Infectious Agent Evolution in Bornean Orangutans

by

Michael John Charles Reid

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Anthropology
University of Toronto

© Copyright by Michael Reid 2016
Infectious Agent Evolution in Bornean Orangutans

Michael J. C. Reid

Doctor of Philosophy

Department of Anthropology
University of Toronto

2016

Abstract

In this thesis I provide among the first detailed studies of orangutan host/parasite co-evolution by bridging the disciplines of molecular virology/parasitology and biological anthropology. I used the tools of both disciplines to study the evolution of three infectious agents: simian foamy virus (SFV), simian T-lymphotropic virus 1 (STLV-1), and the protozoan parasite, Plasmodium, in orangutans Genus Pongo. Key to my thesis is the importance placed on the use of accurate calibration dates informed by archaeological and fossil data. Using this approach, I examine the evolutionary history of these infectious agents, including an estimation of the ages of divergence. The results suggest that SFV is the oldest of these infectious agents in orangutans, dating to >4 million years ago. My results challenge a previous hypothesis surrounding the appearance of a unique cladistic pattern for orangutan SFV based on the Mount Toba supervolcanic eruption at ~73 kya. The results of the present research suggest that the events leading to the development of this clade pattern is ~1.6 million years old, which rules out the Toba supereruption. My results also showed that there is a statistically significant level of host-virus coevolution in orangutans. My thesis research indicates the Plasmodium parasites in orangutans are much younger than previously thought. Finally, I estimate the origins of PTLV-1 in orangutans to between 7-343.7 kya. Overall, my thesis shows that the Pleistocene was an important period for the evolution of these infectious agents in orangutans and that it is at least possible that some of this evolution
occurred on the Indochinese mainland before orangutans reached the islands of Sumatra and Borneo.
Acknowledgments

As this PhD thesis and project comes to its end there are a great many people to thank for being a part of this process. I would first like to thank my academic father and PhD. supervisor Dr. Michael Schillaci without his belief in me and tireless work and support all of this would not be possible. I would also like to thank my thesis committee of Dr. Joyce Parga, and Dr. Shawn Lehman for their support and feedback on the papers contained herein. I would like to thank Dr. Mary Silcox for serving as my departmental external examiner and for providing edits and content suggestions for my thesis. Finally, I would like to thank Dr. Tony Goldberg for being my external examiner and providing thought provoking questions regarding my thesis.

I would like to thank each and everyone of the co-authors who contributed to this project. William M. Switzer, Dr. James Brooks, Amy R. Klegarth, Ellsworth Campbell, Manon Ragonnet-Cronin, Isabelle Joanisse, Kyna Caminiti, Dr. Carl Lowenberger, Dr. Birute Mary Galdikas, Dr. Hope Hollocher, and Dr. Paul Sandstrom without their contributions this project would not have been possible. They have taught me so much along this journey. I have also turned to a number of experts outside of the University of Toronto for expertise in particular subject areas including; Dr. David Bulbeck and Dr. Adam Brumm for their help with discussions of Sulawesi archaeology, Dr. David Burley on Oceanic prehistory and archaeology, Dr. Wallace Peters, Dr. David Warhurst, Dr. Severine Tasker and Dr. Emi Barker have also spent a great deal of time discussing orangutan plasmodia with me. I would also like to thank my fellow U of T graduate students especially Sandra Romain, Dejana Nikitovic, Monika Sumra, Erica Tennenhouse, Lianne Trip, Travis Steffens, Kerianne McGoogan, Laura Eastham, Kathy Pitirri and Laura Bolt.

This thesis project was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Ontario Graduate Scholarship (OGS) and various awards, fellowships and research grants from the Department of Anthropology.

My friends have all been a source of support and encouragement throughout this process. Each of you are in my life for a reason. I would like to especially thank Paul Dike, Ally Noormohamed, Dana Hamilton and the boys of MSG. I would also like to thank lifelong family friends who have encouraged me since childhood including Paul Griesbach, Teri and Bob
Herbert, and Tom and Pat Bradshaw. These family friends have contributed to my growth as a person for many years and encouraged me at every stage.

Kayla Russell has been at my side for much of this process, and has been my biggest cheerleader. She has been my rock. For that I will always be grateful. Your love and encouragement have been needed often through this process. Thank you. I am lucky to have you in my life.

Finally, I would like to thank my family as without them none of this would have ever been possible. First, my parents Robert and Charlotte Reid for the love, encouragement and support both mental and financial. My sisters Michelle Reid and Jennifer Reid who have continued to support, love and encouraging me along with their husbands Michael Pudden and Larry Downes who have provided friendship, guidance and support, their children Danica Pudden, McKenzie Downes, and Eric Downes whom I love.
# Table of Contents

## Chapter 1: The Role of Anthropologists in the Study of Orangutan (*Pongo* sp.) Infectious Agent Evolution

1.1 Introduction .......................................................................................................................... 2
1.2 Simian Foamy Virus in Orangutans .................................................................................. 3
1.3 Primate T-Cell Lymphotropic Virus in Orangutans ....................................................... 5
1.4 The Malaria Causing Parasites of Orangutans ............................................................... 6
1.5 Conclusion .......................................................................................................................... 7
1.6 References ......................................................................................................................... 8

## Chapter 2: Bayesian Inference Reveals Ancient Origin of Simian Foamy Virus in Orangutans

2.1 Abstract .............................................................................................................................. 12
2.2 Introduction ......................................................................................................................... 12
2.2.1 SFV in Orangutans ........................................................................................................ 13
2.3 Methodology ...................................................................................................................... 15
2.3.1 Study population and Specimen Collection and Processing ...................................... 15
2.3.2 PCR amplification of longer SFV fragments from PCR-positive individuals .............. 18
2.3.3 *Pongo* sp. mitochondrial COX2 amplification and sequencing ................................. 19
2.3.4 Sequence analyses ......................................................................................................... 20
2.4 Results .................................................................................................................................. 22
2.4.1 Co-evolution of orangutan SFV *pol* and mtDNA sequences ................................... 24
Table 2.1 SFV polymerase (pol) and mitochondrial cytochrome oxidase subunit II (COX2) sequences used in the study. ................................................................. 30

Table 2.2 Maximum composite likelihood genetic distances for Pongo sp. simian foamy virus and mitochondrial sequences. The SFV pol and COX2 distances are above and below the diagonal, respectively. ................................................................. 32

Table 2.3 Bayesian estimated divergence dates for time to most recent common ancestor (TMRCA) in millions of years (mya) for the Pongo sp. simian foamy virus and other ape clades ................................................................. 33

Table 2.4 Genetic estimates for time to most recent common ancestor (TMRCA) of the genus Pongo, and the split dates for P. abelii and P. pygmaeus in millions of years (mya) ................................................................. 34

Table 2.5 Early to Middle Pleistocene fossil localities with well-dated fossil evidence of orangutans .................................................................................................................. 35

Figure 2.1 Maximum likelihood phylogeny of ape SFV polymerase (pol) sequences (423-bp) .................................................................................................................. 37

Figure 2.2 Bayesian-inferred phylogeny using 423-bp polymerase (pol) sequences ......... 38

Figure 2.3 ........................................................................................................................................ 39

Figure 2.3a Phylogenetic relationships of Pongo sp. SFV polymerase (pol) (425-bp) ... 39

Figure 2.3b Phylogenetic relationships of Pongo sp. mitochondrial cytochrome oxidase subunit II (COX2) (546-bp) .................................................................................. 40

Figure 2.4 Least-squares regression analysis of SFV (pol) and orangutan cytochrome oxidase subunit II (COX2) ................................................................................. 41

Chapter 3: Detailed phylogenetic analysis of Primate T-lymphotropic virus type 1 (PTLV-1) sequences from Orangutans (Pongo pygmaeus) reveals new insights into the evolutionary history of PTLV-1 in Asia .................................................................................. 54
3.1 Abstract ................................................................................................................................. 55
3.2 Introduction .......................................................................................................................... 56
3.3 Methods ............................................................................................................................... 59
  3.3.1 Study population and specimen collection ................................................................. 59
  3.3.2 Elution of genomic DNA from FTA® cards ................................................................. 60
  3.3.3 PCR detection of STLV-1 ............................................................................................ 61
  3.3.4 Phylogenetic analyses ................................................................................................. 63
  3.3.5 Bayesian Divergence Dating Models Tested ............................................................... 63
3.4 Results ................................................................................................................................. 67
  3.4.1 Modified Classical Archaeological calibration model (MCA) phylogeny .................. 68
  3.4.2 Most Likely Archaeological calibration (MLA) model phylogeny ............................ 69
  3.4.3 Pedigree Calibration Model (PCM) topology ............................................................. 69
  3.4.4 Divergence Dates inferred from the Bayesian analyses ............................................. 70
3.5 Discussion ............................................................................................................................ 71
  3.5.1 Paleobiogeographical and environmental context for PTLV evolution in Asia ............ 75
  3.5.2 Possible first introduction of HTLV-1 in Asia ............................................................ 79
3.6 Conclusions ........................................................................................................................ 80
3.7 Tables ................................................................................................................................... 82
  Table 3.1 Primate T-lymphotropic virus sequences used in this study ......................... 82
  Table 3.2 Time to most recent common ancestor (TMRCA) inferred from Bayesian analysis of PTLV-1 long terminal repeat (LTR) sequences 1 ........................................... 86
  Table 3.3 Middle/Late Pleistocene and Early Holocene fossil localities with well-dated fossil evidence of orangutans (Pongo sp.) ................................................................. 87
3.8 Figure Legends .................................................................................................................... 89
3.9 Figures ................................................................................................................................ 90
  Figure 3.1 Maximum likelihood PTLV-1 tax tree ............................................................ 90
  Figure 3.2 Maximum likelihood PTLV-1 LTR tree ............................................................ 91
Figure 3.3 Modified Classical Archaeological (MCA Model) Calibration PTLV-1 LTR tree .............................................................. 92

Figure 3.4 Most Likely Archaeological Calibration (MLA Model) PTLV-1 LTR tree ........ 93

Figure 3.5 Pedigree (PCM Model) Calibration PTLV-1 LTR tree ........................................ 94

Chapter 4: Plasmodia Evolution in Orangutans ................................................................. 110

4.1 Abstract ......................................................................................................................... 111

4.2 Introduction .................................................................................................................... 112

4.2.1 History of research on Orangutan plasmodia ............................................................ 112

4.3 Methodology .................................................................................................................. 120

4.3.1 Bayesian Divergence Dating Model Tested ............................................................... 120

4.4 Results ............................................................................................................................ 122

4.4.1 Bayesian Analyses ................................................................................................. 122

4.4.2 Maximum Parsimony ............................................................................................. 123

4.4.3 Bayesian Dates (TMRCA) ...................................................................................... 124

4.5 Discussion ....................................................................................................................... 124

4.5.1 Phylogenetic Tree Comparisons ............................................................................. 124

4.5.2 Bayesian Date Comparisons ................................................................................... 126

4.5.3 Paleobiogeographical context for the evolution of *Plasmodium* in orangutans ...... 129

4.6 Conclusion ...................................................................................................................... 132

4.7 Tables ............................................................................................................................... 133

Table 4.1 *Plasmodium* mitochondrial DNA sequences used in this study and Genbank accession numbers ....................................................................................................................... 133

Table 4.2 Estimated divergence dates (TMRCA) for major *Plasmodium* clades ............ 136

4.8 Figure Legends ............................................................................................................... 138

4.9 Figures ............................................................................................................................ 140
Figure 4.1 *Plasmodium* phylogeny inferred by Bayesian analysis of 136 complete mtDNA sequences showing only the Asian *Plasmodium* species ................................................. 140

Figure 4.2 Maximum Parsimony (MP) tree of 136 complete mtDNA sequences of the *Plasmodium* parasite showing only the Asian *Plasmodium* species ................................................. 141

Figure S4.1 Complete *Plasmodium* phylogeny inferred by Bayesian analysis of 136 complete mtDNA sequences .......................................................................................... 142

Figure S4.2 *Plasmodium* sp. Bayesian showing only the individual sequences that make up the *P. inui* and *Pongo* 2 clades ................................................................................ 143

Figure S4.3 Complete Maximum Parsimony (MP) tree using the complete mtDNA genome of *Plasmodium* parasites ................................................................................ 144

Figure S4.4 Maximum Parsimony (MP) tree only showing the individual sequences that make up the *P. inui* and *Pongo* 2 clades ........................................................................ 145

4.10 References .................................................................................................................. 146

Chapter 5: Conclusions and Future Directions .................................................................. 155

5.1 Introduction .................................................................................................................. 156

5.2 Summary of Results and Implications ........................................................................ 156

5.3 Future Research .......................................................................................................... 160

5.4 Conclusions ................................................................................................................. 162

5.5 References .................................................................................................................. 163
List of Tables

Table 2.1 SFV polymerase (pol) and mitochondrial cytochrome oxidase subunit II (COX2) sequences used in the study. ................................................................. 30

Table 2.2 Maximum composite likelihood genetic distances for Pongo sp. simian foamy virus and mitochondrial sequences. The SFV pol and COX2 distances are above and below the diagonal, respectively......................................................... 32

Table 2.3 Bayesian estimated divergence dates for time to most recent common ancestor (TMRCA) in millions of years (mya) for the Pongo sp. simian foamy virus and other ape clades .................................................................................. 33

Table 2.4 Genetic estimates for time to most recent common ancestor (TMRCA) of the genus Pongo, and the split dates for P. abelii and P. pygmaeus in millions of years (mya) ................................................................. 34

Table 2.5 Early to Middle Pleistocene fossil localities with well-dated fossil evidence of orangutans ........................................................................................................... 35

Table 3.1 Primate T-lymphotropic virus sequences used in this study. ......................... 82

Table 3.2 Time to most recent common ancestor (TMRCA) inferred from Bayesian analysis of PTLV-1 long terminal repeat (LTR) sequences. ............................................. 86

Table 3.3 Middle/Late Pleistocene and Early Holocene fossil localities with well-dated fossil evidence of orangutans (Pongo sp.) ........................................................................ 87

Table 4.1 Plasmodium mitochondrial DNA sequences used in this study and Genbank Accession Numbers ................................................................. 133

Table 4.2 Estimated divergence dates (TMRCA) for major Plasmodium clades .......... 136
List of Figures

Figure 2.1 Maximum likelihood phylogeny of ape SFV polymerase (pol) sequences (423-bp). 37

Figure 2.2 Bayesian-inferred phylogeny using 423-bp polymerase (pol) sequences .............. 38

Figure 2.3 Phyllogenetic relationships of Pongo sp. SFV polymerase (pol) (425-bp) ............ 39

Figure 2.3a Phylogenic relationships of Pongo sp. SFV polymerase (pol) (425-bp) ............ 39

Figure 2.3b Phylogenetic relationships of Pongo sp. mitochondrial cytochrome oxidase subunit II (COX2) (546-bp) .................................................................................................................. 40

Figure 2.4 Least-squares regression analysis of SFV (pol) and orangutan cytochrome oxidase subunit II (COX2) .................................................................................................................. 41

Figure 3.1 Maximum likelihood PTLV-1 tax tree .................................................................... 90

Figure 3.2 Maximum likelihood PTLV-1 LTR tree .................................................................. 91

Figure 3.3 Modified Classical Archaeological (MCA Model) Calibration PTLV-1 LTR tree ... 92

Figure 3.4 Most Likely Archaeological Calibration (MLA Model) PTLV-1 LTR tree .......... 93

Figure 3.5 Pedigree (PCM Model) Calibration PTLV-1 LTR tree ........................................ 94

Figure 4.1 Plasmodium phylogeny inferred by Bayesian analysis of 136 complete mtDNA sequences showing only the Asian Plasmodium species ................................................................. 140

Figure 4.2 Maximum Parsimony (MP) tree of 136 complete mtDNA sequences of the Plasmodium parasite showing only the Asian Plasmodium species ................................................................. 141

Figure S4.1 Complete Plasmodium phylogeny inferred by Bayesian analysis of 136 complete mtDNA sequences .......................................................................................................................... 142

Figure S4.2 Plasmodium sp. Bayesian showing only the individual sequences that make up the P. inui and Pongo 2 clades ........................................................................................................... 143

Figure S4.3 Complete Maximum Parsimony (MP) tree using the complete mtDNA genome of Plasmodium parasites. ............................................................................................................. 144

Figure S4.4 Maximum Parsimony (MP) tree only showing the individual sequences that make up the P. inui and Pongo 2 clades. ............................................................................................................. 145
Chapter 1: The Role of Anthropologists in the Study of Orangutan (*Pongo* sp.) Infectious Agent Evolution
1.1 Introduction

Traditionally, much of the research on the infectious agents of non-human primates (NHP) has attempted to answer questions surrounding human medicine, veterinary medicine, and ecosystem health. Virologists and parasitologists have also asked basic questions regarding the evolutionary relationships of infectious agents. These questions have included, 1) how old is the origin of a given infectious agent? 2) in which primate species did this infectious agent originate? and, 3) how did this infectious agent jump from NHPs to humans? For primatologists trained in anthropology, the traditional questions regarding infectious agents tend to be related to ecology or conservation (Chapman et al., 2005; Howells et al., 2011; Kowalzik et al., 2010). Surprisingly, anthropologists working with non-human primate infectious agents have generally not been concerned with evolutionary questions or with host/parasite co-evolution (but see Nunn and Altizer, 2006). Considering that one of the traditional essential objectives of biological anthropology is to examine and understand evolutionary processes on a greater scale than most other disciplines, it would seem to be the ideal discipline to combine with new molecular virology techniques to study host/parasite co-evolution in NHP. Yet, to date very little research has attempted to bridge the fields of virology/parasitology and anthropological primatology, to answer questions regarding the evolution and coevolution of infectious agents and their NHP hosts by using the primate fossil record, paleobiogeography, geology, and even archaeology.

This project is unique in that it studies host/parasite co-evolution by bridging the disciplines of molecular virology/parasitology and biological anthropology. I used the tools of both disciplines to study the evolution of three infectious agents by estimating the evolutionary history of these infectious agents, including estimating ages of divergence using published research on the fossil record. This approach has the potential to provide information regarding
not only the evolutionary history of infectious agents, but also that of the NHP hosts. In addition to the fossil record, I used paleobiogeography and archaeology to better understand and reconstruct the evolutionary history of these infectious agents. Until recently, this approach has not often been used in a comprehensive way by molecular virologists/parasitologists interested in primate infectious agents. The use of biogeographic data for dating the origins of primate retroviruses (Worobey et al., 2010), or explaining unlikely phylogenetic relationships (Katzourakis et al., 2014) has become more common as of late. But, has not yet become the norm. This research focuses on the evolution of two retroviruses, simian foamy virus (SFV), and simian T-lymphotropic virus 1 (STLV-1), and the protozoan parasite that causes malaria (*Plasmodium* sp.) in orangutans. Orangutans are ideal hosts to study because in comparison to the African great apes, their fossil history over the past two million years is fairly well known, and they are the only surviving great ape in Asia.

### 1.2 Simian Foamy Virus in Orangutans

The first paper (Chapter 2) of my thesis examines simian foamy virus (SFV) in orangutans. To date only three published studies have looked at orangutan SFV. In 1994, (McClure et al., 1994) published the first discovery of SFV in an orangutan. Ten years later a team led by Ernst Verschoor published two articles on SFV in orangutans (Verschoor et al., 2003; 2004). Verschoor et al. (2003) were the first to use phylogenetics to study the relatedness of SFV variants in orangutans. Verschoor et al. (2004) found a unique phylogenetic pattern during their analysis of the orangutan SFV *pol* gene, with three distinct clades apparent. The first clade comprised SFV sequences from orangutans found only on the island of Borneo (*Pongo pygmaeus*), while the second contained SFV sequences from orangutans only found on the island of Sumatra (*Pongo abelii*). The third clade contained SFV sequences from both Bornean and
Sumatran orangutans. The obvious question was whether recent cross-species transmission of SFV from Bornean to Sumatran orangutans had occurred in captive settings such as zoos. Each of the three clades contained sequences from completely wild or rehabilitant orangutans housed on the islands of Borneo and Sumatra. This led Verschoor et al. (2004) to hypothesize that this phylogenetic pattern was the result of paleobiogeographic influences on ancestral (fossil) orangutans. The authors hypothesized that the eruption of the Mount Toba Volcano on the island of Sumatra at around 74 kya severely reduced the orangutan population on the island. After the orangutan population collapsed on the island of Sumatra, Verschoor et al. (2004) believed that the lowered Pleistocene sea levels forming the Sunda Shelf allowed Bornean orangutans carrying their SFV to move onto the island of Sumatra, increasing the orangutan population and introducing Bornean SFV to Sumatra creating this mixed clade.

In 2005, Switzer et al. (2005) published an article on the co-evolutionary relationship between SFV and non-human primates (NHP) hosts by looking at the degree of co-evolution between NHP mtDNA COII and the SFV pol genes. What they found was that primates and foamy viruses had been co-evolving for millions of years. This finding explained why Verschoor et al. (2003 and 2004) found two primary SFV clades in orangutans, one from Bornean orangutans and one from Sumatran orangutans. This left me with two major objectives for my doctoral research with regard to SFV in orangutans. The first concerned the date of the origins of the orangutan SFV clades. This would allow me to determine whether the unique tri-clade phylogenetic pattern of SFV in orangutans could have been the result of the eruption of Mt. Toba and subsequent influx of Bornean orangutans into Sumatra. My second objective was to examine the degree of co-evolution between orangutan mtDNA (COII gene) and SFV (pol gene).
1.3 Primate T-Cell Lymphotropic Virus in Orangutans

My second paper (Chapter 3) focuses on another primate retrovirus called simian T-lymphotropic virus 1 (STLV-1). STLV-1 was first identified in captive orangutans (Pongo sp.) at Indonesian zoos on the island of Java (Ibuki et al., 1997). As part of that study, a segment of the STLV-1 long terminal repeat (LTR) region was sequenced and compared to other primate STLV-1 sequences. It was found that orangutan STLV-1 was closely related to an STLV-1 from M. tonkeana (Ibuki et al., 1997). A 1998 study by Verschoor et al. (1998) was able to isolate and obtain sequences from the STLV-1 tax/ rex region in blood samples obtained from a wild-caught orangutan. This sequence showed a close evolutionary relationship with the STLV-1 sequences from macaques (M. mulatta and M. nemestrina). These genetic relationships were subsequently confirmed in a separate study using sequences from the envelope (env) and LTR regions by Van Dooren et al. (2007).

Prior to my study, none had attempted to date the origin of PTLV-1 in orangutans. In order to properly establish calibration points for dating the evolution and origin of this virus in orangutans, I noticed potential problems with the way previous studies had set these calibrations (e.g. Andonov et al., 2012; Lemey et al. 2005; Van Dooren et al. 2001). By carefully evaluating the anthropological and archaeological literature I was able to establish more accurate calibration points needed for dating other nodes within the phylogenetic tree describing evolutionary relationships. Of particular importance to my study was the establishment of an accurate calibration for the node describing the common ancestor of PTLV-1 in humans (HTLV-1). Previous studies used an HTLV-1 sequence from Melanesia (Solomon Islands) called MEL5 (GenBank accession number L20534) as a calibration point set to 50-60 kya as the origin of HTLV-1. Using the archaeological literature, I discovered that the earliest possible dates for
human occupation of the Solomon Islands is 32 kya (Sheppard, 2011). Further research of the literature suggested that it is more likely that HTLV-1 was introduced from monkeys to humans and spread during the Neolithic expansion in Southeast Asia, which is believed to have begun in Taiwan 12-5 kya according to human genetic and linguistic data (Gray et al., 2009; Kayser et al., 2006). As a result of this anthropological approach I developed three models for the origins of STLV-1 in orangutans and HTLV-1 in humans. The first model was a modified version of the previously established approach which sets a calibration date of 50-60 kya for the node representing the most recent common ancestor (MRCA) of Solomon Islands HTLV-1. For my modified version I set the calibration at 45 to 55 kya for the MRCA of the PNG-1/Australia HTLV-1 clade which included a sequence from Papua New Guinea, where there is some evidence for an early occupation between 44 and 49 kya (Summerhayes et al., 2010). The second model used the same node but I modified the dates to coincide with the Austronesian expansion by setting the calibration node to 3-8 kya. The final model used an evolutionary rate (mutation rate) for the LTR region inferred by Van Dooren et al. (2004) based on a study of HTLV-1 pedigree data. My study is the first to use multiple models to test the date of origin for STLV-1 in orangutans and HTLV-1 in humans, while also exploring the biogeographic contexts associated with the origins of PTLV-1 in humans and orangutans.

1.4 The Malaria Causing Parasites of Orangutans

The final paper of my thesis (Chapter 4) is a re-examination of the dates surrounding the origin of the malaria causing parasites (Plasmodium sp.) in orangutans. In general, the orangutan plasmodia are much better known to science, with a greater number of articles discussing their discovery, lifecycle and microbiology than either SFV or STLV-1. There are five publications that have used genetic data to look at phylogenetic relationships and have attempted to date the
origins of the parasite in orangutans. What is clear from these articles is that based on multiple
genes, including the entire mtDNA genome, orangutan plasmodia are closely related to Asian
plasmodia found in macaques, leaf monkeys, gibbons, and humans (*P. vivax*). It also appears that
this group of parasites came to Asia with an African ancestor (Muehlenbein et al., 2015) likely
with the dispersal of macaques from Africa and Eurasia (Alba et al., 2014; Liedigk et al., 2014;
Stewart and Disotell, 1998).

Two previous studies which attempted to date the origins of plasmodia in orangutans
used detailed methodologies to calibrate their dating models. Nonetheless, as discussed in my
third paper, important changes to the calibration models were needed in order to obtain more
accurate estimates for the divergence of orangutan malaria parasites. Again, I use
anthropological data to help set calibration points to better estimate divergence dates and
decrease potential error while reducing confidence intervals.

