Bayesian Analysis (2012) **7**, Number 4, pp. 917–974

A Bayesian Method for Estimating Evolutionary History

Joungyoun Kim *, Nicola M. Anthony † and Bret R. Larget ‡

Abstract. Phylogeography is the study of evolutionary history among populations in a species associated with geographic genetic variation. This paper examines the phylogeography of three African gorilla subspecies based on two types of DNA sequence data. One type is HV1, the first hyper-variable region in the control region of the mitochondrial genome. The other type is nuclear mitochondrial DNA (Numt DNA), which results from the introgression of a copy of HV1 from the mitochondrial genome into the nuclear genome. Numt and HV1 sequences evolve independently when in different organelles, but they share a common evolutionary history at the same locus in the mitochondrial genome prior to introgression. This study estimates the evolutionary history of gorilla populations in terms of population divergence times and effective population sizes. Also, this study estimates the number of introgression events. The estimates are obtained in a Bayesian framework using novel Markov chain Monte Carlo methods. The method is based on a hybrid coalescent process that combines separate coalescent processes for HV1 and Numt sequences along with a transfer model for introgression events within a single population tree. This Bayesian method for the analysis of Numt and HV1 sequences is the first approach specifically designed to model the evolutionary history of homologous multi-locus sequences within a population tree framework. The data analysis reveals highly discordant estimates of the divergence time between eastern and western gorilla populations for HV1 and Numt sequences. The discordant east-west split times are evidence of male-mediated gene flow between east and west long after female gorillas stopped this migration. In addition, the analysis estimates multiple independent introgression events.

Keywords: divergence time, Mitochondrial sequence, HV1, Numt, Introgression, Homologous sequences, Coalescence, Phylogeography, Population genetics, Phylogeny

1 Introduction

A central question of evolutionary biology since the time of Charles Darwin is the origin of species. More generally, biologists are interested in the forces that shape evolutionary diversity. Population genetics and phylogenetics are two major studies for a better understanding of the evolutionary history of the variety species. Population genetics is the study of the gene frequency distribution in populations and its change under the influence of the four evolutionary forces: natural selection, genetic drift, mutation, and gene flow (Hartl 2000). Phylogenetics is the study of evolutionary relatedness among vari-

*Cancer Research Institute, Seoul National University, Republic of Korea, joungyoun@gmail.com
†Department of Biological Sciences, University of New Orleans, LO, nanthony@uno.edu
‡Department of Statistics, University of Wisconsin - Madison, Madison, WI, larget@stat.wisc.edu

 DOI:10.1214/12-BA732

ous groups of organisms. The estimate of main interest in phylogenetics is a graphical tree illustrating the ancestor-descendant relationship among groups. Phylogenetic inference methods can be applied to any polymorphic information, for example, nucleotide DNA sequences, amino acid sequences and genetic or morphological markers. As an integration of population genetics and phylogenetics, phylogeography aims to infer the evolutionary past of contemporary populations, by estimating population divergence times, population sizes, migration rates, and other aspects of historical populations on the basis of molecular sequence data sampled from modern-day populations.

In this paper, we describe a model for the phylogeography of African gorilla populations based on the molecular evolution of mitochondrial DNA (mtDNA) sequence data. Gorillas are restricted to tropical forest habitats and are not observed to cross open savannah. As a consequence, historical changes in suitable habitat due to expansion and retraction of forested areas during glacial cycles may have had a profound effect on gorilla diversification and their current geographic distribution. The Pleistocene refuge hypothesis (Haffer 1969; Anthony et al. 2007) states that during times of extensive glaciation when large amounts of the planet's water are trapped as ice, the climate in tropical areas such as central Africa would have been much more arid than at present, causing forests to fragment into isolated refugia. Forest dwelling species such as gorillas would have become genetically isolated and undergone diversification over time. The Pleistocene Epoch was a period of repeated glaciation dating from about 1.8 million to 11,000 years before present. It is, therefore, of interest to see if the modern biogeographical distribution and genetic diversity of gorillas is consistent with the Pleistocene refuge hypothesis. It is of particular interest to examine the question of whether or not the time of divergence between eastern and western gorilla populations occurred during the Pleistocene.

Jensen-Seaman and Kidd (2001) and Clifford et al. (2004) studied the phylogeography of gorillas based on mtDNA. The latter authors estimated the east-mountain split time as 0.316–0.443 million years ago (MYA) and even pointed out a bottleneck in eastern gorillas and mountain gorillas around 0.25 MYA and 0.22 MYA, respectively. Additional studies based on non-coding nuclear DNA revealed stronger geographic structure in mtDNA than nuclear DNA, implying male-mediated gene flow after an initial split between eastern and western gorilla populations (Jensen-Seaman et al. 2001; Thalmann et al. 2007). Thalmann et al. (2007) estimated an initial east-west split time around at 0.9–1.6 MYA and subsequent male-mediated gene flow until 0.08–0.2 MYA. Furthermore, they estimated that the effective population sizes for all gorillas, western gorillas, and eastern gorillas as 26,600, 24,100, and 13,600, respectively.

All the estimates of population divergence times mentioned above are obtained from phylogenetic inference with multiple sequences from each population, without considering population structures on sequence data. Therefore, the estimates of the east-west split time are gene divergence times rather than population divergence times. Also, in the studies above, population sizes are estimated by converting nucleotide diversity (Nei and Li 1979) to population size. Nucleotide diversity, π, measures the degree of genetic polymorphism within a population. It is defined as the proportion of nucleotide differences per site between any two DNA sequences chosen randomly from the sam-

ple population. The conversion from π into population size N is by the relationship $\pi = 4N\mu$ (Tajima 1983), where μ is the mutation rate per generation. The rate μ must be assumed to have some known value. Since mutation rates are not known, phylogeographic inference methods can be improved by using new models which explicitly describe population structures of sampled sequences and infer population divergence times and population sizes jointly based on DNA sequence data.

Markov chain Monte Carlo (MCMC)-based Bayesian inference allows one to jointly estimate population sizes and population divergence times. The Bayesian estimates are based on samples from the posterior distribution which is the combination, using Bayes' Theorem, of the prior distribution for the population structure including divergence times and population sizes, the gene genealogy under the coalescent model, and the parameters from nucleotide substitution, and the likelihood of the data. The software BEAST (Drummond et al. 2002; Drummond and Rambaut 2007) is widely used for MCMC analysis of molecular sequences from a population. BEAST does not consider multiple populations; however, it can provide estimates of time to the most recent common ancestor (tMRCA) for each user defined taxa group. Also, BEAST implements a variety of nucleotide substitution models, for example, combining multiple genes, partitioning data into three codon positions, various site-heterogeneity models, various models for the changes in the population size over time, and so on. Rannala and Yang (2003) estimate divergence times and effective population sizes for a four-taxon tree of great apes. Their more recent work (Yang and Rannala 2006; Rannala and Yang 2007) accommodates larger trees, soft constraints induced by fossils, and relaxations of the molecular clock. Nielsen and Wakeley (2001) and Hey and Nielsen (2007) developed the Isolation with Migration (IM) model which accounts for both equilibrium isolation and migration. Their approach estimates divergence times between two populations, effective population sizes of the two descendant populations and their ancestor population, as well as migration rates, but their method can not easily be extended to larger trees. All of the above models can work with multi-gene data provided that each gene is assumed to be independent with a separate gene-genealogy in the common population tree.

The model we develop in this paper advances related work by others for reconstructing past population dynamics by accounting for a special feature of gorilla mitochondrial DNA (mtDNA) sequences. Efforts to sample mtDNA sequences from gorillas frequently produce nuclear sequences that are homologous to mtDNA from past introgressions of mitochondrial sequences into the nuclear genome. These nuclear sequences of mitochondrial origin are known as Numts (Nuclear mitochondrial DNA, pronounced new might). An important contribution of this study is the development of models that describe the homologies (relatedness by descent from a common ancestor) among Numt and mtDNA sequences by accounting for these ancient introgressions and the use of these models to infer the phylogeny of populations on the basis of homologous mtDNA and Numt sequences. Furthermore, the new models described in this paper are practical for more complex trees than those allowed by other previous methods. The computational foundation necessary for the implementation of these new models is a class of new Bayesian Markov chain Monte Carlo (MCMC) proposal algorithms designed specifically to match

characteristics of samples of mtDNA and homologous Numt sequences.

In section 2, we describe the background of gorilla populations and the special character of our data set. Section 3 illustrates detailed descriptions of the models and parameters. Also the details of simulation study is explained. The results are summarized in section 4, and then discussion follows in section 5. The supplemented contents are in appendices.

2 Data

Our inference about gorilla phylogeography is based on DNA sequence data. As mentioned before, we have two types of DNA sequence data available for this study; HV1, the first hyper-variable region of the control region of the mitochondrial genome, and Numt sequences homologous to HV1 from one or more past introgressions. More details about HV1 and Numt sequences, and the relationship between these two kinds of sequences follow.

2.1 HV1

HV1 is a widely-used genetic marker for studies of populations of great apes because the rate of nucleotide substitution is fairly high, providing information about relatively recent events on an evolutionary time scale. Another advantage to using HV1 is that, as a part of mtDNA, it is believed to represent a single non-recombining locus, so that genealogical trees depicting relationships among variants may be constructed and traced back to an ancestral type. In addition, HV1 is expected to evolve under neutral selection because it is in a non-coding region and has no known regulatory function.

2.2 Numt

Numt sequences arise from the introgression of mtDNA into the separate nuclear genome of a eukaryote organism. The few Numts that arise from HV1 will be homologous to HV1 (Lopez et al. 1994). As whole genome sequencing projects accumulate, increasingly many Numts have been detected in many diverse eukaryotic organisms (Bensasson et al. 2001; Ricchetti et al. 2004; Richly and Leister 2004).

Numts have the potential to improve inferences about the evolutionary history of mtNDA because they are thought to behave as a molecular fossil. Since the substitution rate in nuclear DNA is typically much slower than in mitochondrial sequences (Brown et al. 1982; Zischler et al. 1995), Numt sequences today may be more similar to ancestral HV1 sequences than the HV1 sequences we measure today (Thalmann et al. 2005).

In addition to the substitution rate, Numt sequences have different inheritance mechanisms than HV1 due to the different inheritance mechanisms between genomes in the nucleus and the mitochondria. Almost always, mtDNA is maternally inherited as a single copy allele, known as a haploid. In contrast, the nucleus is passed to the de-

scendant with two copies, known as diploids, one from a mother and another from a father. Therefore, the numbers of alleles participating in the inheritances are different for each of Numt and HV1 sequences, even in the same population. For example, if we assume an equal ratio of numbers of males and females in a population, the total number of Numt alleles available for inheritance is four times larger than the number of HV1 sequences.

For these reasons, special care is needed in an analysis using both Numts and HV1. The model in this paper takes account of the separate mutation and inheritance processes between Numt and HV1 sequences as well as their homologous evolution history.

2.3 DNA data from gorilla, human and chimpanzee

This paper analyzes the 125 gorilla DNA sequences described in Anthony et al.(2006) who compiled the data from several previously published sources, removed putative polymerase chain reaction (PCR) recombinants, and classified remaining sequences as Numt or HV1. Thus, 125 gorilla DNA sequences would be provided in a supplemented file.

Some sequences are extracted from shed hair samples collected from night nests of gorillas, and several of the sequences are from zoo specimens. Samples through the entire gorilla range were collected over a period of several years through collaborations with researchers from existing study sites (Clifford et al. 2004). We follow a classification of gorillas into three subspecies (Groves 1967, 1970), western lowland gorillas (*Gorilla gorilla gorilla*), eastern lowland gorilla (*G. g. graueri*), and mountain gorillas (*G. g. beringei*), recognizing that western lowland gorillas are comprised of many natural populations with significant diversity among them (World Wildlife Foundation 2008).

We add to this data set six chimpanzee HV1 sequences and ten human HV1 sequences obtained from GenBank(Benson et al. 2011) to allow us to calibrate the gorilla divergence times with prior information about human and chimpanzee divergence. The chimpanzee sequences (Hu et al. 2001; Thalmann et al. 2004) are from three subspecies: *Pan troglodytes verus* or western common chimpanzee (AF315497, AF315499, AJ586557); *P. t. vellerosus* Nigerian chimpanzee (AF315498); and *P. t. troglodytes* or central common chimpanzee (AF315500 and AJ586556). The human sequences (AF346963, AF346972, AF346977, AF346978, AF346981, AF346983, AF346987, AF346993, AF347009, and AF347015) are a subset of 53 sequences sampled from humans of diverse origins (Ingman et al. 2000).

We aligned the sequences using Clustal W (Thompson et al. 1994) and, following Anthony et al. (2006), removed a 26-base pair (bp) portion of the alignment containing the poly-cytosine (poly C) motif which contains many gaps useful for distinguishing Numt and HV1 sequences, but is difficult to align with confidence. The final alignment, which consists of 236 sites for 141 sequences partitioned into seven species or subspecies and two types, is summarized in Figure 1.

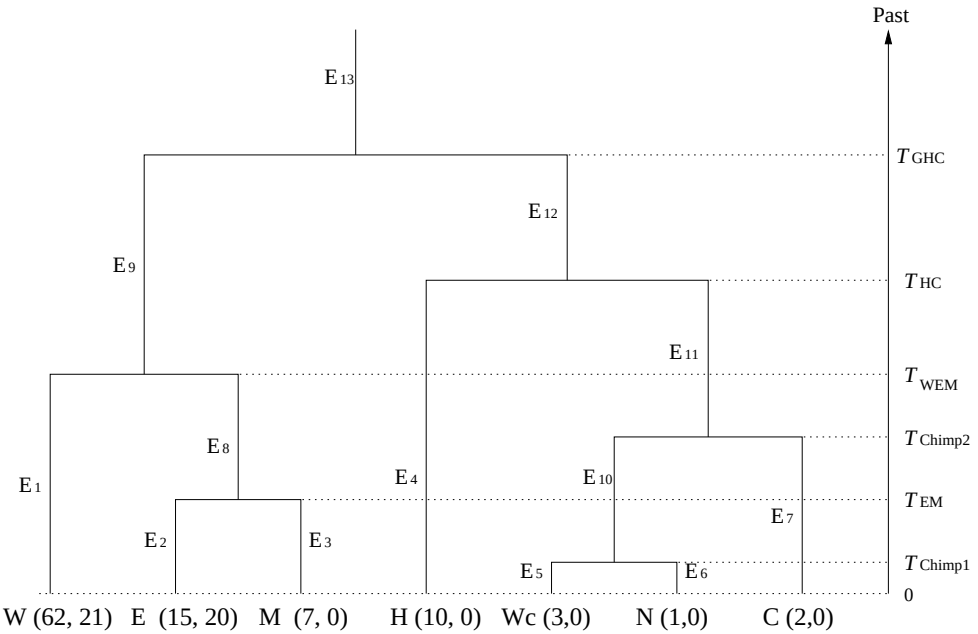

Figure 1: **The population tree for seven populations.** The displayed tree shows the assumed relationship among three gorilla populations, W, E and M, three chimpanzee populations, W_c, N and C, and one human population, H. The numbers in the parentheses are the number of HV1 sequences and the number of Numt sequences sampled from each population, respectively. E_i indicates a population edge, and T_x is the population divergence time of descendant of population x (measured in MYA).

3 Model parameters

3.1 Divergence times

The main interests in this study are the divergence times of three gorilla populations. The estimates of divergence times can be evidence of the change of forest habitat distribution during a period of glaciation and ice ages. In addition to gorillas, human and chimpanzee are included in this phylogeograpy model as populations for calibration, because human and chimpanzee have abundant literature about their divergence time.

Our Bayesian approach requires specification of a prior distribution on the divergence times in the population tree. The population tree and divergence times are shown in Figure 1. We assume that the split time between human and chimpanzee, T_{HC}, is uniformly distributed in the interval (4, 6) million years ago (MYA) (Tamura and Nei 1993; Takahata et al. 1995), and that the divergence time of gorilla from the human/chimpanzee ancestor, T_{GHC}, is also uniformly distributed in the interval (T_{HC}, 9) MYA (Vigilant et al. 1991).

Within gorilla, three gorilla populations have been recognized, western lowland gorilla (W), eastern lowland gorilla (E), and mountain gorilla (M). Western lowland gorillas are separated from eastern subspecies of eastern lowland gorilla (E) and mountain gorilla (M) by more than 850km. Eastern lowland gorilla and mountain gorilla are so close ecologically and genetically that they are called eastern gorilla as opposed to western gorilla. Therefore, we assume that western lowland gorillas diverged first at time T_{WEM} and then eastern lowland gorillas and mountain gorillas diverged more recently at time T_{EM} (i.e, $T_{\mathrm{EM}} \leq T_{\mathrm{WEM}}$).

We assign a scaled joint Beta prior distribution on the times T_{EM} and T_{WEM} given T_{GHC} with $T_{\mathrm{EM}} \leq T_{\mathrm{WEM}}$

$$\begin{aligned} T_{\mathrm{EM}} &= \min(b_1, b_2) \times T_{\mathrm{GHC}}, \\ T_{\mathrm{WEM}} &= \max(b_1, b_2) \times T_{\mathrm{GHC}}, \end{aligned}$$

where b_1 and $b_2 \sim$ i.i.d. Beta(0.1, 1.0). The Beta(α, β) density is

$$\mathrm{Beta}(x \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}.$$

Similar to the subdivisions in gorilla, we have two subdivisions within chimpanzee, resulting in western chimpanzee ($\mathrm{W_c}$), Nigerian chimpanzee (N), and central chimpanzee (C). The structure among human and the three chimpanzee populations is assumed to be (((W$_\mathrm{c}$, N), C), H). The times T_{Chimp1} and T_{Chimp2} represent divergence times between W$_\mathrm{c}$ and N and between (W$_\mathrm{c}$, N) and C, respectively. Given the human-chimp split time T_{HC} and $T_{\mathrm{Chimp1}} \leq T_{\mathrm{Chimp2}}$, the prior distributions on T_{Chimp1} and T_{Chimp2} are

$$\begin{aligned} T_{\mathrm{Chimp1}} &= \min(b_3, b_4) \times T_{\mathrm{HC}}, \\ T_{\mathrm{Chimp2}} &= \max(b_3, b_4) \times T_{\mathrm{HC}}, \end{aligned}$$

where b_3 and b_4 are distributed i.i.d. from the Beta(0.1, 1.0) distribution.

Collectively, we let $\boldsymbol{T} = \{ T_{\mathrm{EM}}, T_{\mathrm{WEM}}, T_{\mathrm{Chimp1}}, T_{\mathrm{Chimp2}}, T_{\mathrm{HC}}, T_{\mathrm{GHC}}\}$ represent all six divergence times. Then the joint prior distribution of $\boldsymbol{T}$ is

$$f(\boldsymbol{T}) = f(T_{\mathrm{HC}}) f(T_{\mathrm{GHC}} \mid T_{\mathrm{HC}}) f(T_{\mathrm{EM}}, T_{\mathrm{WEM}} \mid T_{\mathrm{GHC}}) f(T_{\mathrm{Chimp1}}, T_{\mathrm{Chimp2}} \mid T_{\mathrm{HC}}).$$

3.2 Effective population size and coalescent rate

The effective population size, N, is defined as the number of breeding individuals in an idealized population that would show the same amount of dispersion of allele frequencies under the neutral selection model (Wright 1931, 1938), and is a basic parameter in many models in population genetics. For example, in the standard continuous-time coalescent theory, coalescent times are measured in units of $2N$ generations for diploid alleles. That is, the standard coalescent time t_k, the waiting time until the next coalescent event when there are k current sequences, has the exponential density function

$$f(t_k) = \binom{k}{2} \mathrm{e}^{-\binom{k}{2} t_k},$$

where time t_k is measured in units of $2N$ generations.

In most literature on population genetics, population sizes are estimated through nucleotide diversity (Nei and Li 1979), based on the relationship $\pi = 4N\mu$ (Tajima 1983). However, in this paper, an effective population size is estimated by converting a coalescent rate. A coalescent rate is a scale factor to change the unit of coalescent time from $2N$ generations into MYA which is easier to interpret. With a coalescent rate θ, the time is converted into MYA and the density function of the newly scaled time, t_k, is

$$f(t_k) = \binom{k}{2} \theta \mathrm{e}^{-\binom{k}{2}\theta t_k}.$$

For this conversion, a coalescent rate is defined as

$$\theta = \begin{cases} 1/(2N \times g \times 10^{-6}) & \text{for Numt sequences,} \\ 1/(N_f \times g \times 10^{-6}) & \text{for HV1 sequences,} \end{cases}$$

where N is the effective population size of a population and N_f is the effective population size of female individuals in the same population. A larger coalescent rate results in a shorter coalescent time. Since our model allows a separate population size for each population, a separate coalescent rate is used in each population. Let θ_i denote the HV1 coalescent rate of population edge i in Figure 1. Then the Numt coalescent rate of edge i is $\theta_i/4$ provided $N_f = N/2$. The effective population size N_i of edge i is estimated through the estimate of θ_i and an assumed generation time.

