Structure and evolution of the repeated region of S100 'fused' type genes across primates and filaggrin variations in an Ecuadorian pediatric population

Romero Aguilar, Vanessa Isabel

Doctor of Philosophy

Department of Genetics
School of Life Science
SOKENDAI (The Graduate University for Advanced Studies)

Structure and evolution of the repeated region of S100 'fused' type genes across primates and filaggrin variations in an Ecuadorian pediatric population

Romero Aguilar, Vanessa Isabel

Doctor of Philosophy

Department of Genetics
School of Life Science
SOKENDAI (The Graduate University of Advanced Studies)

Date of creation: June 2017
Date of graduation: September 2017

Table of Contents

List of Tables

Filaggrin

Hornerin

Filaggrin-2

Repetin

Cornulin

Filaggrin variation in Ecuadorian pediatric population

List of Figures

Summary

Background

The S-100 ‘fused’ type genes (SFTP) are members of a gene family part of the Epidermal Differentiation Complex (EDC). EDC is a cluster of genes important for skin structure and were previously described as rapidly divergent. All SFTPs have a similar structure with a tandem repeated region on their third exon.

The evolutionary dynamics of repeated sequences is quite complex, with some duplicates never having differentiated from each other. Two models can explain the evolutionary process for repeated genes—concerted and birth-and-death, of which the latter maintains similarity by purifying selection and has a high level of diversity across repeats. The result of random duplications and losses in repeated regions might modulate molecular pathways and therefore affect phenotypic characteristics in a population. The effect of repeats variation is of interest for the SFTP family as all members share a repetitive exon structure and are located on a rapidly divergent region.

Filaggrin, a member of the SFTP, contains repeat variations across and within species. In human, the filaggrin repeat number variation affects its function with fewer repeats resulting in a higher risk for atopic dermatitis (AD). Globally, Ecuador has the second highest prevalence of AD in children but no studies associated to filaggrin.

I investigated whether the variation in the number of tandem repeats could be found in all SFTPs. Next, I examined which model, concerted or birth-and-death fits best for each member of the SFTP. Finally, I searched for filaggrin variations associated to AD in pediatric Ecuadorian cases.

Materials and methods

The members of the SFTP family are cornulin (*CRNN*), filaggrin-2 (*FLG-2*), filaggrin (*FLG*), hornerin (*HRNR*), repetin (*RPTN*) and trichohyalin (*TCHH*). I obtained DNA sequences from the NCBI database for *FLG, HRNR, RPTN* and *CRNN* of human, chimpanzee, gorilla, orangutan, and macaque, and for *FLG-2* of baboon and marmoset. *TCHH* was not used because it consists of domains with a low conservation rather than repeats. In *FLG*, I obtained new sequences by combining short and long length-sequencing methods.

For each member of the SFTPs, I performed multiple alignment, and phylogenetic analyses across species. Next, I estimated the nucleotide variation between repeats within a species. Finally I searched for codons under selection by maximum likelihood analyses.

Ecuadorian pediatric samples were obtained from CEPI-center (Quito-Ecuador). I extracted DNA from dry blood card by a modified extraction protocol, performed PCR for the repeated region and sequenced using MiSeq system. I evaluated possible associations by using multiple logistic regression.

Results and discussion

I found that SFTP had variation in the number of repeats across species. Overall, the nucleotide variation for each gene ranged similarly and was comparable with that of birth-and-death model. In addition, most of variations are synonymous which is the consequence of the birth-and-death model. Under the birth-and-death model, duplicates, with enough divergence time, can lead to lineage-specific expansions which I observed in the phylogeny of *FLG, FLG-2, HRNR* and *CRNN*.

The nucleotide diversity and phylogeny analyses supported a previously described SFTP division. This division grouped SFTP intro three groups, filaggrin-type (*FLG, FLG-2* and *HRNR*), trichohyalin-type (TCHH and RPTN) and cornulin. Filaggrin-type had high nucleotide variation and duplicates within species, which allow variability in the cornified epithelium. The trichohyalin-type had low nucleotide changes, its repetitive domains are conserved and single nucleotide variations have a phenotypic effect. Cornulin is the most conserved with the lowest nucleotide variability due to its role in cell cycle.

Additionally, I found individual findings for each SFTPs. For *FLG*, I observed length variation of the repeated region in chimpanzee and crab-eating macaque. For *HRNR*, I clarified the repeated unit formation starting with 39 bp (primary unit) which duplicated to form a 351 unit (secondary unit) and then duplicated to form a 702 unit (tertiary unit) which further replicated and form the quartic or longest unit for human, chimpanzee, gorilla and orangutan. *FLG-2* had two different repeated regions that evolved separately but still under the birth-and-death model. For *CRNN*, previously was suggested to evolve under positive selection however my analysis demonstrated misalignment in that analysis and clarified purifying selection.

I analyzed the filaggrin variations in Ecuadorian pediatric population. Most cases and controls have 12-filaggrin repeats and were not associated with the risk for

AD. LOF mutations are also associated with AD. The frequency of the five most common variations in Northern Europeans was low in Ecuadorians. However, I associated 2 new non-synonymous damaging variations (E2250Q and E2652D) with cases.

Conclusions

I concluded that SFTPs evolved under the birth-and-death model by showing high nucleotide diversity within a species and duplication and loss events within and across species. Also, I concluded that filaggrin repeat variation was not associated to the risk of AD in Ecuadorian children; however, I detected two new non-synonymous damaging variants associated to cases.

Chapter 1

OVERALL INTRODUCTION

The evolution of gene families and tandem repeat regions has been a controversial issue in evolutionary genetics because most of the repeat regions do not follow the conventional divergent model. In the divergent model, each duplicate acquires a new function and gradually separates. However, repeated genes often maintain their original functions and are more similar within species than among related species (Nei and Rooney 2005) (Figure 1).

Two alternative models of evolution have been proposed to explain this non-divergent evolutionary pattern in repeated gene families: concerted and birth-and-death. In the concerted model, mutations that occur in one repeat spread to the other repeats by unequal crossover or gene conversion, maintaining homogeneity among the repeats (Figure 1). This results in a low level of diversity of repeats within a species, and the sequences of the repeats being more similar within a species than between species. In the birth-and-death model, new repeat genes diversify by synonymous nucleotide substitutions. Some repeats are maintained and amplified for a long time under purifying selection, whereas others are deleted or become nonfunctional as a result of deleterious mutations, which can lead to a change in the number of copies between species. Thus, there is a high level of synonymous diversity between repeats, and genes have a close interspecies pattern (Nei and Rooney 2005; Austen and Kobayashi 2007; Eirín-López et al. 2012; Sabbagh et al. 2013).

Polyubiquitin genes and rDNA genes are representative multigene families that have evolved under birth-and-death evolution and concerted evolution respectively. The main difference between these two models is the nucleotide diversity between repeats which is low in rDNA genes (range = 0.01×10^{-3} to 0.18×10^{-3}) and high in polyubiquitin genes (range = 88.6×10^{-3} to 197×10^{-3}) (Nei et al. 2000; Austen and Kobayashi 2007). Most of repeated regions are known to evolve by birth-and-death model.

The S100-Fused Type Protein (SFTP) family, whose members share a tandem repeat structure on their third exon, are located on chromosome 1 in a cluster-manner and have a related function in the cornified epithelium (CE) (Henry et al. 2012,

Kypriotou et al. 2012, Zhihong et al. 2009) (Figure 2). The SFTPs consist of an amino-terminal S100 domain that is followed by the highly repetitive region. The members of the SFTP family are cornulin (*CRNN*), filaggrin-2 (*FLG-2*), filaggrin (*FLG*), hornerin (*HRNR*), repetin (*RPTN*) and trichohyalin (*TCHH*) (Kypriotou et al. 2012) (Figure 2 and Figure 3 A). In this section, I will briefly explain the current knowledge on the evolutionary characteristics of the SFTP.

The SFTP can be divided into three groups, filaggrin-type (*FLG, FLG-2* and *HRNR*), trichohyalin-type (*TCHH* and *RPTN*) and cornulin (*CRNN*), according to amino acid composition, expression pattern and function (Mlitz et al. 2014) (Figure 2). The SFTP genes appear firstly in amniotes. Sauropsids species have two SFTP-like genes, a homolog of mammalian cornulin and a trichohyalin-like gene named scaffoldin (SCFN). Like their mammalian counterparts, these sauropsids genes have similar gene structure, which suggests a conserved function in epithelial scaffoldings of growing skin appendages like claws and nails, feathers and hair. Scaffoldin in sauropsids species and chicken have variation in length on their repeated region and in the total percentage of aminoacid residues, indicating a functional involvement of repeats. In contrast, filaggrin-type proteins are only found in mammals, suggesting a mammal-specific gene duplication with subsequent modifications (Figure 3 B) (Mlitz et al. 2014). These differences can be recognized as an example of structural variations between species resulting from species-specific duplication or loss event. Structural variations are a major source of morphological differences, which may allow individuals to adapt to new environments and accelerate divergence between species (Fondon and Garner 2004; Feuk et al. 2006; Gazave et al. 2011; Paudel et al. 2013; Sudmant et al. 2013).

The SFTPs have a repeated region in their third exon and the repeat sequences for each SFTP were characterized in humans and in mouse (Brown and Irwin 2012, Fallon et al. 2009, Hansmann et al. 2012, Huber et al. 2005, Krieg et al. 1997, Little et al. 2007, Makino et al. 2001, Takaishi et al. 2005, Wu et al. 2009, Xu et al. 2000). In human filaggrin repeats vary in number and length within species and between mouse and dog (Fallon et al. 2009; Kanda et al. 2013). Within humans, the number of repeats affects the amount of filaggrin degradation products which are a component of the skin natural moisturizing factor (Kezic et al. 2011; Brown et al. 2012; Dipankar De 2012). The filaggrin variation among closer species, like primates, has not been studied. Additionally from previous studies, the number of repeats differs between

human and mouse in filaggrin (human = 10-12 and mouse = 10-20) (Brown and Irwin 2012, Zhang et al. 2002), and in repetin (human = 28 and mouse = 48) (Krieg et al. 1997, Huber et al. 2005); between human and chimpanzee in cornulin (human = 2 and chimpanzee = 4) (Little et al. 2007, Xu et al. 2000) but is the same for hornerin (human = 6 and mouse = 6) (Makino et al. 2001, Takaishi et al. 2005), and for filaggrin-2 (Filaggrin-2 repeats similar to filaggrin human = 14 and mouse = 14) (Hansmann et al. 2012, Wu et al. 2009). Additionally, hornerin repeats in human divide into three subunits (Takaishi et al. 2005) but those in mouse divide into two subunits (Makino et al. 2001). However, whether the subunit structure is conserved among closer species is not known. Variation in the number of repeats, as in the case of filaggrin, can have a functional effect. At the same time, the nucleotide compositions and variation among repeat units can be important especially in the interference about the evolutionary pathway of repeated regions because the main difference between the concerted and birth-and-death models is the range of the nucleotide variation.

In human, the number of *FLG* complete repeats varies from 10 to 12 within individuals and between populations. The variation in the number of repeats affects filaggrin degradation products with fewer repeats resulting in dry skin and increased number of repeats being protective (Kezic et al. 2011; Brown et al. 2012). Additionally, loss-of-function mutations have also been strongly associated with atopic dermatitis and differ among populations (Gimalova et al. 2016, Li et al. 2016). Globally, Ecuador has the second highest prevalence of atopic dermatitis (22.5%) for children between 6 to 7 years, however filaggrin alterations associated to atopic dermatitis have not been analyzed (Odhiambo et al. 2009).

In this thesis, I investigated variations within repeat-units of each member of the SFTP family within an individual and across primate species and inquired whether the concerted model or the birth-and-death model best fits the evolutionary pathway for the tandem repeat region. Also, I searched for *FLG* repeat-unit variation and mutations associated to atopic dermatitis in Ecuadorian pediatric population.

Chapter 2

SFTP VARIATION IN PRIMATES

CHAPTER SUMMARY

In this chapter, I focused on the variation and evolution of the tandem repeat region of a family of genes known as the S100 domain fused gene type proteins (SFTP) across primates. The SFTP are part of the Epidermal Differentiation Complex (EDC), a dense cluster of genes important for skin structure and previously categorized as rapidly divergent. SFTP have a similar structure with a tandem repeat region on their third exon. The number of tandem repeats of one of the SFTP, filaggrin, has been reported to vary across and within a species probably due to duplications and losses. Taking this into consideration, I investigated if this variation in the number of tandem repeats could be found in all the members of the SFTP. The variation in the number of repeats can be recognized as a structural variation that may allow individuals to adapt to new environments and accelerate divergence between species. The evolution of the tandem repeated regions differs from classic divergent evolution and can be explained by two evolutionary models: concerted and birth-and-death. The major differences between these two models are based on nucleotide diversity between repeats within a species, clusters observed in phylogenetic analysis, and the presence or absence of gene conversion events and pseudogenes. I searched which the model that fits best for each of the SFTPs in primates.

2.1. Filaggrin

2.1.1 Introduction

FLG encodes profilaggrin protein which is dephosphorylated and degraded to monomeric filaggrin repeats and then further proteolyzed to aminoacids (Figure 4). The complete repeats are responsible for the function of this protein in the skin. The number of complete repeats varies from 10 to 12 between human (*Homo sapiens*) individuals and populations. This copy number variation is associated with the expression level of filaggrin in the epidermis (Gan et al. 1990; Redon et al. 2006; Kezic et al. 2011; Brown et al. 2012; Dipankar De 2012). Several nonsynonymous variants in the repeat region have also been reported to be associated with various dry skin disorders in humans, particularly atopic dermatitis and ichthyosis vulgaris (Smith et al. 2006; Brown et al. 2012).

Comparative genomic studies have revealed that the structure of *FLG* is similar across human, mouse (*Mus musculus*), and dog (*Canis lupus familiaris*). However, the number of complete repeats and the length of repeats differ between species (Fallon et al. 2009; Kanda et al. 2013). These differences can be the result of species-specific duplication or loss events (Fondon and Garner 2004; Feuk et al. 2006; Gazave et al. 2011; Paudel et al. 2013; Sudmant et al. 2013).

2.1.2 Materials and Methods

2.1.2.1 Database search for primate *FLG* sequences

I obtained full-length *FLG* DNA sequences from the NCBI gene database (http://www.ncbi.nlm.nih.gov/gene/) for the following primates: *H. sapiens* (NC 000001.11), *P. troglodytes* (NC 006468.3), *G. gorilla* (NC 018424.1), *P. abelii* (NC 012591.1), and *M. fascicularis* (NC 022272.1). In addition I obtained *FLG-like* DNA sequences for *P. troglodytes* (NW 003460518.1) and *M. fascicularis* (NW 005093011.1 and NC 022274.1).

2.1.2.2 Primate samples

Macacca fascicularis (ID: 181 and 246) and *Macacca mulatta* (ID: 725 and 970) samples were derived from cell line established from peripheral blood cells by EBV transformation provided by Dr. Takafumi Ishida from the Department of Biological Sciences, Graduate School of Science, University of Tokyo. Two *G. gorilla* DNA samples (Primate ID: 1943 and 3846) were provided by Dr. Osamu Takenaka from

the Primate Research Institute of Kyoto. Five *Pongo pygmaeus,* one *G. gorilla* and fourteen *P. troglodytes* DNA samples were obtained through the Great Ape Information Network (GAIN): *P. pygmaeus* (GAIN ID: 0091, 0031) from Tennoji Zoo, *P. pygmaeus* (GAIN ID: 0110) from Kobe City Zoo, *P. abelli* (GAIN ID: 0010) from Nagoya Higashiyama Zoo, *G. gorilla* (GAIN ID: 0080) from Ehime Tobe Zoo, and *P. troglodytes* (GAIN ID: 0143, 0158, 0170, 0212 0204, 0211, 0131, 0279, 0169, 0276, 0159, 0345 and Primate ID: 954 and 956) from Sanwa Kagaku Primate Park. All of the aforementioned zoos and parks are located in Japan.

Human samples were Ecuadorian and recruited at Centro de la Piel, Quito-Ecuador. All the participants provided written informed consent. The Ethics Committee of National Institute of Genetics (nig 1419, 2014.11) approved the study protocol.

2.1.2.3 Sequencing of *FLG* in primates using PacBio RSII and MiSeq

I performed long-range PCR of the repeated region and non-repeated region of *FLG* for the crab-eating macaque, orangutan, gorilla, chimpanzee and human samples. Primers were designed at specific conserved regions in each species and the human repeat variation region, the information about the primers is summarized in Table 1. The human primers for the non-repeated region did not match any area on the crab-eating macaque genome database and we needed to design specific primers for macaque from a different database (Table 1). PCR reactions were performed using a PrimeSTAR® GXL DNA Polymerase kit, with a total reaction volume of 10 μl that included: genomic DNA (20 ng), PrimeSTAR GXL 5X buffer, dNTP mixture (200 μM), forward and reverse primers (0.2 μM each), PrimeSTAR GXL polymerase (0.25 U/10 μl), and water. The PCR conditions were as follows: 94°C initial denaturation for 2 min, followed by 30 cycles each of denaturation (98°C for 10 s) and elongation (68°C for 10 min).

PacBio RS II (Pacific Biosciences, California, USA) is a single molecule, real-time DNA sequencing system that provides exceptionally long read lengths. I used all sequence reads generated by PacBio RSII that were longer than 8 kb. For the *de novo* and reference analyses, I used the Amplicon module in SMRT Portal and the Hierarchical Genome Assembly Process (Koren et al. 2012).

The MiSeq system (Illumina, San Diego, California, USA) was employed to generate high-throughput short reads for the identification of sequence variants. The

DNA libraries were sequenced on the MiSeq platform with 350- and 250-bp paired-end modules.

The sequence reads generated in this study are available in DDBJ database, http://www.ddbj.nig.ac.jp/ with the following accession number macaque (LC096129), orangutan (LC096130), gorilla (LC096131) and chimpanzee (LC096132).

2.1.2.4 Identification of *FLG* repeats in primates

I performed a dot-matrix analysis to detect repeat *FLG* sequences in *FLG-like* genes in chimpanzee and macaque using the Harrplot 2.1.8 program from GENETYX Corporation (Software Development Co., LTD, Tokyo, Japan). Initially, I compared the *FLG* sequence in chimpanzee and macaque with that in FLG-like respectively (Figure 5 A-E). The matched sequences were then reanalyzed, using chimpanzee-repeat 1 and macaque-repeat 1 as references, to search for possible repeat regions. Finally, I manually curated the possible repeat regions to determine the correct repeat sequences.

2.1.2.5 Multiple alignment and phylogenetic analyses

I performed the multiple sequence alignment using the profile alignment for nucleotide and codon in ClustalW implemented in MEGA 6.06 (Tamura et al. 2011). I then constructed maximum likelihood trees and neighbor-joining trees. Neighbor-joining trees were constructed by considering pairwise deletions, proportional nucleotide differences (*p*-distance), and 1,000 bootstrap resampling. I performed best DNA/Protein Model detection included in MEGA 6.06 which was Tamura-Nei + Gamma Distributed with Invariant sites model (Table 2). A cutoff value for the condensed tree was set at 50%.

2.1.2.6 Estimation of polymorphic/variant sites, nucleotide diversity, and ratio of synonymous and nonsynonymous sites using the DNAsp5 program

*K*a/*K*s is the ratio of the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon (*K*a), to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon (*K*s) (Hu and Banzhaf 2008; Librado and Rozas 2009). I estimated the *K*a/*K*s ratio for each pair of within-species repeats excluding gaps using the program DNAsp5. The level of purifying selection was then calculated as $(1 - Ka/Ks) \times 100$.

Nucleotide diversity (π), the average number of nucleotide differences per site between sequences, was calculated using the DNA polymorphism option in DNAsp5 (Librado and Rozas 2009).

2.1.2.7 Multiple alignment of repeated and non-repeated regions

I performed the multiple sequence alignment for the repeated and non-repeated *FLG* sequences using the profile alignment for nucleotide and codon in ClustalW implemented in MEGA 6.06 (Table 3) (Tamura et al. 2011). Sandilands et al. (2007) previously described the nucleotide and amino acid sequences for each repeat in human. I used their sequence of human-repeat 1 as a reference to perform nucleotide and codon multiple alignments with the repeated region previously recognized by dot-matrix analysis and identified the repeats for each species.

2.1.2.8 Recombination and gene conversion events

I used a range of nonparametric methods in the RDP4 software package (Martin et al. 2015) to characterize recombination and gene conversion events in the sequence alignments. Six phylogenetic and nucleotide substitution models [RDP applies pairwise scanning approach to detect recombination (Martin and Rybicki 2000), GENECONV finds the most likely candidates for aligned gene converstion events between pairs of sequences in the alignment (Padidam et al. 1999), Bootscan algorithm includes a Bonferroni corrected statistical test of recombination (Martin et al., 2005), MaxChi uses a sliding-window approach along pairwise comparisons to identify discrepancies (Smith M, 1992), Chimaera uses an approach similar to MaxChi but uses triplets of sequences (Posada and Crandall, 2001), and SiScan compares multiple sequences using a window-based approach, randomizes positions within a window and calculates z-scores (Gibbs et al. 2000)] were implemented (Martin et al. 2010). The general parameters were set to linear sequences at an acceptable *P* value of 0.01 after Bonferroni correction. I set phylogenetic evidence, polish breakpoints, and check for alignment consistency as prerequisites for the data processing. The RDP model was used with internal and external references. The G-scale mismatch penalty of GENECONV was changed to 2. Bootscan settings were modified to perform 1,000 bootstrap replicates with a cutoff percentage of 95%. SiScan, Chimaera, and MaxChi were run using the default settings (Maynard-Smith 1992; Padidam et al. 1999; Martin and Rybicki 2000; Posada and Crandall 2001; Martin et al. 2005).

2.1.2.9 Inference of gene duplication and loss in the species tree

The parsimony-based algorithm in the NOTUNG program (Durand et al. 2006) reconciles the species tree that has the better fit for duplication, transfer, loss, and incomplete lineage sorting events from a set of gene trees. The maximum likelihood tree described above was used as the gene tree (Durand et al. 2006; Vernot et al. 2008; Stolzer et al. 2012).