1.5 Conclusion

The three papers of my doctoral thesis have contributed something unique to the study of
infectious agents in nonhuman primates. While virologists and molecular parasitologists have
traditionally used fossil data to help set calibration points in their research of infectious agent
evolution, very few have used multiple sources of data from fields such as archaeology,
linguistics, human evolutionary genetics, bioarchaeology, paleontology, and paleobiogeography
to create a more complete picture of the evolution of these organisms. My thesis research not
only used such data sets, but also examined how the evolution of these infectious agents
corresponded to what is known regarding the evolution of the primate host organism. My
research has highlighted the importance of taking a more holistic and evolutionary approach to
the study of host and parasite evolution in nonhuman primates.
1.6 References


Chapter 2: Bayesian Inference Reveals Ancient Origin of Simian Foamy Virus in Orangutans

**Author Contributions:** Conceived and designed the study: Michael J.C. Reid, Michael A. Schillaci, James I. Brooks. Performed the experiments: Michael J.C. Reid, William M. Switzer, Michael A. Schillaci, James I. Brooks, Amy R. Klegarth, Manon Ragonnet, Isabelle Joanisse, Kyna Caminiti, Hope Hollocher. Analyzed the data: Michael J.C. Reid, Michael A. Schillaci, William M. Switzer, Ellsworth Campbell. Contributed reagents/materials/analysis tools: Amy R. Klegarth, Manon Ragonnet-Cronin, Isabelle Joanisse, Kyna Caminiti, Carl A. Lowenberger, Birute Mary F. Galdikas, Hope Hollocher, Paul A. Sandstrom. Wrote the paper Michael J.C. Reid, Michael A. Schillaci, William M. Switzer, James I. Brooks.
2.1 Abstract
Simian foamy virus (SFV) infects most nonhuman primate species and appears to co-evolve with its hosts. This co-evolutionary signal is particularly strong among great apes, including orangutans (Genus *Pongo*). Previous studies have identified three distinct orangutan SFV clades. The first clade is composed of SFV from *P. abelii* from Sumatra, the second consists of SFV from *P. pygmaeus* from Borneo, while the third clade is mixed, comprising SFV from both species. The existence of the mixed clade has been attributed to an expansion of *P. pygmaeus* into Sumatra following the Mount Toba super-volcanic eruption about 73,000 years ago. Divergence dating however, has not been performed to establish a temporal association with the Toba eruption. Here, we use a Bayesian framework and a relaxed molecular clock model with fossil calibrations to test the Toba hypothesis and to gain a more complete understanding of the evolutionary history of orangutan SFV. Our results show a similar three-clade orangutan SFV phylogeny as with previous studies, along with strong statistical support for SFV-host co-evolution in orangutans. Using Bayesian inference, we date the origin of orangutan SFV at >4 million years ago (mya), while the mixed species clade dates to approximately 1.6 mya, more than 1.5 million years older than the Toba super-eruption. These results, combined with fossil and paleogeographic evidence, suggest that the origin of SFV in Sumatran and Bornean orangutans, including the mixed species clade, likely occurred on the mainland of Indo-China during the Late Pliocene and Calabrian stage of the Pleistocene, respectively.

2.2 Introduction
Retroviridae are characterized by an RNA genome that is reverse transcribed into DNA prior to integration into the host genome (Murphy et al., 2006). Spumavirinae, or foamy viruses (FV), including simian foamy virus (SFV), are unique among retroviruses by having a late-
occurring reverse transcription step allowing for packaging of both DNA and RNA into virions, unlike other genera that package only RNA (Voevodin and Marx, 2009). SFV has received significant public health interest because of the potential ease of transmission to workers handling nonhuman primates (NHPs) in research facilities, and at zoos (Brooks et al., 2002; Heneine et al., 1998; Huang et al., 2012; Murphy et al., 2006; Sandstrom et al., 2000; Stenback et al., 2014; Switzer et al., 2004), the transmission of SFV through hunting and butchering of NHPs in the wild (Betsem et al., 2011; Calattini et al., 2007; 2011; Gessain et al., 2013; Mouinga-Ondémé and Kazanji 2013; Smith et al., 2012; Switzer et al., 2012; Wolfe et al., 2004), zoonotic transmission to tourists or those living in close association with free-ranging NHPs and performing monkeys in Asia (Engel et al., 2013; Jones-Engel et al, 2005, 2008; Schillaci et al., 2005), and the potential threat of SFV to the safety of the blood supply from infected persons who donate blood (Boneva et al., 2002; Brooks et al., 2007, Schillaci et al., 2008).

SFV has been of scientific interest given the strong evidence of ancient co-evolution within and among host species (e.g. Switzer et al., 2005). Switzer et al. (2005) found a tight agreement between primate-host mitochondrial (mtDNA) and SFV evolutionary histories across different Old World anthropoid primates over millions of years, providing the strongest evidence thus far for primate host/virus co-evolution, including a positive correlation between the host and virus phylogenies, with approximately 72% of the variation in viral branch lengths attributable to variation in host branch lengths. SFV has also been shown to have co-evolved at the sub-species level in NHPs, including chimpanzees (Liu et al., 2008; Switzer et al., 2004).

2.2.1 SFV in Orangutans

Orangutans, currently found on the islands of Borneo and Sumatra, are the only remaining great apes in Asia. There are currently two widely recognized species, the Sumatran
orangutan (*P. abelii*), and the Bornean orangutan (*P. pygmaeus*). The Bornean orangutan is further subdivided into the Northwest Bornean (*P. p. pygmaeus*), Northeast Bornean (*P. p. morio*) and Central Bornean (*P. p. wurmbii*) subspecies (Groves, 2001; Singleton et al., 2004). Fossil remains from the late Pliocene through the early Holocene indicate that orangutans once ranged over much of Southeast Asia from China and India to the Sunda Islands of Sumatra, Java and Borneo (Bacon et al., 2008, 2011; Bae, 2010; Drawhorn, 1995; Esposito et al., 2002; Harrison et al., 2006, 2014; Hooijer, 1948; Huang, 1979; Ibrahim et al., 2013; Jones et al., 2004; Kahlke, 1972; Long et al., 1996; Rijksen and Meijaard, 1999; Rink, 2008; Schwartz et al. 1994, 1995; Shen et al., 2002; Takai et al., 2014; Zeitoun et al., 2010).

SFV in orangutans was first described in 1994 and named SFV-11 following the serotyping nomenclature used at that time (McClure et al., 1994). Since then, the complete genome of SFV-11 (renamed SFVora-GenBank Number AJ544579) has been obtained (Verschoor et al., 2003) and used to study orangutan host and SFV evolutionary histories (Verschoor et al., 2004). Orangutan SFV was reported to exist in three independent phylogenetic clades; the first two clades contained strains from either Bornean or Sumatran orangutan species likely reflecting host speciation, while the third clade consisted of SFV from both Bornean and Sumatran orangutans, indicative of cross-species transmission between both orangutan species (Verschoor et al., 2004). Wild-born orangutans were present in each clade indicating that the potential cross-species transmission likely resulted in nature and not during captivity. Hence, Verschoor et al. (2004) suggested that this mixed SFV clade potentially reflected a recent movement of orangutans from what is now the island of Borneo back into Sumatra after the supervolcanic eruption of Mount Toba around 73 kya. Verschoor et al. (2004), however, did not utilize the available SFV sequence data to infer divergence dates for the orangutan SFV clades to test their Toba hypothesis. Furthermore, despite the enormity of the Toba super-eruption, current
evidence suggests that it was not necessarily an extinction level event for Sumatran wildlife. For example, on Sumatra only one mammalian species, a bovid (*Bos javanicus*), appears to have gone extinct around the time of the eruption (Louys, 2007). Sumatra is also home to a number of species of mammals that are found in other parts of mainland Southeast Asia but not on other Indonesian islands (Whitten et al., 2000), including relatively large species such as the tapir (*Tapirus indicus*), the mountain goat (*Capricornis sumatraensis*), and the Asian golden cat (*Felis temmincki*), along with primate species such as the white-handed gibbon (*Hylobates lar*), the siamang (*Symphalangus syndactylus*), and the banded langur (*Presbytis femoralis*) (Whitten et al., 2000). These results suggest that even in the ancient past, barriers were present that limited the flow of mammalian species between Borneo and Sumatra, and that the Mount Toba supereruption may not have been as ecologically destructive as some have suggested (Williams, 2012; Williams et al., 2009).

The objective of our study was to explore Verschoor et al.’s (2004) three clade-Toba supereruption hypothesis through detailed phylogenetic analyses of new and published orangutan SFV sequences. In addition, we used orangutan mtDNA cytochrome oxidase subunit II (COX2) sequences to further examine SFV/host co-evolution in these three clades. The results of these analyses are interpreted within a paleobiogeographical context inferred from the fossil record and sea level fluctuations linked to glaciation.

### 2.3 Methodology

#### 2.3.1 Study population and Specimen Collection and Processing

Blood samples were collected in 2003 from 68 (35 females, 33 males) rehabilitant orangutans (*Pongo pygmaeus*) housed at the Orangutan Care Center and Quarantine (OCC&Q)
in the village of Pasir Panjang in the Central Kalimantan Province of Indonesian Borneo. Blood samples were opportunistically collected as part of semi-annual medical exams, or as part of routine quarantine screenings done on newly arrived wild born orangutans who generally arrive as orphans, sick or injured animals. Orangutan blood was drawn by OCC&Q veterinary staff using a 25-gauge × 1 5/8-inch Precision Glide needle (Becton, Dickinson and Company, Oakville, Ontario, Canada) and a 3-mL syringe. Small aliquots of blood were pipetted onto Whatman FTA® Classic Cards (Whatman Inc., Florham Park, NJ, USA) and dried overnight at room temperature for transport to Canada (Reid et al., 2006). The research protocol was approved by Simon Fraser University's Animal Ethics Review Committee (UACC # 675B-03), and the required Convention on International Trade in Endangered Species (CITES) permits (08692/IV/SATS-LN/2004) were obtained for the transportation of blood samples. Serological testing for SFV was not performed on the FTA blood samples due to the limited amount of specimen.

Following the manufacturer’s instructions genomic DNA (gDNA) was eluted from the FTA cards spotted with orangutan blood. Briefly, using a 1.2 mm Harris Uni-Core punch, 3 discs were punched from each dried blood spot (DBS) and ejected into a sterile 1.5 mL microcentrifuge tube. One punch from an unused FTA® card was taken between each sample to prevent cross-contamination of specimens. Discs were washed prior to gDNA elution via incubation at room temperature with 3 x 200 µL FTA® Purification Reagent (Whatman International Limited) for 5 minutes each with wash buffer removed after each wash using a 200µl pipette (Mandel Scientific, Guelph, Canada). Then the discs were rinsed in 2 x 200 µL TE⁻¹ buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) as above for an additional 5 minutes each. To elute gDNA from the discs, samples were incubated with 35 µL of Solution 1 (0.1 N NaOH, 0.3 mM EDTA, pH 13.0) for 5 minutes at room temperature. 65 µL of Solution 2 (0.1 M Tris-HCl,
pH 7.0) was added to the eluate and the samples were pulse vortexed 5 times to ensure proper mixing. Samples were allowed to incubate at room temperature for 10 minutes, after which they were pulse vortexed an additional 10 times. Punches were removed from each sample and squeezed with tweezers to recover a maximum volume of eluate, which contained gDNA in TE buffer (66 mM Tris-HCl, 0.1 mM EDTA).

SFV polymerase (pol) sequences were amplified by PCR and then detected using a nested real-time (RT) PCR assay. 10µL of extracted gDNA was used in the standard PCR test to generate a 465-bp pol sequence with the primers MS3.pongo.fwd (nt 5817-5836) (5’ CCT GGA TGC AGA GTT GGA TC 3’), and MS4.pongo.rev (nt 6262-6281) (5’ GAG GGA GCC TTT GTG GGG TA 3’) for 95°C for 10 minutes, followed by 35 cycles of 20 seconds at 94°C, 30 seconds at 51°C, and 1 minute at 72°C, with a final extension at 72°C for 10 minutes. All reactions were performed in an ABI GeneAmp PCR System 9700 thermocycler (Applied Biosystems).

For RT-PCR, five µL of first round amplification product was used as template. The RT-PCR primer and probes MS5.pongo.fwd (nt 5995-5015) (5’ ASY GGA AGA GAR GCT ACI TTR 3’), MS6.pongo.rev (nt 6049-6073) (5’ TGA CTT TAA CCA CAT CCT TIC GCA TA 3’), and SFV.Pongo.P (nt 6017-6046) (5’ FAM-TAA AAC TGT CAA ACA YTT ATT GGT GGC CYA-MGBNFQ 3’) were designed within a conserved area of the pol gene (102-bp) of SFVora (GenBank accession number AJ544579) (Verschoor et al., 2003) using Primer Express (Applied Biosystems). RT-PCR was done using an ABI Prism 7000 sequence detection system (Applied Biosystems). Cycling conditions included a single cycle of 15 minutes at 95°C to activate the HotStarTaq® DNA polymerase (Quantitect Probe PCR Kit, Qiagen, Mississauga, Ontario, Canada), followed by 45 cycles of 94°C for 15 seconds, and 1 minute at 60°C. Standard
curves for SFVora consisted of 10-fold serial dilutions of a linearized plasmid (pCR2.1-TOPO; Invitrogen, Burlington, Ontario, Canada) containing a cloned fragment of the pol gene. Both PCR assays could reliably detect 10 copies of target DNA in a background of 1µg of SFV-negative macaque (Macaca fascicularis) genomic DNA.

2.3.2 PCR amplification of longer SFV fragments from PCR-positive individuals

10 µL of orangutan gDNA was used as template for PCR amplification of longer SFV pol sequences (878-bp) for the phylogenetic analyses. Following an initial hot-start incubation of 15 minutes at 95ºC for activation of Taq DNA polymerase (Applied Biosystems, Streetsville, Ontario, Canada), the following cycling conditions were used for the amplification: denaturation for 20 seconds at 94ºC, annealing for 1 minute at 56ºC and extension for 2 minutes 30 seconds at 72ºC. The reaction was cycled 45 times and was completed with a final hold at 72ºC for 10 minutes. All reactions were performed in an ABI GeneAmp PCR System 9700 thermocycler (Applied Biosystems). The following primer sets were used for orangutan SFV pol amplification: MS3.outer.pongo.fwd (nt 58117-5836) (5’ CCT GGA TGC AGA GTT GGA TC 3’) / SFVoraczp-1.rev (nt 6675-6694) (5’CAC GAA TTT CCT GTA AAA AGA 3’). PCR products were separated by electrophoresis gel. PCR products containing only a single band were sequenced directly, while PCR products containing multiple bands had the band of the correct size excised and purified using QIAquick Gel Extraction kit (QIAGEN) and sequenced. Sequencing was performed on a capillary sequencer (ABI 3130, Applied Biosystems) with dye terminator chemistry (Big Dye v3.1, Applied Biosystems) according to the manufacturer’s instructions. The primer and probe sets outlined above were derived from previously published sources, and were modified to target viruses isolated from orangutan hosts (Moens et al., 2009;
Schweizer and Nuemann-Haefelin, 1995). As a result, we obtained a longer pol sequences from only three individuals PP10-060 Santi, PP10-039 Ogeg, and PP10-069 Jimmi.

2.3.3 *Pongo* sp. mitochondrial COX2 amplification and sequencing

One µl gDNA was used to amplify a 585-bp COX2 sequence from 14 individuals using the following primers: Pongo_COII_FA (5'-AATCTGCTTCTAGTCCTGTACGC-3') and Pongo_COII_RA (5'-ACGGGCCCTATTTCGAAGATTT-3'). PCR reactions were prepared in 25 µL volumes containing 1x buffer, 37.5 mM MgCl2, 20 mM dNTP, 62.5 uM of each primer, and 0.75 units Taq polymerase (Bulldog Bio). Amplification was performed with the following parameters: initial denaturation at 94°C for 5 minutes followed by 40 cycles of 94°C for 15 seconds, 52°C for 10 seconds, and 72°C for 40 seconds, followed by a one-time extension period of 72°C for 20 minutes and held at 4°C.

Sequencing was conducted at the University of Washington’s (Seattle, WA) sequencing facility, where the PCR product was purified via Exo-SAP. Samples were sequenced in both the forward and reverse directions using the PCR primers and BigDye chemistry (Applied Biosystems) and products were analyzed using a 3730xl DNA Genetic Analyzer (Applied Biosystems). Sequences were trimmed and assembled automatically by Sequencher 5.1 (Gene Codes Corp.) using default parameters and manually checked against the chromatograms. Sequences were aligned using MAFFTv7’s E-INS algorithm (Katoh et al. 2005; Katoh and Standley 2013) and edited manually in MEGA v.6.06 (Tamura et al. 2013). Minority SNPs were re-verified with Sequencher v5.1 and sequences were realigned again using MAFFTv7.
2.3.4 Sequence analyses

Comparative SFV pol (425-bp) sequences from apes (gibbons, gorillas, chimpanzees, bonobos and orangutans) were downloaded from GenBank (Table 2.1) and aligned with our new Pongo pygmaeus sequences (878-bp) using ClustalW in MEGA6 and edited manually (Tamura et al., 2013) to give a final alignment length of 423-bp. Using jModelTest 2 (Darriba et al., 2012; Guindon and Gascuel 2003; Posada 2008), the Hasegawa-Kishino-Yano model (Hasegawa et al., 1985) with a gamma distribution (HKY+G) was identified as the best-fitting nucleotide substitution model according to the Bayesian Information Criteria (BIC). We conducted Bayesian analyses using BEAST v2.1.2 to simultaneously infer evolutionary histories, divergence dates, and nucleotide substitution rates (Bouckaert et al. 2014). The Bayesian analyses were performed using an uncorrelated lognormal relaxed molecular clock and the Birth-Death tree priors with 100 million Markov Chain Monte Carlo (MCMC) iterations with a 10% burn-in. The relaxed clock was calibrated using fossil evidence and genetically inferred dates. For the first analysis, we set the origin of the Hominidae clade at between 17-15.5 mya (Begun et al., 2012), assuming that Griphopithecus (17-14 mya) is the most recent common ancestor (MRCA) of all extant and fossil hominids (i.e. African and Asian great apes). In addition, genetic calibration dates were included in the first analysis based on nuclear DNA split estimates from Perelman et al. (2011). Using the Perelman et al. (2011) estimates we set the Pan/Gorilla split at 8.3 mya (± 1.1 mya), and the Pan troglodytes/Pan paniscus split at 2.17 mya (± 0.55 mya). For the second analysis, we used the same fossil calibration as the first analysis and used divergence dates inferred from mtDNA by Raumm et al., (2005). We set the Pan/Gorilla split at 8.1 mya (± 0.56 mya) and the Pan troglodytes/Pan paniscus split at 2.4 mya (± 0.179 mya). Convergence of the chain sampling was checked in the program Tracer v1.6 for high effective sample sizes (ESS), all of which were > 276. Trees were saved every 100,000 generations and the tree with
the maximum product of the posterior clade probabilities (maximum clade credibility tree) was chosen from the posterior distribution of 10,001 sampled trees after burning in the first 1,000 sampled trees with the program TreeAnnotator version 2.1.2. Trees were visualized with FigTree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

To evaluate virus/host co-evolution we created a ML tree using orangutan mtDNA COX2 sequences obtained in our study, and from comparative COX2 sequences downloaded from GenBank (Table 2.1). jModelTest 2 (Darriba et al., 2012; Guindon and Gascuel 2003; Posada 2008) identified the Hasegawa-Kishino-Yano model (HKY) (Hasegawa et al., 1985) as the most appropriate nucleotide substitution model via the BIC. Seven orangutans had both SFV and COX2 sequences available, including Ogeg, Santi, and Jimmi (this study), Otis (SFV pol, AY686203; COX2, AY686168), Minyak (SFV pol AY686204; COX2, AY686169), Biji (SFV pol, AY686205; COX2, AY686171), and Sibu (SFV pol, AY686206; COX2, AY686172). Open reading frames were verified in the COX2 sequences using MEGA6 to prevent the use of defective nuclear mtDNA sequences in our analyses which have been shown to have a different evolutionary history and can complicate phylogenetic analyses of mtDNA sequences. We used two different tests to determine the extent of co-evolution between the two organisms. First, branch lengths were calculated from both the SFV pol and COX2 ML trees and were compared using ordinary least-squares regression. We also performed a permutation test of host-parasite co-speciation using a modified version of the Mantel test (Mantel, 1967). The permutation test determines a Pearson's correlation coefficient between host and parasite distances (Table 2.2) of which a probability is determined using 100,000 permutations of the labels while retaining the observed interacting hosts and parasites compared to that from randomized data (Hommola et al., 2009). This permutation test was shown to perform better than the original Mantel test and the Parafit method of Legendre et al. (2002) for assessing host-parasite co-speciation. We used both
the Tamura-Nei (TN93) and Kimura 2-parameter (K2P) nucleotide substitution models for generating the distance matrices since the HKY model is not available in the DNA distance method used in the permutation test. The SFV pol alignment was also checked for recombination using Bootscan, Geneconv, MaxChi, Chimera, and RDP within the program RDP v3 using the parameter defaults (Martin et al., 2010) to confirm that recombination does not affect our results.

2.4 Results

Of the 68 orangutans, 12 (17.6%) were positive for SFV by RT-PCR, including six males and six females. We were also able to amplify 286-bp pol sequences from ten of the twelve animals (data not shown), and 878-bp pol sequences from three orangutans. ML analysis of the 423-bp SFV pol sequences exhibited a strong host-specific phylogenetic signal (Fig. 2.1), with hylobatids forming a sister clade to the African and Asian hominid SFV sequences. Bootstrap support values (BSV) for internal nodes within the African hominid clade were all above 80. As with Verschoor et al. (2004), all Pongo pol sequences were monophyletic with strong support (BSV = 99). Within the Pongo sp. SFV clade are three strongly supported sub-clades consisting of SFV sequences from 1) Pongo pygmaeus, 2) Pongo abelii, and 3) both P. abelii and P. pygmaeus. The Pongo abelii-only clade is a sister group to the remaining two Pongo sub-clades and had a BSV of 98, compared to BSVs of 71 and 75 for the P. pygmaeus and mixed clades, respectively. The node representing the split of the mixed species clade and the P. pygmaeus only clade was not well supported with a BSV of 50 (Fig. 2.1). One of the new Pongo pygmaeus SFV sequences from this study from a wild-born orphaned orangutan. This orangutan is part of a semi-captive population of sick, injured and orphaned animals that spend their days in the forest (if possible), and nights in cages at the OCC&Q (hereafter referred to as rehabilitant), PP10-069-Jimmi, clustered between the P. abelii clade and the other Pongo subclades with
significant BSV support (97). Both of the other new *Pongo* sp. SFV pol sequences from this current study, PP10-060 Santi and PP10-039 Ogeg, rehabilitant apes, clustered within the mixed Bornean/Sumatran SFV clade.

The SFV pol relationships inferred with Bayesian analysis using both the combined fossil and nuclear DNA (Fig. 2.2), or mtDNA (tree not shown), calibrations gave identical topologies, very similar TMRCA estimates (Table 2.3), and showed a nearly identical internal branching structure as the ML tree with two minor exceptions. The orangutan Jimmi PP10-069 SFV sequence shifted to a position within the mixed Bornean/Sumatran SFV clade in both Bayesian inferred trees. In addition, one *Pan paniscus* sequence (AY686195) formed a sister lineage to other *Pan* sequences in contrast to the ML tree in which all four *P. paniscus* sequences form a monophyletic clade (Figs. 2.1, 2.2). According to our Bayesian analyses the SFV clade comprising the time to most recent common ancestor (TMRCA) for all *Pongo* sequences is about 4.1 - 4.5 million years ago (mya; combined 95% highest posterior density [HPD] interval of 8.183 - 2.023 mya) (Table 2.3), with a posterior probability (PP) of 1. The TMRCA of the Bornean and the mixed Bornean/Sumatran SFV sub-clades is approximately 2.05 - 2.23 mya (combined 95% HPD 3.756 - 1.094 mya), with very strong support (PP = 1). The well supported (PP > 0.99) Bornean only (*P. pygmaeus*) sub-clade has a TMRCA of about 1.07 – 1.18 mya (combined 95% HPD 2.173 - 0.493 mya). The main mixed sub-clade, which is important for testing the hypothesis of Verschoor et al. (2004), has a TMRCA of ~1.5 - 1.7 mya (combined 95% HPD 2.746 - 0.822 mya), but was only weakly supported (PP 0.67 – 0.69) due to a single divergent SFV sequence (Jimmi), which clustered outside the main grouping of the mixed clade. The rest of the mixed clade (minus Jimmi) was strongly supported with a PP of > 0.99. The strongly supported (PP = 1) Sumatran (*P. abelii*) only clade was the youngest, with a TMRCA of 0.41 – 0.45 mya (combined 95% HPD 0.971-0.123 mya).
2.4.1 Co-evolution of orangutan SFV \textit{pol} and mtDNA sequences

Orangutan-host (mtDNA) and SFV \textit{(pol)} only ML trees were created to assess host-viral co-evolution (Figure 2.3a, 2.3b). As with the more inclusive SFV trees, the orangutan-only SFV phylogenetic tree showed the same general evolutionary relationships, with the formation of three distinct \textit{Pongo} sp. SFV clades. In contrast, the COX2 phylogeny contained only two clades corresponding to the Bornean or Sumatran species. Seven individuals had both SFV and COX2 sequences, one was from \textit{P. pygmaeus} whose SFV was from the Bornean-only clade, four were \textit{P. pygmaeus} whose SFV sequences belonged to the mixed clade, and two \textit{P. abelii} whose SFVs were in the Sumatran-only clade.

The least-squares regression analysis revealed a positive linear relationship between SFV and host COX2 branch lengths from the ML phylogenies ($r^2 = 0.8242 \ p<0.0001$), a finding consistent with host-viral co-evolution (Figure 2.4). The residuals from the regression model were normally distributed (Wilk-Shapiro $P=0.073$) and exhibited constant variance (Modified Leven'sTest $P=0.437$). However, we did not detect a co-evolutionary signal in the branch lengths within the mixed orangutan SFV clade ($r^2=0.2519$, $p=0.26$). The results of the least-squares regression analysis were confirmed by results from the Mantel Test comparing SFV \textit{pol} and mitochondrial COX2 genetic sequences. Both nucleotide substitution models tested produced statistically significant results (TN93 model $p= 0.01$ and K2P model $p = 0.01$), indicating coevolution of these two markers. Furthermore, the Bayesian-inferred nucleotide substitution rate for the ape SFV sequences of $2.71 \ E-08$ substitutions/site/year (95\% HPD 2.1 – 3.4 E-08 substitutions/site/year) is consistent with previous estimates for both SFV and mtDNA sequences (Switzer et al 2004, etc.), which would be expected with an observed phylogenetic pattern of co-evolution.
2.5 Discussion

Bayesian frameworks are now routinely used for the co-inference of phylogenetic relationships and divergence dates while investigating the evolutionary histories of both viral and host populations, including SFV, which has been shown to co-evolve with NHPs (Liu et al., 2008, Muniz et al., 2013; Switzer et al, 2005). Verschoor et al (2004), using only genetic relationships and a single paleogeological event hypothesized that the mixed SFV clade was the result of a re-introduction of SFV from Bornean orangutans into Sumatran orangutans as Sumatra was re-populated via land bridges following the Toba super-catastrophe ~ 73,000 ya. Our results confirm the existence of the three orangutan SFV clades reported by Verschoor et al. (2004), including the putative mixed SFV clade, however, the inferred divergence dates for orangutan SFV clades are not consistent with the former hypothesis. Our inferred divergence dates derived from the mixed SFV clade using two different calibration models suggest that the mixed clade appeared approximately 1.6 mya, much earlier than the last Toba super-eruption.