We consider θ_i as random and assign a prior distribution for θ_i that is the distribution function of the ratio of independent exponential random variables. The distribution of the ratio of two independent exponentials is

$$\mathrm{P}\left(\frac{X_1}{X_2} \leq x\right) = 1 - \frac{1}{1+\lambda x},$$

where $X_1 \sim \exp(\lambda)$ and $X_2 \sim \exp(1)$. Then the density function of this ratio of exponentials prior distribution with a positive hyper parameter λ is

$$f(x \mid \lambda) = \frac{\lambda}{(1+\lambda x)^2}, x \geq 0.$$

The proof is in Appendix 1.

The advantage of this ratio prior distribution is that it is weakly informative because it is relatively flat on $(0, \infty)$, but is proper. It does not have a mean value, but it is easy to calculate quantiles, and in particular, the median is $1/\lambda$. Therefore, the prior median of θ_i with the hyper parameter λ_θ is $1/\lambda_\theta$, implying the prior median of N_i is $\lambda_\theta \times 100,000$ if the generation time g is assumed as 20 years. To loosen the dependence of the inferences on a particular choice of λ_θ, we model $\lambda_\theta \sim \mathrm{Beta}(1,2)$. The density function of $\mathrm{Beta}(1,2)$ is linearly decreasing from two to zero with the mean 0.33. This corresponds to a marginal prior median for N_i of about 33,000.

Collectively, let $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_{13}, \lambda_\theta)$. Then the joint distribution of $\boldsymbol{\theta}$ is

$$\begin{aligned} f(\boldsymbol{\theta}) &= f(\lambda_\theta) f(\theta_1, \cdots, \theta_{13} \mid \lambda_\theta) \\ &= \text{Beta}(\lambda_\theta \mid 1, 2) \times \left\{ \prod_{i=1}^{13} f(\theta_i \mid \lambda_\theta) \right\}. \end{aligned}$$

More details about the use of coalescent rates are described in the hybrid coalescent process of section 3.5.

3.3 Transfer rate

Coalescent rates introduced in the previous section are used in the generation of an inheritance relationship among homologous alleles in a single lineage. In addition to the separate coalescent processes for Numt and HV1 sequences, the model needs to account for the pre-introgression homologies. For this purpose, we develop a new model, called a transfer model. An introgression event occurs when mtDNA is copied into a nuclear genome in the past. However, in the transfer model, the introgression event is viewed backward in time to be able to jointly model it with the coalescent processes. We model the transfer waiting time with an exponential distribution with a parameter η, called a transfer rate. The transfer event is available on the oldest ancestral lineages of each Numt group after all the members in the group coalesce into a single Numt lineage. Let l be the number of the oldest ancestral Numt lineages which are available for the next transfer event. Then the density function of the waiting time until the next transfer event is defined as

$$f(t_l) = \binom{l}{1} \eta \mathrm{e}^{-\binom{l}{1}\eta t_l},$$

where a transfer rate η is interpreted as the mean number of introgressions per MY, then $1/\eta$ means the duration time until the next introgression events measured in MY.

A common transfer rate is used for all populations. We also use a ratio of exponentials prior distribution for the transfer rate η. The hyper parameter for this distribution, λ_η, should reflect the range of likely values (we use $\lambda_\eta = 10$). The ratio of exponentials prior distribution on η with the hyper parameter 10 corresponds to the ratio of exponentials prior distribution on $1/\eta$ with the hyper parameter $1/10$, resulting in the posterior median 10. The right skewed distribution of the ratios of the exponentials with median 10 seems to cover quite a broad range for the duration time until an introgression. It covers the range 0–10 MY with probability 0.5.

Section 3.5 provides details about how to develop a homologous evolutionary relationship among Numt and HV1 sequences based on the coalescent theory and the transfer model.

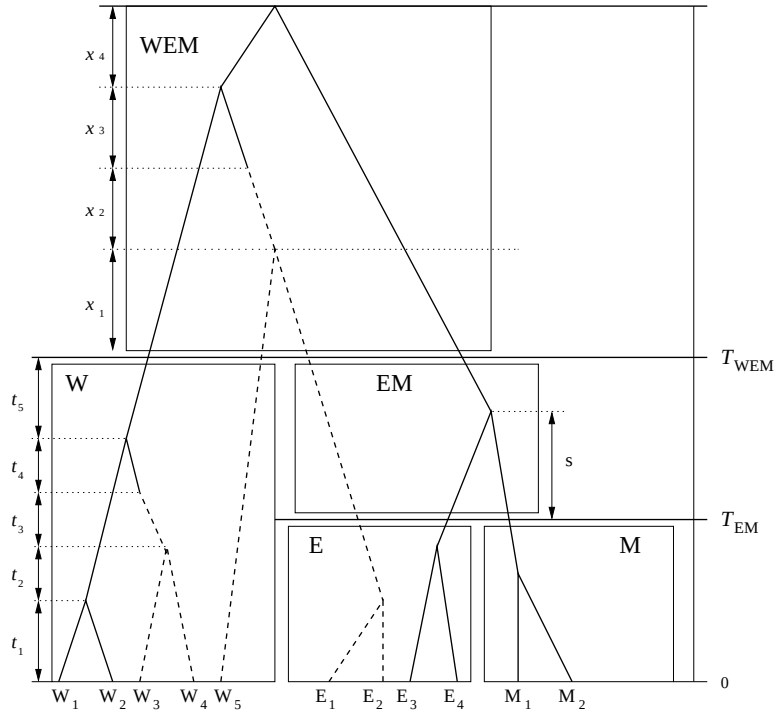

Figure 2: **An example of hybrid coalescent processes in a population tree with three taxa.** Solid lines represent HV1 evolution, and dashed lines are for Numt evolution. The change from a dashed line into a solid line represents the transfer event. Interevent time t_i's and population divergence time T_x's are measured in units of MY.

3.4 Numt partition

We assume a single introgression event on a lineage and do not allow the inverse of it. Since introgression events are assumed to be independent of each other, we do not allow for descendants of an introgression to coalesce with lineages under another introgression event given introgression events. In other words, a Numt coalescent event can occur only within the same Numt group. Also a Numt transfer event is available only on a lineage which is the oldest ancestor of each Numt group. Therefore, a Numt transfer event can happen only after all Numt lineages coalesce into a single lineage.

We model this partition of n Numt taxa given k introgression events, $P_t(n \mid k)$, using the Dirichlet process prior, Dirichlet$(P_t(n \mid k) \mid \alpha)$=Dirichlet$(n_1, \cdots, n_x, \cdots, n_k \mid \alpha)$ with probability density $\frac{\alpha^k \prod_{i=1}^{k}(n_i-1)!}{A_n(\alpha)}$ where $A_n(\alpha) = \alpha(\alpha+1)(\alpha+2)\cdots(\alpha+n-1)$. Let n_i denote the number of taxa belonging to i–th Numt group, and $n = \sum_1^k n_i$. In the example of Figure 2, there are 5 Numt taxa at 0 MYA, that is, $n = 5$ in total, and two Numt groups (i.e, $k = 2$). One group has two Numt taxa, W_3 and W_4 with the group size $n_1 = 2$. The other group has three Numt taxa, W_5, E_1 and E_2 with the group size $n_2 = 3$. Then the Numt partition in Figure 2 is denoted as $P_t(5 \mid 2) = (2, 3)$.

3.5 Gene genealogy derived from the hybrid coalescent process

In this section, we describe how to build a gene genealogy, representing the inheritance relationship among alleles, for Numt and HV1 sequences. The gene genealogy, G, is represented by the tree topology and branch lengths. In our model, the tree topology and branch lengths are specified by histories of HV1 coalescence, Numt coalescence, and Numt transfer events over the population tree. Events are generated from the present time to the ancestral state looking backward in time.

We call the combination of the HV1 coalescent process, Numt coalescent process and Numt transfer event the "hybrid coalescent process". A hybrid coalescent process is a Markov process going backward in time. Suppose there are m HV1 sequences and the partition of n total Numt taxa into k Numt groups is $P_t(n \mid k) = (n_1, \cdots, n_k)$, where $n = \sum_{i=1}^{k} n_i$. The population is assumed to have coalescent rates θ_{HV1} and θ_{Numt} for HV1 and Numt sequences, respectively, and a transfer rate η. Suppose there are $m(c)$ HV1 lineages, $n_i(c)$ Numt lineages for the i–th Numt group ($1 \le i \le k$) at the current state in a population, we only count lineages available for the next hybrid coalescent events in the population no matter what the histories other populations have. For example, at the beginning of population WEM at time T_{WEM} MYA in Figure 2, two HV1 lineages ($m(c) = 2$) and two Numt lineages are entering the population. And both of the two Numt lineages belong to the Numt group with taxa (W_5, E_1, E_3) ($n_2 = 3$). According to the order in the previous section, we have $n_1(c) = 0$ and $n_2(c) = 2$. Then, in Figure 2, the state of population WEM at time T_{WEM} is denoted as $(m(c), n_1(c), n_2(c)) = (2, 0, 2)$.

Let the current state be denoted as $(m(c), n_1(c), \cdots, n_x(c), \cdots, n_k(c))$. The allowed

transitions from the current state are

$$\begin{cases} (\boldsymbol{m(c)-1} & ,n_1(c), & \cdots, & n_x(c) & ,\cdots, & n_k(c)) & \text{if } m(c)\geq 2 & : (\boldsymbol{S1}), \\ (m(c) & ,n_1(c), & \cdots, & \boldsymbol{n_x(c)-1} & ,\cdots, & n_k(c)) & \text{if } n_x(c)\geq 2 & : (\boldsymbol{S2}), \\ (\boldsymbol{m(c)+1} & ,n_1(c), & \cdots, & \boldsymbol{n_x(c)-1} & ,\cdots, & n_k(c)) & \text{if } n_x(c)=1 & : (\boldsymbol{S3}). \end{cases}$$

We use bold to emphasize the change after each transition. ($\boldsymbol{S1}$) corresponds to the state that the total number of HV1 lineages are reduced into $m(c)-1$ from $m(c)$ after an HV1 coalescence. ($\boldsymbol{S2}$) is the result of a Numt coalescence in the Numt group with $n_x(c)$ Numt lineages at the current time. The transition into ($\boldsymbol{S3}$) illustrates a Numt transfer event on the ancestral edge of the Numt group with n_x taxa. We assumed a Numt transfer event is available on the oldest ancestral lineages of each Numt group. The condition $n_x(c)=1$ means all the n_x Numt taxa in the group have coalesced into a single lineage, and the oldest ancestral Numt lineage is waiting for a transfer event.

In population WEM of Figure 2, the state at time T_{WEM} is $(m(c), n_1(c), n_2(c)) = (2,0,2)$. At time $T_{\text{WEM}}+x_1$, the state of $(m(c), n_1(c), n_2(c))$ changes into $(2,0,1)$ after a transition (S2). At time $T_{\text{WEM}}+x_1+x_2$, the state $(m(c), n_1(c), n_2(c)))$ becomes $(3,0,0)$ after a transition (S3). And then sequential transitions of (S1) at time $T_{\text{WEM}}+x_1+x_2+x_3$ and at time $T_{\text{WEM}}+x_1+x_2+x_3+x_4$ result in the states $(2,0,0)$ and $(1,0,0)$, respectively.

We model the duration time for a transition with an exponential density. Note $\min(X,Y)\sim\exp(a+b)$ when $X\sim\exp(a)$, $Y\sim\exp(b)$ and X and Y are independent. We assume the three transitions $(S1)$, $(S2)$ and $(S3)$ are independent given the current state in a population edge. Then the duration time for a transition has an exponential density with rate $(\binom{m(c)}{2}\theta_{\text{HV1}}+\sum_{i=1}^{k}\binom{n_i(c)}{2}\theta_{\text{Numt}}+\{\sum_{i=1}^{k}\text{I}_{n_i(c)=1}\}\eta)$ where $\text{I}_{n_i(c)=1}$ is 1 if $n_i(c)=1$ and 0 otherwise. If a transition is chosen, this change is reflected on the construction of a gene-genealogy either by combining two randomly selected lineages into a single coalescent lineage or by changing a Numt lineage into an HV1 lineage.

Following is the hybrid coalescent algorithm to generate a gene genealogy in a single population edge with $m(c)$ HV1 lineages and $(n_1(c),\cdots,n_k(c))$ Numt lineages available at the present. B denotes the edge length of the population.

1. Set the cumulative interevent time $S=0$.
2. Calculate
$$a=\binom{m(c)}{2}\theta_{\text{HV1}}, b=\sum_{i=1}^{k}\binom{n_i(c)}{2}\theta_{\text{Numt}} \text{ and } c=\{\sum_{i=1}^{k}\text{I}_{n_i(c)=1}\}\eta.$$
3. Generate a waiting time (i.e., interevent time) for a hybrid coalescent process from an exponential distribution with a rate $(a+b+c)$ from 2 and add it to S.
4. If the cumulative interevent time S exceeds the population edge length, B, ignore the event and pass all the lineages to the ancestral population, then stop. Otherwise, continue.

5. Choose an event type randomly with probability $(\frac{a}{a+b+c}, \frac{b}{a+b+c}, \frac{c}{a+b+c})$ for HV1 coalescent, Numt coalescent, and Numt transfer, respectively. If Numt coalescent is selected, we need to pick up a Numt group at random with probabilities $\frac{\binom{n_i(c)}{2}}{\sum_{i=1}^{k}\binom{n_i(c)}{2}}$. If Numt transfer event is the next transition, we also need to pick up a Numt group uniformly at random among the Numt groups satisfying $n_x(c) = 1$.

6. Randomly choose gene-lineage(s) corresponding to the event. If the chosen event is a coalescent event, then two lineages with the same sequence type are selected uniformly at random, and the two sequences are combined to a single lineage, representing their ancestral lineage. Now, the two lineages are not considered in the remaining process. Instead, their ancestral lineage joins the rest of the processes, therefore, the number of lineages is reduced by one (either $m(c) \to m(c) - 1$ or $n_x(c) \to n_x(c) - 1$ for some x). If the chosen event is a Numt transfer event, the Numt lineage corresponding to the Numt group selected in 5 is treated as an HV1 lineage ($m(c) \to m(c)+1$ and $n_x(c) \to 0$ for some x satisfying $n_x(c) = 1$).

7. Stop when $m = 1$, $n = \sum_{i=1}^{k}\binom{n_i(c)}{2} = 0$ and $\sum_{i=1}^{k} \mathrm{I}_{n_i(c)=1} = 0$. Otherwise, go to 2.

Now, we use H to refer to the whole hybrid coalescent process history to determine a gene genealogy G over the population tree. Then the probability to have a gene genealogy G is equal to the probability to have such hybrid coalescent process history over the population tree, that is, $f(G|\boldsymbol{T}, \boldsymbol{\theta}, \eta, P_t) = f(H|\boldsymbol{T}, \boldsymbol{\theta}, \eta, P_t)$. Then, $H = \cup_i H_i$ where H_i is the hybrid coalescent process history of population edge E_i. We also define H_i^D as the histories of all edges descendant to edge i. Given H_i^D, we assume the hybrid coalescent process history H_i in population edge i is independent of H_j if $i \neq j$. For example, in Figure 2, the history in population W is independent of the history in population EM. Then the probability density of a gene genealogy G generated by the hybrid coalesce history H is obtained as

$$\begin{aligned} f(G|\boldsymbol{T}, \boldsymbol{\theta}, \eta, P_t) &= f(H|\boldsymbol{T}, \boldsymbol{\theta}, \eta, P_t) \\ &= \prod_i f(H_i|\boldsymbol{T}, \boldsymbol{\theta}, \eta, H_i^D, P_t). \end{aligned}$$

For example, the probability in the population WEM, $f(H_{\mathrm{WEM}} \mid \boldsymbol{T}, \boldsymbol{\theta}, \eta, H_{\mathrm{WEM}}^D, P_t)$ is

as following,

$$\begin{aligned} f(H_{\mathrm{WEM}}|\boldsymbol{T},\boldsymbol{\theta},\eta,H^{D}_{\mathrm{WEM}},P_t) &= \frac{\binom{2}{2}\theta_{\mathrm{Numt}}}{r_1}\times r_1\mathrm{e}^{-r_1x_1}\times\frac{1}{\binom{2}{2}} \\ &\times \frac{\binom{1}{1}\eta}{r_2}\times r_2\mathrm{e}^{-r_2x_2}\times\frac{1}{\binom{1}{1}} \\ &\times \frac{\binom{3}{2}\theta_{\mathrm{HV1}}}{r_3}\times r_3\mathrm{e}^{-r_3x_3}\times\frac{1}{\binom{3}{2}} \\ &\times \frac{\binom{2}{2}\theta_{\mathrm{HV1}}}{r_4}\times r_4\mathrm{e}^{-r_4x_4}\times\frac{1}{\binom{2}{2}} \\ &= \mathrm{e}^{-(r_1x_1+r_2x_2+r_3x_3+r_4x_4)}\times\theta_{\mathrm{Numt}}\times{\theta_{\mathrm{HV1}}}^2\times\eta, \end{aligned}$$

where

$$\begin{aligned} r_1 &= \binom{2}{2}\theta_{\mathrm{HV1}}+\binom{0}{2}\theta_{\mathrm{Numt}}+\binom{2}{2}\theta_{\mathrm{Numt}}+\binom{0}{1}\eta, \\ r_2 &= \binom{2}{2}\theta_{\mathrm{HV1}}+\binom{0}{2}\theta_{\mathrm{Numt}}+\binom{1}{2}\theta_{\mathrm{Numt}}+\binom{1}{1}\eta, \\ r_3 &= \binom{3}{2}\theta_{\mathrm{HV1}}+\binom{0}{2}\theta_{\mathrm{Numt}}+\binom{0}{2}\theta_{\mathrm{Numt}}+\binom{0}{1}\eta, \\ r_4 &= \binom{2}{2}\theta_{\mathrm{HV1}}+\binom{0}{2}\theta_{\mathrm{Numt}}+\binom{0}{2}\theta_{\mathrm{Numt}}+\binom{0}{1}\eta. \end{aligned}$$

Note $\binom{0}{2}=0$ and $\binom{1}{2}=0$. So far, we have introduced how to generate a gene-genealogy with the hybrid coalescent process given the population structure, the number of sequences of each in each tip, and the coalescent and transfer rates. In section 3.6, we describe how this evolutionary tree is used in the evaluation of the likelihood of DNA sequence data.

3.6 Computing the Likelihood

Let $D=\{D_i\}$ be the entire data set, where D_i represents the sequence alignment at site i, (i=1, $\cdots$, 236 for our data). The probability of DNA sequence data (D) given gene genealogy (G), divergence times ($\boldsymbol{T}$), Numt partition (P_t) and all other parameters ($\Theta=\{\theta_1,\ \cdots,\ \theta_{13},\ \lambda_\theta,\ \eta,\ \mu_{\mathrm{HV1}},\ \mu_{\mathrm{Numt}},\ \kappa_{\mathrm{HV1}},\ \kappa_{\mathrm{Numt}}\}$) is the traditional likelihood in phylogenetics,

$$f(D\mid\boldsymbol{T},\Theta,G,P_t)=\prod_{i=1}^{236}f(D_i\mid\boldsymbol{T},\Theta,G,P_t),$$

where each site D_i is assumed to be independent of one another, and $f(D_i\mid G,\boldsymbol{T},\Theta,P_t)$ is the likelihood for i–th site based on the nucleotide substitution model (Felsenstein

1981). Here we use the HKY model (Hasegawa et al. 1985) allowing separate transition and transversion rates. We incorporate substitution rates and transition-transversion rates for each of HV1 and Numt sequences, and they are $\mu_{\text{HV1}}, \mu_{\text{Numt}}, \kappa_{\text{HV1}}$ and κ_{Numt}.

Let's say the transition probability $P_{ij}(t \mid \mu, \kappa)$ is the probability that the base at a single site changes from state i to state j after time t MY(million years) based on the HKY model with a substitution rate μ and a transition-transversion ratio κ. If a sequence lineage has a transfer event on it, then the lineage has two parts with different evolutionary processes. The part of the lineage close to the present time evolves as a part of a nucleus genome and the other part evolves in a mitochondrial genome. If the partial branch length as an HV1 is t_1 and the partial branch length as a Numt is t_2, the transition probability on this lineage is

$$P_{ij}(t_1 + t_2 \mid \boldsymbol{\mu}, \boldsymbol{\kappa}) = \sum_{k \in \{A,C,G,T\}} P_{ik}(t_1 | \mu_{\text{HV1}}, \kappa_{\text{HV1}}) \times P_{kj}(t_2 | \mu_{\text{Numt}}, \kappa_{\text{Numt}}),$$

where $\boldsymbol{\mu}$=(μ_{HV1}, μ_{Numt}) and $\boldsymbol{\kappa}$=(κ_{HV1}, κ_{Numt}). We model μ_{HV1}, μ_{Numt}, κ_{HV1} and κ_{Numt} with independent ratio of exponentials prior distributions using separate hyper parameter values.

The pruning algorithm Felsenstein (1973) and Felsenstein (1981) enables rapid and practical computation of the likelihood of a phylogeny. Even using the partial likelihood modified by Larget and Simon (1999), we can speed up the computation.

