# MITOCHONDRIAL GENOME RECONSTRUCTION OF *Rattus rattus* OBTAINED FROM AN OWL PELLET FROM TRES BOCAS CAVE IN THE DOMINICAN REPUBLIC

by

Marlys E. Massini Espino

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
In
BIOLOGY

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2018

Approved by:

| | |
|---|---|
| _____ | _____ |
| Alex Van Dam, Ph.D.<br>President, Graduate Committee | Date |
| _____ | _____ |
| Carlos J. Santos-Flores, Ph.D.<br>Member, Graduate Committee | Date |
| _____ | _____ |
| Sean Locke, Ph.D.<br>Member, Graduate Committee | Date |
| _____ | _____ |
| Jaime Acosta-Martínez, Ph.D.<br>Member, Graduate Committee | Date |
| _____ | _____ |
| Ivette Cruzado-Vélez<br>Representative of Graduate Studies | Date |
| _____ | _____ |
| Ana V. Vélez, M.S.<br>Chairperson of the Department | Date |

# ABSTRACT

Molecules of DNA are ubiquitous in the environment and a rich source of biological information, but because DNA degrades rapidly in tropical climates studies of ancient DNA in the Caribbean have been limited until recently. Fossilized owl pellets are one potential source of ancient DNA with useful phylogenetic information. In this study, we obtained a fossilized *Rattus rattus* sample from an owl pellet excavated from Tres Bocas Cave in the Dominican Republic. We sequenced DNA pooled from three rat maxillae from the pellet, annotated the mitochondrial genome and compared the sequence to contemporary data from *R. rattus* from other studies. Maximum Likelihood phylogenetic analysis showed the sample from the Dominican Republic was nested within a clade of *R. rattus* from Madagascar and East Africa. The sequence from *R. rattus* from the Dominican Republic had few nucleotide misincorporations, suggesting a recent origin, which was corroborated with $^{14}C$ isotope ratios that indicated the sample dates from 1960-70. We have demonstrated that phylogenetic information can be obtained from DNA from bones in decades-old pellets in a tropical region, which validates an ancient DNA extraction technique and indicates further sampling in the Caribbean may prove fruitful.

## RESUMEN

El ADN sufre daños moleculares a un ritmo más rápido en climas tropicales que en climas templados. Esta es una razón por la que los estudios en el Caribe han sido limitados hasta años más recientes. Se ha ignorado que los depósitos de egagrópilas de búho son una fuente de ADN antiguo viable que provee información filogenética. En esta investigación, se obtuvo una muestra de *Rattus rattus* de una egagrópila de búho, colectado de la Cueva De Tres Bocas en República Dominicana. Se secuenciaron tres maxilas de esta especie de rata que provinieron de una sola egagrópila de búho, se anotó el genoma mitocondrial y se construyó el árbol filogenético de la muestra. La muestra se ubicó dentro de un clado monofilético de Madagascar y del Este de África y proveyó pocos eventos de sustitución que son característicos del ADN antiguo. Esto significa que la muestra no es lo suficientemente vieja para considerarse como ADN antiguo, lo cual se corroboró con la datación de fracción D14 que indicó que la muestra es de los años 1960-70. Al lograr construir el genoma mitocondrial y el árbol filogenético a partir de la muestra, se demostró que el Caribe puede proveer material genético viable que puede expandir las posibilidades de muestreo para incluir esta región y probar una técnica de extracción para ADN antiguo obtenido en el Caribe.

*Knowing where you came from is no less important than knowing where you are going.*

*-Neil DeGrasse Tyson*

# ACKNOWLEDGEMENTS

I would sincerely like to thank Dr. Van Dam for his endless patience, skill and knowledge of how to fix all my thesis problems, never regarding them as only mine but also carrying the load as his own. I need to thank my Committee: Dr. Acosta, Dr. Santos and Dr. Locke, for their valuable input and impeccable corrections.

I must thank the University of Puerto Rico, Mayagüez Campus, for being my home and giving me all the tools I needed to carry on.

To my mother: for her words of wisdom and eternal love. You will never know how many times you have brought me back from the edge. Thank you forever, mom. I wouldn't be half of who I am without you.

To my friends: Keysha, Kenneth, Paloma, Marialis, Selena, Lyann, Samalis and Kritzia. You have never failed me with your unshakable support and willingness to listen.

To my brother, Julio, for being himself and not letting me take myself too seriously. To my sister, Omayra, for being incredibly strong in the face of adversity and for being my constant reminder to laugh because everything is not as bad as it seems.

To Gene: Thank you. For everything. You'll never know the peace I've found with you.

To all of those who have read my thesis with the purpose of helping: Derek Soto, Beatriz Romero and Zongie Rodríguez come to mind.

This thesis was for me but also, a little bit, for all of you.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# INTRODUCTION

Most studies of ancient DNA have taken place in temperate regions, as DNA is better preserved at low temperatures (Bressan et al., 2014; Gutiérrez-García et al., 2014). DNA tends to become degraded more rapidly in warm climates due to enzyme and microbe activity (Wade, 2015). In more recent years, several studies of environmental DNA (eDNA) and ancient DNA (aDNA) have been conducted in the tropics on tropical frogs, Bahamian giant tortoise, human indigenous groups, etc. (Bálint et al., 2018; Kehlmaier et al., 2017; Kuzmina et al., 2018; Lopes et al., 2017; Mendisco et al., 2015; Robson et al., 2016). In the Caribbean however, other potential sources of tropical aDNA have not been explored. For example, a variety of sub-fossil deposits occur in karst caves. One type of cave deposit that is especially rich in small mammals is owl pellets (Gutiérrez-García et al., 2014). Some Caribbean owl pellet deposits date back thousands of years, making them useful in reconstructing past relationships and completing the topography of phylogenies (Gutiérrez-García et al., 2014). The present study follows these recent advances in aDNA sequencing techniques by applying these techniques to small mammal deposits from owl pellets found in karst caves in the Dominican Republic.

The emergence of Next Generation Sequencing (NGS) has advanced research in aDNA from short mitochondrial sequences to whole genomes (Damgaard et al., 2015). NGS technology, such as Illumina, has allowed retrieval of short segments of mtDNA and characterization of whole genomes from aDNA (Damgaard et al., 2015; Morozova &

Marra, 2008). In the past, many specimens could not be sequenced because of extensive degradation caused by depurination (Hansen et al., 2006). NGS and aDNA extraction techniques can reveal functions and phenotypes of ancient genomes from a wide variety of specimens due to their sensitivity to low frequency variants which are common in aDNA (Der Sarkissian et al., 2015; Kehlmaier et al., 2017; Mendisco et al., 2015).

Owl pellets are a rich source of small mammal bones that have utility in reconstructing past ecosystems (Fumagalli et al., 1996; Gutiérrez-García et al., 2014; Martin et al., 2009; Reed & Reed, 1928; Sharikov et al., 2018; Vigne & Valladas, 1996). The majority of previous work using owl pellets has focused on morphological and molecular features to identify small rodents to species (Galan et al., 2012; Martin et al., 2009; Matisoo-Smith & Allen, 2001). More recently environmental DNA sequencing has been applied to owl pellets to rapidly reconstruct small mammal prey communities (Ogada, 2018; Sharikov et al., 2018; Rocha et al., 2015). Developing methods that could identify past communities would be of utility to investigating these bone deposits further. A number of extinction events have occurred with the arrival of domestic and peridomestic pests during European colonization of the western hemisphere in the 16th century (Hingston et al., 2007; Lund, 2015; Thompson, 2015). The black rat, *Rattus rattus*, is one of the most widely distributed pests due to its opportunistic abilities (Backhans et al., 2012; Ewer, 1971; Koma et al., 2013; Aplin et al., 2011; Terkel, 1996). In this project, the utility of owl pellets as a source of aDNA was assessed using three rat skulls extracted from a single owl pellet collected from Tres Bocas Cave, a karst cave in

the Dominican Republic. The study has potential historical relevance in that the results revealed an unexpected geographical origin of this rat.

My first aim was to show that small mammal bones deposited in owl pellets from humid karst caves can serve as a viable source of genetic material for aDNA analyses. Secondly, I performed a phylogenetic reconstruction of mitochondrial DNA of *R. rattus* to determine this sample's geographic origin. This investigation demonstrates that genetic material viable for NGS can be obtained from prey in decades-old rat pellets in the Caribbean using a technique for aDNA extraction that until now had only been applied in temperate regions. The present work also serves as future reference that the extraction and bioinformatics analysis used for this investigation is versatile for many kinds of eDNA, phylogenetic and even pathogen detection investigations.

# LITERATURE REVIEW

In this section, I will present a literature review of the relevant topics that will be discussed in the rest of my thesis.

**1. Owl Pellet Genetic Deposits**

Limited research is available on genetic analysis of prey in pellets of birds of prey but interest in using owl pellets for DNA extractions has recently grown (Bobola et al., 2018; Comay & Dayan, 2018; Horváth et al., 2018; Malhotra & Singla, 2018; Moysi et al., 2018). Owl pellets can be used as a non-invasive sampling method for genetic studies of small mammal prey. The first to succeed in DNA extraction from bones in owl pellets were Fumagalli et al. (1996). Poulakakis et al. (2004) demonstrated that mitochondrial DNA (mtDNA) of small mammals can be isolated and amplified from owl pellets. Poulakakis et al. (2004) targeted the cytochrome *b* gene region, which is a highly conserved region within a species and variable among different species (Castresana, 2001). Cytochrome *b* is the gene most widely sequenced among mammals, which makes it useful for comparisons in phylogenetic studies (Avise & Aquadro, 1982; Irwin et al., 1991; Meyer, 1994). Pellet analysis has become a popular way of obtaining information on the food habits of raptors (Ogada, 2018; Sharikov et al., 2018). Poulakakis et al. (2005) tested different DNA extraction methods and discovered that the silica purification originally proposed by Boom et al. (1990) yielded more successful amplifications compared to the protocol proposed by Holmes & Bonner (1973), though Holmes &

Bonner's method is faster, simpler and more cost effective. Studies of owl pellets indicate DNA is best preserved in bones from the cranium, but DNA can also be extracted from tibiae, femora and coxae (Sharikov et al., 2018).

**2. Ancient DNA for analysis**

Extraction techniques for ancient DNA have enabled the phylogenetic analysis of extinct species and populations and the study of long-term temporal genetic changes (Bandelt et al., 2005). Ancient DNA is highly modified due to oxidative processes which are responsible for low recovery rate of undamaged DNA (Kalmar, 2000). Additionally, the study of aDNA includes problems such as accurate assessment of specimen provenance and dating. For example, the production of miscoding lesions in the sequence can lead to sequence errors, or physical destruction such as insertions, substitutions and deletions of the molecule, which can increase the chance of amplifying a contaminant sequence (Bandelt et al., 2005; Jónsson et al., 2013). Poor DNA preservation is the most limiting factor in ancient genomic research (Damgaard et al., 2015). The presence of PCR inhibitors, which vary in concentration among samples, adds to the difficulties of working with aDNA (Rohland & Hofreiter, 2007). Some of the PCR inhibitors are: silica, environmental remains such as humic acid, fulvic acid, and in the case of bones, collagen type I and Maillard products (Hänni et al., 1995; Höss & Pääbo, 1993; Kalmar, 2000). Rohland & Hofreiter (2007) propose the use of a buffer of EDTA and proteinase K only, in a silica-binding purification process, to reduce further degradation.

Free DNA molecules are ubiquitous in the environment, released from skin, saliva, mucus, feces, urine, blood, fruit, pollen, regurgitations, rotting bodies, etc., and are collectively referred to as environmental DNA (Bohmann et al., 2014). Environmental DNA is extracted from media such as soils, water or air without first isolating target organisms (Taberlet et al., 2012). The integration of eDNA techniques alongside traditional monitoring can increase the quality and extent of natural resource monitoring programs. Environmental DNA is often used to detect the spread of invasive species, quantify populations, identify distributions and for biodiversity assessments (Merkes et al., 2014; Nevers et al., 2018).