Importantly, our SFV divergence dates are supported by similar evidence dating the split of the genus *Pongo* into the two recognized species. When 95% confidence intervals are considered these *Pongo* divergence estimates range from between 5.36 - 0.038 mya depending on the gene region and type of analysis performed (Table 2.4; e.g. Chattergee et al., 2009; Finstermeier et al., 2013; Locke et al., 2011; Ma et al., 2013; Mailund et al. 2011; Nater et al., 2011; Perelman et al., 2011; Pozzi et al., 2014; Raaum et al., 2005; Springer et al., 2012; Steiper, 2006). Our inferred ancient dates for the MRCA of SFV in orangutans (4.14 - 4.45 mya; 95% C.I. 8.183 - 2.023 mya) suggests an early Pliocene (Zanclean 5.333 - 3.6 mya) (Cohen et al., 2013) origin for SFV in *Pongo*. This agrees with many of the dates suggested by host mitochondrial and nuclear DNA dates (e.g. Chattergee et al., 2009; Finstermeier et al., 2013; Ma
et al., 2013; Nater et al., 2011; Pozzi et al., 2014; Raaum et al., 2005; Springer et al., 2012; Steiper, 2006). Not surprisingly, the SFV divergence date estimates for the other hominoids are also congruent with the inferred divergence dates of their hosts. For example, Perelman et al. (2011) using 54 nuclear gene regions in a Bayesian framework estimated the Hylobates/Hominidae apes, Gorilla/Pan, and Pan troglodytes/Pan paniscus splits to be 20.32 mya (95% HPD 16.59 – 24.22 mya), 8.3 mya (95% HPD 6.59 – 10.07 mya), and 2.17 mya (95% HPD 1.28 – 3.21 mya), respectively, which are strongly consistent with the SFV divergence dates from our analysis for these same splits (Table 3). These orangutan SFV divergence dates are further supported by our co-evolutionary analyses showing that both SFV and host mtDNA sequences have accumulated genetic diversity over an equivalent period of time.

### 2.5.1 Paleobiogeography

Examination of the demographic history of orangutans and the paleogeological history of Southeast Asia, combined with our robust divergence dates, may help answer how and when the three distinct SFV clades originated. The orangutan fossil record corresponding to our estimated date for the MRCA of SFV in orangutans is sparse when compared to that of the Late Pleistocene (Upper Pleistocene 126 -11.7 kya) when orangutans ranged throughout Southeast Asia, including Java, Sumatra, Borneo, Vietnam, and southern China (Table 2.5). Despite Steiper’s (2006) assertion that Pliocene fossil orangutan sites should exist on both the Islands of Sumatra and Borneo, the oldest fossil orangutan sites in the Sundaic Islands, including Borneo and Sumatra, date to ~1 mya on the Indonesian Island of Java (Ibrahim et al., 2013). This date is about 500 kya earlier than the earliest evidence of orangutans on the Malay Peninsula (Ibrahim et al., 2013). Our estimated TMRCA for SFV in Pongo suggests that SFV must have evolved within the orangutan lineage long before orangutans arrived in the Sunda Shelf. The oldest fossil
orangutan sites are found in China and date much earlier than those on the Sunda Shelf to just over 2 mya. For example, Baikong Cave and Mohui Cave date to 2.2 and 2.0 mya (Harrison et al., 2014; Ibrahim et al., 2013; Takai et al., 2014). The fossil record also indicates that orangutans were not present on the island of Sumatra until approximately 0.13 mya (Ibrahim et al., 2013).

Given the relatively recent fossil dates for *Pongo* in Borneo and Sumatra, our Bayesian estimates for the MRCA of SFV in *Pongo* at around 4.25 mya suggest a mainland origin not only for SFV in the genus, but likely also for the mixed species SFV clade estimated herein to have diverged about 1.6 mya. These results therefore suggest the mixed *P. pygmaeus*/*P. abelii* SFV clade was not likely the result of SFV reintroduction by Bornean orangutans moving into Sumatra after the Toba super-eruption ~ 73,000 ya, as proposed by Verschoor et al. (2004). Furthermore, a recent study reconstructing the demographic history of orangutans using approximate Bayesian analysis showed that the Toba eruption had little impact on orangutan populations and other local fauna and flora (Louys, 2007; Nater et al., 2015).

Alternatively, the mixed *Pongo* sp. SFV clade could reflect orangutan species diversity on the Indochinese mainland during the Plio-Pleistocene, sometime around 1-2 million years ago, or perhaps even earlier. Extinct orangutan species may have had some range overlap leading to episodic contact and perhaps even hybridization. Male orangutans are known to disperse over long distances to find females and establish territories (Nater et al., 2011, 2013; Nietlisbach et al., 2012). Host mobility would facilitate the movement of an SFV strain from a now extinct mainland *Pongo* species or subspecies such as *P. p. ciochoni, P. p. weidenreichi*, and *P. p. fromageti* (Cameron, 2001; Drawhorn, 1995; Ibrahim et al., 2013; Schwartz et al., 1995) into both the ancestral lineages of Bornean and Sumatran orangutans. Because recombination of SFV
strains has been reported (Feeroz et al., 2013; Galvin et al., 2013; Liu et al., 2008), we also examined the possibility that the mixed Sumatran/Bornean clade may reflect an ancient viral recombination event between the Bornean and Sumatran orangutan SFV strains. Using five different methods of recombination detection including Bootscan, Geneconv, MaxChi, Chimera, and RDP, we found no evidence of SFV recombination for the pol sequences for any of the Pongo sp. SFV clades. This would suggest that an ancient viral recombination event is less likely to be a cause of the mixed SFV Pongo clade. Our analyses indicate that the MRCA of the mixed SFV clade and the Bornean orangutan clades diverged about 1.6 mya, which predates the earliest evidence of fossil orangutans in Sundaland and supports our suggestion that this strain originated within mainland Southeast Asia, subsequently expanding south into Sundaland with proto-P. pygmaeus and proto-P. abelii populations. Interestingly, recent genetic studies have identified a population of P. abelii from Batang Toru in Sumatra, whose mtDNA is more closely related to P. pygmaeus than to other individuals from Sumatra (Nater et al., 2011), and that these orangutans diverged about 2.09 mya (95% C.I, 3.00 - 1.27 mya). This mtDNA divergence date generally corresponds to our date estimates for the MRCA of the P. pygmaeus/mixed SFV clade at ~1.5 - 1.7 mya (95% HPD 2.53-0.82 mya) providing additional support for what we believe to be the most parsimonious explanation of the tri-clade pattern of SFV evolution in orangutans.

2.6 Conclusions

In summary, through detailed phylogenetic analysis of new SFV sequences from wild born, semi-captive orangutans living on the island of Borneo and those available at GenBank, we inferred an ancient origin for SFV in Pongo. Although our results confirm three orangutan SFV clades, two of which co-evolved with each main Pongo species and a third mixed species clade, our inferred ancient divergence dates and absence of genetic recombination are inconsistent with
the prior hypothesis of a more recent introduction of SFV in the mixed orangutan subspecies clade precipitated by the Mount Toba eruption. By integrating genetic, fossil and biogeographic evidence we provide new evidence that the phylogenetic pattern of SFV in modern orangutans reflects the ancient diversity of the genus *Pongo* throughout the Plio-Pleistocene on the Indochinese mainland before an expansion south into Sundaland.
### 2.7 Tables

**Table 2.1** SFV polymerase (*pol*) and mitochondrial cytochrome oxidase subunit II (COX2) sequences used in the study.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>GenBank Accession Number</th>
<th>Gene Region</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Hylobates pileatus</em></td>
<td>AF516486</td>
<td><em>pol</em></td>
</tr>
<tr>
<td><em>Hylobates concolor leucogenys</em></td>
<td>AF516487</td>
<td><em>pol</em></td>
</tr>
<tr>
<td><em>Pongo pygmaeus</em></td>
<td>AJ627527-AJ627540, AY686203-AY686204</td>
<td><em>pol</em></td>
</tr>
<tr>
<td><em>Pongo abelii</em></td>
<td>AJ627541-AJ627547, AY686205, AY686206</td>
<td><em>pol</em></td>
</tr>
<tr>
<td><em>Pongo pygmaeus</em></td>
<td>PP10-039, PP10-069, PP10-060, PP10 022,</td>
<td><em>pol</em>, COX2</td>
</tr>
<tr>
<td></td>
<td>PP10 004, PP10 042</td>
<td></td>
</tr>
<tr>
<td><em>Pongo pygmaeus</em></td>
<td>PP10 046, PP10 018, PP10 023, PP10 050,</td>
<td>COX2</td>
</tr>
<tr>
<td></td>
<td>VM20, VM46, VM84, VL73, D38115,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U12703, X97716, AY686168, AY686169,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>X97712</td>
<td></td>
</tr>
<tr>
<td><em>Pongo abelii</em></td>
<td>U12704, AY686168, X97707, AY686170-AY686172,</td>
<td>COX2</td>
</tr>
<tr>
<td></td>
<td>CR926488, EU835077-EU835084, EU835090-EU835095</td>
<td></td>
</tr>
<tr>
<td><em>Pan paniscus</em></td>
<td>AJ627550, AJ627551, AY686195, EU527595</td>
<td><em>pol</em></td>
</tr>
<tr>
<td><em>Pan troglodytes schweinfurthii</em></td>
<td>AJ627552, AJ627553, EU527542</td>
<td><em>pol</em></td>
</tr>
<tr>
<td><em>Pan troglodytes troglodytes</em></td>
<td>AY639122-AY639124, AY639126,</td>
<td><em>pol</em></td>
</tr>
<tr>
<td></td>
<td>AY639128, AY639130, AJ627555,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EU527494-EU527495</td>
<td></td>
</tr>
<tr>
<td>Species</td>
<td>Accession Numbers</td>
<td>Type</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td><em>Pan troglodytes verus</em></td>
<td>AJ627556, AJ627560, EU527503, EU527517-EU527518, EU527525</td>
<td>pol</td>
</tr>
<tr>
<td><em>Pan troglodytes vellerosus</em></td>
<td>AY639133, AY639141, EU527484, EU527487, EU527492-EU527493, EU527528</td>
<td>pol</td>
</tr>
<tr>
<td><em>Gorilla gorilla</em></td>
<td>AY686191-AY686194, AY603409, AY603410, AY583782, EU527593</td>
<td>pol</td>
</tr>
</tbody>
</table>
Table 2.2 Maximum composite likelihood genetic distances for *Pongo* sp. simian foamy virus and mitochondrial sequences. The SFV *pol* and COX2 distances are above and below the diagonal, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Minyak¹</th>
<th>Biji²</th>
<th>Sibu²</th>
<th>Ogeg¹</th>
<th>Santi¹</th>
<th>Jimmi¹</th>
<th>Otis¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minyak¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AY686169</td>
<td>-</td>
<td>0.181128</td>
<td>0.180479</td>
<td>0.065526</td>
<td>0.071593</td>
<td>0.065127</td>
<td>0.098481</td>
</tr>
<tr>
<td>Biji²</td>
<td>0.046965</td>
<td>-</td>
<td>0.005169</td>
<td>0.192965</td>
<td>0.177974</td>
<td>0.137052</td>
<td>0.170593</td>
</tr>
<tr>
<td>AY686171</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sibu²</td>
<td>0.05662</td>
<td>0.013048</td>
<td>-</td>
<td>0.192314</td>
<td>0.17747</td>
<td>0.136807</td>
<td>0.170254</td>
</tr>
<tr>
<td>AY686172</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ogeg¹</td>
<td>0.006466</td>
<td>0.044624</td>
<td>0.054279</td>
<td>-</td>
<td>0.04316</td>
<td>0.068731</td>
<td>0.095485</td>
</tr>
<tr>
<td>Santi¹</td>
<td>0.006466</td>
<td>0.044624</td>
<td>0.054279</td>
<td>0</td>
<td>-</td>
<td>0.071639</td>
<td>0.098491</td>
</tr>
<tr>
<td>Jimmi¹</td>
<td>0.006466</td>
<td>0.044624</td>
<td>0.054279</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0.059532</td>
</tr>
<tr>
<td>Otis¹</td>
<td>0.002145</td>
<td>0.044624</td>
<td>0.054279</td>
<td>0.004321</td>
<td>0.004321</td>
<td>0.004321</td>
<td>-</td>
</tr>
<tr>
<td>AY686168</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ Indicates *Pongo pygmaeus*  
² Indicates *Pongo abelii*
Table 2.3 Bayesian estimated divergence dates for time to most recent common ancestor (TMRCA) in millions of years (mya) for the Pongo sp. simian foamy virus and other ape clades

<table>
<thead>
<tr>
<th>Clade</th>
<th>Node ID</th>
<th>Posterior probability</th>
<th>TMRCA mtDNA model (95% HPD)</th>
<th>TMRCA nuclear DNA model (95% HPD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pongo</td>
<td>1</td>
<td>1</td>
<td>4.138 (7.527 - 2.023)</td>
<td>4.447 (8.183 - 2.075)</td>
</tr>
<tr>
<td>Mixed/P. pygmaeus</td>
<td>2</td>
<td>1</td>
<td>2.051 (3.348 – 1.094)</td>
<td>2.231 (3.756 – 1.164)</td>
</tr>
<tr>
<td>Mixed</td>
<td>3</td>
<td>&gt;0.67</td>
<td>1.516 (2.526 - 0.828)</td>
<td>1.665 (2.746 - 0.822)</td>
</tr>
<tr>
<td>P. pygmaeus</td>
<td>4</td>
<td>&gt;0.99</td>
<td>1.073 (2.012 - 0.493)</td>
<td>1.179 (2.173 - 0.493)</td>
</tr>
<tr>
<td>P. abelii</td>
<td>5</td>
<td>1</td>
<td>0.413 (0.893 - 0.123)</td>
<td>0.447 (0.971 - 0.134)</td>
</tr>
<tr>
<td>Hylobates/Hominidae</td>
<td>6</td>
<td>1</td>
<td>15.03 (10.85 - 16.65)</td>
<td>15.33 (11.09 - 16.75)</td>
</tr>
<tr>
<td>Gorilla/Pan</td>
<td>7</td>
<td>1</td>
<td>8.03 (6.96 - 9.1)</td>
<td>8.12 (6.06 - 9.95)</td>
</tr>
<tr>
<td>Pan troglodytes/Pan paniscus</td>
<td>8</td>
<td>1</td>
<td>2.5 (2.16 - 2.83)</td>
<td>2.85 (2.04 - 3.71)</td>
</tr>
</tbody>
</table>

1 Node IDs correspond to Bayesian tree presented in Figure 2.
2 HPD, high posterior density ranges
Table 2.4 Genetic estimates for time to most recent common ancestor (TMRCA) of the genus *Pongo*, and the split dates for *P. abelii* and *P. pygmaeus* in millions of years (mya)\(^1\)

<table>
<thead>
<tr>
<th>Split dates (mya) for <em>P. abelii</em> and <em>P. pygmaeus</em></th>
<th>Genetic basis for date estimates</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 (3.5 - 4.7)</td>
<td>mtDNA</td>
<td>Raaum et al., 2005</td>
</tr>
<tr>
<td>2.7-5</td>
<td>mtDNA and Nuclear DNA meta-analysis</td>
<td>Steiper, 2006</td>
</tr>
<tr>
<td>4.5 (3.9 - 5.0) and 4.7 (3.2 - 6.1)</td>
<td>Nuclear DNA and mtDNA (Relaxed and strict clock, respectively)</td>
<td>Chattergee et al., 2009</td>
</tr>
<tr>
<td>1.31 (0.61 - 2.2)</td>
<td>Nuclear DNA</td>
<td>Perelman et al., 2011</td>
</tr>
<tr>
<td>3.5 (2.31 - 4.75)</td>
<td>mtDNA</td>
<td>Nater et al., 2011</td>
</tr>
<tr>
<td>0.168 (0.038 – 0.375)</td>
<td>Y chromosome</td>
<td></td>
</tr>
<tr>
<td>0.334 (0.189 - 0.479)</td>
<td>Hidden Markov model of nuclear DNA (no Bayesian analysis)</td>
<td>Mailund et al. 2011</td>
</tr>
<tr>
<td>0.4</td>
<td>Nuclear DNA model (no Bayesian analysis)</td>
<td>Locke et al., 2011</td>
</tr>
<tr>
<td>1.28 (0.51 - 2.99)</td>
<td>mtDNA and nuclear DNA</td>
<td>Springer et al., 2012</td>
</tr>
<tr>
<td>3.67</td>
<td>mtDNA</td>
<td>Ma et al., 2013</td>
</tr>
<tr>
<td>0.4</td>
<td>Autosomal DNA</td>
<td></td>
</tr>
<tr>
<td>3.99 (2.78 - 5.36) and 3.74 (2.62 - 4.93)</td>
<td>mtDNA</td>
<td>Finstermeier et al., 2013</td>
</tr>
<tr>
<td>4.2 (3.39 - 5.14)</td>
<td>mtDNA</td>
<td>Pozzi et al., 2014</td>
</tr>
</tbody>
</table>

\(^1\)95% confidence intervals or Bayesian highest posterior density ranges in parentheses
Table 2.5 Early to Middle Pleistocene fossil localities with well-dated fossil evidence of orangutans

<table>
<thead>
<tr>
<th>Locality</th>
<th>Dates (mya)</th>
<th>Epoch-Stage (Cohen et al. 2013)</th>
<th>Fossil Orangutan Species</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baikong Cave, China</td>
<td>2.2</td>
<td>Pleistocene-Gelasian</td>
<td><em>Pongo sp.</em></td>
<td>Harrison et al., 2014; Ibrahim et al., 2013; Takai et al., 2014</td>
</tr>
<tr>
<td>Mohui Cave, China</td>
<td>2</td>
<td>Pleistocene-Gelasian</td>
<td><em>Pongo sp.</em></td>
<td>Harrison et al., 2014; Ibrahim et al., 2013; Takai et al., 2014</td>
</tr>
<tr>
<td>Sanheda Cave, China</td>
<td>1.6 - 1.2</td>
<td>Pleistocene-Calabrian</td>
<td><em>Pongo sp.</em></td>
<td>Ibrahim et al., 2013; Wang et al., 2014</td>
</tr>
<tr>
<td>Gigantopithecus Cave, China</td>
<td>1.2 - 0.9</td>
<td>Pleistocene-Calabrian</td>
<td><em>Pongo sp.</em></td>
<td>Ibrahim et al., 2013; Rink et al., 2008</td>
</tr>
<tr>
<td>Sangiran, Java, Indonesia</td>
<td>1</td>
<td>Pleistocene-Calabrian</td>
<td><em>Pongo pygmaeus</em></td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Trinil, Java, Indonesia</td>
<td>0.5</td>
<td>Pleistocene-Middle</td>
<td><em>Pongo pygmaeus</em> sp.</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Tham Khuyen, Vietnam</td>
<td>0.475</td>
<td>Pleistocene-Middle</td>
<td><em>Pongo pygmaeus</em> kahlkei</td>
<td>Ciochon et al., 1996; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Tham Hai, Vietnam</td>
<td>0.475</td>
<td>Pleistocene-Middle</td>
<td><em>Pongo pygmaeus</em></td>
<td>Ibrahim et al., 2013; Olsen and Ciochon, 1990</td>
</tr>
<tr>
<td>Lang Trang Caves, Vietnam</td>
<td>0.5 - 0.146</td>
<td>Pleistocene-Middle</td>
<td><em>Pongo pygmaeus</em></td>
<td>Drawhorn, 1995; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Kao Pah Nam, Thailand</td>
<td>~0.690</td>
<td>Pleistocene-Middle</td>
<td><em>Pongo sp.</em></td>
<td>Ibrahim et al., 2013; Pope et al., 1981</td>
</tr>
<tr>
<td>Badak Caves, Malaysia</td>
<td>0.5</td>
<td>Pleistocene-Middle</td>
<td><em>Pongo sp.</em></td>
<td>Ibrahim et al., 2013</td>
</tr>
</tbody>
</table>

1MYA=Million Years Ago
2.8 Figure Legends

**Figure 2.1.** Maximum likelihood phylogeny of ape SFV polymerase (*pol*) sequences (423-bp) using the HKY+G nucleotide substitution model. 1,000 bootstrap replicates were used to test the robustness of the observed clades; bootstrap values > 60 are shown at nodes. Letter after animal name indicates sample origin; Z=Zoo, R=Wild-born rehabilitant, W=Wild-born, C=Captive born, U=Unknown.

**Figure 2.2.** Bayesian-inferred phylogeny using 423-bp polymerase (*pol*) sequences and a relaxed molecular clock with a host fossil calibration for the Hominidae set to 17 - 15.5 million years ago (mya), and nuclear DNA calibrations for the *Pan/Gorilla* split at 8.3 mya and the *Pan troglodytes/Pan paniscus* split at 2.17 mya. Nodes with posterior probabilities (PP) greater than 0.60 are indicated. Numbers at nodes indicate splits with inferred divergence dates provided in Table 2.3. Letter after animal name indicates sample origin; Z=Zoo, R=Wild-born rehabilitant, W=Wild-born, C=Captive born, U=Unknown.

**Figure 2.3.** Phylogenetic relationships of *Pongo* sp. (a) SFV polymerase (*pol*) (425-bp) and (b) mitochondrial cytochrome oxidase subunit II (COX2) (546-bp) sequences. Phylogenies were inferred using the maximum likelihood method. 1,000 bootstrap replicates were used to test the robustness of the observed clades; bootstrap values > 60 are shown at nodes. Letter after animal name indicates sample origin; Z=Zoo, R=Wild-born rehabilitant, W=Wild-born, C=Captive born, U=Unknown.

**Figure 2.4.** Least-squares regression analysis of SFV polymerase (*pol*) and orangutan host mitochondrial cytochrome oxidase subunit II (COX2) maximum likelihood inferred branch lengths.
2.9 Figures

**Figure 2.1** Maximum likelihood phylogeny of ape SFV polymerase (*pol*) sequences (423-bp)
Figure 2.2 Bayesian-inferred phylogeny using 423-bp polymerase \((pol)\) sequences
Figure 2.3

Figure 2.3a Phylogenetic relationships of *Pongo* sp. SFV polymerase (*pol*) (425-bp)
Figure 2.3b Phylogenetic relationships of *Pongo* sp. mitochondrial cytochrome oxidase subunit II (COX2) (546-bp)
**Figure 2.4** Least-squares regression analysis of SFV (*pol*) and orangutan cytochrome oxidase subunit II (COX2)

\[ y = 4.2943x + 0.0723 \]

\[ R^2 = 0.8242 \]
2.10 References


Heneine, W., Switzer, W.M., Sandstrom, P., Brown, J., Vedapuri, S., Schable, C.A., Khan, A.S.,

Hooijer, D.A., 1948. Prehistoric teeth of man and of the orangutan from central Sumatra, with
notes on the fossil orang utan from Java and southern China. Zoologische Mededeelingen,
29, 175-301.


Vol. 17, 327-43.


Ibrahim, Y.K., Tshen, L.T., Westaway, K.E., Cranbrook, E.O., Humphrey, L., Muhammad, R.F.,
Zhao, J., Peng, L.C., 2013. First discovery of Pleistocene orangutan (Pongo sp.) fossils in

Jones, H.L., Rink, W.J., Schechart, L.A., Miller-Antonio, S., Weiwen, H., Yamei, H., Wei, W.,
2004. Coupled electron spin resonance (ESR)/uranium-series dating of mammalian tooth


Chapter 3: Detailed phylogenetic analysis of Primate T-lymphotropic virus type 1 (PTLV-1) sequences from Orangutans (*Pongo pygmaeus*) reveals new insights into the evolutionary history of PTLV-1 in Asia

**Author Contributions:** Conceived and designed the study: Michael J.C. Reid, Michael A. Schillaci, James I. Brooks. Performed the experiments: Michael J.C. Reid, William M. Switzer, Michael A. Schillaci, James I. Brooks Manon Ragonnet, Isabelle Joanisse, Kyna Caminiti, Analyzed the data: Michael J.C. Reid, Michael A. Schillaci, William M. Switzer. Contributed reagents/materials/analysis tools: Manon Ragonnet-Cronin, Isabelle Joanisse, Kyna Caminiti, Carl A. Lowenberger, Birute Mary F. Galdikas, Paul A. Sandstrom. Wrote the paper Michael J.C. Reid, Michael A. Schillaci, William M. Switzer, James I. Brooks.
3.1 Abstract

While human T-lymphotropic virus type 1 (HTLV-1) originates from ancient and contemporaneous cross-species transmission of simian T-lymphotropic virus type 1 (STLV-1) from infected nonhuman primates, much debate exists on whether the first HTLV-1 occurred in Africa or Asia during early human evolution and migration. This topic is complicated by a lack of representative Asian STLV-1 to infer PTLV-1 evolutionary histories. In this study we obtained new STLV-1 LTR and tax sequences from a wild-born Bornean orangutan (Pongo pygmaeus) and performed detailed phylogenetic analyses using both maximum likelihood and Bayesian inference of available Asian PTLV-1 and African STLV-1 genetic sequences. Phylogenies, divergence dates and nucleotide substitution rates were co-inferred and compared using three different molecular clock calibrations in a Bayesian framework, including both archaeological and nucleotide substitution rate calibrations. We then combined our molecular results with paleobiogeographical and ecological data to infer the most likely evolutionary history of PTLV-1. Our analyses robustly inferred an Asian source for PTLV-1 with cross-species transmission of STLV-1 likely from a macaque (Macaca sp.) to an orangutan about 207.5 - 17.2 kya, and to humans between 125.9 – 10.4 kya. An orangutan diversification of STLV-1 commenced approximately 37.5 – 3.2 kya. Our analyses also inferred that HTLV-1 was first introduced into Australia ~3.7 kya, corresponding to both genetic and archaeological changes occurring in Australia at that time. Finally, HTLV-1 appears in Melanesia at ~2.7 kya corresponding to the migration of the Lapita peoples into the region. Our results also provide an important future reference for calibrating information essential for PTLV evolutionary timescale inference.
3.2 Introduction

Deltaretroviruses are oncogenic retroviruses that include at least four viral groups, primate T-lymphotropic virus (PTLV) types 1-4 (PTLV-1, PTLV-2, PTLV-3, and PTLV-4) (Wolfe et al., 2005). Each PTLV group is composed of a simian variant (simian T-lymphotropic virus (STLV)) and a human counterpart (human T-lymphotropic virus (HTLV)), which is the result of cross-species transmissions from an STLV-infected nonhuman primate (NHP). More recently a fifth PTLV group (PTLV-5) has been proposed for a highly divergent STLV found in *Macaca arctoides* (Ayoub et al., 2013; Liegeois et al., 2008; Mahieux and Gessain, 2011), though an HTLV-5 has not yet been identified. STLV-1 has a wide geographic distribution and has been identified in many species of monkeys and apes in both Asia and Africa (Sintasath et al., 2009; Van Dooren et al., 2007). Epidemiological data from an estimated 1.5 billion people suggests that between 4.5 and 9.3 million individuals are infected with HTLV-1, indicating that the total global infection rate is likely much higher (Gessain and Cassar, 2012). There are seven HTLV-1 subtypes: subtype A (or Cosmopolitan) which is found throughout Africa, Asia, and the Americas, subtype B which is found in central Africa, subtype C which is found in Melanesia and Australia, and subtypes D, E, F and G all being found in west-central Africa (Gessain and Cassar, 2012). Approximately 5% of those infected with HTLV-1 are susceptible to development of two severe illnesses associated with severe morbidity, adult T-cell leukemia/lymphoma (ATLL), and HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP) (Gessain, 2011; Gessain and Cassar, 2012; Ibrahim et al., 1995; Locatelli and Peters, 2012).

Given the wide geographic distribution of STLV-1, the precise origin of HTLV-1 has been the subject of much discussion and debate (Andonov et al., 2012; Kelsey et al., 1999;
Vandamme et al., 1998; Van Dooren et al., 2005, 2007; Verdonck et al., 2007). An African origin for HTLV-1 was suggested based on the phylogenetic divergence of African STLV-1s from Asian strains (Kelsey et al., 1999). In 1998 more complete phylogenetic analyses, combined with a larger number of PTLV-1 sequences from different NHPs available for comparison, led to the inference of an Asian origin of PTLV-1 (Vandamme et al., 1998).