To take account of variability in substitution rates over sites, we implemented the discrete Gamma rate variation model (Yang 1994). Then the complete likelihood is

$$f(D|G, \Theta, \boldsymbol{T}, P_t) = \prod_{i=1}^{236} \left\{ \sum_{x=1}^{4} \frac{1}{4} f(D_i \mid G, \boldsymbol{T}, q_x \times \boldsymbol{\mu}, \boldsymbol{\kappa}, P_t) \right\},$$

where q_x is the $\frac{2\times x-1}{8}$ quantile of a Gamma (0.2, 0.2) distribution, which has 1 as its mean value and density function

$$\text{Gamma}(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} \text{e}^{-\beta x} x^{\alpha-1}.$$

In the next section, we describe how the prior distributions of parameters and the likelihood from a sequence tree are integrated in Bayesian inference.

3.7 Bayes estimation of parameters

The state space of our model can be represented as ($\boldsymbol{T}$, Θ, G, P_t), where $\boldsymbol{T}$ is the population divergence times, Θ= $\{\theta_1, \cdots, \theta_{13}, \lambda_\theta, \eta, \mu_{\text{HV1}}, \mu_{\text{Numt}}, \kappa_{\text{HV1}}, \kappa_{\text{Numt}}\}$ are other numerical parameters, G is the sequence genealogy, and $P_t(\cdot \mid \cdot)$ is the Numt partition. Bayesian inference over the state space ($\boldsymbol{T}$, Θ, G, P_t) given data is based on the posterior distribution $f(\boldsymbol{T}, \Theta, G, P_t \mid D)$. Based on the Bayes Theorem, $f(\boldsymbol{T}, \Theta, G, P_t \mid D)$ can be evaluated as

$$f(\boldsymbol{T}, \Theta, G, P_t \mid D) = \frac{f(D \mid \boldsymbol{T}, \Theta, G, P_t) f(G \mid \boldsymbol{T}, \Theta, P_t) f(\boldsymbol{T}) f(\Theta) f(P_t)}{\int_{(\boldsymbol{T}, \Theta, G, P_t)} f(D \mid \boldsymbol{T}, \Theta, G, P_t) f(G \mid \boldsymbol{T}, \Theta, P_t) f(\boldsymbol{T}) f(\Theta) f(P_t)}.$$

However, exact calculation of $\int_{(\boldsymbol{T},\Theta,G,P_t)} f(D \mid \boldsymbol{T},\Theta,G,P_t)f(G \mid \boldsymbol{T},\Theta,P_t)f(\boldsymbol{T})f(\Theta)f(P_t)$ is infeasible. Therefore, we use Markov chain Monte Carlo (MCMC) (Hastings 1970) which only requires calculation of $f(D \mid \boldsymbol{T},\Theta,G,P_t)f(G \mid \boldsymbol{T},\Theta,P_t)f(\boldsymbol{T})f(\Theta)f(P_t)$, the product of the likelihood and the prior distributions. Note

$$f(\Theta) = \{\prod_{i=1}^{13} f(\theta_i \mid \lambda_\theta)\}f(\lambda_\theta)f(\eta)f(\mu_{\mathrm{HV1}})f(\mu_{\mathrm{Numt}})f(\kappa_{\mathrm{HV1}})f(\kappa_{\mathrm{Numt}}).$$

For the Bayesian inference, the MCMC algorithm constructs a Markov chain whose stationary distribution is $f(\boldsymbol{T},\Theta,G,P_t \mid D)$ by proposing a new state $(\boldsymbol{T}^*, \Theta^*, G^*, P_t^*)$ with proposal density $q(\cdot \mid \boldsymbol{T},\Theta,G,P_t)$. The new state is accepted with probability

$$R = \min\left\{1, \frac{f(\boldsymbol{T}^*,\Theta^*,G^*,P_t^* \mid D)}{f(\boldsymbol{T},\Theta,G,P_t \mid D)} \times \frac{q(\boldsymbol{T},\Theta,G,P_t \mid \boldsymbol{T}^*,\Theta^*,G^*,P_t^*)}{q(\boldsymbol{T}^*,\Theta^*,G^*,P_t^* \mid \boldsymbol{T},\Theta,G,P_t)}\right\}.$$

The proposal density q can be rather flexible as long as it specifies an aperiodic and irreducible Markov chain. The proposal density we use for $\boldsymbol{T}$, G and P_t is a mixture of many small proposals, each of which proposes changes to small subsets of the full set of parameters, like modifying population divergence times, modifying the hybrid coalescent history in a population, changing the number of transfer events, pruning and re-grafting of gene genealogy and re-scaling the whole trees. More details about the complete set of MCMC proposals are in Appendix 2. For a numerical parameter in Θ, a parameter is chosen at random with a probability 0.5/13 for N_i and a probability 0.5/6 for λ_θ, η, μ_{HV1}, μ_{NUMT}, κ_{HV1} and κ_{NUMT} right after every proposal of ($\mathbf{T}$, G, P_t). Given the choice of a parameter in Θ, a new state is proposed as the product of the current value and a random multiplier from a gamma(2, 2) distribution, independently from the others.

3.8 Verification

We have implemented the new method in a program written in C++. Implementation of complex MCMC methods requires careful validation to ensure both that the desired acceptance probabilities are correct and the proposal methods are accurately coded. To validate the MCMC algorithm, we undertook several approaches. First, we verified that the method could sample from the prior distribution ignoring the sequence data by evaluating its likelihood to be one across the entire state space. The sample was consistent with the prior distribution. Secondly, we used Gelman and Rubin's convergence diagnostic (Gelman and Rubin 1992) implemented in R, which monitors convergence of MCMC output based on a comparison of within-chain and between-chain variances. Approximate convergence is diagnosed when the upper limit is close to one. Six MCMC chains with different starting points are sufficiently consistent for us to be confident in the good mixing in MCMC chains with the eight proposal methods.

The results presented in the next sections are from the combined samples from six MCMC runs. Each MCMC run discards the initial 10,000,000 sample points as burn-in,

Population	Numt I	Numt II	HV1
W	1, 2, 3	4, 5, 6	7, 8, 9
E	10, 11, 12	13, 14, 15	16, 17, 18
M	19, 20, 21		
H	22		
W_c	23		
N	24		
C	25		

Table 1: **The sequence id's of the simulated data by populations and by types.**

and then subsamples every 1000–th state from the next 10,000,000 states in each chain. Inferences are based on the remaining 60,000($=10,000\times 6$) sampled trees. See Appendix 3 for the result from each chain.

3.9 Simulation study

We performed a simulation study to check the efficacy of our method. We use the same population structure as Figure 1. We generated 25 DNA taxa in total, and assumed that each sequence has 100 nucleotide bases. Three HV1 sequences were assigned in each of the gorilla population W, E and M. In the other populations, a single HV1 sequence was assigned in each. We consider two groups of Numt sequences, defined by separate introgression events. The Numt groups are named Numt I and Numt II, respectively. Each of the gorilla populations W and E has six Numt sequences, in total, three from Numt I and another three from Numt II. We mimic the real data which do not have any Numt sequences in the populations other than the gorilla populations W and E. Table 1 shows the distribution of simulated data over the seven populations, providing the id of each taxon. After assigning the sequences into the populations and setting the model parameters, we ran the hybrid coalescent process to generate the gene-genealogy with 25 taxa.

On each site of the root node, a nucleotide base was determined at random with the probabilities 0.3156, 0.3577, 0.1077 and 0.2190 for A, C, G and T, respectively. These probabilities are obtained empirically from our gorilla data. Given the gene-genealogy and the nucleotide bases of the root node, we can generate the nucleotide bases of all the internal and terminal nodes based on the transition probabilities introduced in section 3.6. The list of nucleotide bases of each taxon becomes the sequence data, on which our method can be applied. This simulated data is available in a supplemented file. For the hyperparameter α in the Dirichlet process about the Numt partition, we used 0.5, accounting for the fewer sequences than the real data.

We ran six MCMC chains with the different starting points, selected randomly from each prior distribution. Each chain has length 1,000,000 and the first 100,000 were discarded as burn-in. We sampled every 1000–th state, then each chain generated 900 states as posterior samples. We checked the convergence of the six MCMC chains with

	Numt-HV1 case			only-HV1 case		
	mean	median	95% C.R.	mean	median	95% C.R.
	Population divergence time					
T_{EM}	0.485	0.409	(0.105, 1.22)	0.673	0.634	(0.133, 1.5)
T_{WEM}	1.14	1.08	(0.54, 2.02)	3.8	3.75	(1.33, 6.62)
T_{chimp1}	0.481	0.34	(5.98e-08, 1.6)	0.515	0.372	(5.9e-08, 1.7)
T_{chimp2}	1.45	1.45	(0.209, 2.55)	1.49	1.48	(0.00765, 2.7)
T_{HC}	4.77	4.67	(4.03, 5.88)	4.92	4.88	(4.04, 5.93)
T_{GHC}	7.97	8.18	(5.85, 8.97)	7.61	7.79	(5.35, 8.95)
	Effective population sizes					
N_1 (W)	128000	122000	(73100, 213000)	118000	113000	(61900, 203000)
N_2 (E)	40500	36800	(15800, 87000)	22800	20100	(7400, 53100)
N_3 (M)	12700	10200	(2580, 37900)	12700	10100	(2450, 38700)
N_4 (H)	81400	73000	(31600, 180000)	78700	69900	(29000, 177000)
N_5 ($\mathrm{W_c}$)	657000	71300	(4550, 2010000)	656000	63800	(3430, 1330000)
N_6 (N)	435000	49200	(1030, 2010000)	577000	40800	(831, 1630000)
N_7 (C)	1530000	83600	(17400, 986000)	158000	75200	(15000, 749000)
N_8 (EM)	562000	87500	(2460, 2070000)	56900	25800	(717, 295000)
N_9 (WEM)	59600	56000	(29100, 111000)	129000	38100	(854, 578000)
N_{10} (Chimp1)	144000	69800	(2490, 611000)	217000	65600	(1640, 681000)
N_{11} (Chimp2)	59200	29500	(864, 296000)	54800	25500	(825, 281000)
N_{12} (HC)	70000	24000	(649, 366000)	111000	24900	(552, 474000)
N_{13} (GHC)	271000	110000	(2520, 1260000)	148000	50800	(997, 818000)
	Other parameters					
μ_{HV1}	0.0394	0.0386	(0.0252, 0.0586)	0.0471	0.0455	(0.0267, 0.0773)
μ_{Numt}	0.00307	0.00289	(0.00138, 0.00579)	NA	NA	NA
κ_{HV1}	19.7	19.4	(13.9, 27.7)	21.5	21	(14.3, 31.7)
κ_{Numt}	8.99	8.36	(4.39, 17.3)	NA	NA	NA
λ_θ	0.501	0.488	(0.182, 0.879)	0.423	0.4	(0.126, 0.835)
η	0.595	0.521	(0.218, 1.41)	NA	NA	NA

Table 2: **The posterior estimates of the parameters.** The estimates on the left are from the analysis with both Numt and HV1 sequences and the estimates on the right are from the analysis with only HV1 sequences.

Gelman and Rubin's convergence diagnostic, and they seem to converge. See Appendix 3 for more details.

4 Results

Table 2 summarizes the posterior estimates of the population divergence times, effective population sizes, HKY parameters, the hyperparameter λ_θ for the coalescent rates and transfer rate η. 95% C.R. stands for the 95% credible region from a posterior sample. More details of results for each parameter follow.

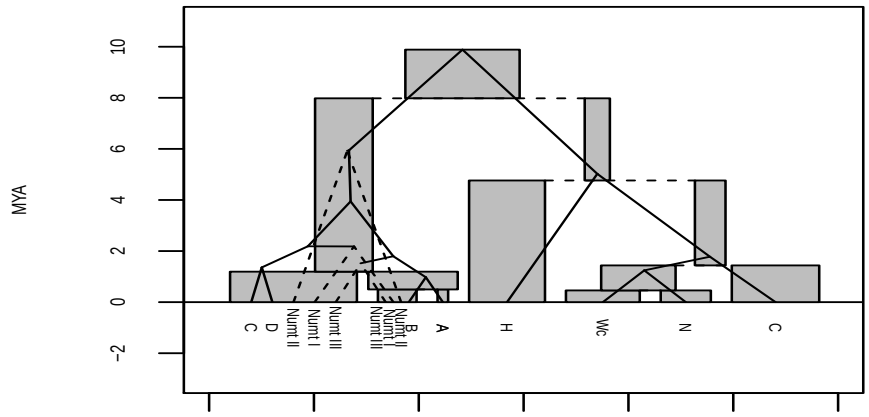

Figure 3: **The summary result from the analysis with Numt and HV1 sequences.** The linkage of boxes represents the population tree. The height of a box represents the lasting period of the population and the width is proportional to the estimated effective population size N. The solid line represents the evolutionary sequence tree for HV1 haplogroups and the dashed line represents the evolutionary sequence relationship among Numt sequences.

4.1 Population divergence times

Figure 3 summarizes the joint posterior samples from the analysis with the hybrid coalescent model. The connection of boxes represents the estimated population tree. The heights of boxes correspond to the posterior means of population divergence times. The widths of boxes are proportional to the posterior medians of the effective population sizes.

The connection of lines represents a sequence tree which simplifies the consensus sequence tree by replacing tips with haplogroups. A consensus tree summarizes the trees from the posterior samples. In biological taxonomy, a clade is a group that includes a common ancestor and all the descendants (internal and terminal) of that ancestor and a haplogroup is a group of sequences at each taxon whose alignments are very similar to one another. In this paper, only those clades present in the specified fraction ($\geq$50%) of the sampled trees will be present in the consensus tree. Solid lines are for HV1 sequences and dashed lines are for Numt sequences. We observe in Figure 3 that the relatively recent east-west split time may strongly depend on the recent coalescent event of eastern and western sequences in haplogroup Numt III. In contrast, note that the first estimated east-west coalescent event of HV1 sequences occurs around 4 MYA.

From the analysis with only HV1 sequences, the east-west split time T_{WEM} is estimated to have posterior mean 3.8 MYA with 95% credible region (1.33, 6.62) MYA. Thalmann et al. (2005) estimated 1.36 MYA as the divergence time of HV1 sequences

from the western and eastern gorillas applying a mutation rate 0.075/site/MY (Tamura and Nei 1993), whereas the mean substitution rate estimated from the only-HV1 case is 0.0471/site/MY. In terms of the number of substitutions, the east-west divergence time corresponds to a mean of 0.179 with 95% credible region (0.063, 0.312) contrary to the estimate 0.102 ($= 0.075 \times 1.36$) from Thalmann et al. (2005). The posterior distribution of T_{EM} from the only-HV1 case has mean 0.673 MYA with 95% credible region (0.133, 1.5) MYA. This interval is much wider than the interval (0.316, 0.443) MYA estimated in Jensen-Seaman and Kidd (2001).

When both Numt and HV1 sequences are used, the posterior distribution of T_{WEM} is estimated to have mean 1.14 MYA with 95% credible region (0.54, 2.02) MYA, which is at least two times smaller than the estimates from the only-HV1 analysis. Based on 16 noncoding nuclear loci, Thalmann et al. (2007) estimated east-west split time as 0.9–1.6 MYA followed by gene-flow until 0.08–0.2 MYA.

Other than T_{EM} and T_{WEM}, the rest of divergence times have similar estimates in the two cases. Chen and Li (2001) estimated the human-chimp divergence time as (4.6, 6.2) MYA and gorilla-human-chimp divergence time as (6.2, 8.4) MYA when taking the orangutan speciation time as (12, 16) MYA (orangutan is known to have split from gorilla-human-chimp first). In our analysis, the human-chimp divergence time and the gorilla-human-chimp divergence time are estimated at mean 4.77 MYA with 95% credibility region (4.03, 5.88) MYA and mean 7.97 MYA with 95% credible region (5.85, 8.97) MYA, respectively.

The Pleistocene Epoch of frequent glaciation is known to have lasted from around 1.8 MYA until about 11,000 years ago. 93.58% of the posterior samples for the east-west split time from Numt-HV1 case fall in this period. In contrast, only 6.26% of the posterior samples of the east-west split time from the only-HV1 case fall in the Pleistocene period. In the only-HV1 case, we find it to be highly likely that the divergence of HV1 sequences from the western and eastern gorillas predates the beginning of the Pleistocene period. More discussion will follow in section 5.

For sensitivity analysis of the prior distributions, especially that of the divergence times, we tried some other distributions on T_{WEM}, T_{EM}, T_{Chimp1} and T_{Chimp2}, then compared the results based on the simulated data. Instead of scaled Beta distributions, we used Uniform(0, T_{GHC}), Uniform(0, T_{WEM}), Uniform(0, T_{HC}) and Uniform(0, T_{Chimp2}) distributions, for T_{WEM}, T_{EM}, T_{Chimp2} and T_{Chimp1}, respectively, in orders. For each population divergence time, we compared the distributions of the independent Monte Carlo sample and the posterior sample, using each of the old prior and new prior distributions. With the original prior distributions, all these four divergence times tended to zero in both the prior and posterior sample. With the new prior distributions, the posterior divergence times showed the same tendency favoring values close to zero. Interestingly, the posterior distributions of T_{WEM} and T_{Chimp2} using the new prior distributions look similar to the posterior distributions of each using the old prior distributions, respectively, although the new prior distributions have much broader area with even probabilities than the old ones. See Appendix 3 for more results.

4.2 Effective population size

Note N_i denotes the effective population size of edge i in Figure 1. The posterior distribution of N_1 and N_2, the effective population sizes of populations W and E, have median 122,000 with 95% credible region (73100, 213000) and median 36800 with 95% credible region (15800, 87000), respectively. These estimates are much larger than those of 24,100 and 13,600 for western and eastern gorillas, respectively, from Thalmann et al. (2007). Not only are the estimates of N_1 and N_2 from our model larger than those from Thalmann et al. (2007), but also the ratio N_1/N_2 is 3.32 (= 122000/36800), much larger than 1.77 (= 24100/13600) from Thalmann et al. (2007).

The effective population size of the human-chimp ancestor was median 24,000 with 95% credible region (649, 366000) which is much wider than the estimates (52000, 96000) by Chen and Li (2001) The interesting finding in Table 2 is that the effective population size of western lowland gorillas is larger than any other populations, even larger than human, which agrees with the estimates in Yu et al. (2004). More discussion will follow in section 5.

4.3 tMRCAs of sequences and sequence trees

This section provides the estimates of divergences of DNA sequences. Figure 4 is the consensus sequence tree summarizing 60,000 tree samples from the analysis with both Numt and HV1 sequences. Tree samples are summarized by the **summarize** code in **BADGER** (Simon and Larget 2004). This consensus tree is comparable with the sequence tree in Figure 1 estimated in Clifford et al. (2004) in which those haplogroups were defined. In Figure 4, there are five haplogroups within population W and populations E and M have a single haplogroup each. It is notable that there are more complex subdivisions in HV1 sequences from population W, contrary to populations E and M. For Numt sequences, there are three Numt haplogroups, Numt I, Numt II, and Numt III. The HV1 haplogroups agree with gorilla populations, that is, haplogroup A for population M, haplogroup B for population E, and the rest of the haplogroups for population W. In contrast, each Numt haplogroup consists of sequences from both populations W and E.