**3. Mitochondrial DNA**

In vertebrates, mtDNA is transmitted maternally with a mutation rate 5-10 times higher than nuclear DNA (Brown et al., 1979; Mandal et al., 2014). This high mutation rate in mtDNA is due to limited repair systems, although the mitochondrial genome also has highly conserved regions that allow universal primers to be developed (Kocher et al., 1989; Mandal et al., 2014). Mitochondrial DNA does not recombine and is present in high numbers of copies within cells, which facilitates recovery of this molecule from degraded samples (Mandal et al., 2014). Once the mtDNA genome is sequenced it is possible to correctly identify the specimen that was collected. An option for this is by identifying the well conserved mitochondrial genes that have shown to be sensitive genetic markers of species relationships and genetic differences among taxa (Muraji et al., 2000). *Cytochrome c oxidase I* (COI) mtDNA gene is used as a species identification tool

and it has been shown to provide robustness in some groups such as arthropods, birds, fish and more recently mammals (Hebert et al., 2003; Hebert et al., 2004; Barrett et al., 2005; Cywinska et al., 2006; Ward et al., 2005; Clare et al., 2006; Robins et al., 2007; Pages et al., 2010; Francis et al., 2010; Clare et al., 2011; Nicolas et al., 2012). The DNA barcoding technique has been proposed as a method for species identification in a standard gene region on the basis of pairwise evolutionary divergence, such as the CO1 gene (Hebert et al., 2003; Hingston et al., 2007). DNA barcoding is based on the observation that intraspecific genetic divergences are significantly smaller than divergences between species and as a consequence species identities can be assigned to an unknown specimen based on its relative genetic divergence from known species (Hingston et al., 2007). False-positive barcode identifications can arise from paraphyly, which are groups consisting of a common ancestor and most but not all of its descendants, poor species sampling or small interspecific genetic divergences (Moritz & Cicero, 2004; Will & Rubinoff, 2004) but DNA barcoding has proven to be a valuable method for species identification in animals (Hingston et al., 2007; Ondrejicka et al., 2014). Cytochrome b (*cyt b*) is another marker that is commonly used to determine species boundaries in mammals as it has been shown that it accurately reconstructs the mammalian phylogeny and gives better resolution for separating species (Bradley & Baker, 2001; Nicolas et al., 2012).

**4. Historic importance**

The genus *Rattus* originated in Southeast Asia, and are known to stow away in ships and colonize new territories as the ships dock in ports (Aplin & Singleton, 2003; Wang et al., 2000). Since the 16$^{th}$ century, these ships that likely contained rats have made their way into the Caribbean as it is a strategic location for maritime transport, between the Americas and Europe (Thompson, 2015). The phylogeographic histories of rats and mice can also be valuable for tracing connections between human history and introductions of alien species. Phylogenies of mitochondrial DNA sequences have provided useful information for inferring the history of *R. rattus*, tracking human movement, rodent pathways of colonization, origins, and time frames of introduction (Tollenaere et al., 2010). During the 60's-70's, studies of chromosomes and blood proteins identified patterns of geographic variation that separated the *Rattus* species into two weakly differentiated lineages that came from Europe and India (reviewed by Aplin et al., 2011). With this in mind, the samples obtained in the Caribbean could have been originally from any part of the Eastern Hemisphere. However, given that the Caribbean has served as a maritime hub between Europe and North America, we expected a European or possibly West-African origin for *R. rattus* samples in the Dominican Republic.

# METHODOLOGY

In this section, I present the materials and methods that were used throughout my investigation.

**1. Sample collection**

The Tres Bocas Cave, located in Pedernales Province in the Dominican Republic (see Figure 1), was excavated and dry sifted to obtain the owl pellets (J. Almonte and A. Mychajliw). This cave has three vertical entrances: the first is 3 m tall, 2.3 m wide and 7 m deep; the second is 3 m tall, 2.8 m wide and 7.5 deep; and the third is 12 m tall, 9 m wide and 12 m deep. The owl pellets were located in the third entrance.

Inside the cave, a spot chosen at a random was outlined with a 60 cm x 60 cm square marked on the floor with nylon rope. Within the square, an 18 cm hole was dug with a small shovel using sequence stratigraphy method. The owl pellet, which contained three *R. rattus* skulls, was collected from the first stratum of the sediment that was extracted at a depth of 0-5 cm. The pellet is thought to originate from *Tyto alba*, commonly known as the barn owl, which occurs in the region (Almonte, personal communication). The other local strigiform, the endemic Dominican ashy-faced owl (*Tyto glaucops*), eats smaller prey and lives in more wooded areas.

Material covering the rat skulls was brushed off with a small, soft-bristled brush at the "Museo Nacional de Historia Natural Profesor Eugenio de Jesús Marcano" Paleontology Lab, Santo Domingo, Dominican Republic.

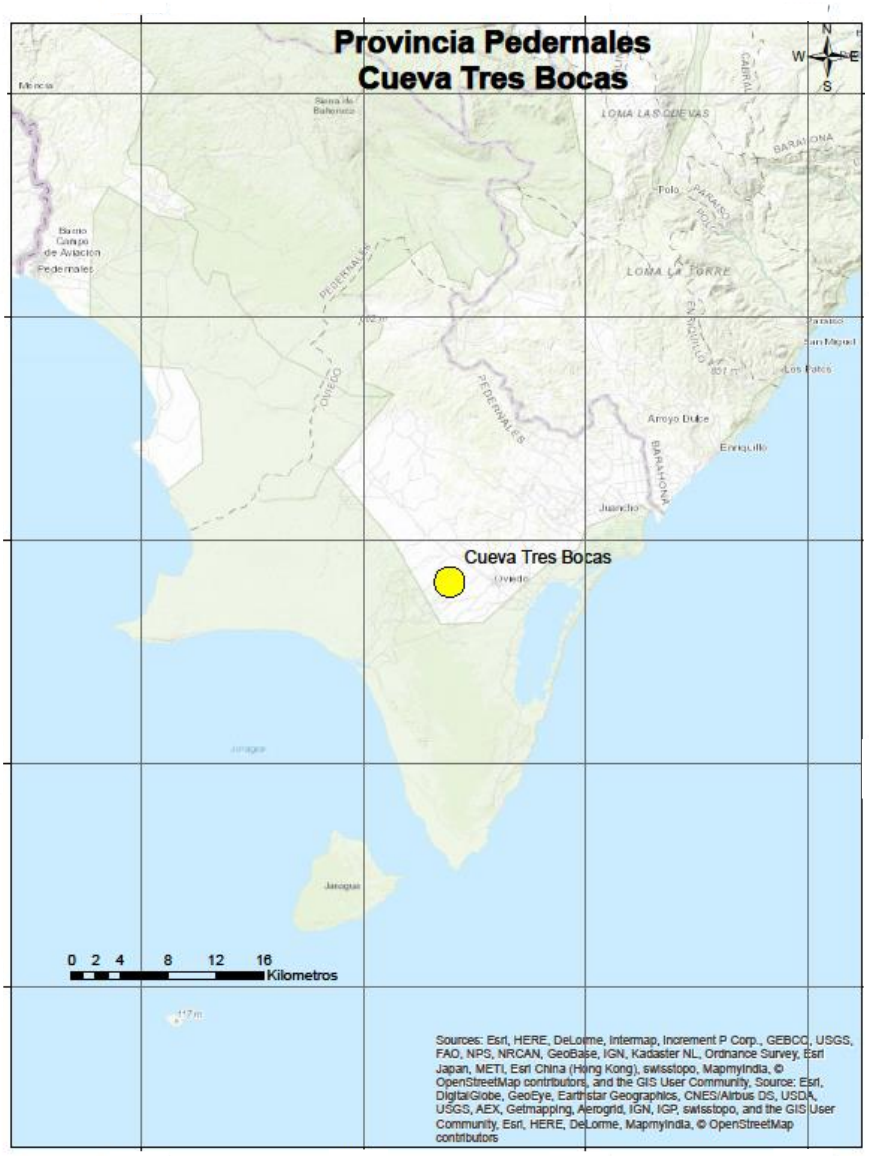


**Figure 1:** Map of the cave in the Pedernales Province of the Dominican Republic.

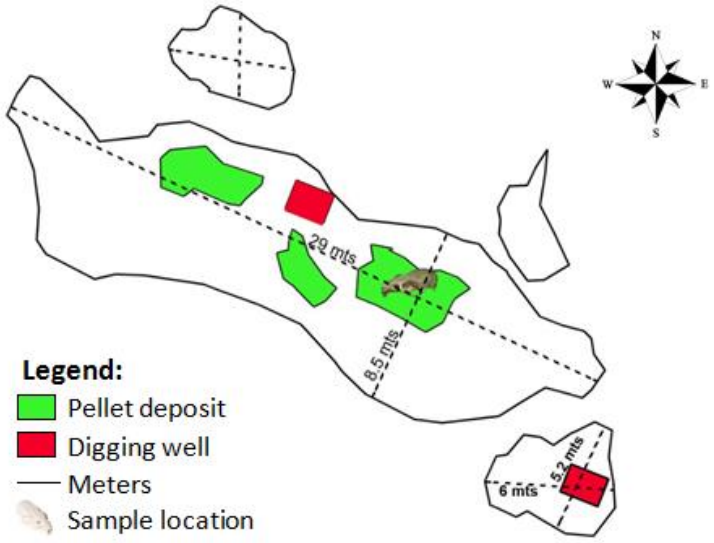


**Figure 2:** Diagram of the Tres Bocas Cave.

**2. Radiocarbon dating**

Dr. A. Mychajliw conducted the D14 fraction dating, also known as $^{14}$C-dead material fraction, using CALIB (Stuiver, 2018). Radiocarbon dating is mostly performed by accelerator mass spectrometry (AMS), which counts the atoms of $^{14}$C and provides a final result of an isotope ratio that is converted to a fraction of modern carbon (Hole et al., 2001). This radiocarbon dating technique is for modern $^{14}$C, post-1950 date. The smoothing filter estimate for collagen turnover was of 2 years and no marine carbon correction was done. The fraction modern calculations are done by dividing the ratio of normalized carbon by the ratio of modern carbon.

AMS $^{14}C$ measurements began by converting the carbon in the sample to $CO_2$, the $CO_2$ was converted to filamentous carbon, known as graphite, by using a catalytic process using metal as the catalyst. The sample was pressed into a target holder and placed in an accelerator's ion source, called a carousel. The carbon atoms within the sample were converted to single negatively charged ions by bombarding them with Cs ion beam. The final isotope selection process occurred in a detector, which discriminates particles based on each particle's energy. The charge from the atoms collected was converted to pulses using Coulomb's law and the charge of an electron and the ratio of $^{14}C$ were calculated. The $^{14}C$ ratio was used to calculate the fraction modern (Hole et al., 2001).

**3. DNA extraction**

At the GeoGenetics Ancient DNA Lab, Copenhagen Denmark, DNA was extracted from the maxillae of three rat skulls from a single owl pellet and pooled together. The maxilla from each skull was crushed with pliers that had been previously soaked in 15% bleach. The DNA extraction was performed by Dr. M. Allentoft using a method similar to Damgaard et al. (2015). The sample was homogenized and was transferred to a 15 mL tube. The sample was treated with a digestion buffer that contained 4.7 mL 0.5M EDTA, 50 µL recombinant Proteinase K, and 250 µL 10% N-Laurylsarcosyl, and incubated at 50 ºC. After 15 minutes, the extraction was centrifuged and supernatant removed. The same digestion buffer was added to the sample, was vortexed, and incubated for 24 hours.