Subsequent studies have shown that the Asian clades exhibit much longer branch lengths than those of the African PTLV-1, reflecting deeper evolutionary histories, which is further supported by older estimated molecular dates for the Asian (269-156 thousand years ago [kya]) compared to African (35-19 kya) PTLV-1 clades (Van Dooren et al., 2005; 2007). These findings point to a possible Asian origin for PTLV-1, a suggestion which contrasts sharply with the current out-of-Africa paradigm for other simian retrovirus zoonotic infections (Katzourakis et al., 2014; Sharp and Hahn, 2011).

Unlike HTLV in Asia, comparatively little is known about the evolutionary history of STLV-1 in Asia. As many as 14 Asian macaque species are known to harbour STLV-1-like viruses based on serological studies (Ibrahim et al., 1995). However, our search of GenBank showed that only eight macaque species have STLV-1 long terminal repeat (LTR) or tax sequences available, which are two of the most common PTLV regions analyzed phylogenetically, including Macaca nemestrina (Richards et al., 1998), M. tonkeana (Ibrahim et al., 1995), M. arctoides (Mahieux et al., 1997; Vandamme et al., 1998; Van Dooren et al., 2005), M. nigra, M. maura (Van Dooren et al., 2007), M. mulatta, M. fascicularis and M. fuscata (Meertens and Gessain, unpublished).

STLV-1 in orangutans (Pongo sp.) was first identified at Indonesian zoos on the island of Java (Ibuki et al., 1997). Of 41 orangutans sampled for that initial study, three (7.3%) showed
serological evidence for STLV-1 infection, with a long terminal repeat (LTR) sequence reported from one animal (INA004). The authors showed that the orangutan STLV-1 LTR sequence was most closely related to STLV-1 from *M. tonkeana* (Ibuki et al., 1997). In another study, Verschoor et al., (1998) sampled a large group (n=143) of wild-caught orangutans housed at a rehabilitation center in Kalimantan (Indonesian Borneo), and confirmed the low seroprevalence of STLV-1 in orangutans (2/143, 1.4%). Although genetic testing of the two seropositive individuals was not reported in that study, STLV-1 tax/rex sequences from a third wild-caught orangutan (OU-KA) yielded a phylogeny similar to that presented by Ibuki et al., (1997), with orangutan STLV-1 showing a closer evolutionary relationship with macaques (*M. mulatta* and *M. nemestrina*) than with STLV-1 sequences from African apes (Verschoor et al., 1998). These genetic relationships were subsequently confirmed in a separate study using envelope (*env*) and LTR sequences from OU-KA (Van Dooren et al., 2007). Only two gibbon STLV-1 sequences have been reported: One each in a captive Symphalangus syndactylus (GenBank # DQ076651) and a wild caught Hylobates pileatus (GenBank # KC283222) from Cambodia (Ayouba et al. 2013; Van Dooren et al. 2007). However, both gibbon STLV-1 sequences clustered with macaque STLV-1 in those studies complicating determination of the evolutionary history of the gibbon viruses.

STLV-1 in African apes is much better known than in Asian apes. STLV-1 has been identified in gorillas (*Gorilla* sp.) and chimpanzees (*Pan troglodytes*) (Junglen et al., 2010; F. Leendertz et al., 2004; S. Leendertz et al., 2010; Liégeois et al., 2008; Nerrinet et al., 2001; 2004). For example, a study examined chimpanzee (*Pan troglodytes verus*) and Western red colobus (*Piliocolobus badius*) STLV-1 at Taï National Park in Côte d’Ivoire, and discovered that chimpanzees not only had their own STLV-1 strain, but were also infected with STLV-1 from Western red colobus monkeys (F. Leendertz et al., 2004). Another study of chimpanzees (*Pan
troglodytes verus) in Taï National Park, and a single chimpanzee from Sierra Leone, found three distinct chimpanzee STLV-1 clades in one geographic area suggesting that STLV-1 clusters by host phylogeny (Junglen et al., 2010) for P. troglodytes rather than by geography as has become the paradigm for STLV (Sintasath et al., 2009; Vandamme et al., 1998; Van Dooren et al., 2001).

To better understand the evolutionary history of STLV-1 in orangutans, we report herein the results of detailed phylogenetic analyses of available Asian PTLV-1 LTR and tax sequences, including new STLV-1 LTR and/or tax sequences from two orangutans. The first was derived from blood samples collected in June/July of 2003 from a wild-born orangutan in Kalimantan Tengah, Indonesia, while the second is an unpublished LTR sequence available in GenBank (Ppy-Ph711, AY141165). Using a Bayesian approach, we inferred divergence dates and evolutionary histories for STLV-1 in orangutans and PTLV-1 clades with various Asian and African origins. Setting the most accurate molecular clock calibrations involves undertaking detailed reviews of anthropological evidence (e.g., archaeology, human and nonhuman primate fossils, historical linguistics, human and primate genetics etc.).

3.3 Methods

3.3.1 Study population and specimen collection.

Blood samples were collected in 2003 from 68 (35 females, 33 males) wild-born orangutans (Pongo pygmaeus) housed at the Orangutan Care Center and Quarantine (OCC&Q) located in the village of Pasir Panjang in the Central Kalimantan Province of Indonesian Borneo (see Reid et al., 2006; Pacheco et al., 2012). The relative age of each orangutan was estimated using body mass. Blood samples were collected as part of semi-annual health checks done at the center, or as part of routine health checks done on newly arrived orangutans associated with a screening and quarantine protocol. Blood was drawn from the orangutans by OCC&Q veterinary
staff using a 25-gauge × 1 5/8-inch PrecisionGlide needle (Becton, Dickinson, and Company, Oakville, Ontario, Canada) and a 3-mL syringe. About 160-200 µl of orangutan blood preserved with EDTA was pipetted onto Whatman FTA® Classic Cards (Whatman Inc., Florham Park, NJ, USA) and dried overnight at room temperature for transport to Canada. Animal ethics approval (Simon Fraser University UACC # 675B-03) and a Convention on International Trade in Endangered Species (CITES) permit (08692/IV/SATS-LN/2004) were obtained for this research. Serology was not performed on the FTA® blood specimens because of limited specimen volume available for testing.

### 3.3.2 Elution of genomic DNA from FTA® cards.

Genomic DNA (gDNA) was eluted from the Whatman FTA® Classic Cards spotted with orangutan (*P. pygmaeus*) blood following the manufacturer’s instructions. Briefly, using a 1.2 mm Harris Uni-Core punch, 3 discs were punched from each dried blood spot (DBS) and ejected into a sterile 1.5 mL microcentrifuge tube. To avoid cross contamination, one punch from an unused FTA® card was taken between each sample. Discs were washed prior to gDNA elution via incubation at room temperature with 3 x 200 µL FTA® Purification Reagent (Whatman International Limited) for 5 minutes each, and 2 x 200 µL TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) for an additional 5 minutes each. gDNA was eluted from the discs by incubation with 35 µL of Solution 1 (0.1 N NaOH, 0.3 mM EDTA, pH 13.0) for 5 minutes at room temperature. To this, 65 µL of Solution 2 (0.1 M Tris-HCl, pH 7.0) was added, and the samples were flash vortexed 5 times to ensure proper mixing. Samples were allowed to incubate at room temperature for 10 minutes, after which they were flash vortexed an additional 10 times. Punches were removed from each sample and squeezed with tweezers to recover a maximum volume of eluate, which contains gDNA in TE buffer (66 mM Tris-HCl, 0.1 mM EDTA).
3.3.3 PCR detection of STLV-1

A 246-bp fragment of the STLV-1 \textit{tax} gene (nucleotides 7521-7765 of STLV-1\_Tan90 from a tantalus monkey, GenBank number NC\_000858, Saksena et al., [1993]) was PCR-amplified and quantified with real-time PCR as a qualitative assay to detect STLV-1 in the 68 \textit{Pongo pygmaeus} samples. Briefly, 10 µL orangutan gDNA was used as template for the PCR reactions. Following an initial hot-start incubation of 10 minutes at 95°C for activation of the Taq DNA polymerase (Applied Biosystems, Streetsville, Ontario, Canada), the following cycling conditions were used for the 1\textsuperscript{st} round of template amplification: denaturation for 20 seconds at 94°C, annealing for 30 seconds at 51°C and extension for 1 minute at 72°C. The reaction was cycled 35 times, and was completed with a final hold at 72°C for 10 minutes. The following orangutan specific primers were developed using STLV-1\_Tan90 as a reference; \textit{tax}.outer.fwd (5’ GTT TGG AGA CTG TGT ACA AGG C 3’; nucleotides 7350-7371) and \textit{tax}.outer.rev (5’ GGT AGA GGT ACA KGC AGA CAA CAG 3; nucleotides 7742-7765).

For real time PCR, 5 µL of first round amplification product was used as template. Real-time PCR primer and probe sets were optimized with computer software (Primer Express, Applied Biosystems) for orangutan STLV-1 with previously published primer sequences (Moens et al., 2009; Salemi et al., 1998) and a target located within a conserved area of the \textit{tax} gene using the primers and probe \textit{tax}.inner.fwd (5’ C TCC TTC CCC ACC CAG AG 3’; nucleotides 7521-7538), \textit{tax}.inner.rev (5’ GGG TGG GYT CCA TRT ATC CRT T 3’; nucleotides 7642-7663) and \textit{tax} probe (5’ FAM-CAT ACA ACC CCC AAC ATT CCA CCC TCC T-MGBNFQ 3’; nucleotides 7579-7606). Cycling conditions included denaturation for 15 seconds at 94°C and anneal-extend for 1 minute at 60°C for 45 cycles. Prior to cycling, an incubation of 15 minutes at 95°C was performed as per manufacturer’s recommendations for HotStarTaq\textsuperscript{®} DNA.
polymerase activation (Quantitect Probe PCR Kit, Qiagen, Mississauga, Ontario, Canada). The standard curve for the STLV-1 \textit{tax} sequences consisted of 10-fold serial dilutions of a linearized plasmid (pCR2.1-TOPO; Invitrogen, Burlington, Ontario, Canada) containing a cloned fragment of the LTR and \textit{tax} regions, respectively. The assays could reliably detect 10 copies of target DNA in a background of STLV-negative macaque (*Macaca fascicularis*) gDNA with three replicates. Real-time PCR procedures were carried out in an ABI Prism 7000 sequence detection system (Applied Biosystems).

A 460-bp fragment of the STLV-1 LTR region between nucleotides 8246-8719 of the complete HTLV-1 genome (GenBank accession number J02029, Seiki et al., [1983]) was amplified using nested PCR with modifications of primer and probe sets to target STLV-1 from orangutans (Moens et al., 2009; Salemi et al, 1998). Briefly, 10 µL orangutan gDNA was used as template for the first round of PCR amplification, using primer set AV121.fwd (5'CTC ACA CGG CCT CAT ACA G 3'; J02029 nucleotides 8196-8214) and/ AV122.rev (5'ACG CAG TTC AGG AGG CAC C 3'; J02029 nucleotides 8749-8767). Following an initial hot-start at 95°C for 10 mins, the following cycling conditions were used for the amplification: denaturation for 30 seconds at 95°C, annealing for 30 seconds at 55°C and extension for 45 seconds at 72°C. The reaction was cycled 35 times, with a final hold at 72°C for 10 minutes. For nested PCR, 4 µL of the first round PCR was used as template with primer set AV123MOD (5' TTG AAG AAT ACA CCA ACA TCC 3'; J02029 nucleotides 8246-8266) and AV124MOD (5’ACT CAA CCG GCG TGG ATG G 3’; J02029 nucleotides 8699-8717) and conditions similar to those outlined above except with an annealing temperature of 50°C and 30 cycles of amplification. All PCR reactions were performed in an ABI GeneAmp PCR System 9700 thermocycler (Applied Biosystems, Waltham, MA).
STLV-1 LTR sequences (460-bp) were cloned into TopoTA cloning kit (pCR2.1-TOPO/Top 10 cells; Invitrogen, Burlington, Ontario, Canada), purified using QIAprep Spin Miniprep Kit (Qiagen) and Sanger sequenced in both directions with an ABI 3130 (Applied Biosystems) using the AV123MOD and AV124MOD primers.

### 3.3.4 Phylogenetic analyses

Comparative PTLV-1 LTR (n=151) and tax (n=68) sequences were obtained from GenBank (Table 3.1) and aligned separately using ClustalW in MEGA6 (Tamura et al., 2013). 35 HTLV-1 LTR sequences from Papua-New Guinea (n=1), Australia (n=25), Vanuatu (n=8) and the Solomon Islands (n=1) were included to aid in dating the origin of STLV-1 in orangutans and the origin of HTLV-1 in the region and globally. HTLV-1 sequences from Africa were not included in the analyses since they have been previously shown to cluster strongly with African STLV-1 and also to reduce the computational complexity of the dataset.

jModelTest 2 (Darriba et al., 2012; Guindon and Gascuel 2003; Posada 2008) was used to infer the best-fitting evolutionary model for maximum likelihood (ML) phylogenetic analysis. For the PTLV-1 LTR and tax datasets the Hasegawa-Kishino-Yano (HKY) model (Hasegawa et al., 1985) with a gamma distribution (+G), and the Tamura-Nei model (Tamura and Nei, 1993) with a gamma distribution (TrN+G) were identified, respectively, as the two most appropriate models via the Bayesian Information Criteria. ML trees were generated in MEGA 6, with bootstrap support values based on 1,000 nonparametric replicates.

### 3.3.5 Bayesian Divergence Dating Models Tested

In order to more accurately explore PTLV divergence dating and evolutionary histories in Asia, we tested three different Bayesian models using BEAST v2.1.3 (Bouckaert et al. 2014).
Two of our models used archaeological calibration points, while our third used evolutionary rate estimates for the LTR region as presented by Van Dooren et al., (2004). For the first model, the “Modified Classical Archaeological” (MCA) calibration, we used a different calibration node for HTLV-1 then that used in other studies (e.g. Andonov et al., 2012; Lemey et al. 2005; Van Dooren et al. 2001). In our model we used the most recent common ancestor of the HTLV-1 strain PNG-1 (GenBank accession number M85207), and the Australia HTLV-1 clade as our calibration node at 45-55 kya. Previous studies have used the node directly ancestral to the HTLV-1 strain Mel-5 (GenBank accession number L20534) from the Solomon Islands as the calibration with a date of 40-60 kya. The earliest possible archaeological dates for the Solomon Islands is 32 kya, though a more recent date range associated with the Lapita culture (3-3.5 kya) is far more likely (Green et al., 2008; Sheppard, 2011; Wickler, 2001). Only one archaeological site on the Soloman Islands has been dated to 32 kya, while there are number of sites associated with the Lapita culture suggesting an increase in the movement of people into the Soloman Islands. We chose the HTLV-1_PNG-1 sequence as a calibrating node given recent genetic and archaeological findings suggesting modern humans first arrived in Papua New Guinea around 50 kya (Summerhayes et al., 2010; Vilar et al., 2008). Additional evidence suggests that Australia was occupied shortly thereafter by 45 kya (O’Connell and Allen, 2004). Because non-human primates have not been found in the region of Australasia/Melanesia, our archaeological calibration date assumes that infected individuals brought HTLV-1 with them as they migrated from Southeast Asia.

Our second “Most Likely Archaeological (MLA) calibration”, set the calibration of the PNG-1/Australia HTLV-1 clade (Node G) to 3-8 kya. While there is great debate as to the origins of the Austronesian peoples (Larson et al., 2007; Oppenheimer and Richards, 2001), we believe that our MLA calibration model is likely more accurate than the MCA calibration
because the time period between 3-8 kya corresponds to the Neolithic expansion and population growth associated with the development of agriculture in Papua New Guinea and Australia (Larson et al., 2007; Sacks et al., 2013). The Neolithic expansion is believed to have begun in Taiwan 12-6 kya according to human genetic data, or around 5 kya according to linguistic data (Gray et al., 2009; Kayser et al., 2006). This is supported by archaeological evidence indicating that between 12-10 kya there is a surge in the number of recorded archaeological sites in Thailand, northern Vietnam, Laos, Burma, Malaysia, Philippines and Talaud Islands, Borneo, Sumatra, Java, South Sulawesi, Roti/East Timor and the Northern Moluccas (O’Connor and Bulbeck, 2013). Assuming the increase in the number of radiometric dates from archaeological contexts from SE Asian sites at the Pleistocene/Holocene boundary could be taken as a marker of population increase, this would have led to a large wave of immigrants eventually expanding southeast into Papua New Guinea and Australia. The earliest boundaries of our model coincide with the origins of agriculture in Papua New Guinea, with true agriculture being visible in the archaeological record by ~7 kya (Denham et al., 2003). Also, paleodemographic research has identified a population surge in Australia between 4.4-3.7 kya based on radiocarbon data from archaeological sites (Williams, 2013). This surge in population also coincides with the introduction of the dingo (*Canis lupus dingo*) to Australia between 5-3.5 kya (Oskarsson et al., 2012; Sacks et al., 2013; Savolainen et al., 2004).

The most recent boundary for the MLA calibration model is based on the expansion of the Lapita peoples into Melanesia. One of the oldest known sites (site: SE-SZ-8) of the Lapita culture on the Santa Cruz Islands region of the Solomon Islands dates to 3.3-3.0 kya (Green et al., 2008). Another Solomon Island site called Kolombangara has been dated to at least 2.9-2.6 kya, but these dates are only based on Lapita pottery styles and not radiometric dates (Summerhayes and Scales, 2005). The site of Teouma is the oldest well-recognized Lapita site
(ca. 3.0 kya) on the Vanuatu island of Efate (Bedford et al., 2006; Kinasten et al., 2014; Storey et al., 2010). Thus, we feel confident that a date of 3.0 kya is appropriate as the lower boundary for our calibration of HTLV-1 in this model.

Our final model, called the “Pedigree Calibration Model” (PCM), used previously estimated evolutionary rates for the LTR region inferred by Van Dooren et al. (2004) based on pedigree data from families with a history of HTLV-1 infection. We set this evolutionary rate to $1.77 \times 10^{-6}$ to $3.65 \times 10^{-6}$ in order to account for both horizontal and vertical transmission of HTLV-1 (see discussion in Van Dooren et al., 2004). Because horizontal and cross-species transmission of PTLV-1 may be associated with higher substitution rates (Duffy et al., 2008; Van Dooren et al., 2004), and previous phylogenetic analyses have indicated cross-species transmission of PTLV-1, we used Van Dooren et al.’s estimates as evolutionary rate boundaries. We assigned a uniform distribution to this evolutionary rate, and assumed an arbitrary 50/50 split in the occurrence of vertical vs. horizontal transmission.

For the BEAST analyses, the LTR sequences were aligned using MAFFT v7 (multiple sequence alignment based on the fast Fourier transform) (Katoh and Standley, 2013) and then edited manually. The final alignment contained 151 taxa covering 500 nucleotide positions, including indels. Each of the calibration models utilized a log normal relaxed molecular clock, a uniform distribution for the calibration dates, and a birth-death tree prior and were run in duplicate for 100 million Markov Chain Monte Carlo (MCMC) iterations with a 10% burn-in. Log and tree files from each run and calibration model were combined using LogCombiner v1.8.2. Convergence of the chain sampling was checked in the program Tracer v1.6 for effective sample sizes (ESS), all of which were > 423. Trees were saved every 100,000 generations and the tree with the maximum product of the posterior clade probabilities (maximum clade
credibility tree) was chosen from the posterior distribution of 10,001 sampled trees after burning in the first 1,000 sampled trees with the program TreeAnnotator v1.8.2. Trees were viewed in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

3.4 Results

Of the 68 individual orangutans sampled, only one, a wild-born sub-adult male named Botak (1/68, 1.5%), tested positive by qPCR for STLV-1 tax sequences. STLV-1 LTR sequences were also obtained from Botak confirming infection of this animal. Phylogenetic analysis of the tax sequences using ML analysis revealed a statistically well-supported orangutan clade positioned sister to a clade comprising Asian HTLV-1 (Mel-5), STLV-1 (M. tonkeana) sequences, as well as African STLV-1 sequences (Fig. 3.1). It should be noted however, that this statistically well-supported orangutan clade is limited by comprising only two sequences.

ML analysis of the LTR sequences revealed a well supported monophyletic African STLV-1 clade weakly positioned sister to a strongly supported Asian STLV-1 clade (bootstrap=99) containing macaques (Fig. 3.2). Short branch lengths for the African clade likely reflect a shorter evolutionary history for PTLV-1 on that continent confirming previous analyses reported by Van Dooren et al. (2007) assuming equivalent evolutionary rates. Alternatively, the longer branch lengths for Asian STLV-1 could also reflect a faster evolution for those strains. A poorly supported HTLV-1 clade is sister to a single STLV-1 sequence from of a macaque species from Sulawesi (M. nigra). The two new orangutan STLV-1 sequences VL04 (this study) and Ph711 (AY141165) form a very robust monophyletic clade (bootstrap=94) with the two previously published Pongo STLV-1 sequences (U53562 and AY899816). The orangutan clade shares a MRCA with all PTLV-1 LTR sequences except those of M. arctoides. The M. arctoides STLVs are believed by some to represent a new strain (STLV-5) (Ayoub et al., 2013; Liegeois
et al., 2008; Mahieux and Gessain, 2011). Interestingly, in this analysis LTR sequences from *M. mulatta* are paraphyletic, appearing in three distinct clades within Asian STLV-1. Unfortunately, there is a lack of provenance information for these individuals which would help to determine any geographical/subspecies variations in *M. mulatta*. The two hylobatid STLV-1 LTR sequences are closely associated with *Macaca* STLV-1 sequences. The *H. pileatus* sequence forms a very strongly supported clade (bootstrap=99) with two *M. mulatta* sequences, while the *S. syndactylus* LTR sequence forms a clade (bootstrap=93) with a macaque species from Sulawesi (*M. maura*). The phylogenetic relationship between those two hylobatid clades remains unresolved. Both the *S. syndactylus* and *H. pileatus* STLV-1 are likely the result of captive cross-species transmissions, as suggested previously for *S. syndactylus* by Van Dooren et al., (2007). The HTLV-1 region of the tree is only moderately supported (bootstrap=61), and the New Guinea PTLV-1 appears as an unresolved branch withing the Melanesian HTLV-1s.

### 3.4.1 Modified Classical Archaeological calibration model (MCA) phylogeny

Our Bayesian-inferred tree based on the MCA calibration of 45 – 55 kya for the node representing the divergence of the HTLV-1 PNG strain (Node G) (Fig. 3.3) is broadly similar to the ML tree, though a few important differences are apparent. In the ML tree, Node I for the *Pongo pygmaeus* clade shares a MRCA with all other PTLV-1s (Node H). This is not the case in the MCA tree as the *Pongo* clade is now sister to a macaque clade. Like the ML tree, the MCA Bayesian tree does not exhibit a monophyletic *M. fuscata* clade, with a single *M. mulatta* sequence (AY141175) again nested within this grouping (PP=1). *H. pileatus/M. mulatta* clade is resolved in this tree with strong support (PP=1) and shares a MRCA with a clade composed of African STLV-1, HTLV-1, and STLV-1s from two macaques and a siamang. The Austromelanesian HTLV-1 clade is well supported (Node D, PP=1). The PNG sequence is now
resolved within the strongly supported (PP=1) HTLV-1 clade. Another major difference between the ML and MCA trees is the position of the African STLV-1 clade, which is now sister to the HTLV-1 clade but with only weak support. As with the ML tree, *M. mulatta* continues to be paraphyletic appearing in three distinct clades in the Bayesian-inferred tree.

### 3.4.2 Most Likely Archaeological calibration (MLA) model phylogeny

The Bayesian-inferred tree (Fig. 3.4) using the MLA calibration of 3-8 kya for the node defining the divergence of the HTLV-1 PNG sequence (Node G), shares topological characteristics with both the ML and MCA trees. The first major difference is in the position of the mixed clades containing gibbon and macaque sequences. In the MLA tree the *M. mulatta/H. pileatus* clade is sister to a clade containing sequences from *M. tonkeana, M. maura* and *S. syndactylus*. Another major difference is the placement of the HTLV-1 clade which in the MCA tree was sister to the African STLV-1 clade (Node C). In the MLA tree the HTLV-1 clade has is sister to a clade of macaque STLV-1 sequences, and this larger clade shares a MRCA with the African STLV-1s. *M. mulatta* remains paraphyletic appearing in three clades in the MLA model tree.

### 3.4.3 Pedigree Calibration Model (PCM) topology

The Pedigree Calibration tree (Fig. 3.5) inferred using the evolutionary rates reported by Van Dooren et al., (2004) differs in several important ways from both the MCA and MLC trees. The *M. arctoides* clade (PP=1) remains at the base of the tree, and the *Pongo* STLV-1 sequences again form a strongly supported clade (PP=1). The African STLV-1 clade is sister to an Asian STLV-1 clade comprising sequences from *M. mulatta, M. fuscata, M. nemestrina*, and *M. nigra*. The HTLV-1 clade now shares a MRCA with this African STLV-1 and macaque clade. The
HTLV-1 topology remains the same with strongly supported clades. Like the other Bayesian trees, *M. mulatta* sequences remain paraphyletic, appearing in three distinct clades, while the *S. syndactylus* and *H. pileatus* sequences are mixed with various *Macaca* STLV-1 sequences, and thus may reflect cross-species transmission events in captivity. Our inferred evolutionary rate for this tree was 2.83E-06 (95%HPD 1.83E-06-3.68E-06) which is in broad agreement with those published by Van Dooren et al., (2004) (i.e. 1.77E-06 to 3.65E-06).

### 3.4.4 Divergence Dates inferred from the Bayesian analyses

Bayesian analyses allowed us to co-infer the divergence dates with the phylogenetic relationships and further explore the evolutionary histories of PTLV-1. For this study we focused on estimating the time to most recent common ancestor (TMRCA) for the following nodes and clades they define: A) African STLV-1s, B) HTLV-1, STLV-1 Africa, and Asian STLV-1 including those from *M. fuscata, M. mulatta, M. nemestrina, M. nigra, M. maura, M. tonkeana, H. pileatus*, and *S. syndactylus*, C) macaque STLV-1 sequences, or African STLV-1 and HTLV-1. D) HTLV-1, E) HTLV-1 in Melanesia, F) HTLV-1 in Australia, G) Australian and Papua New Guinea HTLV-1, H) *Macaca/Pongo* STLV-1, and I) *Pongo* STLV-1. In all three of our Bayesian models Node B defines a clade containing all HTLV-1, STLV-1 Africa, and STLV-1 sequences from *M. fuscata, M. mulatta, M. nemestrina, M. nigra, M. maura, M. tonkeana, H. pileatus*, and *S. syndactylus* sequences. These inferred divergence dates using each calibration model are summarized in Table 3.2. For all nodes the MCA model yielded divergence time estimates roughly 4 to 13 times older than the MLA and PCM models. The PCM consistently yielded estimates 1.4 to 3 times older than the MLA model but the 95% HPDs overlapped in these two models suggesting the MLA model is more comparable to the PCM than the MCA which comprises older calibration points. Likewise, the nucleotide substitution rates of the MLA and
PCM are in broad agreement with each other, with both showing evolutionary rates one log faster than that of the MCA, accounting for the increased divergence dates of the MCA model. Important to the present study, the TMRCA of the MacacalPongo STLV-1 clade varies between 17.2-207.5 kya, while the TMRCA of Pongo STLV-1 ranges from 3.2-37.5 kya depending on the calibration model. Our three models also allowed us to estimate the timing of the introduction of PTLV-1 into humans in Australomelanesia. In contrast with the older TMRCA dates inferred using the MCA calibration model of 54.7 kya, we found a much earlier origin of PTLV-1 in humans between 4.6-9.7 kya for the PCM and MLA models, respectively.

