5-2012

# Single Nucleotide Polymorphism Identification and Phylogenetic Analysis of Pan Troglodytes Using Mitochondrial DNA Coding Region Sequences

Amanda Faris
*University at Albany, State University of New York*

# "Single Nucleotide Polymorphism Identification and Phylogenetic Analysis of *Pan troglodytes* using Mitochondrial DNA Coding Region Sequences"

An honors thesis presented to the
Department of Biological Sciences
University at Albany
State University at New York
In partial fulfillment of the
Honors Program Requirements

Amanda Faris
2012

# The Honors College

**University at Albany**

## BIOLOGICAL SCIENCES

Certification of the Completion of a Senior Honors Thesis

- Successful completion of a senior honors thesis is required for graduation from The Honors College.
- The determination of whether a student's work satisfies the criteria for a senior honors thesis is left to each department.
- Each honors student must return this completed certification, along with an electronic copy of his or her senior honors thesis, to the honors college office.
- This certification must be signed by the student's thesis advisor and by either the Director of the department's honors program or the Chair of the department.

**This Honors Thesis has been read and approved by the signatures below and, is hereby recommended for acceptance.**

Student Name: Amanda Faris — Student ID 000931574

Thesis Title: "Single Nucleotide Identification Polymorphism Identification and Phylogenetic Analysis of *Pan troglodytes* using Mitochondrial DNA Coding Region Sequences"

Department in which Thesis was completed: **BIOLOGICAL SCIENCES**

**Honors Thesis Presentation Date** **5-8-2012**

**Honors Thesis Research Advisor:**

Print name | Signature | Date

**Honors Thesis Committee Member:**

Print name | Signature | Date

**Honors Comm. 2nd Member: (optional)**

Print name | Signature | Date

CHOOSE ONE: ☒ **Department Chair** OR ❒ Director of Department Honors Committee

**Biology Honors Committee Director & Department Chair** RICHARD S. ZITOMER

Print name | Signature | Date

Original – Dr. Jeffrey Haugaard LC-31

Copy –Department of Biological Sciences BI-126

**Abstract**

Despite the close relationship humans share with chimpanzees, the evolutionary history of this species is relatively unknown. Due to its high copy number and high mutation rate, population genetic analyses of chimpanzees have focused on a 1.1 kilo-base segment of mitochondrial DNA (mtDNA) known as the control region. Earlier work based on analyses of the hypervariable region 1 (HVR-1), one of two fragments that make up the control region, support the separation of Upper Guinea and Gulf of Guinea chimpanzees into a monophyletic group separate from Central and East African chimpanzees. I aimed to compare and contrast the chimpanzee mtDNA coding region phylogeny to the mtDNA complete genome phylogeny to analyze the effects of removing the control region. Through this investigation it can be inferred whether analysis of the coding region maintains the relationships between taxa without losing substantial resolution. The second goal of this research was to identify fixed differences outside the control region. In the present study, 59 chimpanzee and bonobo mtDNA genomes were analyzed. Alignments were generated in preparation for the creation of four phylogenies, (complete genome, coding region, entire control region, hyper variable region 1), using the program *Geneious*. The program *Network* was also applied to examine the patterns of diversity within each subspecies and to identify fixed differences. The mtDNA coding region tree inferred the same general relationships between taxa as that proposed by the mtDNA complete genome. The branches of the mtDNA complete genome tree, however, are slightly better resolved due to the high mutation rate of the control region. Although previous and current analysis of mtDNA yields informative phylogenies, the conclusions drawn from this research and past investigations is limited to the relationship between females of this species as mtDNA is maternally inherited. Through subsequent studies of several types of molecular data, a better understanding of the chimpanzee lineage can be achieved. In addition 143 fixed differences were identified that could aid in subspecies classification in future investigations of chimpanzee fecal samples.

**Acknowledgements**

I am indebted to many individuals for their support with this project and to them I would like to extend my gratitude. This research was carried out under the supervision of Prof. Katherine Gonder. The work presented in this thesis was only possible through her guidance, support, and patience. I am thankful to her for allowing me to participate in the Gonder lab; the experiences and knowledge I have gained these past two years will forever be cherished. Furthermore, I am grateful for her permission to work with samples she retrieved from the Limbe Wildlife Centre, Cameroon. Additionally, I would like to give thanks to Matt Mitchell and Paul Sesink Clee for integral feedback and assistance. The transition to laboratory work was eased by their combined help with data analysis programs and presentations. Many thanks also go to George Karakolev, who aided in my understanding of the differences and similarities of the evolutionary histories of chimpanzees, bonobos, and humans. I would like to give a special thanks to Caro-Beth Stewart for assisting with the presentation and approval of this thesis paper. Finally, the following presentation of my investigation would not have been possible without the Limbe Wildlife Centre whom aided in data collection.

**Contents**

**Introduction**

In recent years the evolutionary history and population structure of *Pan troglodytes* (chimpanzee) has been heavily debated. Previous evolutionary studies have mainly centered on *Homo sapiens* (human) nuclear and mtDNA sequences differences, with the aim of discerning our origins (Gonder *et al.* 2007; Ingman *et al*. 2000; Mishmar *et al*. 2002). The information available concerning nucleotide diversity in chimpanzees and bonobos, our close relatives, pales in comparison to humans (Kaessmann *et al.* 1999). Studies of these species, however, can prove to be integral to the understanding of our own, due to the close relationship humans share with chimpanzees and bonobos (Sarich & Wilson 1967). Recent research of ape demography has revealed that chimpanzees and bonobos have populations that are characterized by deep separation times (Bjork *et al.* 2010, Gonder *et al.* 2011, Stone *et al.* 2010, Zsurka *et al.* 2010). Genetic differences between populations of chimpanzees and bonobos even surpasses that separating human populations on separate continents (Bowden *et al.* 2012).

Currently, four subspecies of chimpanzees are recognized (Figure 1): *P.t. verus* inhabits Upper Guinea; *P.t. ellioti* inhabits Nigeria and adjacent parts of Cameroon (Gonder *et al.* 2006); *P.t troglodytes* inhabits central Africa; and *P.t. schweinfurthii* inhabits eastern Africa (Gonder *et al.* 2011). At present, it is generally accepted that bonobos are distinct enough from chimpanzees, both morphologically and molecularly to be considered its own species, *Pan paniscus* (Coolidge 1933). Although four subspecies of chimpanzees are now recognized, the relationships between them are still debated (Bjork *et al.* 2010, Fischer *et al.* 2006,Gonder *et al.* 2011). Different models for population structure can be inferred, depending on the type of data used and samples available. For example, a study of Y-chromosome variation in chimpanzee has led to the suggestion that *P.t.* ellioti are not a distinct group; this study, however, did not include multiple individuals classified as belonging to this subspecies (Stone *et al.*

2002). Moreover, some authors claim that chimpanzees across Africa may be characterized by clinal genetic variation, and thus, suggest that this species should not be divided into subspecies (Fischer *et al.* 2006).

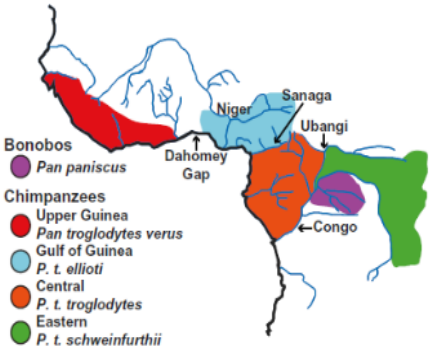


Figure 1) Map of the subspecies distribution of chimpanzees. Range of bonobos is also included. [From Gonder *et al.* 2011.]

Given that different loci and molecular data can lead to the suggestion of different evolutionary histories, several chimpanzee phylogenies have been proposed. Of particular interest has been the use of mtDNA control region sequences (a rapidly evolving non-coding 1.1 kilo-base (Kb) fragment of the mtDNA genome that is prone to mutational saturation and homoplasy) and microsatellites (areas of nuclear DNA that contain short repeats) in chimpanzee studies (Gagneux *et al.* 2001; Gonder *et al.* 2006; Goldberg & Ruvolo 1997; Gonder *et al.* 2011, Morin *et al.* 1994). In many population genetic studies, mtDNA has been analyzed due to such traits as its high copy number, lack of recombination, rapid mutation rate compared to nuclear DNA, and maternal mode of inheritance (Ballard and Whitlock 2004). Until recently, population genetic research on ape mtDNA has focused on a segment of the control region known as hypervariable region 1 (HVR-1) (Gagneux *et al.* 2001, Gonder *et al.* 2006., Goldberg & Ruvolo 1997, Morin *et al.* 1994). Current research has focused more often on studying chimpanzee and bonobo mtDNA complete genomes (Bjork *et al.* 2010, Gonder *et al.* 2011, Stone *et al.* 2010, Zsurka *et al.* 2010). The mtDNA HVR-1 phylogeny of chimpanzees and bonobos is given in figure 2 (modified from Gonder *et al.* 2011). Complete mtDNA genomes and hypervariable control region mtDNA data have yielded a discontinuous population structure for the four chimpanzee subspecies.

According to the mtDNA phylogeny, *P.t. verus* and *P.t. ellioti* share a monophyletic relationship (Bjork *et al.* 2010, Gagneux 2001, Gonder *et al.* 2006, Stone *et al.* 2010). Additionally, East African chimpanzees and Central African chimpanzees form a monophyletic group (Gagneux 2001, Gonder *et al*. 2006, Stone *et al*. 2010).

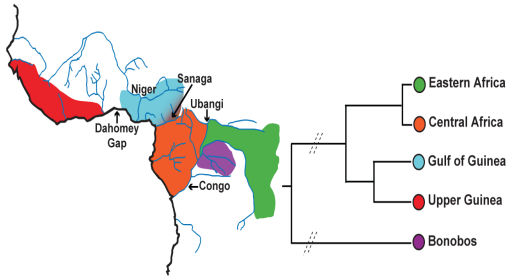


Figure 2) Map and phylogeny displaying the subspecies divisions suggested by analysis of chimpanzee and bonobo mtDNA [modified from Gonder *et al.* 2011].