For extracting the DNA from the supernatant, the binding buffer was prepared by mixing 118.2 g Guanidium Thiocyanate with 10 mL Tris 1M, 1 mL NaCl 5M, 8 mL EDTA 0.5M, 1 g N-Lauryl-Sarcosyl and water to a total volume of 200 mL. Of that volume, 20 mL were transferred to the sample and left rotating for 3 hours with 100µL silica powder in solution to bind to the DNA. After this, the silica was centrifuged and washed twice with 80% cold ethanol and the DNA eluted in 80 µL EB Buffer (Qiagen). Using the TruSeq DNA Sample Preparation Kit, we generated single-end sequencing libraries with the universal Illumina adapters that were used as primers for PCR amplification in the modern lab.

**4. DNA amplification and sequencing**

Following addition of the sequencing and barcode primers samples were moved to the modern lab at the University of Copenhagen National High-throughput DNA Sequencing Center. The concentration of the DNA was measured with Qubit Fluorometric Quantitation (Life Technologies, Grand Island, NY) (Damgaard et al., 2015).

The sample was subjected to 10 cycles of PCR amplification using the standard protocol to bring the total concentration of DNA up to sequencing thresholds. For the Illumina, single end DNA molecule is attached to the bead surface using an adapter; the molecules subsequently bend over and hybridize to complementary adapters thereby forming the template for the synthesis of their complementary strands. After the amplification step, the solid surface has approximately 1000 clonal copies of the template

which then are sequenced. Samples were loaded on a Hiseq 3000 at the National High-throughput DNA Sequencing Centre. This is done using DNA sequencing-by-synthesis approach that uses reversible terminators with removable fluorescent moieties and DNA polymerases that can incorporate these terminators into growing oligonucleotide chains (Morozova & Marra, 2008). These terminators are marked with different colors to differentiate among the bases at the sequence positions. The template sequence is recorded by reading the different colors of each nucleotide addition step. Our sample was pooled at equal-molar concentration with 11 other samples before loading onto a single lane of the Illumina Hi-Seq 3000.

**5. Bioinformatics analysis**

Initially, we loaded the data into the FastQC program (Andrews, 2010) to obtain a qualitative assessment of the quality of our reads. The Pittsburgh Supercomputing Center (PSC), a remote supercomputer access, provided the computing power for the bioinformatics analysis. Quality filtering was completed with cutadapt 1.12 to remove adapter sequences from the sample and keeping the sequences with a minimal length of 30bp and a minimum quality of score of 30 as this is the minimum sequence size and score for Illumina (Martin, 2011).

Initially using *stampy*-1.0.28 (Lunter & Goodson, 2011) with a flexible substitution rate of 0.05, to accommodate putative DNA base misincorporation, and alignment quality score of 93 to make sure we had accurate alignments, was used to initially map the

Formatted: Indent: First line: 0"

Formatted: Indent: First line: 0.44"

Formatted: Font: Italic

quality filtered reads to the Muridae RefSeq mtDNA genomes. This initial alignment was used to generate summary statistics (Table 2). Following this initial alignment in order to gather as many reads as possible we varied the acceptable substitution rate in *stampy* between: 0.12, 0.08, 0.05, 0.02, 0.008, and 0.001. Next each of these six different alignments were subjected SAMtools 0.1.19 (Li et al., 2009) rmdup to remove PCR duplicates. Next each alignment was merged into a single bam file. To remove additional PCR duplicates and duplicate reads between the different alignments a second round of samtools rmdup was performed followed by bbMAP dedupe.sh (Bushnell, 2015) script to remove additional PCR duplicates and duplicate reads between alignments. Using these combined reads we loaded them as input reads into MITObim 1.8 (Hahn et al., 2013) using the *Rattus rattus* mtDNA genome (RefSeq: NC_012374.1) as a reference guide. MITObim was ran for five iterations until no new bases were added to the assembly.

Formatted: Font: Italic

Formatted: Font: Italic

Following the assembly using MITObim to evaluate the level of DNA base misincorporation we used mapDamage, which reports DNA damage patterns in sequences generated from aDNA using NGS platforms (Ginolhac et al., 2011). To run mapDamage we used the *bwa mem* (v. 0.7.7-r441) algorithm to align our reads used in the assembly back to the resulting assembly from MITObim.

Finally to verify that our assembly assembly was in fact from *R. rattus* we ran *blastn* search via NCBI BLAST tool (Johnson et al., 2008) to verify that the data was of *R. rattus*.

**6. Phylogenetic Analysis**

Our *R. rattus* assembly from MITObim was mapped against the *cyt b* gene from Aplin et al. (2011) (GenBank PopSet: 357089839), via lastal-833 (Frith et al, 2010). Using the alignment of the representative *cyt b* gene we found the positions in our mtDNA that corresponded to *cyt b* and extracted them using *bedtools* (Quinlan & Hall, 2010). This resulted in 954bp from our Tres Bocas *R. rattus* assembly. Finally the Tres Bocas cyt b sequence a long with the *cyt b* sequences from Aplin et al. (2011), were re-aligned using MUSCLE (Edgar, 2004) via CIPRES Science Gateway.

Following the nucleotide alignment PartitionFinder2-2.1.1 (Lanfear et al., 2016) was used to determine which parameters should be used to do a phylogenetic study of the sequences. The set-up for our file in PartitionFinder2 was one of linked branch lengths, in a greedy search, following the BEAST2-2.1 (Drummond & Rambaut, 2007) models on Bayesian Information Criterion (BIC), which is a model with high penalty terms for over-fitting parameters that could be too stringent for the analysis (Schwarz, 1978). The best partitioning parameters were determined as:

**Table 1:** Parameters used for phylogenetic study of *R. rattus* as determined by PartitionFinder2

| **Subset:** | **Model:** |
|---|---|
| 1 | GTR+I+G |
| 2 | GTR+I |
| 3 | GTR+I+G |

The Generalized Time-Reversible (GTR) model is the most general parameter for phylogenetic analysis. GTR model gives the frequency of each base at each site of the sequence. Once the Gamma distribution (+G) and Invariant sites (+I) options are added to address rate heterogeneity over the sites. +G accommodates for a varying degree of rate heterogeneity whereas +I accounts that a proportion of sites are do not vary(Strimmer & Haeseler, 2009). Once these parameters were set, they were used in CIPRES (Miller et al., 2010) to build a phylogenetic tree using RAxML-HPC2 8.2.10 (Stamatakis, 2014). The specifications for running RAxML selected the GTR+G model for the bootstrapping model phase in the Nucleic Acid Options and set the bootstrapping iterations to 1000. The maximum likelihood tree was built using the *cyt b* gene, as was done for the PopSet 357089839 used by Aplin et al. (2011). Once the tree was built it was edited using FigTree for some reorganization of the tree (Rambaut, 2007). For the finishing aesthetic touches, the tree was edited using Adobe Illustrator CC 2018 (22.0).

A second RAxML tree was built with the addition of the D-loop region with PopSet: 260780207 from Tollenaere et al. (2010). The PopSet consisted of the *cyt b*, tRNA-Theonine, tRNA- Proline and the displacement loop (D-loop) from countries that border the Indian Ocean. The sequence was manually edited to contain the same mtDNA gene as in Tollenaere et al. (2010): 954 bp in the *cyt b* gene, 67 bp in tRNA- Theonine, 59 bp in tRNA- Proline and 194 bp in the partial d'loop region. The same protocol was used for this D-loop phylogeny as was done for *cyt b*.

# RESULTS

In this section, I will present the results obtained from the assembly, coverage analysis, annotation and phylogenetic analysis.

**1. Quality Filtering and Assembly**

We obtained 21.4 million raw reads, of which 18.2 million were retained after quality filtering (see Table 2). Post quality filtering, we then aligned an existing mitochondrial genome of *R. rattus* via *stampy*, resulting in 2,395 uniquely mapped reads, which are the reads that map to one location on the reference genome. Once the PCR-duplicated reads were removed, 1,833 reads remained. The clonality was of 23%, a calculation of the lost reads because of PCR contamination, which is obtained by subtracting the duplicate mapped reads from the unique reads divided by the unique reads. The endogenous percent was of 0.01. This metric of endogenous DNA is obtained by dividing the after trimmed reads by the uniquely mapped reads; which represents the proportion of reads that mapped on one location of our *R. rattus* mtDNA genome. The efficiency was of 0.01%, which is the number of duplicates, was calculated by dividing the rmdup reads by the number of raw reads before they have been trimmed.

**Table 2:** Reads from the *Rattus rattus* sample

| **Character:** | **Numbers:** |
|---|---|
| Total | 21,427,688 |
| After trim | 18,165,028 |
| Unique | 2,395 |
| Removed duplicates (rmdup) | 1,833 |
| Clonality % | 23.47 |
| Endogenous % | 0.01 |
| Efficiency % | 0.01 |

**2. Genome assembly coverage analysis results**

Coverage are estimated numbers of DNA sequence reads of the three pooled rat maxillae plotted against nucleotide position in the mitochondrial genome of the mitochondrial genome of *R. rattus* from an owl pellet. Coverage information is crucial to detect copy number variation and structural variation which are two variables that can cause problems such as missing data (Pedersen et al., 2017). The coverage has many 1x-5x peaks throughout the sequence with a notable spike in the coverage of nearly 40x close to the 1000$^{th}$ position which corresponds to the small subunit of the mitochondrial ribosome 12S gene (see Figures 3 and 4). Many of the positions of the genome present coverage of 5x (see Figure 4). A custom script using R samtools library was made to calculate the average coverage of the sequence reads. The average coverage was of 1.372916, which equates to an average of 1x coverage throughout the sequence.

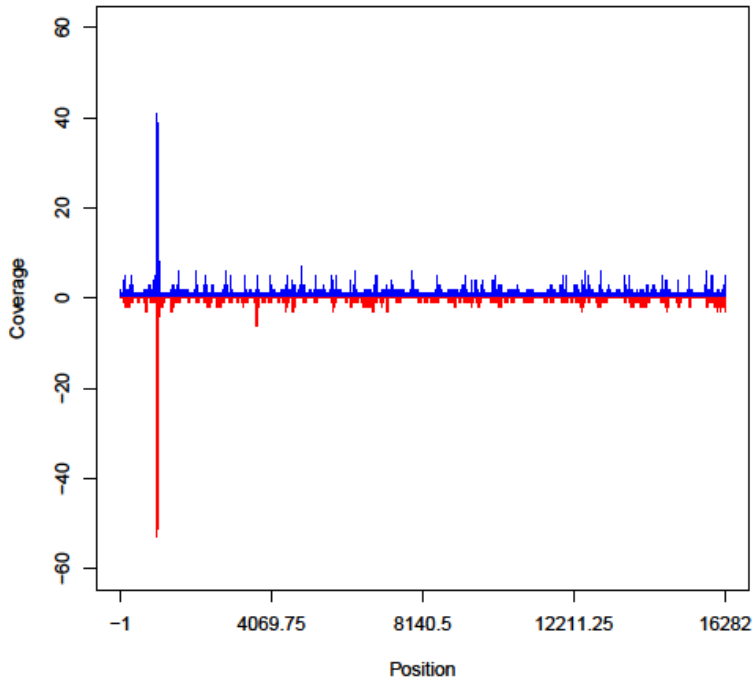


**Figure 3:** Coverage plot of *R. rattus* shows a peak approximately in the 1000$^{th}$ position that corresponds within the mitochondrial 12S ribosomal RNA gene. Reads of the positive strand is in blue and the negative strand is in red.