2.1.2.10 Repeat duplication and divergence time of *FLG* repeats

Bayesian Evolutionary Analysis Sampling Trees is a cross-platform program for the Bayesian analysis of molecular sequences using Markov chain Monte Carlo to weight potential trees according to their posterior probability. The resulting tree has branch lengths that are proportional to divergence time (Hahn et al. 2005; Drummond et al. 2012; Bouckaert et al. 2014). I used high and low estimates derived from the dates based on the young and old calibration estimates. Crab-eating macaque repeats have previously been estimated to diverge between 16.3 and 20.8 Mya (Steiper and Young 2006).

The parameters with the highest effective sample size were the site model TN93, estimation of substitution rate with gamma distribution = 4, log-normal relaxed clock, and birth-and-death model for all three nucleotide positions.

2.1.2.11 Likelihood ratio tests for positive selection

Selection on the repeat region of *FLG*-like was evaluated using the codon substitution models (Codeml) tool in the PAML package (version 4.7). I compared several pairs of nested models with and without positive selection to be tested in a likelihood ratio framework to examine whether adaptive evolution had occurred (Sabbagh et al. 2013).

I used the repeat sequences aligned to human-repeat 1 and performed three separate analyses: site-based test, branch-based test, and branch-site test. The site-based test compared models with neutral/negative selection (M1a and M7 (beta)) and positive selection (M2a and M8 (beta and ω)). The branch-based test considered heterogeneity among lineages and compared a free ratio model (where each branch has a different ω) with a one-ratio null model (M0, where ω is fixed to 1 for all branches). The branch-site test was used to evaluate whether the signature of positive selection was evident at a specific branch (foreground) against the rest of the branches (background) in the phylogeny. The Bayes Empirical Bayes approach was used to detect positive selection sites (Yang 1997, 2007).

In all of the analyses, statistical comparisons of the models were performed using likelihood ratio tests, in which twice the log-likelihood difference between the alternative and null models was compared with the critical values from a χ^2 distribution, with the number of degrees of freedom equal to the difference in the number of parameters between the two models. In the branch-site test, the Bonferroni correction was employed for the derived *P* value, to control for an increased type I error rate when comparing multiple foreground lineages (Yang 1997, 2007; Anisimova et al. 2001; Sabbagh et al. 2013).

2.1.3 Results

2.1.3.1 *FLG* sequences in primates

The DNA sequences of *FLG* were obtained from the National Center for Biotechnology Information (NCBI) gene database (http://www.ncbi.nlm.nih.gov/gene/) for the following primate species: *H. sapiens* (NC 000001.11), *P. troglodytes* (NC 006468.3), *G. gorilla* (NC 018424.1), *P. abelii* (NC 012591.1), and *M. fascicularis* (NC 022272.1). These were compared with the repeat region in human by carrying out a dot-matrix analysis using the Harrplot program, and the matched sequences were then compared with individual partial and complete repeats of human *FLG*. This allowed me to reconstruct the number and order of the repeats in these primate species.

The estimated number of complete repeats differed between species, as shown in Figure 6. It has previously been reported that the length of these complete repeats ranges from 972 to 975 bp and flanked by partially similar repeats (Brown and Irwin 2012). Based on this, I noted that there were several gaps in some of the complete repeats in orangutan from the NCBI database, and one of the partial repeats was not identified in gorilla. Furthermore, the complete repeats with gaps resulted in unframed repeats, which may produce truncated proteins (Figure 5 A-C). This finding implies that the sequences of *FLG* for these primates that were obtained from the NCBI database are also likely to be incomplete.

Sequencing of large and nearly identical repeated structures is a challenging task, and the presence of exonic copy number variation increases complexity of the sequence analysis (Brown and Irwin 2012). To overcome these difficulties and determine the complete nucleotide sequences of *FLG*, I combined two different sequencing platforms: PacBio RS II and Illumina MiSeq. The long reads (>8 kb)

generated by PacBio RS II were used to determine the overall structure of the repeat regions as an initial reference. To compensate for the high error rate of PacBio RS II, reads generated by MiSeq were then mapped to the initial reference to determine error-corrected sequences of *FLG* for each of the primates (Martin and Wang 2011; Koren et al. 2012). In the resulting sequences, there were no gaps in the complete repeats, and two partial repeats were retrieved for all species. The sequences I acquired altered the number of complete repeats to 10 repeats for human, chimpanzee, and gorilla, 9 repeats for orangutan, and 12 repeats for crab-eating macaque (Figure 6). In the subsequent analyses, each of the repeats was considered as an independent unit.

2.1.3.2 Phylogenetic analysis of complete repeats across species

FLG repeats have "cluster-per-species" pattern in both neighbor-joining, and maximum likelihood methods. All complete repeats in crab-eating macaque and orangutan grouped into their own clusters ("crab-eating macaque cluster" and "orangutan cluster," respectively), whereas the complete repeats in gorilla, chimpanzee, and human scattered across species in a large cluster ("gorilla/chimpanzee/human cluster") (Figure 7 and Figure 8). This large cluster contained nine subclusters of repeats among species. However, with the exception of subclusters "F" and "H", their counterparts were not in order—for example, chimpanzee-repeat 4 gathered with human-repeat 3 (Figure 7). To further confirm the intra and inter species clustering in primates, I added repeats of mouse and dog and then constructed neighbor-joining, and maximum likelihood methods, with a cutoff value of 50% and observed the same phylogeny (Figure 10 A-B).

The clusters of crab-eating macaque and orangutan repeats suggest that unique ancestral repeats were duplicated in these two primate species. The finding that human, chimpanzee, and gorilla repeats clustered together, but with limited conservation in the order of the repeat sequences, suggests that these three species share several ancestral repeats, and that random duplication and loss events occurred independently in each of the three species (Figure 9).

The main difference between the concerted and birth-and-death models of evolution is the high levels of intragenic nucleotide diversity that are only found in the latter (Nei and Rooney 2005; Austen and Kobayashi 2007; Eirín-López et al. 2012; Sabbagh et al. 2013). I calculated the total number of sites showing variation and the nucleotide diversity (π) between complete repeats of *FLG* using the DNAsp5

program, as described in the Materials and Methods section. The estimated π for human, chimpanzee, gorilla, orangutan, and crab-eating macaque was 7.7×10^{-2}, 7.5×10^{-2}, 6.1×10^{-2}, 7.7×10^{-2}, and 3.1×10^{-2}, respectively (Table 4 and Table 5). I then compared the π of *FLG* with that of polyubiquitin genes and rDNA genes, which are representative multigene families that have evolved under birth-and-death evolution and concerted evolution, respectively (Nei et al. 2000; Austen and Kobayashi 2007). The π of *FLG* was comparable with that of polyubiquitin genes (range = 88.6×10^{-3} to 197×10^{-3}) and much larger than those of rDNA genes (range = 0.01×10^{-3} to 0.18×10^{-3}) (Nei et al. 2000; Austen and Kobayashi 2007). This suggests that *FLG* repeats have evolved under the birth-and-death model.

2.1.3.3 Selection within the repeats of each species

The major driving force of the birth-and-death model is purifying selection. Under purifying selection, the number of nonsynonymous variations in a gene is expected to be smaller than the number of synonymous variations. Purifying selection can be measured by making pairwise comparisons ($(1 - Ka/Ks) \times 100$), which indicate the percentage of nonsynonymous mutations that have been eliminated (Nei 2007).

I found that the average percentage of eliminated nonsynonymous substitutions for human, chimpanzee, gorilla, orangutan, and crab-eating macaque was 32.23%, 38.14%, 23.47%, 35.49%, and 71.26%, respectively (Table 5), suggesting that *FLG* repeats have evolved under strong selective constraints. Although nucleotide substitutions were generally synonymous, I found 12 pairs of complete repeats within species with higher nonsynonymous variations: two pairs in human repeats (R3–R7 and R3–R10), two pairs in chimpanzee repeats (R3–R6 and R3–R10), and eight pairs in gorilla repeats (R1–R8, R2–R7, R2–R8, R2–R9, R3–R8, R5–R8, R7–R8, and R8–R9). None of the pairs of repeats in crab-eating macaque and orangutan had $Ka/Ks > 1$ (Table 6).

2.1.3.4 Crab-eating macaque and orangutan have evolved under the birth-and-death model by recent gene duplication

Under the birth-and-death model of evolution, multigene families are not only expected to have high levels of nucleotide variation and purifying selection but also an interspecies gene-clustering pattern in the phylogenetic analysis. I only found high nucleotide variation, purifying selection, and an interspecies repeat cluster in the complete repeats in gorilla, chimpanzee, and human. The other two species (crab-

eating macaque and orangutan) had high nucleotide variation and purifying selection but a intraspecies repeat clusters.

Intraspecies repeat clusters can occur if the repeats are under recent duplication. Therefore, to further characterize the evolutionary process behind the intraspecies repeat clusters detected in crab-eating macaque and orangutan, I constructed reconciled and divergence trees (Durand et al. 2006; Vernot et al. 2008; Stolzer et al. 2012), the parameters for which are described in the Materials and Methods section. These two trees suggest that the original crab-eating macaque repeat duplicated 28 Mya, and that each branch has undergone a series of five duplications in the past 5.5 Mya. By contrast, the original orangutan repeat separated 21 Mya, and the current nine repeats were created by one divergence event and seven subsequent duplications during the last 18 Mya (Figure 9 and Figure 11). Previously, ubiquitin C was suggested to evolve by concerted evolution due to lineage-specific homogenization in primates (Tachikui et al. 2003), however duplication events with enough divergence time was not considered and in further studies confirmed birth-and-death evolution (Nei et al. 2000; Austen and Kobayashi 2007).

I also found that gorilla, chimpanzee, and human have their own duplication events. For example, the four repeats in gorilla (R3, R5, R7, and R9) gathered within the gorilla/chimpanzee/human cluster in the phylogenetic tree (Figure 7). By combining the information about the subclusters from the phylogenetic tree with the duplication and loss events from the reconciled tree, I were able to infer the evolution of the complete repeats in each species. I named each of the subclusters from "A" to "I," according to the order of the repeat sequences in human, for example, the subcluster containing the first repeat in human was defined as subcluster "A". I found that the order of these subclusters was largely conserved between human and chimpanzee (Figure 12). In contrast, the order of the subclusters in gorilla differed from those of human and chimpanzee, suggesting specific duplications in this species (Figure 12). Here I named chimpanzee-repeat 9 as "I", this repeat did not cluster with any other repeats but the reconciled analysis showed that the counterpart of human was lost. These results suggest that during a period of divergence, each species has gone through independent duplication and loss events, which has led to the creation of a unique set of repeats and their own clusters.

2.1.3.5 Codons and branches under selection across species

In humans, it has been suggested that several amino acid substitutions of *FLG* may have evolved under positive selection (Sandilands et al. 2009; Brown and Irwin 2012; Dipankar De 2012). While homozygous null-mutations of *FLG* are associated with a variety of skin disorders, individuals expressing heterozygous null-mutations show a less severe phenotype and could develop immunity via the skin barrier. Population-specific and shared mutations have also been reported, which rules out genetic drift (Sandilands et al. 2009; Brown and Irwin 2012). In this study, I detected high *K*a/*K*s ratios (>1) in our within-species comparisons. Therefore, I investigated whether similar variations in *FLG* were also present in other primates by searching for signatures of positive selection in the complete repeats across species using phylogenetic analysis by maximum likelihood (PAML) software (Yang 1997, 2007; Sabbagh et al. 2013), as described in the Materials and Methods section.

First, I carried out two site-based tests for each codon of the complete repeats, in which we compared models with positive selection and neutral/negative selection (M1a (nearly neutral) vs M2a (positive selection) models and M7 (beta) vs M8 (beta and ω)). In both of these tests, the results supported the presence of positive selection for 14.13% and 10.80% of codons, respectively; in *FLG* complete repeats (Table 7).

Second, I scrutinized whether selective pressures differed between branches of the phylogenetic tree by a branch-based test. The result of this test was not significant, suggesting that each branch has not evolved independently, and that either all branches have evolved at the same rate or certain branches have evolved at similar rates (Table 7).

Third, I performed branch-site tests to detect signatures of positive selection at specific branches. Since the phylogenetic analyses outlined previously showed that the complete repeats were grouped into three clusters (crab-eating macaque cluster, orangutan cluster, and gorilla/chimpanzee/human cluster), I searched for the positively selected codons in the main branches of these clusters. I identified positively selected codons in the branches of the orangutan cluster and gorilla/chimpanzee/human cluster (Table 2), with the signature of positive selection identified at codon 26 in both of these branches. I used the repeat sequences aligned to human-repeat 1 when I performed the search for signatures of positive selection by PAML software; therefore the positively selected codon refers to codon 26 in human-repeat 1. The known functional domains of filaggrin are located in a linker region at amino acids 11–20 and a cleavage region for caspase-14 at amino acids 162–165 and

171–174 (Sandilands et al. 2007; Sandilands et al. 2009; Hoste et al. 2011). The positively selected amino acids identified in this study were not located in these domains (Table 7). However, it is possible that the positively selected codons detected in this study are located in as-yet-unknown functional domains.

2.1.3.6 Repeat variation found in crab-eating macaque and chimpanzee

In humans, the number of *FLG* complete repeats varies from 10 to 12 within individuals and populations (Kezic et al. 2011 and Brown et al. 2012). Taking the variation described in humans into consideration, I performed PCR of the repeated region for additional primate samples. Gel electrophoresis result showed that the size in chimpanzee and macaque samples varied suggesting the possibility of repeat variation in these species as seen in human (Figure 13).

2.1.3.7 Filaggrin-like gene evolution in chimpanzee and crab-eating macaque

My study focuses on the evolution of the repeated region of filaggrin under the birth-and-death model, which is localized within an exon. The birth-and-death model suggests the presence of pseudogenes. In the NCBI gene database, I did not find any filaggrin pseudogenes but there are two types of genes that have a similar repeated region as filaggrin; filaggrin2 and filaggrin-like. Filaggrin-2 repeats are shorter than complete filaggrin repeats (225bp) and I will explain it in the filaggrin-2 section. I found one filaggrin-like gene located in chromosome 1 of chimpanzee and two filaggrin-like genes located in chromosomes 1 and 3 of macaque. I compared these sequences with repeat of the corresponding species by a dot-matrix analysis using the Harrplot program (Figure 5 D-E). Filaggrin-like gene in chimpanzee includes 1 complete repeat between two incomplete repeats. Only one complete filaggrin repeat was found in the chimpanzee filaggrin-like sequence thus I could not perform any evolutionary analysis. Filaggrin-like gene in macaque located in chromosome 3 does not include any filaggrin repeat and the one located in chromosome 1 includes 5 complete repeats. In crab-eating macaque filaggrin-like and filaggrin have a similar total variation, average *K*a/*K*s, nucleotide variation; in addition no positively selected sites were detected (Table 8, Table 9 and Table 10). I concluded that filaggrin and filaggrin-like repeats have been evolving under a similar birth-and-death model with strong selective constrains and no positively selected codon.

2.1.3.8 'Non-repeated' region gene structure

Additionally I wanted to verify if the non-repeated region of filaggrin in primates was similar to the one in human. Human filaggrin gene includes 3 exons and 2 introns,

with the repeat region being found on the third. The rest of the human filaggrin gene comprises exon 1 with 54 bp, intron 1 with 9613 bp, exon 2 with 159 bp, an intron 2 with 570 bp and, in exon 3 a domain A, which contains two calcium-binding domain, and domain B, which facilitates the translocation in terminally differentiating keratinocytes, with 879 bp, for a total of 11275 bp which I further refer this region as the "non-repeated region". Neither exons other than exon 3 nor introns have filaggrin repeated units.

I obtained the sequences of the non-repeated region of primates from the National Center for Biotechnology Information (NCBI) gene database (http://www.ncbi.nlm.nih.gov/gene/) and performed the multiple sequence alignment using the profile alignment for nucleotide in ClustalW implemented in MEGA 6.06. I added the alignment result and additional result for the non-repeated region at Table 3. The alignment result indicated that the sequences from the database were much shorter than the one in human and to confirm this difference I performed PCR of the non-repeated region. The expected length of the PCR product for human primers was 15249 bp and for the macaque primers 16042 bp. Gel electrophoresis result showed that the non-repeated region from human, chimpanzee, gorilla, orangutan, and macaque have the similar length showing that the sequences from the database seemed to be incomplete and that the overall non-repeated structure should be similar (Figure 13). The non-repeated region in primates does not include filaggrin units thus the evolution of this non-repeated region is not likely to have an effect on the repeated region.

2.1.4. Discussion on filaggrin repeats

The repeat region of *FLG* consists of nearly identical complete repeats. However, the number of complete repeats differs between species, as previously reported in mouse and dog (Fallon et al. 2009; Kanda et al. 2013). It has previously been reported that gene-associated tandem repeats act as an accelerator of evolution by generating variation in structure and functionality (Fondon and Garner 2004). Birth-and-death model has been used to explain multigene families evolution. Under the birth-and-death model, duplicates vary by silent nucleotide variations, which, with enough divergence time, can lead to lineage-specific expansions. I detected this pattern of evolution in the repeats of crab-eating macaque and orangutan using both reconciled and divergence trees (Figure 7, Figure 8, Figure 9 and Figure 11) and was

corroborated by including the repeats of mouse and dog in the reconciled tree (Figure 10 C) (Fallon et al. 2009; Kanda et al. 2013). The clusters of species-specific repeats may reflect the high nucleotide diversity together with duplication and loss events in each species (Table 5), which seems to fit the birth-and-death model (Nei et al. 2000; Nei 2007; Sabbagh et al. 2013). Also, I detected evidence of purifying selection on *FLG*, which is the main driving force of the birth-and-death model. Macaque repeats had the highest purifying rate, most likely due to the recent gene duplication demonstrated in my analysis (Table 5 and Figure 11). High synonymous variation has previously been found in several genes under the birth-and-death model of evolution, including actin in flagellate protists (dinoflagellates) and the ubiquitin gene family (Nei et al. 2000; Zhu et al. 2013). Under the birth-and-death model of evolution, duplication and loss events are not caused by gene conversion. Gene conversion is an unlikely event for *FLG* repeats because they are located in one exon, none of the repeats are incomplete, and there is no gene family. Although several algorithms for detection of recombination events were examined, I could not identify any major recombination event (Table 11) (Sawyer 1999; Martin et al. 2010). Furthermore, reconstruction of the phylogenetic tree after removing putative recombinant regions did not largely change the topology, confirming a low possibility of gene conversion (Figure 14) (Chen et al. 2007). A similar finding has previously been reported for the NAT gene family, in which there was no evidence of gene conversion following the combined use of phylogenetic analysis and fine-scale synteny mapping (Sabbagh et al. 2013).

Several studies have demonstrated that copy number variations due to the duplication or loss of repeats are important in modulating molecular pathways. This includes morphological modifications in dog breeds and craniofacial or digit defects in humans, which have been shown to be due to variation in the length of developmental genes (Fondon and Garner 2004; Caburet et al. 2005); an increased in the risk for schizophrenia (Hosak et al. 2012); and several complex diseases and syndromes in human caused by heterozygosity, such as mild dysmorphic features (Feuk et al. 2006; Gemayel et al. 2010) or dosage compensation (Aldred et al. 2005).

FLG encodes profilaggrin, a long polyprotein. Profilaggrin is dephosphorylated and cleaved into filaggrin repeat units. In the outer layers of skin, the repeat units of filaggrin are degraded into free amino acids that are a major determinant of the natural moisturizing factor. Glutamine is converted into

pyrrolidone-5-carboxylic acid (PCA) and histidine is deiminated to cis-urocanic acid (UCA). PCA and UCA serve important functions as protecting against UV irradiation, maintaining the acidic pH, being involved in the local immune response, and maintaining overall homeostasis (Kezic et al. 2011). The amount of filaggrin degradation products can be altered by *FLG* loss-of function mutations and variation in the number of complete repeats. Loss-of-function mutations strongly associate with atopic dermatitis. Atopic dermatitis patients with null-mutations were reported to have decreased levels of PCA and UCA (Kezic et al. 2011; Brown et al. 2012; Dipankar De 2012). The number of *FLG* complete repeats varies from 10 to 12 within human. I found variation in the length of the repeated region within species such as chimpanzee and crab-eating macaque and suggest the possibility of repeat variation in these species as seen in human (Figure 13). In human, it was previously reported that the variation in repeats are due to duplicates of repeat 8 and 10. I designed primers for a product from human-repeat 7 to human-repeat 11 and gel electrophoresis showing more clearly this variation (Figure 13). The variation in the number of repeats affects the amount of filaggrin degradation products. Increased number of *FLG* repeats associate with a decreased risk for atopic dermatitis (Kezic et al. 2011; Brown et al. 2012). Therefore, random copy number variation in the *FLG* repeats within individuals could affect filaggrin degradation products and its functions, and will allow them to adapt more readily to a new environment (Nei 2007). In human, the effect of repeat variation was explained as mentioned before; however similar skin disorder or the functional significance of the repeat variation within primates has not been reported. Therefore, further studies are required to verify the functional significance of copy number variation of *FLG* repeats in primates.

I concluded that *FLG* repeats evolved under the birth-and-death model, by showing species-specific clusters in crab-eating macaque and orangutan; high nucleotide diversity together with duplication and loss events in each species and the unlikeliness of gene conversion. I conclude that the copy number variation in the complete repeats of *FLG* across primates is a consequence of species-specific expansions following a long period of divergence.

2. Hornerin

2.2.1. Introduction

Human hornerin is a 2850 amino acid (aa) protein, containing abundant serine (33.7%), glycine (24.4%), glutamine/glutamate (12.1%) and histidine (9.5%) residues (Takaishi et al. 2005, Tsu et al. 2011). Hornerin is expressed in regenerating and normal skin. Hornerin is detected in skin biopsies from different sites as head, trunk, legs, hands and feet. In normal skin, the mRNA expression level of *HRNR* is four times lower than that of *FLG* (Wu et al. 2009).

The structure of the repetitive region of human hornerin was reported by two groups: Takaishi et al. (2005) and Paar et al. (2011). Takaishi et al. (2005) described hornerin formation as follow: starting with 39 aa which later amplified 4-times (156 aa) and then triplicated to form three segments of 468 aa in tandem, these further amplified 6-fold to form a longer repeat unit of 2808 aa (Figure 16). On the other hand, Paar et al. described a different hornerin formation found by Global Repeat Map algorithm as the following 39 bp amplified by nine as a 'primary repeat unit' sequence, a 'secondary repeat unit' sequence of two sequences of 0.35 kb, two 'tertiary repeat unit' of 0.70 kb and finally, five 1.4kb 'quartic repeat unit' (Figure 16). I examined both of the possibilities, clarified which structure is found in primates and performed the nucleotide diversity analysis using the longest unit or quartic repeat unit since both have the same length.