Recently, autosomal microsatellites that are informative for examining chimpanzee population structure have been analyzed by Gonder *et al*. 2011. This type of molecular data suggests a history of chimpanzee populations that is different from that proposed from analysis of mtDNA data (Figure 3). Analysis of the microsatellite data suggests a more distant relationship between *P.t. ellioti* and *P.t. verus* than suggested by mtDNA (Figure 2). Furthermore, *P.t. ellioti* and *P.t. troglodytes* form a hybrid area surrounding the upper Sanaga River (Gonder *et al.* 2011), which complements earlier data from the mtDNA HVRI suggesting that a zone of overlap between subspecies may exist in central Cameroon (Gagneux et al. 2001; Gonder et al. 2006). Similarly, *P.t. troglodytes* and *P.t. schweinfurthii* do not form two distinctive monophyletic groups, but instead show evidence of clinal variation (Gonder *et al.* 2011), as suggested previously (Gagneux et al. 2001; Gonder et al. 2006).

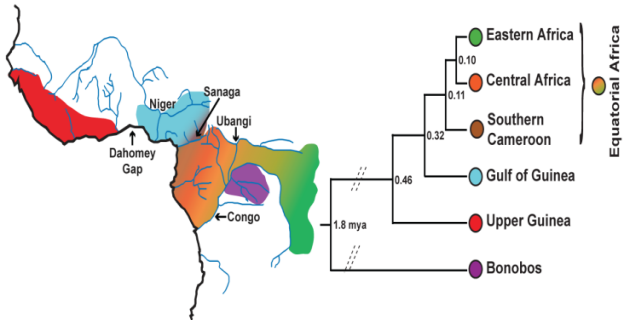


Figure 3) Map and phylogeny displaying the subspecies divisions as suggested by analysis of microsatellites of chimpanzees and bonobos (from Gonder *et al.* 2011)

Two issues I sought to address in this investigation were the following: (1) to investigate if the chimpanzee mtDNA coding region contains the necessary phylogenetic resolution to discern the evolutionary history of the chimpanzee entire mtDNA genome; and (2) to identify informative single nucleotide polymorphisms (SNPs) outside the control region that might be useful for future population genetic studies. In this research, 54 chimpanzee and bonobo mtDNA genomes were retrieved from GenBank that were previously reported in three different studies (Bjork *et al.* 2010; Stone *et al*. 2010; Zsurka *et al.* 2010). These published sequences were aligned to 5 newly sequenced mtDNA genomes that were produced in the Gonder lab using standard protocols. Neighbor joining (NJ) phylogenetic trees — a type of tree that uses distance information to create branches with a minimal total length (Saitou & Nei 1987) — were subsequently formed using the program *Geneious* (Drummond *et al*., 2010), in order to compare and contrast the mtDNA complete genome phylogeny to the phylogeny created using only coding region sequences. Through this investigation it can be inferred whether analysis of the coding region alone recapitulates the relationships between taxa without losing substantial resolution from the more quickly evolving HVR-I. NJ trees were additionally generated for the control region and the HVR-1. A median joining tree — a network that utilizes median vectors (hypothetical ancestors) to reduce tree length (Bandelt *et al*. 1999) — was additionally generated in the program *Network* (Fluxus Technology Ltd 2010) for SNP (a single base differences between taxa) identification. These phylogenetically informative SNPs will allow the mtDNA genotyping of wild chimpanzees without full mtDNA sequencing.

## Methods

### mtDNA Samples from Genbank

MtDNA genomic sequences were retrieved from GenBank for 54 chimpanzees and bonobos. The group of sequences was comprised of 7 from *P.t. schweinfurthii*, 14 from *P.t. troglodytes*, 5 from *P.t. ellioti* , 6 from *P.t verus*, and 22 from *P. paniscus* genomes. These genome sequences can be found under the accession numbers GU112738-GU112745 (Stone *et al*. 2010), GU18957-GU189677, HM015213 (Zsurka *et al.* 2010), and HM068570-HM068593 (Bjork *et al*. 2010).

### Selection of Samples for Further Sequencing

Primary mtDNA sequence data in the form of chromatograms from 25 individuals previously partially sequenced in the Gonder lab were assembled using the software *Sequencher* version 4.9 (GeneCodes Corp., Ann Arbor, MI, USA). Sequencing had been carried out by splitting the genome into 24 different regions for sequencing in the reverse and forward direction. The primers used are listed in Appendix II. All available sequenced regions were aligned to a reference chimpanzee mtDNA genome, accession number GU112739 (Stone *et al.* 2010*)*, for sequence assembly. Subsequently, base calls were manually edited through analysis of chromatograms of overlapping sequences. Regions where further sequencing was needed due to missing areas in the sequenced genome or due to low quality base calls were annotated. From this annotation, the five most complete genomes were selected for further analysis. Four of these five individuals were selected for further mtDNA sequencing.

### Touchdown Long-range PCR

Four mtDNA *P. troglodytes* genomes were selected from the 25 partially sequenced individuals previously assembled in the Gonder lab for amplification: *P. t. troglodytes* (2), *P. t. ellioti* (1), and transition zone *P. troglodytes* (1). This individual was included because it likely came from a chimpanzee hybrid zone between the confluence of the Sanaga and Mbam Rivers in central Cameroon.

Hifi long-range Platinum Taq kits [Invitrogen, Corp.] were used to complete PCR, following the manufacturer's protocol. Completing PCR reactions involved making a master mix of 18.75 microliters (µl) of water, 2.5 µl of 10x standard PCR buffer, 1 µl of $MgSO_4$, 0.5 µl of 10 M dNTPs, 0.5 µl of 10µM/µL each primer, and 0.25 µl *Taq* polymerase for each 1µl of 50ng/ml mtDNA sample amplified. Two primers (PCR region 1 primer and PCR region 2 primer) were used to amplify the genome into 2 long fragments (Appendix II). The temperature conditions of the touchdown PCR amplification were: initialization at 94 C° for 2 minutes (min); 20 cycles of denaturation at 94 C° for 30 seconds (s), annealing at 53 C° for 45 s, and elongation at 68 C° for 10 min 30 s; followed by another 20 cycles of denaturation at 94 °C for 30 s, annealing at 47.9 °C for 45 s, and elongation at 68 °C for 10 min 30 s; and final extension at 68 C° for 11 min. The PCR reaction products were electrophoresed through 1% agarose gels to estimate their sizes and qualities.

**SAP/EXO 1 mtDNA Cleanup**

The resulting PCR products were treated with SAP and EXO 1 enzymes to purify them from excess dNTPs and primers. A master mix was made with 0.25 µl SAP enzyme, 0.25 µl EXO enzyme and 0.5µl dilution buffer [reagents from USB Co./Amersham] for each 5 µl sample of PCR product used. The temperature conditions of the digestion of the surplus nucleotides were: 37 °C for 30 min; 80 °C for 15 min; 12 °C for 5 min; and 4 °C for 10 min.

**Sanger Sequencing**

The purified amplified PCR products were sequenced using the 48 primers listed in Appendix II. A master mix was made containing 3 µl of standard sequencing buffer, 3.67 µl water, 0.33 µl of 1µM/µL primer, 2 µl Big Dye, and 5 µl SAP/EXO [Applied Biosystems]. Prior to the addition of primers to the mix, primers were diluted from 10 µM to 1 µM for cycle sequencing. The temperature conditions of the sequencing reactions were: 96 °C for 1 min; 30 cycles at 96 °C for 10 s; 50 °C for 5 s; and 4 °C for 55 min.

The sequencing reactions were subsequently prepared for analysis on an ABI 3730 using a standard isopropanol precipitation protocol to remove sequencing primers and unincorporated nucleotides.

**Assembly of Sample Sequences**

The newly-generated mtDNA sequences were added to the previously assembled genomes using *Sequencher* version 4.9 (GeneCodes Corp., Ann Arbor, MI, USA). Areas sequenced in this investigation are indicated in Table I. Areas that remain to be sequenced for all 25 individuals present in the Gonder lab were annotated in Table II.

| Names of Individuals Sequenced | Regions Sequenced [labeled by primer used (see Appendix II)] |
|---|---|
| Nemo (*P.t. ellioti)* | 1F, 1R, 3F, 4F, 4R, 5R, 6F, 6R, 9F, 10F, 12R, 16F, 16R, 17F, 17R, 18R, 24F |
| Maya (*P.t. troglodytes)* | 1F, 1R, 2R, 3F, 4F, 4R, 5R, 6F, 6R, 9R, 12R, 15F, 16F, 16R, 17F, 17R, 18R, 24F |
| Jules (*P.t. troglodytes)* | 6R, 9R, 12R, 15F, 16R |
| Lada (Transition zone *P. troglodytes*) | 4F |

**Table I:** Table I displays a summary of the mtDNA sequencing of 4 individuals. Column one indicates the names of the chimpanzees whose mtDNA were sequenced. Column II indicates the regions of the mtDNA genomes that were sequenced from each individual.

**Alignment Generation**

An alignment was generated using the program *Geneious* version 4.7 (Drummond *et al*. 2010) that contained the 54 retrieved mtDNA genomes, the 4 newly assembled genomes and 1 additional *P.t. ellioti* previously sequenced in the Gonder Lab. Manual editing was completed in *MacClade* version 4.07 (Maddison & Maddison 2005).