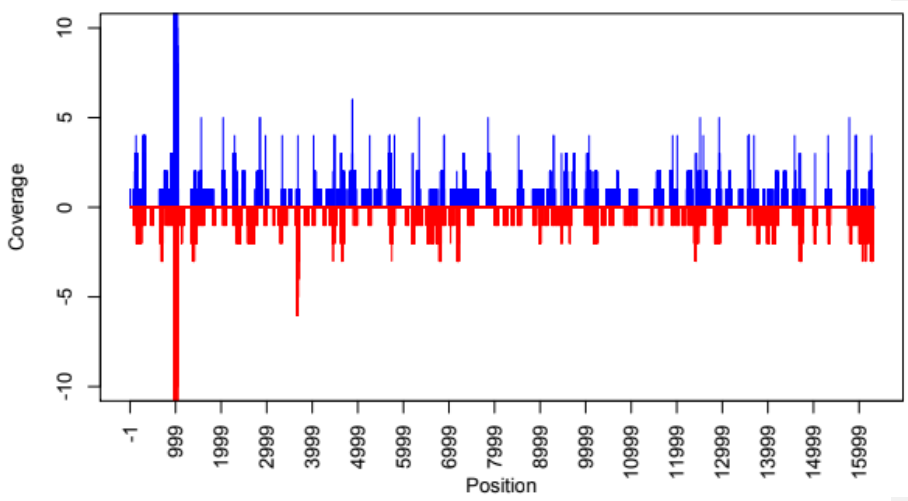


**Figure 4:** Close-up of *R. rattus* coverage plot shows a peak approximately in the 1000$^{th}$ position that corresponds within the mitochondrial 12S ribosomal RNA gene. The positive strand is in blue and the negative strand is in red.

The fragmentation and a lack of misincorporation patterns in the sample were measured after the duplicates were removed from the sample (see Figure 5). The sample has a slight higher frequency of purines (A and G) before the strand break, at the beginning of the sequence, but no peaks can be seen before the strand break in the pyrimidines (C and T). The misincorporation frequencies of the sample in the reads are higher for the pyrimidines than for the purines. In the -1 position there is a higher frequency of guanine than of adenine in the 5' end.

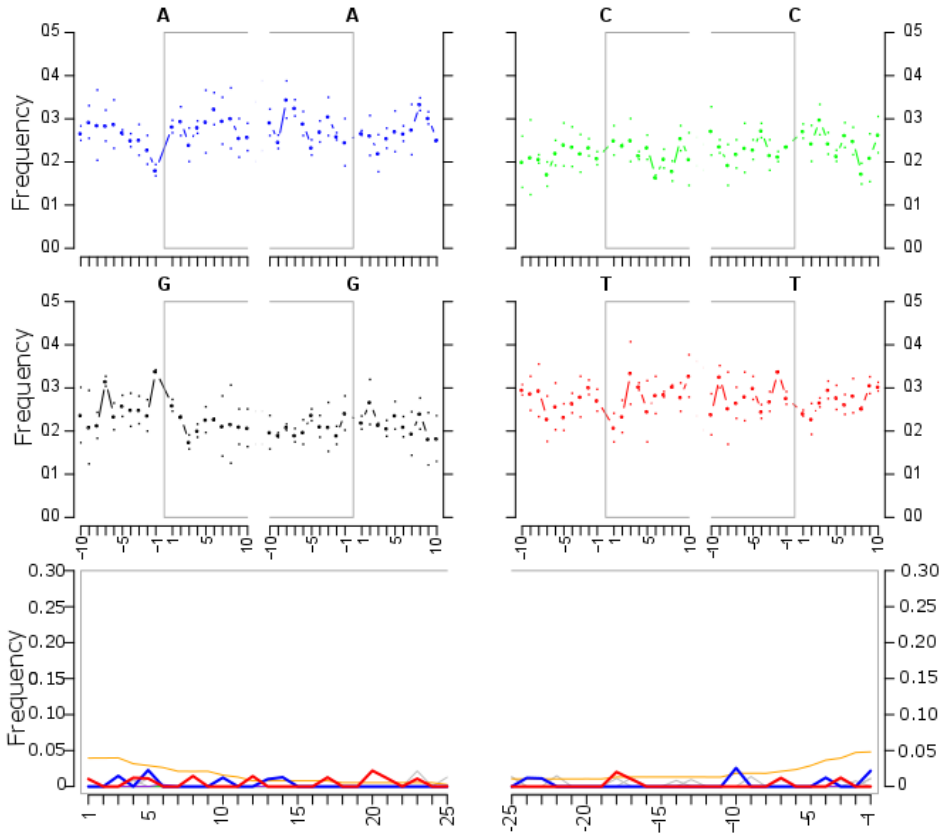


**Figure 5:** Fragmentation and misincorporation patterns plot obtained from mapDamage. This plot was obtained after computing the number of occurrences of mutations divided by the number of occurrences of the reference genome. The four upper plots show the base frequency outside and in the strand break, the open grey box corresponds to the read. In the Y-axis the frequencies are displayed for A, G , C, and T for the bases in 5'and 3'of the read site, which are presented in the X-axis. For the top four plots, the X-axis represents the first and last 10 bases, and in the bottom two plots the first and last 25 bases**.** The bottom plots are the positions' specific substitutions from the 5' (on the left) and the 3' end (on the right). The red line relates to the C-T substitutions, blue relates to the G-A substitutions, grey are all the other substitutions, orange are the soft-clipped bases. Both the red and the blue lines have low frequencies of less than 0.05. The soft-clipped base frequency is higher than the ones for the C->T and A->G misincorporations starting at 0.05 for the first few positions.

The sample's parameters were tested with five different Markov chain Monte Carlo (MCMC) models: Theta, Rho, Delta D Delta S, Lambda and LogLikelihood. The parameters proved to be well calibrated after presenting stationarity (see Figure 6). The acceptance ratios were: Theta 0.24, Rho 0.21, DeltaD 0.23, DeltaS 0.18, Lambda 0.18, and LogLikelihood were 0.69.

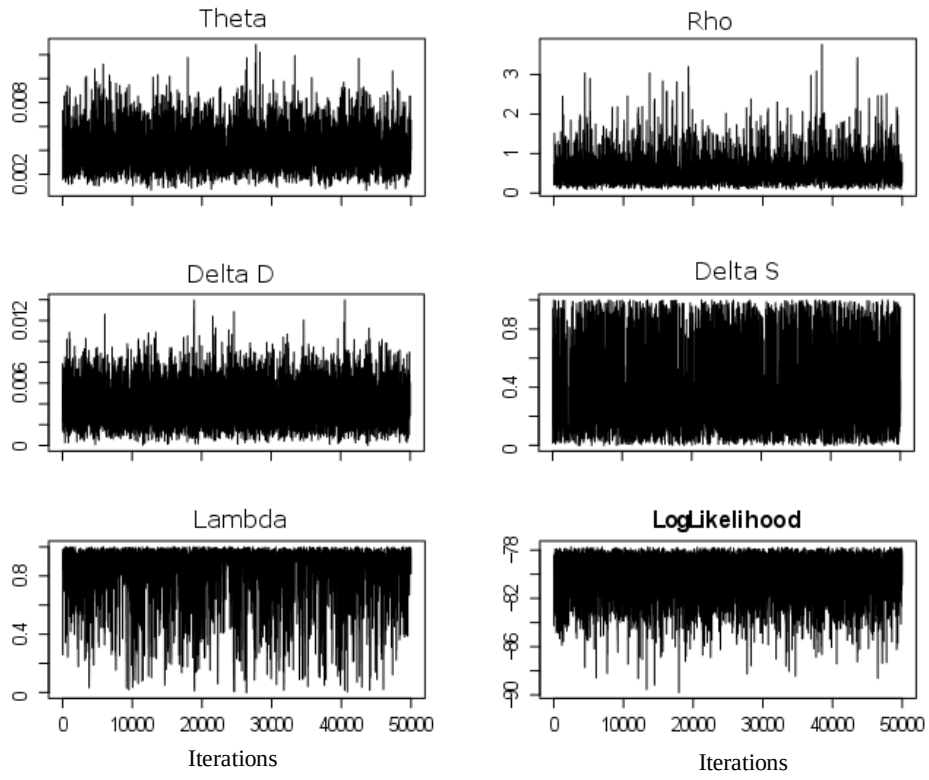


**Figure 6:** Trace graphs statistics of the MCMC iterations of five model parameters: Theta, Rho, Delta D, Delta S, Lambda and LogLikelihood, obtained from mapDamage. They present stationarity on all the models.

### 3. Annotation

The assembled mitochondrial sequence (see Figure 7) obtained from MITOS web server listed all the genes and their positions in the genome, including their length (See Table 3). There were several genes that were not found in the genome such as: atp8, trnA, trnD, trnF, trnI, trnK, trnQ, trnR, trnV. Another peculiarity was that the CO1, nad3 and nad4 genes were split.

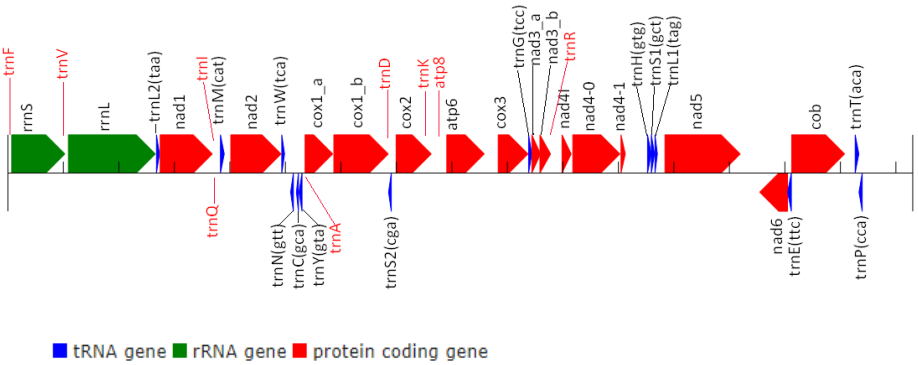


**Figure 7**: Mitochondrial genome diagram produced by MITOS using the raw sample assembly. The protein coding genes are in red, the ribosomal RNA are in green and the transfer ribonucleic acids are in blue. The genes depicted in red letters are in approximate locations of the missing genes according to the *R. rattus* genome from GenBank NC_12374

**Table 3:** Genes present in the raw *R. rattus* sample that was obtained from the owl pellet. The missing genes are in red letters in their approximate locations according to the *R. rattus* genome from GenBank NC_12374

| Gene | Abbreviation | Start | Length | Strand |
|---|---|---|---|---|
| tRNA codon Phenylalanine | trnF | 1 | 67 | + |
| Small ribosomal RNA-12S | rrnS | 70 | 958 | + |
| tRNA codon Valine | trnV | 1026 | 68 | + |
| Large ribosomal RNA-16S | rrnL | 1091 | 1567 | + |
| tRNA codon Leucine 2 | trnL2(taa) | 2678 | 55 | + |
| NADH oxidoreductase chain 1 | nad1 | 2745 | 939 | + |
| tRNA codon Isoleucine | trnI | 3695 | 69 | + |
| tRNA codon Glutamine | trnQ | 3761 | 71 | - |
| tRNA codon Methionine | trnM(cat) | 3831 | 69 | + |
| NADH dehydrogenase chain 2 | nad2 | 4017 | 906 | + |
| tRNA codon Tryptophan | trnW(tca) | 4936 | 51 | + |
| tRNA codon Alanine | trnA | 5010 | 69 | - |
| tRNA codon Asparagine | trnN(gtt) | 5095 | 51 | - |
| tRNA codon Cysteine | trnC(gca) | 5200 | 37 | - |
| tRNA codon Tyrosine | trnY(gta) | 5247 | 55 | - |
| Cytochrome c oxidase subunit 1 | cox1_a | 5356 | 498 | + |
| Cytochrome c oxidase subunit 1 | cox1_b | 5874 | 981 | + |
| tRNA codon Serine | trnS2(cga) | 6860 | 50 | - |
| tRNA codon Aspartic acid | trnD | 6937 | 68 | + |
| Cytochrome c oxidase subunit 2 | cox2 | 6996 | 633 | + |
| tRNA codon Lysine | trnK | 7693 | 64 | + |
| ATP synthase 8 | atp8 | 7758 | 49 | + |
| ATP synthase 6 | atp6 | 7909 | 675 | + |
| Cytochrome c oxidase subunit 3 | cox3 | 8835 | 537 | + |
| tRNA codon Glycine | trnG(tcc) | 9383 | 58 | + |
| NADH dehydrogenase chain 3 | nad3_a | 9441 | 156 | + |
| NADH dehydrogenase chain 3 | nad3_b | 9587 | 192 | + |
| tRNA codon Arginine | trnR | 9800 | 67 | + |
| NADH oxidoreductase chain 4L | nad4L | 9985 | 165 | + |
| NADH oxidoreductase chain 4 | nad4-0 | 10179 | 165 | + |
| NADH oxidoreductase chain 4 | nad4-1 | 11048 | 78 | + |
| tRNA codon Histidine | trnH(gtg) | 11526 | 68 | + |
| tRNA codon Serine | trnS1(gct) | 11594 | 59 | + |
| tRNA codon Leucine | trnL1(tag) | 11660 | 44 | + |
| NADH oxidoreductase chain 5 | nad5 | 11840 | 1356 | + |
| NADH oxidoreductase chain 6 | nad6 | 13548 | 501 | - |
| tRNA codon Glutamic acid | trnE(ttc) | 14049 | 69 | - |
| Cytochrome b | cob | 14123 | 954 | + |
| tRNA codon Theonine | trnT(aca) | 15267 | 67 | + |
| tRNA codon Proline | trnP(cca) | 15336 | 59 | - |