I will refer to the description by Takaishi et al. (2005) as follows: to the 468 aa repeats as longest units from 1 to 6, to the second-longest units (156 aa) as subunits A, B and C and to the third or initial-117 bp unit (39 aa) from I-IV, and to the sequences described by Paar et al. (2011) as 39 bp 'primary', 350 bp 'secondary', 700 bp 'tertiary' and 1400 bp 'quartic'.

2.2.2 Materials and Methods

2.2.2.1 Database sequences of hornerin

I obtained full-length of hornerin DNA sequences from the NCBI gene database (http://www.ncbi.nlm.nih.gov/gene/) for the following species *H. sapiens* (NC 000001.11), *P. troglodytes* (NC 006468.3), *G. gorilla* (NC 018424.1), *P. abelii* (NC 012591.1), *M. fascicularis* (NC 022272.1). and *M. musculus* (NC 000069.6).

2.2.2.2 Identification of repeats in primates

I used the amino acid description of repeat 4 in human for the longest, second-longest (includes subunit A, B and C), and initial-117 bp units described by Takaishi et al (2005) and by using Align Sequences Nucleotide BLAST (Zheng et al 2000) detected the rest of the nucleotide sequence repeats in human (Figure 16). As described by Takaishi et al (2005), the sixth longest-unit does not include subunit C. Next, using BLAST (Zheng et al 2000), I compared the repeated sequences from human with those in crab-eating macaque, orangutan, gorilla and chimpanzee. Finally, the matched sequences were reanalyzed by BLAST and then manually curated (Figure 16 and Figure 17). Hornerin description by Takaishi et al. divided the longest units into 468 bp subunits (A, B and C) due to amino acid similarity and confirmed by phylogeny analysis between subunits A, B and C respectively; and additionally described each subunit being composed of four repeats ~39 amino acids in length (2005). In the case of crab-eating macaque and mouse, I also acquired the dot-matrix plot comparing each initial-117 bp unit of human with the longest-4 of crab-eating macaque and with mouse-1 subunit A and B (Zheng et al 2000) (Figure 18 A-C). Each of the identified longest, second-longest and initial-117 bp units was considered as an independent unit for the subsequent analyses. Additionally, I used the 39 bp sequence of the primary unit described by Paar et al (2011) and by using BLAST detected the primary units for human, chimpanzee, gorilla, orangutan and crab-eating macaque, and then reconstructed the secondary, tertiary and quartic units.

Multiple alignment viewer Mview (Brown et al 1998) allowed me to observe the alignment location between each of the initial-117 bp unit of human with the longest-4 of crab-eating macaque, orangutan and chimpanzee and the longest-2 of gorilla (Figure 19).

2.2.2.3 Multiple alignment and phylogenetic analyses

I performed nucleotide multiple sequence alignment, detecting the best DNA/Protein model and then constructed neighbor-joining trees and maximum likelihood trees as described in the *FLG* section by using MEGA 6.06 (Tamura et al. 2011). Additionally I used NOTUNG program to detect species-gene reconciled trees as mentioned in the *FLG* section (Durand et al. 2006; Vernot et al. 2008; Stolzer et al. 2012).

2.2.2.4 Estimation of polymorphic/variant sites, nucleotide diversity, and ratio of synonymous and nonsynonymous sites using the DNAsp5 program

I estimated the *K*a/*K*s ratio, level of purifying selection, nucleotide diversity (π), and the average number of nucleotide differences per site between sequences as mentioned in the *FLG* section by using DNAsp5 (Hu and Banzhaf 2008; Librado and Rozas 2009).

2.2.2.5 Likelihood ratio tests for positive selection

Selection on the repeat region of hornerin was evaluated using the codon substitution model (Codeml) tool in the PAML package (version 4.7). I performed site-based test, branch-based test, and branch-site test as described in the *FLG* section (Yang 1997, 2007; Anisimova et al. 2001; Sabbagh et al. 2013).

2.2.3 Results

2.2.3.1. Repeat number and similarity

I started with the DNA sequences of *HRNR* were obtained from the National Center for Biotechnology Information (NCBI) gene database which I aligned with the longest-4 in human by BLAST, and the matched sequences were then compared with the second-longest-units of human, and further to the initial-117 bp unit (Figure 16 and Figure 17). This allowed me to reconstruct the number and order of the longest, second-longest units and initial-117 bp unit in these primate species. In the subsequent analyses, each of the repeats was considered as an independent unit.

The number of the longest hornerin units was 6 for human, chimpanzee, orangutan, and crab-eating macaque and 4 for gorilla (Figure 17). The length of each longest-unit varied from 936 to 1443 bp as mentioned before due to the lack of subunit C in the last longest-unit (Figure 17). The percentage of similarity between the longest units within a species ranged from 75.76% in crab-eating macaque to 99.36% in human (Table 12).

2.2.3.2. Nucleotide variation within a species

I compared the nucleotide variation of the longest units within a species and the diversity ranged from 1.13×10^{-2} – 1.7×10^{-2} (Table 13). The nucleotide variation was comparable to the previously described birth-and-death model such as poly ubiquitin gene (range = 88.6×10^{-3} to 197×10^{-3}) and filaggrin repeated unit (range = 3.2×10^{-2} to 7.7×10^{-2}). Overall repeat comparison and pairwise repeat comparison revealed that most of variations are synonymous in concordance with purifying selection. Pairwise comparison detected higher non-synonymous variations in 3 out of 15 pairs in orangutan (Table 13 and 14).

2.2.3.3. Phylogeny analysis

I constructed all of the subsequent maximum likelihood trees and neighbor-joining trees using the detected best-fit model which was Kimura's two-parameter substitution model + Gamma distribution (Table 15).

Both methods gave similar phylogenetic trees (Figure 20 A and B). The longest-units, from 1 to 4 in crab-eating macaque (cluster named 'α'), from 2 to 5 in orangutan units (cluster named 'δ-3') and human 3 and 5 units (cluster named 'ε'), gathered within their species. Some of gorilla and chimpanzee repeats clustered across species as follow: chimpanzee longest-6 and gorilla longest-4 (cluster 'β'), chimpanzee longest-3 and gorilla longest-2 (cluster 'γ'). The rest of the repeats that gathered across species formed a big cluster named 'δ' which was further subdivided into sub-clusters as follow: chimpanzee longest-2 and human longest-2 named 'δ-1', chimpanzee longest-5 and gorilla longest-3 named 'δ-2', orangutan repeats from 2 to 5 as 'δ-3', and gorilla longest-1, chimpanzee longest-1 and human longest-1 named 'δ-4' (Figure 20 A). Chimpanzee longest-4 and orangutan longest-1 were included in cluster 'δ' but did not gather with other repeats. Finally, human longest-4 and longest-6, orangutan longest-6, crab-eating macaque longest-5 and crab-eating macaque longest-6 did not cluster with any repeat. Maximum likelihood and neighbor-joining analysis had the same results for cluster 'α', 'γ' and 'δ' and in addition neighbor-joining tree further grouped human longest-4 into cluster 'ε', and orangutan longest-6, crab-eating macaque longest-5 and crab-eating macaque longest-6 into cluster 'β' (Figure 20 B).

Additionally, I constructed a species-gene tree and identified the duplications and losses in the phylogeny. Cluster 'α', 'δ-3' and 'ε' were the result of duplications within crab-eating macaque, orangutan and human separately. Seven duplications occurred in cluster 'δ'. Gorilla lost a repeat similar to cluster 'δ-1', human lost a repeat similar to cluster 'δ-2', human and gorilla lost a repeat similar to chimpanzee longest-4, gorilla, the ancestor from chimpanzee and human lost a repeat similar to orangutan longest-1, orangutan lost a repeat similar to cluster 'δ-4', human lost a repeat similar to cluster 'γ', and human lost a repeat similar to cluster 'β' (Figure 20 D).

2.2.3.4. Positive selected codons across species

I searched for signature of positive selection in the longest repeats across species. While comparing all repeats across species, 17 codons were detected which

accounted for 0.4% (M2a) and 1.1% (M8) of the codons in a repeat unit of *HRNR* (Table 16). All branches evolved with a different selective pressure (Table 16). When comparing specific branches, *HRNR* macaque and repeat-1 branches had sites under positive selection (Table 16).

2.2.3.5 High order structure in primates

The model of hornerin formation has been controversial with two groups suggesting different formations. The proposed formation by Takaishi et al. is {[(117bp x4) x3 subunits] x6 units} and the formation by Paar et al. is {[(‘‘39 bp x9 primary’’ x2 secondary) x2 tertiary] x5 quartic} (Figure 16). The formation by Paar et al. was analyzed for both human and chimpanzee but was only detected in human (2011). I examined both possibilities and clarified which structure is found in primates.

First, I focused on the proposed by Takaishi et al. and estimated, using BLAST, the longest-units and second-longest units for all primates and the 117-bp for all primates except crab-eating macaque (Figure 16). The length of all units varies slightly in chimpanzee, gorilla, and orangutan and moderately in crab-eating macaque (Figure 17). Crab-eating macaque had low conservation to human at the 117-bp units and was excluded for further analysis (Figure 18 A). Multiple alignment and phylogenetic analyses for the longest-unit was described on a previous section, and as for the second-longest units, all subunits A, B and C clustered within themselves (Figure 21).

According to the hornerin formation proposed by Takaishi el al, a phylogenetic tree of the sequences would group into clusters I-IV of the initial-117 bp units from I-IV of each subunits A to C (Figure 16). Taking this into consideration, I constructed a phylogenetic tree for the initial-117 bp units for human, chimpanzee, gorilla and orangutan separately. The clustered pattern for the description by Takaishi et al. was not detected, instead I observed the following overall clusters: A-I, A-IV, B-III and C-II (named ‘Group-1st’), A-II, B-I, B-IV and C-III (named ‘Group-2nd’), and A-III, B-II, C-I and C-IV (named ‘Group-3rd’) with the exceptions of a few initial-117 bp units in human B-III, chimpanzee A-I and IV, B-I to III and C-III and IV, gorilla B-I and II and C-I, orangutan A-I and III, B-I to IV and C-I and III, and all orangutan C-II and IV (Figure 22 A-D). The phylogenic clustering description implies the following sequential pattern a ‘Group-1st’, ‘Group-2nd’ and a ‘Group-3rd’, which is the length of the secondary unit (351 bp) description by Paar et al. (2011) and not the second-longest unit (Figure 16 and Figure 22 A-D).

I next focused on the primary, secondary and tertiary units proposed by Paar et al. (2011). I started with the proposed 39 bp sequence and estimated, using BLAST, the primary units for all primates (Figure 23 A-B). The expected phylogeny of the primary units should divide them into 9 clusters (order from 1 to 9), as proposed by Paar et al. (2011). This pattern was observed for human, chimpanzee, gorilla and orangutan however the branches had low bootstrap value and when adjusting to a cut-off 50% the nodes were lost probably due to low similarity across primary units (~30% in human described by Paar et al. 2011) (Figure 24 A-E). Crab-eating macaque primary units did not divide into 9 clusters, as various primary units were misplaced into different order unit, therefore I constructed a phylogenic tree including all primary units of all priamtes. The misplacement of primary units suggests a different duplication pattern to the other primates with independent duplications and losses thus resulting in the length variation observed in the longest unit (Figure 23 B and Figure 24 F-G). As mentioned in the comparison to the formation by Takaishi et al., crab-eating macaque was also excluded as the alignment showed very low similarity and the initial-117 bp unit in human was not found (Figure 18 A). As described before, the primary unit phylogeny detected weak 1 to 9 division nodes however the phylogeny of the initial 117-bp unit formation was consistent even with a cut-off value of 50%, suggesting a probable duplication pattern relevance. Therefore, I suggest that the primary units are the conserved starting duplication unit which further duplicate to form the secondary unit (Figure 18 A, Figure 22 A-D, Figure 23 A-B, Figure 24 A-B).

Starting with the primary units, I then reconstructed the secondary and tertiary units. The order of the primary units of crab-eating macaque was not sequential therefore I could not reconstruct the secondary and tertiary units, and excluded it for further analysis. Human, chimpanzee and orangutan had 23 secondary units and 12 tertiary units; and gorilla had 19 secondary units and 8 tertiary units (Figure 17 and Figure 23 A). I constructed a phylogenetic tree for both secondary and tertiary units for human, chimpanzee, gorilla and orangutan within species (Figure 25 and Figure 26). The description by Paar et al. divides secondary units into ‘n01’ and ‘n02’ types and the tertiary units into ‘v01’ and ‘v02’ types. The phylogeny clustering for the secondary units gathered n01 type, n02 type and the last units for gorilla (except for w04-v01-n01 and w04-v02-n01), orangutan, chimpanzee and human secondary units (Figure 24). Similarly to the phylogeny of the secondary units, the tertiary units clustered v01 type, v02 types and the last units for chimpanzee (except for w03-v01),

gorilla (except for w01-v02, w02-v01, w02-v02 and w03-v01), orangutan, and human tertiary units (Figure 25). In the phylogeny of the secondary and tertiary units, I also observed duplications within species, more common in orangutan and human (Figure 24 and Figure 25).

As for the quartic structure, Takaishi et al. described 6 longest-units and Paar et al. described 5 longest-units. I considered that since the last unit contains only subunits A and B, it was not detected by Global Repeat Maker algorithm developed by Paar et al. (2011) (Figure 20 A-B). I concluded, same as the formation for human by Paar et al. (2011), that the primary units duplicated to form the secondary units that once again duplicated to construct the tertiary units and duplicated once again to form the quartic units. The description by Paar et al. (2011) only detected this formation in human but not chimpanzee, however my analyses detected it to be conserved in chimpanzee, gorilla and orangutan (Figure 16).

2.2.3.6. Mouse hornerin comparison

The structure of the repeated region of hornerin in mouse was described by Makino et al. (2001) as having two types A (5 units) and B (7 units), which consists of around 170 aa. The maximum-likelihood phylogeny analysis with the addition of mouse gave the following results, all mouse repeats clustered apart and divided by either mouse-type A or B subunits, clusters 'α' to 'δ' remain the same, and cluster 'ε' regrouped with human longest-4 supportive with the neighbor-joining analysis (Figure 20 C and Figure 34). In addition, I searched the conservation between mouse units and human longest and initial-117 bp units. I aligned mouse unit 1 A and B to human longest-1. Mouse unit 1 A was aligned to three segments, the first from 590 to704 bp, the second from 625 to 662 bp and the third from 1333 to 1369, and mouse unit B did not show any align match (Figure 18 B). Next, I aligned mouse unit 1 A and B to each of the initial-117 bp units to the mouse unit 1 A and B matched at a low threshold with short areas between 150 to 350 bp (Figure 18 C). I conclude that there is some similarity between human longest-unit and mouse unit A, however the initial-117 bp were not conserved.

2.2.3.7. Chimpanzee and human comparison

Chimpanzee and human repeats have six longest units and have the same length variation. HRNR repeat 1 (Cluster 'δ-4') and repeat 2 (Cluster δ-1) are conserved between human and chimpanzee. I observed two duplications in human branch (cluster 'ε') but none in chimpanzee (Figure 34). I identified the 39 sequences

in both chimpanzee and human with 6 units varying in human. Furthermore, the secondary units and tertiary varied one unit in chimpanzee respectively (Figure 23 A).

2.2.4. Discussion on hornerin repeats

The longest unit and quartic unit length is shared by the descriptions of hornerin of both Takaishi et al. and Paar et al., therefore I used them to understand hornerin evolution. As mentioned on the previous section, the birth-and-death model can be used for the understanding of multigene families evolution and similarly to filaggrin, I observed most of the changes on the longest units in hornerin being synonymous and various lineage specific expansions on the longest and secondary units analysis. The synonymous variations and duplications per species found in hornerin fit the birth-and-death model.

Hornerin has a unique and complex duplicate formation, different from the other SFPTs. The hornerin formation described by Takaishi et al. (2005) {[(39 aa x4 initial) x3 subunits] x6 units} and by Paar et al. (2011) {[(‘‘39 bp x9 primary’’ x2 secondary) x2 tertiary] x5 quartic} differed apart from the ~1404 bp length units (Figure 16).

I analyzed both descriptions and found that the sequence by Paar et al. (2011) of 39 bp or primary unit could be detected in chimpanzee, gorilla, orangutan and crab-eating macaque (Figure 23 A-B). The phylogeny of the primary units per species divided them into 9 clusters in chimpanzee, gorilla and orangutan, however crab-eating macaque did not follow this pattern and was excluded for further analysis (Figure 23 A-B and Figure 24 A-B). The phylogeny of the primary units for the rest of the primates had low bootstrap nodes however it placed nearby primary units on a 1 to 9 pattern which can suggest that primary units duplicated in ninths (Figure 24 A-B). In chimpanzee, gorilla and orangutan, the primary unit duplicated, formed the initial 117-bp which triplicated into the secondary units, and then the secondary units duplicated twice constructing the tertiary units described by Paar et al. The phylogeny for most of human, chimpanzee, gorilla and orangutan secondary units clustered n01 n02 and last units and gathered the tertiary units in v01, v02 and last units. I observed the clustering within species in human and orangutan suggesting unique duplications in these species (Figure 25). Furthermore, I searched for the conservation between human to crab-eating macaque and mouse and found moderate resemblance to crab-eating macaque and low to mouse; which suggests partial conservation with further species (Figure 18 A-C). I conclude that the primary unit is conserved and detectable

in all primates but the duplication clustering order from 1 to 9 was not found in crab-eating macaque. The primary units duplicated to make secondary structure, and the secondary units duplicated twice again to form the tertiary structure described by Paar et al. (Figure 16).

As mentioned for filaggrin, duplication can act as accelerator of evolution. The biological importance of tandem repeats was mentioned by Paar et al. (2011) as a rapidly evolving type of DNA sequences that could contribute to phenotypic differences even between closely related species, such as humans and chimpanzees. The global repeat map by Paar et al. could not find the same high order repeat for human in chimpanzee however my analysis detected the 39 bp length to be conserved in all primates but not the duplication pattern in crab-eating macaque. My analysis also detected unique duplications for all primates, especially for human and orangutan in the secondary and tertiary phylogeny. Neuroblastoma break-point family (NBFP) copy number variability is another example of high order repeat in human (Paar et al. 2011). The NBFP monomer repeat is related to the evolutionary level of higher primates and the high order repeat pattern shows a discontinuous jump in the evolutionary step from 48 monomers in chimpanzee to 165 monomers in human, possibly related to a regulatory function of high order repeats (2011). Indeed, there is no change in the number hornerin copies between chimpanzee and human, however the duplications found in the phylogeny could possibly contribute to a different phenotype. The functional significance of copy number variation of *HRNR* repeats in primates requires further studies but I suggest that, like NBFP and *FLG*, *HRNR* high repeat order pattern might contribute to a different phenotype.

2.3. Filaggrin-2

2.3.1. Introduction

Filaggrin-2 is a 2391 amino acid protein and is found in thymus, stomach, tonsils, testis and placenta. It is probably regulated at the transcriptional level because the corresponding mRNA level is dramatically increased in granular compared with basal keratinocytes (Hsu et al. 2011, Zhihong et al. 2009). In the epidermis, *FLG-2* co-localizes with filaggrin but is expressed differently from that of hornerin or trichohyalin in hair follicles. *FLG-2* inhibition induces pH increase with decrease urocanic acid and pyrrolidone carbozylic acid concentration and *FLG-2* expression is decreased in psoriatic lesions (Pellerin et al. 2013).

The large repetitive region of *FLG-2* contains two types of tandem repeats (Wu et al. 2009). The first set of repeats is homologous (50-77%) to repeats of *HRNR*; I refer to these repeats as 'FLG-2-HRNR-like' repeats. In contrast the second set of repeats is closer to filaggrin (28-39%); I refer to these repeat as 'FLG-2-FLG-like' repeats. I performed separate analyses for each region of *FLG-2*.

2.3.2. Materials and Methods

2.3.2.1 Database sequences of the filaggrin-2

I obtained full-length of filaggrin-2 DNA sequences from the NCBI gene database (http://www.ncbi.nlm.nih.gov/gene/) for the following species *H. sapiens* (NC 000001.11), *P. troglodytes* (NC 006468.4), *M. fascicularis* (NC 022272.1), *P. anubis* (NC 018512.1), *C. jacchus* (NC013913.1) and *M. musculus* (NC 000069.6). NCBI database does not provide sequences for orangutan and macaque, instead I used baboon and marmoset sequences.

2.3.2.2 Identification of repeats in primates

I performed a dot-matrix analysis comparing the repeat sequence in human with those in marmoset, baboon, crab-eating macaque, and chimpanzee (Harrplot 2.1.8 program). The matched sequences were then reanalyzed to search for possible repeat regions and were manually curated (Figure 27). Each of the identified repeats was considered an independent unit for the subsequent analyses.

2.3.2.3 Multiple alignment and phylogenetic analyses

I performed nucleotide multiple sequence alignment, detected the best DNA/Protein model and then constructed neighbor-joining trees and maximum likelihood trees as described in the *FLG* section in MEGA 6.06 (Tamura et al. 2011).

2.3.2.4 Estimation of polymorphic/variant sites, nucleotide diversity, and ratio of synonymous and nonsynonymous sites using the DNAsp5 program

I estimated the *K*a/*K*s ratio, level of purifying selection, nucleotide diversity (π), and the average number of nucleotide differences per site between sequences as mentioned in the *FLG* section in DNAsp5 (Hu and Banzhaf 2008; Librado and Rozas 2009).

2.3.2.5 Likelihood ratio tests for positive selection

Selection on the repeat region of filaggrin-2 was evaluated using the codon substitution models (Codeml) tool in the PAML package (version 4.7). I performed site-based test, branch-based test, and branch-site test as described in the filaggrin section (Yang 1997, 2007; Anisimova et al. 2001; Sabbagh et al. 2013)

2.3. Results

❖ **2.3.3.1 'FLG-2-FLG-like' repeats results**

2.3.3.1.1. Repeat number and similarity

The number of FLG-2-FLG-like repeats identified was 14 for human and macaque, 13 for chimpanzee, 22 for baboon and 8 in marmoset (Figure 27). The length of each repeat unit varied from 225 to 231 (Figure 27). The percentage of similarity between repeats within a species ranged from 73.18% in chimpanzee to 99.11% in baboon and marmoset (Table 17).