**Phylogenetic Analysis**

Four roote NJ trees were inferred using the program *Geneious* version 4.7 (Drummond *et al*. 2010). One tree included 54 complete *P. troglodytes* and *P. paniscus* mtDNA sequences, plus the 5 sequenced individuals from the Gonder lab. The second tree included these same 59 genomes, minus their control regions. The third phylogeny generated analyzed these same sequences using only the

control region. The final phylogeny produced included only the HVR-1 region of all 59 sequences. These analyses were carried out in order to determine if the coding region phylogeny displays accurately the taxon relationships within the mtDNA genome phylogeny without losing significant resolution. *Geneious* does not conduct pairwise comparisons at nucleotide positions where one of the two sequences contains missing data. Pairwise comparisons in the same column are still performed as long as the pair of sequences contains no missing data (Drummond *et al.* 2010).

**Network and SNP annotation**

In order to proceed with SNP identification, the previous manually edited alignment containing complete genomes was reduced to contain only coding region sequences. Further reduction was carried out by editing the alignment so that only variable sites were displayed, again using *MacClade* version 4.07 (Maddison & Maddison 2005). This edited alignment data was then used to generate a median joining tree in the program *Network* (Fluxus Technology Ltd, 2010). Patterns of diversity were analyzed within each subspecies and informative SNPs were then identified along major branches. Characteristics of these polymorphisms were annotated and summarized in Appendix III.

**Results**

This study involved the compilation of 54 mtDNA genomes from the species *P troglodytes* and *P. paniscus* and 5 newly-assembled *P. troglodytes* mtDNA genomes that were generated in the Gonder laboratory. Genomes published by Bjork and colleagues (Bjork *et al.* 2010) have several bases that are represented by the letters W (A or T), R (A or G), Y (C or T), S (C or G), K (G or T), M (A or C) and N (unknown). These ambiguous regions were infrequent and were ignored in pairwise comparisons when forming NJ phylogenies (Drummond *et al.* 2010). Ambiguous base calls were manually converted to N's for network generation; *Network* converts ambiguities to the most common base state (Fluxus Technology Ltd, 2010).

In addition to the previously published mtDNA genome sequences, four newly sequenced and one individual previously sequenced by members of the Gonder Laboratory were also included in this analysis. Areas that were sequenced in this study are indicated in Table I. Due to this new sequencing, these 4 genomes are mostly complete. These new data, along with annotations concerning previously analyzed data, are summarized in Table II. The colors in Table II denote successfully sequenced regions to date and areas that should be sequenced in future studies. Failed sequences, indicated in red, include regions that had low quality sequencing or had short sequences and therefore lacked coverage. The 5 individuals that were selected for this analysis are additionally indicated in pale green. These individuals were chosen because they were the most complete out of the original 25 samples.

Two NJ trees [complete genome tree (Appendix I, Figure A1) and coding region tree (Appendix 1, Figure A2)] were generated to demonstrate that phylogenetic analysis without the control region yielded accurate mtDNA relationships between taxa without sacrificing significant resolution. The two phylogenies shown in Appendix I (Figure A1 and A2) are similar, indicating that analysis of only coding region data can be carried out in subsequent work. Both trees display similar divisions, dividing *P.*

*troglodytes* into three major groups: (1) *P.t. verus* (Upper Guinea); (2) *P.t. ellioti* (Gulf of Guinea); and (3) *P.t. troglodytes* (Central African) plus *P.t. schweinfurthii* (East African). Bonobos are included in their own monophyletic outgroup. As with most prior studies on mtDNA, these trees suggest that *P.t. ellioti* and *P.t. verus* are sister taxa. Additionally, two phylogenies were made of the control region [entire control region (Appendix 1, Figure A3), and HRV-1 (Appendix 1, Figure A4)]. Analysis of the control region alone yielded a phylogeny similar to the microsatellite phylogeny. On the other hand, removal of hypervariable region 2 (HRV-2) resulted in a tree topology the relationships inferred were similar to that of the entire mtDNA genome.

Subsequent to phylogenetic analysis, a median joining tree was generated for later SNP identification. The phylogeny generated by the *Network* software (Appendix 1, Figure A5) shows the same three groupings for chimpanzees as indicated in the previous trees. Homoplasy, the same change happening independently, was not observed within this phylogeny; this can be seen through the absence of cubes joining the haplotypes. The presence of cubes in the control region network supports the existence of homoplasy within this mtDNA segment (Appendix 1, Figure A6). Coding region informative sites (nucleotide positions in a genome that resolve lineages within a phylogeny) were then annotated as fixed (all of the individuals in one group have the same base at a certain site, while individuals of another group have a different base that is characteristic to them) or not fixed (bases are not specific to certain groups). Whether the base change was a transition or a transversion was also noted; this information is tabulated in Appendix III.

**Table II: Annotation of Assembled Sequences**

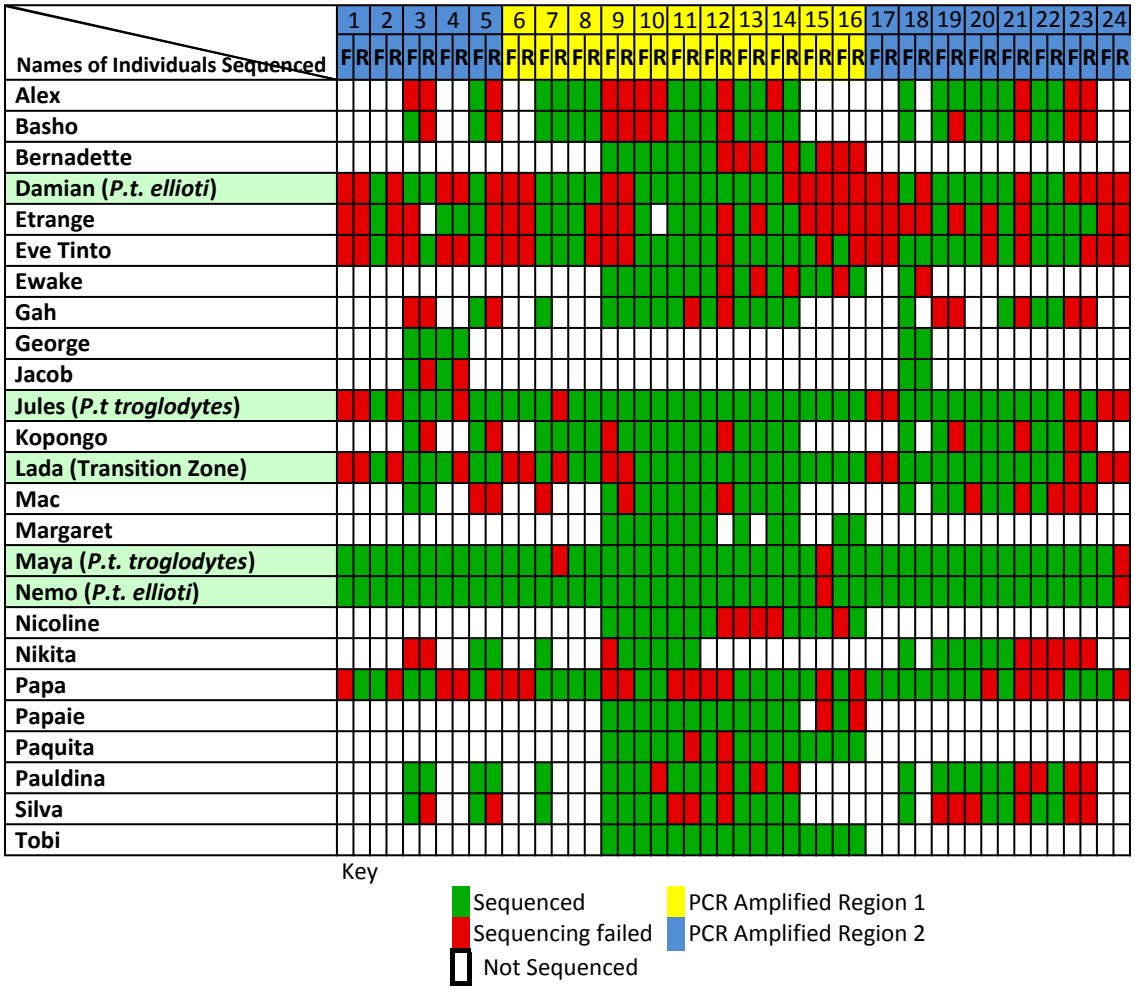


Table II: Summary of mtDNA sequencing from 25 chimpanzees by Gonder lab members. The names in the first column indicate the names of the chimpanzees from which the samples were retrieved. Subspecies designation of the individuals included in the phylogenetic analyses are indicated in partentheses. Numbers at the top of the table indicate the number of the primers, (see Appendix III for a list of the primers). F and R are used to indicate sequencing in the forward and reverse direction respectively.

**Discussion**

Recently, there has been an abundance of chimpanzee and bonobo published mtDNA genome analyses (Bjork *et al.* 2010, Gonder *et al.* 2011, Stone *et al.* 2010, Zsurka *et al.* 2010). The overall aim of this research was to gain a better understanding of the evolutionary history of chimpanzee mitochondrial DNA. Specifically, the first goal of this study was to test whether the mtDNA coding region phylogeny was consistent with the complete mtDNA genome phylogeny. Five new genomes sequenced in the Gonder lab were aligned with 54 genomes retrieved from GenBank. These 5 individuals were selected from 25 genomes previously assembled in Sequencher (GeneCodes Corp., Ann Arbor, MI, USA). In future investigations, the other 20 individuals I assembled should be targeted for sequencing so that a larger data set can be analyzed to facilitate searching for rare SNP variants and minimize the risk of ascertainment biased in larger population samples.

Mitochondrial DNA genome samples under the accession numbers HM068570-HM068593 contained several areas with low quality sequences. These errors most likely emerged due to the many issues that arise from the use of DNA extracted from fecal samples. These difficulties include degradation and DNA contamination (Taberlet et al. 1999). These regions were not problematic in subsequent genetic analysis since these ambiguous base call areas were limited. In the program *Network*, base calls outside of the standard A, T, C, and G are impossible to interpret; consequently, it was necessary to convert all ambiguous base call to 'N's.