**4. Phylogenetic Results**

Phylogenetic analysis of *R. rattus* mitochondrial genomes, using *cyt b*, revealed the relationship between the clades (see Figure 8). Clade A presents high bootstrap support as its recent ancestor has 99% support and the clades themselves have 99% and 100% support, respectively. The clades have been separated into Clade A, subclades A-I-VI, and subclades B-IV-VI belong in Clade B, according to the taxa on the tree. The Caribbean sample from Tres Bocas Cave was nested within the Madagascar cluster in Clade A-I. Clade A-I is comprised partly of samples obtained from India, Madagascar, Europe, America, Australia and several Pacific Islands. Clade A-II occurs widely in Indonesia, Myanmar, including the Philippines, Japan, Papua New Guinea and South Africa. Clade A-III includes Pakistan and Nepal. Clade B-IV has overlaps with Clade A-II and has many of the same distributions. Clade B-V has only samples from Thailand and Laos while Clade B-VI has samples from Java and Borneo. The rest of the clades represent *Rattus losea*, *R. pyctoris*, *R. argentiventer*, *R. exulans* and *R. andamanensis*.

In a second phylogenetic tree (see Figure 10), the Caribbean sample was located within the Madagascar and Ethiopian *R. rattus* clade and not sister to the rest of the East African-Malagasy *R. rattus*. The first tree used the *cyt b* gene, whereas the second tree used the displacement loop (D-loop) region.

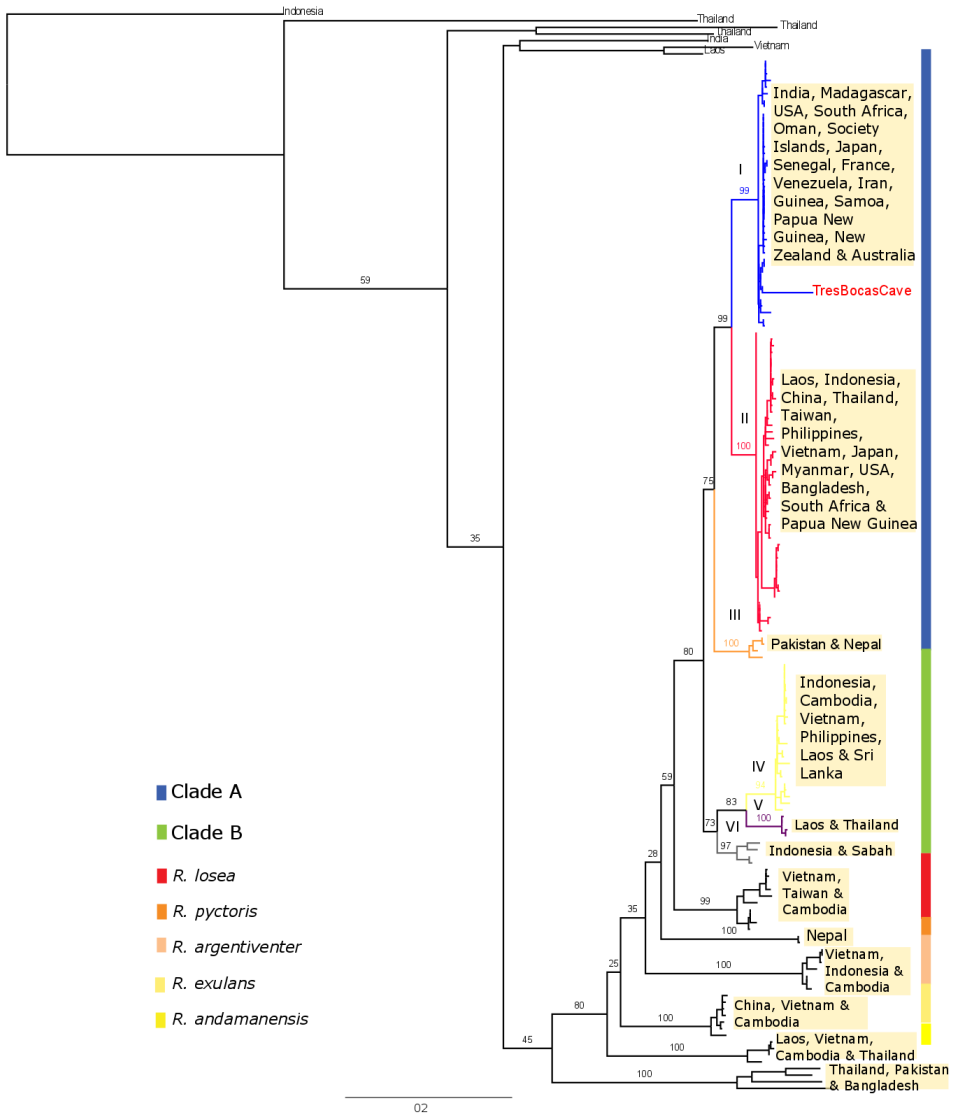


**Figure 8:** Maximum Likelihood tree using *cyt b* gene for alignment with bootstrap support of the relationship of the Tres Bocas Cave sample; marked with red letters in subclade I within Clade A. In the beige boxes next to the branches are the countries of the samples organized from the most recurring location to the least.

The phylogenetic tree's clades have been colored for convenience to resemble the phylogeny in Aplin et al. (2011). The Caribbean sample is nested within Subclade I in Clade A.

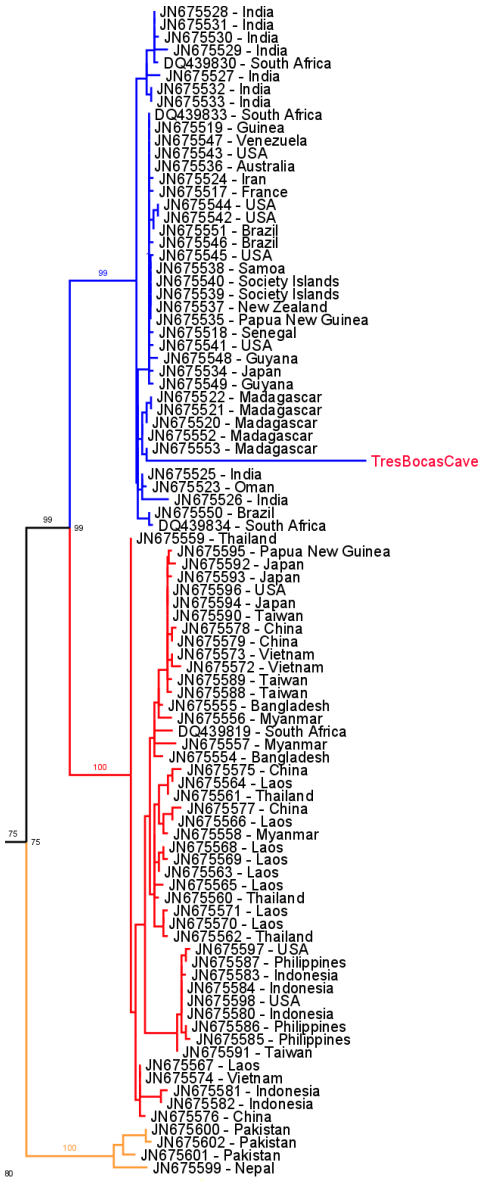


**Figure 9:** Maximum Likelihood tree zoom of Clade A. The Tres Bocas Cave sample is marked in red letters. The beige boxes next to the branches are the countries of the samples organized from the most recurring location to the least. Tree built with RAxML.

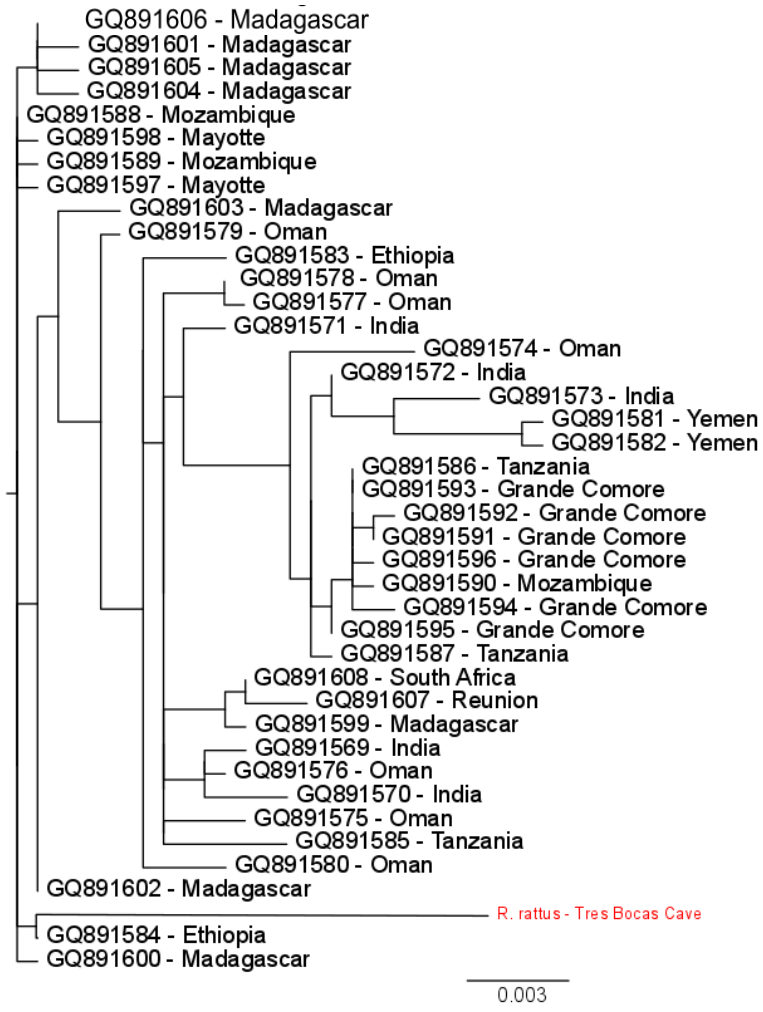


**Figure 10:** Phylogenetic tree using cyt b, tRNA- Theonine, tRNA- Proline and the displacement loop region of the sample with the PopSet 260780207. The Tres Bocas Cave sample, in red, nested in with the Ethiopian and Madagascar clades.

**5. Radiocarbon dating results**

The D14 fraction dating revealed the sample dates ~1961-1978 AD. The fraction modern is 1.3079, which is the unit of $^{14}C$ measurement, with an uncertainty of 0.0047.