2.3.3.1.2. Nucleotide variation within a species

I compared the nucleotide variation for units within a species and the diversity ranged from $1.2 \times 10^{-2} - 1.5 \times 10^{-2}$ (Table 18). The nucleotide variation is comparable to the previously described nucleotide variation examples, poly ubiquitin gene (range = 88.6×10^{-3} to 197×10^{-3}), filaggrin repeated unit (range = 3.2×10^{-2} to 7.7×10^{-2}) and hornerin repeated unit (range = $1.13 \times 10^{-2} - 1.7 \times 10^{-2}$) indicating birth-and-death model. Overall repeat comparison and pairwise repeat comparison identified that most of variations are synonymous which suggest purifying selection (Table 18). FLG-2-FLG-like pairwise comparison repeats had the highest number of non-synonymous pairs. Pairwise comparison detected high non-synonymous

variations; 37 out of 91 pairs in human, 33 of 78 in chimpanzee, 19 of 91 in macaque, 31 of 231 in baboon and 4 of 28 in marmoset (Table 17 and Table 19).

2.3.3.1.3. Phylogeny analysis

I constructed neighbor-joining trees, and maximum likelihood trees using the detected best-fit model (General Time Reversible + Gamma Distributed) (Table 20).

All marmoset-repeats formed a distinct cluster (Cluster named 'Marmoset'). Interestingly, four consecutive repeats at the N-terminal in the rest of the species formed their own clusters (Clusters named 'A', 'B', 'C', and 'D'), indicating the order of these four repeats are conserved from baboon to human (Figure 28 A and C).

Apart from the above mentioned repeats, baboon-repeat 9 and 13 (Cluster named 'R'), and baboon-repeat 10 and 14 (Cluster named 'S') gathered within a species in independent clusters (Figure 28 A and C), suggesting species-specific duplication evolution. The rest of the repeats gathered across species or were not diverged enough to form a cluster.

Baboon and macaque repeats formed several clusters (Clusters 'N', 'O', 'P', and 'Q') (Figure 28 A and C). Similarly, chimpanzee and human repeats formed various clusters (Clusters 'F', 'G', 'H', 'I', 'J', 'K', 'L', and 'M') (Figure 28 A and C).

To identify the duplications and losses events and validate the clusters suggested by the phylogeny analysis, I constructed a species-gene tree. Marmoset branch underwent independent duplications from other primates which resulted in the current repeats, baboon-repeat 9 and 13, and baboon-repeat 10 and 14 are also the result of recent duplications, and chimpanzee lost one repeat which was similar to current human-repeat 10 (Figure 28 E).

2.3.3.1.4. Positive selected codons across species

I searched for signature of positive selection in the repeats across species. While comparing all repeats across species, 7 codons were detected which accounted for FLG-2-FLG-like 2.7% (M2a) and 3.1% (M8) of the codons in one repeat (Table 21). All branches have independent evolutionary pressure (Table 21). While comparing specific branches, Cluster 'Marmoset', repeat-1 or cluster 'A' and repeat-2 or cluster 'B' branches had sites under positive selection (Figure 28 A and Table 21).

❖ **2.3.3.2'FLG-2-HRNR-like' repeats results**

2.3.3.2.1. Repeat number and similarity

The number of FLG-2-HRNR-like repeats was 9 for human and chimpanzee, 7 for macaque and baboon, and 5 in marmoset (Figure 27). The length of each repeat unit varied from 225 to 231(Figure 27). The percentage of similarity between repeats within a species ranged from 76.19% in baboon to 96% in human (Table 17).

2.3.3.2.2. Nucleotide variation within a species

I compared the nucleotide variation for units within a species and the diversity ranged from $1.1 \times 10^{-2} - 1.5 \times 10^{-2}$ (Table 18). The nucleotide variation is comparable to the previously described birth-and-death model such as, poly ubiquitin gene (range = 88.6×10^{-3} to 197×10^{-3}), filaggrin repeated unit (range = 3.2×10^{-2} to 7.7×10^{-2}) and hornerin repeated unit (range = $1.13 \times 10^{-2} - 1.7 \times 10^{-2}$). Overall repeat comparison and pairwise repeat comparison identified that most of variations are synonymous suggesting purifying selection (Table 18). The location of synonymous and non-synonymous variations distributed along the repeat unit and was not specific. Pairwise comparison detected higher non-synonymous variations in 1 out of 21 pairs in baboon (Table 17 and Table 19).

2.3.3.2.3. Phylogeny analysis

I constructed neighbor-joining trees, and maximum likelihood trees using the detected best-fit model (General Time Reversible + Gamma Distributed) (Table 20). The repeat 1 from all species clustered together and was named ‘A’, and repeat 4 from all species except marmoset gathered into a branch named ‘D’. Macaque-repeat 3, chimpanzee-repeat 3, human-repeat 3, chimpanzee-repeat5 and human-repeat 5 clustered together and was named ‘C’, and marmoset-repeat 4 and 5, baboon-repeat and macaque-repeat 7 and chimpanzee-repeat and human-repeat 9 gathered into a cluster named ‘G’. Marmoset-repeat 4 and 5 are the only repeats that gathered within marmoset species (Figure 28 B and D).

Chimpanzee and human repeats gathered together in distinct clusters as follows: chimpanzee-repeat 2 and human-repeat 2 (Cluster named ‘B’), and chimpanzee-repeat 7 and human-repeat 7 (Cluster named ‘F’) (Figure 28 B and D).

Additionally, I constructed a species-gene tree and identified the duplications and losses in the phylogeny. Marmoset branch underwent recent duplications for marmoset-repeat 4 and 5, moreover marmoset, baboon and macaque lost a repeat similar to repeat ‘E’, and baboon also lost repeat ‘C’ (Figure 28 F).

2.3.3.2.4. Positive selected codons across species

I searched for signature of positive selection in the repeats across species. No positively selected codon was detected across branches and in specific branches (Figure 28 A and Table 21). All branches are evolving under an independent selective pressure (Figure 28 A and Table 21).

2.3.3.3. Mouse filaggrin-2 comparison

The mouse filaggrin-2 gene structure differs from the one described in human. Human filaggrin-2 has a N-terminal part and the first repeated region similar to hornerin, followed by a second repeated region and a C-terminus similar to filaggrin. In contrast, mouse filaggrin-2 structure lacks the first repeated region similar to hornerin but conserves the repeated region similar to filaggrin (Hansmann et al. 2012) (Figure 35). Mouse and human filaggrin-2 are processed by calpain-1 and are localized in keratinizing epithelia in the upper cell layers.

I used the repeats described by Hansmann et al. (2012) and compared the FLG-2-FLG-like repeats from mouse to human. The number of repeats in mouse was 14 and the length of the repeat was similar to the length of repeats in primates (219-240 bp in mouse and 225-231 bp in primates) (Figure 35). I constructed a phylogenic tree using primates and mouse repeats and I observed that mouse repeats cluster apart from primates which suggests low conservation of filaggrin-2 (Figure 28 G).

2.3.3.4. Chimpanzee and human comparison in *FLG-2*

Chimpanzee and human have the same length of FLG-2-FLG-like repeats and FLG-2-HRNR-like repeats, however the number of repeats of FLG-2-FLG-like is 13 in chimpanzee and 14 in human. *FLG-2* repeats between chimpanzees and human are the most conserved in the SFTP-genes. The only difference between human and chimpanzee is the loss of human FLG-2-FLG-like repeat 10 in chimpanzee. Therefore, human FLG-2-FLG-like repeat 11 clustered with chimpanzee FLG-2-FLG-like repeat 10, and human FLG-2-FLG-like repeat 10 formed separate branch (Figure 28 A). I did not identify any duplication between chimpanzee and human (Figure 34).

2.3.4. Discussion on filaggrin-2 repeats

FLG-2-FLG-like repeats are similar to the nucleotide sequence of *FLG* repeats ranging from 28-39% and have been described to have similar functions. Both, colocalized in the epidermis, are degraded by calpain-1 and their amino acids composition affects the pH of skin. Taking this into consideration, I hypothesized that both repeated regions might share a similar evolutionary process. Indeed, both of

them have high nucleotide variation corresponding to birth-and-death model of evolution with most variations being synonymous and a high similarity (>73%) across species. However, I found that FLG-2-FLG-like repeats have a lower percentage of similarity within a species, a lower nucleotide variation, and even if most of the variations are synonymous, pairwise comparison detected more non-synonymous pair variations compared to that of *FLG* (Table 4, Table 5, Table 17, and Table 18). Also, the phylogeny in FLG-2-FLG-like only partially resembles the one described for *FLG*. Within a species cluster 'marmoset' in FLG-2-FLG-like was similar to cluster 'orangutan' and 'macaque' in *FLG*. To further confirm this, I additionally identified the filaggrin repeats from baboon and marmoset sequences of NCBI database, however baboon sequence had undetermined regions. Baboon had 2 repeats and marmoset 6. I constructed neighbor-joining trees, and maximum likelihood trees using the detected best-fit model Tamura-Nei 93 substitution model + Gamma distribution (Table 2). Baboon repeats gathered with macaque suggesting that even after divergence some similarity is shared; in contrast marmoset repeats cluster apart (Figure 15 A-B).

Recent duplications only occurred in baboon (Cluster 'R' and 'S' in FLG-2-FLG-like) but were found in all species in *FLG* (Cluster 'B', 'C', and 'D'). FLG-2-FLG-like repeats 1 to 4 or clusters 'A' to 'D' gathered but the gathering of initial repeats was not found in *FLG* (Figure 7, Figure 8 and Figure 28 A). I suggest FLG-2-FLG-like repeated region and *FLG* repeated region do not evolve in the same manner.

Although *FLG-2* expresses differently from hornerin in hair follicles, the similarity between these genes is high, which ranges from 50-77%. FLG-2-HRNR-like repeats within a species are >76% similar and just a few of pairs have a higher non-synonymous variation, however the nucleotide variation is similar to hornerin (Table 12, Table 13, Table 17, and Table 18). The phylogenetic analysis between FLG-2-HRNR-like and hornerin are not alike. FLG-2-HRNR-like trees do not have a big cluster within a species as observed for hornerin of orangutan and macaque (clusters 'α' and 'δ-3) (Figure 18 A and Figure 28 B). I suggest FLG-2-HRNR-like repeats and *HRNR* repeats within a species seems to be evolving on a similar manner, however across species they are evolving separately.

I expected that FLG-2-FLG-like and FLG-2-HRNR-like regions would evolve similarity to *FLG* and *HRNR*, respectively. However my findings suggest that the

nucleotide variation along a repeat is independent of nucleotide composition, number of repeats and length.

Filaggrin-2 is the unique gene that included two different repeated regions with similarity to other SFTPs. Both types of repeats have a comparable nucleotide variation along their repeats, however the FLG-2-FLG-like repeats had a remarkably high number of pairs with non-synonymous variations (Table 18 and Table 19). The phylogenic reconstructed trees found recent duplications in marmoset and baboon in FLG-2-FLG-like repeats, but only one recent duplication in marmoset FLG-2-HRNR-like. Additionally, I observed that the first repeats in both regions are more likely to be grouped as found in FLG-2-FLG-like repeat 1, 2, 3 and 4, and FLG-2-HRNR-like repeat 1 (Figure 28 A-B). This finding suggests that in *FLG-2* repeated regions, the repeats closer to the N-terminal are conserved and the latter ones can be duplicated or lost. I conclude that both of the repeated regions contained in *FLG-2* are evolving separately from each other.

2.4. Repetin and trichohyalin

2.4.1 Introduction

Repetin consists of 784 amino acids. Human has 28 repeats of 12 amino acids that are rich in glutamine 23.8%, serine 15.5% and glycine 14%. It shows a high expression in keratinized area of filiform papilli of the tongue, overlapping expression with loricrin, and a low expression in the trunk and foreskin epidermis. Additionally it cross-links with trichohyalin in the inner root sheath of human hair and to trichohyalin and loricrin in mouse (Huber et al. 2005).

Trichohyalin has a low expression in sparse keratinocytes in the granular and cornified layers of the epidermis and in the filiform papillae of the tongue epithelium (Kypriotou et al. 2012). It is abundant in the inner root sheath cells and in the medulla of the hair follicle (Henry et al. 2012). Trichohyalin consists of various domains that contain short repetitive residues. *TCHH* sequences were not used in the analyses because the repeated region was separated into domains rather than repeats and those domains include repeats that are too divergent.

2.4.2. Materials and Methods

2.4.2.1 Database sequences of repetin

I obtained full-length of repetin DNA sequences from the NCBI gene database (http://www.ncbi.nlm.nih.gov/gene/) for the following species *H. sapiens* (NC 000001.11), *P. troglodytes* (NC 006468.3), *G. gorilla* (NC 018424.1), *P. abelii* (NC 012591.1), *M. fascicularis* (NC 022272.1) and *M. musculus* (NC 000069.6).

2.4.2.2 Identification of repeats in primates

I performed a dot-matrix analysis comparing the repeat sequence in human with those in crab-eating macaque, orangutan, gorilla and chimpanzee (Harrplot 2.1.8 program). The matched sequences were then reanalyzed to search for possible repeat regions and were manually curated (Figure 29). Each of the identified repeats was considered an independent unit for the subsequent analyses.

2.4.2.3 Multiple alignments and phylogenetic analyses

I performed nucleotide multiple sequence alignment, detected the best DNA/Protein model and then constructed neighbor-joining trees and maximum likelihood trees as described in the *FLG* section in MEGA 6.06 (Tamura et al. 2011).

2.4.2.4 Estimation of polymorphic/variant sites, nucleotide diversity, and ratio of synonymous and nonsynonymous sites using the DNAsp5 program

I estimated the *K*a/*K*s ratio, level of purifying selection, nucleotide diversity (π), and the average number of nucleotide differences per site between sequences as mentioned in the *FLG* section in DNAsp5 (Hu and Banzhaf 2008; Librado and Rozas 2009).

2.4.2.5 Likelihood ratio tests for positive selection

Selection on the repeat region of repetin was evaluated using the codon substitution models (Codeml) tool in the PAML package (version 4.7). I performed site-based test, branch-based test, and branch-site test as described in the filaggrin section (Yang 1997, 2007; Anisimova et al. 2001; Sabbagh et al. 2013)

2.4.3. Results

2.4.3.1. Repeat number and similarity

Repetin repeats are 28 in human and macaque, 19 in chimpanzee, 27 in gorilla and 16 in orangutan (Figure 29). The length of all repeat units was 36 bp (Figure 29). The percentage of similarity between repeats within a species ranged from 36.67% in human, chimpanzee, orangutan and macaque, to 100% in human, gorilla, orangutan and macaque (Table 22).

2.4.3.2. Nucleotide variation within a species

I compared the nucleotide variation for units within a species and the diversity ranged from 1.9×10^{-2} to 2.6×10^{-2} (Table 23). The nucleotide variation is comparable to the previously described birth-and-death model examples, poly ubiquitin gene (range = 88.6×10^{-3} to 197×10^{-3}) and filaggrin repeated unit (range = 3.2×10^{-2} to 7.7×10^{-2}). Overall repeat comparison and pairwise repeat comparison identified that most of variations are synonymous in concordance with purifying selection. The location of synonymous and non-synonymous variations distributed along the repeat unit and was not specific. Pairwise comparison detected higher non-synonymous variations in 17 out of 378 pairs in human, 12 of 171 in chimpanzee, 22 of 351 in gorilla, 5 out of 120 in orangutan, and 21 out of 378 pairs in macaque (Table 22 and Table 24).

2.4.3.3. Phylogeny analysis

I constructed neighbor-joining trees, and maximum likelihood trees using the detected best-fit model (Kimura's two-parameter substitution model + Gamma

distribution) (Table 25). All of the repeats either gathered across species or were too similar to form a cluster. The first consecutive repeats formed their own clusters (Cluster 'A', 'B' and 'C') indicating the order of these three repeats is conserved from macaque to human. Human, chimpanzee and gorilla repeats formed four clusters (cluster 'D', 'E', 'F', and 'J'), human and gorilla formed three clusters (cluster 'H', 'I', 'K' and 'M'), human and orangutan formed cluster 'H' and chimpanzee and macaque cluster was named 'L' (Figure 30 A-B). The constructed species-gene tree identified no duplications within a species and many losses in the phylogeny across species. The original ancestor of repetin in primates duplicated and diverged into the ancestor of repetin-1 (cluster 'A') and repetin-3 (cluster 'B'), and the ancestor of the rest of repeats. A duplication event created repetin-1 (cluster 'A') and repetin-3 (cluster 'B'), and additional duplications in repetin-3 (cluster 'B') created repetin-3 in macaque and orangutan, human and gorilla, and chimpanzee. Orangutan species lost repeat 'L', gorilla species lost repeat 'L' and 'M', chimpanzee lost repeat from 'H' to 'M', and human lost repeat 'L' (Figure 30 C).

2.4.3.4. Positive selected codons across species

While comparing all repeats across species, 1 codon was detected. I further analyzed the specific branch in which the codon locates and it was detected in branch repeat-3 or 'B' (Table 26).

2.4.3.5. Mouse repetin comparison

I compared the repeats from mouse to human and constructed a phylogenic tree. The length of the repeat is the same as primates (36 bp). However the number of repeats in mouse is 48, which is more than those in primates. The only repeat that shares similarity with primates is mouse-repeat 1 which resembles with repeat 2 in primates. Furthermore, the phylogenetic analysis gathered many mouse repeats together in clusters. The mouse repeats that gather together are Mouse repeat 10 and 13, repeat 18 and 42, repeat 19, 22, 25, 28, 43 and 46, repeat 20, 23 and 44, repeat 21, 24 and 27, repeat 26, 29 and 47, and repeat 35, 36 and 39. The conservation of the length of repeat supports the importance of the amino acid positions and repeat length rather than the number of repeats (Figure 30 D and Figure 35).

2.4.3.6. Chimpanzee and Human comparison

Chimpanzee has 19 repeats and human 28. Clusters named 'A', 'B', 'C', 'D', 'E', 'F', and 'G' gathered between human and chimpanzee. Chimpanzee lost repeats

‘H’, ‘I’, ‘J’, ‘K’, and ‘M’, and Human lost repeat ‘L’. I did not identify any duplication between chimpanzee and human (Figure 34).

2.4.4. Discussion on repetin and trichohyalin repeats

Repetin gene is also found in mouse, however it is localized in chromosome 3. *RPTN* repeats of human and mouse have positional conservation of glutamine at positions 1, 5 and 7, glycine at positions 6 and 12, serine at positions 8 and 9, and histidine at position 10 (Krieg et al. 1997). The predicted secondary structure for the central domain of human and mouse is a β-sheet structure which probably requires a conserved arrangement of specific residues within the 12 amino acid repeat for stability and structural functions (Huber et al. 2005). Furthermore, repetin does not undergo proteolytical processing during epidermal differentiation that further suggests that amino acid positions and repeat length rather than the total number of repeats have been conserved during evolution (Huber et al. 2005). I identified 1 positive selected codon (located in codon 2 of human repetin sequence) in repeat 3 cluster named ‘B’ (Table 26), this codon was not conserved with mouse repetin sequences and I speculate the variation in this codon could have a functional effect.

As mentioned before, *TCHH* repeated segments are irregular and were not analyzed. However, the *TCHH* central region is predicted to form a flexible single stranded α-helical domain that associated with keratin filaments to reinforce the hair shaft, and single nucleotide polymorphism in *TCHH* is associated with straight hair in Europeans and different strains of hair in sheep hair (Henry et al. 2011, Pospiech et al. 2015). I suggest that the change of the phenotypic variation from a single aminoacid variation in *TCHH* might have a similar result for repetin however I could not identify any report that supports my hypothesis.

2.5. Cornulin

2.5.1. Introduction

Cornulin encodes a protein of 495 amino acids. It is rich in glutamine and threonine. It is detected in fetal brain, lung, kidney, uterus, skeletal muscle and heart but preferentially in esophagus, fetal bladder, scalp and foreskin keratinocytes. *CRNN* prevents apoptosis against deoxycholate cytotoxicity. Cornulin has been implicated in esophageal squamous cell carcinoma and atopic eczema. Loss of *CRNN* expression correlates with deep and long tumor invasion depth and length, advanced lymph node metastasis and poor survival (Contzler et al. 2009 and Hsu et al. 2014).

2.5.2. Materials and Methods

2.5.2.1 Database sequences of cornulin

I obtained full-length of cornulin DNA sequences from the NCBI gene database (http://www.ncbi.nlm.nih.gov/gene/) for the following species *H. sapiens* (NC 000001.11), *P. troglodytes* (NC 006468.3), *G. gorilla* (NC 018424.1), *P. abelii* (NC 012591.1), *M. mulatta* (NC 027893.1) and *M. musculus* (NC 000069.6).

2.5.2.2 Identification of repeats in primates

I performed a dot-matrix analysis comparing the repeat sequence in human with those in crab-eating macaque, orangutan, gorilla and chimpanzee (Harrplot 2.1.8 program). The matched sequences were then reanalyzed to search for possible repeat regions and were manually curated (Figure 31). Each of the identified repeats was considered an independent unit for the subsequent analyses.

2.5.2.3 Multiple alignment and phylogenetic analyses

I performed nucleotide multiple sequence alignment, detected the best DNA/Protein model and then constructed neighbor-joining trees and maximum likelihood trees as described in the *FLG* section in MEGA 6.06 (Tamura et al. 2011).

2.5.2.4 Estimation of polymorphic/variant sites, nucleotide diversity, and ratio of synonymous and nonsynonymous sites using the DNAsp5 program

I estimated the *K*a/*K*s ratio, level of purifying selection, nucleotide diversity (π), and the average number of nucleotide differences per site between sequences as mentioned in the *FLG* section in DNAsp5 (Hu and Banzhaf 2008; Librado and Rozas 2009).