Two rooted NJ trees were inferred using the program *Geneious,* in order to interpret whether the mtDNA coding region tree yielded relationships between taxa consistent with the mtDNA complete genome phylogeny. Comparison of the trees in Figure A1 and A2 shows that the phylogeny using only coding data matches closely to that inferred from complete mtDNA genome sequences. By eliminating the hypervariable control region, homoplasy is eliminated. Evidence of the absence of homoplasy can be seen through the lack of cubes on the median joining network displayed in Figure A5 and the

presence of cubes in Figure A6. The removal of the control region, however, does lead to a reduction in the resolution of individuals within the major clades.

Based on the phylogenies generated, three conclusions can be made. Firstly, the Upper Guinea and Gulf of Guinea chimpanzee mtDNA genomes are sister groups that form a monophyletic group that excludes Central and East African mtDNA genomes. Secondly, the Central and East African chimpanzee mtDNA genomes do not form monophyletic sister groups, but rather are intermixed in the trees, consistent with a genetic cline spanning central Africa. These conclusions support previous phylogenies of the chimpanzee hvr1 region (Gagneux *et al.* 2001, Gonder *et al.* 2006). Finally, there may be unrecognized reservoirs of genetic diversity amongst bonobos. Visual analysis of the phylogenies in figure A1 and A2 reveals several deep mtDNA lineages within the bonobos. It can therefore be suggested from this data that subspecies structure may exist for this species. However, it should be noted that the geographic origins of the bonobos included here remain unknown since the samples were from individuals housed at sanctuaries. Despite this limitation, these results suggest that additional analyses of bonobos from samples of known origin might be beneficial in to examine if these 'pockets' of genetic diversity correspond to geographical subdivisions, and ultimately correspond to significantly different populations and/or possibly subspecies.

The second part of this investigation involved the identification of informative SNPs outside the control region that might be useful for studies using fecal DNA samples. This task was carried out by constructing a median joining network containing only chimpanzee sequences. The network created, shown in Figure A5, shows the same three groupings of chimpanzees inferred from the previously-mentioned NJ trees. The longest branch separates *P.t. troglodytes* and *P.t. schweinfurthii* from *P.t. ellioti and P.t verus*. These results suggest that equatorial African chimpanzees diverged from populations located west of the Sanaga River a long time ago. This claim is further supported by the

data presented in Table IIA, which indicates that there are 112 fixed differences that separate these two mtDNA lineages.

This research should be extended in subsequent investigations. Twenty samples previously retrieved from the Limbe Wildlife Centre, Cameroon, remain to be included in this analysis. By using the fixed differences identified in this research, these partially sequenced samples can be categorized. Finally, more genetic analysis should be carried out analyzing the base differences between bonobos. Bonobos appear to be more greatly diverged within their own lineage than initially believed.

**Literature Cited**

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**: 147.

Ballard JW, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol Ecol* **13**: 729-744.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies *Mol Biol Evol* **16**: 37-48.

Bjork A, Liu W, Wertheim JO, Beatrice HH, Worobey M. 2010. Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. *Mol Biol Evol* **28:** 615-623.

Bowden R, MacFie TS, Myers S, Nerrienet E, Bontrop RE, Freeman C, Donnelly P, Mundy NI. 2012. Genomic Tools for Evolution and Conservation in the Chimpanzee: *Pan troglodytes ellioti* Is a Genetically Distinct Population. *PLoS Genet* **8**: e1002504. doi: 10.1371/journal.pgen. 1002504

Coolidge HJ. 1933. *Pan paniscus*. Pigmy chimpanzee from south of the Congo river. *Am J Phys Anthropol* **18**: 1-59.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir, R, Stones- Havas S Sturrock S, Thierer T, Wilson A. 2010. Geneious v5.1, Available from http://www.geneious.com.

Fischer A, Pollack J, Thalmann O, Nickel B, Pääbo S. 2006. Demographic history and genetic differentiation in apes. *Curr Biol* **16**: 1133-1138.

Fluxus Technology Ltd 2010. Network, Available at http://www.fluxus-engineering.com/

Gagneux P, Gonder MK, Goldberg TL, Morin PA. 2001. Gene flow in wild chimpanzee populations what genetic data tell us about chimpanzee movement over space and time. *Philos Trans R Soc L and B Biol Sci* **356**: 889-897.

Gene Codes Corporation. 2009. *Sequencher* v4.10.1, Available from http://www. genecodes.com/.

Goldberg TL, Ruvolo M. 1997. The Geographic Apportionment of Mitochondrial Genetic Diversity in East African Chimpanzees, *Pan troglodytes schweinfurthii. Mol Biol Evol* **14**: 976-984.

Gonder MK, Disotell T, Oates J. 2006. New genetic evidence on the evolution of chimpanzee populations and implications for taxonomy. *Int J Primatol* **27**: 1103-1127.

Gonder MK, Locatelli S, Ghobrial L, Mitchell M, Kujawski J, Lankester F, Stewart C-B, Tishkoff S. 2011. Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations. *Proc Natl Acad Sci U S A.* **108**: 4766-4771.

Gonder MK, Mortensen HM, Reed FA, de Sousa, A, Tishkoff SA. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol* **24**: 757-768.

Gonder MK, Oates JF, Disotell TR, Forstner MR, Morales JC, Melnick DJ. 1997. A new west African chimpanzee subspecies? *Nature* **388**: 337.

Groves CP. 2005. Geographic variation within eastern chimpanzees. *Australas Primatol* **17**: 19-46.

Hey J. 2009. The divergence of chimpanzee species and subspecies as revealed in multi-population isolation-with-migration analyses. *Mol Biol Evol* **27:** 921-933.

Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**: 708-713.

Kaessmann H, Wiebe V, Pääbo S. 1999. Extensive nuclear DNA sequence diversity among chimpanzees *Science* **286**: 1159-1161.

Maddison WP, Maddison DR. 2005. *MacClade: Analysis of Phylogeny and Character Evolution* Version 4.07. Sinauer Associates, Sunderland, MA.

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E Brown MD, Sukernik RI, Olckers A, Wallace, DC. 2003. Natural selection shaped regional mtDNA

variation in humans. *Proc Natl Acad Sci U S A* **100**: 171-176.

Morin PA, Moore JJ, Chakraborty R, Jin L, Goodall J, Woodruff DS. 1994. Kin selection, social structure gene flow, and the evolution of chimpanzees. *Science.* **265**: 1193-1201.

Saitou N & Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425.

Sarich VM, Wilson AC. 1967. Immunological Time Scale for Hominid Evolution. *Science*. **158**: 1200-1203.

Stone AC, Battistuzzi FU, Kubatko LS, Perry Jr. GH, Trudeau E, Hsiuman L, Kumar S. 2010. More reliable estimates of divergence times in *Pan* using complete mtDNA sequences and accounting for population structure. *Philos. Trans R Soc B Biol Sci* **365**: 3277-3288.

Stone AC, Griffiths RC, Zegura SL, Hammer MF. 2002. High levels of Y-chromosome nucleotide diversity in the genus *Pan*. *Proc Natl Acad Sci U S A* **99**: 43-48.

Taberlet P, Waits LP, Luikart G. 1999. Noninvasive genetic sampling: look before you leap. *Trends Ecol Evol.* **14**: 323-327.

Zsurka G, Kudina T, Peeva V, Hallmann K, Elger CE, Khrapko K, Kunz WS. 2010. Distinct patterns of mitochondrial genome diversity in bonobos (*Pan paniscus*) and humans. *BMC Evolutionary Biology* **10:** 270.