Table 3 summarizes the tMRCAs. An interesting finding is that Numt I and Numt III coalesce with the ancestral lineages of HV1 sequences before they coalesce with Numt lineages from other Numt haplogroups. Numt I coalesces to the ancestral lineage of the HV1 sequences from population W, and Numt III coalesces to the ancestral lineage of the HV1 sequences from populations E and M. This implies that Numt I is close to the ancestral state of HV1 sequences of population W, and Numt III is close to the ancestral state of HV1 sequence of population E and M. These observations are best explained, perhaps, if two introgression events occurred after female gorillas were isolated into the east and west. Numt I could have arisen from an introgression event in the west with a male gorilla carrying the Numt to the east, and Numt III may have arisen in the east and been carried west by male migration. Numt II is estimated as an outgroup to the rest of the gorilla sequences with a posterior probability 0.55, consistent with

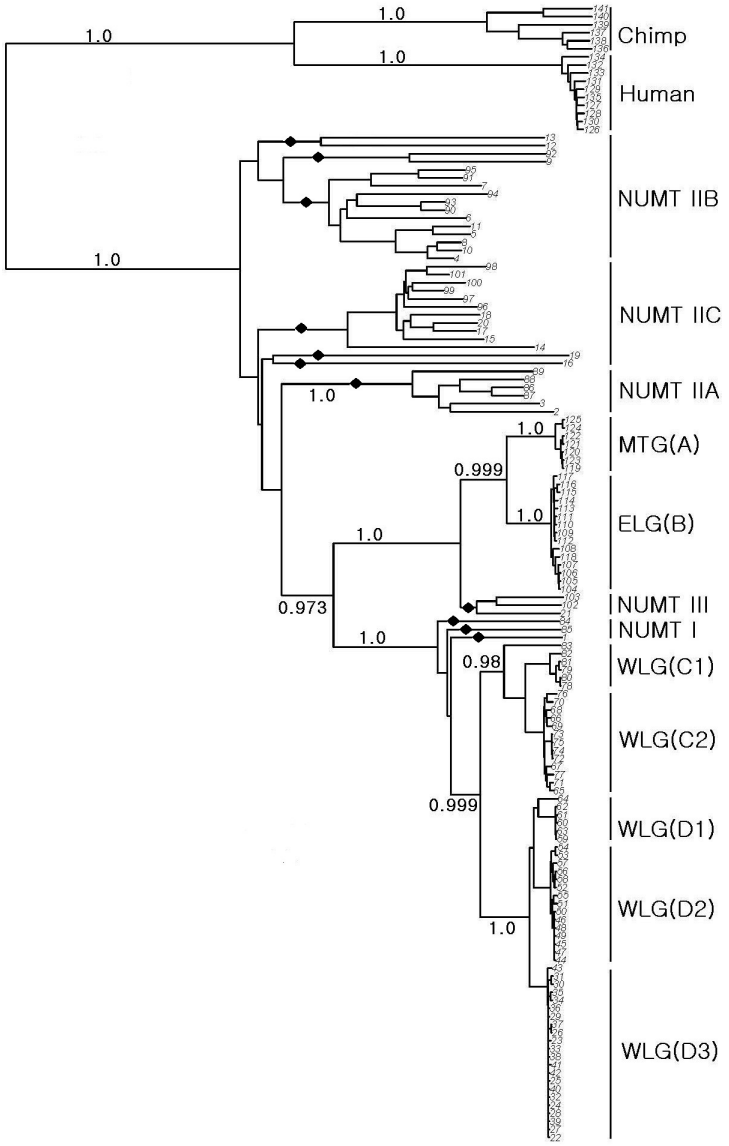

Figure 4: **The Consensus sequence tree from the posterior tree samples.** 60,000 sequence trees were sampled from the six MCMC runs. The numbers on the edges are posterior probabilities of the clades.

introgression prior to east-west isolation.

Another finding is discordant tMRCA of east-west samples between Numt and HV1 sequences. That is, tMRCA of east-west Numt sequences is 1.45 MYA with 95% credible

	Numt-HV1 case			only-HV1 case		
	mean	median	95% C.R.	mean	median	95% C.R.
A	0.149	0.134	(0.0432, 0.341)	0.15	0.133	(0.0397, 0.361)
B	0.225	0.208	(0.0938, 0.451)	0.186	0.17	(0.0696, 0.398)
C1	0.286	0.271	(0.135, 0.517)	0.265	0.249	(0.114, 0.511)
C2	0.965	0.927	(0.539, 1.59)	1	0.955	(0.513, 1.76)
D1	0.352	0.331	(0.131, 0.692)	0.367	0.342	(0.125, 0.764)
D2	0.157	0.147	(0.0645, 0.303)	0.167	0.155	(0.0651, 0.338)
D3	0.172	0.159	(0.0714, 0.346)	0.155	0.141	(0.0613, 0.332)
Numt I	2.2	2.12	(1.3, 3.53)	NA	NA	NA
Numt IIA	2.65	2.51	(1.12, 5)	NA	NA	NA
Numt IIB	5.08	5	(2.73, 7.88)	NA	NA	NA
Numt IIC	5.43	5.34	(3.55, 7.8)	NA	NA	NA
Numt II (IIA, IIB, IIC)	5.89	5.79	(3.84, 8.33)	NA	NA	NA
Numt III	1.45	1.38	(0.731, 2.46)	NA	NA	NA
A, B	0.957	0.917	(0.501, 1.65)	0.936	0.884	(0.443, 1.72)
C	0.965	0.927	(0.54, 1.59)	1	0.955	(0.513, 1.76)
D1, D2	0.47	0.448	(0.25, 0.814)	0.549	0.518	(0.254, 1.03)
D	0.54	0.518	(0.296, 0.899)	0.581	0.554	(0.279, 1.05)
C, D	1.37	1.32	(0.835, 2.17)	1.66	1.6	(0.898, 2.76)
Numt I, C, D	2.21	2.13	(1.31, 3.54)	NA	NA	NA
Numt III, A, B	1.73	1.68	(0.984, 2.74)	NA	NA	NA
Numt IIB, IIC	5.74	5.64	(3.7, 8.19)	NA	NA	NA
A, B, C, D, NumtI, III	3.98	3.88	(2.53, 5.99)	NA	NA	NA
All Gorilla	5.91	5.81	(3.87, 8.34)	4.44	4.3	(2.46, 7.21)
Human	0.515	0.486	(0.258, 0.93)	0.5	0.471	(0.233, 0.944)
P.t.Verus	0.501	0.476	(0.237, 0.901)	0.494	0.466	(0.22, 0.926)
P.t.Troglodytes	0.828	0.79	(0.403, 1.47)	0.792	0.752	(0.364, 1.45)
P.t.Verus, P.t.Vellerosus	1.25	1.21	(0.701, 2.06)	1.27	1.22	(0.657, 2.18)
All Chimp	1.79	1.74	(1.08, 2.79)	1.82	1.76	(1.01, 2.93)
Human, Chimp	5.03	4.93	(4.1, 6.54)	5.26	5.19	(4.13, 6.98)
Gorila, Human, Chimp	9.89	9.32	(7.21, 14.8)	8.68	8.52	(6.07, 12.9)

Table 3: **The posterior estimates of the tMRCAs.** The estimates on the left are from the analysis with both Numt and HV1 sequences and the estimates on the right are from the analysis with only HV1 sequences.

region (0.731, 2.46), which is the tMRCA of Numt III sequences. In contrast, tMRCA of east-west HV1 sequences is 4.4 MYA with 95% credible region (2.46, 7.21), almost three times larger compared to the Numt sequences.

4.4 Number of transfer events

Table 4 shows how often (in %) two Numt taxa are in the same Numt group. Note a Numt group is defined by an introgression event on the ancestral edge. We call this matrix as pairwise membership matrix. Then, we can identify 11 Numt groups in which taxa have the same membership for more than 70% of the 60,000 posterior samples, and

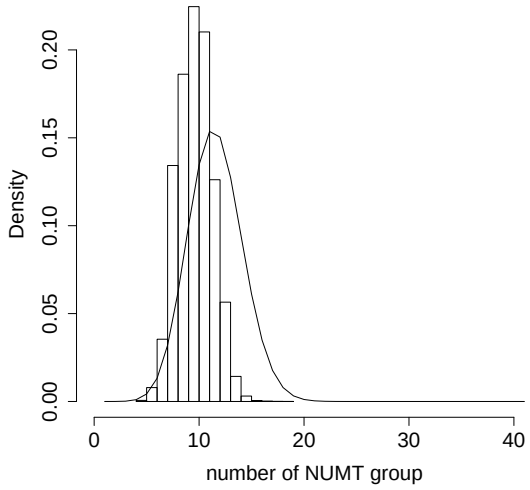

Figure 5: **The prior and posterior distributions of the number of Numt transfer events.** The histogram of the posterior samples for the number of Numt transfer events. The curve represents the prior density function.

those 11 groups are Group 1 (sequence 1), Group 2 (sequence 84), Group 3 (sequence 85), Group 4 (Numt IIA), Group 5 (Sequences 12 and 13), Group 6 (sequences 9 and 92), Group 7 (all sequences in Numt IIB other than sequences 9, 12, 13 and 92), Group 8 (sequence 16), Group 9 (sequence 19), Group 10 (all sequences in Numt IIC but for sequences 16 and 19), and Group 11 (Numt III). Sequences 1, 84 and 85 tend to have separate introgression events on each lineage, about 99% out of the 60,000 posterior samples. Sequences in Numt IIA belong to the same Numt group about 100% of the time. Sequences 12 and 13 are under the same Numt group about 98% of the time. Sequences 9 and 92 are in the same group almost 100% of the time. More than 86% of the time, sequences in Numt IIB except for sequences 9, 12, 13 and 92 are in the same Numt group. For sequences 16 and 19, they have separate introgressions on each lineage 73% of the time, and they are under the same introgression 27% of the time. Sequences in Numt IIC except for sequences 16 and 19 are in the same Numt group more than 92% of the time. Among Numt III sequences, sequences 21 and 102 are in the same group 92% of the time, and all sequences in Numt III are in the same group about 72% of the time. This implies two separate introgression events are estimated on each parent lineage of (sequence 21 and sequence 102) and sequence 103 for 20% in the posterior sample.

Figure 5 provides the posterior distribution of the number of introgression events, which is identical to the number of Numt groups. The relative frequency for the range from 8 to 13 is 93.79% of the time. The merge and division among (12, 13) (9, 92) and

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	84
1	-	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
2	-	-	**100**	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
3	-	-	-	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
4	-	-	-	-	**99**	**91**	**92**	**100**	43	**100**	**99**	44	45	≤	≤	≤	≤	≤	≤	≤	≤	≤
5	-	-	-	-	-	**91**	**92**	**99**	43	**99**	**100**	44	45	≤	≤	≤	≤	≤	≤	≤	≤	≤
6	-	-	-	-	-	-	**86**	**91**	49	**91**	**91**	43	44	≤	≤	≤	≤	≤	≤	≤	≤	≤
7	-	-	-	-	-	-	-	**92**	43	**92**	**92**	44	44	≤	≤	≤	≤	≤	≤	≤	≤	≤
8	-	-	-	-	-	-	-	-	43	**100**	**99**	44	45	≤	≤	≤	≤	≤	≤	≤	≤	≤
9	-	-	-	-	-	-	-	-	-	43	43	21	21	1.9	1.9	1.9	1.9	1.9	≤	1.9	≤	≤
10	-	-	-	-	-	-	-	-	-	-	**99**	44	45	≤	≤	≤	≤	≤	≤	≤	≤	≤
11	-	-	-	-	-	-	-	-	-	-	-	44	45	≤	≤	≤	≤	≤	≤	≤	≤	≤
12	-	-	-	-	-	-	-	-	-	-	-	-	**98**	≤	≤	≤	≤	≤	≤	≤	≤	≤
13	-	-	-	-	-	-	-	-	-	-	-	-	-	≤	≤	≤	≤	≤	≤	≤	≤	≤
14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**93**	17	**94**	**93**	≤	**93**	≤	≤
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	**99**	**99**	≤	**99**	≤	≤
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	17	27	18	≤	≤
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**100**	≤	**100**	≤	≤
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	≤	**100**	≤	≤
19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	≤	≤	≤
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	≤	≤
21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	≤
84	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
85	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
86	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
87	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
88	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
89	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
90	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
91	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
92	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
93	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
94	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
95	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
96	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
97	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
98	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
99	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
101	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
102	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
103	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103
1	1.2	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
2	≤	**100**	**100**	**100**	**99**	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
3	≤	**100**	**100**	**100**	**99**	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
4	≤	≤	≤	≤	≤	**98**	**93**	43	**98**	**94**	**93**	≤	≤	≤	≤	≤	≤	≤	≤
5	≤	≤	≤	≤	≤	**98**	**93**	43	**98**	**93**	**93**	≤	≤	≤	≤	≤	≤	≤	≤
6	≤	≤	≤	≤	≤	**92**	**87**	49	**92**	**89**	**87**	≤	≤	≤	≤	≤	≤	≤	≤
7	≤	≤	≤	≤	≤	**92**	**98**	43	**92**	**87**	**98**	≤	≤	≤	≤	≤	≤	≤	≤
8	≤	≤	≤	≤	≤	**99**	**93**	43	**98**	**94**	**93**	≤	≤	≤	≤	≤	≤	≤	≤
9	≤	≤	≤	≤	≤	43	43	**100**	43	43	43	1.9	1.9	1.9	1.9	1.9	1.9	≤	≤
10	≤	≤	≤	≤	≤	**99**	**93**	43	**98**	**94**	**93**	≤	≤	≤	≤	≤	≤	≤	≤
11	≤	≤	≤	≤	≤	**98**	**93**	43	**98**	**93**	**93**	≤	≤	≤	≤	≤	≤	≤	≤
12	≤	≤	≤	≤	≤	45	44	21	45	45	44	≤	≤	≤	≤	≤	≤	≤	≤
13	≤	≤	≤	≤	≤	45	44	21	45	45	44	≤	≤	≤	≤	≤	≤	≤	≤
14	≤	≤	≤	≤	≤	≤	≤	1.9	≤	≤	≤	**92**	**93**	**93**	**94**	**93**	**93**	≤	≤
15	≤	≤	≤	≤	≤	≤	≤	1.9	≤	≤	≤	**98**	**99**	**99**	**99**	**99**	**99**	≤	≤
16	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	18	17	17	17	17	17	≤	≤
17	≤	≤	≤	≤	≤	≤	≤	1.9	≤	≤	≤	**98**	**100**	**100**	**100**	**100**	**100**	≤	≤
18	≤	≤	≤	≤	≤	≤	≤	1.9	≤	≤	≤	**98**	**100**	**100**	**100**	**100**	**100**	≤	≤
19	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
20	≤	≤	≤	≤	≤	≤	≤	1.9	≤	≤	≤	**98**	**100**	**100**	**100**	**100**	**100**	≤	≤
21	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	**92**	**72**
84	11	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
85	-	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
86	-	-	**100**	**100**	**99**	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
87	-	-	-	**100**	**99**	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
88	-	-	-	-	**99**	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
89	-	-	-	-	-	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤	≤
90	-	-	-	-	-	-	**93**	43	**100**	**94**	**93**	≤	≤	≤	≤	≤	≤	≤	≤
91	-	-	-	-	-	-	-	43	**93**	**88**	**100**	≤	≤	≤	≤	≤	≤	≤	≤
92	-	-	-	-	-	-	-	-	43	43	43	1.9	1.9	1.9	1.9	1.9	1.9	≤	≤
93	-	-	-	-	-	-	-	-	-	**94**	**93**	≤	≤	≤	≤	≤	≤	≤	≤
94	-	-	-	-	-	-	-	-	-	-	**88**	≤	≤	≤	≤	≤	≤	≤	≤
95	-	-	-	-	-	-	-	-	-	-	-	≤	≤	≤	≤	≤	≤	≤	≤
96	-	-	-	-	-	-	-	-	-	-	-	-	**98**	**98**	**98**	**98**	**98**	≤	≤
97	-	-	-	-	-	-	-	-	-	-	-	-	-	**100**	**100**	**100**	**100**	≤	≤
98	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**100**	**100**	**100**	≤	≤
99	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**100**	**100**	≤	≤
100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**100**	≤	≤
101	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	≤	≤
102	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**73**
103	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 4: **The posterior estimates of pairwise membership.** Each entry x_{ij} represents the relative frequency (%) for Numt taxa i and j to belong to the same Numt group out of the 60,000 posterior MCMC samples. The values which are greater than 70 are in bold, and "≤" replaces values less than 1.

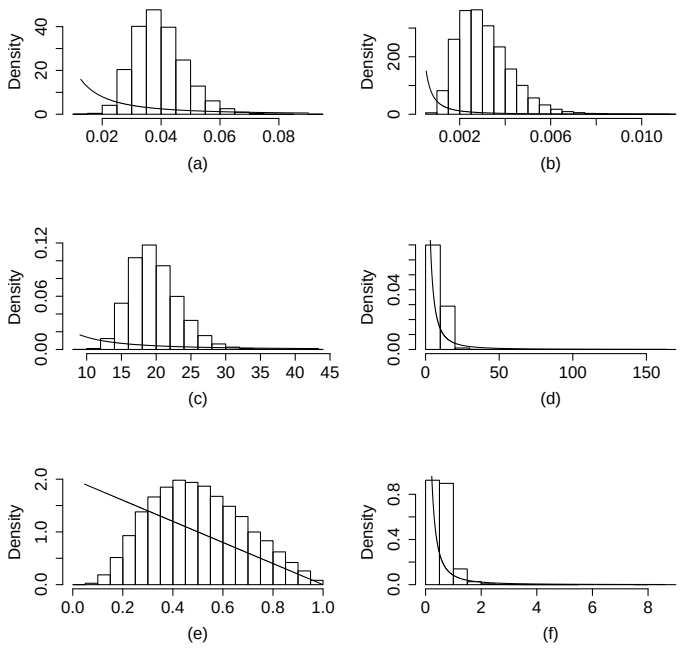

Figure 6: **Prior and posterior distributions of the HKY parameters, λ_θ, and η.** In each panel, the histogram shows the posterior distribution and the curve represents the prior density function for each parameter. (a), (b), (c), (d), (e) and (f) are for $\mu_{\rm HV1}$, $\mu_{\rm Numt}$, $\kappa_{\rm HV1}$, $\kappa_{\rm Numt}$, λ_θ and η, respectively, from the analysis with both Numt and HV1 sequences. The hyperparameter λ in the ratio of exponentials prior distributions is 200, 20,000, 0.5, 0.5 and 10 for $\mu_{\rm HV1}$, $\mu_{\rm Numt}$, $\kappa_{\rm HV1}$, $\kappa_{\rm Numt}$ and η, in the order. The prior distribution of λ_θ is the Beta(1, 2) distribution.

Numt IIB except for 9, 12, 13 and 92, the merge and division among (16), (19) and the rest of Numt IIC, and the merge and division between (21, 102) and (103) in Numt III can cover the high proportion over the range from 8 to 13 in Figure 5.

The α - clumpiness parameter in the Dirichlet process prior was set to 5 for the result in this paper. We explored the robustness of the method to misspecification of α. We did this with a prior mean of number of transfer events=4, 11 and 16 (corresponding to α=1, 5 and 10, respectively). We obtained the best mixing of MCMC running with $\alpha = 5$. With $\alpha = 10$, the mixing was not as good as the result with $\alpha = 5$, but the posterior estimates were close to the estimates with $\alpha = 5$. For the case using $\alpha = 1$, the mixing of MCMC running were too poor to use for inferences.

4.5 Other parameters

The substitution rates μ_{HV1} and μ_{Numt} are mutation rates measured in numbers of substitutions per site per MY for HV1 and Numt sequences, respectively. The prior distribution of μ_{HV1} is a ratio of exponentials prior distribution with median 0.005. The prior distribution of μ_{Numt} is also a ratio of exponentials prior distribution with median 0.00005. The posterior distribution for μ_{HV1} has mean 0.0394 with the 95% credible region (0.0252, 0.0586), smaller than the estimate 0.075 from Tamura and Nei (1993). The posterior estimate of μ_{Numt} has mean 0.00307 and 95% credible region (0.00138, 0.00579), larger than the estimate 0.00096 estimated from Thalmann et al. (2007) which is based on 16 noncoding nuclear loci under the assumption that humans and gorillas diverged 8 million years ago. The estimate of μ_{HV1} is larger than the estimate of μ_{Numt}, and this is consistent with the rate of mtDNA evolution in primates being higher than in nuclear DNA (Brown et al. 1982). Both κ_{HV1} and κ_{Numt} have ratio of exponentials prior distributions with median values 2. The posterior distribution κ_{HV1} has mean 19.7 with the 95% credible region (13.9, 27.7) and κ_{Numt} has posterior mean 8.99 with the 95% credible region (4.39, 17.3) in Numt-HV1 case. Jensen-Seaman and Kidd (2001) estimated 19 as the transition to transversion ratio based on 15 eastern gorilla mtDNA sequences.

The posterior distribution of μ_{HV1} for the only-HV1 case has mean 0.0471 with the 95% credible region (0.0267, 0.0773), which is a little larger than μ_{HV1} in the Numt-HV1 case. The Bayes estimate of κ_{HV1} in the only-HV1 case has mean 21.5 with the 95% credible region (14.3, 31.7), which is a little larger than the estimates from the Numt-HV1 case.

The hyperparameter λ_θ has a Beta(1, 2) prior distribution, with mean 0.33. The posterior distribution of λ_θ is estimated to have mean 0.501 with 95% credible region (0.182, 0.879). The transfer rate η has a ratio of exponentials prior distribution with median value 0.1. The posterior distribution for η has mean 0.595 with the 95% credible region (0.218, 1.41), where η is the expected number of transfer events per million years. Figure 6 compares the prior and posterior distributions of these six parameters.

4.6 Comparison with BEAST

BEAST (Bayesian evolutionary analysis by sampling trees, Drummond and Rambaut (2007)) is a software package providing a general Bayesian framework for parameter estimation and hypothesis testing of evolutionary models from molecular sequence data. BEAST aims to bring a large number of complementary evolutionary models (substitution model, insertion-deletion models, demographic model, tree shape prior, relaxed clock models, node calibration models) into a single coherent framework for evolutionary inference. However, the demographic parameters estimated in BEAST are for a single population, so it does not handle population subdivision structure. For this reason, it is not appropriate to use BEAST for direct comparison with our methods.

	tMRCA (in MYA)			Population size		
Groups	mean	median	95% C.R.	Mean	median	95% C.R.
W–HV1	2.1	2.1	(1.4, 3.4)	74000	73000	(47000, 110000)
E–HV1	0.21	0.19	(0.085, 0.4)	12000	11000	(4300, 27000)
M–HV1	0.14	0.13	(0.037, 0.33)	6100	4800	(1100, 19000)
W–Numt	47	45	(30, 73)	980000	940000	(560000, 1600000)
E–Numt	52	49	(29, 92)	550000	520000	(290000, 960000)
W–Numt II	19	18	(12, 28)	620000	590000	(330000, 1100000)
E–Numt II	12	11	(6.6, 23)	230000	210000	(110000, 450000)

Table 5: **The estimates of tMRCAs and effective population sizes from BEAST.**

Analysis of subsets of the data by sequence types and sampled populations

We apply BEAST to sequence data set by sequence types (Numt and HV1) and by sampled gorilla populations (W, E and M). There are five available combinations because Numt sequences are not available in population M. In this BEAST analysis, we use the molecular clock rate model and the HKY model for substitution process, and also take account of among-site rate variation with the discrete gamma prior as in our models. To convert estimated branch length from the number of mutations into calendar time (MYA), we use the substitution rate 0.0394×10^{-6}subs/site/year and 0.00307×10^{-6}subs/site/year for HV1 and Numt, respectively, which are the posterior means from our method (Table 2). For each data set, we ran six MCMC chains with different seed numbers. Each MCMC chain uses 10,000,000 proposals discarding the first 1,000,000 as burn-in and subsampling every 1,000–th state afterwards for inference.