2.5.2.5 Likelihood ratio tests for positive selection

Selection on the repeat region of cornulin was evaluated using the codon substitution models (Codeml) tool in the PAML package (version 4.7). I performed site-based test, branch-based test, and branch-site test as described in the filaggrin section (Yang 1997, 2007; Anisimova et al. 2001; Sabbagh et al. 2013)

2.5.3. Results

2.5.3.1. Repeat number and similarity

Cornulin repeats include 2 in human, orangutan and macaque, 4 in chimpanzee and 1 in gorilla (Figure 31). The length of all repeat units was 180 bp (Figure 31). The percentage of similarity between repeats within a species ranged from 89.44% in orangutan to 98.33% in chimpanzee (Table 27).

2.5.3.2. Nucleotide variation within a species

I compared the nucleotide variation for units within a species and the diversity ranged from 0.5×10^{-2} to 1.1×10^{-2} (Table 28). Cornulin had the lowest nucleotide diversity within a species from all SFTPs. The nucleotide variation is high and comparable to the previously described birth-and-death model examples, poly ubiquitin gene (range = 88.6×10^{-3} to 197×10^{-3}) and filaggrin repeated unit (range = 3.2×10^{-2} to 7.7×10^{-2}). Overall repeat comparison and pairwise repeat comparison identified that most of variations are synonymous in concordance with purifying selection. The location of synonymous and non-synonymous variations distributed along the repeat unit and was not specific. Pairwise comparison did not detect any pairs with higher non-synonymous variations.

2.5.3.3. Phylogeny analysis

I constructed neighbor-joining trees, and maximum likelihood trees using the detected best-fit model Tamura-Nei 93 substitution model + Gamma distribution (Table 29). All of the repeats gathered across species and divided into two clusters. Cluster ‘A’ included human, chimpanzee, orangutan and macaque repeat-1 and chimpanzee repeat-2 and repeat-3, and cluster ‘B’ included human, orangutan and macaque repeat-2, chimpanzee repeat-4 and gorilla repeat-1 (Figure 32 A-B). Additionally, the constructed species-gene tree identified one duplication in chimpanzee and one loss of repeat ‘A’ in gorilla (Figure 32 C).

2.5.3.4. Positive selected codons across species

While comparing all repeats across species and specific branches, no positive selected codon was detected (Table 30).

2.5.3.5. Mouse cornulin comparison

I compared the *CRNN* repeat sequences from mouse to human, and constructed a phylogenic tree. The number of repeats found in mouse was 2, same as macaque, orangutan and human, and the length of the repeat was shorter to those in primates (177 bp in mouse and 180 bp in primates). Mouse repeats cluster apart from primates which suggests low conservation of cornulin between mouse and primates (Figure 32 D and Figure 35).

2.5.3.6. Chimpanzee and Human

Chimpanzee and human repeats have the same length, however the number of repeats is 4 in chimpanzee and 2 in human. Chimpanzee was the only species that had one duplication in primates (Figure 34). Cluster 'A' included human, chimpanzee, orangutan and macaque repeat-1 and chimpanzee repeat-2 and repeat-3, and cluster 'B' included human, orangutan and macaque repeat-2, chimpanzee repeat-4 and gorilla repeat-1 (Figure 32 C).

2.5.4. Discussion on cornulin repeats

CRNN, functions between the signal cascade initiated by exposure and the release of calcium leading to apoptosis, and its overexpression in oral cancer cells cause a decline in cell proliferation by G1/S phase arrest (Hsu et al. 2014). The role of *CRNN* in the cell cycle further supports my hypothesis of the strict amino acid conservation. In the paper "Adaptive Evolution of a Stress Response Protein", *CRNN* was suggested to evolve by adaptive evolution, with the role in skin being the most significant, actively participating on the host-pathogen arms race (Little et al. 2007). However, I found this not to be the case. First, the authors used the incorrect repeated sequence. Xu et al. (2000) described the sequence found in human repeat-1 (TQTVEQDSSHQTGRTSKQEATNDQNRGTETHGQGRSQTSQAVTGGHAQIQA GTHTQTP) and repeat-2 (TQTVEQDSSHQTGSTSTQTQESTNGQNRGTEIHGQGRSQTSQAVTGGHTQIQ AGSHTETV). Little et al. (2007) did not describe how they delimited each repeat unit. I aligned the protein sequences described by Xu et al. with the ones mentioned by Little et al. using Clustal W. in Mega 6 and observed that Little et al. sequences were misaligned (Figure 33). The incorrect sequence of the repeated region doubts the results of nucleotide variation and positive selection. Seconds, in addition to the repeated region, they included the N-terminal domain and C-terminal domain which

has been known to be divergent in all SFTP across species. Third, like in my study, they searched for positive selected codons using likelihood ratio tests for sequences, including the N-terminal and C-terminal domains, and found 8% of codons being positively selected; most of them located on the C-terminal. They also performed the same analyses considering repeats as units and found almost half of the sites under positive selection contradicting the more than 90% similarity across repeats. Fourth, none of their pairwise Ka/Ks ratio result was higher than 1. I considered that the adaptive evolution description is mostly based on the variation on the C-terminal domain and does not affect the function of cornulin repeats. Additionally, later studies demonstrated the importance of cornulin function on the cell cycle, rather than skin, further supporting my results.

Chapter three

FILAGGRIN VARIATION IN ECUADORIAN PEDIATRIC POPULATION

3.0 Chapter Summary

In human, the number of *FLG* complete repeats varies from 10 to 12 among individuals and populations. The variation in the number of repeats affects filaggrin degradation products with fewer repeats resulting in dry skin. Additionally, loss-of-functions variants have been strongly associated with atopic dermatitis and vary depending on population. Most of *FLG* variants have been described in Northern European and Asian populations; although Latin America has higher prevalence of atopic dermatitis there are no reports of *FLG* variants. Ecuador has the second highest country prevalence for children between 6 to 7 years. Therefore in this chapter I focus on *FLG* alterations associated to atopic dermatitis in pediatric Ecuadorian population.

3.1 Introduction

Disturbances in cornified envelope (CE) proteins are closely related to defective skin barrier and disorders, such as atopic dermatitis (AD) (Trzeciak et al. 2016). AD predisposition has been strongly associated with *FLG* mutations. *FLG* mutations alter the process of filaggrin formation and it is further decomposition resulting in dry skin (Cabanillas and Novak 2016).

AD is the most common form of eczema in childhood, manifestations as a chronic, relapsing itchy rash that usually starts in early life and, in many children, wanes in severity later in childhood (Odhiambo et al. 2009). Early-onset AD is caused by a skin barrier abnormality frequently caused by filaggrin gene (*FLG*) mutations with a high frequency in Northern Europeans and Asians (Thyssen et al. 2014 and 2015). The prevalence of AD was positively correlated with a farther distance from the equator line and negatively correlated with mean annual outdoor temperature (Thyssen et al. 2014 and 2015). However globally, Ecuador (latitude 0) has the second highest country in the prevalence of AD (22.5%) for children between 6 to 7 years, and the eight highest one for children between 13 to 14 years (16.5%) (Odhiambo et al. 2009). Furthermore, within Latin American cities, Quito has the highest risk of current symptoms of eczema in children between 6-7 years and the fifth highest risk for children between 13-14 years (Sole et al. 2010); but there are no studies analyzing a genetic relation.

As mentioned before, filaggrin copy number variation (CNV) and loss-of-function (LOF) mutations have been associated with AD. Copy number variation refers to the 10-12 nearly identical repeats allelic variation in exon 3. Fewer copies contribute to the risk of AD; with a correlation between the number of *FLG* repeats, and with the amount of filaggrin degradation products and each additional repeat decreases the risk by a 0.88 in odds ratio (Brown et al. 2012, Li et al. 2013).

Loss-of-functions (LOF) mutations have been strongly associated with AD. In general populations, about 10% of Northern Europeans and 5% of Asians carry heterozygous *FLG* LOF. In AD, more than 40 LOF mutations in *FLG* have been associated and many of them are specific to a particular population. In Northern Europeans more than 80% of LOF mutations are accounted by two mutations (R501X and 2882del4). In contrast, in Asian populations, eight different mutations are associated (Akiyama 2010, Thyssen et al. 2014).

In this chapter, I aimed to quantify the CNVs in filaggrin repeated region, compare the frequency of the most common LOF mutations and identify population specific mutations between AD patients and controls from Ecuadorian children.

3.2 Materials and Methods

3.2.1 Human samples

One hundred seventy five cases with AD and 154 control samples come from Ecuadorian children between the ages of 2 months to 18 years old (average age 8.35 years old for cases and 11.04 for controls) and were recruited at Centro de la Piel (CEPI), Quito-Ecuador. CEPI is a specialized dermatology center that focuses on AD. Dermatologists performed the diagnosis of AD according to the criteria of Hanifin and Rajka (Rajka 1989), acquired medical information, extracted blood and transferred it to a FTA card (Whatman, Florham Park, NJ, USA). Controls with no personal or family history of AD were recruited. Variables acquired were age, sex, race and allergic history for cases and controls. In cases, additional variables were acquired as the age of diagnoses, sub-type, current treatment and complication of AD. All the participants' parents provided written informed consent. CEPI follows the Sociedad Ecuatoriana de Bioetica regulations and the study was approved by the Ethics Committee of Centro de la Piel (CEPI 141014). The Ethics Committee of National Institute of Genetics (nig 1419, 2014.11) approved the study protocol.

3.2.2 Determination of atopic disease severity and sub-type

Severity of AD was assessed using the SCORing Atopic Dermatitis index (SCORAD index).

The SCORAD scale categorized AD into mild (<25), moderate (25-50) or severe (>50). Fifty-three cases were categorized as mild, 93 were moderate and 29 were severe.

3.2.3 DNA card extraction

Dry blood card extraction was performed using DNA purification from dried blood spots (QIAamp DNA Mini Kit) with the following modifications for each sample. I started with 2 tubes with 1/12 piece of card, added 180 µl of buffer ATL, and incubated the tubes at 85°C for 15 minute following 10 minutes at room temperature. Next, I added 20 µl of proteinase K, vortex the tubes, and incubated at 56°C for 1 hour. Then, I centrifuge the tubes for 1 min, added 200 µl of buffer AL, vortex, and incubated at 70°C for 10 min. After the incubation, I took out the FTA card, added

200 µl of ethanol and mixed the two tubes of each sample. Next, I added to the single tube 500 µl of buffer AW1, centrifuged for 1 minute, discarded the supernatant, did this step once again, and continue the protocol by adding 500 µl of buffer AW2, centrifuging for 1 minute, and discarding the supernatant. Finally, I centrifuged the tube for 3 minutes, added 30 µl of buffer AE and centrifuge for 1 min.

3.2.4 CNV detection by PCR

I performed long-range PCR for the repeated region of *FLG* for cases and controls. Primers were designed outside the repeated region and from repeat 7 up to the last repeat and are summarized in Table 1. PCR reactions were performed using a PrimeSTAR® GXL DNA Polymerase kit, with a total reaction volume of 10 µl that included: genomic DNA (20 ng), PrimeSTAR GXL 5X buffer, dNTP mixture (200 µM), forward and reverse primers (0.2 µM each), PrimeSTAR GXL polymerase (0.25 U/10 µl), and water. The PCR conditions were as follows: 94°C initial denaturation for 2 min, followed by 30 cycles each of denaturation (98°C for 10 s) and elongation (68°C for 10 min). Detection of CNV was observed by gel electrophoresis for the PCR product from repeat 7 to the last repeat.

3.2.5 Sequencing of *FLG* by MiSeq

The MiSeq system (Illumina, San Diego, California, USA) was employed to generate high-throughput short reads for the identification of sequence variants. The DNA libraries were sequenced on the MiSeq platform with 350- and 250-bp paired-end modules. I performed sequencing only from the repeated regions. I sequenced 175 cases and 154 controls.

3.2.6 NGS data processing and variant calling

The reads containing the Illumina adapter sequences were trimmed using Trimmomatic (Bolger et al. 2014). After trimming the sequence reads were aligned to human reference genome (hg19) via BWA version 0.7.13 (Li and Durbin 2009). The information was compressed into binary form (BAM format), sorted by using SAMtool version 1.3 (Li et al. 2009), and processed by Picard tools. Base quality recalibration and single nucleotide variants (SNVs), insertions and deletions (indels) were detected by GATK (McKenna et al. 2010, DePristo et al. 2011). Bioinformatics pipeline can be seen in Ahmadloo et al. (2017).

3.2.7 Functional annotation of identified variants

Known and novel variants were categorized according to NCBI dnSNP. The potential effects of the mutation were estimated by the online prediction PolyPhen2 (Adzhubei et al. 2010).

3.2.8 Statistical analyses

Possible association of *FLG* mutation with AD was evaluated by multiple logistic regression between the affected status and non-synonymous, frame-shift and stop-codon variation found in the Miseq data.

I estimated odds ratio and conducted χ^2 test for confirming the frequency differences of repeat-allele and repeat number between AD patients and normal controls, and to determine clinical relevance. All analyses were performed using R Development Core Team (2008). A p-value of less than 0.05 was considered as statistical significant.

3.3 Results

3.3.1 CNV detection

I searched for filaggrin repeat variation as either a duplication of 8^{th} and/or 10^{th} repeat units in cases and controls. In both groups, the most prevalent was 12 repeat-allele 43.51% in cases and 48.87% in controls) and the total number of repeats was 22 (30.52% in cases and 30.29% in controls). I calculated odds ratio and χ^2 in comparison to the most common variations for both repeat-allele and repeat-number but I did not detected any significant association for cases or controls (Table 31 A and B).

3.3.2 CNV and SCORAD

SCORAD scale is used to classify the severity of AD. I compared the repeat variation to the previously clinical categorized mild, moderate or severe SCORAD. The most prevalent repeat-allele for mild, moderate and severe cases was 12. As done before for repeat variation in cases and controls, I calculated odds ratio χ^2 between 10, 11 and 12 repeat allele to the mild, moderate and severe cases however there was no significant difference (Table 32 A and B).

3.3.3 Analysis of common European *FLG* mutations

The most prevalent mutations in Northern Europeans are 2882del4, R501X, R2247X, 3702delG and S3247X (Sandilands et al. 2007). I searched for the 5 mutations and identified one heterozygote case (Allele frequency = 0.004) for R501X and none for the rest of the mutations in cases or controls (Table 33).

3.3.4 Analysis of Ecuadorian *FLG* mutation

Since none of the most common European mutations were significantly present in my samples, I searched for other variations in Ecuadorian population. Two novel non-synonymous variations, E2250Q (Frequency cases = 0.15 controls = 0.08) and E2652D (Frequency cases = 0.27 controls = 0.18), were identified. I calculated the allele frequency in cases and controls and all of them were significantly associated with AD (E2250Q p-value = 0.003 and E2652D p-value = 0.014). (Table 33 and 34).

3.4 Discussion on filaggrin variation in Ecuadorian pediatric population

Two types of filaggrin alterations are associated with AD; first, the variation in the number of repeats, and second, LOF mutations. I analyzed both types of alterations in cases and controls from Ecuadorian pediatric population as Ecuador has the second highest prevalence of pediatric AD worldwide.

Filaggrin repeat variation differs between populations. In Irish population, the most prevalent allele was 11 (Brown et al. 2012). In Korean population, the most common total repeat-number was 22 for cases and 24 for controls and in Russian population; the most prevalent was 20 for both groups (Gimalova et al. 2016, Li et al. 2016). In Ecuador, I identified that the most prevalent repeat-number was 22 for cases and controls (Table 31 B).

Fewer filaggrin repeats were associated with fewer amounts of filaggrin breakdown products in Irish population with AD and with the risk of AD in Korean population (Brown et al. 2012, Li et al. 2016). However, the fewer repeats association was not observed in Russian population of Russian and Tatar origin (Gimalova et al. 2016). Similarly to the Russian population, my analysis did not associate fewer filaggrin repeats with the risk of AD (Table 31 A). I concluded that the variation in filaggrin repeats is not associated with atopic dermatitis in my samples and I focused on the mutations of filaggrin.

Five filaggrin mutations (2882del4, R501X, R2247X, 3702delG and S3247X) have been strongly associated with AD in European population (Sandilands et al. 2007). Latin American populations are the result of past admixture between European, African and Amerindian population (Salzano and Sans 2014). Therefore, I searched whether the most common filaggrin variants are prevalent in Ecuadorian population. I analyzed the frequency of these variants in Ecuadorians and identified one heterozygote-case for the R501X variation and none for the other two mutations

(Table 33). The most common European variations are not related to AD in my samples, but I identified 2 significantly associated non-synonymous variations in cases. The 2 non-synonymous variations were categorized as possibly damaging by polyphen (Table 34) (Adzhubei et al. 2010).

Population specific filaggrin variants were associated with AD in different populations (European variations vs Asian variations). This population specific pattern could explain the low frequency of the most common European mutations in Ecuadorians. Additionally, the higher prevalence of LOF mutations was previously observed on a geographical gradient distribution (Cascella et al. 2011). This geographical gradient distribution suggests that LOF is latitude-dependent with more LOF mutations found in higher latitudes, especially in northern European countries (Cascella et al. 2011). The increased prevalence of mutations found in northern Europeans, regardless of having AD or not, was suggested to insure adequate vitamin D3 status in high latitudes locations and in cultures where seafood was not extensively consumed (Thyssen et al. 2014). Ecuador location at latitude 0 could explain the absence of the filaggrin LOF mutations meanwhile non-synonymous variations could damage the functionality of filaggrin resulting in the high frequency of AD.

I conclude that the repeat variation was not associated with the risk of AD and that the most common associated European mutations have a low frequency in Ecuadorian population. Instead, I identified 2 new non-synonymous damaging variations significantly associated in AD Ecuadorian cases.

Chapter 4

Overall Discussion

Birth-and-death model and concerted model have been used to explain multigene families evolution. The main difference between the concerted and birth-and-death models of evolution is the high levels of intragenic nucleotide diversity that are only found in the latter (Nei and Rooney 2005; Austen and Kobayashi 2007; Eirín-López et al. 2012; Sabbagh et al. 2013). Overall, the nucleotide variation for each gene from SFTPs ranged similarly independently from the species or genes (Table 5, Table 13, Table 18, Table 23 and Table 28) and was comparable with that of previously described birth-and-death model as an example for polyubiquitin genes (range = 88.6 $\times 10^{-3}$ to 197 $\times 10^{-3}$) (Nei et al. 2000; Austen and Kobayashi 2007). This suggests that SFTP family repeats have evolved under the birth-and-death model. In addition, most of variations are synonymous (Table 5, Table 13, Table 18, Table 23 and Table 28) which is the consequence of birth-and-death model.

It has previously been reported that gene-associated tandem repeats act as an accelerator of evolution by generating variation in structure and functionality (Fondon and Garner 2004). Under the birth-and-death model, duplicates vary by silent nucleotide variations, which, with enough divergence time, can lead to lineage-specific expansions. I detected lineage-specific expansions in the repeats of filaggrin, filaggrin-2, hornerin and cornulin using reconciled trees (Figure 9, Figure 18 D, Figure 28 E and F, Figure 30 C and Figure 32 C) and was further corroborated by including the repeats of mouse in the reconciled tree (Figure 10, Figure 18 C, Figure 28 G, Figure 30 D, Figure 32 D and Figure 35).

Many multigene families have been identified to undergo birth-and-death model. One example is the chemosensory system family of the insect (Vieira et al. 2007). This multigene family can be classified into 2 main functional groups, the odorant-binding (OBPs) and chemosensory (CSPs) protein, and the chemosensory receptors. In insects, there are 3 chemosensory receptor gene families: the olfactory (ORs), gustatory (GRs) receptors and the inotropic receptors (IRs). The size of the chemosensory gene family differs markedly across species and the disparity can provide an insight into the role of natural selection and adaptation with gene family size. Additionally, the CSPs have a large number of gene gains, gene losses and

pseudogenization events although these events differ among gene families. Similarly in my findings of SFTPs, the number of repeats, and duplication and loss events vary across species and genes independently (Figure 9, Figure 18 D, Figure 28 E and F, Figure 30 C and Figure 32 C). Also in the CSPs family, the number of orthologous groups gradually decreased with increasing diverging time with several gene expansions and contractions occurring across large but not across short evolutionary times. I analyzed each SFTP gene independently and observed similar species-specific duplications between primates with enough divergent time (Figure 9, Figure 18 D, Figure 28 E and F, Figure 30 C and Figure 32 C). Finally, in the olfactory multigene family there is no evidence for a major impact of gene conversion in the evolution of paralogous genes which is the same case for the SFTPs as repeats are located within an exon (Figure 2).

The chemosensory system family study classified their members by functionality; therefore I hypothesized a similar evolutionary pattern depending on functionality of the SFTPs. All SFTP share the same organization at the protein level and are specifically expressed in the stratified cornified epithelia and/or the hair follicles (Henry et al. 2012, Kypriotou et al. 2012, Zhihong et at. 2009). Additionally, the SFTP cross-linked by transglutaminases to maintain an intact physical barrier which provides the characteristic resistance and insolubility of the cornified epithelium. Within the SFTP they interact with each other during cornification of skin. Cornification is the elimination of all organelles and of the nucleus by the aggregation of intermediate filaments to form an intracellular fibrous matrix and by the assembly of a resistant protein-shell at the keratinocyte periphery, the cornified cell envelope (Henry et al. 2012). CE composition and shape vary according to the tissue and its state of maturation (Kypriotou et al. 2012). The SFTP family is part of the EDC together with the cornified envelope precursor family and the calcium-binding proteins (S100) family (Kypriotou et al. 2012). Involucrin, a member of the cornified envelope precursor family, also has an intraexonic repeated region containing 39 repeats of 10 aa. Involucrin comparison between human and gorilla demonstrated allele variation across and within species and the addition of repeats was suggested to likely conferred distinctive properties on the primate epidermis (Teumer and Green 1989).

On a previous study, Mlitz et al. cathegorized the SFTP into three groups filaggrin-type (*FLG, FLG-2* and *HRNR*), trichohyalin-type (*TCHH* and *RPTN*) and

cornulin (2014). My findings support the SFTP division from a new evolutionary point of view.