**Appendix I**

**Figure A1**

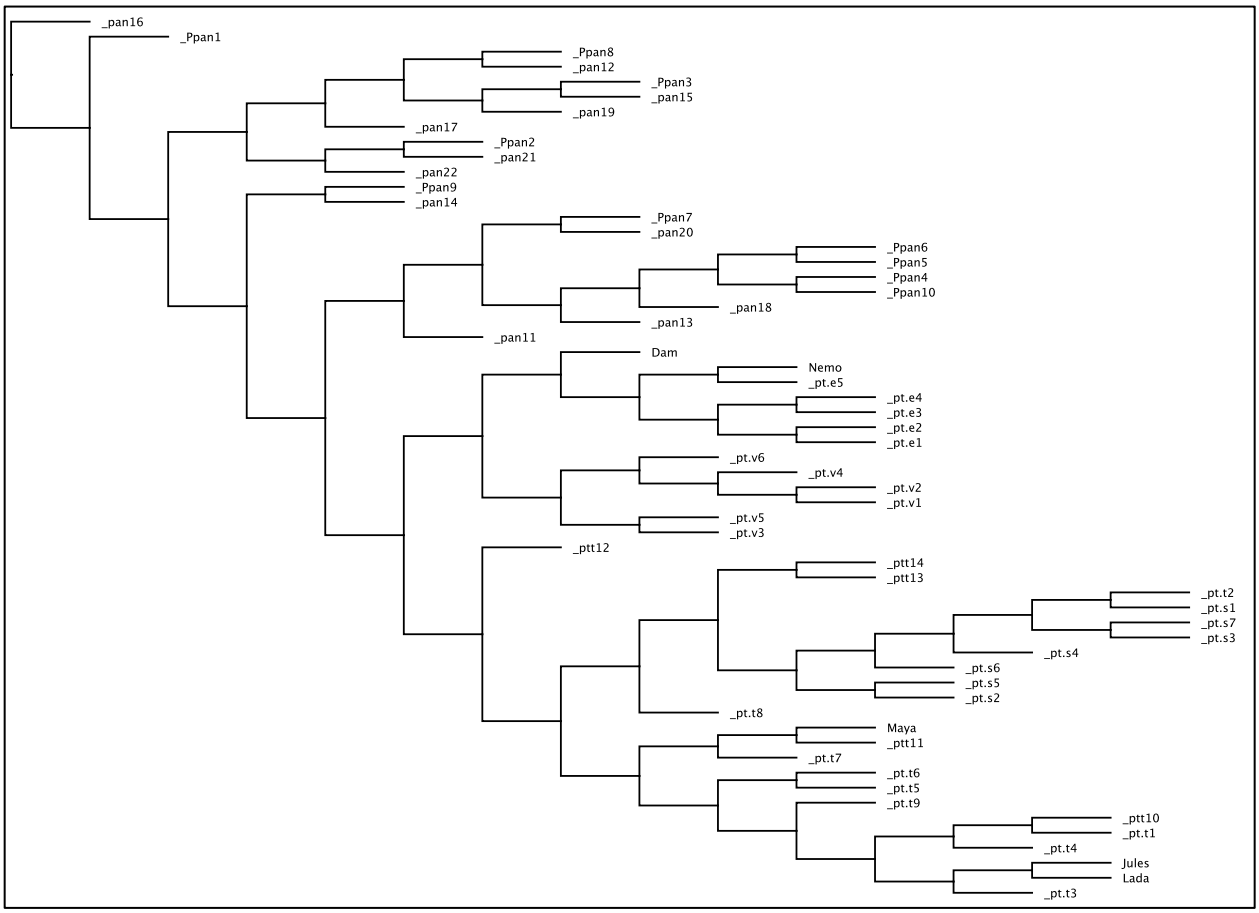


Figure A1- Displayed in figure A1 is a rooted neighbor joining tree containing 59 complete mtDNA genome sequences. Taxon labels: [Ppan# & pan# sequences are *P.paniscus*; Pt.e# sequences are *P.t. ellioti*; Pt.v# sequences are *P.t verus*; Ptt# & Pt.t# sequences are *P.t. troglodytes*; Pt.s# sequences are *P.t. scweinfurthii*]. Gonder lab sequences are indicated by name: [Dam= Damian (*P.t. ellioti*); Nemo (*P.t. ellioti*); Maya (*P.t. troglodytes*), Jules (*P.t. troglodytes*)].

**Figure A2**

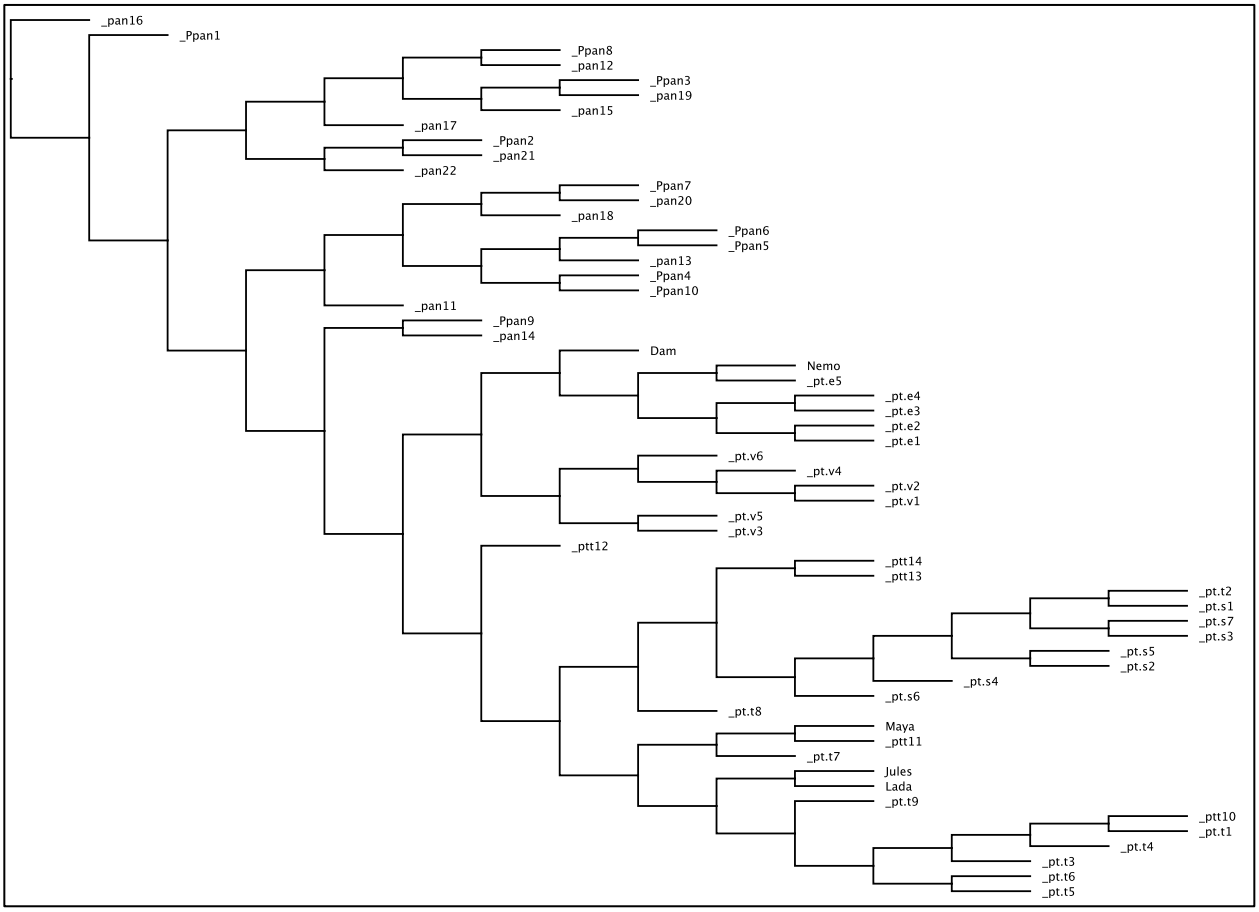


Figure A2- Displayed in figure A2 is a rooted neighbor joining tree containing 59 coding region mtDNA sequences. Taxon labels: [Ppan# & pan# sequences are *P.paniscus*; Pt.e# sequences are *P.t. ellioti*; Pt.v# sequences are *P.t verus*; Ptt# & Pt.t# sequences are *P.t. troglodytes*; Pt.s# sequences are *P.t. scweinfurthii*]. Gonder lab sequences are indicated by name: [Dam= Damian (*P.t. ellioti*); Nemo (*P.t. ellioti*); Maya (*P.t. troglodytes*), Jules (*P.t. troglodytes*)].

**Figure A3**

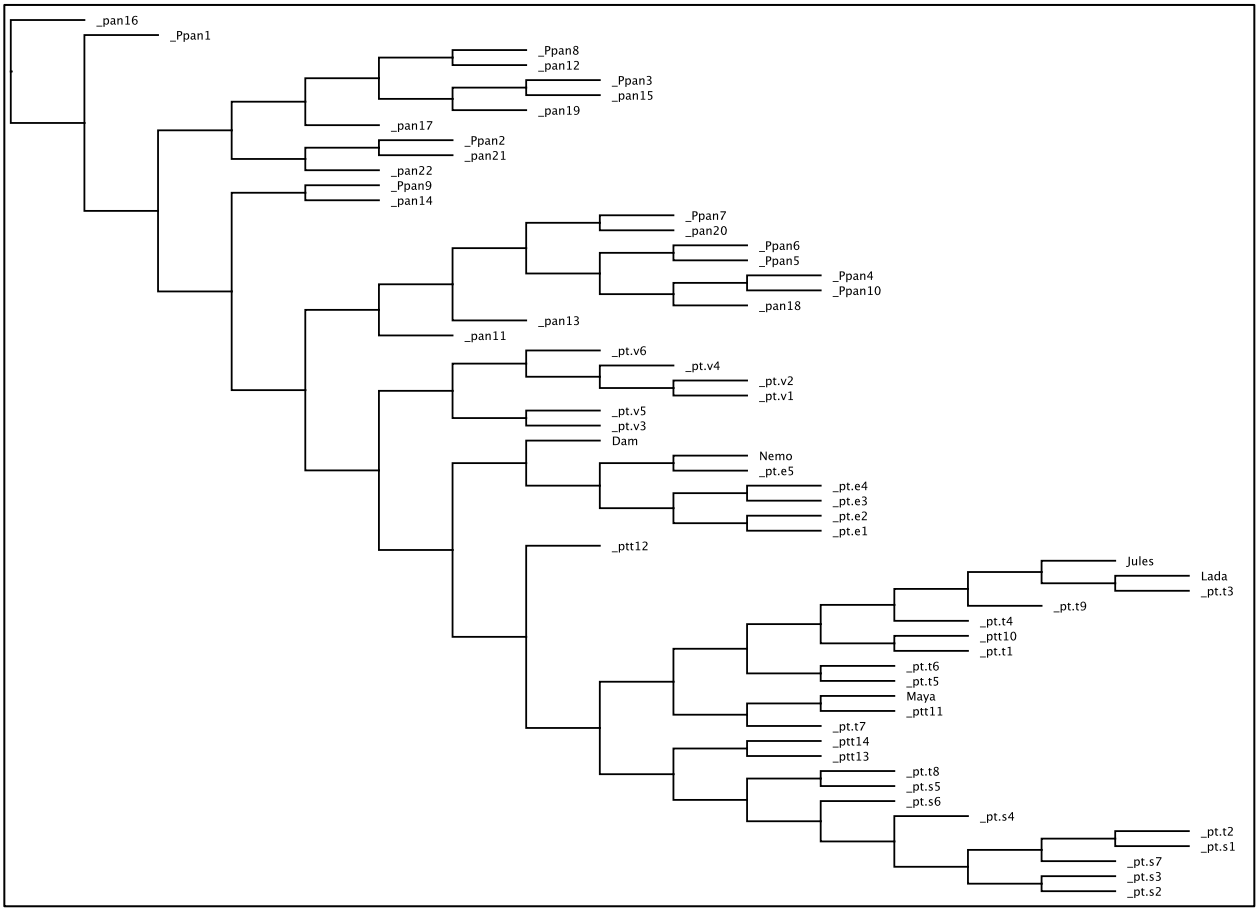


Figure A3- Displayed in figure A3 is a rooted neighbor joining tree containing 59 control region mtDNA sequences. Taxon labels: [Ppan# & pan# sequences are *P.paniscus*; Pt.e# sequences are *P.t. ellioti*; Pt.v# sequences are *P.t verus*; Ptt# & Pt.t# sequences are *P.t. troglodytes*; Pt.s# sequences are *P.t. scweinfurthii*]. Gonder lab sequences are indicated by name: [Dam= Damian (*P.t. ellioti*); Nemo (*P.t. ellioti*); Maya (*P.t. troglodytes*), Jules (*P.t. troglodytes*)].