3.5 Discussion

The PCR prevalence of STLV-1 in orangutans in our study (1.6%, n=68) was similar to that described by Verschoor et al. (1998) (1.4%, n=143) based on Western blot (WB) testing of animals at a rehabilitation centre in Sabah, Malaysia. Our observed prevalence was, however lower than the 7.3% (3/41) reported by Ibuki et al. (1997) using serological assays, but similar to their PCR results (2.4%, 1/41). Our phylogenetic analysis of new and existing orangutan STLV-1 tax and LTR sequences showed that the orangutan STLV-1 clade was positioned within the diversity of macaque STLV-1 clades similar to that observed by others (e.g. Ibuki et al., 1997; Van Dooren et al., 2007; Verschoor et al., 1998). In all phylogenies the orangutan (Pongo sp.) STLV-1 sequences formed a strongly supported monophyletic clade nested within macaque (Macaca sp.) STLV-1 clades suggesting the origin of STLV-1 in orangutans may be the result of ancient cross-species transmission from macaques. This may not be surprising in that orangutans and macaques are sympatric both as extant and fossil species, and that STLV-1 clades are known to cluster geographically (Vandamme et al., 1998; Van Dooren et al., 2001). Importantly, however, this interpretation may be limited by the absence of phylogenetic analyses of STLV-1
sequences from other NHPs known to be sympatric with orangutans, such as colobines
(*Presbytis* and *Nasalis*), and gibbons (*Hylobates* and *Symphalangus*). Without extensive
sampling of these extant genera it is impossible to identify macaques as the definitive source of
orangutan STLV-1.

*Macaca mulatta* STLV-1 was found to be paraphyletic in all LTR tree topologies which
may be explained by the complex evolutionary history of macaques. Mitochondrial DNA
(mtDNA) studies have shown two distinct genetic populations referred to as the Eastern and
Western haplotypes (Hoelzer and Melnick, 2005). These two mtDNA haplotypes have been
hypothesized to result from the major glacial expansion within the Bramaputra River valley in
India and Bhutan approximately 180 kya (Hoelzer and Melnick, 2005). Additional evidence
supporting this date can be seen in the genetic variation between Indian and Chinese subspecies.
For example, by examining genomic SNPs, Hernandez et al. (2007) estimated that the Indian and
Chinese populations separated at approximately 162 kya. The clade identified in our study
containing both *M. mulatta* and *M. fascicularis* STLV-1 sequences may reflect contact within the
hybrid zone between these two species in northern Thailand, and subsequent introgression down
the Malay Peninsula to the Isthmus of Kra (Fooden, 1980; Tosi et al., 2003). Unfortunately,
provenance information for the eight individuals from these two STLV-1-infected macaque
species was not reported in GenBank, and to date no peer-reviewed publication is associated with
these sequences (Meertens and Gessain, unpublished). Thus, the phylogenetic position of these
macaque LTR sequences may either reflect host evolutionary history, potential cross-species
transmission of STLV-1 associated with hybridization in the wild, or interspecies transmission
during captivity.
One main goal of our study was to use Bayesian inference to estimate the TMRCA of PTLV-1 in orangutans given divergence dates were not reported in previous studies examining PTLV-1 in *Pongo* (e.g. Ibuki et al., 1997; Van Dooren et al., 2007; Verschoor et al., 1998). We sought to explore the evolutionary histories of PTLV-1 using new and modified calibration dates based on recent archaeological information regarding population history and migration of ancient humans (Bedford et al., 2006; Denham et al., 2003; Gray et al., 2009; Green et al., 2008; Kayser et al., 2006, 2008; Kinaston et al., 2014; Larson et al., 2007; O’Connor and Bulbeck, 2014; Oskarsson et al., 2012; Sacks et al., 2013; Savolainen et al., 2004; Sheppard, 2011; Summerhayes and Scales, 2005; Storey et al., 2010; Wickler, 2001; Williams, 2013). Newer bioinformatics methods, including Bayesian analysis, have facilitated inference of viral divergence dates of nucleotide sequences by utilizing various molecular clock calibrations or nucleotide substitution rates (Drummond et al., 2003; Faria et al., 2011; Holmes, 2011; Pybus and Rambaut, 2009). Studies using these methods have shed light on the emergence and evolutionary histories of various pathogens, including retroviruses (Holmes and Grenfell, 2009; Katzourakis et al., 2014; Lemey et al., 2006; Switzer et al., 2009). Despite the development of these new analytical techniques, comparatively little is known about the evolutionary history of STLV-1 in orangutans (Ibuki et al., 1997; Van Dooren et al., 2007; Verschoor et al., 1998). Of the research examining PTLV-1 evolution, use of a 40–60 kya calibration for the divergence of HTLV-1 Melanesia may have inflated the estimated origin of HTLV-1 by 1,000 fold. Others have shown the importance of selecting priors for accurately inferring divergence times using Bayesian methods (Heled and Drummond 2012; Inoue et al., 2010; Warnock et al., 2015).

Our three Bayesian models yielded different divergence date estimates for orangutan STLV-1, with some overlap in the 95% HPDs. Although we believe our modification (i.e., MCA) to the often-used archaeological calibration of 45 – 55 kya is more accurate, estimated
dates for any given node generated using this model must be considered as the earliest possible boundary. Although it is likely accurate that the early appearance of modern humans in Australia and New Guinea occurred between 45-55 kya (O’Connell and Allen, 2004; Summerhayes et al., 2010; Vilar et al., 2008), it is far more likely that NHP zoonotic viruses were introduced to these regions in conjunction with the much larger population sizes associated with the Austronesian expansion to Melanesia and coastal New Guinea, probably starting in Taiwan just over 5 kya based on linguistic data (Gray et al., 2009), or perhaps as early as 6-12 kya according to human genetic data (Kayser et al., 2006). Similarly, it is just as likely that HTLV-1 was brought to Australia as people migrated from India to Australia (Pugach et al., 2013) at the time of the introduction of the dingo between 3.5 and 5 kya (Oskarsson et al., 2012; Sacks et al., 2013; Savolainen et al., 2004). Recent studies have suggested that population sizes increased dramatically in Australia only about 3-5 kya (Williams, 2013). As such, our MCA model, which is very similar to what has become the standard in the field (e.g., Andonov et al., 2012; Lemey et al., 2005; Van Dooren et al., 2001, 2005), sets a calibration that is likely far too early, perhaps by as much as 30 to 40 kya.

Our estimated dates for the diversification of orangutan-specific STLV-1 fell into one of two geological age categories. Our PCM and MLA models indicated that this strain arose during the Holocene, about ~7,300 and ~3,200 years before present (ybp), respectively, or Holocene to terminal Late-Pleistocene if 95% HPDs are considered (95% HPD, 1-13.4 kya). The MCA model provides a Late Pleistocene date for the diversification of orangutan STLV-1, between 12.3 and 75.4 kya. We estimate that the origin of orangutan STLV-1 is possibly the product of cross-species transmission from macaques that occurred sometime between 107.5 and 343.9 kya based on the MCA model, and between about 7 and 77.1 kya based on both the PCM and MLA models. These estimates (including 95% HPDs) span a portion of the Holocene (11.7 kya to
present), the entire Late Pleistocene (126-11.7 kya) and about 33% of the Middle Pleistocene (781-126 kya). Although it is difficult to be more precise regarding when STLV-1 transmission to orangutans may have occurred, it seems likely that this transmission occurred well after the separation of Borneo and Sumatran Pongo species, for which some of the earliest molecular date estimates are between 400 kya (Locke et al., 2011) and 1.31 mya (Perelman et al., 2011). It is important to note, however, that currently STLV-1 sequences are not available from Pongo abelii to test this hypothesis. Nonetheless, analysis of additional STLV-1 from larger numbers of different Asian primate species are required to more fully explore these scenarios.

3.5.1 Paleobiogeographical and environmental context for PTLV evolution in Asia

The date range for the origin of orangutan STLV-1 from our MCA model encompasses much of the Middle Pleistocene when 95% HPDs are considered (~108-344 kya). Given the low probability that HTLV-1 was introduced to Australia, New Guinea, and Melanesia 45-55 kya, as assumed by the classical archaeological calibration model, we feel more confident in assigning the origin of orangutan STLV-1 to the Late Pleistocene or Holocene as indicated by our PCM and MLA models. A description of paleoenvironmental trends during the Pleistocene and early Holocene, therefore, may help to better understand the ecological contexts of STLV-1 evolution in orangutans.

The paleoenvironmental conditions during the Pleistocene of Southeast Asia were estimated to be cooler and drier during glaciation events, which shrunk rainforests and created more open savannah or mixed forest-savannah type habitats (Bird et al., 2005; Harrison et al., 2006; Louys and Meijaard, 2010; Meijaard, 2003; Page et al., 2004; Tougard, 2001; Wong, 2011; Wurster et al., 2010). During the Late Pleistocene, at 50 kya, sea levels would have been
approximately 50-80 metres below those presently observed (Chappell and Shackleton, 1986; Rohling et al., 1998). Lower sea levels of this magnitude would have, at a minimum, connected the Malay Peninsula, Java, Sumatra and Borneo to form Sundaland (Voris, 2000), allowing species to migrate freely around the region prior to the early Holocene at about 10 kya (Meijaard, 2003 and references therein), after which Borneo began experiencing increased temperatures and humidity (Hunt and Premathilake, 2012). The colder, drier periods were also characterized by sea levels that were much lower, allowing for the migration of plants and animals across much of Sundaland (Bird et al., 2005; Chappell and Shackelton, 1986; Harrison et al., 2006; Louys and Meijaard, 2010; Meijaard, 2003; Page et al., 2004; Tougard, 2001; Voris 2000). Despite this cooler and drier climate, there is extensive evidence for rainforest refugia during the Late Pleistocene throughout Southeast Asia, from southern China to Borneo (Wong, 2011). Glacial periods were interrupted by shorter interglacial periods when climates in the region were warmer and wetter, allowing for rainforest expansion (Harrison et al., 2006; Louys and Meijaard, 2010; Wong, 2011). Perhaps one of the most significant events in Sundaland was the eruption of Mount Toba about 73 kya (Williams et al., 2009; Williams, 2012). All of our more recent Bayesian estimates place the MRCA of orangutan STLV-1 in the Late Pleistocene after the eruption of the Mount Toba Volcano. It has been suggested that this major geologic event caused the climate to cool leading to massive deforestation in South Asia (Williams et al., 2009).

Paleoenvironmental reconstructions indicate that the Late Pleistocene Malay Peninsula was at least always partly forested, with either open woodlands or closed evergreen rainforest depending on the warming/cooling trends of the last glaciation (Louys and Meijaard, 2010). The Southeast Asian forests and woodlands, however, may have been broken up in Sundaland by a north-south running savannah corridor through its interior (Bird et al., 2005; Heaney, 1991). The commencement of the Holocene brought a warmer, wetter climate to Southeast Asia and sea
levels began to rise dramatically by as much as 75-85 meters (Hanebuth, 2000). Consequently, there was an increase in the number of swamps, lakes and tropical broadleaf forests (Hope et al., 2004). The Pleistocene/Holocene transition in northeastern Thailand saw the pine/oak forests that had been dominant for the previous 20-30 kya replaced by tropical broadleaf forests (Penny, 2001).

There are a number of fossil orangutan localities dating from the end of the Middle Pleistocene through to the Holocene boundary that correspond to our estimates of orangutan STLV-1 origins and diversification (Table 3.3). In China, the Wuyun and Loushan Caves contain orangutan fossils found in association with macaques (*Macaca* sp.), leaf monkeys (*Presbytis* sp.), gibbons (*Nomascus* sp.), langurs (*Pygathrix* sp.), and lutungs (*Trachypithecus* sp.) that date to our inferred time period. In Southeast Asia, there are a number of well-dated Pleistocene fossil locations yielding several orangutan species including *Pongo pygmaeus*, *P. p. ciochoni*, *P. p. weidenreichi*, and *P. p. fromageti* (Cameron, 2001; Drawhorn, 1995; Ibrahim et al., 2013; Schwartz et al., 1995). In Vietnam, the sites of Lang Trang Caves (500-146 kya), Tham Om Cave (250-140 kya), Hang Hum Cave (140-80 kya); Duoi U’oi Cave (66 kya), Keo Leng (30-20 kya) and Nguom Rock Shelter (~23 kya) all date to the Middle or Late Pleistocene (Bacon et al., 2008; Ibrahim et al., 2013; Olsen and Ciochon, 1990). A number of sympatric fossil primates are also found at these sites, including macaques (*Macaca* sp., *Macaca mulatta*, *Macaca assamensis*), langurs (*Seminopithecus* sp.), gibbons (*Hylobates* sp.), leaf monkeys (*Presbytis* sp.) and several unidentified colobines (Bacon et al., 2008; Olsen and Ciochon, 1990).

Orangutan fossil sites in Laos, Thailand and mainland Malaysia including Tam Hang, Thum Wiman Nakin (both older than 169 kya), and Ban Fa Sui I or “The Cave of the Monk” (32-19 kya), all date to the Middle or Late Pleistocene (Bacon et al., 2011; Esposito et al., 2002;
Sympatric fossil primates from these sites include macaques (*Macaca* sp., *Macaca nemestrina, Macaca andersoni, Macaca assamensis*), colobines (*Presbytis obscura, Pygathrix neamus*), and gibbons (*Hylobates* sp.) (Bacon et al., 2011; Chaimanee and Jaeger, 1993; Zeitoun et al., 2010). The site of Badak Cave C in peninsular Malaysia has yielded fossil remains of orangutans and *Macaca nemestrina*. In Sundaland, fossil locations on the islands of Java (Punung 128-115 kya), Sumatra (the Padang Highlands [128-118 kya], Lida Ajer Cave [80 kya], and Djamboe Cave [70-60 kya]) and Borneo (Niah Cave [45 kya] and the Holocene sites at Bau and Madai Caves, and Guah Sireh have yielded orangutan fossils, all dating to the time period within our estimates for the origin of orangutan STLV-1 (Barker et al., 2007, 2009; Cranbrook, 1986; Drawhorn, 1995; Harrison, 1996, 2000; Ibrahim et al., 2013; Piper and Rabett, 2009; Westaway et al., 2007). There are a number of potential fossil orangutan species at these Indonesian and Malaysian sites, including *Pongo* sp, *P. deroisi djamboensis, P. deroisi, P. pygmaeus paleosumatrensis*, and *P. deroisi lidaajarensis*. A number of sympatric primate species are also found at these sites, including macaques (*Macaca* sp., *Macaca nemestrina, Macaca fascicularis*), leaf monkeys (*Presbytis* sp., *Trachypithecus cristatus*), and hylobatids (*Hylobates muelleri, Symphalangus syndactylus*) (Barker et al., 2007; 2009; Cranbrook, 1986; Drawhorn, 1995; Harrison, 1996, 2000; Ibrahim et al., 2013; Piper and Rabett, 2009; Westaway et al., 2007).

The combined fossil and paleoenvironmental evidence indicates orangutans were present on both the Indochinese Mainland and Sundaland from the Middle Pleistocene to the Late Pleistocene/Early Holocene boundary. Forests may have expanded and contracted, but suitable forest refugia and mixed forest habitats must have been present to support orangutans and other primates. We also know that orangutans shared Pleistocene habitats with various primate taxa including *Macaca, Presbytis, Semnopithecus, Trachypithecus*, and at least two genera of
hylobatids. Sympatry with these other primate taxa suggests that the cross-species transmission of STLV-1 to Pongo sp. could have occurred through any of the above primate taxa. Although current evidence points to the genus Macaca, more complete sampling of other primates known to be sympatric with orangutans is required to determine with more precision the exact source of STLV-1 in Pongo sp.

3.5.2 Possible first introduction of HTLV-1 in Asia

The HTLV-1 Australian clade dates to between 1.7 and 20.8 kya based on our models, an estimate which is similar to that proposed by Cassar et al. (2013) (9.1 kya) for HTLV-1 in Australia. Strong evidence supporting a more recent origin of HLV-1 in Australia can also be found in our pedigree model (PCM). We speculate therefore that the human expansion into Australasia around 3-5 kya, a period when the dingo was introduced to Australia, may also be responsible for introducing HTLV-1 to the continent. In our estimation the PCM model conforms most closely with what is known about human prehistory in Australasia. Despite their reservations, using the pedigree-based substitution rates presented in Van Dooren et al. (2004) may provide the best means for estimating PTLV-1 clade divergence. Further support for our PCM model can be inferred from our estimated dates for the introduction of HTLV-1 into Melanesia at ~2.7 kya. This corresponds closely with the movement of the Lapita peoples into Melanesia. Early Melanesian Lapita sites date anywhere from between 3-3.3 kya (Green et al., 2008) to between 2.6-2.9 kya (Summerhayes and Scales, 2005) showing strong support for our HTLV-1 dates using this model.

Recent research has shed new light on the age and evolutionary history of HTLV-1. Indigenous people in Canada have been found to have been host to HTLV-1 since ancient times, dating to between 5.4 and 11.9 kya. The HTLV-1 found in aboriginal Canadians exhibits recent
shared ancestry with East-Asian HTLV-1s (Andonov et al., 2012). Parenthetically, in that the study by Andonov et al., (2012) the pedigree-based substitution rate was used to estimate HTLV-1a divergence from East Asian HTLV-1a LTR sequences (2.1 kya) based on analyses of samples obtained from two Canadian Inuit women. Their estimated divergence dated matches closely the mtDNA divergence dates between Inuit and Siberian populations (2.7±1.8 kya), thus supporting our suggestion that the pedigree model provides a generally accurate estimate of the PTLV-1 substitution rate (Andonov et al., 2012).

3.6 Conclusions

Using various Bayesian models with different molecular clock calibrations we obtained dates placing the orangutan STLV-1 clade within the Late Pleistocene to Early Holocene. The potential cross-species transmission of STLV-1 to orangutans, possibly from macaques, likely occurred during the Late Pleistocene. Analysis of STLV-1 sequences from primate species known to be sympatric with orangutans today and during the Pleistocene, such as gibbons and langurs, is needed to better characterize the timing and nature of this cross-species transmission event. Our models have also allowed us to study the evolution of HTLV-1 in Australasian populations. Our results support an early Asian origin for HTLV-1 that likely resulted from an ancient cross-species transmission from macaques. Using the PCM model, our inferred dates for the MRCA of HTLV-1 in Australia at ~3.7 kya corresponds closely to the appearance of human genome-wide single nucleotide polymorphisms (SNPs) whose origins are from the Indian subcontinent, and possibly coincided with the introduction of dingoes to the island, and strongly support the utility of this model based on HTLV-1 nucleotide substitution rates. Similarly, the introduction of HTLV-1 into Melanesia at ~2.7 kya corresponds to the movements of the Lapita
peoples into the region. Our results also provide an important future reference for calibrating information essential for PTLV evolutionary timescale inference.
3.7 Tables

Table 3.1 Primate T-lymphotropic virus sequences used in this study.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Genbank Accession Number</th>
<th>Location</th>
<th>Viral Region</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Allenopithecus nigroviridis</em></td>
<td>AY026856-AY026858</td>
<td>Africa-captive</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Cercopithecus mona</em></td>
<td>AY026853</td>
<td>Africa-wild orphan</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Cercocebus agilis</em></td>
<td>AY496607-AY496611, AY496613-AY496614, AM746641-AM746645, AM746628</td>
<td>Africa-wild</td>
<td>tax</td>
</tr>
<tr>
<td><em>Cercocebus agilis</em></td>
<td>AY496626-AY496627, AY496629-AY496634, EU152273-EU152274, AF384872</td>
<td>Africa-wild</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Cercocebus atys</em></td>
<td>AY876036-AY876043</td>
<td>Africa-captive</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Cercocebus torquatus</em></td>
<td>AY026855</td>
<td>Africa-wild orphan</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Cercocebus torquatus</em></td>
<td>HE589477</td>
<td>Africa-wild</td>
<td>tax</td>
</tr>
<tr>
<td><em>Cercopithecus ascanius</em></td>
<td>AY899812, JN210960</td>
<td>Africa-wild</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Cercopithecus ascanius</em></td>
<td>JN210942, JN210945</td>
<td>Africa-wild</td>
<td>tax</td>
</tr>
<tr>
<td><em>Cercopithecus cephus</em></td>
<td>AM712677</td>
<td>Africa-local captive or wild</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Cercopithecus cephus</em></td>
<td>AM746629, AM746632, AM746636, AM746638-AM746639, HE589473, HE589479, HE589483</td>
<td>Africa-wild</td>
<td>tax</td>
</tr>
<tr>
<td><em>Cercopithecus mona wolfi</em></td>
<td>JN210956, JN210958</td>
<td>Africa-wild</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Cercopithecus mona wolfi</em></td>
<td>JN210936, JN210939, JN210954-JN210955</td>
<td>Africa-wild</td>
<td>tax</td>
</tr>
<tr>
<td>Species</td>
<td>Accession Numbers</td>
<td>Location</td>
<td>Taxonomic Notes</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>--------------------</td>
<td>---------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td><em>Cercopithecus neglectus</em></td>
<td>JN210937</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cercopithecus nictitans</em></td>
<td>EU152272, EU152275, EU152276, AM712674, AM712679, AM712680</td>
<td>Africa-wild, orphan, local captives</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Cercopithecus nictitans</em></td>
<td>AM746635, AM746637, AM746640, HE589471-HE589472, HE589474-HE589476, HE589481</td>
<td>Africa-wild</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Cercopithecus pogonias</em></td>
<td>AM712675, AM496628</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cercopithecus pogonias</em></td>
<td>HE589478, HE589472, HE589474-HE589476, HE589481</td>
<td>Africa-wild</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Chlorocebus aethiops</em></td>
<td>AY026846-AY026852, AF074966</td>
<td>Africa-wild, wild orphan</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Chlorocebus tantalus</em></td>
<td>AF074966</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Gorilla gorilla</em></td>
<td>AY899813, AM496635</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Gorilla gorilla gorilla</em></td>
<td>AY263382</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Homo sapiens</em></td>
<td>M85207</td>
<td></td>
<td>Papua New Guinea</td>
</tr>
<tr>
<td><em>Homo sapiens</em></td>
<td>L02534</td>
<td></td>
<td>Melanesian-Mel 5</td>
</tr>
<tr>
<td><em>Homo sapiens</em></td>
<td>KC786962-KC786965, AM549875-AY549878</td>
<td>Vanuatu</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Homo sapiens</em></td>
<td>JX891478-JX891479, KC786899-KC786919, KF242505-KF242506</td>
<td>Australia</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Hylobates pileatus</em></td>
<td>KC283222</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lophocebus albigena</em></td>
<td>AM712678</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lophocebus albigena</em></td>
<td>AM746631</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Macaca arctoides</em></td>
<td>AY141170-AY141174, AY590142</td>
<td>Asia-captive</td>
<td>LTR</td>
</tr>
<tr>
<td>Species</td>
<td>Accession Numbers</td>
<td>Location</td>
<td>Type</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Macaca arctoides</td>
<td>AY590142, U76625</td>
<td>Asia-captive</td>
<td>tax</td>
</tr>
<tr>
<td>Macaca fascicularis</td>
<td>AY141156</td>
<td>Asia-unknown</td>
<td>LTR</td>
</tr>
<tr>
<td>Macaca fuscata</td>
<td>AY141162-AY141169</td>
<td>Asia-unknown</td>
<td>LTR</td>
</tr>
<tr>
<td>Macaca maura</td>
<td>DQ076650</td>
<td>Asia-wild</td>
<td>LTR</td>
</tr>
<tr>
<td>Macaca mulatta</td>
<td>AY141157-AY141161, AY141175</td>
<td>Asia-unknown</td>
<td>LTR</td>
</tr>
<tr>
<td>Macaca nemestrina</td>
<td>M57591, M11373</td>
<td>Asia</td>
<td>LTR</td>
</tr>
<tr>
<td>Macaca nigra</td>
<td>DQ076656</td>
<td>Asia-captive</td>
<td>LTR</td>
</tr>
<tr>
<td>Macaca sylvanus</td>
<td>AY899810-AY899811, DQ076652-DQ076655</td>
<td>Africa-wild</td>
<td>LTR</td>
</tr>
<tr>
<td>Macaca tonkeana</td>
<td>Z46900</td>
<td>Asia</td>
<td>LTR, tax</td>
</tr>
<tr>
<td>Mandrillus sphinx</td>
<td>AF045932-AF045933 AY496637-AY496638</td>
<td>Africa-captive, Africa-pet</td>
<td>LTR</td>
</tr>
<tr>
<td>Mandrillus sphinx</td>
<td>HE589480, HE589482, FJ755931</td>
<td>Africa-wild</td>
<td>tax</td>
</tr>
<tr>
<td>Miopithecus ogouensis</td>
<td>AY496636</td>
<td>Africa-wild</td>
<td>LTR</td>
</tr>
<tr>
<td>Pan troglodytes</td>
<td>L75787, U86376, AY267833- AY267835,</td>
<td>Africa</td>
<td>LTR</td>
</tr>
<tr>
<td>Pan troglodytes vellerosus</td>
<td>AY263381, AF384871</td>
<td>Africa-wild</td>
<td>LTR</td>
</tr>
<tr>
<td>Pan troglodytes verus</td>
<td>AY899815, AF539747</td>
<td>Africa-wild, local captives</td>
<td>LTR</td>
</tr>
<tr>
<td>Pan troglodytes verus</td>
<td>AF539747</td>
<td>Africa-wild</td>
<td>tax</td>
</tr>
<tr>
<td>Papio anubis</td>
<td>AY026842-AY026844, AF230485</td>
<td>Africa-local captive, wild</td>
<td>LTR</td>
</tr>
<tr>
<td>Species</td>
<td>Accession Numbers</td>
<td>Region</td>
<td>Type</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------------------</td>
<td>--------------------</td>
<td>---------</td>
</tr>
<tr>
<td><em>Papio hamadryas cynocephalus</em></td>
<td>L60024</td>
<td>Africa</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Piliocolobus badius</em></td>
<td>AY333376, AY267837</td>
<td>Africa-wild</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Piliocolobus tholloni</em></td>
<td>JN210959, JN210961-JN210965</td>
<td>Africa-wild</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Piliocolobus tholloni</em></td>
<td>JN210938, JN210940-JN210941, JN210943, JN210946-JN210952</td>
<td>Africa-wild</td>
<td><em>tax</em></td>
</tr>
<tr>
<td><em>Pongo pygmaeus</em></td>
<td>AY141165, AY899816, U53562, This study</td>
<td>Asia</td>
<td>LTR</td>
</tr>
<tr>
<td><em>Pongo pygmaeus</em></td>
<td>This study, Y13146</td>
<td>Indonesian Borneo-wild born, unknown</td>
<td><em>tax</em></td>
</tr>
<tr>
<td><em>Symphalangus syndactylus</em></td>
<td>DQ076651</td>
<td>Asia-captive</td>
<td>LTR</td>
</tr>
</tbody>
</table>
Table 3.2 Time to most recent common ancestor (TMRCA) inferred from Bayesian analysis of PTLV-1 long terminal repeat (LTR) sequences\(^1\).