## DISCUSSION

In this section, I will discuss the results from the previous section in greater detail, providing definitions for terms presented but not explained in the results and, also, using references as tools to corroborate my findings.

### 1. Overview

In this study the mitochondrial genome from *R. rattus* bones found within an owl pellet dating from 1961-1978 were reconstructed. This demonstrates that DNA from 60-year old samples collected from the Caribbean, known for its warm temperatures and humidity, which makes the conservation of DNA difficult, can indeed be sequenced. A phylogenetic analysis also revealed the sample is nested in a Madagascar clade.

### 2. Assembly and annotation

The mitochondrial genome (see Figure 7 and Table 3) of the sample from Tres Bocas Cave, in the Dominican Republic, is missing a number of genes which could be due to degradation. Another reason for the missing genes could be the coverage was too low during the sequencing and these genes were not recognized, which impeded their assembly (Briggs et al., 2007; Paabo, 1989).

A large number of PCR duplicates were removed due to the degradation in the sample (see Table 2). The PCR duplicates that were removed are also reflected by the quality control values of the clonality, endogenous and efficiency percentages. Each one indicates that the 1,833 reads that were left after the clean-up process are of acceptable

quality. The 23% clonality represents the PCR clones that were present during the sequencing step of the methods, which gave false positives for the presence of reads. The endogenous percentage of 0.01 is due to degradation and the need for PCR amplification (Seguin-Orlando et al., 2015). This low endogenous percentage is similar to other previously observed sequences such as the *Myotragus* cave goat, lynx and Neanderthal, which all presented a <1 endogenous percent (Gigli et al., 2011). The efficiency percentage of 0.01 is low because the number of duplicates was high, but was still enough to reconstruct the mitochondrial genome and reveal its ancestral population.

**3. Bioinformatics analysis**

Coverage plots have been largely used to illustrate the number of correct bases at the correct positions; as in, the amount of times each base in the sequence align with each base of the reference sequence. In figures 3 and 4, the coverage plot depicts an average of 1x coverage throughout the whole mitochondrial genome, many 5x peaks and a specific peak of about 40x coverage at about the $1000^{th}$ position. This peak is within the mitochondrially encoded 12S ribosomal RNA gene and is due to a tandem repeat region in the gene. Although aDNA coverage tends to be poor and less than 1x, the coverage in the *R. rattus* sample is an on average of 1x (Le & Durbin, 2011; Li et al., 2011; Pasaniuc et al., 2012).

The misincorporations in our sample are lower than what is characteristic of aDNA, as aDNA tends to have over representation of purines at the 5' end and pyrimidines towards the 3' end, indicating that the Dominican Republic sample contains modern

DNA. Misincorporations in the complementary bases account for the majority of potential errors in aDNA (Stiller et al., 2006). A comparison in the complementary bases to represent the misincorporations of our sample shows that the adenine plot (see Figure 5) has a peak before strand breaks which matches the peak in the guanine plot. This higher frequency of purines is typically seen in aDNA samples (Briggs et al., 2007; Brotherton et al., 2007; Matisoo-Smith & Allen, 2001; Sawyer et al., 2012; Weiß et al., 2016). It has been suggested that this higher frequency of purines before strand breaks in 5' is due to the depurination that contributes to the degradation process which is one of the major fragmentation processes in DNA (Briggs et al., 2007). Adenine residues predominate over guanine residues in samples younger than 100 years old; both purines predominate in samples from 500-2,000 years old and in Neanderthal samples that are 40,000 years old Guanine is predominant (Sawyer et al., 2012). Higher frequency of purines attributed to depurination has been observed in aDNA from Neanderthal, humans, mammoths, herbarium samples, cave goats, lynxes, and cave bears (Briggs et al., 2007; Gigli et al., 2011; Krause et al., 2010; Weiß et al., 2016). In contrast, others have stated that the higher frequency of purines before the strand breaks must be due to a chemical mechanism other than depurination (Sawyer et al., 2012), while others say that it could be the result of contaminant sequences being fragmented by bacterial enzymes, and also that bleach treatment of the bones could convey ancient characteristics (Gigli et al., 2011). These studies conclude that more research is necessary in the biochemical process of the depurination and fragmentation patterns because they are still not fully understood.

The C-T and the G-A substitutions are at a low frequency well below 0.05, but the soft-clip frequency is higher towards the positions 1-5 and also the -1-5 positions at about 0.05 (see Figure 5, in the bottom two plots). Normally a high frequency of soft-clipped bases indicates that the bases should be trimmed; the frequency in the sample is not high enough to need that. To corroborate, a second figure was produced which presents low level of misincorporation at both ends of the sample (see Appendix I). These figures show that the frequencies of misincorporation and fragmentation in the sample are too low with what is typical from aDNA, which indicates the sample is not ancient enough to be considered ancient (Briggs et al., 2007; Gigli et al., 2011; Gutiérrez-García et al., 2014; Krause et al., 2010; Mendisco et al., 2015; Paabo, 1989; Sawyer et al., 2012; Stiller et al., 2006; Weiß et al., 2016) These low misincorporation frequencies are in accordance with the results from the D14 fraction which estimated the sample to be from the 1960's-70's, which is considered modern DNA and will not present the amount of misincorporations that are customary of ancient samples (Briggs et al., 2007; Gigli et al., 2011; Krause et al., 2010; Weiß et al., 2016).

Length distribution of single-end reads per strand and cumulative frequencies of C-T at the 5'-end and G-A at the 3'-end per strand (see Appendix II) corroborate the parameters set for a minimum of 30 bp and the automatic 100 bp maximum. The frequency plots of misincorporation show that the lack of a smooth curve means we obtained too few reads. With higher number of sequences, a smoother curve can be achieved (Schuenemann et al., 2011).

The trace plots (see Figure 6) present that the sample has converged. That means that the sequences have mixed well and have reached stationarity with the correct variance. This is an indication that the MCMC is efficiently sampling from a maximum in the underlying distribution. Trace plots show us the history of a parameter value across the iterations at the chain. This shows where the chain has been exploring. The acceptance rates should be around 23% for each model, which was close to what was obtained except for the DeltaS and Lambda models, which both were 18%. Lower acceptance rates usually mean that the convergence was slower due to the suggested proposals by the models that could sample the chain in regions of lower probability density.

**4. Phylogenetic analysis**

The sample obtained from the Tres Bocas Cave is nested within the Madagascar-East African subclade A-I. As can be seen in Figure 8, the clade is well supported by a bootstrap value of 99. To verify that our sample belongs to the Madagascar- East African clade, we built a second Maximum Likelihood tree (see Figure 10) in which the sample was nested within the Madagascar taxa and also an Ethiopia taxon, which means that the *R. rattus* from Tres Bocas Cave belongs within the East African clade. Clade A-I, where the sample is nested, is a monophyletic *R. rattus* clade; whereas, Clade B includes *R. losea* from Thailand and Laos.

Aplin et al. (2011) stated that *R. rattus* had two main groups, one from Europe and Middle-East (EME) and the other from India. These two groups were the potential

sources for the dispersion of the rat, with India being the ancestral haplotype from which all the lineages derive. A third, smaller group was described as a Madagascar clade which has a common haplotype with the EME group but also is linked with the Indian group through a different haplotype. The *R. rattus* sample from the Tres Bocas Cave nested within the Clade A-I, which is part of the EME group. The sample being nested in Clade A-I was expected from the historical introduction of rats to the Caribbean in the 16$^{th}$ century through the Age of Exploration. However, it was not expected to be nested in the Madagascar clade.

Madagascar has had a long history of immigration and the entire Indian Ocean was part of a trading network that connected China and India with the Mediterranean (Tollenaere et al., 2010). These immigration events introduced the black rat from its origins in southern India to Africa (Tollenaere et al., 2010) and, as trading continued, *R. rattus* was introduced to the rest of the world (Aplin et al., 2011). Other studies have not investigated the geographic origins of *R. rattus* in the Caribbean. Here we can conclude that some of the *R. rattus* in the Caribbean have origins from East Africa. The paper by Aplin et al. (2011) concluded that rats from Europe and the Middle-East populated the western hemisphere based on 164 samples collected from 76 localities in 32 countries. Here we indicate that East African rats were introduced as well. Multiple origins of *R. rattus* in the western hemisphere are reasonable to conclude from our results.

Many of the branches ended in polytomies. This indicates that these branches are lacking samples to resolve these relationships (Coddington & Scharff, 1996, Purvis &

Garland, 1993). Heavier sampling and the inclusion of fossils could make the topology of these relationships robust and possibly resolve the polytomies (Rothwell et al., 2018).

## CONCLUSION

*Rattus rattus* bones obtained from an owl pellet collected from Tres Bocas Cave in the Dominican Republic provided a mitochondrial genome. This genome was subjected to aDNA damage analysis, which determined there were minimal misincorporations within the sample. The *R. rattus* sample dated to the 1960's-70's. The maximum likelihood tree from the phylogenetic analysis using *cyt b* revealed that the *R. rattus* sample nested within a monophyletic clade of rats and is sister taxa to a Madagascar clade. The second maximum likelihood tree, which used the displacement loop region, revealed the sample was sister taxa with an Ethiopia clade. Both trees presented that the mtDNA of the sample nests in East African clades.

For future recommendations, I suggest heavier geographic sampling of *R. rattus* in order to add more data to this research. The largest amount of samples collected was from Asia, but the Americas were insufficiently sampled and the Caribbean had only one collection spot, including present work. Secondly, a more robust phylogenetic analysis by using more genes would improve any kind of population demographic reconstruction. Lastly, more studies of owl pellets in the Caribbean and subfossils from tropical regions are promising for future investigations of small mammals.

## LITERATURE CITED


Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aplin, K. P., & Singleton, G. R. (2003). Balancing rodent management and small mammal conservation in agricultural landscapes : challenges for the present and the future. In Grant R. Singleton, L. A. Hinds, C. J. Krebs, & D. M. Spratt (Eds.), *Rats , mice and people : rodent biology and management* (pp. 80–88). Canberra: Australian Center for International Agricultural Research.

Aplin, K. P., Stephen, C., Suzuki, H., Chinen, A. A., Chesser, R. T., Austin, J., … Herbreteau, V. (2011). Multiple Geographic Origins of Commensalism and Complex Dispersal History of Black Rats, *6*(11). https://doi.org/10.1371/journal.pone.0026357

Avise, J. C., & Aquadro, C. F. (1982). A comparative summary of genetic distances in the vertebrates - patterns and correlations. *Evolutionary Biology*, *15*(11), 151–185.

Backhans, A., Jacobson, M., Hansson, I., Lebbad, M., Lambertz, S. T., Gammelgård, E., … Fellström, C. (2012). Occurrence of pathogens in wild rodents caught on Swedish pig and chicken farms. *Epidemiology and Infection*, 1–7. https://doi.org/10.1017/S0950268812002609

Bálint, M., Nowak, C., Márton, O., Pauls, S. U., Wittwer, C., Aramayo, J. L., … Jansen, M. (2018). Accuracy, limitations and cost efficiency of eDNA-based community survey in tropical frogs. *Molecular Ecology Resources*, *0*(0). https://doi.org/doi:10.1111/1755-0998.12934

Bandelt, H., Hofreiter, M., Barnes, I., & Gilbert, M. T. P. (2005). Assessing ancient DNA studies Assessing ancient DNA studies, (November). https://doi.org/10.1016/j.tree.2005.07.005

Bobola, E., Goutner, V., & Liordos, V. (2018). Foraging habitat selection and differentiation among coexisting raptors across an estuarine landscape (Evros Delta, northern Greece). *Estuarine, Coastal and Shelf Science*, *213*, 108–114. https://doi.org/https://doi.org/10.1016/j.ecss.2018.08.009

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., … de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, *29*(6), 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Boom, R., Sol, C., Salimans, M. M. M., Jansen, C. L., Wertheim-van Dillen, P., & van der Noordaa, J. (1990). Rapid and Simple Method for Purification of Nucleic Acids. *Journal of Clinical Microbiology*, *28*(3), 495–503. https://doi.org/10.2116/analsci.25.941

Bressan, E. A., Rossi, M. L., Gerald, L. T., & Figueira, A. (2014). Extraction of high-quality DNA from ethanol-preserved tropical plant tissues. *BMC Research Notes*, *7*(1), 1–6. https://doi.org/10.1186/1756-0500-7-268

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prufer, K., … Paabo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences*, *104*(37), 14616–14621. https://doi.org/10.1073/pnas.0704665104

Brotherton, P., Endicott, P., Sanchez, J. J., Beaumont, M., Barnett, R., Austin, J., & Cooper, A. (2007). Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Research*, *35*(17), 5717–5728. https://doi.org/10.1093/nar/gkm588

Brown, W. M., George, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA, *76*(4), 1967–1971.