Table 5 shows the estimates of both tMRCA and population sizes for each five subsets of data. Table 5 is compared with Table 2. tMRCA of HV1 in E and HV1 in M are almost identical to the estimates from our method. The estimates of tMRCA for HV1 in W by BEAST has mean 2.1 MYA and 95% credible region (1.4, 3.4), which is larger than the estimates with mean 1.66 MYA and 95% credible region (0.898, 2.76) from our method. The estimates of tMRCA for Numts are huge, mean 47.0 MYA and 52.0 MYA for W and E, respectively, whereas the posterior mean tMRCA for all gorillas was estimated as 5.91 MYA in our analysis.

The estimates for the effective population sizes from our method (Table 2) are between the BEAST estimates from HV1 and from Numt. For HV1 sequences, the estimates by BEAST are about the half of the corresponding estimates from our methods. In contrast, for Numt sequences, BEAST provided about 10 times larger estimates than our method. We can observe the very inconsistent estimates between HV1 sequences and Numt sequences when they are analyzed separately by BEAST. In contrast, our method can provide the estimate incorporating both Numt and HV1 sequences.

Analysis of Numt II in western and eastern lowland gorilla populations

The analysis shows there are three main Numt groups (Numt I, Numt II and Numt III) which are likely the result of independent transfer events, possibly into different places in the nuclear genome. If three Numt transfers occurred independently at different places, the population sizes must have been overestimated by ignoring the subdivision within a Numt population.

In this section, by applying BEAST separately on each Numt group in a population, we can see how the estimates of population sizes are changed by accounting for the subdivision within a population. Coalescent-based estimation of population sizes is possible when at least two sequences are available in the populations. Since few sequences are available for Numt I and Numt III in each W and E, the additional analysis within a Numt group is performed on Numt II only. Then W and E have 18 and 16 Numt sequences belonging to Numt II, respectively. In this analysis, we assume the same model setting as in the previous section.

The last two lines of Table 5 show the estimates of population sizes and tMRCA for Numt II sequences in each population E and W. The posterior estimates of tMRCAs for Numt II have mean 19 MYA with 95% credible region (12.0, 28) and mean 12 MYA with 95% credible region (6.6, 23) for Numt II from population W and E, respectively. Therefore, the huge tMRCAs of W-Numt and E-Numt in Table 5 can be thought to be caused by long coalescent time between Numt haplogroups in a gorilla population.

Also, the estimates of population sizes are reduced up to half in the inference based on Numt II. These reduced estimates of population sizes support the hypothesis that the huge estimates of population sizes of Numt groups in Table 5 may be caused by ignoring the subdivision within a population. This subdivision in a population represents Numt introgressions in multiple loci, not variation in geography.

Analysis of simulated data

We applied BEAST to the simulated data, after dividing the data according to the combinations of the sequence types and gorilla populations. There are nine available combinations: W-HV1, E–HV1, M–HV1, W–Numt, W–Numt I, W–Numt II, E–Numt, E–Numt I and E–Numt II. The analysis is not available on the subsets like W–Numt II, E–Numt I, E–Numt II, because the sequences in each set are identical. Therefore, we can apply BEAST only for the other six data sets.

In the BEAST analysis, we used the same model setting as in the previous sections. To convert BEAST estimates to be commeasurable with the estimates from our method, we used the substitution rate 0.075×10^{-6}subs/site/year and 0.0096×10^{-6}subs/site/year for HV1 and Numt sequences, respectively, which were used for simulation. For each data, we ran six MCMC chains with different seed numbers. Each MCMC chain generated 1,000,000 proposals, and we sampled every 1,000–th state after discarding the initial 100,000 proposals as burn-in. Then each MCMC chain provides 900 states, and our inference is based on 5,400 posterior samples from the six MCMC runs.

Note, our method can provide estimates by integrating both Numt and HV1 sequences over seven populations while BEAST is applicable on the subset defined by sequence types and populations. All the 95% credible regions of tMRCAs for W–HV1, E–HV1 and M–HV1 contain the true tMRCAs of each haplogroup The tMRCAs for the Numt groups are not comparable with our result. The estimates for the effective population sizes look very different from the true values as well as the estimates from our method. BEAST provides consistent results about the subdivisions in Numt sequences. The mean tMRCA of W–Numt is estimated as 5.3 MYA when the mean tMRCA of W–Numt I is 0.33 MYA. In terms of the effective population size, the estimates from W–Numt are at least two times larger than the estimates from W–Numt I. See Appendix 3 for more results.

5 Discussion

5.1 Eastern-western gorilla divergence time

Figure 7 provides the prior and posterior distributions of the east-west split times from two analyses, the Numt-HV1 case and the only-HV1 case. Figure 7(a) is the prior distribution of T_{WEM}, which is approximated by 10,000 independent samples from the prior distribution as described in section 3.1. Figure 7(b) and Figure 7(c) are the posterior distributions of T_{WEM} from the analyses of the Numt-HV1 case and the only-HV1 case, respectively. Figure 7(b) has much lower mean value with narrower range than those of Figure 7(c), although the same prior distribution was used on T_{WEM} in both cases. The 95% credible region of Figure 7(b), (0.54, 2.02) MYA, and 93.58% of posterior sample fall in the Pleistocene period, 0.11–1.8 MYA. In contrast, only 6.26% of samples in Figure 7(c) fall in the Pleistocene period, 0.11–1.8 MYA. This means the east-west split of HV1 sequences occurred before Pleistocene, while Numt sequences diverged into the eastern and western gorillas during Pleistocene.

The most important finding in this study is that the Bayesian analysis with two different data sets, the Numt-HV1 case and the only-HV1 case, provided discordant estimates for the east-west split time of gorilla populations. The estimated recent coalescent event in Numt III seems to cause the more recent east-west divergence time estimated in the Numt-HV1 case. The discordant coalescent times between Numt and HV1 sequences can be explained with different isolation times or last gene-flow times between Numt and HV1 sequences.

In population genetics, gene flow (i.e., gene migration) is the transfer of an allele from one population to another, and it is one of four evolutionary forces: natural selection, genetic drift, mutation and gene flow (Hartl 2000). Gene flow between two populations can lead to a combination of the two gene pools, reducing the genetic variation between the two groups. There are two kinds of common models for gene flow between populations. First, an isolation model assumes no gene flow between populations since the populations diverged. Therefore, each population evolves independently within each genetic pool since after the split time. However, after the isolation of populations, there

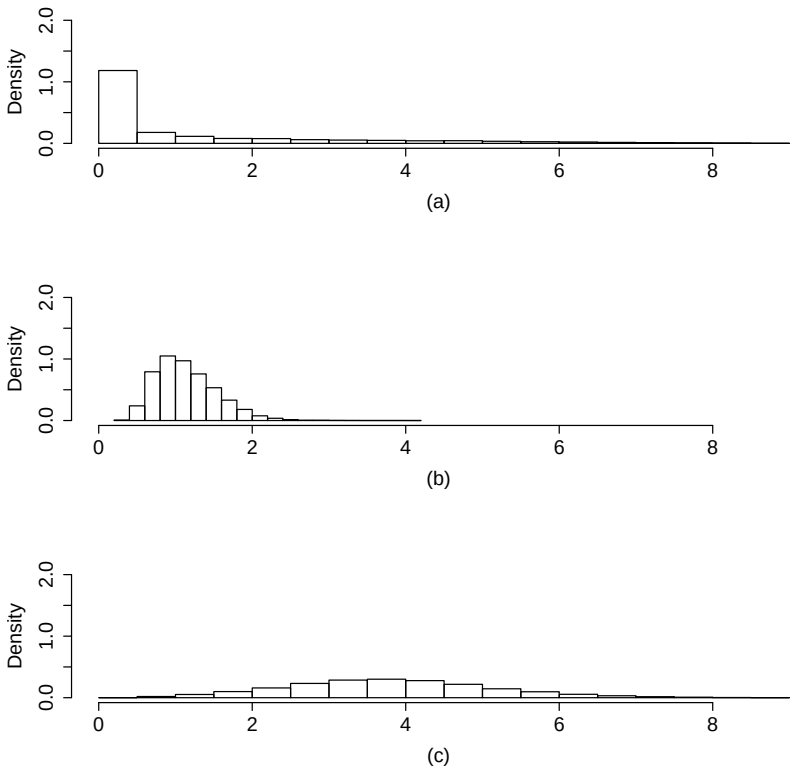

Figure 7: **Prior and posterior distributions of T_{WEM}.** (a) is the distribution of the Monte Carlo samples from the prior distribution. (b) and (c) are distributions of the posterior samples from the analysis of the Numt-HV1 case and the only-HV1 case, respectively.

are still chances to trade genes between populations by migration of individuals. A migration model is a model that allows gene-migration between populations after two populations are mostly isolated from each other.

In our model, we assumed an isolation model with seven populations and then estimated divergence times of these populations. Therefore, the population divergence time corresponds to the time of the last gene flow between two populations. In two analyses, we estimate that the last east-west gene flow for mitochondrial genes was 4.44 MYA (2.46–7.21) while nuclear genomes had some gene flow much more recently, 2.2 MYA (1.3–3.53). This discordance requires explanation, and the most plausible explanation is that male and female gorillas have very different migration behavior. It would be highly informative to confirm this conclusion by, for example, obtaining Y-chromosome sequence data from natural gorilla populations as Y-chromosomes are paternally inher-

ited.

We consider possible reasons for more recent gene-flow in male gorillas than in female gorillas. There are a number of factors that affect the rate of gene flow between different populations. One of the most significant factors is mobility, as greater mobility of an individual tends to give it greater migratory potential. Our results suggest that male gorillas may tend to migrate further from their place of birth than female gorillas do. Observation studies about gorillas also suggest that male gorillas migrate farther than females (Yamagiwa 1986; Yamagiwa and Mwanza 1994). Male gorillas begin to leave their original troop, traveling alone or with a group of other males for 2–5 years before being able to form a new group. In contrast, female gorillas just travel around their nests, up to a couple of hours distance.

A similar scenario was invoked in the study of different loci in gorillas (Thalmann et al. 2007). They investigated eastern-western divergence gorilla populations with 16 non-coding autosomal sequences. The comparison with the result of Jensen-Seaman et al. (2003) suggested the discrepancy among nuclear and mitochondrial phylogenies. The discrepancy could be partly a result of male-mediated gene flow between western and eastern gorillas (Jensen-Seaman and Kidd 2001), and this scenario has previously been invoked for other species (Tosi et al. 2000; Pidanciera et al. 2006). Such a hypothesis got additional support from some analysis of Y-chromosome variation, showing a single shared haplotype in western and eastern gorillas (Burrows and Ryder 1997; Altheide 2002). Thalmann et al. (2007) estimated a single ancestral gorilla population until 0.9–1.6 MYA followed by possibly male-mediated gene flow at around 0.77 MYA.

In the study of Thalmann et al. (2007), more than twice as much gene flow was found from eastern to western gorillas than vice versa based on 16 noncoding autosomal loci. We can infer directions of Numt migrations with the consensus tree in Figure 4 and the results in section 3.3. In Figure 4, Numt I is close to the ancestral lineage of HV1 sequences from population W with a posterior probability 1 and Numt III is close to the ancestral lineage of HV1 sequences from populations E and M. The introgression times on each ancestral lineage of Numt I and Numt III are estimated after HV1 east-west split time and before Numt east-west split time. Therefore, Numt I sequences sampled in population E are considered as offspring of a Numt sequence carried from population W to population E when HV1 sequences are isolated in populations W and E. In contrast, Numt III sequences sampled in population W are thought to be the result of gene-flow from population E to population W when HV1 sequences are isolated in each population. The gene-flow direction in Numt II sequences can be more complex since Numt II sequences are estimated as an outgroup to the rest of gorilla sequences.

5.2 Estimates of effective population sizes

Effective population size estimates of gorillas and chimpanzees are much larger than the estimate of human population size. Yu et al. (2004) found that gorillas and chimpanzees have effective population sizes at least twice as large as humans. The relatively small effective population size in humans, considering their large census size, could have been

contributed to a large expansion, possibly following a bottleneck (Harpending et al. 1998) in human history.

The larger effective population size of gorillas is likely due to their greater population subdivision. Present gorillas have a more restricted geographic range than even chimpanzees. Chimpanzees are able to live in a wider range of habitats including open woodland and savanna (Kortland 1983) and therefore may be capable of maintaining long distance gene flow between forests. However, gorilla populations are restricted to forests, therefore may be unable to allow frequent migrants with other populations across open habitats. Indeed, genetic studies have revealed that chimpanzees share mtDNA haplotypes over 900 km (Morin et al. 1994; Goldberg and Ruvolo 1997). The same has not been found for gorillas (Jensen-Seaman and Kidd 2001; Clifford et al. 2003). With the same analogy, we can consider why population W has much larger population size than other gorilla populations. The 62 HV1 sequences from population W were sampled from 17 sites over western Africa, contrary to 15 HV1 sequences from 3 sites in population E and 7 HV1 sequences from 2 sites in population M. In the model of this article, the effective population size is effectively proportional to the expected time of coalescence for two randomly chosen individuals in the populations, and population subdivision requires this time to be large.

Appendix 1: Proof of the distribution of a ratio of exponentials prior

Suppose $X_1 \sim \exp(\lambda)$, $X_2 \sim \exp(1)$ and X_1 and X_2 are independent. The joint pdf of (X_1, X_2) is $f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = \lambda e^{-\lambda x_1} e^{-x_2} = \lambda e^{-(\lambda x_1 + x_2)}$. Let $Z = \frac{X_1}{X_2}$, and $W = X_2$. Then, the Jacobian is

$$J = \begin{vmatrix} \frac{\partial z}{\partial x_1} & \frac{\partial z}{\partial x_2} \\ \frac{\partial w}{\partial x_1} & \frac{\partial w}{\partial x_2} \end{vmatrix}$$

$$= \begin{vmatrix} \frac{1}{x_2} & -\frac{x_1}{{x_2}^2} \\ 0 & 1 \end{vmatrix}$$

$$= \frac{1}{x_2} = w;$$

$$\begin{aligned} f_{Z,W}(x, w) &= f_{(X_1,X_2)}(x_1, x_2) \times w \\ &= f_{X_1}(x_1) f_{X_2}(x_2) \times w \\ &= \lambda e^{-\lambda(zw)} e^{-w} \times w \\ &= \lambda w e^{-(1+\lambda z)w}; \end{aligned}$$

$$\begin{aligned} f_Z(z) &= \int_0^\infty f_{Z,W}(z,w)dw \\ &= \int_0^\infty \lambda w e^{-(1+\lambda z)w} dw \\ &= \frac{\lambda}{(1+\lambda z)^2}; \end{aligned}$$

$$\begin{aligned} Pr(Z \leq x) &= \int_0^x \frac{\lambda}{(1+\lambda z)^2} dz \\ &= 1 - \frac{1}{1+\lambda x}. \end{aligned}$$

Appendix 2: MCMC Algorithm

This section describes an MCMC algorithm to propose a new state of ($\mathbf{T}$, G, P_t), where $\mathbf{T}$ is the population divergence times, G is the gene-genealogy and P_t is the Numt partition. The description of the MCMC algorithm is organized as follows. Section A2.1 describes two update methods to propose new population divergence times. Section A2.2 includes three ways to propose a new hybrid coalescent history in a population. Section A2.3 illustrates how to change the number of transfer events. Last, the proposal method in section A2.4 changes the total height of both the population tree and the sequence tree with a common scale factor. Table 6 at the end of Appendix 2 summarizes the acceptance rates for MCMC methods, averaging over the six MCMC runs.

To propose a new state given the current state, we pick one out of these eight (2+3+2+1=8) update methods. An update method is selected randomly with probabilities 0.06, 0.1, 0.1, 0.06, 0.16, 0.25, 0.25, 0.02, for each method, respectively, in orders. These probabilities are tuned to obtain good mixing, by giving the higher probabilities to update methods with the lower acceptance rate. Given an update method, the details are described in each following section.

Each proposal method is illustrated using the example in Figure 2.

A2.1 Update population divergence times

A population divergence time in the population tree corresponds to the depth of a population internal node. By relocating the population internal node, a new divergence time and new population edge lengths are proposed. In proposing a new divergence time, we use two distinct proposals that differ on how they handle the sequence tree embedded within the affected population edges. One proposal changes the edge lengths of affected sequence tree lineages, proportionally to the change of each population edge length. The second proposal leaves the sequence tree unchanged.

A2.1.1 Update divergence time while changing the sequence tree

This method proposes a new population divergence time. For example, suppose we propose changing T_{EM} in Figure 2. Then the new divergence time T^*_{EM} is

$$T^*_{\mathrm{EM}} = u,$$

where $u \sim$Uniform$(0, T_{\mathrm{WEM}})$. Then partial lineages of the sequence tree in populations EM, E, and M are scaled by factors u/T_{EM}, u/T_{EM} and $(T_{\mathrm{WEM}} - u)/(T_{\mathrm{WEM}} - T_{\mathrm{EM}})$, respectively. The proposal ratio is 1 because

$$\frac{1/T_{\mathrm{WEM}}}{1/T^*_{\mathrm{WEM}}} = \frac{T^*_{\mathrm{WEM}}}{T_{\mathrm{WEM}}} = \frac{T_{\mathrm{WEM}}}{T_{\mathrm{WEM}}} = 1.$$

The Jacobian (Green 1995) is

$$\begin{aligned} J &= \frac{\partial(T^*_{\mathrm{EM}}, T^*_{\mathrm{WEM}}, t^*_{E1}, t^*_{E2}, t^*_{M1}, t^*_{EM1})}{\partial(T_{\mathrm{EM}}, T_{\mathrm{WEM}}, t_{E1}, t_{E2}, t_{M1}, t_{EM1})} \\ &= \left(\frac{T^*_{\mathrm{EM}}}{T_{\mathrm{EM}}}\right)^1 \left(\frac{T^*_{\mathrm{EM}}}{T_{\mathrm{EM}}}\right)^2 \left(\frac{T_{\mathrm{WEM}} - (T^*_{\mathrm{EM}})}{T_{\mathrm{WEM}} - T_{\mathrm{EM}}}\right)^1, \end{aligned}$$

in the example since there are 1, 2 and 1 events on each population edge. The acceptance probability is

$$\begin{aligned} & \min\left\{1, \frac{f(\boldsymbol{T}^*, \Theta, G^*, P_t) \mid D)}{f(\boldsymbol{T}, \Theta, G, P_t \mid D)} \times 1 \times J\right\} \\ = \quad & \min\left\{1, \frac{f(D \mid \boldsymbol{T}^*, \Theta, G^*, P_t) f(H^*_{\mathrm{W}} \mid \boldsymbol{T}^*, \Theta, P_t) f(H^*_{\mathrm{M}} \mid \boldsymbol{T}^*, \Theta, P_t) f(H^*_{\mathrm{EM}} \mid \boldsymbol{T}^*, \Theta, P_t) f(\boldsymbol{T}^*)}{f(D \mid \boldsymbol{T}, \Theta, G, P_t) f(H_{\mathrm{W}} \mid \boldsymbol{T}, \Theta, P_t) f(H_{\mathrm{M}} \mid \boldsymbol{T}, \Theta, P_t) f(H_{\mathrm{EM}} \mid \boldsymbol{T}, \Theta, P_t) f(T)} \times J\right\}. \end{aligned}$$

In general, the Jacobian is

$$\left(\frac{V^*_1}{V_1}\right)^{k_1} \left(\frac{V^*_2}{V_2}\right)^{k_2} \left(\frac{V^*_3}{V_3}\right)^{k_3},$$

where the $\{k_i\}$ are the number of hybrid coalescent events on the three affected population edges, $\{V_i\}$ are current edge lengths of the populations and $\{V_i{}^*\}$ are the proposed edge lengths of the populations.