Filaggrin-type genes are subjected to a complex processing and have common and individual biological functions. The sequence and function similarities of the filaggrin-type genes, hornerin, filaggrin-2 and filaggrin, are the following: first, the repeated region of filaggrin shares a high homology with hornerin (41.5%) and filaggrin-2 (42%). Second, the amino acid composition is similar with high levels of serine, glycine, arginine, histidine and glutamine, accounting for more than 70% of the overall constitution (Henry et at. 2012). Third, *FLG* is colocalized with *HRNR* and *FLG-2* in the cytoplasm of the upper granular keratinocytes. But only *FLG* is expressed on lower granular keratinocytes (Molin et al. 2014, Pellerin et al. 2013). Fourth, while *FLG* and *FLG-2* are distributed uniformly within the granules, *HRNR* is detected on the periphery. Fifth, in the lower corneocytes layer, proFLG is cleaved into *FLG* monomers that aggregate to the keratin cytoskeleton to form the corneocytes fibrous matrix. *FLG-2* is also found in the matrix of lower corneocytes and has similar histidine composition, necessary for aggregation, suggesting a similar role (Hsu et al. 2011). Sixth, *HRNR, FLG* and *FLG-2* are synthesized as high-molecular-mass precursors that are proteolytically processed to fragments during cornification. In the upper stratum corneum, filaggrin and filaggrin-2-filaggrin-like repeats are deiminated by peptidyl deiminase (PAD) enzymes which reduces the affinity of the filaggrin/keratin complex. Next, *FLG* units are degraded to free amino acids by calpain 1, caspase 14 and bleomycin hydrolase (Hsu et al. 2011). FLG-2-FLG-like repeats are degraded also degraded by calpain1. In the case of *HRNR*, it undergoes spontaneous degradation at neural pH to release free amino acids. The most important released amino acids are histidine and glutamine functioning as the source of the natural moisturizing factor. Histidine is deiminated to trans-urocanic acid (UCA) and glutamine is metabolized to pyrrolidone-5-carboxylic acid (PCA); together they maintain the pH gradient of the epidermis. The acidity of the stratum corneum has a well-known antimicrobial effect and demonstrates an inhibitory effect on the growth of *S. aureus*. This is also necessary for ceramide metabolism and to control the activity of the serine protease cascade required for EDC coordination. Additionally, UCA decreases the UV-induced apoptosis providing skin photoprotection (Henry et al. 2011, Kezic et al. 2011, Levin et al. 2013, Molin et al.

2014, Zhihong et al. 2009). In conclusion, the first group interacts with keratin and their amino acid composition having the highest influence in the NMF.

In my study, the first group or filaggrin-type has high nucleotide variation between repeats within individuals. In the case of *FLG*, the nucleotide variation is the highest and the FLG-2-FLG-like repeats have the highest pairwise values of Ka/Ks (Table 6, Table 14 and Table 19). Likewise, all positively selected codons detected between species, except for one in *RPTN*, were found in this group. I suggest that the nucleotide variation found in these genes is due to the fact that a global rather than a precise amino acid sequences is necessary for its function, except for the domain that probably interacts with keratin which remains unknown. Additionally, I suggest that the initial repeats are more likely to be conserved, as I observed this in FLG-2-FLG-like repeat 1, 2, 3 and 4, and FLG-2-HRNR-like repeat 1 and the latter repeats are allowed to duplicate which was observed in the phylogenetic analyses (Figure 7, Figure 8, Figure 18 and Figure 28 A-B).

The second group or trichohyalin-type and the third group, which is only cornulin, have a more active function in the inner sheath of hair and a more compensatory one in the CE (Contzler et at 2005, Henry et at 2011, Kypriotou et al. 2012, Lieden et al. 2009). *TCHH* and *RPTN* are detected at very low levels in the granular keratinocytes, in contrast to *CRNN* which is expressed at higher levels and prevents apoptosis against deoxycholate acid implicated in esophageal reflux. *TCHH*, *CRNN* and *RPTN* undergo post-translational modifications but no proteolytic processing during epidermal differentiation. The expression of *CRNN, RPTN* and *TCHH* increases during stress which suggests that their function in the CE is more a compensatory mechanism allowing the organism to repair an altered epidermal barrier (Contzler et at 2005, Henry et at 2011, Kypriotou et al. 2012).

RPTN and *CRNN* repeats have shorter length (36 bp and 180 bp per repeat unit respectively) and have no variation in length within or across species (Figure 29 and Figure 31). *RPTN* and *CRNN* have low nucleotide variation and most of them being synonymous within species and only one positively selected codon in repetin-repeat 2 was found across species (Table 26). Additionally, neither duplication nor species-specific cluster was observed on the phylogenetic analyses (Figure 30 C and Figure 32 C). I suggest that in comparison with the first group, the amino acid compositions of *RPTN* and *CRNN* are necessary to be conserved for function and additional duplications are not permitted nor needed.

In conclusion, the ancestor of the SFTP appeared in amniotes and was diverged into three groups, the filaggrin-type, trichohyalin-type and cornulin. The filaggrin-type has high nucleotide variation within the repeats within and across species, especially filaggrin, suggesting that a global rather than a precise amino acid sequences is necessary for function, and allowing variability in the CE. The trichohyalin-type in contrast has low nucleotide changes as its repetitive domains are conserved and single nucleotide variations have a phenotypic effect. Cornulin is the most conserved with the lowest nucleotide variability most likely due to its role in cell cycle, contrary to the adaptive evolution described previously.

As mentioned before, the number of *FLG* complete repeats varies from 10 to 12 within human due to duplicates of repeat 8 and 10 as previously determined by Sandilands et al. (2007). The variation in the number of repeats affects filaggrin degradation products. Fewer repeats are associated with dry skin and increased number of repeats is associated with a decreased risk for AD (Kezic et al. 2011; Brown et al. 2012). AD is the most common form of eczema in childhood and globally, Ecuador has the second highest prevalence of AD in children between 6 to 7 years old and the eight highest one for children between 13 to 14 years (16.5%) (Odhiambo et al. 2009). Previous studies in Korean and Irish populations found significant difference between CNV in repeat-allele and the risk of AD; in contrast a study performed in Russian population did not replicate this association (Brown et al. 2012, Gimalova et al. 2016, Li et al. 2016,). I obtained samples from Ecuadorian pediatric population, analyzed the number of filaggrin repeats but could not associate the number of filaggrin repeats to the risk of developing AD. In previous studies, population-specific LOF mutations have been strongly associated with AD (Akiyama 2010, Thyssen et al. 2014). Therefore, I searched for the most common variations associated with AD in European populations and identified low frequency of the most prevalent variations (1 heterozygote case for R501X). LOF mutations follow a geographical gradient distribution with the prevalence of filaggrin variations increasing as latitude increases. The latitude-dependent distribution of *FLG* LOF mutations is suggested to insure an adequate vitamin D3 status in high latitudes locations. As Ecuador is located on latitude 0, LOF mutations are less frequent; instead I identified new non-synonymous variations significant in cases. The 2 new non-synonymous variations are predicted as possibly damaging by polyphen. I

conclude that the new non-synonymous damaging variations are associated to AD in Ecuadorian pediatric population.

Final conclusions

The SFTP are part of the Epidermal Differentiation Complex (EDC), a dense cluster of genes important for skin structure. I concluded that SFTPs evolved under the birth-and-death model by showing high nucleotide diversity within species and duplication and loss events within and across species. These evolutionary characteristics further categorized the SFTPs similarly to a previously described division based on amino acids composition, expression pattern and function. This division grouped the SFTP into three groups filaggrin-type (*FLG, FLG-2* and *HRNR*), trichohyalin-type (*TCHH* and *RPTN*) and cornulin. My results concluded that filaggrin-type has high nucleotide variation within the repeats within and across species and species-specific duplications, trychohyalin-type has lower nucleotide changes as single nucleotide variations have a phenotypic effect, and cornulin has the lowest nucleotide variability most likely due to its role in cell cycle, contrary to an adaptive hypothesis previously described.

Additionally, I found variation in the length of the *FLG* repeated region within species in chimpanzee and crab-eating macaque.

I also conclude that the 39 bp unit of *HRNR* is conserved and detectable in primates. I clarify the formation of hornerin repeats as follow: the primary unit (39 bp) duplicated nine times to form the secondary unit. The secondary unit duplicated again to form the tertiary unit and then further duplicated to made the quartic unit or longest-unit. The formation of hornerin repeats was conserved in all primates except macaque whose primary units did not duplicate in ninths.

I expected that FLG-2-FLG-like and FLG-2-HRNR-like regions would evolve similarity to *FLG* and *HRNR*, respectively. However my findings conclude that FLG-2 evolved separately from *FLG* and *HRNR*.

Moreover, repetin is the only SFTPs in which some repeats are conserved and gathered from human to mouse.

Furthermore, my analyses demonstrated that cornulin, which previously was suggested to evolver under positive selection due to a misalignment in the analysis, actually evolved by purifying selection.

Finally, I conclude filaggrin repeat variation was not associated with the risk of atopic dermatitis in Ecuadorian children and I detected 2 new non-synonymous damaging variants associated to cases.

References

- Ahmadloo S, Nakaoka H, Hayano T, Hosomichi K, You H, Utsuno E, Sangai T, Nishimura M, Matsushita K, Hata A, Nomura F, Inoue I. Rapid and cost-effective high-throughput sequencing for identification of germline mutations of BRCA1 and BRCA2. *J Human Genet.* 2017; 62 (5): 561-567.
- Akiyama M. FLG mutations in icthyosis vulgaris and atopic eczema: spectrum of mutations and population genetics. *Br J Dermatol.* 2010; 162:472-477
- Aldred P, Hollox E, Armour J. Copy number polymorphism and expression level variation of the human α-defensin genes DEFA1 and DEFA3. *Hum Mol Genet*. 2005;14:2045–2052.
- Anisimova M, Bielawski J, Yang Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol*. 2001;18:1585–1592.
- Austen R, Kobayashi T. Highly efficient concerte evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole genome shotgun sequence data. Genome Res. 2007;17:184–191.
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–249.
- Bolger, AM, Lohse, M and Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30: 2114–2120
- Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu C, Xie D, Suchard MA, Rambaut A, Drummond AJ. BEAST2: A software platform for Bayesian evolutionary analysis. *PLOS Comput Biol*. 2014;10.
- Brown N, Lery C, Sander C. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics.* 1998; 14: 380-381
- Brown S, Irwin W. One remarkable molecule: Filaggrin. *J Invest Dermatol*. 2012;132:751–762.
- Brown S, Kroboth K, Sandilands A, Campbell LE, Pohler E, Kezic S, Cordell H, McLean I, Irvine A. Intragenic Copy Number Variation within Filaggrin contributes to the risk of Atopic Dermatitis with a Dose-Dependent Effect. *J Invest Dermatol*. 2012;133:98–104.
- Cabanillas B and Novak N. Atopic dermatitis and filaggrin. *Curr Opin Immunol.* 2016; 17:1-8

- Caburet S, Cocquet J, Vaiman D, Veitia RA. Coding repeats and evolutionary "agility". BioEssays 2005;581–587.
- Cascella R, Foti V, Lepre T, Galli E, Moschese V, Chini L, Mazzanti C, Fortugno P, Novelli G, Giardina E. Full Sequencing of the FLG Gene in Italian Patients with Atopic Eczema: Evidence of New Mutations, but Lack of an Association. *J. Invest Dermatol.* 2011; 131: 982-984
- Chen JM, Cooper D, Chuzhanova N, Férec C, Patrinos G. Gene conversion: mechanisms, evolution and human disease. *Nature* 2007;8:762–775
- Contzler R, Favre B, Huber M, Hohl D. Cornulin, a New Member of the "Fused Gene" Family, Is Expressed During Epidermal Differentiation. *J Invest Dermatol. 2005 May; 124:990-997*
- DePristo M A, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 2011; 43: 491–498
- Dipankar D. Handa S. Filaggrin mutations and the skin. *Indian J Dermatol Venereol Leprol*. 2012;78:545–551.
- Drummond A, Suchard M, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012;29:1969–1973.
- Durand D, Halldórsson BV, Vernot B. A hybrid micro-macroevolutionary approach to gene tree reconstruction. *J Comput Biol*. 2006;13:320–335.
- Eirín-López J, Rebordinos L, Rooney A, Rozas J. The birth-and-death evolution of multigene families revisited. *Genome Dyn*. 2012;7:170–196.
- Fallon P, Sasaki T, Sandilands A, Campbell L, Saunders S, Mangan N, Callanan J, Kawasaki H, Shiohama A, Kubo A, et al. A homozygous frameshift mutation in the murine filaggrin gene facilitates enhanced percutaneous allergen priming. *Nat Genet*. 2009;41:602–608.
- Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006;7:85–97.
- Fondon JW, Garner HR. Molecular origins of rapid and continuous morphological evolution. *PNAS*. 2004;101:18058–18063.
- Gan SQ, McBride W, Idler W, Markov N, Steinert P. Organization, structure and polymorphisms of the human profilaggrin gene. *Biochemistry* 1990;29:9432–9440.

- Gazave E, Darre F, Morcillo-Suarez C, Petit-Marty N, Carreno A, Marigorta U, Ryder O, Blancher A, Rocchi M, Bosch E. Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. *Genome Res*. 2011;21:1626–1639.
- Gemayel R, Vinces M, Legendre M, Verstrepen K. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu. Rev. Genet.* 2010; 44:445-77
- Gibbs M, Armstrong J, Gibbs A. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics.* 2000; 16 (7): 573-82
- Gimalova G, Karunas A, Fedorova Y, Khusnutdinova E. The study of filaggrin gene mutatins and copy number variation in atopic dermatitis patients from Volga-Ural region of Russia. *Gene.* 2016; 591: 85-89
- Hahn M, De Bie T, Stajich J, Nguyen C, Cristianini N. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res*. 2005;15:1153–1160.
- Hansmann B, Ahrens K, Wu Z, Proksch E, Meyer-Hoffert U, Schroder JM. Murine filaggrin-2 is involved in epithelial barrier function and down-regulated in metabolically induced skin barrier dysfunction. *Experimental Dermatology.* 2012; 21:271-276
- Henry J, Tsu CY, Haftek M, Nachat R, de Koning HD, Gardinal-Galera I, Hitomi K, Balica S, Jean-Decoster C, Schmitt AM et al. Hornerin is a component of the epidermal cornified cell envelopes. *FASEB J.* 2011 May; 25: 1567-1576.
- Henry J, Toulza E, Hsu C, Pellerin L, Balica S, Mazereeuw-Hautier J, Paul C, Serre G, Jonca N, Simon M. Update on the epidermal differentiation complex . *Frontiers in Bioscience.* 2012; 17:1517-1532.
- Hosak L, Silhan P, Hosakova J. Genomic copy number variations: A breakthrough in our knowledge on schizophrenia etiology? *Neuro Endocrinol Lett*. 2012;33:183–190.
- Hoste E, Kemperman P, Devos M, Denecker G, Kezic S, Yau N, Gilbert B, Lippens S, De Groote P, Roelandt R, et al. Caspase-14 is required for filaggrin degradation to natural moisturizing factors in the skin. *J Invest Dermatol.* 2011 November; 121: 2233-2241

- Hsu CY, Henry J, Raymond AA, Mechin MC, Pendaries V, Nassar D, Hansmann B, Balica S, Burlet-Schiltz O, Schmitt AM et al. Deimination of human filaggrin-2 promotes its proteolysis by calpain 1. *J Biol Chem.* 2011 July;286:23222-33
- Hsu PK, Kao HL, Chen Hy, Yen CC, Wu YC, Hsu, WH, Chou TY. Loss of CRNN expression is associated with advanced tumor stage and poor survival in patients with esophageal squamous cell carcinoma. *J Thorac Cardiovasc Surg.* 2014 May; 147: 1612-1618
- Hu T, Banzhaf W. Nonsynonymous to Synonymous Substitution Ratio ka/ks: Measurement for Rate of Evolution in Evolutionary Computation In R. *Parallel Problem Solving from Nature - PPSN X*. Dortmund: Springer. 2008.
- Huber M, Siegenthaler G, Mirancea N, Marenholz I, Nizetic D, Breitkreutz D, Mischke D, Hohl D. Isolation and Characterization of Human Repetin, a Member of the Fused Gene Family of the Epidermal Differentiation Complex. *J Invest Dermatol.* 2005 May; 124: 998-1007.
- Kanda S, Sasaki T, Shiohama A, Nishifuji K, Amagai M, Iwasaki T, Kudoh J. Characterization of canine filaggrin: gene structure and protein expression in dog skin. *Vet Dermatol*. 2013;24:25–32.
- Kezic S, O'Regan GM, Yau N, Sandilands A, Chen H, Campbell LE, Kroboth K, Watson R, Rowland M, McLean WH, et al. Levels of filaggrin degradation products influenced by both filaggrin genotype and atopic dermatitis severity. *Allergy.* 2011 July;66:934-940.
- Krieg P, Schuppler M, Koesters R, Mincheva A, Litcher P, Marks F. Repetin (Rptn), a New Member of the ''Fused Gene'' Subgroup within the S100 Gene Family Encoding a Murine Epidermal Differentiation Protein. *Genomics.* 1997; 43: 339-348
- Koren S, Schatz M, Walenz B, Martin J, Howard J, Ganapathy G, Wang Z, Rasko D, McCombie R, Jarvis E, et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol*. 2012;30:693–700
- Kypriotou M, Huber M, Hohl D. The human epidermal differentiation complex: cornified envelope precursors, S100 proteins and the 'fused genes' family. *Experimental dermatology.* 2012; 21:643-649.
- Levin J, Friedlander SF, Del Rosso JQ. Atopic Dermatitis and the Stratum Corneum Part 1: The Role of Filaggrin in the Stratum Corneum Barrier and Atopic Skin. *J Clin Aesthet Dermatol.* 2013 October; 6: 16-22.

- Li H, and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. July 2009; 25 (14): 1754-60
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25: 2078–2079.
- Li K, Seok J, Park KY, Yoon Y, Kim KH, Seo SJ. Copy-number variation of the filaggrin gene in Korean patients with atopic dermatitis: what really matters, 'number' or 'variation'?. *Br J Dermatol. 2*016; 174: 1098-1100.
- Li M, Liu Q, Liu J, Cheng R, Zhang H, Xue H, Bao Y, Yao Z. Mutations analysis in filaggrin gene in northern China patients with atopic dermatitis. *J. Eur. Acad. Dermatol. Venereol.* 2013; 27: 169-174
- Lieden A, Ekelund E, Kyo IC, Kockum I, Huang CH, Mallbris L, Lee SP, Seng LK, Chin GY, Wahlgren CF, et al. Cornulin, a marker of late epidermal differentiation, is down-regulated in eczema. *Allergy* 2009; 64: 304-311.
- Librado P, Rozas J. DnaSP v5:A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009;25:1451–1452.
- Little T, Nelson L, Hupp T. Adaptive evolution of a stress response protein. *Plos One.* 2007; 10: e1003.
- Makino T, Takaishi M, Morohashi M, Huh N. Hornerin, a novel profilaggrin-like protein and differentiation-specific marker isolated from mouse skin. *The journal of biological chemistry.* 2001; 276: 7445–47452
- Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 2000;16:562–563.
- Martin D, Posada D, Crandall K, Williamson C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses*. 2005;21:98–102.
- Martin D, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 2010;26:2462–2463.
- Martin D, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 2015; 1: vev003
- Martin J, Wang Z. Next-generation transcriptome assembly. *Nat Genet*. 2011;12:671–682.

- Maynard-Smith J. Analyzing the mosaic structure of genes. *J Mol Evol*. 1992; 34:126–129.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A *et al*. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010; 20: 1297–1303
- Mlitz V, Strasser B, Jeager K, Hermann M, Ghannadan M, Buchberger M, Alibardi L, Tschachler E, Eckhart L. Trichohyalin-like proteins have evolutionarily conserved roles in the morphogenesis of skin appendages. *J Invest Dermatol.* 2014 November; 134: 2685-2692.
- Molin S, Merl J, Kietrich KA, Regauer M, Flaig M, Letule V, Sauckle T, Herzinger T, Ruzicka T, Hauck SM. The hand eczema proteome: imbalance of the epidermal barrier proteins. *British Journal of Dermatology.* 2014; 994-1001.
- Nei M. The new mutation theory of phenotypic evolution. *PNAS* 2007;104:12235–12242.
- Nei M, Rooney A. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet*. 2005;39:121–152.
- Nei M, Rogozin I, Piontkivska H. Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci*. 2000;97:10866–10871.
- Odhiambo J, Williams H, Clayton T, Robertson C, Asher I, ISAAC Phase Three Study Group. Global variations in prevalence of eczema symptoms in children from ISAAC Phase Three. *J Allergy Clin Immunol.* 2009; 124: 1251-1258
- Padidam M, Sawyer S, Fauquet C. Possible emergence of new geminiviruses by frequent recombination. *Virology*. 1999;265:218–225.
- Paudel Y, Madsen O, Megens HJ, Frantz L, Bosse M, Bastiaansen J, Crooijmans R, Groenen M. Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC Genomics* 2013;14:449.
- Pellerin L, Henry J, Hsu CY, Balica S, Jean-Decoster C, Cechin MC, Hansmann B, Rodriguez E, Weindinger S, Schmitt AM, Serre G, Paul C, Simon M. *J Allergy Clin Immunol* 2013 April; 131:1094-1102
- Posada D, Crandall K. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc Natl Acad Sci*. 2001;98:13757–13762.