**Figure A4**

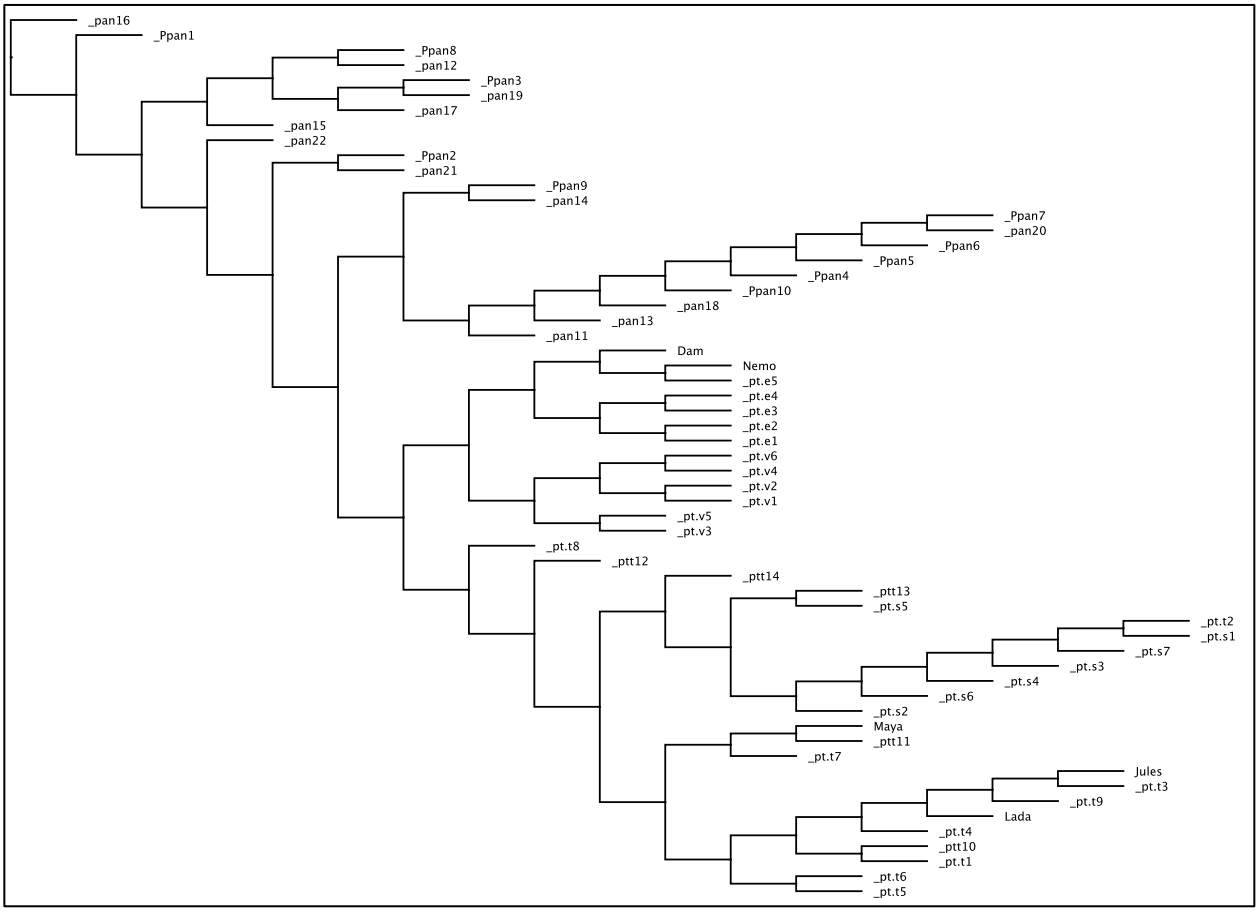


Figure A4- Displayed in figure A4 is a rooted neighbor joining tree containing 59 hvr1 mtDNA sequences. Taxon labels: [Ppan# & pan# sequences are *P.paniscus*; Pt.e# sequences are *P.t. ellioti*; Pt.v# sequences are *P.t verus*; Ptt# & Pt.t# sequences are *P.t. troglodytes*; Pt.s# sequences are *P.t. scweinfurthii*]. Gonder lab sequences are indicated by name: [Dam= Damian (*P.t. ellioti*); Nemo (*P.t. ellioti*); Maya (*P.t. troglodytes*), Jules (*P.t. troglodytes*)].

**Figure A5**

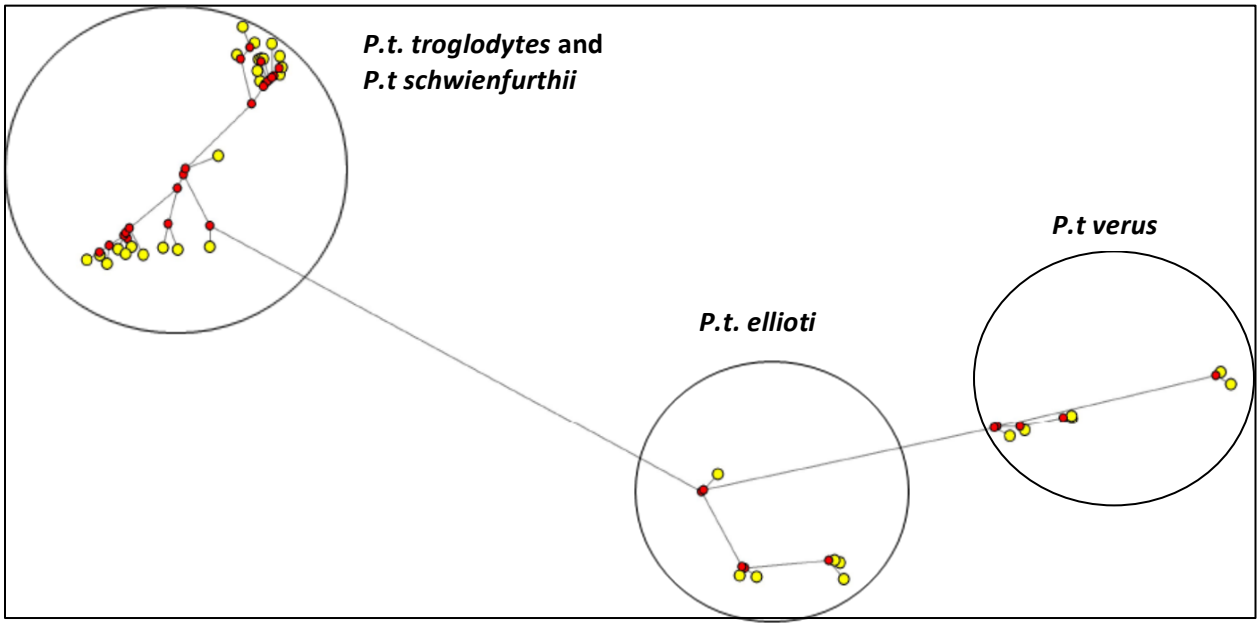


Figure A5- Median joining network displaying the relationships between the 59 coding region sequences. Median vectors (hypothetical ancestors) are indicated as red dots. Taxa are indicated as yellow dots.

**Figure A6**

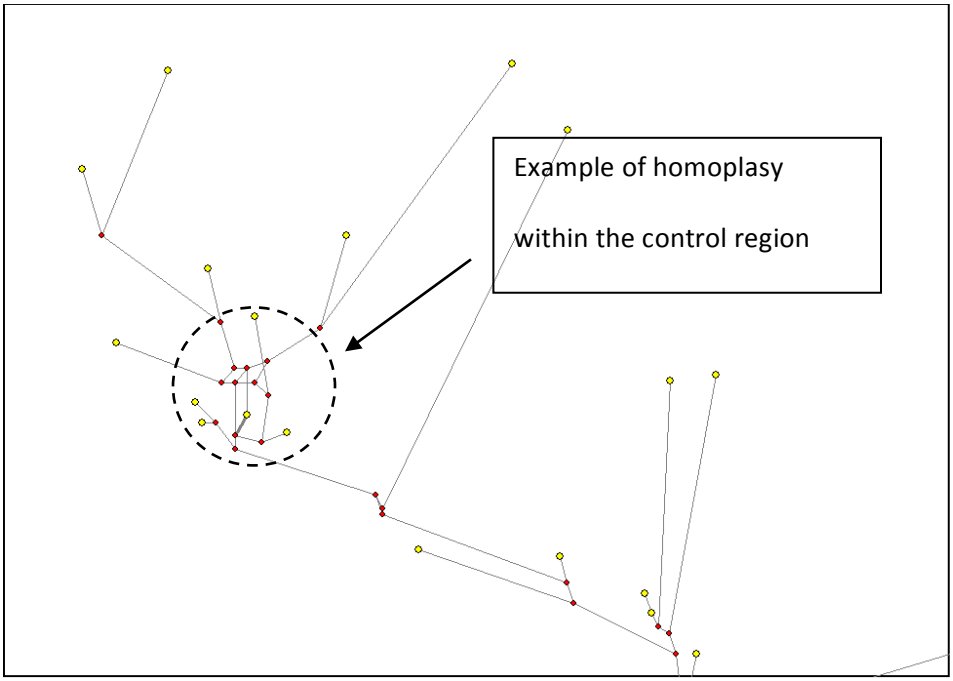


Figure A6- Figure A6 shows a section of the median joining network displaying the relationships between the control region sequences of the 59 mtDNA samples. Median vectors (hypothetical ancestors) are indicated as red dots. Taxa are indicated as yellow dots. Homoplastic mutations are indicated as cubes as emphasized on Figure A6.