A2.1.2 Update divergence time without changing the sequence tree

This alternative method to propose a new population divergence time differs from the previous update method as this method does not change the sequence tree at all. This proposal simply restricts the upper limit of potential new divergence times to prevent impossible coalescent events. For example, suppose we move T_{EM} in Figure 2. Then the upper bound for T^*_{EM} is $T_{\mathrm{EM}} + s$, since the coalescent event in population EM at time $T_{\mathrm{EM}} + s$ combines two lineages from different populations. We assumed sequence lineages can coalesce only in the same or common ancestral populations. The new divergence time T^*_{EM} is

$$T^*_{\mathrm{EM}} = u,$$

where $u \sim$Uniform$(0,\, T_{\mathrm{EM}} + s)$. In this update, gene genealogy is not affected at all. However, the histories of populations E, M and EM are modified relative to the new location of T_{EM}. The proposal ratio is

$$\frac{1/T_{\mathrm{EM}} + s}{1/T_{\mathrm{EM}} + s} = \frac{T_{\mathrm{EM}} + s}{T_{\mathrm{EM}} + s} = 1.$$

The acceptance ratio is

$$\begin{aligned} & \min\left\{1, \frac{f(\boldsymbol{T}^*, \Theta^*, G^*, P_t \mid D)}{f(\boldsymbol{T}, \Theta, G, P_t \mid D)}\right\} \\ = \quad & \min\left\{1, \frac{f(H_{\mathrm{E}}^* \mid \boldsymbol{T}^*, \Theta, P_t) f(H_{\mathrm{M}}^* \mid \boldsymbol{T}^*, \Theta, P_t) f(H_{\mathrm{EM}}^* \mid \boldsymbol{T}^*, \Theta, P_t) f(\boldsymbol{T}^*)}{f(H_{\mathrm{E}} \mid \boldsymbol{T}, \Theta, P_t) f(H_{\mathrm{M}} \mid \boldsymbol{T}, \Theta, P_t) f(H_{\mathrm{EM}} \mid \boldsymbol{T}, \Theta, P_t) f(\boldsymbol{T})}\right\}. \end{aligned}$$

Since the gene genealogy is not affected by this proposal, we do not have to consider the likelihood ratio LR in the acceptance ratio.

A2.2 Update histories in a population edge

The gene genealogy, G, is determined by a hybrid coalescent process over the population tree. Therefore, a new gene genealogy G^* can be proposed by changing the hybrid coalescent history in a single population. This section describes three ways to change the history within a population: generating new interevent times, relocating an event, and switching one of a coalescent event pair.

A2.2.1 Generate new event times in a population edge

This update method proposes new event times in a population edge. The way to generate new event times is different depending on whether the population is the parent of the root or not.

1. **Root edge case**
   First, let's consider the case of the parental population of the root. This proposal method generates new event times as independent exponential random variables that match the prior distribution given the sequences of events. For example, in Figure 2, population WEM corresponds to the parent of the root, and it has four hybrid coalescent events with interevent times x_i, $i \in \{1, \cdots, 4\}$. Let $a = (\binom{2}{2}\theta_{\mathrm{HV1}} + \binom{2}{2}\theta_{\mathrm{Numt}} + \binom{1}{1}\eta)$, $b = (\binom{2}{2}\theta_{\mathrm{HV1}} + \binom{1}{1}\eta)$, $c = \binom{3}{2}\theta_{\mathrm{HV1}}$ and $d = \binom{2}{2}\theta_{\mathrm{HV1}}$. These rates are same as the rates for the hybrid coalescent processes in the population. Then new interevent times are

$$\begin{aligned} x_1^* &\sim \exp(a), \\ x_2^* &\sim \exp(b), \\ x_3^* &\sim \exp(c), \\ x_4^* &\sim \exp(d). \end{aligned}$$

The proposal ratio is

$$\frac{ae^{-ax_1}be^{-bx_2}ce^{-cx_3}de^{-dx_4}}{ae^{-ax_1^*}be^{-bx_2^*}ce^{-cx_3^*}de^{-dx_4^*}} = \frac{f(H_{\text{WEM}} \mid \boldsymbol{T}, \Theta, P_t)}{f(H^*_{\text{WEM}} \mid \boldsymbol{T}, \Theta, P_t)}.$$

Because this proposal ratio is the inverse of the prior ratio of the two population histories, the acceptance ratio is simply

$$\min\left\{1, \frac{f(D \mid \boldsymbol{T}, \Theta, G^*, P_t)}{f(D \mid \boldsymbol{T}, \Theta, G, P_t)}\right\};$$

$$\begin{aligned} R &= \min\left\{1, \frac{f(\boldsymbol{T}^*, \Theta^*, G^*, {P_t}^* \mid D)}{f(\boldsymbol{T}, \Theta, G, P_t \mid D)} \times \frac{q(\boldsymbol{T}, \Theta, G, P_t \mid \boldsymbol{T}^*, \Theta^*, G^*, P_t^*)}{q(\boldsymbol{T}^*, \Theta^*, G^*, P_t^* \mid \boldsymbol{T}, \Theta, G, P_t)}\right\} \\ &= \min\left\{1, \frac{f(D \mid \boldsymbol{T}, \Theta, G^*, P_t) f(H^*_{\text{WEM}} \mid \boldsymbol{T}, \Theta, P_t)}{f(D \mid \boldsymbol{T}, \Theta, G, P_t) f(H_{\text{WEM}} \mid \boldsymbol{T}, \Theta, P_t)} \times \frac{f(H_{\text{WEM}} \mid \boldsymbol{T}, \Theta, P_t)}{f(H^*_{\text{WEM}} \mid \boldsymbol{T}, \Theta, P_t)}\right\} \\ &= \min\left\{1, \frac{f(D \mid \boldsymbol{T}, \Theta, G^*, P_t)}{f(D \mid \boldsymbol{T}, \Theta, G, P_t)}\right\}. \end{aligned}$$

2. **Non-root case**
Now, consider the case of a non-root parent edge. For this proposal method, we keep the events in the same order and use the order statistics of an independent uniform random sample to propose new times. For example, if we generate new event times for population W in Figure 2, we generate four random numbers from Uniform $(0, T_{\text{WEM}})$, u_1, u_2, u_3 and u_4. Order them $u_{(1)} \le u_{(2)} \le u_{(3)} \le u_{(4)}$. Then the new interevent times are

$$\begin{aligned} t_1^* &= u_{(1)}, \\ t_2^* &= u_{(2)} - u_{(1)}, \\ t_3^* &= u_{(3)} - u_{(2)}, \\ t_4^* &= u_{(4)} - u_{(3)}. \end{aligned}$$

The proposal ratio is

$$\frac{\frac{1}{4}(\frac{1}{T_{WEM}})^4}{\frac{1}{4}(\frac{1}{T_{WEM}})^4} = 1.$$

The acceptance ratio is

$$\begin{aligned} &\min\left\{1, \frac{f(\boldsymbol{T}^*, \Theta, G^*, P_t \mid D)}{f(\boldsymbol{T}, \Theta, G, P_t \mid D)}\right\} \\ = &\min\left\{1, \frac{f(D \mid \boldsymbol{T}, \Theta, G^*, P_t) f(H^*_{\text{W}} \mid \boldsymbol{T}, \Theta, P_t)}{f(D \mid \boldsymbol{T}, \Theta, G, P_t) f(H_{\text{W}} \mid \boldsymbol{T}, \Theta, P_t)}\right\}. \end{aligned}$$

A2.2.2 Relocate an event in a population edge

This proposal changes a single event time by selecting a new time uniformly from its

possible range, conditional on the times of other events on the edge. Events cannot leave the edge or occur out of order with related events. For example, there are four events in population W of Figure 2. The HV1 coalescent event at time t_1 can occur between times 0 and $t_1 + t_2 + t_3 + t_4$, but not at a later time since this coalescent event must precede the coalescent event at time $t_1 + t_2 + t_3 + t_4$. Similarly, the Numt coalescent at time $t_1 + t_2$ and the Numt transfer event at time $t_1 + t_2 + t_3$ and HV1 coalescent event at time $t_1 + t_2 + t_3 + t_4$ can be relocated in the intervals $(0,\ t_1 + t_2 + t_3)$ and $(t_1 + t_2,\ t_1 + t_2 + t_3 + t_4)$, and $(t_1 + t_2 + t_3,\ T_{\mathrm{WEM}})$, respectively.

In this example, one of the four events would be selected uniformly at random. Conditional on being selected, the proposed event would be one of following:

$$\begin{aligned}
\text{event at time } t_1 &\sim \text{Uniform}(0, \sum_1^4 t_i),\\
\text{event at time } \sum_1^2 t_i &\sim \text{Uniform}(0, \sum_1^3 t_i),\\
\text{event at time } \sum_1^3 t_i &\sim \text{Uniform}(\sum_1^2 t_i, \sum_1^4 t_i),\\
\text{event at time } \sum_1^4 t_i &\sim \text{Uniform}(\sum_1^3 t_i, \sum_1^5 t_i).
\end{aligned}$$

The proposal ratio is 1 for any case, therefore the acceptance probability is

$$\begin{aligned}
& \min\left\{1, \frac{f(\boldsymbol{T}^*, \Theta, G^*, P_t \mid D)}{f(\boldsymbol{T}, \Theta, G, P_t \mid D)}\right\}\\
= \ & \min\left\{1, \frac{f(D \mid \boldsymbol{T}, \Theta, G^*, P_t) f(H_{\mathrm{W}}^* \mid \boldsymbol{T}, \Theta, P_t)}{f(D \mid \boldsymbol{T}, \Theta, G, P_t) f(H_{\mathrm{W}} \mid \boldsymbol{T}, \Theta, P_t)}\right\}.
\end{aligned}$$

A2.2.3 Modify the sequence topology

This update prunes a subtree of gene genealogy and then regrafts it back on a feasible branch at the same time within the same population. In population W of Figure 2, if we want to move either W_3 or W_4, W_5 is the only feasible branch in the population because it is the unique Numt sequence at time $(t_1 + t_2)$ other than W_3 and W_4. If W_4 moves, then the new state shown in Figure 8 is obtained given the old state. This update changes the topology of the gene tree without modification of event order and interevent times. The proposal ratio is 1 because the number of feasible branches is the same in both directions. Note the sequence lineages affected by this proposal method may change their Numt groups. Therefore the state for the Numt partitions is also newly proposed. For example, given the state in Figure 2, five Numt taxa are partitioned into two Numt groups as $(W_3,\ W_4)$ and $(W_5,\ E_1,\ E_2)$ in each (i.e, $P_t(5 \mid 2) = (2, 3)$). However, if the new state in Figure 8 is newly proposed for population W in Figure 2, the Numt taxa are partitioned into two Numt groups as (W_3), $(W_4,\ W_5,\ E_1,\ E_2)$ in

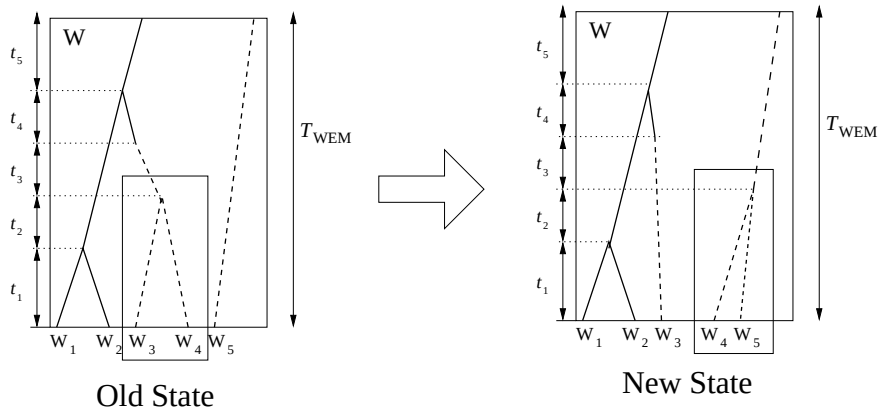

Figure 8: **An example of pruning-regrafting algorithm.** The internal node at time $(t_1 + t_2)$ is made by the coalescent pair (W_3, W_4). The Pruning-regrafting algorithm changes the coalescent pair into (W_4, W_5).

each, and this corresponds to the Numt partition $P_t(5 \mid 2)^* = (1, 4)$. Sequentially, the histories in the descendent edges of the population under consideration are affected by this method. Therefore, in general, the acceptance ratio is

$$\begin{aligned} & \min\left\{1, \frac{f(\boldsymbol{T}^*, \Theta, G^*, P_t^* \mid D)}{f(\boldsymbol{T}, \Theta, G, P_t \mid D)}\right\} \\ = \quad & \min\left\{1, \frac{f(D \mid \boldsymbol{T}^*, \Theta, G^*, P_t^*)\{\prod_i f(H_i^* \mid \boldsymbol{T}, \Theta, P_t^*)\} f(P_t^*)}{f(D \mid \boldsymbol{T}, \Theta, G, P_t)\{\prod_i f(H_i \mid \boldsymbol{T}, \Theta, P_t)\} f(P_t)}\right\}. \end{aligned}$$

A2.3 Update the number of transfer events

In this section, we explain two proposal methods that change the number of transfer events by one.

A2.3.1 Increase the number of transfer events by one This proposal eliminates a transfer event from an internal edge of the sequence tree and creates two new transfer events, one on each child edge. For this proposal, we have to pick up a Numt edge, among which are the oldest ancestral edges from each Numt group and also which are non-leaves, uniformly at random from the whole sequence tree, with probability $\frac{1}{x}$, where x denotes the number of Numt groups having at least two taxa. If there is no candidate, the proposal is rejected. Figure 9 describes the change in population W of Figure 2 by this proposal method which increases the number of transfer events by one. For example, population W in Figure 2(a) has a transfer event at time $(\sum_1^3 t_i)$. In Figure 9(b), after removing this transfer event, two new events are created on each child edge at time ${t_1}^*$ and at time ${t_1}^* + {t_2}^*$, respectively. The locations of the new transfer events are determined uniformly at random on each child edge and the event type of the Numt coalescent event at time $(\sum_1^2 t_i)$ in Figure 9(a) becomes a HV1 coalescent event.

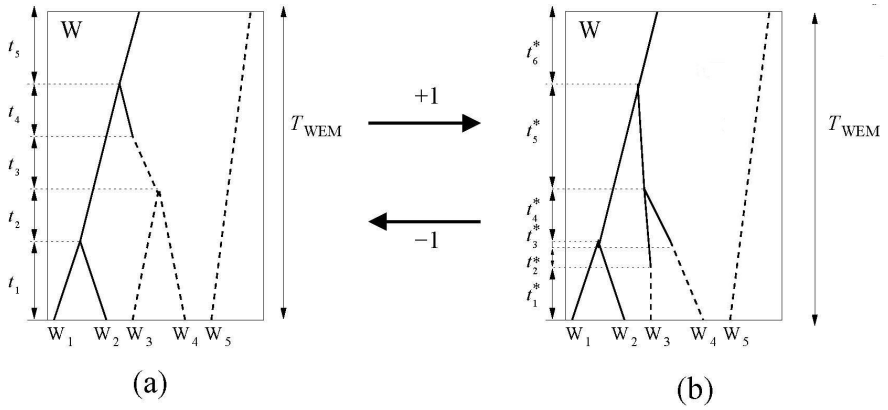

Figure 9: **An example to change the total number of transfer events by one.** (a) → (b): the number of transfer events is increased by one by creating two new transfer events on each W_3 and W_4, at time t_1^* and $t_1^*+t_2^*$, respectively, after deleting the transfer event at $(t_1 + t_2)$ in (a). (b) → (a): the number of transfer events is decreased by one by creating a new transfer event at time $(t_1 + t_2 + t_3)$ after deleting the transfer events on each child edge, at time t_1^* and $t_1^* + t_2^*$, respectively in (b).

The proposal probability for the transition from (a) into (b) is $\frac{1}{x}\frac{1}{t_1+t_2}\frac{1}{t_1+t_2}$, where $x = 2$ given the state of Figure 2.

Since this proposal method always increases the number of transfer events by one, we need to introduce the next proposal method to decrease the number of transfer events before discussing the acceptance probability.

A2.3.2 Decrease the number of transfer events by one

The new proposal method decreases the number of transfer events by one. Pick an HV1 coalescent event such that both children edges of the HV1 coalescent node contain transfer events on each, uniformly at random from the whole sequence tree. If there is no case satisfying the condition, the proposal is rejected. This proposal method removes these original transfer events from the children edges and puts a new transfer event on their ancestral edge. Then the HV1 coalescent event becomes a Numt coalescent event. Also two Numt groups are merged as a single group. The location of the new transfer event is chosen uniformly at random on the parent edge of the selected coalescent node. The change from (b) into (a) in Figure 9 is an example showing the application of this proposal. Let y be the number of the HV1 coalescent nodes whose children have a transfer event on each. Given the state in Figure 9 (b) as the current state of population W in Figure 2, we have $y = 1$, and the proposal probability from (b) into (a) is $\frac{1}{y}\frac{1}{t_3+t_4}$.

Note it is obvious that the transition in the number of transfer events leads to the new Numt partition. Therefore, these two methods can affect the hybrid coalescent process histories of other population edges as well. For example, if Figure 9 (a) is

the current state, two Numt groups exist as (W_3, W_4) and (W_5, E_1, E_2) in each (i.e, $P_t(5 \mid 2) = (2, 3)$). In contrast, if Figure 9 (b) is newly proposed, there are three Numt groups as (W_3), (W_4) and (W_5, E_1, E_2) with $P_t(5 \mid 3) = (1, 1, 3)$. Unlike the previous two update methods for changing the event history, this proposal can affect the histories of other population edges with the new Numt partitions.

To calculate the proposal ratio with these two proposal methods, recall the following definition:

- x=the number of Numt lineages which are the oldest ancestors in each Numt group and which are non-leaves at the current state.
- y= the number of HV1 coalescent events whose children have transfer events on each edge at the current state.

In Figure 2, we obtain $x = 2$ and $y = 0$. Then the proposal ratio for the transition from (a) to (b) in the example of Figure 9 is

$$PR = \frac{\frac{1}{y+1}\frac{1}{t_3+t_4}}{\frac{1}{x}\frac{1}{t_1+t_2}\frac{1}{t_1+t_2}} = \frac{(t_1+t_2)^2}{t_3+t_4}\frac{x}{y+1}. \tag{1}$$

Note, if the state (b) is accepted as a new state, the number of HV1 coalescent events whose children have a transfer event on each edge increased to $y + 1$ from y.

In contrast, suppose Figure 9 (b) is the current state of population W in Figure 2. Then, we obtain $x = 1$ and $y = 1$. The proposal ratio for the transition from (b) into (a) in Figure 9 is

$$PR = \frac{\frac{1}{x+1}\frac{1}{t_1+t_2}\frac{1}{t_1+t_2}}{\frac{1}{y}\frac{1}{t_3+t_4}} = \frac{t_3+t_4}{(t_1+t_2)^2}\frac{y}{x+1}.$$

The acceptance ratio is

$$\begin{aligned} & \min\left\{1, \frac{f(\boldsymbol{T}, \Theta, G^*, P_t^* \mid D)}{f(\boldsymbol{T}, \Theta, G, P_t \mid D)} \times PR\right\} \\ & \min\left\{1, \frac{f(D \mid \boldsymbol{T}, \Theta, G^*, P_t^*) f(G^* \mid \boldsymbol{T}, \Theta, P_t^*) f(\boldsymbol{T}) f(\Theta) f(P_t^*)}{f(D \mid \boldsymbol{T}, \Theta, G, P_t) f(G \mid \boldsymbol{T}, \Theta, P_t) f(\boldsymbol{T}) f(\Theta) f(P_t)} \times PR\right\} \\ = \quad & \min\left\{1, \frac{f(D \mid \boldsymbol{T}, \Theta, G^*, P_t^*)\{\prod_i f(H_i^* \mid \boldsymbol{T}, \Theta, P_t^*)\} f(P_t^*)}{f(D \mid \boldsymbol{T}, \Theta, G, P_t)\{\prod_i f(H_i \mid \boldsymbol{T}, \Theta, P_t)\} f(P_t)} \times PR\right\}. \end{aligned}$$

A2.4 Rescale both trees

This proposal method multiplies all edges in both the population and sequence trees by a common factor $m = \mathrm{e}^{-\lambda(U-0.5)}$, where $U \sim$Uniform(0, 1) and λ is a tuning parameter.