- Pospiech E, Karlowska-Pik J, Marcinska M, Abidi S, Andersen JD, van den Berge M, Carracedo A, Eduardoff M, Feire-Aradas A, Morling N et al. Evaluation of the predictive capacity of DNA variants associated with straight hair in Europeans. *Forensic Sci Int Genet.* 2015; 19: 280-288.
- R Development Core Team. R: A Language and Enviroment for Statistical Computing. *R Foundation for Statistical Computing*. 2008. http://www.R-project.org.
- Rajka G. On definition and framework of atopic dermatitis. *Acta Derm Venereol Suppl (Stockh)*. 1989; 44:10-2
- Redon R, Ishikawa S, Fitch K, Feuk L, Perry G, Andrews D, Fiegler H, Shapero M, Carson A, Chen W, et al. Global variation in copy number in the human genome. *Nature* 2006;444:444–454.
- Sabbagh A, Marin J, Vyssiére C, Lecompte E, Boukouvala S, Poloni ES, Darlu P, Crouau-Roy B. Rapid birth-and-death evolution of the xenobiotic metabolizing NAT gene family in vertebrates with evidence of adaptive selection. *BMC Evol Biol* 2013;13:62.
- Salzano FM, Sans M. Interethnic admixture and evolution of Latin American populations. *Genet Mol Biol.*2014 Mar; 37: 151-170.
- Sandilands A, Terron-Kwiatkowski A, Hull P, O'Regan G, Clayton T, Watson R, Carrick T, Evans A, Liao H, Zhao Y, et al. Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema. *Nature Genetics* 2007;39:650-654
- Sandilands A, Sutherland C, Irvine A, McLean H. Filaggrin in the frontline: role in skin barrier function and disease. J Cell Sci. 2009;122:1285–1294.
- Sawyer SA. GENECONV: a computer package for the statistical detectin of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis. 1999. http://www.math.wustl.edu/~sawyer.
- Smith FJ, Irvine AD, Terron-Kwiatkowski A, Sandilands A, Campbell LE, Zhao Y, Liao H, Evans AT, Goudie Dr, Lewis-Jone S, et al. Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat Genet*. 2006;38:337–342.
- Smith M. Analyzing the mosaic structure of genes. *J Mol Evol.* 1992; 34(2):126-9
- Sole D, Mallol J, Wandalsen GF, Aguirre V, Latin American ISAAC Phase 3 Study Group. Prevalence of symptoms of eczema in Latin America: results of the

International Study of Asthma and Allergies in Childhood (ISAAC) Phase 3. *J Investig Allergol Clin Immunol.* 2010; 20: 311-323.

- Steiper M, Young N. Primate molecular divergence dates. *ScienceDirect*. 2006;41:384–394.
- Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. Inferring duplications, losses, transfers, and incomplete lineage sorting with non-binary species trees. *Bioinformatics*. 2012;28:409–415.
- Sudmant P, Huddleston J, Catacchio C, Malig M, Hillier L, Baker C, Mohajeri K, Kondova I, Bontrop R, Persengiev S, et al. Evolution and diversity of copy number variation in the great ape lineage. *Genome Res*. 2013;23:1373–1382.
- Tachikui H, Naruya S, Nakajima T, Hayasaka I, Ishida T, Inoue I. Lineage-specific homogenization of the polyubiquitin gene among human and great apes. *Journal of Molecular Evolution.* 2003 57;737-744
- Takaishi M, Makino T, Morohashi M, Huh NH. Identification of Human Hornerin and Its Expression in Regenerating and Psoriatic Skin. *J Biol Chem.* 2005 February; 280: 4696-703
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–2739.
- Thyssen JP, Bikle DD, Elias PM. Evidence That Loss-of-Function Filaggrin Gene Mutations Evolved in Northern Europeans to Favor Intracutaneous Vitamin D3 Production. *Evol Biol.* 2014; 41: 388-396
- Thyssen Jp, Zirwas MJ, Elias PM. Potential role of reduced environmental UV exposure as a driver of the current epidemic of atopic dermatitis. *J Allergy Clin Immunol.* 2015; 136: 1163-1169.
- Trzeciak M, Wesserling M, Bandurski T, Glen J, Nowicki R, Pawelczyk T. Association of a single nucleotide polymorphism in a late cornified envelope-like proline-rich 1 gene (LELP1) with atopic dermatitis. *Acta Derm Venereol.* 2016; 96: 459-463.
- Vernot B, Stolzer M, Goldman A, Durand D. Reconciliation with non-binary species trees. *J Comput Biol*. 2008;15:981–1006.
- Viera F, Sánchez-Gracia A and Rozas J. Comparative genomic analysis of the odorant-binding protein family in 12 Drosophila genomes: purifying selection and birth-and-death evolution. *Genome Biol.* 2007;8(11):R235

- Wu Z, Meyer-Hoffert U, Reithmayer K, Paus R, Hansmann B, He Y, Bartels J, Glaser R, Harder J, Schroder JM. Highly Complex Peptide Aggregates of the S100 Fused-Type Protein Hornerin Are Present in Human Skin. *J Invest Dermatol*. 2009 June; 129: 1446-1458
- Xu Z, Wang M, Xu X, Cai Y, Han Y, Wu K, Wang J, Chen B, Wang X, Wu M. Novel Human Esophagus-specific gene C1orf10: cDNA cloning gene structure and frequent loss of expression in esophageal cancer. *Genomics*. 2000; 69: 322-330
- Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24:1586–1591.
- Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 1997;13:555–556.
- Zhang D, Karunaratne S, Kessler M, Mahony D, Rothnagel JA. Characterization of mouse profilaggrin: evidence for nuclear engulfment and translocation of the profilaggrin B-domain during epidermal differentiation. *J Invest Dermatol.* 2002; 119 (4): 905-912
- Zheng Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. *J Comput Biol.* 2000; 7:203-214
- Zhihong W, Hansmann B, Meyer-Hoffert U, Glaser R, Schroder J. Molecular identification and expression analysis of filaggrin-2, a member of the S100 fused-type protein family. *PLoS ONE.* 2009; e5227.
- Zhu L, Zhang Y, Hu Y, Wen T, Wang Q. Dynamic actin gene family evolution in primates. *Biomed Res Int*. 2013;11.

Tables

Table 1. Primer sets for the filaggrin gene repeat region and non-repeated region in human, chimpanzee, gorilla, orangutan, crab-eating macaque, and human variation in repeated region.

REPEATED REGION	
Human	
Forward	5′-CTTGTCATATGGCTAACTGGCTTTCAGAGA-3′
Reverse	5′-ATTGTGGGACAGTGATTATGTTGGAGAAAA-3′
Human variation in repeated region	
Forward	5′ -GTGCAAGCAGAAAAACATATGACA -3′
Reverse	5′ -CCTGTTTCGTGATCTGCCTTTGACATGG -3′
Chimpanzee	
Forward	5′-TGTTCACATATAACACCTGAATGAGGATGAGAC-3′
Reverse	5′-GGATAATAGAGAAAGATGTGCTAGCCCTG-3′
Gorilla	
Forward	5′-GTAACACCTGAATGAGGATGAGACGAAA-3′
Reverse	5′-TAGCCCTGATGTTGATATAGCCACTTTG-3′
Orangutan	
Forward	5′-AAGAAAAAGCGTCACTTACCCCATCAAA-3′
Reverse	5′-TTGACATGGCTTAATCACCACCTAAGTT-3′
Crab-eating macaque	
Forward	5′-ACACCTGAGCGAGGATGAGATGAAAAAG-3′
Reverse	5′-TCTGGCCATGGGGAAGTATGTAATTTGG-3′
NON-REPEATED REGION	
Human	
Forward	5′ -TGAGAAAGGAGAGTAGAAGCTTATTTCA-3′
Reverse	5′ -TTTGGCAATAAATGTGAACCTAGAAAGA-3
Macaque	
Forward	5′ -TAGAATGCCTCTGGCCATCTCTGTA -3′
Reverse	5′ -GATGACTGTGCTTTCTGTGCTTGTG -3′

Table 2. Best-fit model comparison for filaggrin gene

Filaggrin			
Model	**#Parameters**	**AICc**	**lnL**
TN93+G+I	106	15792.76	-7790.15
TN93+G	105	15797	-7793.28
GTR+G+I	109	15797.15	-7789.33
GTR+G	108	15801.25	-7792.39
HKY+G+I	105	15803.62	-7796.58
HKY+G	104	15807.12	-7799.34
T92+G+I	103	15854.64	-7824.1
T92+G	102	15858.75	-7827.16
K2+G+I	102	15872.82	-7834.2
K2+G	101	15875.5	-7836.54
TN93+I	105	16001.73	-7895.64
GTR+I	108	16004.65	-7894.09
HKY+I	104	16011.18	-7901.37
T92+I	102	16062.56	-7929.07
K2+I	101	16073.61	-7935.6
JC+G+I	101	16259.3	-8028.44
JC+G	100	16261.94	-8030.77
JC+I	100	16455.1	-8127.35
GTR	107	16547.98	-8166.76
TN93	104	16561.66	-8176.61
HKY	103	16574.9	-8184.23
K2	100	16609.66	-8204.63
T92	101	16645.8	-8221.69
JC	99	16984.97	-8393.28

INCLUDING DOG AND MOUSE REPEATS

Filaggrin			
Model	**#Parameters**	**AICc**	**lnL**
GTR+G	148	15848.28	-7775.71
TN93+G	145	15848.61	-7778.89
GTR+G+I	149	15850.29	-7775.71
TN93+G+I	146	15850.62	-7778.89
HKY+G	144	15878.45	-7794.82
HKY+G+I	145	15880.46	-7794.82
T92+G	142	15928.78	-7821.99
T92+G+I	143	15930.79	-7821.99
K2+G	141	15987.01	-7852.11
K2+G+I	142	15989.02	-7852.11
GTR+I	148	16165.21	-7934.17

TN93+I	145	16171.63	-7940.40
HKY+I	144	16200.70	-7955.94
GTR	147	16231.13	-7968.14
TN93	144	16235.74	-7973.46
T92+I	142	16247.84	-7981.52
K2+I	141	16263.26	-7990.24
HKY	143	16274.55	-7993.87
T92	141	16332.26	-8024.74
K2	140	16342.98	-8031.10
JC+G	140	16347.52	-8033.38
JC+G+I	141	16349.54	-8033.38
JC+I	140	16608.63	-8163.93
JC	139	16680.85	-8201.04

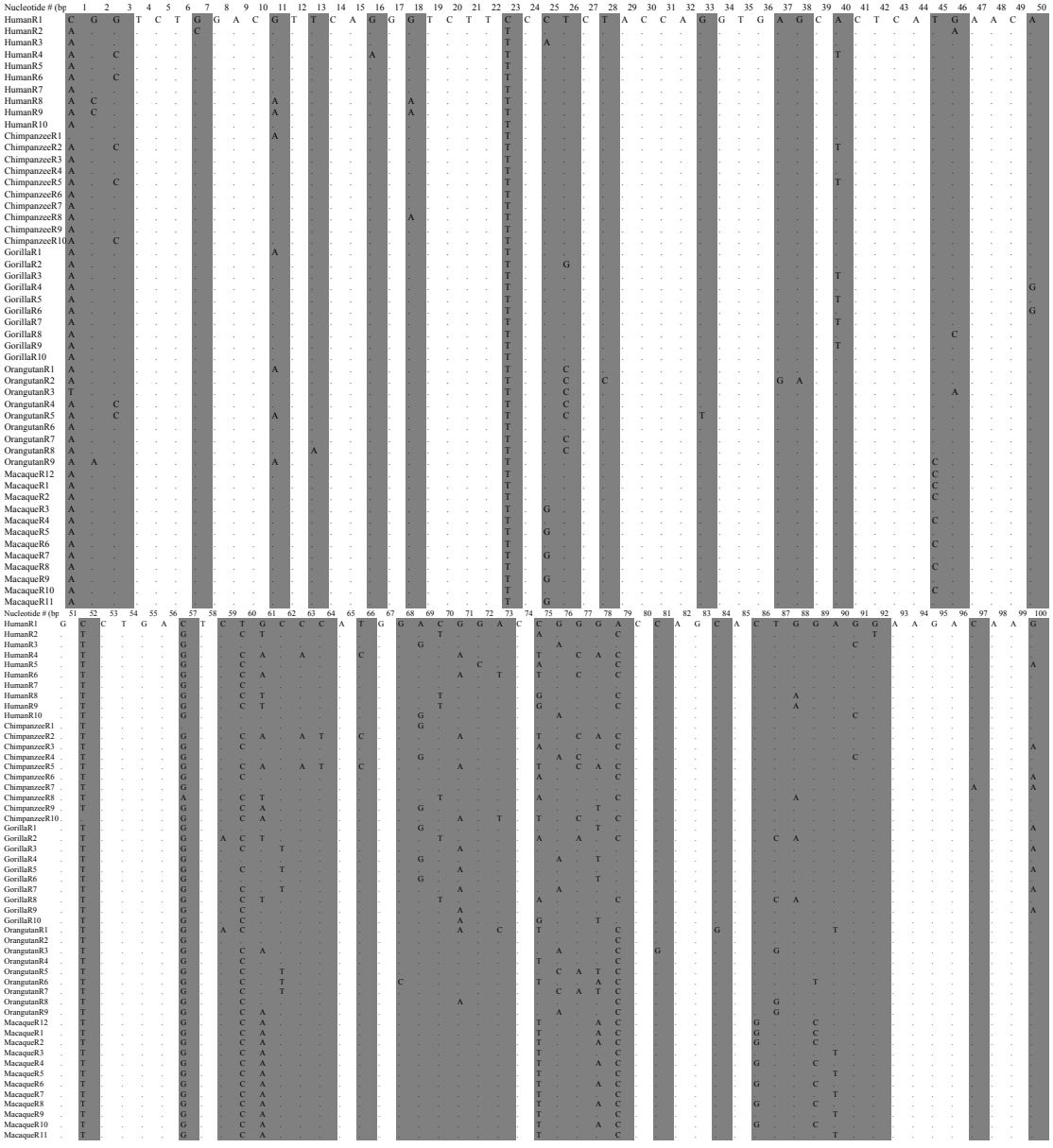
INCLUDING BABOON AND MARMOSET REPEATS

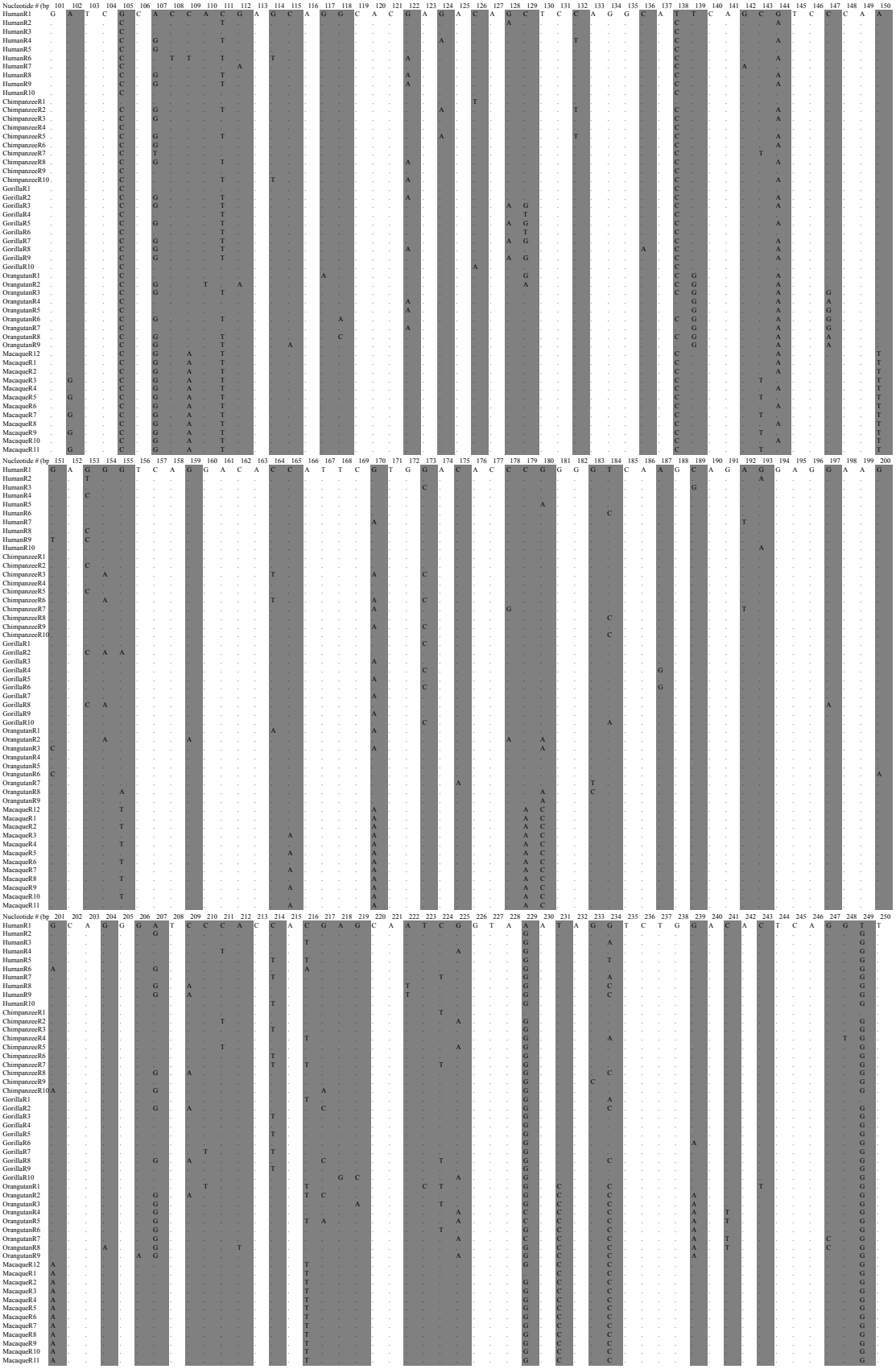
Filaggrin

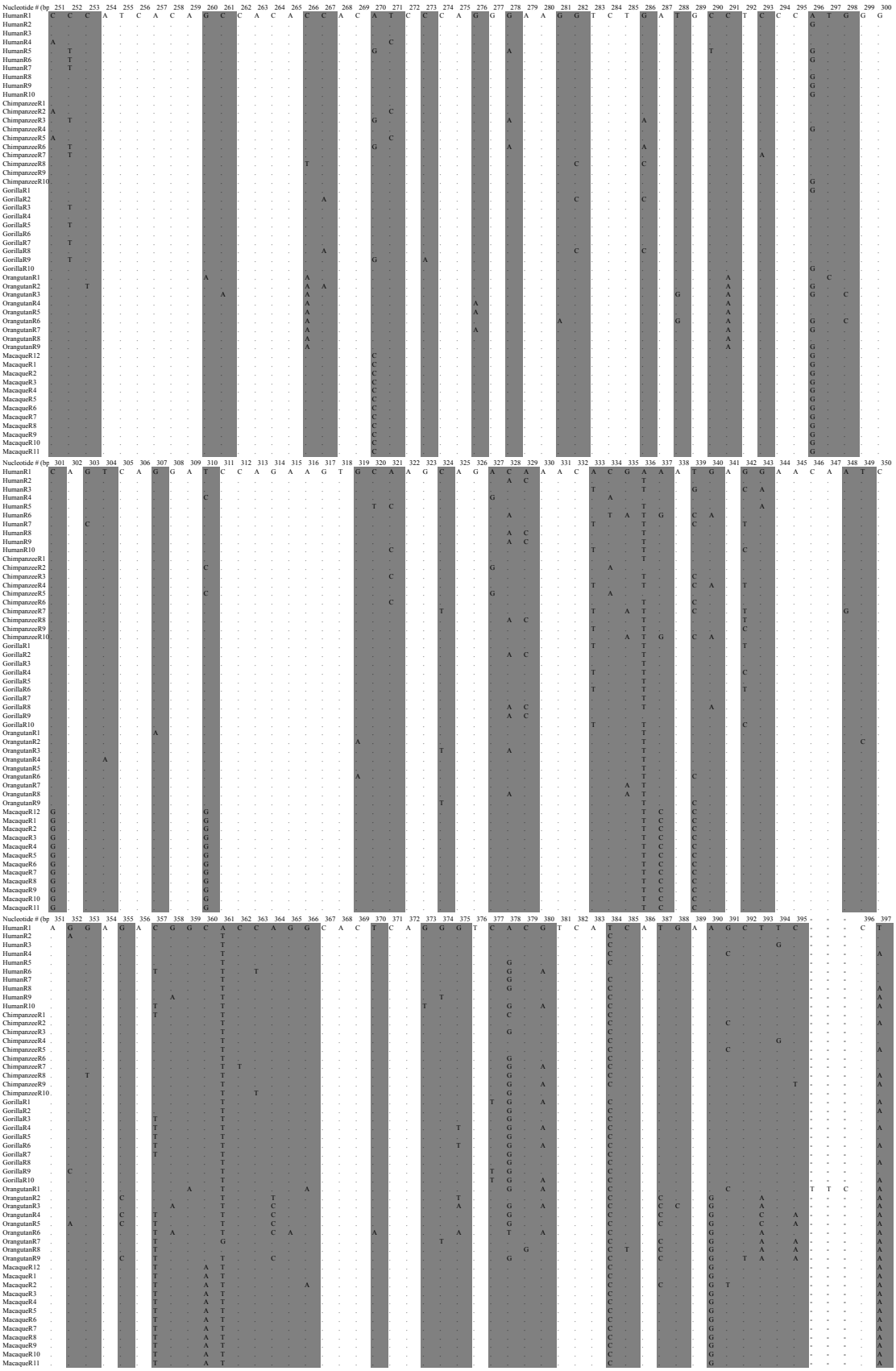
Model	#Param	AICc	lnL
TN93+G	121	17830.43	-8793.96
TN93+G+I	122	17832.44	-8793.96
GTR+G	124	17834.64	-8793.05
GTR+G+I	125	17836.65	-8793.05
HKY+G	120	17850.60	-8805.05
HKY+G+I	121	17852.61	-8805.05
T92+G	118	17896.39	-8829.95
T92+G+I	119	17898.39	-8829.95
K2+G	117	17922.94	-8844.23
K2+G+I	118	17924.94	-8844.23
TN93+I	121	18103.20	-8930.34
GTR+I	124	18103.20	-8927.33
HKY+I	120	18119.57	-8939.53
T92+I	118	18168.31	-8965.91
K2+I	117	18177.35	-8971.43
JC+G	116	18340.00	-9053.76
JC+G+I	117	18342.01	-9053.76
GTR	123	18485.53	-9119.50
TN93	120	18499.01	-9129.25
HKY	119	18519.82	-9140.66
K2	116	18551.55	-9159.54
T92	117	18580.09	-9172.80
JC+I	116	18589.11	-9178.32
JC	115	18957.72	-9363.63

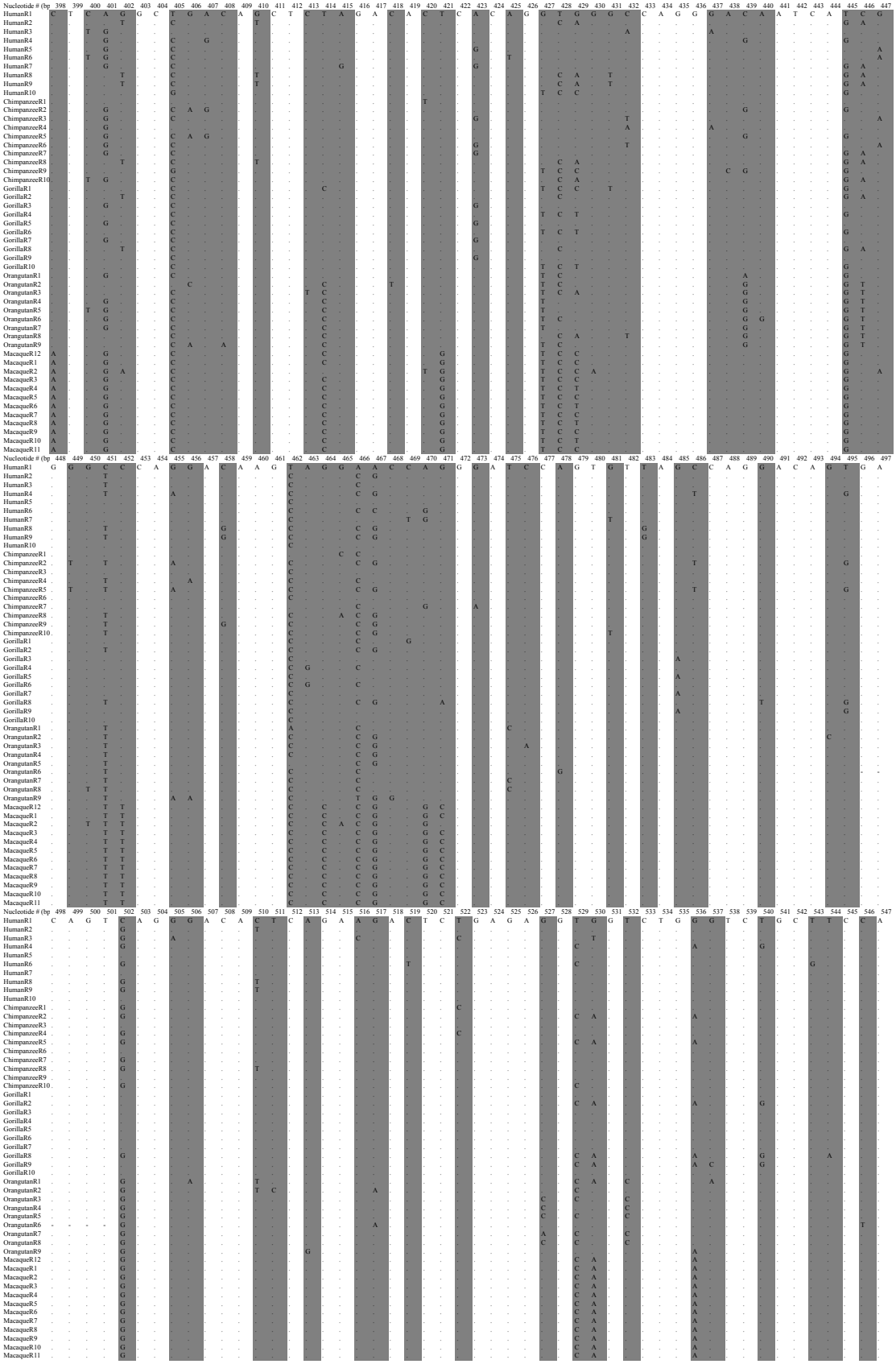
Table 3.