**Appendix II**

| Name of Primer | Primer Sequence |
|---|---|
| 1F | CAG GTT TGG TCC TAG CCT TTC |
| 1R | GCT AAA TCC ACC TTT GAC CCT TAG G |
| 2F | CCA CTA TGC TTA GCC CTA AAC TTC |
| 2R | GGA CAA CCA GCT ATC ACC AG |
| 3F | CTA ACC CCT GTA CCT TTT GC |
| 3R | GGG TAG GTC AAT TTC ACT GGT TG |
| 4F | CAT TAC CAG TAT TAG AGG CAC CGC |
| 4R | GGA ATG CTA TTG CGA TTA GGA TGG |
| 5F | CCA ACA CCC ACT CAA GAA CAG |
| 5R | GGA AGA TTG TAG CAG TGA GGG TG |
| 6F | CTC CTT CAA TCT CTC TAC CCT TGT C |
| 6R | GAG TAT GCT GTT GGA GAG GAT AG |
| 7F | CAT CTT TAC AGG CAC GCT CAT TAC |
| 7R | GTG GTA AGG GTG ATG AGT GTA GG |
| 8F | CCT CAT CAT CC CAC CAT CAT AGC |
| 8R | GTA GAT GTG GTC ATT ACC TAG |
| 9F | CCT GTC TTT AGA TTT ACA GTC C |
| 9R | GGA GTA ATA AGT TAC AAT GTG GG |
| 10F | CTT CGT CTG ATC CGT CCT AAT C |
| 10R | GGA GGG CTC TTC TAC TAT TAG G |
| 11F | CAG GCT ATA CCC TAG ACC AAA C |
| 11R | GGT GAA AGT GGT TTG GTT TAG G |
| 12F | CGG GCT AAT CTT CAA CTC CTA C |
| 12R | GAT AGT TGG GTG GTT GGT GTG |
| 13F | CCC ACT TCT AAA CAT CTC ATC AAC |
| 13R | GGT ATG TGC CTT CTC GTA CAA CAT C |
| 14F | CAA CAC ATA ATG ACC CAC CAA TCA C |
| 14R | GTT GAG TTG TGG TAG TCA GAA TGT G |
| 15F | VTA TCT ACT GAT GAG GAT CTT ACT C |
| 15R | GGC TGT GAG TGG TTG TGT TG |
| 16F | CAG CAG TAG GTC TAG CAC TAC |

| | |
|---|---|
| 16R | GGA TGA GTG TGA GGC GTA TTA TG |
| 17F | CCC AAG AAC TAT CAA ACA CCT GAG C |
| 17R | GAT TGA GGG ATA GAA GGA GGA TGG C |
| 18F | CCT ACT CAC AGG ATT CAA CAT AC |
| 18R | GCC GAT GAA GAG TTG GAA TAG |
| 19F | CAA CCC AAA CAA CCC AAC TCT C |
| 19R | GAT GAT GAG GTC TTT GGA GTA G |
| 20F | CCA CAC CTA GCA TTC CTT CAC ATC |
| 20R | GAT TAT GGG CGT TGG TTA GTA G |
| 21F | CCA AAT CTC CGC TTC CAT TAC C |
| 21R | GAG GAT AAT GCC AAT GTT TCA GG |
| 22F | CTA TAC ACT ACT CAC CAG ACG CC |
| 22R | GCT TTG GGT GCT GAT GGT GAA G |
| 23F | CCT AAT ACC AAT CGC CTC TCT AAT C |
| 23R | CGT TAA TAG GGT GAT AGA CCT GTG |
| 24F | CAC CAT CCT CCG TGA AAT CAA TAT C |
| 24R | GCT AGG CTA AGC GTT TTG AGC TG |
| | |
| PCR 1F | CCA ACA TTG TAG GTC CTT ACG GGC |
| PCR 1R | GCT ACG ACT ATG GTG CTT GAG TGG |
| PCR 2F | GAT AAC TCA TTT GGG TAA GAA GCC |
| PCR 2R | CTG AAC GCA GGT ACA TAC TTC CTA |

Table IA- Table IA displays the primers used for sequencing and amplification. Column 1 shows the names of the primers used [sequencing primers: 1F-24R; amplification primers PCR 1F-PCR 2R]. Numbers within the names of the primers indicate the region targeted. The letters F and R within the primer names indicate the direction of sequencing or amplification, forward or reverse.