The reverse proposal requires $U^* = 1 - U$, so the proposal ratio is

$$\mathrm{e}^{-\lambda(-(U-0.5))} = \frac{m}{(1/m)} = m^2 . \frac{1/1}{1/1} = 1.$$

The Jacobian is

$$J = \frac{\partial(T^*_{\mathrm{EM}}, T^*_{\mathrm{WEM}}, T^*_{\mathrm{Chimp1}}, T^*_{\mathrm{Chimp2}}, T^*_{\mathrm{HC}}, T^*_{\mathrm{GHC}}, t_{ij}(i = 1, \cdots, 13, j = 1, \cdots, I_i), U^*)}{\partial(T_{\mathrm{EM}}, T_{\mathrm{WEM}}, T_{\mathrm{Chimp1}}, T_{\mathrm{Chimp2}}, T_{\mathrm{HC}}, T_{\mathrm{GHC}}, t_{ij}(i = 1, \cdots, 13, j = 1, \cdots, I_i), U)}$$

$$= \begin{vmatrix} \ddots & & & & & 0 \\ & \cdot & & & & 0 \\ & & \mathrm{e}^{-\lambda(U-0.5)} & & & 0 \\ & & & \cdot & & 0 \\ & & & & \ddots & 0 \\ \cdots & -\lambda T \mathrm{e}^{-\lambda(U-0.5)} & \cdots & -\lambda t_{ij} \mathrm{e}^{-\lambda(U-0.5)} & \cdots & -1 \end{vmatrix}$$

$$= m^{\#\text{ of divergence times} + \sum_i I_i},$$

where I_i is the number of events in population edge i. Then the acceptance ratio is

$$R = \min\left\{1, \frac{f(\boldsymbol{T}^*, \Theta, G^*, P_t \mid D)}{f(\boldsymbol{T}, \Theta, G, P_t \mid D)} \times J\right\}.$$

Appendix 3: Other results

This section provides supplementary information. First, Table 7, Table 8, Table 9 and Table 10 provide the estimates from each MCMC chain. For each parameter in an MCMC chain, there are two values: the upper one is the mean and the lower one is the standard deviation for the posterior samples from each chain. Table 7 and Table 8 are from the analysis of the only-HV1 case, and Table 9 and Table 10 are for the Numt-HV1 case. The estimates of each parameter from the six MCMC chains are consistent with one another in both the only-HV1 case and the Numt-HV1 case.

Next, Table 11, Table 12, Table 13 and Figure 10 are the results obtained from the analysis of the simulation data described in section 3.9. The data were generated based on the values of the parameters listed in the second column of Table 11. All the results were obtained by summarizing 54,000 posterior samples from the six MCMC chains when our method was applied to the simulated data. We checked whether the true value of each parameter fell into the 95% credible region of the posterior sample, provided by our method (Table 11). Also we checked that the estimated consensus tree is identical with the original gene-genealogy (Figure 10). The haplogroups are well defined in the consensus tree with high posterior probabilities (Table 12). Table 13 provides the results from the analysis of the simulated data with BEAST. The table contains the posterior mean, median and 95% credible region for the tMRCA and the effective population size from each data set, the subset of the simulated data.

Parameters	num of accepts	num of visits	Accept Rate
A.2.1.1	1744246	7195799	0.242
A.2.1.2	4433969	12001947	0.369
A.2.2.1	1949604	12001416	0.162
A.2.2.2	2896288	7203681	0.402
A.2.2.3	1706571	19187309	0.0889
A.2.3.1	675822	30013374	0.0225
A.2.3.2	675771	25407628	0.0266
A.2.4	2166486	2402527	0.902
N_1	94820	461924	0.205
N_2	164397	461308	0.356
N_3	241411	461058	0.524
N_4	205822	461802	0.446
N_5	318911	461584	0.691
N_6	352970	461526	0.765
N_7	310379	462486	0.671
N_8	269163	461461	0.583
N_9	132724	461501	0.288
N_{10}	297913	461656	0.645
N_{11}	311625	460676	0.676
N_{12}	324386	460774	0.704
N_{13}	308264	460850	0.669
μ_{HV1}	142118	1000468	0.142
μ_{Numt}	234959	1000901	0.235
κ_{HV1}	262177	1001442	0.262
κ_{Numt}	443582	999754	0.444
λ_θ	455072	999612	0.455
η	464643	999217	0.465

Table 6: **The acceptance rate of each MCMC update method.** The first eight rows are from the MCMC methods described from A.2.1 to A.2.4 in order of appearance. The remaining rows are acceptance rates for numerical parameters in $\Theta = \{\theta_1, \cdots, \theta_{13}, \lambda_\theta, \eta, \mu_{\mathrm{HV1}}, \mu_{\mathrm{Numt}}, \kappa_{\mathrm{HV1}}, \kappa_{\mathrm{Numt}}\}$.

Finally, Figure 11 and Figure 12 show the graphical results, in regard to the sensitivity analysis of the prior distributions mentioned in section 4.1. Figure 11 is for the divergence times of gorillas and Figure 12 is for the divergence times of chimpanzees.

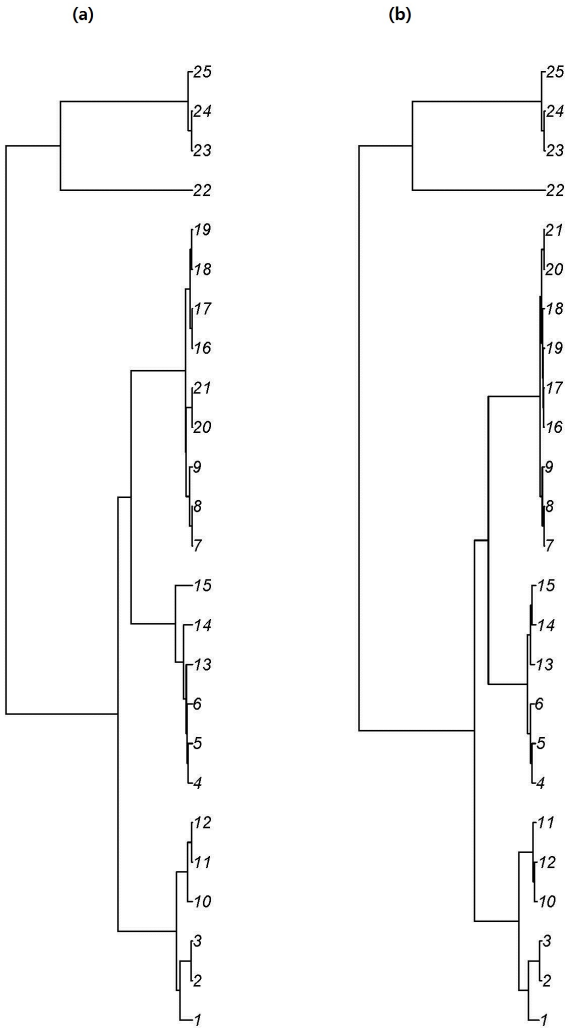

Figure 10: **Comparison of the true topology and the consensus topology from MCMC analysis.** (a) is the true gene-genealogy generating the sequence data. (b) is the consensus tree summarized from 5400 posterior sampled trees. Population W has Numt I (1, 2, 3), Numt II (4, 5, 6) and HV1 (7, 8, 9) sequences. Population E has Numt I (10, 11, 12), Numt II (13, 14, 15), HV1 (16, 17, 18) sequences. Population M has HV1 (19, 20, 21) sequences. Human and Chimpanzee populations have a single HV1 sequence labeled as 22, 23, 24 and 25, respectively.

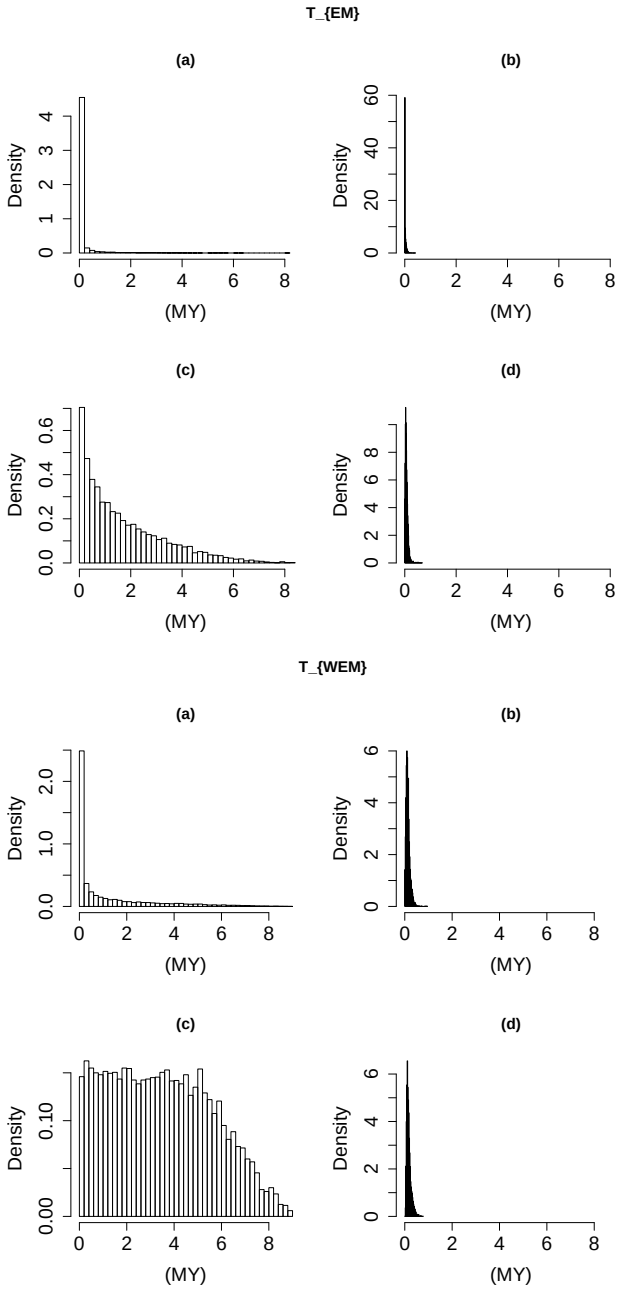

Figure 11: **Comparison of the distributions of prior and posterior samples for gorilla divergence times.** In each panel, (a) and (b) are for the Monte Carlo samples and the posterior samples from the original prior distribution, and (c) and (d) are for the Monte Carlo samples and the posterior samples from the new prior distribution.

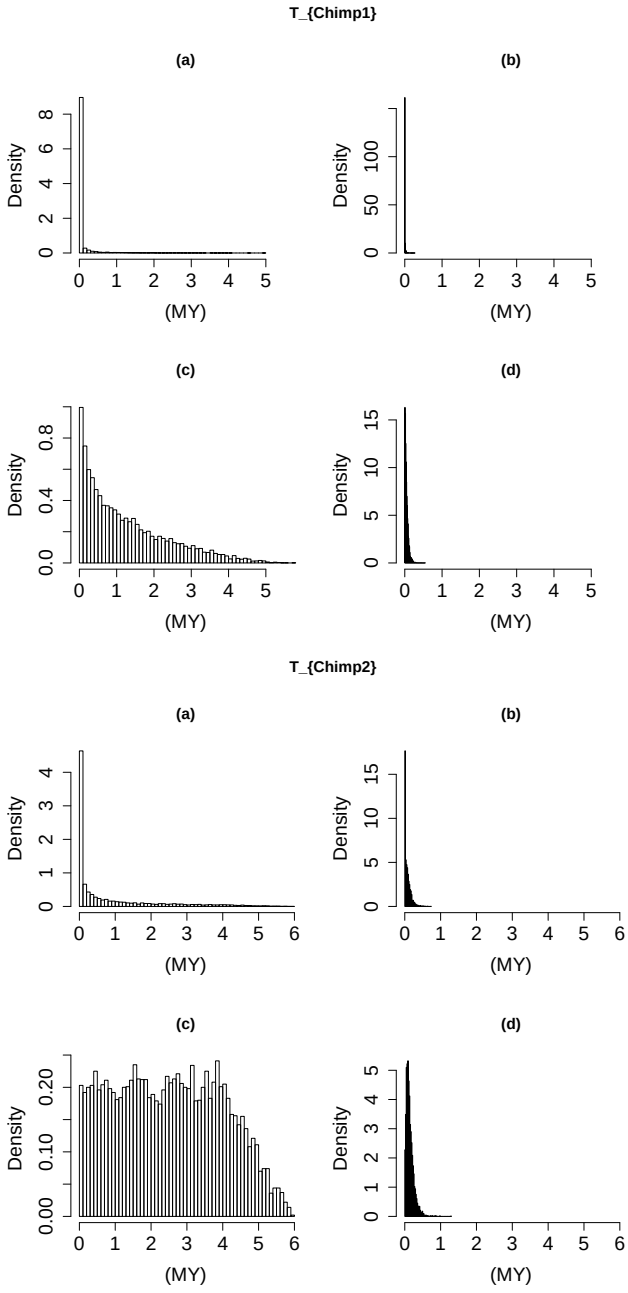

Figure 12: **Comparison of the distributions of prior and posterior samples for chimpanzee divergence times.** In each panel, (a) and (b) are for the Monte Carlo samples and the posterior samples from the original prior distribution, and (c) and (d) are for the Monte Carlo samples and the posterior samples from the new prior distribution.

Parameters	Chain 1	Chain 2	Chain 3	Chain 4	Chain 5	Chain 6
T_{EM}	0.688	0.664	0.67	0.671	0.688	0.686
	0.365	0.36	0.354	0.363	0.369	0.355
T_{WEM}	3.88	3.77	3.79	3.77	3.77	3.85
	1.36	1.34	1.32	1.3	1.33	1.32
T_{chimp1}	0.548	0.55	0.451	0.512	0.58	0.617
	0.541	0.538	0.539	0.542	0.553	0.555
T_{chimp2}	1.54	1.51	1.39	1.52	1.55	1.58
	0.561	0.581	0.674	0.583	0.587	0.549
T_{HC}	4.93	4.9	4.92	4.93	4.92	4.93
	0.565	0.562	0.568	0.567	0.567	0.568
T_{GHC}	7.61	7.58	7.62	7.62	7.61	7.6
	1	1.01	1	0.999	1	1
N_1 (W)	0.923	0.965	0.922	0.912	0.915	0.889
	0.289	0.314	0.262	0.304	0.29	0.28
N_2 (E)	5.71	5.79	5.6	5.63	5.63	5.5
	3.09	3.37	2.92	3.14	3.09	3.12
N_3 (M)	12.3	13	13.3	12.9	12.8	11.4
	9.41	10.6	10.9	11.4	10.5	8.47
N_4 (H)	1.53	1.6	1.58	1.6	1.55	1.51
	0.728	0.759	0.732	0.828	0.762	0.712
N_5 ($\mathrm{W_c}$)	3.81	10.8	6.69	5.94	7.02	4.98
	21.7	251	34.4	47.6	79.7	42.6
N_6 (N)	444	17.4	21	24.8	15.4	21
	16300	94.1	218	234	77.3	193
N_7 (C)	1.82	1.91	2.2	2.03	1.73	1.72
	1.98	3.02	5.74	9.5	2.65	1.57
N_8 (EM)	25.3	22.1	29.2	17.8	23.5	26.6
	158	191	218	89.2	285	181
N_9 (WEM)	19.9	18.7	17.8	37.3	20.3	27.6
	196	124	190	497	189	256
N_{10} (Chimp1)	21.9	13.4	36.1	9.25	8.44	8.25
	206	149	633	91.7	49.6	42.4
N_{11} (Chimp2)	69.4	33.9	18.3	23.2	27.9	17.3
	1100	307	142	169	241	59.4
N_{12} (HC)	23	52.3	40.6	37.3	35.2	29.2
	145	483	519	291	419	219
N_{13} (GHC)	19.7	20.9	16.4	45.4	11	20.9
	212	160	206	929	59	182
μ_{HV1}	0.0463	0.0481	0.0473	0.0467	0.0463	0.0454
	0.0128	0.0135	0.0122	0.0135	0.013	0.0125
μ_{Numt}	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA
κ_{HV1}	21.5	21.6	21.6	21.4	21.5	21.2
	4.51	4.53	4.46	4.43	4.47	4.43
κ_{Numt}	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA
λ_θ	0.429	0.411	0.43	0.42	0.424	0.431
	0.189	0.188	0.185	0.191	0.187	0.189
η	NA	NA	NA	NA	NA	NA
	NA	NA	NA	NA	NA	NA

Table 7: **The estimates of the parameters from each of six MCMC chains in the analysis of the only-HV1 case.**

References

Altheide, T. (2002). "Comparative population genetics of the Hominoidea: an investigation of locus-specific and genome wide influence." Ph.D. thesis, Department of

Parameters	Chain 1	Chain 2	Chain 3	Chain 4	Chain 5	Chain 6
A	0.153	0.152	0.146	0.151	0.152	0.161
	0.0834	0.0887	0.0812	0.0831	0.0843	0.0868
B	0.19	0.182	0.184	0.189	0.189	0.194
	0.0901	0.0843	0.0817	0.0851	0.0867	0.0888
C1	0.272	0.25	0.27	0.268	0.266	0.279
	0.104	0.0958	0.104	0.102	0.104	0.102
C2	1.03	0.976	0.997	1	1.02	1.03
	0.33	0.326	0.31	0.322	0.34	0.321
D1	0.377	0.349	0.357	0.386	0.367	0.384
	0.167	0.163	0.153	0.17	0.174	0.171
D2	0.174	0.175	0.161	0.158	0.179	0.165
	0.0696	0.0785	0.0662	0.068	0.079	0.074
D3	0.147	0.152	0.153	0.17	0.155	0.161
	0.0678	0.0645	0.0665	0.0816	0.0701	0.0698
A, B	0.955	0.924	0.927	0.938	0.951	0.959
	0.333	0.338	0.316	0.337	0.341	0.323
C	1.03	0.976	0.997	1	1.02	1.03
	0.33	0.326	0.31	0.322	0.34	0.321
D1, D2	0.56	0.549	0.538	0.548	0.576	0.56
	0.201	0.208	0.199	0.191	0.218	0.195
D	0.586	0.583	0.571	0.585	0.601	0.589
	0.2	0.204	0.195	0.195	0.215	0.191
C, D	1.69	1.64	1.65	1.65	1.69	1.68
	0.479	0.489	0.467	0.48	0.512	0.465
All Gorilla	4.51	4.41	4.44	4.4	4.45	4.49
	1.21	1.24	1.19	1.22	1.24	1.21
Human	0.505	0.489	0.496	0.509	0.498	0.521
	0.184	0.186	0.177	0.193	0.183	0.189
P.t.Verus	0.501	0.489	0.489	0.498	0.504	0.506
	0.186	0.184	0.175	0.186	0.187	0.185
P.t.Troglodytes	0.801	0.78	0.789	0.798	0.804	0.812
	0.281	0.282	0.277	0.29	0.29	0.278
P.t.Verus, *P.t.Vellerosus*	1.29	1.25	1.26	1.27	1.29	1.31
	0.395	0.392	0.381	0.395	0.403	0.386
All Chimp	1.84	1.8	1.81	1.83	1.85	1.86
	0.492	0.503	0.479	0.501	0.515	0.493
Human, Chimp	5.27	5.22	5.27	5.27	5.29	5.29
	0.77	0.778	0.796	0.773	0.816	0.788
Gorila, Human, Chimp	8.74	8.62	8.67	8.68	8.73	8.78
	1.66	1.69	1.63	1.68	1.75	1.72

Table 8: **The estimates of tMRCAs from each of six MCMC chains in the analysis of the only-HV1 case.**

Parameters	Chain 1	Chain 2	Chain 3	Chain 4	Chain 5	Chain 6
T_{EM}	0.515	0.55	0.46	0.468	0.483	0.435
	0.337	0.309	0.277	0.319	0.308	0.268
T_{WEM}	1.31	1.01	1.15	1.29	1.06	1.01
	0.393	0.349	0.372	0.391	0.363	0.398
T_{chimp1}	0.477	0.386	0.463	0.51	0.513	0.534
	0.505	0.512	0.488	0.493	0.521	0.516
T_{chimp2}	1.46	1.48	1.39	1.42	1.47	1.5
	0.528	0.605	0.537	0.503	0.538	0.559
T_{HC}	4.76	4.8	4.75	4.74	4.76	4.81
	0.545	0.552	0.536	0.537	0.548	0.554
T_{GHC}	8	7.94	7.95	8.03	7.95	7.94
	0.825	0.867	0.862	0.812	0.86	0.871
N_1 (W)	0.84	0.83	0.865	0.879	0.823	0.825
	0.224	0.244	0.238	0.23	0.217	0.229
N_2 (E)	2.91	2.92	2.98	3.13	3.04	2.93
	1.32	1.31	1.31	1.46	1.42	1.29
N_3 (M)	12.3	12	13.4	12.8	12.3	12.6
	9.65	9.12	10.7	10	10.4	10.2
N_4 (H)	1.48	1.46	1.54	1.57	1.47	1.47
	0.671	0.676	0.695	0.707	0.654	0.664
N_5 ($\mathrm{W_c}$)	4.49	4.63	6.9	3.18	4.04	4.45
	37.5	19	114	11.5	25.9	37
N_6 (N)	18.6	320	14.7	25.9	14.2	14.6
	249	14300	146	334	88.4	75.2
N_7 (C)	1.74	1.81	1.93	1.71	1.6	1.59
	3.91	5.81	6.08	2.32	1.62	1.5
N_8 (EM)	2.12	26.9	6.12	1.97	3.14	7.09
	5.08	147	15.8	3.83	5.88	42.5
N_9 (WEM)	1.99	1.79	1.86	2.03	1.93	1.7
	0.655	0.658	0.62	0.707	0.67	0.546
N_{10} (Chimp1)	9.29	4.63	6.66	7.45	10.2	20.9
	118	25.5	44.9	44.3	136	442
N_{11} (Chimp2)	41.7	28.5	14.5	41.9	26.2	21.7
	414	241	79.8	355	258	209
N_{12} (HC)	27.6	25.6	25.1	19.2	63.2	27.1
	225	156	128	117	690	196
N_{13} (GHC)	5.83	7.5	5.34	41.4	7.47	4.89
	36.4	58	34.8	1400	59.9	20.2
μ_{HV1}	0.0394	0.0384	0.0404	0.0406	0.0392	0.0386
	0.0085	0.0089	0.0084	0.00855	0.00873	0.0081
μ_{Numt}	0.00275	0.00355	0.00297	0.00229	0.0036	0.00329
	0.000879	0.00121	0.00113	0.000797	0.00122	0.00106
κ_{HV1}	19.8	19.8	19.6	19.4	20	19.8
	3.57	3.63	3.49	3.4	3.6	3.52
κ_{Numt}	8.81	9.12	9	8.86	9.08	9.07
	3.29	3.34	3.41	3.56	3.25	4.13
λ_θ	0.502	0.499	0.494	0.494	0.501	0.514
	0.184	0.189	0.184	0.183	0.188	0.186
η	0.622	0.561	0.617	0.529	0.632	0.609
	0.373	0.332	0.322	0.225	0.36	0.336