Bushnell, B. (2015). *BBMap short-read aligner, and other bioinformatics tools*.

Castresana, J. (2001). Cytochrome *b* phylogeny and the taxonomy of great apes and mammals. *Molecular Biology and Evolution*, *18*(4), 465–471. https://doi.org/10.1093/oxfordjournals.molbev.a003825

Coddington, J. A., & Scharff, N. (2018). Problems with " soft " polytomies, (June 1996). https://doi.org/10.1111/j.1096-0031.1996.tb00198.x

Comay, O., & Dayan, T. (2018). Taphonomic signatures of owls: New insights into micromammal assemblages. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *492*, 81–91. https://doi.org/https://doi.org/10.1016/j.palaeo.2017.12.014

Damgaard, P. B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., & Allentoft, M. E. (2015). Improving access to endogenous DNA in ancient bones and teeth. *Scientific Reports*, *5*, 11184. https://doi.org/10.1038/srep11184

Der Sarkissian, C., Allentoft, M. E., Ávila-Arcos, M. C., Barnett, R., Campos, P. F., Cappellini, E., … Orlando, L. (2015). Ancient genomics. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1660), 20130387. https://doi.org/10.1098/rstb.2013.0387

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, *7*(1), 1–8. https://doi.org/10.1186/1471-2148-7-214

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Ewer, R. (1971). *The Biology and Behaviour of a Free-Living Population of Black Rats (Rattus rattus). Animal Behaviour Monographs* (Vol. 4). https://doi.org/10.1016/S0066-1856(71)80002-X

Frith, M. C., Wan, R., & Horton, P. (2010). Incorporating sequence quality data into alignment improves DNA read mapping, *38*(7). https://doi.org/10.1093/nar/gkq010

Fumagalli, L., Taberlet, P., Favre, L., & Hausser, J. (1996). *Fumagalli L, Taberlet P, Favre L, Hausser J. Origin and evolution of homologous sequences in the mitochondrial DNA control region of shrews. Mol Biol Evol 13: 31-46. Molecular biology and evolution* (Vol. 13). https://doi.org/10.1093/oxfordjournals.molbev.a025568

Galan, M., Pagès, M., & Cosson, J. F. (2012). Next-Generation Sequencing for Rodent

Barcoding: Species Identification from Fresh, Degraded and Environmental Samples. *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0048374
Gigli, E., Sanchez-quinto, F., Ramirez, O., Calafell, F., & Garcia, M. (2011). Fragmentation of Contaminant and Endogenous DNA in Ancient Samples Determined by Shotgun Sequencing ; Prospects for Human Palaeogenomics, *6*(8). https://doi.org/10.1371/journal.pone.0024161
Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., & Orlando, L. (2011). mapDamage: Testing for damage patterns in ancient DNA sequences. *Bioinformatics*, *27*(15), 2153–2155. https://doi.org/10.1093/bioinformatics/btr347
Gutiérrez-García, T. A., Vázquez-domínguez, E., Arroyo-cabrales, J., Kuch, M., King, C., Poinar, H. N., … Enk, J. (2014). Ancient DNA and the tropics : a rodent ' s tale. *Biology Letters*, (June).
Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - A baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13). https://doi.org/10.1093/nar/gkt371
Hänni, C., Brousseau, T., Laudet, V., & Stehelin, D. (1995). Isopropanol precipitation removes PCR inhibitors from ancient bone extracts. *Nucleic Acids Research*, *23*(5), 881–882. https://doi.org/10.1093/nar/23.5.881
Hansen, A. J., Mitchell, D. L., Wiuf, C., Paniker, L., Brand, T. B., Binladen, J., … Willerslev, E. (2006). Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics*, *173*(2), 1175–1179. https://doi.org/10.1534/genetics.106.057349
Hingston, M., Matisoo-smith, L., & Ross, H. A. (2007). Identifying Rattus species using mitochondrial DNA Article in Molecular Ecology Notes · September 2007 Identifying Rattus species using mitochondrial DNA, (May 2014). https://doi.org/10.1111/j.1471-8286.2007.01752.x
Hole, W., Hole, W., & Building, P. (2001). Converting ams data to radiocarbon values: considerations and conventions. McNichol, *43*(2), 313–320.
Holmes, D. S., & Bonner, J. (1973). Preparation, molecular weight, base composition, and secondary structure of giant nuclear ribonucleic acid. *Biochemistry*, *12*(12), 2330–2338.
Horváth, A., Morvai, A., & Horváth, G. F. (2018). Food-niche pattern of the Barn Owl (Tyto alba) in intensively cultivated agricultural landscape×. *Ornis Hungarica*, *26*(1), 27–40. https://doi.org/10.1515/orhu-2018-0002
Höss, M., & Pääbo, S. (1993). DNA extraction from pleistocene bones by a silica-based purification method. *Nucleic Acids Research*, *21*(16), 3913–3914. https://doi.org/10.1093/nar/21.16.3913
Irwin, D. M., Kocher, T. D., & Wilson, A. C. (1991). Evolution of the cytochromeb gene of mammals. *Journal of Molecular Evolution*, *32*(2), 128–144. https://doi.org/10.1007/BF02515385
Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, *36*(Web

Server issue).
Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193
Kalmar, T. (2000). A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. *Nucleic Acids Research*, *28*(12), 67e–67. https://doi.org/10.1093/nar/28.12.e67
Kehlmaier, C., Barlow, A., Hastings, A. K., Vamberger, M., Paijmans, J. L. A., Steadman, D. W., … Fritz, U. (2017). Tropical ancient DNA reveals relationships of the extinct bahamian giant tortoise Chelonoidis alburyorum. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1846). https://doi.org/10.1098/rspb.2016.2235
Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers, *86*(August), 6196–6200.
Koma, T., Yoshimatsu, K., Yasuda, S. P., Li, T., Amada, T., Shimizu, K., … Arikawa, J. (2013). A survey of rodent-borne pathogens carried by wild Rattus spp. in Northern Vietnam. *Epidemiology and Infection*, *141*(09), 1876–1884. https://doi.org/10.1017/s0950268812002385
Krause, J., Briggs, A. W., Kircher, M., Maricic, T., Zwyns, N., & Derevianko, A. (2010). Report A Complete mtDNA Genome of an Early Modern Human from Kostenki , Russia, 231–236. https://doi.org/10.1016/j.cub.2009.11.068
Kuzmina, M. L., Braukmann, T. W. A., & Zakharov, E. V. (2018). Finding the pond through the weeds: eDNA reveals underestimated diversity of pondweeds. *Applications in Plant Sciences*, *6*(5), 1–8. https://doi.org/10.1002/aps3.1155
Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*. https://doi.org/dx.doi.org/10.1093/molbev/msw260
Le, S. Q., & Durbin, R. (2011). SNP detection and genotyping from low coverage sequencing data on multiple diploid samples. *Genome Research*, *21*, 952–960. https://doi.org/10.1101/gr.113084.110
Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Li, Y., Sidore, C., Kang, H. M., Boehnke, M., & Abecasis, G. R. (2011). Low-coverage sequencing: implications for design of complex trait association studies. *Genome …*, *21*, 940–951. https://doi.org/10.1101/gr.117259.110.individuals
Lopes, C. M., Sasso, T., Valentini, A., Dejean, T., Martins, M., Zamudio, K. R., & Haddad, C. F. B. (2017). eDNA metabarcoding: a promising method for anuran

surveys in highly diverse tropical forests. *Molecular Ecology Resources*, *17*(5), 904–914. https://doi.org/doi:10.1111/1755-0998.12643
Lunter, G., & Goodson, M. (2011). Stampy : A statistical algorithm for sensitive and fast mapping of Illumina sequence reads, (3), 936–939. https://doi.org/10.1101/gr.111120.110.tions
Malhotra, R., & Singla, N. (2018). Analysis of regurgitated pellets of Spotted Owlet Athene brama (Temminck, 1821) (Aves: Strigiformes: Strigidae) from Punjab, India. *Journal of Threatened Taxa*, *10*(6 SE-Communications). https://doi.org/10.11609/jott.4020.10.6.11717-11724
Mandal, S. De, Chhakchhuak, L., & Gurusubramanian, G. (2014). Mitochondrial markers for identification and phylogenetic studies in insects – A Review, (April). https://doi.org/10.2478/dna-2014-0001
Martin, J. M., Raid, R. N., & Branch, L. C. (2009). Barn Owl (Tyto alba). *University of Florida*, *IFAS Extension*, *185*(April), 1–3.
Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/http://dx.doi.org/10.14806/ej.17.1.200
Matisoo-Smith, E., & Allen, J. S. (2001). Name that rat: Molecular and morphological identification of Pacific rodent remains. *International Journal of Osteoarchaeology*, *11*(1), 34–42. https://doi.org/10.1002/oa.544
Mendisco, F., Pemonge, M. H., Leblay, E., Romon, T., Richard, G., Courtaud, P., & Deguilloux, M. F. (2015). Where are the Caribs? Ancient DNA from ceramic period human remains in the Lesser Antilles. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1660), 1–8. https://doi.org/10.1098/rstb.2013.0388
Merkes, C. M., McCalla, S. G., Jensen, N. R., Gaikowski, M. P., & Amberg, J. J. (2014). Persistence of DNA in carcasses, slime and avian feces may affect interpretation of environmental DNA data. *PLoS ONE*, *9*(11). https://doi.org/10.1371/journal.pone.0113346
Meyer, A. (1994). *Shortcomings of the cytochrome b gene as a molecular marker*.
Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees.
Morozova, O., & Marra, M. A. (2008). Genomics Applications of next-generation sequencing technologies in functional genomics. *Genomics*, *92*, 255–264. https://doi.org/10.1016/j.ygeno.2008.07.001
Moysi, M., Christou, M., Goutner, V., Kassinis, N., & Iezekiel, S. (2018). Spatial and temporal patterns in the diet of barn owl ( Tyto alba ) in Cyprus. *Journal of Biological Research-Thessaloniki*, 1–8. https://doi.org/10.1186/s40709-018-0080-8
Muraji, M., Kawasaki, K., & Shimizu, T. (2000). Phylogenetic utility of nucleotide sequences of mitochondrial 16S ribosomal RNA and cytochrome b genes in anthocorid bugs (Heteroptea: Anthocoridae). *Appl. Entomol. Zool.* https://doi.org/10.1303/aez.2000.301
Nevers, M. B., Byappanahalli, M. N., Morris, C. C., Shively, D., Przybyla-kelly, K.,

Spoljaric, A. M., … Roseman, E. F. (2018). Environmental DNA ( eDNA ): A tool for quantifying the abundant but elusive round goby ( Neogobius melanostomus ). *PloS One*, 1–22. https://doi.org/10.5066/F7GH9H6F.Funding

Nicolas, V., Schaeffer, B., Missoup, A. D., Kennis, J., Colyn, M., Denys, C., … Laredo, C. (2012). Assessment of Three Mitochondrial Genes ( 16S , Cytb , CO1 ) for Identifying Species in the Praomyini Tribe ( Rodentia : Muridae ), *7*(5), 1–11. https://doi.org/10.1371/journal.pone.0036586

Ogada, D. (2018). Insights into the natural history of the little known maned rat Lophiomys imhausi through examination of owl pellets and prey remains. *Journal of East African Natural History*, *107*(1), 1–7.