**Appendix III**

**Table IIA**

| SNP | Location | Ts | Tv | Mixed | | | | |
|---|---|---|---|---|---|---|---|---|
| 2 | 468 | | | C-T-N | | | | |
| 3 | 547 | | | A-G-N | | | | |
| 4 | 670 | A-G | | | | | | |
| 5 | 733 | A-G | | | | | | |
| 7 | 779 | C-T | | | | | | |
| 8 | 968 | C-T | | | | | | |
| 9 | 970 | C-T | | | | | | |
| 11 | 1129 | C-T | | | | | | |
| 15 | 1332 | C-T | | | | | | |
| 17 | 1431 | C-T | | | | | | |
| 20 | 1503 | A-G | | | | | | |
| 21 | 1511 | C-T | | | | | | |
| 23 | 1676 | C-T | | | | | | |
| 25 | 1850 | A-G | | | | | | |
| 30 | 2191 | C-T | | | | | | |
| 33 | 2336 | A-G | | | | | | |
| 37 | 2806 | C-T | | | | | | |
| 39 | 2845 | | | C-T-A | | | | |
| 40 | 2860 | C-T | | | | | | |
| 41 | 2863 | C-T | | | | | | |
| 44 | 2914 | A-G | | | | | | |
| 45 | 2926 | C-T | | | | | | |
| 46 | 2966 | C-T | | | | | | |
| 47 | 3010 | A-G | | | | | | |
| 49 | 3112 | A-G | | | | | | |
| 51 | 3134 | C-T | | | | | | |
| 53 | 3172 | C-T | | | | | | |
| 55 | 3208 | C-T | | | | | | |
| 57 | 3233 | A-G | | | | | | |
| 60 | 3301 | C-T | | | | | | |
| 61 | 3313 | A-G | | | | | | |
| 63 | 3355 | A-G | | | | | | |
| 64 | 3373 | C-T | | | | | | |
| 67 | 3505 | A-G | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 69 | 3538 | C-T | | | | | | |
| 71 | 3589 | A-G | | | | | | |
| 72 | 3601 | C-T | | | | | | |
| 74 | 3657 | C-T | | | | | | |
| 76 | 3770 | A-G | | | | | | |
| 77 | 3885 | C-T | | | | | | |
| 80 | 3943 | A-G | | | | | | |
| 81 | 3951 | A-G | | | | | | |
| 82 | 4066 | C-T | | | | | | |
| 86 | 4161 | C-T | | | | | | |
| 87 | 4170 | C-T | | | | | | |
| 88 | 4267 | C-T | | | | | | |
| 89 | 4279 | C-T | | | | | | |
| 90 | 4294 | C-T | | | | | | |
| 91 | 4315 | C-T | | | | | | |
| 94 | 4379 | C-T | | | | | | |
| 96 | 4489 | C-T | | | | | | |
| 97 | 4496 | A-G | | | | | | |
| 98 | 4513 | C-T | | | | | | |
| 105 | 4633 | C-T | | | | | | |
| 106 | 4663 | A-G | | | | | | |
| 108 | 4684 | A-G | | | | | | |
| 109 | 4696 | C-T | | | | | | |
| 114 | 4766 | C-T | | | | | | |
| 116 | 4820 | C-T | | | | | | |
| 117 | 4826 | C-T | | | | | | |
| 118 | 4873 | C-T | | | | | | |
| 120 | 4925 | C-T | | | | | | |
| 121 | 4947 | A-G | | | | | | |
| 125 | 5349 | C-T | | | | | | |
| 126 | 5410 | C-T | | | | | | |
| 127 | 5416 | | | A-G-N | | | | |
| 128 | 5482 | C-T | | | | | | |
| 129 | 5560 | C-T | | | | | | |
| 130 | 5590 | C-T | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 133 | 5650 | A-G | | | | | | |
| 134 | 5656 | A-G | | | | | | |
| 135 | 5674 | C-T | | | | | | |
| 136 | 5722 | A-G | | | | | | |
| 138 | 5791 | A-G | | | | | | |
| 139 | 5804 | A-G | | | | | | |
| 140 | 5869 | A-G | | | | | | |
| 141 | 6004 | C-T | | | | | | |
| 142 | 6037 | A-G | | | | | | |
| 144 | 6121 | C-T | | | | | | |
| 145 | 6184 | | | C-T-A | | | | |
| 146 | 6244 | | A-C | | | | | |
| 147 | 6373 | C-T | | | | | | |
| 148 | 6385 | C-T | | | | | | |
| 150 | 6478 | C-T | | | | | | |
| 151 | 6481 | C-T | | | | | | |
| 152 | 6514 | A-G | | | | | | |
| 154 | 6532 | C-T | | | | | | |
| 155 | 6598 | C-T | | | | | | |
| 156 | 6625 | A-G | | | | | | |
| 158 | 6682 | C-T | | | | | | |
| 159 | 6718 | C-T | | | | | | |
| 162 | 6805 | C-T | | | | | | |
| 163 | 6809 | A-G | | | | | | |
| 168 | 7047 | C-T | | | | | | |
| 170 | 7095 | C-T | | | | | | |
| 173 | 7266 | | | A-G-C | | | | |
| 176 | 7512 | A-G | | | | | | |
| 181 | 7696 | C-T | | | | | | |
| 182 | 7712 | C-T | | | | | | |
| 185 | 7822 | C-T | | | | | | |
| 189 | 7911 | | | A-G-C | | | | |
| 191 | 8001 | C-T | | | | | | |
| 192 | 8016 | A-G | | | | | | |
| 194 | 8064 | C-T | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 195 | 8157 | A-G | | | | | | |
| 198 | 8203 | A-G | | | | | | |
| 200 | 8280 | C-T | | | | | | |
| 201 | 8304 | A-G | | | | | | |
| 202 | 8309 | C-T | | | | | | |
| 204 | 8355 | A-G | | | | | | |
| 205 | 8361 | C-T | | | | | | |
| 206 | 8367 | C-T | | | | | | |
| 208 | 8382 | C-T | | | | | | |
| 210 | 8389 | A-G | | | | | | |
| 212 | 8410 | C-T | | | | | | |
| 216 | 8493 | C-T | | | | | | |
| 217 | 8530 | A-G | | | | | | |
| 219 | 8669 | C-T | | | | | | |
| 220 | 8672 | C-T | | | | | | |
| 221 | 8690 | A-G | | | | | | |
| 222 | 8714 | A-G | | | | | | |
| 224 | 8747 | A-G | | | | | | |
| 225 | 8801 | A-G | | | | | | |
| 228 | 8879 | C-T | | | | | | |
| 231 | 8987 | A-G | | | | | | |
| 232 | 9035 | A-G | | | | | | |
| 234 | 9056 | A-G | | | | | | |
| 235 | 9068 | A-G | | | | | | |
| 236 | 9086 | C-T | | | | | | |
| 238 | 9159 | C-T | | | | | | |
| 239 | 9165 | C-T | | | | | | |
| 243 | 9246 | C-T | | | | | | |
| 247 | 9339 | C-T | | | | | | |
| 249 | 9411 | C-T | | | | | | |
| 250 | 9429 | | | C-T-N | | | | |
| 251 | 9446 | A-G | | | | | | |
| 253 | 9482 | A-G | | | | | | |
| 255 | 9528 | | | C-T-N | | | | |
| 256 | 9555 | C-T | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 257 | 9558 | C-T | | | | | | |
| 261 | 9741 | A-G | | | | | | |
| 262 | 9750 | A-G | | | | | | |
| 264 | 9783 | C-T | | | | | | |
| 265 | 9790 | C-T | | | | | | |
| 267 | 9825 | A-G | | | | | | |
| 269 | 9915 | C-T | | | | | | |
| 271 | 9975 | C-T | | | | | | |
| 272 | 10029 | A-G | | | | | | |
| 273 | 10062 | C-T | | | | | | |
| 275 | 10131 | C-T | | | | | | |
| 276 | 10164 | C-T | | | | | | |
| 283 | 10388 | C-T | | | | | | |
| 285 | 10499 | C-T | | | | | | |
| 286 | 10541 | | | A-G-C | | | | |
| 290 | 10697 | C-T | | | | | | |
| 291 | 10724 | C-T | | | | | | |
| 292 | 10733 | A-G | | | | | | |
| 293 | 10746 | A-G | | | | | | |
| 295 | 10802 | A-G | | | | | | |
| 298 | 10880 | A-G | | | | | | |
| 299 | 10883 | A-G | | | | | | |
| 302 | 10976 | C-T | | | | | | |
| 303 | 10991 | A-G | | | | | | |
| 306 | 11175 | C-T | | | | | | |
| 308 | 11282 | A-G | | | | | | |
| 310 | 11312 | C-T | | | | | | |
| 311 | 11330 | | | A-G-N | | | | |
| 313 | 11357 | C-T | | | | | | |
| 315 | 11411 | | | | | | | |
| 316 | 11414 | A-G | | | | | | |
| 318 | 11420 | C-T | | | | | | |
| 319 | 11429 | A-G | | | | | | |
| 322 | 11480 | A-G | | | | | | |
| 327 | 11558 | C-T | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 330 | 11678 | C-T | | | | | | |
| 331 | 11724 | C-T | | | | | | |
| 332 | 11725 | C-T | | | | | | |
| 334 | 11840 | A-G | | | | | | |
| 335 | 11851 | C-T | | | | | | |
| 338 | 11957 | A-G | | | | | | |
| 339 | 11974 | A-G | | | | | | |
| 345 | 12158 | C-T | | | | | | |
| 348 | 12284 | C-T | | | | | | |
| 349 | 12296 | A-G | | | | | | |
| 350 | 12299 | C-T | | | | | | |
| 351 | 12368 | C-T | | | | | | |
| 354 | 12391 | A-G | | | | | | |
| 355 | 12398 | C-T | | | | | | |
| 356 | 12413 | A-G | | | | | | |
| 357 | 12416 | A-G | | | | | | |
| 359 | 12546 | A-G | | | | | | |
| 361 | 12650 | C-T | | | | | | |
| 362 | 12656 | C-T | | | | | | |
| 364 | 12683 | A-G | | | | | | |
| 365 | 12734 | C-T | | | | | | |
| 366 | 12797 | C-T | | | | | | |
| 368 | 12818 | C-T | | | | | | |
| 369 | 12830 | C-T | | | | | | |
| 373 | 12905 | G-C | | | | | | |
| 375 | 12926 | A-G | | | | | | |
| 376 | 12929 | C-T | | | | | | |
| 377 | 12941 | C-T | | | | | | |
| 380 | 13070 | A-G | | | | | | |
| 381 | 13161 | C-T | | | | | | |
| 382 | 13163 | A-G | | | | | | |
| 388 | 13217 | A-G | | | | | | |
| 389 | 13233 | C-T | | | | | | |
| 391 | 13292 | C-T | | | | | | |
| 394 | 13337 | C-T | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 395 | 13338 | C-T | | | | | | |
| 396 | 13346 | C-T | | | | | | |
| 397 | 13349 | C-T | | | | | | |
| 398 | 13368 | A-G | | | | | | |
| 399 | 13385 | A-G | | | | | | |
| 401 | 13421 | A-G | | | | | | |
| 402 | 13424 | C-T | | | | | | |
| 403 | 13430 | C-T | | | | | | |
| 404 | 13442 | | C-G | | | | | |
| 407 | 13496 | C-T | | | | | | |
| 408 | 13526 | A-G | | | | | | |
| 412 | 13590 | C-T | | | | | | |
| 414 | 13683 | A-G | | | | | | |
| 416 | 13749 | C-T | | | | | | |
| 418 | 13791 | C-T | | | | | | |
| 420 | 13829 | A-G | | | | | | |
| 422 | 13899 | | A-C | | | | | |
| 423 | 13965 | A-G | | | | | | |
| 428 | 14038 | A-G | | | | | | |
| 431 | 14077 | C-T | | | | | | |
| 432 | 14085 | C-T | | | | | | |
| 439 | 14323 | C-T | | | | | | |
| 440 | 14344 | C-T | | | | | | |
| 441 | 14356 | A-G | | | | | | |
| 443 | 14416 | C-T | | | | | | |
| 444 | 14434 | C-T | | | | | | |
| 446 | 14443 | C-T | | | | | | |
| 447 | 4485 | C-T | | | | | | |
| 448 | 14491 | C-T | | | | | | |
| 449 | 14512 | C-T | | | | | | |
| 450 | 14539 | C-T | | | | | | |
| 453 | 14638 | A-G | | | | | | |
| 454 | 14644 | C-T | | | | | | |
| 457 | 14674 | A-G | | | | | | |
| 462 | 14746 | C-T | | | | | | |
| 463 | 14752 | C-T | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 467 | 14909 | A-G | | | | | | |
| 468 | 14917 | A-G | | | | | | |
| 470 | 14944 | C-T | | | | | | |
| 471 | 14956 | C-T | | | | | | |
| 472 | 14972 | C-T | | | | | | |
| 474 | 15029 | C-T | | | | | | |
| 475 | 15059 | C-T | | | | | | |
| 476 | 15069 | C-T | | | | | | |
| 477 | 15074 | | | A-G-T | | | | |
| 478 | 15086 | A-G | | | | | | |
| 479 | 15093 | A-G | | | | | | |
| 480 | 15113 | C-T | | | | | | |
| 482 | 15140 | C-T | | | | | | |
| 483 | 15177 | A-G | | | | | | |
| 484 | 15188 | C-T | | | | | | |
| 488 | 15246 | A-G | | | | | | |
| 490 | 15293 | C-T | | | | | | |
| 491 | 15294 | A-G | | | | | | |
| 493 | 15299 | | | C-G-T | | | | |
| 494 | 15326 | A-G | | | | | | |
| 499 | 15413 | | A-T | | | | | |
| 500 | 15420 | C-T | | | | | | |

Table IIA- Table IIA summarizes the SNPs that were identified on major branches of the median joining tree generated in *Network.* Column 1 indicates the SNP number. Column 2 indicates the location of the SNP according to the human Cambridge reference sequence (Andrews *et al.* 2010). Column 3, 4 and 5 show if the polymorphism is a transition (Ts), transversion (Tv), or mixed state (more than two bases). Column 6 (red) shows polymorphisms that distinguish *P.t. ellioti* and *P.t. verus* from *P.t. schweinfurthii* and *P.t. trogolodytes*. Column 7 (blue) and 8 (pale purple) indicate base substitutions that are unique to *P.t. ellioti* and *P.t. verus* respectively. The final column (pale green) shows base changes that are not fixed.

Summary:

Total Fixed Base substitutions: Ts= 250; Tv = 4; Mixed= 13
Total Fixed Base Substitutions: Ts= 143; Tv=3; Mixed = 0
Column 6: Ts= 111 Tv=1
Column 7: Ts= 26 Tv= 1
Column 8: Ts= 6 Tv= 1