<table>
<thead>
<tr>
<th>TMRCA</th>
<th>Node ID(^2)</th>
<th>Evolutionary Model</th>
<th>Pedigree Calibration Model (PCM) Van Dooren et al.(^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Modified Classical Archaeological calibration model (MCA)</td>
<td>Most Likely Archaeological calibration model (MLA)</td>
</tr>
<tr>
<td>Clock date calibration</td>
<td>HTLV-1_PNG (45-55)</td>
<td>HTLV-1_PNG (3-8)</td>
<td>None</td>
</tr>
<tr>
<td>Clock rate calibration(^3)</td>
<td>None</td>
<td>None</td>
<td>1.8E-06 - 3.7E-06</td>
</tr>
<tr>
<td>PTLV root</td>
<td>335.5 (146.8-585.2.4)</td>
<td>28.2 (9.9-58.8)</td>
<td>69.7 (37.2-115.5)</td>
</tr>
<tr>
<td>African STLV-1</td>
<td>A</td>
<td>71.3 (36.9-116.2)</td>
<td>5.9 (2.5-12)</td>
</tr>
<tr>
<td>HTLV-1, STLV-1 Africa, and STLV-1 from \textit{M. fuscata, M. mulatta, M. nemestrina, M. nigra, Macaca maura, Macaca tonkeana, Symphalangus syndactylus, and Hylobates pileatus}</td>
<td>B</td>
<td>167.9 (87.3-270.7)</td>
<td>14 (6.4-28.5)</td>
</tr>
<tr>
<td>Macaque STLV-1 sequence or African STLV-1 and HTLV-1</td>
<td>C</td>
<td>125.9 (69.8-196.9)</td>
<td>10.4 (4.6-20.1)</td>
</tr>
<tr>
<td>HTLV-1</td>
<td>D</td>
<td>54.7 (45.3-73.5)</td>
<td>4.6 (3-8.2)</td>
</tr>
<tr>
<td>Melanesia HTLV-1</td>
<td>E</td>
<td>14.4 (4.7-32.3)</td>
<td>1.2 (0.3-3.1)</td>
</tr>
<tr>
<td>Australia HTLV-1</td>
<td>F</td>
<td>20.8 (9.5-34.2)</td>
<td>1.7 (0.7-3.5)</td>
</tr>
<tr>
<td>Australia/PNG HTLV-1 split(^4)</td>
<td>G</td>
<td>*49.3 (45-54.2)</td>
<td>*3.9 (3-7)</td>
</tr>
<tr>
<td>\textit{Pongo/Macaca} STLV-1</td>
<td>H</td>
<td>207.5 (107.5-343.9)</td>
<td>17.2 (7-34.6)</td>
</tr>
<tr>
<td>\textit{Pongo} STLV-1</td>
<td>I</td>
<td>37.5 (12.3-75.4)</td>
<td>3.2 (1-7.2)</td>
</tr>
<tr>
<td>Nucleotide substitutions/site/year</td>
<td>5.18E-07 (2.81E-7, 7.98E-07)</td>
<td>6.22E-06 (2.43E-6, 1.05E-5)</td>
<td>2.83E-06 (1.83E-06, 3.68E-6)</td>
</tr>
<tr>
<td>All ESS(^5)</td>
<td>&gt;423</td>
<td>&gt;1950</td>
<td>&gt;4675</td>
</tr>
</tbody>
</table>

1. Clock calibrations, TMRCA and 95% highest probability densities (HPD, parentheses) estimates are presented in thousands of years (kya).
2. Node IDs are shown on the trees in Figures 3.2-3.5.
3. LTR nucleotide substitution rate inferred from pedigree familial HTLV-1 transmission (Van Dooren et al. 2004).
4. * indicates node was used for date calibrations.
5. ESS, effective sample size of all run parameters.
Table 3.3 Middle/Late Pleistocene and Early Holocene fossil localities with well-dated fossil evidence of orangutans (*Pongo* sp.)

<table>
<thead>
<tr>
<th>Locality</th>
<th>Dates</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shuangtan, China</td>
<td>400-320 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td><em>Gigantopithecus</em> Cave, China</td>
<td>380-308 kya</td>
<td>Huang, 1979; Rink et al., 2008; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Hei Cave, China</td>
<td>380-300 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Da Cave, China</td>
<td>300-120 kya</td>
<td>Jones et al., 2004; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Wuyun Cave, China</td>
<td>279-76 kya</td>
<td>Rink et al., 2008; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Huiyan Cave, China</td>
<td>240-206 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Guanyin Cave, China</td>
<td>240-40 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Shiziyan, China</td>
<td>~237 kya</td>
<td>Bae 2010; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Liujiang Fossil Locality, China</td>
<td>153-68 kya</td>
<td>Huang, 1979; Shen et al., 2004; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Huangyan Cave, China</td>
<td>145-127 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Ganqianyan Tubo, China</td>
<td>139-22 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Mulantun, China</td>
<td>110 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Nongbashankou Cave, China</td>
<td>~110 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Baxian Cave, China</td>
<td>~110 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Xiaren Cave, China</td>
<td>105-47 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Yixiantian, China</td>
<td>100 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Gongishan Cave, China</td>
<td>~100 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Loushan Cave, China</td>
<td>Holocene</td>
<td>Takai et al., 2014</td>
</tr>
<tr>
<td>Lang Trang Caves, Vietnam</td>
<td>500-146 kya</td>
<td>Olsen and Ciochon, 1990; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Thanm Om Cave, Vietnam</td>
<td>250-140 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Hang Hum Cave, Vietnam</td>
<td>140-80 kya</td>
<td>Olsen and Ciochon, 1990; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Duoi U’oi Cave, Vietnam</td>
<td>66 kya</td>
<td>Bacon et al., 2008; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Keo Leng, Vietnam</td>
<td>30-20 kya</td>
<td>Olsen and Ciochon, 1990; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Nguom Rock Shelter, Vietnam</td>
<td>~20 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Tam Hang, Laos</td>
<td>&gt;169 kya</td>
<td>Bacon et al., 2011; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Thum Wiman Nakin, Thailand</td>
<td>&gt;169 kya</td>
<td>Chaimanee and Jaegar, 1993; Esposito et al., 2002; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Ban Fa Sui I (Cave of the Monk), Thailand</td>
<td>32-19 kya</td>
<td>Zeitoun et al., 2010; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Badak Cave C, Peninsular Malaysia</td>
<td>500-130 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Site</td>
<td>Age</td>
<td>References</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Punung, Java, Indonesia</td>
<td>128-115 kya</td>
<td>Westaway et al., 2007; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Padang Highlands, Sumatra, Indonesia</td>
<td>128-118 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Lida Ajer Cave, Sumatra, Indonesia</td>
<td>80 kya</td>
<td>Drawhorn, 1995; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Djamboe Cave, Sumatra, Indonesia</td>
<td>70-60 kya</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Niah Cave, Sarawak (Borneo)</td>
<td>45 kya</td>
<td>Harrison, 1996; Barker et al., 2007; Barker et al., 2009; Barker et al., 2009; Piper and Rabett, 2009; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Bau Caves, Sarawak (Borneo)</td>
<td>Late Pleist/Early Holocene</td>
<td>Cranbrook, 1986; Harrison, 2000; Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Madai Caves, Sabah (Borneo)</td>
<td>Holocene</td>
<td>Ibrahim et al., 2013</td>
</tr>
<tr>
<td>Gua Sireh, Sarawak (Borneo)</td>
<td>Holocene</td>
<td>Ibrahim et al., 2013</td>
</tr>
</tbody>
</table>
3.8 Figure Legends

**Figure 3.1.** Maximum likelihood (ML) tree inferred using primate T-lymphotropic virus type 1 (PTLV-1) *tax* sequences (246 bp). The ML tree was created using the TrN+G model. Bootstrap support values greater than 60 are shown at major nodes in the tree.

**Figure 3.2.** Primate T-lymphotropic virus type 1 (PTLV-1) phylogeny inferred using maximum likelihood (ML) analysis of the 500-bp LTR sequences. ML tree was created using the HKY+G nucleotide substitution model. Bootstrap support values for nodes with values greater than 60 are indicated.

**Figure 3.3.** Primate T-lymphotropic virus type 1 (PTLV-1) phylogeny inferred by Bayesian analysis of 500-bp LTR sequences using our Modified Classical Archaeological Calibration Model (MCA) by setting the PNG-1/Australia HTLV-1 clade (Node G) at 45-55 kya with a uniform distribution. Nodes with Bayesian posterior probabilities greater than 0.60 are indicated. * indicates Node that was used for date calibrations. Capital letters at some nodes indicate divergence dates were inferred at that split and are provided in Table 2.3.

**Figure 3.4.** Primate T-lymphotropic virus type 1 (PTLV-1) phylogeny inferred by Bayesian analysis of 500-bp LTR sequences using our Most Likely Archaeological Calibration Model (MLA) calibrated by setting the PNG-1/Australia HTLV-1 clade (Node G) at 3-8 kya with a uniform distribution. Nodes with Bayesian posterior probabilities greater than 0.60 are indicated. * indicates Node that was used for date calibrations. Capital letters at some nodes indicate divergence dates were inferred at that split and are provided in Table 3.2.

**Figure 3.5.** Primate T-lymphotropic virus type 1 phylogeny (PTLV-1) inferred by Bayesian analysis of 500-bp LTR sequences using our Pedigree Calibration Model (PCM) calibrated using the pedigree substitution rates from Van Dooren et al., (2004) $1.77 \times 10^{-6}$ to $3.65 \times 10^{-6}$ with a uniform distribution. Nodes with Bayesian posterior probabilities greater than 0.60 are indicated. Capital letters at some nodes indicate divergence dates were inferred at that split and are provided in Table 3.2.
3.9 Figures

**Figure 3.1** Maximum likelihood PTLV-1 *tax* tree
Figure 3.2 Maximum likelihood PTLV-1 LTR tree
Figure 3.3 Modified Classical Archaeological (MCA Model) Calibration PTLV-1 LTR tree
Figure 3.4 Most Likely Archaeological Calibration (MLA Model) PTLV-1 LTR tree
Figure 3.5 Pedigree (PCM Model) Calibration PTLV-1 LTR tree
3.10 References

Andonov, A., Coulthart, M.B., Pérez-Losada, M., Crandall, K.A., Posada, D., Padmore, R.,
cell lymphotrophic virus type 1 based on new strains from aboriginal people of Canada.

Ayouba, A., Duval, L., Liégeois, F., Ngin, S., Ahuka-Mundeke, S., Switzer, W.M., Delaporte, E.,
Ariey, F., Peeters, M., Nerrienet, E., 2013. Nonhuman primate retroviruses from Cambodia:
high simian foamy virus prevalence, identification of divergent STLV-1 strains and no

Bacon, A.-M., Demeter, F., Duringer, P., Helm, C., Bano, M., Kim Thuy, N.T., Antoine, P.-O.,
Thi Mai, B., Huong, N.T.M., Dodo, Y., Chabaux, F., Rihs, S., 2008. The late Pleistocene
Duoi U’Oi cave in northern Vietnam: palaeontology, sedimentology, taphonomy and

Bacon, A.-M., Duringer, P., Antoine, P.-O., Demeter, F., Shackelford, L., Sayavongkhamdy, T.,
Sichanthongtip, P., Khamdalavong, P., Nokhamaomphu, S., Sysuphanh, V., Patole-
Edoumba, E., Chabaux, F., Pelt, E., 2011. The middle Pleistocene mammalian fauna from
Tam Hang karstic deposit, northern Laos: new data and evolutionary hypothesis. Quat. Int.
245, 315–32.

Bae, C.J., 2010. The late middle Pleistocene hominin fossil record of eastern Asia: synthesis and

Barker, G., Barton, H., Bird, M., Daly, P., Datan, I., Dykes, A., Farr, L., Gilbertson, D.,
Harrisson, B., Hunt, C., Higham, T., Kealhofer, L., Krigbaum, J., Lewis, H., McLaren, S.,
Paz, V., Pike, A., Piper, P., Pyatt, B., Rabett, R., Reynolds, T., Rose, J., Rushworth, G.,
lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans

Barker, G., Piper, P.J., Rabett, R.J., 2009. Zooarchaeology at the Niah Caves, Sarawak: context


Moens, B., López, G., Adaui, V., González, E., Kerremans, L., Clark, D., Verdonck, K.,
Gotuzzo, E., Vanham, G., Cassar, O., Gessain, A., Vandamme, A.-M., Van Dooren, S.,
2009. Development and validation of a multiplex real-time PCR assay for simultaneous
genotyping and human T-lymphotropic virus type 1, 2, and 3 proviral load determination. J.

Nerrienet, E., 2004. Simian T cell leukaemia virus type I subtype B in a wild-caught gorilla
(Gorilla gorilla gorilla) and chimpanzee (Pan troglodytes vellerosus) from Cameroon. J.

epidemiology of simian T-lymphotropic virus (STLV) in wild-caught monkeys and apes
from Cameroon: a new STLV-1, related to human T-lymphotropic virus subtype F, in a


The Oxford handbook of the archaeology and anthropology of hunter-gatherers.1-12.


islanders. Sci. Prog. 84, 157–81.

Oskarsson, M.C.R., Klütsch, C.F.C., Boonyaprakob, U., Wilton, A., Tanabe, Y., Savolainen, P.,
2012. Mitochondrial DNA data indicate an introduction through mainland Southeast Asia

Pacheco, M.A., Reid, M.J.C., Schillaci, M.A., Lowenberger, C.A., Galdikas, B.M.F., Jones-
7, e34990.


Chapter 4: Plasmodia Evolution in Orangutans

Author Contributions: Conceived and designed the study: Michael J.C. Reid, Michael A. Schillaci. Performed the experiments: Michael J.C. Reid, William M. Switzer, Michael A. Schillaci. Analyzed the data: Michael J.C. Reid, Michael A. Schillaci, William M. Switzer, Carl A. Lowenberger. Contributed reagents/materials/analysis tools: William M. Switzer, Carl A. Lowenberger. Wrote the paper: Michael J.C. Reid, Michael A. Schillaci.
4.1 Abstract

During the past 15 years, researchers have shown a renewed interest in the study of the \textit{Plasmodium} parasites infecting orangutans. Advances in molecular genetics have allowed us to isolate these parasites from orangutan blood samples, and conduct phylogenetic studies. Most recently, work has focused on determining phylogenetic relationships, and dating the origin of these parasites in orangutans using complete mtDNA genomes. Questions regarding the dating of these parasites, however, remain. In the present study we provide a new calibration model for dating the origins of \textit{Plasmodium} parasites in orangutans using a modified date range for the origin of macaques in Asia. Our phylogenetic results indicate that there are two clades of \textit{Plasmodium} parasites in orangutans (which we now refer to as \textit{Pongo} 1, and \textit{Pongo} 2), and that these clades likely represent previously identified species \textit{P. pitheci} and \textit{P. silvaticum}. However, we can not identify which clade is representative of the morphologically described species. The most recent common ancestor of both \textit{Pongo} sp. plasmodia, \textit{P. hylobati} and \textit{P. inui} dates to 2.27 mya (95\% HPD 1.3-3.3 mya). The most recent common ancestor of the second orangutan plasmodia clade (\textit{Pongo} 2) with \textit{P. hylobati} and \textit{P. inui}, dates to 2.01 mya (95\% HPD 1.14-2.93mya). The divergence of various subclades within the \textit{Pongo} 2 \textit{Plasmodium} clade all occurred between 0.07 and 0.83 mya. These estimates for the divergence of orangutan malaria parasites, and subsequent diversification, are all considerably later than previous estimates.
4.2 Introduction

The emergence of modern molecular genetic techniques has generated renewed interest in the evolutionary history of the malaria causing parasites of non-human primates. For example, recent molecular genetic research has led to the discovery that the macaque (Genus *Macaca* sp.) plasmodia, *Plasmodium knowlesi* is an emerging source of human malaria infections, causing significant morbidity and mortality in Southeast Asia (Cox-Singh et al., 2008; Divis et al., 2015; White, 2008). Genetic studies have also allowed researchers to identify a number of new *Plasmodium* parasites in great apes, including *P. gaboni* (Ollomo et al., 2009), *P. billcollinsi*, and *P. billbrayi* (Kaiser et al., 2010; Krief et al., 2010), as well as unnamed gorilla strains such as *Plasmodium GorA*, and *Plasmodium GorB* (Prugnolle et al., 2010). These newly described African ape parasites are all closely related to the most pathogenic human plasmodia *P. falciparum*. The discovery of new plasmodia species and the identification of the evolutionary relationships among these species has led to debate regarding the origins and evolution of *P. falciparum* in humans. Some believe that *P. falciparum* arose in humans via a host switch from the chimpanzee parasite *P. reichenowi* (Duval et al., 2010, Hughes and Verra, 2010; Prugnolle et al., 2010), while other research has suggested that *P. falciparum* is the result of the transfer of a *P. falciparum* orthologue from gorillas to humans (Kaiser et al., 2010; Liu et al., 2010; Prugnolle et al., 2011), or bonobos (Krief et al., 2010). Despite the growing body of research on the origins and evolution of African ape plasmodia, Asian ape plasmodia by comparison remains relatively unknown.

4.2.1 History of research on Orangutan plasmodia

Although it has been more than a century since *Plasmodium pitheci* from a Bornean orangutan (*Pongo pygmaeus*) was described by Halberstaedter and von Prowazek (1907), we still
know comparatively little about plasmodia in orangutans. Soon after the initial identification of
*P. pitheci* by Halberstaedt and von Prowazek (1907), a second report on *P. pitheci* presented by
G. Shibayama at the meetings of the Far Eastern Association of Tropical Medicine in 1910,
identified a *Plasmodium pitheci*-like parasite in an orangutan (Shibayama, 1910). Unlike
Halberstaedter and von Prowazek’s (1907) description of the parasite, Shibayama’s *P. pitheci*
seemed different in that it lacked "stippling of the red blood corpuscles" (Shibayama 1910:189).
Dodd (1913), after a necropsy identified a heavy infection of *Plasmodium pitheci* (called
*Haemoproteus pitheci*) in a newly acquired orangutan at the Sydney Zoological Garden who had
died, as a result of the malaria infection. Dodd’s (1913) findings are important to this day
because they are among only a few describing mortality caused by a *Plasmodium* infection in
orangutans. Currently, rates of morbidity and mortality caused by *Plasmodium* infection in
orangutans are unknown.

*Plasmodium pitheci* is not mentioned again in the scientific literature until 1926 when
C.M. Wenyon briefly discusses *P. pitheci* in his book *Protozoology II*. While not providing a
detailed discussion of *P. pitheci*, Wenyon (1926) does mention some interesting historical facts
about this parasite. According to Wenyon (1926), the first scientist to observe *P. pitheci* was A.
Laveran, rather than Halberstaeder and von Prowazek. Wenyon (1926) credits Laveran with the
original description of *P. pitheci* in 1905 (see Laveran, 1905). Interestingly, Wenyon (1926) also
indicates that E. Reichenow believed that *P. pitheci* was primarily a human malaria parasite (see
Reichenow, 1920). Particularly important to the research we present here, Wenyon (1926)
asserted that *P. pitheci* and *P. inui* closely resemble each other morphologically. Finally,
Wenyon (1926) provides a possible explanation as to why Shibayama's samples lacked the
"stippling of the red blood corpuscles" (or Schüffner's dots), described by Halberstaedter and von
Prowazek, (1907). Wenyon (1926) suggests that Shibayama (1910) had not properly stained the
samples, thus accounting for the lack of such stippling. There were no further accounts of orangutan plasmodia until the 1960s.

In his book titled *Malaria Parasites and other Haemosporidia*, Garnham (1966) provided the first detailed English language description of the asexual life cycle of the malaria parasite, and the formation of gametocytes. Garnham (1966) also noted that the malaria parasites of orangutans had never been studied under natural conditions, and such studies were critical because orangutans were an endangered host species. 1966 also marked the first of two expeditions to Borneo to try and find orangutan malaria, which had not been seen in an orangutan since 1936, when Wenyon had prepared several *P. pitheci* positive blood smears from an orangutan who had died at the Zoological Gardens in London.

During a 1966 expedition to Malaysian Borneo, McWilson Warren collected the first samples from orangutans living in the rainforests of Borneo (Malaysian Borneo). These results published in Coatney et al. (1971) set the stage for future *in situ* orangutan malaria studies. Blood samples were collected from 18 orangutans, with 10 testing positive for *P. pitheci* (Coatney et al., 1971). This study provided the first prevalence data for naturally occurring *P. pitheci* infections and described attempts to inoculate *P. pitheci* into five other primate species including the owl monkey (*Aotus trivirgatus*), the pig-tailed macaque (*Macaca nemestrina*), the rhesus macaque (*Macaca mulatta*), a gibbon (*Hylobates* sp.) and two splenectomized chimpanzees (*Pan troglodytes*). Subsequent *P. pitheci* infection occurred only in the splenectomized chimpanzees.

The 1970s proved to be an important decade for orangutan malaria parasite research, with another expedition supported in part by the Royal Society of Tropical Medicine and Hygiene (Garnham et al., 1972, Killick-Kendrick et al., 1972, 1973; Peters, 1973). Research during the 1970s would eventually produce the most detailed description of *P. pitheci*, and the discovery of
a new orangutan plasmodia, *Plasmodium silvaticum*. In 1972, eight blood samples were collected from orphaned orangutans housed at Sepilok orangutan rehabilitation center in Sabah, Malaysia (Garnham et al., 1972). Of these eight samples, six were found to be infected with *Plasmodium* sp., three of whom were infected with both *P. pitheci* and another new, never before described species of *Plasmodium* parasite, which was subsequently named *Plasmodium silvaticum* in honour of Stanley de Silva, Sabah's Game Warden (Garnham et al., 1972). Garnham et al. (1972) were able to provide a detailed microscopic description of the blood stages of this new parasite, including trophozoites which are shaped like those in *Plasmodium vivax* infections. The authors also provided a re-description of the blood stages of *P. pitheci* and showed that both parasites were found to share a tertian periodicity, or a 48-hour blood stage lifecycle. Garnham et al. (1972) note that the original *P. pitheci* samples were lost during WWII, so there is no type specimen for *P. pitheci* for comparison with these two orangutan plasmodia. As a result, (Garnham et al., 1972) designate the smaller of the parasites as *P. pitheci*, and the larger one as *P. silvaticum*. Garnham et al. (1972) believed that *P. pitheci* and *P. silvaticum* were more closely related to the gibbon plasmodia (*P. youngi*, *P. eylesi* and *P. hylobati*) rather than the plasmodia of African apes.

In 1973, two short papers on orangutan plasmodia would foreshadow a larger, more detailed report to eventually be published in 1976. The first of these papers was a report on the Royal Society of Tropical Medicine and Hygiene 1972 expedition to Borneo written for the Society’s laboratory meeting of November 16, 1972 (Killick-Kendrick et al., 1973). That paper detailed the purpose of the expedition, which was to find *P. pitheci*, and the fortuitous discovery of the second orangutan plasmodia *P. silvaticum* (Killick-Kendrick et al., 1973). The second paper examining the zoonotic potential of all nonhuman primate plasmodia native to Sabah including *P. silvaticum* was presented at the same meeting by Wallace Peters (Peters, 1973).
Peters (1973) also discussed the need for research on the potential for zoonotic infections of *P. vivax*-like parasites such as *P. silvaticum* and the plasmodia of gibbons.

In 1976, Peters et al. published the most complete description to date of orangutan plasmodia, including detailed descriptions and pictures of lifecycle stages for both *P. pitheci* and *P. silvaticum*. The authors of that report detailed the key morphological features distinguishing *P. pitheci* and *P. silvaticum*, mainly that *P. silvaticum* was the larger, more amoeboid shaped parasite, which causes an increase in the size of infected host’s red blood cells (RBC). Both parasites showed some form of stippling, with the stippling of *P. pitheci* being very characteristic. *P. pitheci* showed prolific schizogony in peripheral blood, while *P. silvaticum* rarely showed peripheral blood schizogony. Peters et al. (1976) also attempted to identify the anopheline vector(s) of these two parasites. They noted that while these parasites are very similar in their vector lifecycle and development, there was one key difference, the appearance of sporozoites in the salivary glands of the mosquito vector. In *P. silvaticum* the sporozoites appeared a full day later than those of *P. pitheci*.

Peters et al. (1976) were also the first to classify the orangutan plasmodia using an evolutionary approach. They placed *P. silvaticum* within the *P. vivax*-*P. cynomolgy* group of parasites, noting similarities between it and *P. vivax* (humans), *P. schwetzi* (chimpanzees and gorillas), and *P. eylesi* (gibbon). They concluded that the orangutan malaria parasites are closely related to the plasmodia of gibbons. In their evolutionary scheme *P. pitheci* forms its own group (*P. pitheci* group) closely related to the gibbon parasites *P. hylobati* and *P. youngi*, while *P. silvaticum* belongs to a more ancient *P. vivax* group, being closely related to *P. eylesi* and other monkey plasmodia.
More recent research by N. Wolfe conducted in Sabah, Malaysia in 1996 and 1997, bridged traditional morphological identification with molecular genetic detection of plasmodia parasites (Wolfe et al., 2002; also see Kilbourn et al., 2003). Wolfe et al., used a *Plasmodium* sp. genus specific polymerase chain reaction (PCR) analysis to detect the 18s small subunit rRNA (SSU RNA) gene of this parasite. This PCR methodology proved to be more effective in detecting orangutan plasmodia infections than microscopy alone (Wolfe et al., 2002). According to their results, the orangutans in their study were only infected with *P. pitheci*. The authors of that paper also provided important prevalence information on *Plasmodium* infection rates in both wild and semi-captive orangutans. Of the 74 orangutans sampled, 31 were semi-captive orangutans housed at Sepilok Orangutan Rehabilitation Center (SORC), and 43 were free-ranging animals. Of the 31 SORC animals, 29 were infected with *P. pitheci*, while only 5 of the 43 free-ranging animals tested positive for this parasite. Wolfe et al. (2002) identified a number of possible explanations for this huge difference in prevalence including host stress, higher than natural host densities, increased sources of standing water and drainage ditches, the close proximity of SORC to human settlements, abnormal ecological conditions, aberrant diets, and behavioural changes such as decreased time spent in the canopy, smaller day ranges and the unique social structure provided by grouping so many orangutans together (Wolfe et al., 2002). By proving that the genetic detection of orangutan plasmodia was possible, Wolfe et al. (2002) paved the way for future molecular phylogenetic studies of these parasites in orangutans.