Ondrejicka, D. A., Locke, S. A., Morey, K., Borisenko, A. V., & Hanner, R. H. (2014). Status and prospects of DNA barcoding in medically important parasites and vectors. *Trends in Parasitology*, *30*(12), 582–591. https://doi.org/10.1016/j.pt.2014.09.003

Paabo, S. (1989). Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(6), 1939–1943. https://doi.org/10.1073/pnas.86.6.1939

Pasaniuc, B., Rohland, N., McLaren, P. J., Garimella, K., Zaitlen, N., Li, H., … Price, A. L. (2012). Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nature Genetics*, *44*, 631. Retrieved from http://dx.doi.org/10.1038/ng.2283

Pedersen, B. S., Collins, R. L., Talkowski, M. E., & Quinlan, A. R. (2017). Indexcov : fast coverage quality control for whole-genome sequencing, (September), 1–6. https://doi.org/10.1093/gigascience/gix090

Poulakakis, N., Lymberakis, P., Paragamian, K., & Mylonas, M. (2005). Isolation and amplification of shrew DNA from barn owl pellets. *Biological Journal of the Linnean Society*, *85*(3), 331–340. https://doi.org/10.1111/j.1095-8312.2005.00500.x

Purvis, A., & Garland, T. J. (1993). Society of Systematic Biologists Polytomies in Comparative Analyses of Continuous Characters. *Society of Systematic Biologists*, *42*(4), 569–575.

Reed, C. I., & Reed, B. P. (1928). The mechanism of pellet formation in the great horned owl (*Bubo virginianus*). *Science*, *68*(1763), 359 LP-360. Retrieved from http://science.sciencemag.org/content/68/1763/359.abstract

Robson, H. L. A., Noble, T. H., Saunders, R. J., Robson, S. K. A., Burrows, D. W., & Jerry, D. R. (2016). Fine-tuning for the tropics: application of eDNA technology for invasive fish detection in tropical freshwater ecosystems. *Molecular Ecology Resources*, *16*(4), 922–932. https://doi.org/doi:10.1111/1755-0998.12505

Rohland, N., & Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. *Nature Protocols*, *2*(7), 1756–1762. https://doi.org/10.1038/nprot.2007.247

Rothwell, G. W., Escapa, I. H., & Tomescu, A. M. F. (2018). Tree of death : The role of fossils in resolving the overall pattern of plant phylogeny, *105*(1966), 1239–1242. https://doi.org/10.1002/ajb2.1138

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA.

*PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0034131
Schuenemann, V. J., Bos, K., DeWitte, S., Schmedes, S., Jamieson, J., Mittnik, A., … Poinar, H. N. (2011). Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. *Proceedings of the National Academy of Sciences*, *108*(38), E746–E752. https://doi.org/10.1073/pnas.1105107108
Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136
Seguin-Orlando, A., Gamba, C., Sarkissian, C. Der, Ermini, L., Louvel, G., Boulygina, E., … Orlando, L. (2015). Pros and cons of methylation-based enrichment methods for ancient DNA. *Scientific Reports*, *5*, 1–15. https://doi.org/10.1038/srep11826
Sharikov, A., Kovinka, T., & Bragin, M. (2018). Brief report A comparative laboratory study of the preservation of different rodent bones in pellets of Strigiformes, (March).
Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033
Stiller, M., Green, R. E., Ronan, M., Simons, J. F., Du, L., He, W., … Pääbo, S. (2006). Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *PNAS*, *103*(37), 13578–13584. https://doi.org/https://doi.org/10.1073/pnas.0605327103
Strimmer, K., & Haeseler, A. von. (2009). Genetic distances and nucleotide substitution models. In P. Lemey, M. Salemi, & A.-M. Vandamme (Eds.), *The phylogenetic Hanbook: a Practical Approach to Phylogenetic Analysis and Hpothesis Testing* (2nd ed., pp. 111–141). Cambridge: Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511819049
Stuiver, M., Reimer, P.J., and Reimer, R.W., 2018, CALIB 7.1 [WWW program] at http://calib.org, accessed 2018-12-10
Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, *21*(8), 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x
Terkel, J. (1996). Chapter 2 - Cultural Transmission of Feeding Behavior in the Black Rat (Rattus rattus). In C. M. Heyes & B. G. Galef (Eds.) (pp. 17–47). San Diego: Academic Press. https://doi.org/https://doi.org/10.1016/B978-012273965-1/50003-0
Tollenaere, C., Brouat, C., Duplantier, J. M., Rahalison, L., Rahelinirina, S., Pascal, M., … Cosson, J. F. (2010). Phylogeography of the introduced species Rattus rattus in the western Indian Ocean, with special emphasis on the colonization history of Madagascar. *Journal of Biogeography*, *37*(3), 398–410. https://doi.org/10.1111/j.1365-2699.2009.02228.x
Vigne, J.-D., & Valladas, H. ène. (1996). Small Mammal Fossil Assemblages as Indicators of Environmental Change in Northern Corsica during the Last 2500 Years. *Journal of Archaeological Science*, *23*(2), 199–215. https://doi.org/https://doi.org/10.1006/jasc.1996.0018

Wade, L. (2015). Breaking a tropical taboo. *Science*, *349*(6246), 370 LP-371. Retrieved from http://science.sciencemag.org/content/349/6246/370.abstract

Wang, H., Yoshimatsu, K., Ebihara, H., Ogino, M., Araki, K., & Kariwa, H. (2000). Genetic Diversity of Hantaviruses Isolated in China and Characterization of Novel Hantaviruses Isolated from Niviventer confucianus and Rattus rattus, *345*, 332–345. https://doi.org/10.1006/viro.2000.0630

Weiß, C. L., Schuenemann, V. J., Shirsekar, G., Reiter, E., Gould, B. A., Stinchcombe, J. R., … Burbano, H. A. (2016). Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens Subject Category : Subject Areas :

# APPENDIX I

Figure 8 presents a fitted model about the observed misincorporation patterns and the posterior predictive intervals. In position 1 of the 5'-end, the substitution rate was at its highest for the C->T (in red) and in position 1 of the 3'-end the highest rate of the G->A (in green) substitutions. Both of these substitution types have similar behavior of an exponential distribution.

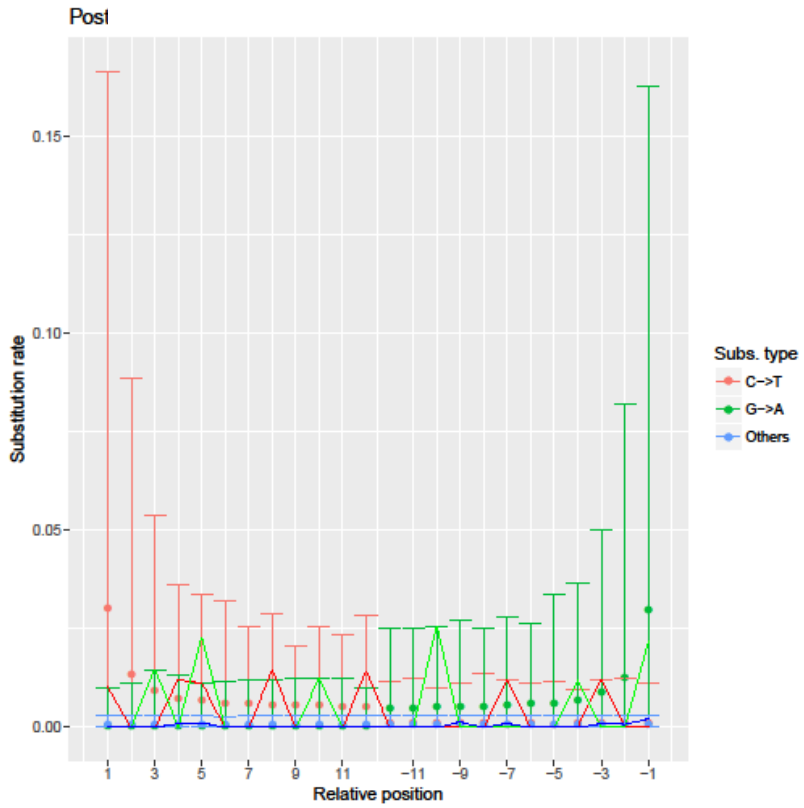


**Figure 8:** Substitution rate graph and posterior predictive intervals obtained from mapDamage

# APPENDIX II

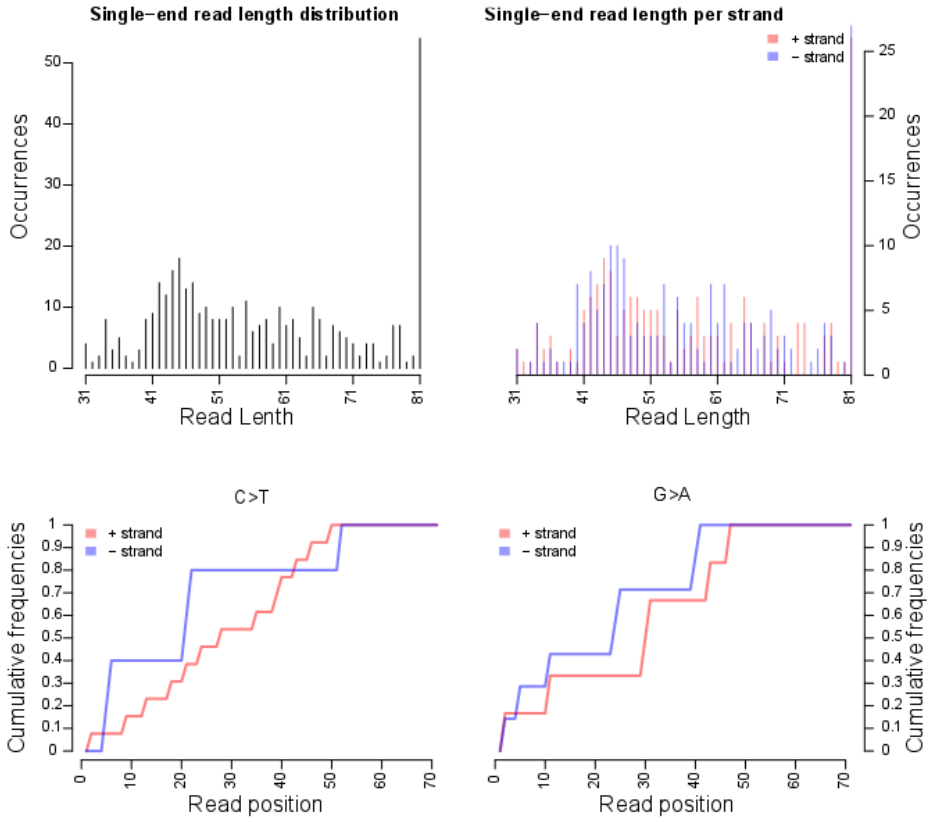


**Figure 9**: The two top plots are length plots. It demonstrates the number of occurrences of the different read lengths throughout the sequence. The longest single-end read was of 81 bp, with many in the 41-51 bp range. The bottom two plots show the cumulative frequency of C->T and G->A misincorporations with a normalization before the first 70 positions. These misincorporations show a tendency of a step like appearance relative to the read position and the frequency. The C->T plot shows steeper steps in the

3’-end strand and smaller ones in the 5’end. For the G->A plot the steps are almost of the same size. This step-like appearance is due to insufficient amount of data, with more data the line would have non-linear tendency.