Table 9: **The estimates of the parameters from each of six MCMC chains in the analysis of the Numt-HV1 case.**

Parameters	Chain 1	Chain 2	Chain 3	Chain 4	Chain 5	Chain 6
A	0.15	0.156	0.142	0.147	0.152	0.148
	0.078	0.079	0.0741	0.0778	0.0828	0.0769
B	0.23	0.238	0.219	0.217	0.221	0.226
	0.0928	0.1	0.0851	0.0871	0.0923	0.0966
C1	0.268	0.295	0.298	0.283	0.293	0.281
	0.081	0.107	0.0923	0.0925	0.11	0.101
C2	0.936	1.03	0.926	0.945	0.961	0.991
	0.252	0.295	0.248	0.261	0.256	0.298
D1	0.346	0.367	0.35	0.334	0.352	0.364
	0.135	0.149	0.141	0.12	0.163	0.147
D2	0.152	0.154	0.143	0.152	0.151	0.187
	0.0631	0.0661	0.0581	0.0535	0.0562	0.0696
D3	0.172	0.182	0.15	0.17	0.189	0.169
	0.0655	0.0744	0.0621	0.0613	0.0725	0.0846
Numt I	1.99	2.66	2.02	2.05	2.21	2.26
	0.459	0.608	0.486	0.456	0.562	0.598
Numt IIA	1.99	2.66	2.02	2.05	2.21	2.26
	0.459	0.608	0.486	0.456	0.562	0.598
Numt IIB	1.99	2.66	2.02	2.05	2.21	2.26
	0.459	0.608	0.486	0.456	0.562	0.598
Numt IIC	1.99	2.66	2.02	2.05	2.21	2.26
	0.459	0.608	0.486	0.456	0.562	0.598
Numt II (IIA, IIB, IIC)	1.99	2.66	2.02	2.05	2.21	2.26
	0.459	0.608	0.486	0.456	0.562	0.598
Numt III	1.99	2.66	2.02	2.05	2.21	2.26
	0.459	0.608	0.486	0.456	0.562	0.598
A, B	1.04	0.949	0.935	0.97	0.924	0.924
	0.309	0.265	0.271	0.287	0.313	0.289
C	0.936	1.03	0.926	0.945	0.961	0.991
	0.252	0.295	0.248	0.261	0.256	0.298
D1, D2	0.466	0.479	0.458	0.436	0.477	0.504
	0.137	0.152	0.139	0.133	0.156	0.168
D	0.537	0.556	0.521	0.504	0.551	0.571
	0.14	0.169	0.149	0.133	0.167	0.168
C, D	1.31	1.51	1.3	1.32	1.37	1.4
	0.31	0.406	0.298	0.312	0.331	0.373
Numt I, C, D	1.99	2.66	2.02	2.05	2.21	2.26
	0.459	0.608	0.486	0.456	0.562	0.598
Numt III, A, B	1.99	2.66	2.02	2.05	2.21	2.26
	0.459	0.608	0.486	0.456	0.562	0.598
Numt IIB, Numt IIC	1.99	2.66	2.02	2.05	2.21	2.26
	0.459	0.608	0.486	0.456	0.562	0.598
A, B, C, D, NumtI, Numt III	1.99	2.66	2.02	2.05	2.21	2.26
	0.459	0.608	0.486	0.456	0.562	0.598
All Gorilla	6.11	5.74	5.84	6.08	5.82	5.88
	1.17	1.18	1.12	1.14	1.21	1.24
Human	0.531	0.538	0.49	0.495	0.524	0.511
	0.177	0.176	0.165	0.163	0.178	0.176
P.t. verus	0.501	0.515	0.485	0.484	0.504	0.515
	0.17	0.178	0.162	0.163	0.173	0.18
P.t. troglodytes	0.828	0.854	0.805	0.802	0.831	0.846
	0.27	0.292	0.256	0.258	0.28	0.288
P.t. troglodytes, *P.t. vellerosus*	1.25	1.3	1.22	1.21	1.26	1.28
	0.34	0.376	0.328	0.332	0.356	0.362
All Chimp	1.79	1.85	1.75	1.73	1.8	1.83
	0.43	0.476	0.406	0.411	0.446	0.467
Human, Chimp	5.01	5.1	5	4.99	5.04	5.07
	0.673	0.738	0.646	0.644	0.71	0.709
Gorila, Human, Chimp	6.1	5.69	5.83	6.08	5.8	5.82
	1.17	1.18	1.12	1.14	1.21	1.23

Table 10: **The estimates of tMRCAs from each of six MCMC chains in the analysis of the Numt-HV1 case.**

Parameters	true values	posterior median	95% C.R.
	Population divergence time		
T_{EM}	3.6e-05	0.0042	(1.1e-07, 0.12)
T_{WEM}	0.206	0.12	(0.017, 0.35)
T_{chimp1}	2.5e-05	3.7e-05	(5.8e-11, 0.052)
T_{chimp2}	0.1554	0.048	(1e-05, 0.31)
T_{HC}	5	5.1	(4.1, 6)
T_{GHC}	7	7.3	(5.2, 8.9)
	Effective population size		
N_1	10	14	(1.3, 95)
N_2	10	11	(0.24, 380)
N_3	10	9.7	(0.35, 320)
N_4	10	8.3	(0.24, 320)
N_5	10	7.5	(0.084, 360)
N_6	10	7.4	(0.21, 270)
N_7	10	7.6	(0.097, 290)
N_8	10	11	(0.76, 100)
N_9	10	10	(2.5, 46)
N_{10}	10	12	(0.22, 350)
N_{11}	10	14	(1.5, 240)
N_{12}	10	8.8	(0.19, 870)
N_{13}	10	7.7	(0.25, 460)
	tMRCA		
West HV1	0.112	0.081	(0.018, 0.27)
East HV1	0.0895	0.067	(0.015, 0.27)
Mountain HV1	0.235	0.11	(0.027, 0.35)
East-Mountain HV1	0.235	0.12	(0.035, 0.37)
Numt I	0.601	0.74	(0.17, 3.2)
Numt II	0.654	0.48	(0.13, 2.1)
Gorilla HV1	0.235	0.18	(0.064, 0.44)
All Gorilla	2.85	2.6	(1.3, 5.6)
Chimp1-Chimp2	0.039	0.032	(0.0032, 0.15)
All Chimp	0.158	0.13	(0.037, 0.4)
Human-Chimp	5.047	5.4	(4.1, 8)
	Other parameters		
μ_{HV1}	0.075	0.13	(0.058, 0.28)
μ_{Numt}	0.0096	0.0042	(0.00071, 0.03)
κ_{HV1}	19	23	(11, 50)
κ_{Numt}	19	22	(2.4, 970)
η	0.5	0.63	(0.11, 6.3)

Table 11: **The numerical results from the analysis of the simulation data.** The estimates are obtained by summarizing the six MCMC chains. The second column provides the true values were used in generating the simulation data. The third and fourth column are the median and 95% credible region from the posterior samples, respectively.

Count	Prob.	Tree topology
5400	1.000	{1-25}
5400	1.000	{1-21}
5400	1.000	{23-25}
5400	1.000	{4-6,13-15}
5400	1.000	{7-9,16-21}
5395	0.999	{1-3,10-12}
5340	0.989	{20-21}
5326	0.986	{7-8}
5194	0.962	{10-12}
5140	0.952	{22-25}
5120	0.948	{23-24}
4715	0.873	{7-9}
3921	0.726	{16-19}
3619	0.670	{16-17}
1811	0.335	{10,12}
1394	0.258	{1,10-12}
1228	0.227	{16-18}
980	0.181	{7-9,20-21}
161	0.030	{6,13}
160	0.030	{5,14}
134	0.025	{1-3,7-12,16-21}
118	0.022	{4,6,13}
99	0.018	{5,14-15}
1	0.000	{1,3,10-12}

Table 12: **The posterior probabilities for the clades after summarizing the 5400 posterior sampled trees.**

	tMRCA (in MYA)			Population size		
Groups	mean	median	95% C.R.	Mean	median	95% C.R.
W-HV1	0.24	0.22	(0.07,0.42)	15000	11000	(3100,31000)
E-HV1	0.15	0.11	(0.095,0.3)	13000	5300	(4200,31000)
M-HV1	0.25	0.26	(0.071,0.45)	16000	8700	(1300,49000)
W-Numt	5.3	5.4	(2.8,7.6)	64000	45000	(17000,160000)
W-Numt I	0.33	0.25	(0.039,0.79)	24000	20000	(300,67000)
E-Numt	4	3.9	(2,6.6)	36000	21000	(12000,89000)

Table 13: **The estimates of tMRCA and effective population sizes from BEAST.**

Ecology and Evolutionary Biology, University of Arizona. 948

Anthony, N. M., Clifford, S. L., Bawe-Johnson, M., Abernethy, K. A., Bruford, M. W., and Wickings, E. J. (2006). "Distinguishing gorilla mitochondrial sequences from nuclear integrations and PCR recombinants: Guidelines for their diagnosis in complex sequence databases." *Molecular Phylogenetics and Evolution*, 43: 553–566. 921

Anthony, N. M., Johnson-Bawe, M., Jeffery, K., Clifford, S. L., Abernethy, K. A., Tutin, C. E., Lahm, S. A., White, L. J. T., Utley, J. F., Wickings, E. J., and Bruford, M. W. (2007). "The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa." *Proceedings of the National Academy of Sciences*, 104: 20432–20436. 918

Bensasson, D., Zhang, D., Hartl, D., and Hewitt, G. (2001). "Mitochondrial pseudogenes: evolution's misplaced witnesses." *Trends in Ecology and Evolution*, 16(6): 314–321. 920

Benson, D., Karsch-Mizrachi, I., Lipman, D., Ostell, J., and Sayers, E. (2011). "GenBank." *Nucleic Acids Research*, 39(Database issue): D32–37. 921

Brown, W., Prager, E., Wang, A., and Wilson, A. (1982). "Mitochondrial DNA sequences of primates: tempo and mode of evolution." *Journal of Molecular Evolution*, 18(4): 225–39. 920, 943

Burrows, W. and Ryder, O. A. (1997). "Y-chromosome variation in great apes." *Nature*, 385: 125–126. 948

Chen, F.-C. and Li, W.-H. (2001). "Genomic Divergences between Humans and Other Hominoids and the Effective Population Size of the Common Ancestor of Humans and Chimpanzees." *American Journal of Human Genetics*, 68(2): 444–456. 936, 937

Clifford, S., Abernethy, K., White, L., Tutin, C., and Bruford, M. (2003). "Genetic studies of western gorillas." In Taylor, A. and Goldsmiths, M. (eds.), *Gorilla biology: a multidisciplinary perspective*, 269–292. Cambridge University Press. 949

Clifford, S. L., Anthony, N. M., Bawe-Johnson, M., Abernethy, K. A., Tutin, C. E. G., White, L. J. T., Bermejo, M., Goldsmith, M. L., Mcfarland, K., Jeffery, K. J., Bruford, M. W., and Wickings, E. J. (2004). "Mitochondrial DNA phylogeography of western lowland gorillas (Gorilla gorilla gorilla)." *Molecular Ecology*, 13: 1551–1565. 918, 921, 937

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G., and Solomon, W. (2002). "Estimating Mutation Parameters, Population History and Genealogy Simultaneously From Temporally Spaced Sequence Data." *Genetics*, 161: 1307–1320. 919

Drummond, A. J. and Rambaut, A. (2007). "BEAST: Baysian evolutionary analysis by sampling trees." *BMC Evolutionary Biology*, 7: 214. 919, 943

Felsenstein, J. (1973). "Maximum likelihood estimation of evolutionary trees from continuous characters." *American Journal of Human Genetics*, 25: 471–492. 931

— (1981). "Evolutionary trees from DNA sequences: A maximum likelihood approach." *Biological Journal of the Linnean Society*, 17: 368–376. 930, 931

Gelman, A. and Rubin, D. (1992). "Inference from iterative simulation using multiple sequences." *Statistical Science*, 7: 457–511. 932

Goldberg, T. L. and Ruvolo, M. (1997). "The geographic apportionment of mitochondrial genetic diversity in East African chimpanzees, Pan troglodytes schweinfurthii." *Molecular Biology*, 14: 976–984. 949

Green, P. J. (1995). "Reversible jump Markov chain Monte Carlo computational and Bayesian model determination." *Biometrika*, 82(4): 711–732. 951

Groves, C. P. (1967). "Ecology and taxonomy of the gorilla." *Nature*, 213: 890–893. 921

— (1970). "Population systematics of the gorilla." *Journal of Zoology*, 161: 287–300. 921

Haffer, J. (1969). "Speciation in Amazonian forest birds." *Science*, 165: 131–137. 918

Harpending, H. C., Dagger, M. A., Gurven, M., Jorde, L. B., Rogers, A. R., and Sherry, S. T. (1998). "Genetic traces of ancient demography." *Proceedings of the National Academy of Sciences*, 95: 1961–1967. 949

Hartl, D. L. (2000). *A Primer of Population Genetics*. Sinauer, 3 edition. 917, 946

Hasegawa, M., Kishino, H., and Yano, T. (1985). "Dating of the human-ape splitting by a molecular clock of mitochondrial DNA." *Journal of Molecular Evolution*, 22: 160–174. 931

Hastings, W. K. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, 57(1): 97–109. 932

Hey, J. and Nielsen, R. (2007). "Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics." *Proceedings of the National Academy of Sciences*, 104: 2785–2790. 919

Hu, X., Javadian, A., Gagneux, P., and Robertson, B. H. (2001). "Paired chimpanzee hepatitis B virus (ChHBV) and mtDNA sequences suggest different ChHBV genetic variants are found in geographically distinct chimpanzee subspecies." *Virus Research*, 79: 103–108. 921

Ingman, M., Kaessmann, H., Pääbo, S., and Gyllensten, U. (2000). "Mitochondrial genome variation and the origin of modern humans." *Nature*, 408: 708–713. 921

Jensen-Seaman, M., Deinard, A., and Kidd, K. (2003). "Mitochondrial and nuclear DNA estimates of divergence between and western gorillas." In Taylor, A. and Goldsmiths, M. (eds.), *Gorilla biology: a multidisciplinary perspective*, 247–268. Cambirdge University Press. 948

Jensen-Seaman, M. and Kidd, K. (2001). "Mitochondrial DNA variation and biogeography of eastern gorillas." *Molecular Ecology*, 10: 2241–2247. 918, 936, 943, 948, 949

Jensen-Seaman, M. I., Deinard, A. S., and Kidd, K. K. (2001). "Modern African Ape Populations as Genetic and Demographic Models of the Last Common Ancestor of Humans, Chimpanzees, and Gorillas." *Journal of Heredity*, 92: 475–480. 918

Kortland, A. (1983). "Marginal habitats in chimpanzees." *Journal of Human Evolution*, 12: 231–278. 949

Larget, B. and Simon, D. L. (1999). "Faster Likelihood calculations on trees." Technical report, Duquesne University. 931

Lopez, J., Yuhki, N., Masuda, R., Modi, W., and O'Brien, S. (1994). "Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat." *Journal of Molecular Evolution*, 39(5): 174–190. 920

Morin, P., Moore, J., and Chakraborty, R. (1994). "Kin selection, social structure, gene flow, and the evolution of chimpanzees." *Science*, 265: 1193–1201. 949

Nei, M. and Li, W.-H. (1979). "Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases." *Proceedings of the National Academy of Sciences*, 76: 5269–5273. 918, 924

Nielsen, R. and Wakeley, J. (2001). "Distinguishing Migration From Isolation: A Markov Chain Monte Carlo Approach." *Genetics*, 158: 885–896. 919

Pidanciera, N., Jordana, S., Luikarta, G., and Taberleta, P. (2006). "Evolutionary history of the genus Capra (Mammalia, Artiodactyla): Discordance between mitochondrial DNA and Y-chromosome phylogenies." *Molecular Phylogenetics and Evolution*, 40: 739–749. 948

Rannala, B. and Yang, Z. (2003). "Bayes Estimation of Species Divergence Times and Ancestral Population Sizes Using DNA Sequences From Multiple Loci." *Genetics Society of America*, 164: 1645–1656. 919

— (2007). "Inferring Speciation Times under an Episodic Molecular Clock." *Systematic Biology*, 56: 453–466. 919

Ricchetti, M., Tekaia, F., and Dujon, B. (2004). "Continued colonization of the human genome by mitochondrial DNA." *PLoS Biol*, 2(9): E273. 920

Richly, E. and Leister, D. (2004). "NUMTs in sequenced eukaryotic genomes." *Molecular Biology and Evolution*, 21(6): 1081–4. 920

Simon, D. L. and Larget, B. (2004). "Bayesian Analysis to Describe Genomic Evolution by Rearrangement (BADGER)." Technical report, Department of Mathematics and Computer Science, Duquesne University. 937

Tajima, F. (1983). "Evolutionary relationship of DNA sequences in finite populations." *Genetics*, 105: 437–460. 919, 924

Takahata, N., Satta, Y., and Klein, J. (1995). "Divergence Time and Population Size in the Lineage Leading to Modern Humans." *Theoretical Population Biology*, 48(2): 198–221. 922

Tamura, K. and Nei, M. (1993). "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees." *Molecular Biology and Evolution*, 10: 512–526. 922, 936, 943

Thalmann, O., Fischer, F., Lankester, F., Pääbo, S., and Vigilant, L. (2007). "The Complex Evolutionary History of Gorillas: Insights from Genomic Data." *Molceular Biology and Evolution*, 24: 146–158. 918, 936, 937, 943, 948

Thalmann, O., Hebler, J., Poinar, H., Pääbo, S., and Vigilant, L. (2004). "Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes." *Molecular Ecology*, 13: 321–335. 921

Thalmann, O., Serre, D., Hofreiter, M., Lukas, D., and Eriksson, J. (2005). "Nuclear insertions help and hinder inference of the evolutionary history of gorilla mtDNA." *Molecular Ecology*, 15: 179–188. 920, 935, 936

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Research*, 22: 4673–4680. 921

Tosi, A. J., Morales, J. C., and Melnick, D. J. (2000). "Comparison of Y Chromosome and mtDAN Phylogenies Leads to Unique Inferences of Macaque Evolutionary History." *Molecular Phylogenetics and Evolution*, 17: 133–144. 948

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., and Wilson, A. C. (1991). "African populations and the evolution of human mitochondrial DNA." *Science*, 253: 1503–1507. 922

World Wildlife Foundation (2008).
URL http://www.wwf.org/ 921

Wright, S. (1931). "Evolution in Mendelian populations." *Genetics*, 16: 97–159. 923

— (1938). "Size of population and breeding structure in relation to evolution." *Science*, 87: 430–431. 923

Yamagiwa, J. (1986). "Activity rhythm and the ranging of a solitary male mountain gorilla (Gorilla gorilla beringei)." *Primates*, 27: 273–282. 948

Yamagiwa, J. and Mwanza, N. (1994). "Day-Journey Length and Daily Diet of Solitary Male Gorillas in Lowland and Highland Habitats." *International Journal of Primatology*, 15: 207–224. 948

Yang, Z. (1994). "Maximum Likelihood Phylogenetic Estimation from DNA sequences with Variable Rates over Sites: Approximate Methods." *Journal of Molecular Evolution*, 39: 306–314. 931

Yang, Z. and Rannala, B. (2006). "Bayesian Estimation of Species Divergence Times Under a Molecular Clock Using Multiple Fossil Calibrations with Soft Bounds." *Molecular Biology and Evolution*, 23: 212–226. 919

Yu, N., M. I. Jensen-Seaman, L. C., Ryder, O., and Lia, W.-H. (2004). "Nucleotide Diversity in Gorillas." *Genetics*, 166: 375–1383. 937, 948

Zischler, H., H, H. G., von Haeseler, A., and Pääbo, S. (1995). "A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans." *Nature*, 378(6556): 489–492. 920

Acknowledgments

We thank Cecile Ane, Michael Newton, Bret Payseur and Sunduz Keles as well as anonymous reviewers for many helpful discussions or comments, which have improved the paper.