Palaeogeogr. Palaeoclimatol. Palaeoecol. 237, 92–109.

Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E. 2011. "ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis." Nucleic Acids Res;39(21):e145.

Ross, Michael G., Carsten Russ, Maura Costello, Andrew Hollinger, Niall J. Lennon, Ryan Hegarty, Chad Nusbaum, and David B. Jaffe. 2013. "Characterizing and Measuring Bias in Sequence Data." Genome Biology 14 (5). https://doi.org/10.1186/gb-2013-14-5-r51.

Sharpton, Thomas J. 2014. "An Introduction to the Analysis of Shotgun Metagenomic Data." Frontiers in Plant Science. Frontiers Research Foundation. https://doi.org/10.3389/fpls.2014.00209.

Shah, Neethu, Haixu Tang, Thomas G Doak, and Yuzhen Ye. 2010. "Comparing Bcterial Communities Inferred from 16S rRNA Gene Sequencing and Shotgun Metagenomics." www.worldscientific.com.

Smiley, F. E. & Robins, M. R. 1994. "The Agricultural Transition in the Northern Southwest: Patterns in the Current Chronometric Data." Kiva 60: 165089. University of Arizona Press, Tucson.

1997. In Early Farmers in the Northern Southwest: "Papers on Chronometry, Social Dynamics, and Ecology. in A brief description of a selection of Basketmaker II rockshelter sites in the northern Southwest" 43–58 (United States Department of the Interior, Bureau of Reclamation, Upper Colorado Region.,).

2002. "Black Mesa before Agriculture: Paleoindian and Archaic Evidence." In Prehistoric Culture Change on the Colorado Plateau: Ten Thousand Years on Black Mesa, edited by Shirley Powell and Francis E. Smiley, pp. 15-34. University of Arizona Press, Tucson.

Smith, Bruce D. 2015. "A Comparison of Niche Construction Theory and Diet Breadth Models as Explanatory Frameworks for the Initial Domestication of Plants and Animals." Journal of Archaeological Research 23 (3): 215–62. https://doi.org/10.1007/s10814-015-9081-4.

2001. "Low-Level Food Production." Journal of Archaeological Research 9(1):1-43.

Stiner, Mary C, and Natalie D Munro. 2002. "Approaches to Prehistoric Diet Breadth, Demography, and Prey Ranking Systems in Time." Source: Journal of Archaeological Method and Theory. Vol. 9.

Swarts, Kelly, Rafal M Gutaker, Bruce Benz, † Michael Blake, † Robert Bukowski, James Holland, Melissa Kruse-Peeples, et al. n.d. "Genomic Estimation of Complex Traits Reveals Ancient Maize Adaptation to Temperate North America." http://science.sciencemag.org/.

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E. 2007. “Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding.” Nucleic Acids Res; 35:e14.

Winterhalder, Bruce. 1997. “An Evolutionary Ecology Perspective on Diet Choice, Risk, and Plant Domestication Maya Hunting with Dogs in Belize View Project.” https://www.researchgate.net/publication/239731232.

Wood DE, Salzberg SL. 2014. “Kraken: ultrafast metagenomic sequence classification using exact alignments.” Genome Biology 15(3): R46.