The first molecular phylogenetic study of orangutan plasmodia was published in 2006 by Reid and colleagues using the same target region of DNA as Wolfe et al. (2002). The Reid et al. (2006) study tested 86 orangutans housed at the Orangutan Care Center and Quarantine (OCC&Q) in Pasir Panjang, Kalimantan Tengah (Indonesian Borneo). Of the 86 individuals tested, 24 (27.9%) were found to be microscopically positive for plasmodia infections. That
study was able to get 18s small subunit rRNA gene sequences from 13 of the 24 microscopically
positive samples, and found four seemingly distinct orangutan plasmodia clades. These four
clades were identified by Reid et al. (2006) to represent a *P. cynomolgi*-like parasite, a *P. inui*-like parasite, *P. cynomolgi* and *P. vivax*. It was concluded that the *P. cynomolgi* and *P. inui*-like parasites probably represented *P. silvaticum* and *P. pitheci*. The presence of *P. cynomolgi* and *P. vivax* sequences were interpreted as potentially reflecting cross-species infections with macaque (*P. cynomolgi*) and human (*P. vivax*) parasites (Reid et al., 2006).

Singh and Divis (2009) subsequently conducted a phylogenetic reanalysis of the sequences obtained by Reid et al. (2006), and indicated that there were several problems with the original conclusions. Singh and Divis (2009) pointed out that in malaria parasites the SSU RNA genes are developmentally regulated with each parasite generally having an A-type and an S-type and they suggested that Reid et al. (2006) likely analyzed a mix of A-type and S-type sequences of one, or both, of *P. silvaticum* and *P. pitheci*. In addition, Singh and Divis (2009) demonstrate that one of the sequences used by Reid et al. (2006) was actually an A-type sequence closely related to *P. hylobati*, not *P. cynomolgi*. They subsequently rule out the infection of the OCC&Q orangutans with *P. cynomolgi* and *P. vivax*. Singh and Divis (2009) emphasize the need to analyze multiple gene targets from multiple closely related parasites to obtain a good phylogenetic picture of orangutan plasmodia.

Taking this advice, Pacheco et al. (2012) conducted a comprehensive reanalysis of the original 24 *Plasmodium* positive samples collected by Reid and colleagues, and used in their 2006 paper. Their reanalysis included multiple gene targets such as the merozoite surface protein 1 (MSP-1), the circumsporozoite protein (CSP), and the complete mitochondrial genome (mtDNA) (Pacheco et al., 2012). Of the 24 samples analyzed 15 were found positive for
Plasmodium, with 11 samples yielding full mtDNA sequences. Pacheco et al. (2012) identified five new haplotypes that were so similar it was concluded that they must represent only one species of Plasmodium. It was, however, still unknown whether these sequences represent P. pitheci, P. silvaticum, or a new previously unknown orangutan malaria parasite (Pacheco et al., 2012). The results from that analysis also provided evidence of a close evolutionary relationship between the orangutan plasmodia and P. hylobati from gibbons, and P. inui from macaques. Importantly, Pacheco et al. (2012) were the first to attempt to date the origin of orangutan plasmodia using a Bayesian approach. Their estimates indicated the orangutan plasmodia sequenced in their analysis arose 2.85-3.35 mya (95% C.I. 2.20-4.38 mya), leading to the conclusion that it is the product of a host switch from a non-ape malaria host (Pacheco et al., 2012).

More recently Muelenbein et al. (2015) returned to the site of the original orangutan plasmodia research, SORC in Sabah Malaysia, and tested another 38 wild-born ex-captive orangutans as part of a larger study of plasmodia in Malaysian primates. In that study, 60.5% of the orangutans sampled (23/38) were found to be infected with a plasmodia parasite (Muelenbein et al., 2015). The authors also identified 10 orangutans with mixed-plasmodia infections using molecular genetic techniques. Three clades (A, B, and C) of orangutan plasmodia were identified in that study, each suggested by the authors to represent a different species (Muehlenbein et al., 2015). One clade (Clade A), contained complete mtDNA sequences from the 2012 study by Pacheco et al. (2012). With a larger data set, including these new orangutan plasmodia, and those recovered from macaques, the authors estimated the age of origin for the orangutan plasmodia using a Bayesian methodology. They date the split of the clade composed of orangutan plasmodia Clade A and B from their most recent common ancestor (MRCA) with P. inui to have
occurred between 2.92-4.80 mya. Clade C diverged from a large clade composed of Clade A, Clade B, and P. inui found in macaques between 3.71 and 5.98 mya (Muehlenbein et al., 2015).

While each of these recent genetic studies have increased our knowledge regarding the temporal and phylogenetic origins of orangutan plasmodia, none have integrated their results with what is known about orangutan evolution and paleobiogeography. In this paper, we use a Bayesian approach to generate estimates for the timing of plasmodia evolution in orangutans using revised calibration points informed by the fossil record, and by molecular estimates based on host genomes. We then identify the key features of orangutan evolutionary history and paleobiogeography associated with our revised estimates for the timing of Plasmodium evolution in orangutans.

4.3 Methodology

136 comparative Plasmodium complete mtDNA sequences were downloaded from GenBank (Table 4.1) and aligned and analyzed in MEGA6 (Tamura et al., 2013). We generated a Maximum Parsimony (MP) tree with bootstrap support values based on 500 nonparametric replicates.

4.3.1 Bayesian Divergence Dating Model Tested

We conducted Bayesian analyses using BEAST v2.1.2 to simultaneously infer evolutionary histories, divergence dates, and nucleotide substitution rates (Bouckaert et al., 2014). The Bayesian analyses were performed using an uncorrelated lognormal relaxed molecular clock and the Birth-Death tree priors with 100 million Markov Chain Monte Carlo (MCMC) iterations and a 10% burn-in. We used the node representing the most recent common ancestor (MRCA) of P. fragile and all the other Asian plasmodia as our calibration point. We
selected a date range of 3.20-5.90 million years ago for this calibration because 3.20 mya represents the earliest fossil evidence of macaques in Asia, while 5.90 is the fossil-derived age for the dispersal of African macaques (\textit{M. sylvanus}) to Eurasia (Alba et al., 2014; Barry, 1987). Given that 5.90 mya represents the dispersal of macaques out of Africa, and the introduction of macaques into Asia must have occurred sometime before the age of the first fossil macaques found in Asia (i.e. 3.20 mya), the origin of \textit{Plasmodium} in Asian macaques must have occurred sometime between 5.90 and 3.20 mya. The earliest fossil evidence for macaques in Asia is \textit{M. paleoindicus} from the Siwaliks of India at 3.20 mya (Barry, 1987). This date corresponds almost identically with molecular estimates based on nuclear sequences for the MRCA of Asian macaques between 5.01 and 3.26 mya (Perelman et al., 2011). Our calibration date range also overlaps considerably with estimates based on mtDNA for the MRCA of Asian macaques at between 6.34 and 4.69 mya (Liedigk et al., 2015).

For the BEAST analyses, the complete mtDNA sequences were aligned using MAFFT v7 (multiple sequence alignment based on the fast Fourier transform) (Katoh and Standley, 2013) and then edited manually. The calibration model utilized a log normal relaxed molecular clock, a normal prior probability distribution for the calibration, and a birth-death tree prior and were run in duplicate for 100 million Markov Chain Monte Carlo (MCMC) iterations, each with a 10% burn-in. Log and tree files from each run were combined using LogCombiner v1.8.2. Convergence of the chain sampling was checked in the program Tracer v1.6 for effective sample sizes (ESS), all of which were > 2,000. Trees were saved every 10,000 generations, and the tree with the maximum product of the posterior clade probabilities (maximum clade credibility tree) was chosen from the posterior distribution of 10,001 sampled trees after burning in the first 1,000 sampled trees with the program TreeAnnotator v1.8.2. Trees were viewed in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).
4.4 Results

4.4.1 Bayesian Analyses

Our discussion of the results from the Bayesian analysis of the complete mitochondrial genome focuses on the Asian primate plasmodia (Figure 4.1). The complete Bayesian tree including all other primate plasmodia, as well as lizard, bird and rodent species is presented as a supplementary document (Figure S4.1). A second Bayesian subtree (Figure S4.2) showing all branches from the *Pongo* 1, *P. hylobati*, *Pongo* 2 and *P. inui* region of the tree is also presented as a supplementary document. The MRCA of *P. fragile* and the rest of the Asian primate plasmodia (Node A) is strongly supported with a posterior probability of 1 (PP=1). The node ancestral to all remaining Asian primate plasmodia (Node B) also showed strong support (PP=1), and represented the divergence of two major clades. The clade, defined by Node C, (PP=1) represents the MRCA of the clades representing *P. knowlesi* (Node D, PP=1) and *P. coatneyi* (Node E, PP=1). Node F is the MRCA for two major phylogenetic groupings, one of which includes *P. cynomolgi*, *P. simium*, and the human malaria parasite *P. vivax*. The second clade (Node J, PP=0.99) comprises *P. fieldi*, *P. semiovale*, *P. inui*, *P. hylobati* and *Pongo* plasmodia (*Pongo* 1 and 2) subclades. Node K defines a strongly supported clade (PP=1) which includes a paraphyletic grouping of *P. fieldi* and a recently diverging *P. semiovale* subclade (Node L, PP=1). The *Pongo* plasmodia are positioned within a strongly supported clade (Node M, PP=1) composed of *Pongo* 1, *P. hylobati*, *Pongo* 2 and *P. inui*. The *P. inui* and *Pongo* 2 subclades within the large clade were both strongly supported (PP=1). Our Bayesian analysis yielded a substitution rate of 4.68X10^{-9} (95% HPD 2.8568X10^{-9}-6.6845X10^{-9}).
4.4.2 Maximum Parsimony

Our maximum parsimony (MP) analysis yielded two equally parsimonious trees. We focus our description of the results on the Asian primate plasmodia (Figure 4.2). The complete MP tree is available as a supplementary document (Figure S4.3), along with a MP subtree (Figure S4.4) showing all branches included in the Pongo 1, P. hylobati, Pongo 2 and P. inui region of the tree. Both MP trees exhibited the exact same topology with respect to the major clades thus only one tree is provided (Figure 4.2, Figures S4.3 and S4.4). The only differences were in terms of branch positions representing individual sequences within several subclades. For example, sequences such as EU880484 from Plasmodium knowlesi (isolate LT26) and KJ569823 Plasmodium sp. Pongo pygmaeus (Rosalita36A) appear at different places within their respective clades. These position shifts were generally associated with low bootstrap support values.

A comparison of the complete MP and Bayesian trees (Figures S1 and S3) reveals one major differences. P. cynomolgi is not grouped with P. vivax as was seen in the Bayesian tree, but instead is positioned as a sister clade to the P. vivax, P. simium, P. fieldi and P. semiovale sequences. Also, P. vivax forms a clade sister to the P. fieldi and P. semiovale clade. Not surprisingly, the node representing the divergence of P. cynomolgi from P. vivax and the other sequences from that clade is not supported (Bootstrap 31%).

Importantly for this study the clades in the region of the MP and Bayesian trees encompassing the orangutan plasmodia species, along with P. hylobati and P. inui, have the exact same structure and similar pattern of clade support. This includes weak to moderate support for the position of P. hylobati in both trees (Bayesian PP=0.71, MP Bootstrap=63). The
node representing the MRCA of *P. inui* and *Pongo* 2 subclades is somewhat less well supported in the MP tree (Bootstrap=86) versus the Bayesian tree (PP=1).

4.4.3 Bayesian Dates (TMRCA)

Bayesian estimates for TMRCA of the major plasmodia clades are presented in Table 2. Of particular interest to this study, the TMRCA of the clade comprising both orangutan plasmodia (*Pongo* 1 and *Pongo* 2), along with *P. hylobati* and *P. inui* is 2.27 mya (95% HPD 1.30-3.30 mya). The TMRCA of the *P. hylobati*, *P. inui*, and *Pongo* 2 clade is 2.01 mya (95% HPD 1.14-2.93 mya). The divergence of various subclades within the *Pongo* 2 clade all occurred between 0.07 and 0.83 mya.

4.5 Discussion

4.5.1 Phylogenetic Tree Comparisons

The results of our analyses differ in several ways from recent studies on orangutan plasmodia by Muehlenbein et al. (2015), and Pacheco et al. (2012). In comparison to results presented by Pacheco et al. (2012) our phylogenetic trees based on the mtDNA variation position *P. fragile* as the sister clade to all other Asian plasmodia, rather than forming a subclade with *P. knowlesi* and *P. coatneyi*. A second major difference is the position of *P. hylobati* within the tree. In the Pacheco et al. (2012) study *P. hylobati* is the sister clade to a clade comprising both *Pongo* plasmodia groups and *P. inui*. In our current study, along with Muehlenbein et al., (2015), there are now two clades of *Pongo* sp. plasmodia, with the first of these clades (*Pongo* 1) sharing an MRCA with *P. hylobati*, *P. inui* and a second *Pongo* sp. plasmodia clade, while, the second orangutan clade (*Pongo* 2) is positioned sister to *P. inui*. There are several important differences between our Bayesian mtDNA tree and the one created by Muehlenbein et al. (2015). Perhaps the most significant of these differences is the position of the *P. fieldi/P. semiovale* and *P.*
vivax/P. cynomolgi clades. In the Muehlenbein et al. (2015) study the P. fieldi/P. semiovale clade is positioned sister to a clade comprising P. vivax and P. cynomolgi, along with Pongo 1, P. hylobati, P. inui and Pongo 2 plasmodia. In our study this structure is reversed, with the P. vivax and P. cynomolgi clade appearing as a sister clade to P. fieldi, P. semiovale, Pongo 1, P. hylobati, P. inui and Pongo 2. Another difference between our Bayesian mtDNA tree and trees produced by Muehlenbein et al. (2015), and Pacheco et al. (2012) concerns the position of P. semiovale. In the two previous studies P. semiovale forms a clade sister to P. fieldi. In our current study P. semiovale is positioned within a paraphyletic P. fieldi clade, with P. semiovale being the result of a recent evolutionary divergence from a P. fieldi ancestor. Each of these clades are strongly supported in our Baysian tree (PP=1), and MP tree (Bootstrap=99).

In addition to differences in the structure, we have a different interpretation of the phylogenetic structure of orangutan plasmodia than that presented by Muehlenbein et al. (2015). In their study, Muehlenbein et al. (2015) describe three clades of Pongo sp. Plasmodium (Clades A, B, and C). We believe, more conservatively, that there are only two monophyletic orangutan plasmodia clades (Pongo 1 and 2) that have the potential to represent distinct species of Plasmodium. Our Pongo 1 clade corresponds to their Clade C, while our Pongo 2 clade corresponds to their Clades A and B. Part of our reasoning involves the fact that given two species of orangutan plasmodia have been previously identified morphologically (P. silvaticum and P. pitheci), it seems most parsimonious that the two clades we describe based on mtDNA variation represents those previously described species, although we can not identify which of the morphologically defined species belongs to these clades. In the present study P. vivax and P. simium form a single clade. Thus we suggest that P. simium is not a phylogenetically supported taxon. This notion is supported by previous studies indicating that P. simium and P. vivax are
genetically indistinguishable (Cornejo and Escalante, 2006; Escalante et al., 2005; Lim et al., 2005).

### 4.5.2 Bayesian Date Comparisons

Our Bayesian estimated divergence dates (TMRCA) differ from those presented in Muehlenbein et al. (2015) and Pacheco et al. (2012). Overall, our estimated TMRCAs are much younger than both of these previous studies, which is not surprising given the difference in calibration models used. Although our estimated date for the origins of *Plasmodium* in mammals (17.36 mya; 95% HPD 9.72-26.24 mya) is much younger than those presented by Muehlenbein et al. (2015) (34.68, or 38.23 mya; 95% HPD 29.89-47.58 mya), and Pacheco et al. (2012) (34.41, or 40.38 mya; 95% HPD 30.02-48.54 mya), ours falls close to the estimate suggested by Ricklefs and Outlaw (2010), who suggest this divergence occurred at 12.82 mya (95% HPD 9.93-19.49 mya).

In both the previous studies attempting to date the origin of orangutan plasmodia (Muehlenbein et al., 2015, Pacheco et al., 2012), the lemur plasmodia have been used as a calibration point in the Bayesian models, and given a minimum date of 20 mya. We did not use the origins of lemur plasmodia as a calibration in our model for two reasons. First, under this model it is assumed that colonization of land mammals to Madagascar ended 15-20 mya (Muehlenbein et al., 2015; Pacheco et al., 2011, 2012). Yet evidence suggests that hippopotami (*H. amphibious*) may have arrived on the island as recently as 2.50 mya (Dewar and Richard, 2012; Stuenes, 1989). It is possible that a malaria parasite could have been introduced to Madagascar with the hippopotamus. Second, as originally discussed by Bensch et al. (2013), mammalian sources may not be needed to introduce plasmodia to Madagascar. Given that plasmodia are vector borne parasites, it is possible that *Plasmodium* was brought to Madagascar
by the vector itself. As such, in our study where the lemur plasmodia were not used as a
calibration point, our date for the MRCA of the lemur plasmodia and *P. ovale* is 8.70 my
younger than the 20 mya calibration date proposed by Muehlenbein et al. (2015) and Pacheco et
al. (2012).

Dating the origin of the clade containing the most pathogenic human plasmodia is of
interest to many researchers. In our study, the radiation of the African ape plasmodia dates to
8.23 mya (95% HPD 3.59-13.43 mya), while the origin date for the clade comprising *P.
reichenowi* and *P. falciparum* dates to 2.60 mya (95% HPD 0.53-4.53 mya). Our date for the *P.
falciparum*/*P. reichenowi* split mentioned above is almost exactly the same as that estimated by
Ricklefs and Outlaw (2010) at 2.49 mya (95% HPD 1.93-3.79 mya). Pacheco et al. (2012) date
the origin of the African ape plasmodia to 15.07, or 16.94 mya (95% HPD 11.32-22.40 mya),
and the split of *P. reichenowi* from *P. falciparum* to 5.05, or 5.72 mya (95% HPD 3.60-7.89
mya). Muehlenbein et al. (2015) date the origin of African ape plasmodia to 14.94-17.70 mya
(95% HPD 11.18-22.80 mya). It is important to note that the early dates for the origin of African
ape plasmodia suggested by Pacheco et al. (2012) and Muehlenbein et al. (2015) requires the
initial host-switch in apes to have occurred outside of Africa given the origins of the family
Hominidae are in Eurasia between 17-13 mya (Begun, 2010). Our estimates for the TMRCA of
*P. vivax* and *P. cynomolgi* (2.05 mya; 95% HPD 0.80-3.33 mya) are also younger than those
presented by both Pacheco et al. (2012) (3.25, or 3.77 mya; 95% HPD 2.17-5.27 mya), and
Muehlenbein et al. (2015) (3.68, or 4.43 mya; 95% HPD 2.5-5.96 mya). It should be noted
however, that there is some overlap between the 95% HPDs of all of these studies. Our results
indicate that the radiation of *P. vivax* began approximately 0.25 mya, which if the 95% HPD is
considered (i.e. 0.07-0.48 mya), is close to many previous estimates. Mu et al. (2005), date this
radiation to between ~0.05-0.25 mya, while Jongwutiwes et al. (2005) date it to ~0.22-0.30 mya.
These dates however are slightly older than those estimated by Escalante et al. (2005), who dated *P. vivax* to ~0.05-0.08 mya.

While the above dates are of great general interest, for this study the most important dates relate to the clade associated with the origins of *P. hylobati*, *P. inui*, *Pongo* 1, and *Pongo* 2 *Plasmodium* parasites. Our estimated TMRCA for this clade is 2.27 mya (95% HPD 1.30-3.30 mya), which is much younger than that of Muehlenbein et al. (2015) who date this node to 4.13-4.97 mya (95% HPD 3.36-6.12 mya). Although based on our estimates the *Pongo* 1 parasite diverged from the MRCA of *Pongo* 2, *P. hylobati* and *P. inui* at 2.27 mya (95% HPD 1.30-3.30 mya), it is worth noting that according to our estimates the *Pongo* 1 *Plasmodium* diversified only 0.23 mya (95% HPD 0.08-0.43 mya), which is well after its divergence from the *P. hylobati*, *P. inui* and *Pongo* 2 clade. This may be due to a lack of sampling of other primates on the island of Borneo such as gibbons, and colobine monkeys such as langurs and proboscis monkeys (Subfamily: Colobinae). We estimate the TMRCA of *P. inui* and *Pongo* 2 to be 1.69 mya (95% HPD 0.95-2.53 mya), which is more than 1 million years younger than previous studies (Muehlenbein et al., 2015; Pacheco et al., 2012). Our estimate for the origin of the *P. inui* clade is 1.20 mya (95% HPD 0.59-1.81 mya), which is also much younger than the 2.55-3.08 mya (95% HPD 2.01-3.90 mya) estimated by Muehlenbein et al. (2015). Our *Pongo* 2 clade which includes the A and B (sub)clades described by Muehlenbein et al. (2015) dates to 0.84 mya (95% HPD 0.39-1.35 mya), this is considerably younger than the 1.37 mya (0.91-2.32 mya) they propose for this clade.

We date Muehlenbein et al.’s (2015) subclade A to 0.51 mya (0.23-0.83 mya) and subclade B to 0.2 mya (0.07-0.34 mya). Interestingly, within our subclade R1 (Figure S4.2), and the subclade B described by Muehlenbein et al. (2015), there is actually a secondary group of
smaller subclades which exhibit geographic structure. One variant subclade (Node R2-Figure S4.2) dates to approximately 0.36 mya and is found in the southcentral-most region of the island of Borneo, within the Indonesian province of Kalimantan Tengah. The second subclade (Node R4-Figure S4.2) dates to approximately 0.18 mya and represents orangutans sampled in the northeastern part of the island of Borneo in the province of Sabah, Malaysia.

4.5.3 Paleobiogeographical context for the evolution of *Plasmodium* in orangutans

Identifying the paleobiogeographical contexts within which orangutan plasmodia diverged and diversified is essential to our understanding of the evolution of this parasite in the only remaining Asian great ape. According to our results, orangutan plasmodia emerged during the early Pleistocene, to a stage known as the Gelasian which dates from 2.59-1.81 mya (Cohen et al., 2013). Even if 95% HPDs are considered, the earliest possible origins for the orangutan plasmodia would date to Piacenzian era of the late Pliocene (3.60-2.59 mya) (Cohen et al., 2013). These dates are much younger than the late Miocene (Messinian stage), or early Pliocene (Zanclean stage) dates proposed by Muehlenbein et al. (2015). During the Gelasian sea levels began to drop due to glaciation, creating land bridges throughout Island Southeast Asia, eventually resulting in a large landmass called Sundaland (Bird et al., 2005; van den Bergh et al., 2001; Voris, 2000). This new landmass allowed for the movement of animals from the mainland to what would eventually become insular Southeast Asia. For example, there is no evidence of a mammalian presence on the island of Java until 2.4 mya (Bird et al., 2005; van den Bergh et al., 2001). In general, the Pleistocene paleoenvironment of Southeast Asia varied between warm/wet periods interglacial periods and cool/dry periods during glaciation events, which shrunk rainforests and created more open savannah, or mixed forest-savannah type habitats (Bird et al., 2005; Harrison et al., 2006; Heaney, 1991; Louys and Meijaard, 2010; Meijaard, 2003;
Page et al., 2004; Tougard, 2001; Wong, 2011; Wurster et al., 2010). Some researchers have suggested that these cooler, drier periods created a large savannah corridor through Sundaland beginning on the eastern slopes of the Sumatran highlands and extending to what is now western Borneo (Bird et al., 2005). This would have resulted in forest refugia in the highlands of what is now Java and Sumatra, along with northeastern and southern Borneo (Bird et al., 2005). This wide savannah corridor may have posed a barrier for the dispersal of orangutans, and the Anopheles sp. mosquito vector, thus shaping the geographic distribution of the orangutan host populations, and making the transmission of Plasmodium sp. parasites more difficult.

Understanding orangutan evolution is also key to understanding the evolution and diversification of their infectious agents. Fossil orangutans are known from sites in China, Vietnam, Java, Thailand, and Peninsular Malaysia during much of the period for the origin and diversification of the orangutan plasmodia (i.e., 3.30-0.08 mya) (Ciochon et al., 1996; Drawhorn, 1995; Huang, 1979; Ibrahim et al., 2013; Olsen and Ciochon, 1990; Pope et al., 1981; Rink et al., 2008; Wang et al., 2014). Our TMRCA for both of our orangutan plasmodia (Pongo 1 and Pongo 2), P. inui and P. hylobati at 2.27 mya corresponds closely with the oldest fossil orangutan sites which are located in China, which include the Baikong and Mohui cave sites dating to 2.20 and 2.00 mya, respectively (Harrison et al., 2014; Ibrahim et al., 2013; Takai et al., 2014; Wang et al., 2014). The diversification of this orangutan Plasmodium sp. does not occur until 0.23 mya (0.08-0.43 mya), meaning that there is approximately 2 million years of evolutionary history missing for this parasite. This could easily represent a sampling bias, as our knowledge of hylobatid and leaf monkey plasmodia are extremely limited. In addition, samples from Sumatran orangutans could also fill this gap. Given that the ancestors of modern orangutans were still living in mainland Southeast Asia during the late Pleistocene, our estimated date for the divergence of the Pongo 1 Plasmodium lineage (2.27 mya) suggests a mainland origin for
this parasite is possible. Similarly, the divergence of *Pongo* 2 and *P. inui* (1.69 mya; 95% HPD 0.95-2.53 mya) also suggests that its evolution (i.e. speciation) may also have occurred on the mainland. Both *Pongo* 1 (0.23 mya; 95% HPD 0.08-0.43 mya) and *Pongo* 2 (0.84 mya; 95% HPD 0.39-1.35 mya) diversified much later than their initial divergence events. Interestingly, these dates correspond to some of the earliest dates for the appearance of orangutans on Sundaland. These early Sunda fossil orangutan sites are on the island of Java, at a site named Sangiran, and date to 0.83-1 mya (Ibrahim et al., 2013). It is likely that orangutans would be present throughout Sundaland. It should be noted though that these dates are earlier than those reported for the earliest fossil orangutans on Sumatra, specifically from the sites of Punung, and the Padang Highland caves which date to ~0.12-0.13 mya (Ibrahim et al., 2013; Westaway et al., 2007). Evidence for fossil orangutans in Borneo is even more recent, with the earliest site on the island being Niah Cave (approximately 0.05 mya) in the Malaysian province of Sarawak (Harrison, 1996; Barker et al., 2007, 2009; Piper and Rabett, 2009; Ibrahim et al., 2013). It is important to recognize that orangutan plasmodia are currently only known from Bornean populations. Obtaining genetic sequences of plasmodia from Sumatran orangutans is critical if we are to improve our understanding of plasmodia evolution in orangutans.

Finally, further support for our divergence dates can be found in what is known about the evolutionary history of macaques. Our estimated divergence dates for orangutan and macaque (*P. inui*) plasmodia correspond with when it is believed the earliest ancient macaque species arrived in Sundaland (Abegg and Thierry, 2002). Also, our dates for the origin of the *Pongo* 1, *Pongo* 2, *P. inui* and *P. hylobati* clade (2.27 mya) corresponds to some of the earliest known dates for fossil macaques in China at 2.20 mya (Takai et al., 2014). Parenthetically, the dates for the origin of Southern Asian primate plasmodia (*P. fragile, P. knowlesi, P. coatneyi, P. vivax, P. cynomolgi, P. fieldi, P. semiovale, P. hylobati, P. inui* and the *Pongo Plasmodium* clades) at
5.69-6.66 mya (95% HPD 4.5-8.3 mya) presented by Muehlenbein et al. (2015) indicate Asian primate plasmodia would have actually not evolved in Asian macaques, as the earliest fossil evidence for macaques in Asia is 3.2 mya (Barry, 1987).

4.6 Conclusion

In this paper we report on the evolutionary relationships and history of *Plasmodium* parasites in orangutans gleaned from the results of a Bayesian analysis of complete mitochondrial DNA sequences, and an evaluation of what is known about orangutan and macaque evolution, and the paleobiogeographical contexts of that evolution. Our Bayesian analysis utilizes a calibration model that is well informed by the fossil record, and corroborated by divergence estimates derived from molecular genetic sequence data. Our results indicate that there are two clades of orangutan plasmodia (*Pongo* 1, *Pongo* 2) which we believe can be most parsimoniously explained as representing the orangutan plasmodia *P. silvaticum* and *P. pitheci*, which have been previously identified morphologically. The clade containing these two orangutan plasmodia along, with *P. inui* and *P. hylobati*, diverged from the other Asian primate plasmodia 2.27 mya. These orangutan plasmodia did not diversify in orangutans until 0.84 mya (*Pongo* 2) and 0.23 mya (*Pongo* 1), suggesting that the evolution of these parasites could have occurred on the Indochinese mainland, and not on the island of Borneo. There is still a great deal of uncertainty about the plasmodia of hylabatids, langurs, and Sumatran orangutans. Future research focusing on these primate hosts would play a significant role in helping us better understand the evolution of Asian primate plasmodia.
## 4.7 Tables

**Table 4.1** *Plasmodium* mitochondrial DNA sequences used in this study and Genbank accession numbers

<table>
<thead>
<tr>
<th>Plasmodium species</th>
<th>Genbank Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Plasmodium cynomolgi</em></td>
<td>AB434919, AB444121-AB444131, KJ569865-KJ569868</td>
</tr>
<tr>
<td><em>Plasmodium knowlesi</em></td>
<td>AB444106-AB444108, AY598141, AY722797, EU880467, EU880470, EU880475, EU880484, EU880489</td>
</tr>
<tr>
<td><em>Plasmodium fieldi</em></td>
<td>AB444132-AB444134, KJ569861-KJ569864, AB354574</td>
</tr>
<tr>
<td><em>Plasmodium sp. (P. pygmaeus)</em></td>
<td>JQ308530-JQ308534, KJ569804 KJ569833</td>
</tr>
<tr>
<td><em>Plasmodium inui</em></td>
<td>KJ569834-KJ569850, GQ355482</td>
</tr>
<tr>
<td><em>Plasmodium sp. (Macaca nemestrina)</em></td>
<td>KJ569851-KJ569854</td>
</tr>
<tr>
<td><em>Plasmodium sp. (Macaca fascicularis)</em></td>
<td>KJ569855-KJ569860</td>
</tr>
<tr>
<td><em>Plasmodium vivax</em></td>
<td>AY598098, AY598107, DQ396547- DQ396548</td>
</tr>
<tr>
<td><em>Plasmodium simium</em></td>
<td>AY800110</td>
</tr>
<tr>
<td><em>Plasmodium simiovale</em></td>
<td>AB434920, AY800109</td>
</tr>
<tr>
<td><em>Plasmodium fragile</em></td>
<td>AB444136</td>
</tr>
<tr>
<td><em>Plasmodium coatneyi</em></td>
<td>AB354575</td>
</tr>
<tr>
<td><em>Plasmodium hylobati</em></td>
<td>AB354573</td>
</tr>
<tr>
<td><em>Plasmodium sp.</em></td>
<td>AY800112</td>
</tr>
<tr>
<td><em>Plasmodium gonderi</em></td>
<td>AY800111</td>
</tr>
<tr>
<td><em>Plasmodium brasilianum</em></td>
<td>GQ355484</td>
</tr>
<tr>
<td>Species</td>
<td>Accession</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td><em>Plasmodium malariae</em></td>
<td>AB354570, AB489194</td>
</tr>
<tr>
<td><em>Plasmodium ovale wallikeri</em></td>
<td>HQ712053</td>
</tr>
<tr>
<td><em>Plasmodium ovale curtisi</em></td>
<td>HQ712052</td>
</tr>
<tr>
<td><em>Plasmodium</em> sp. A-Lemur</td>
<td>HQ712054</td>
</tr>
<tr>
<td><em>Plasmodium</em> sp. B-Lemur</td>
<td>HQ712055</td>
</tr>
<tr>
<td><em>Plasmodium</em> sp. C-Lemur</td>
<td>HQ712056</td>
</tr>
<tr>
<td><em>Plasmodium</em> sp. D-Lemur</td>
<td>HQ712057</td>
</tr>
<tr>
<td><em>Plasmodium</em> sp. E-Lemur</td>
<td>JN131536</td>
</tr>
<tr>
<td><em>Plasmodium yoelii</em></td>
<td>M29000</td>
</tr>
<tr>
<td><em>Plasmodium atheruri</em></td>
<td>HQ712051</td>
</tr>
<tr>
<td><em>Plasmodium chabaudi</em></td>
<td>AF014116</td>
</tr>
<tr>
<td><em>Plasmodium berghei</em></td>
<td>AF014115</td>
</tr>
<tr>
<td><em>Plasmodium vinckei vinckei</em></td>
<td>AB599931</td>
</tr>
<tr>
<td><em>Plasmodium reichenowi</em></td>
<td>GQ355476, NC002235</td>
</tr>
<tr>
<td><em>Plasmodium gaboni</em></td>
<td>JX893151, FJ895307</td>
</tr>
<tr>
<td><em>Plasmodium billcollinsi</em></td>
<td>GQ355479</td>
</tr>
<tr>
<td><em>Plasmodium billbrayi</em></td>
<td>GQ355468</td>
</tr>
<tr>
<td><em>Plasmodium falciparum</em></td>
<td>AY282930</td>
</tr>
<tr>
<td><em>Plasmodium relictum</em></td>
<td>AY733088</td>
</tr>
<tr>
<td>Plasmodium floridense</td>
<td>NC009961</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Plasmodium mexicanum</td>
<td>NC009960</td>
</tr>
<tr>
<td>Plasmodium gallinaceum</td>
<td>NC008279</td>
</tr>
<tr>
<td>Plasmodium juxtanucleare</td>
<td>NC008279</td>
</tr>
</tbody>
</table>
Table 4.2 Estimated divergence dates (TMRCA) for major *Plasmodium* clades

<table>
<thead>
<tr>
<th>Divergence</th>
<th>Node¹</th>
<th>MRCA³</th>
<th>95% HPD³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origins of <em>Plasmodium</em> sp. in birds, mammals, and reptiles</td>
<td></td>
<td>18.3</td>
<td>9.55-27.44</td>
</tr>
<tr>
<td>Origins of <em>Plasmodium</em> sp. in mammals</td>
<td></td>
<td>17.36</td>
<td>9.72-26.24</td>
</tr>
<tr>
<td>Radiation of African ape plasmodia</td>
<td></td>
<td>8.23</td>
<td>3.59-13.43</td>
</tr>
<tr>
<td>Origins of <em>P. falciparum/P. reichenowi</em></td>
<td></td>
<td>2.60</td>
<td>0.53-4.53</td>
</tr>
<tr>
<td>Split of rodent and primate plasmodia</td>
<td></td>
<td>15</td>
<td>8.22-22.34</td>
</tr>
<tr>
<td>Radiation of rodent plasmodia</td>
<td></td>
<td>6.97</td>
<td>2.98-11.61</td>
</tr>
<tr>
<td>MRCA of lemur plasmodia and <em>P. ovale</em></td>
<td></td>
<td>11.30</td>
<td>5.85-16.73</td>
</tr>
<tr>
<td>Radiation of lemur plasmodia</td>
<td></td>
<td>8.92</td>
<td>4.42-13.88</td>
</tr>
<tr>
<td>Origins of <em>P. ovale</em></td>
<td></td>
<td>1.72</td>
<td>0.46-3.44</td>
</tr>
<tr>
<td>MRCA <em>P. gonderi</em> and <em>P.</em> sp. from Africa</td>
<td></td>
<td>2.75</td>
<td>0.64-4.99</td>
</tr>
<tr>
<td>Split of Asian plasmodia from African plasmodia</td>
<td></td>
<td>6.21</td>
<td>3.50-8.98</td>
</tr>
<tr>
<td>Origins of Asian primate plasmodia</td>
<td>A</td>
<td>4.19</td>
<td>2.79-5.63</td>
</tr>
<tr>
<td>Split of Node C from other Asian plasmodia</td>
<td>B</td>
<td>3.57</td>
<td>2.29-4.99</td>
</tr>
<tr>
<td>Split of <em>P. coatneyi</em> and <em>P. knowlesi</em></td>
<td>C</td>
<td>2.88</td>
<td>1.67-4.20</td>
</tr>
<tr>
<td>Origins of <em>P. knowlesi</em></td>
<td>D</td>
<td>0.42</td>
<td>0.19-0.74</td>
</tr>
<tr>
<td>Origins of <em>P. coatneyi</em></td>
<td>E</td>
<td>1.99</td>
<td>0.94-3.13</td>
</tr>
<tr>
<td>MRCA <em>P. vivax/P. cynomolgi</em> and the clade containing <em>P. inui</em>, <em>P. hylobati</em>, <em>P. fieldi</em>, <em>P.</em></td>
<td>F</td>
<td>2.97</td>
<td>1.73-4.18</td>
</tr>
<tr>
<td>Event Description</td>
<td>Node</td>
<td>Time (Mya)</td>
<td>95% HPD (Mya)</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>MRCA of <em>P. vivax</em> and <em>P. cynomolgi</em></td>
<td>G</td>
<td>2.05</td>
<td>0.80-3.33</td>
</tr>
<tr>
<td>Origins of <em>P. cynomolgi</em></td>
<td>H</td>
<td>1.19</td>
<td>0.50-2.03</td>
</tr>
<tr>
<td>Origins of <em>P. vivax</em></td>
<td>I</td>
<td>0.25</td>
<td>0.07-0.48</td>
</tr>
<tr>
<td>MRCA of Nodes K and M</td>
<td>J</td>
<td>2.24</td>
<td>1.63-3.96</td>
</tr>
<tr>
<td>Diversification of <em>P. fieldi</em></td>
<td>K</td>
<td>1.53</td>
<td>0.70-2.70</td>
</tr>
<tr>
<td>Origins of <em>P. semiovale</em></td>
<td>L</td>
<td>0.01</td>
<td>0-0.06</td>
</tr>
<tr>
<td>MRCA <em>P. hylobati, P. inui, Pongo 1, and Pongo 2</em></td>
<td>M</td>
<td>2.27</td>
<td>1.30-3.30</td>
</tr>
<tr>
<td>Diversification of <em>Pongo 1</em></td>
<td>N</td>
<td>0.23</td>
<td>0.08-0.43</td>
</tr>
<tr>
<td>MRCA <em>P. hylobati, P. inui, and Pongo 2</em></td>
<td>O</td>
<td>2.01</td>
<td>1.14-2.93</td>
</tr>
<tr>
<td>MRCA <em>P. inui and Pongo 2</em></td>
<td>P</td>
<td>1.69</td>
<td>0.95-2.53</td>
</tr>
<tr>
<td>Diversification of <em>P. inui</em></td>
<td>Q</td>
<td>1.20</td>
<td>0.59-1.81</td>
</tr>
<tr>
<td>Diversification of <em>Pongo 2</em></td>
<td>R</td>
<td>0.84</td>
<td>0.39-1.35</td>
</tr>
<tr>
<td>MRCA of <em>Pongo 2 variants Sabah 1 and Central Kalimantan</em></td>
<td>R1^2</td>
<td>0.51</td>
<td>0.23-0.83</td>
</tr>
<tr>
<td>MRCA of <em>Pongo 2 variant Central Kalimantan</em></td>
<td>R2^2</td>
<td>0.36</td>
<td>0.16-0.6</td>
</tr>
<tr>
<td>MRCA of <em>Pongo 2 variant Sabah 1</em></td>
<td>R3^2</td>
<td>0.18</td>
<td>NA</td>
</tr>
<tr>
<td>MRCA of <em>Pongo 2 variant Sabah 2</em></td>
<td>R4^2</td>
<td>0.20</td>
<td>0.07-0.34</td>
</tr>
</tbody>
</table>

^1 Node letter corresponds to Bayesian mtDNA tree Figure 4.1
^2 Node letter corresponds to Bayesian mtDNA tree Figure S4.2
^3 Node ages and 95% HPDs are in millions of years
4.8 Figure Legends

**Figure 4.1.** *Plasmodium* phylogeny inferred by Bayesian analysis of 136 complete mtDNA sequences. Only the Asian *Plasmodium* species are shown. The most recent common ancestor (MRCA) of all Asian plasmodia (Node A) was set as the calibration with a normal distribution and a date range of 3.20-5.90 mya. Nodes with Bayesian posterior probabilities greater than 0.60 are indicated. Capital letters at some nodes indicate divergence dates were inferred (Table 4.2).

**Figure 4.2.** Maximum Parsimony (MP) tree inferred from a comparison of 136 sequences using the complete mtDNA genome of *Plasmodium* parasites. Only the Asian *Plasmodium* species are shown. The tree was created using the Subtree-Pruning-Regrafting (SPR) method with 500 bootstrap repetitions. Bootstrap support values greater than 60 are shown at major nodes in the tree.

**Figure S4.1.** Complete *Plasmodium* phylogeny inferred by Bayesian analysis of 136 complete mtDNA sequences. The most recent common ancestor (MRCA) of all Asian plasmodia (Node A) was set as the calibration with a normal distribution and a date range of 3.20-5.90 mya. Nodes with Bayesian posterior probabilities greater than 0.60 are indicated. Capital letters at some nodes indicate divergence dates were inferred (Table 4.2).

**Figure S4.2.** *Plasmodium* sp. phylogeny inferred by Bayesian analysis of 136 complete mtDNA sequences with a normal distribution and calibrated with a date range of 3.20-5.90 million years ago. This tree only shows the individual sequences that make up the *P. inui* and *Pongo* 2 *Plasmodium* clades. Nodes with Bayesian posterior probabilities greater than 0.60 are indicated. Capital letters at some nodes indicate divergence dates were inferred (Table 4.2).
**Figure S4.3.** Complete Maximum Parsimony (MP) tree inferred in a comparison of 136 sequences using the complete mtDNA genome of *Plasmodium* parasites. The MP tree was created using the Subtree-Pruning-Regrafting (SPR) method with 500 bootstrap repetitions. Bootstrap support values greater than 60 are shown at major nodes in the tree.

**Figure S4.4.** Maximum Parsimony (MP) tree inferred in a comparison of 136 sequences using the complete mtDNA genome of *Plasmodium* sp. parasites. The MP tree was created using the Subtree-Pruning-Regrafting (SPR) method with 500 bootstrap repetitions. This tree only shows the individual sequences that make up the *P. inui* and *Pongo 2 Plasmodium* clades. Bootstrap support values greater than 60 are shown at major nodes in the tree.
4.9 Figures

**Figure 4.1** *Plasmodium* phylogeny inferred by Bayesian analysis of 136 complete mtDNA sequences showing only the Asian *Plasmodium* species
Figure 4.2 Maximum Parsimony (MP) tree of 136 complete mtDNA sequences of the *Plasmodium* parasite showing only the Asian *Plasmodium* species.
Figure S4.1 Complete *Plasmodium* phylogeny inferred by Bayesian analysis of 136 complete mtDNA sequences
Figure S4.2 Plasmodium sp. Bayesian showing only the individual sequences that make up the P. inui and Pongo 2 clades
**Figure S4.3** Complete Maximum Parsimony (MP) tree using the complete mtDNA genome of *Plasmodium* parasites.
Figure S4.4 Maximum Parsimony (MP) tree only showing the individual sequences that make up the *P. inui* and *Pongo* 2 clades.
4.10 References


Chapter 5: Conclusions and Future Directions
5.1 Introduction

When I began this thesis my goal was to provide an anthropological perspective to the study of the evolution of primate infectious agents. Prior to this project few studies of primate infectious agent evolution involved evolutionary anthropologists. The results of this thesis examining the evolution of simian foamy virus (SFV), simian T-lymphotropic virus (STLV-1), and the malaria causing *Plasmodium* parasites in Bornean orangutans (*Pongo pygmaeus*) helps demonstrate the potential contributions evolutionary anthropologists may have on the study of the evolution of infectious agents. As trained evolutionary anthropologists, we are able to place the observed phylogenetic relationships among infectious agents in a historical evolutionary context with regard to primate hosts. Similarly, evolutionary anthropologists are able to recognize how infectious agent evolution may inform our understanding of primate host evolution. My doctoral research has made three primary contributions. The first and second major contributions are closely intertwined, with the first being the importance of providing a paleobiogeographic context that is informed by the primate fossil record for interpreting the evolutionary history of orangutan SFV, STLV-1 and *Plasmodium*. The second contribution is examining how infectious agent evolution may inform our understanding of orangutan evolution. The third contribution is the importance of establishing precise calibration points for dating the origins of these infectious agents based on a comprehensive review of the genetic, fossil, archaeological and biogeographic data.

5.2 Summary of Results and Implications

One of the more interesting aspects of looking at these three different orangutan infectious agents is that while each of them exhibit origins of different ages, there is some overlap among the origin dates of these organisms when 95% HPDs are considered, and they all
share the general phylogenetic pattern of a split from the MRCA followed by long periods before diversifying into an orangutan specific clade. SFV is the oldest (95% HPD 2.02-8.18 mya) having dates that overlap with the origins of Plasmodium in orangutans (95% HPD 1.30-3.30 mya). Similarly, the dates for the diversification of the orangutan plasmodia (Pongo 1, 80-430 kya; Pongo 2, 1.35-390 kya) exhibit overlap with our dates for the origins of STLV-1 in orangutans (7-343.9 kya).

Our Bayesian TMRCA estimations for each of the infectious agents suggest that the Pleistocene Epoch (2.588-0.0117 mya) was an important time in the evolution of these organisms. In each of the individual papers comprising this thesis, I have discussed the paleobiogeographic contexts of the dates associated with the origin of each of these infectious agents. For example, the dates for the origin of SFV suggest that the mixed species (Sumatran and Bornean orangutans) clade may have evolved on the Indochinese mainland. It should not be surprising that this time period might have been important for the spread of infectious agents in Asian primates including orangutans. We know that this was a cooler drier period in which sea levels dropped exposing the Sunda Shelf, connecting the Indochinese mainland to Sundaland (Bird et al., 2005; Harrison et al., 2006; Heaney, 1991; Louys and Meijaard, 2010; Meijaard, 2003; Page et al., 2004; Tougard, 2001; van den Bergh et al., 2001; Voris, 2000; Wong, 2011; Wurster et al., 2010). We also see a great diversity of fossil orangutan species (or suspected species), and subspecies during this time (e.g. Ciochon et al., 1996; Drawhorn, 1995; Ibrahim et al., 2013). The oldest fossil orangutan sites are in China at approximately 2.2 mya, with orangutans being present in Sundaland by 1 mya (Ibrahim et al., 2013). Orangutans are, however, still present on the Indochinese mainland as far north as China until the late Pleistocene or early Holocene (Ibrahim et al., 2013; Wang et al., 2014). The fossil and biogeographic contexts tell us that the evolutionary processes working on both orangutans and their infectious agents could
have occurred on the mainland or in Sundaland. There may also have been fairly recent gene flow from the mainland into Sundaland, but not necessarily gene flow from the direct ancestors of the two living species of orangutan species, perhaps from fossil taxa which are now extinct.

The results from the study of orangutan SFV indicate this infectious agent is likely quite old, with the mixed Sumatran/Bornean SFV clade dating to ~1.5-1.7 mya. This is ~0.5 million years before we have any evidence of orangutans in Sundaland. These dates suggest that mainland Southeast Asian orangutan species or subspecies may have experienced periods of gene flow. Or that perhaps the mixed clade is a relict viral strain from an extinct orangutan species. In either case my research on this infectious agent highlights the importance of the evolutionary history of orangutans on the Indochinese mainland prior to their introduction to Sundaland. Many studies suggest that the ancestors to *Pongo abelii* and *Pongo pygmaeus* split ~3-5 million years ago (Chattergee et al., 2009; Finstermeier et al., 2013; Ma et al., 2013; Nater et al., 2011; Pozzi et al., 2014; Raaum et al., 2005; Steiper, 2006), but few of these studies have attempted to reconcile these dates with the orangutan fossil record. My thesis has contributed to this conversation.

My research paper on orangutan plasmodia identified two potentially geographically isolated subclades within a clade we believe to represent with *P. pitheci* or *P. silvaticum*. These two subclades shared an MRCA at 0.51 mya, with the Kalimantan variant diversifying by 0.36 mya and the Sabah variant diversifying by 0.2 mya. This evidence could suggest that Bornean orangutans were present on the island at least 0.5 mya. The fact that the orangutans living in Sabah and Central Kalimantan are now classified by some as separate subspecies (Goossens et al., 2009) lends support to this hypothesis.
My doctoral research illustrates the importance of accurate calibration dates informed by archaeological and fossil data. For example, studies dating the origins of PTLV-1 (i.e. STLV-1 and HTLV-1) had used a sequence from the Solomon Islands (Mel5) as the calibration for Bayesian models with a date set to 40-60 kya. The archaeological record, however, indicates that the earliest possible evidence for people living in the Solomon Islands is about 32 kya, but that a more recent date associated with the appearance of the Lapita culture (3-3.5 kya) is far more likely (Green et al., 2008; Sheppard, 2011; Wickler, 2001). Consequently, we used a sequence of HTLV-1 from Papua New Guinea dated to 45-55 kya as the calibration for the model given that the archaeological evidence suggests that people first arrived on the island at that time (Summerhayes et al., 2010; Vilar et al., 2008). Another example can be taken from our study of orangutan Plasmodium. Previous calibrations (Muehlenbein et al., 2015; Pacheco et al., 2012) dated the divergence (node) of African and Asian primate plasmodia at 6-14 mya. There are a number of potential problems with that calibration, the first being that the host of those plasmodia have been identified as both Cercocebus and Chlorocebus depending on the source. This is problematic as the dates for the MRCA Chlorocebus and macaques is much older than the MRCA of Cercocebus and macaques. This creates a very large calibration range. The model I used dates the estimated for the dispersal of macaques from Africa based on the fossil record (5.9 mya) and the earliest fossil macaque in Asia (3.2 mya) to establish the range for our calibration model. The lemur plasmodia as a calibration, is another potential problem (Muehlenbein et al., 2015; Pacheco et al., 2011, Pacheco et al., 2012). The rational for this potentially problematic calibration is that the plasmodia of lemurs must have arrived on the island of Madagascar prior to the last known migration of land mammals at 15-20 mya. We were not the first to critique this assumption as Bensch et al. (2013) noted that mammalian sources may not have been needed for the introduction of plasmodia to Madagascar because it is a vector borne pathogen from a flying
vector that could be blown, or have flown, to the island. It is also worth noting that the fossil evidence suggests that some large mammals in the form of hippopotami may have arrived on Madagascar during the Early Pleistocene (Dewar and Richard, 2012; Stuenes, 1989).

Importantly, the Bayesian estimates for TMRCA should be supported by what is known about the evolution of host human or nonhuman primates, and the prehistory or biogeography of a region. As such a post hoc comparison of TMRCA estimates with the archaeological or fossil record should be undertaken. The STLV-1/PTLV-1 research presented in this thesis provides an example of such a post-hoc evaluation. For the pedigree calibration model (PCM), the estimated dates for the introduction of HTLV-1 into Papua New Guinea, Australia and the Solomon Islands can all be supported by genetic, and archaeological research. Our Papua dates (8.4 kya; 95% HPD 4.1-14.7 kya) are consistent with the introduction of plant cultivation to the island around 10 kya (Denham et al., 2003), while on Australia our dates are consistent with an influx of people from South Asia (4.2 kya) and the appearance of the Dingo (3.5-5.0 kya) (Oskarsson et al., 2012; Pugach et al., 2013; Savolainen et al., 2004). Finally, the estimated date for HTLV-1 from the Bayesian PCM model in the Solomon Islands (2.7 kya) is consistent with the arrival and migration of Lapita peoples between 2.6 and 3.3 kya (Green et al., 2008; Summerhayes and Scales, 2005).

5.3 Future Research

The need for additional samples or genetic sequences for infectious agents and their primate hosts became very apparent during this course of this research. For STLV-1 and orangutan plasmodia we have no samples or genetic sequences of these infectious agents from Sumatran orangutans (*Pongo abelii*). This creates an enormous gap in our knowledge about the evolution of both of these infectious agents in orangutans. For all three infectious agents
examined in this thesis, additional samples from sympatric host primates are needed, most importantly hylobatids, the proboscis monkey, langurs, and even sympatric prosimians. An argument could even be made for the collection of samples for all Euarchontans living in the region in order to provide the most complete, evolutionarily accurate picture.

In addition to greater sampling efforts from sympatric primates and other Euarchontans, future research needs to incorporate the collection of samples from wild primates, including wild orangutans. Up until recently studies such as mine have required blood samples for the genetic analysis of infectious agents, but recent advances in molecular techniques have allowed researchers to isolate *Plasmodium* sp and retroviral DNA from fecal samples (Li et al., 2012; Sundararaman et al., 2013; 2016). This means that we can now collect samples non-invasively from wild primates without the need for capture and anesthesia. Primates held at rehabilitation centers are important for this type of research and have been critical to its advancement thus far, but we can never be certain the provenance or history of individuals held in these centers. This means that we have less reliability in knowing how much human contact they have had and whether their infectious agents reflect truly wild forms and are not the result of a cross-species transmission. As such not having samples from wild primates is a potential limitation to this research.

The research on orangutan plasmodia presented in this thesis highlights another potential sampling bias. The results of this study showed that the *Pongo* 2 clade exhibited two sub clades that corresponded to the geographic location sampling. One subclade consisted of all samples collected in the southern-most region of Borneo (the Indonesian Province of Central Kalimantan), while the other consists of samples all collected in the northeastern most region of the island (the Malaysian Province of Sabah). To date there are no plasmodia samples from
orangutans, or any other primates living on the island of Borneo outside of these regions. STLV-1 has traditionally been seen as an infectious agent that clusters based on geography (Sintasath et al., 2009; Vandamme et al., 1998; Van Dooren et al., 2001), but recent evidence from wild chimpanzees suggests they also group based on host phylogeny (Junglen et al., 2010). Expanding the geographic coverage for sample collection could ultimately test this hypothesis.

5.4 Conclusions

This thesis combines the study of molecular evolution in infectious agents with anthropological data to help clearly define calibration points, as well as to aid in the interpretation of host and infectious agent evolution. Orangutans were chosen as the host species to study for this research because their infectious agents are less well known, yet their fossil history from 2.2 mya until the Holocene is better known than that of African apes. While all three of our infectious agents dated to different times, each underwent significant evolution during the Pleistocene. It should not be surprising that this would be an important time period for the evolution of hosts and their pathogens given the changing climate and lowered sea levels. This thesis is but a beginning. A great deal of research remains to be completed if we are to obtain a more complete picture of the evolution of orangutan infectious agents.
5.5 References