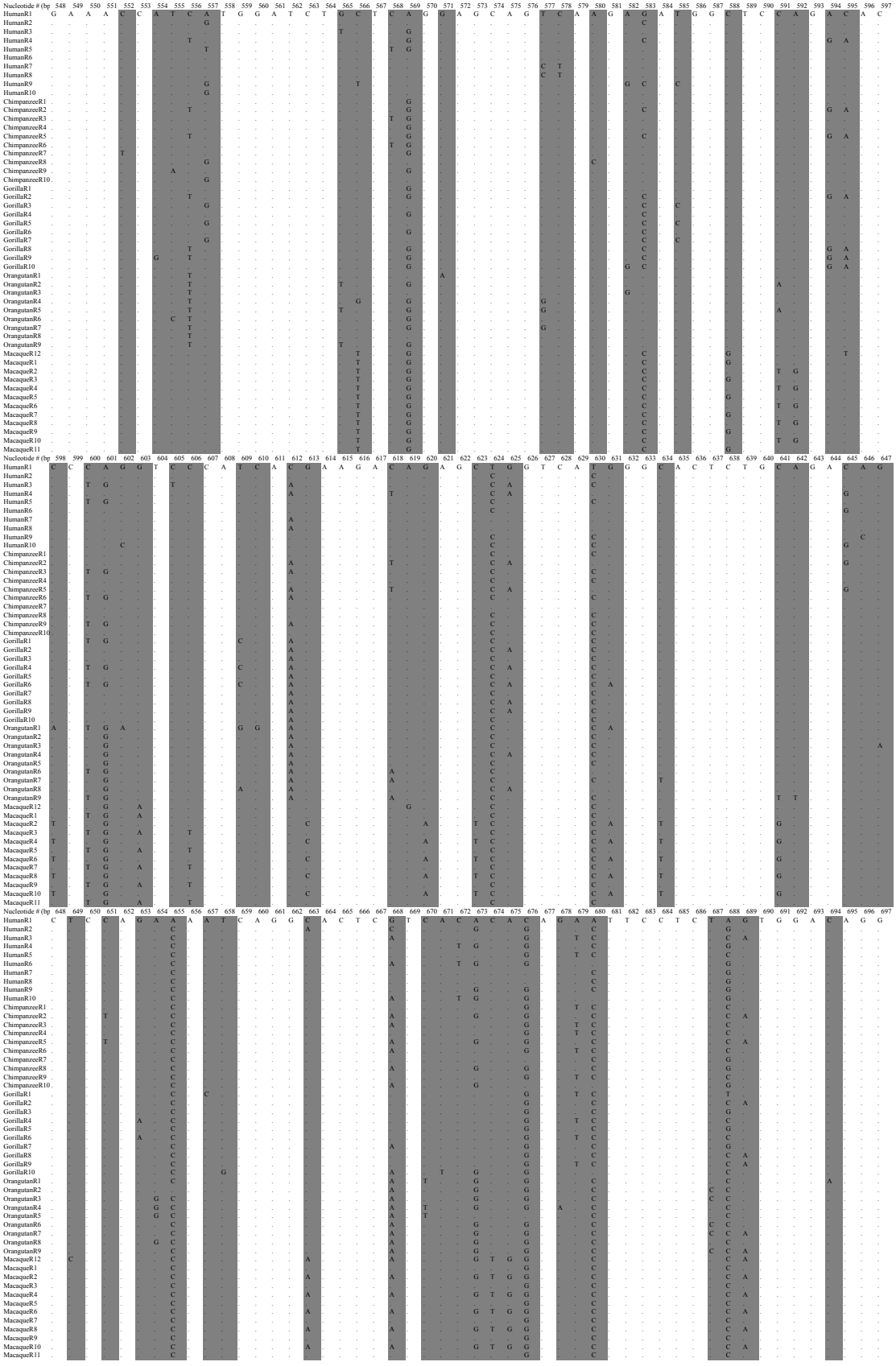
A. Nucleotide alignment of all complete repeat sequences to Human R1 performed by ClustalW.

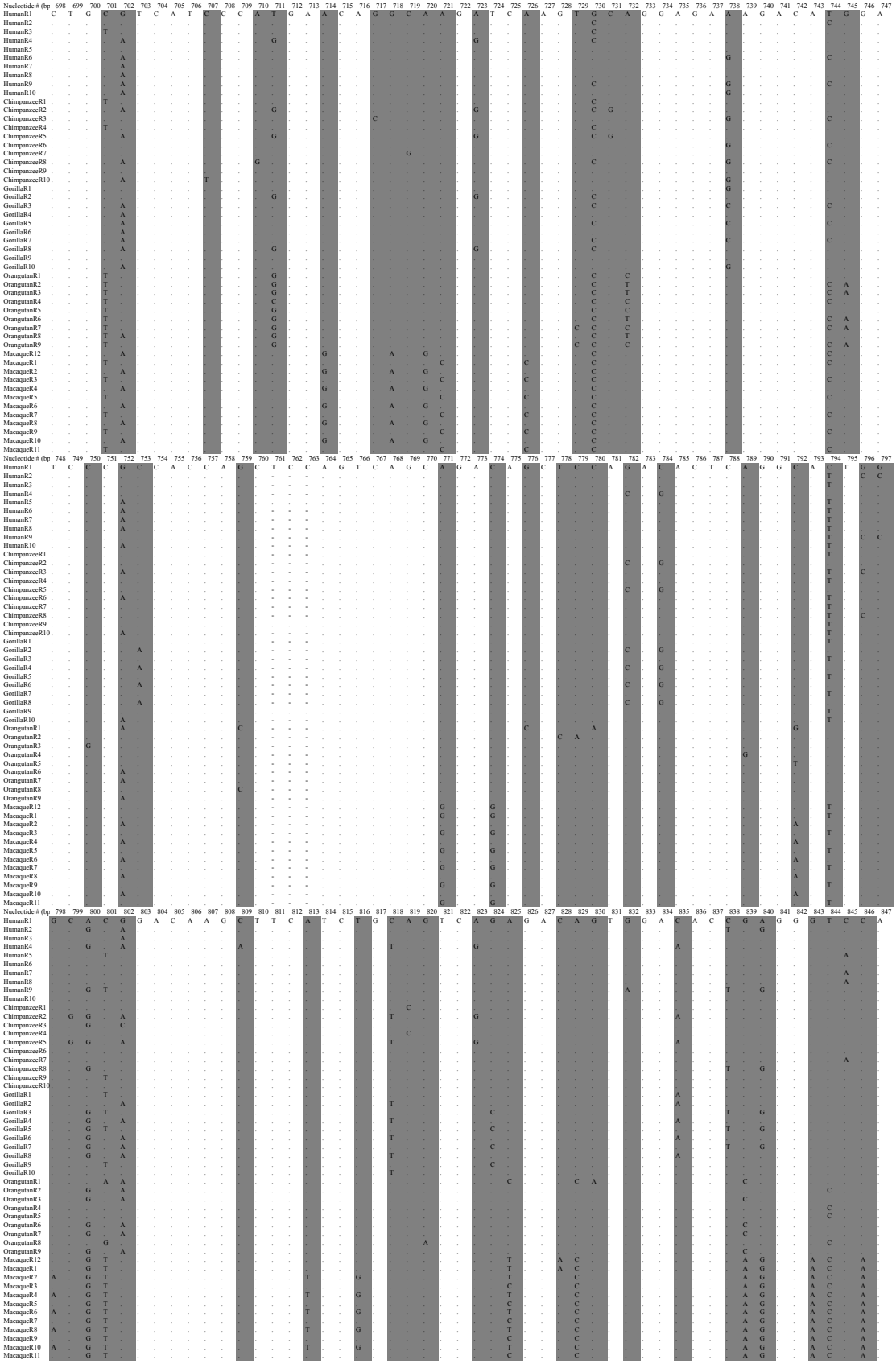


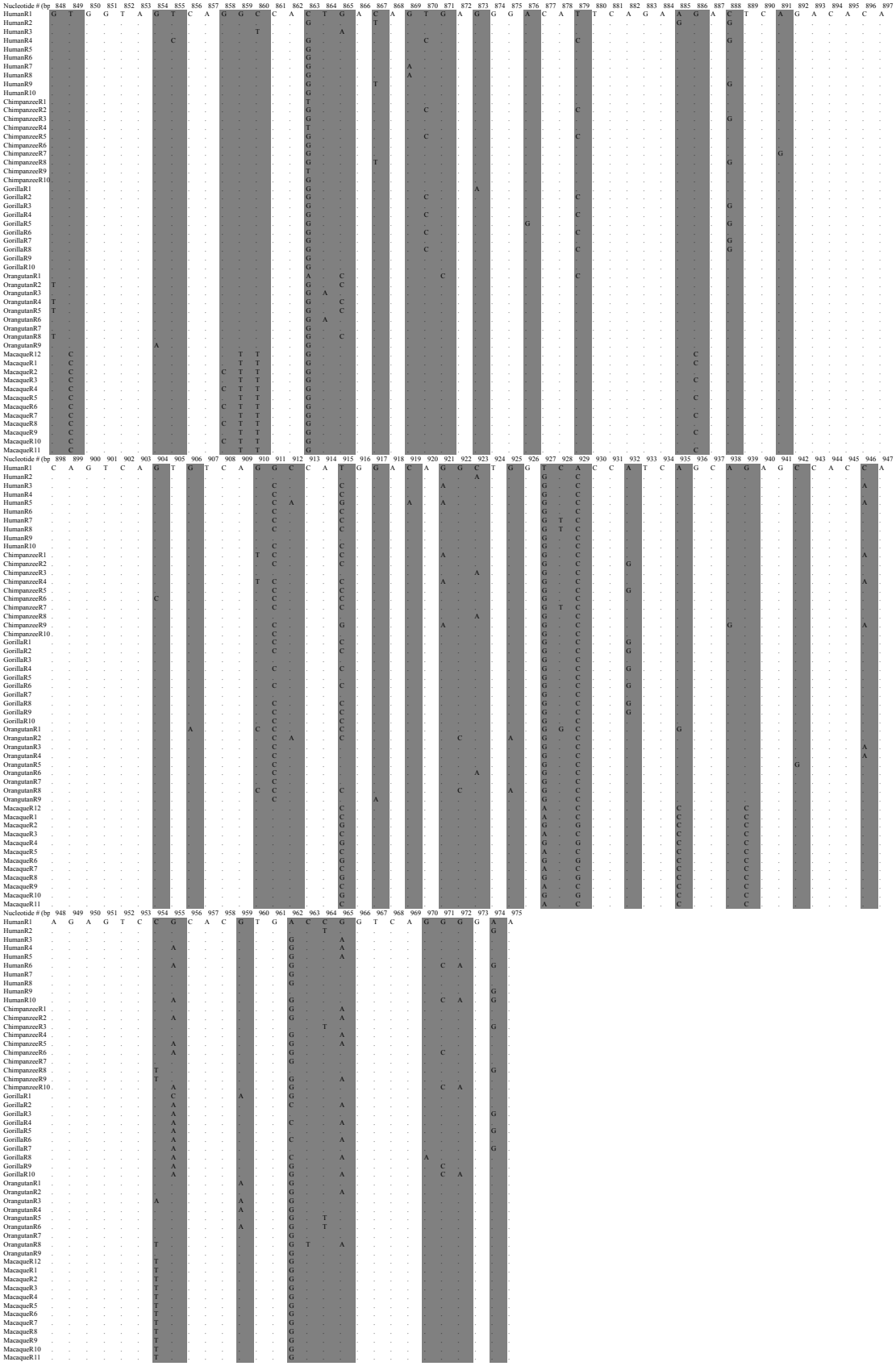




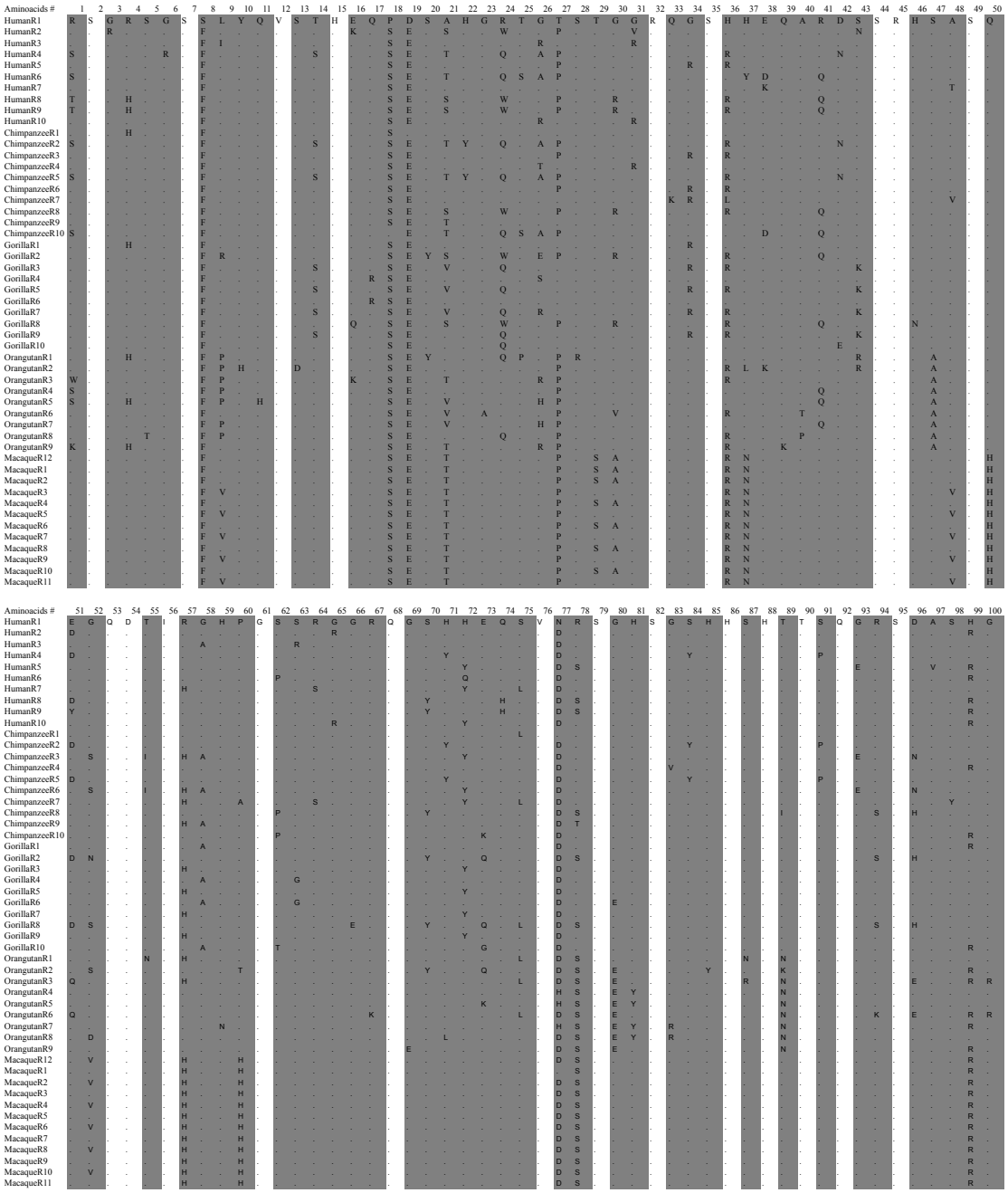


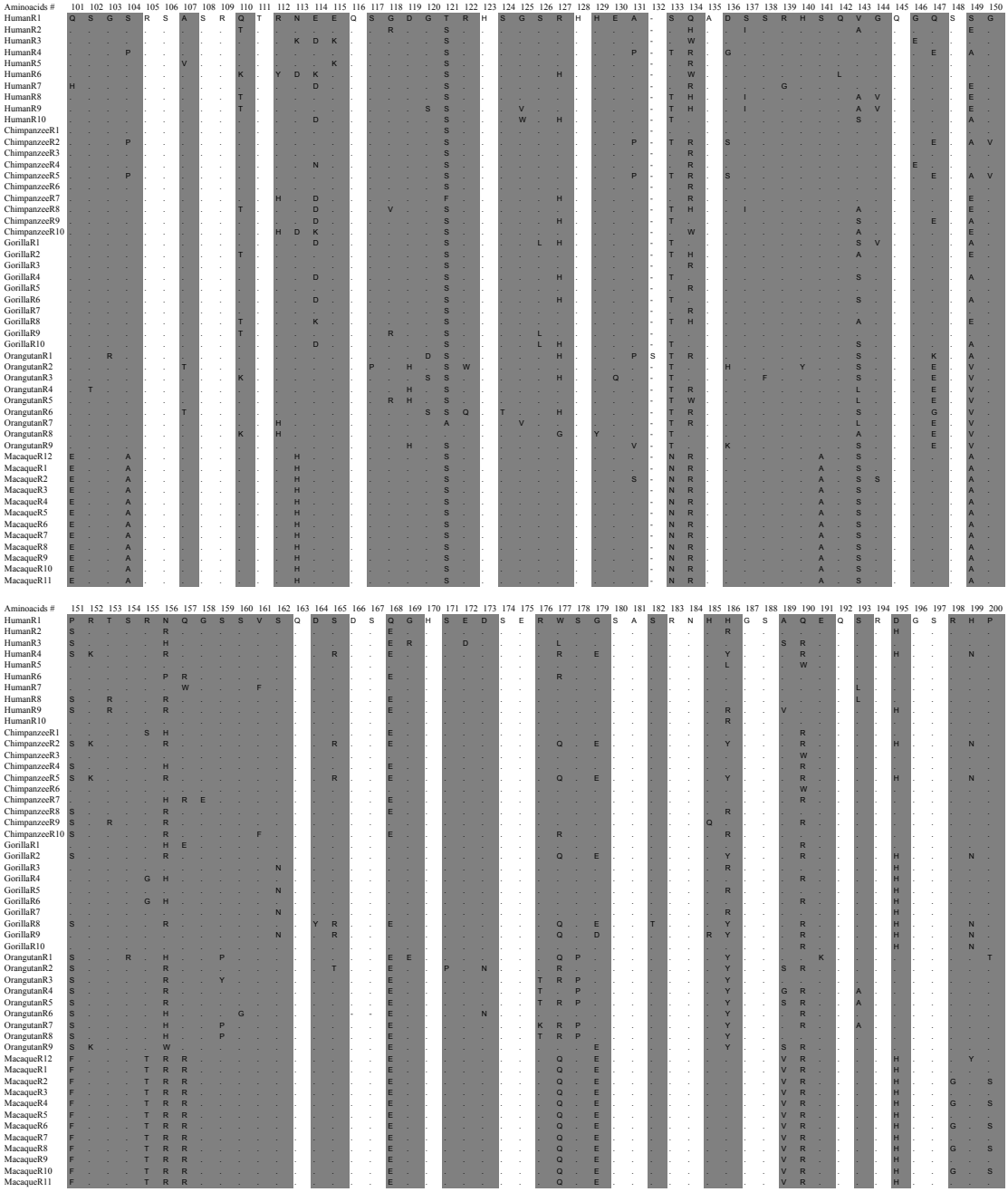


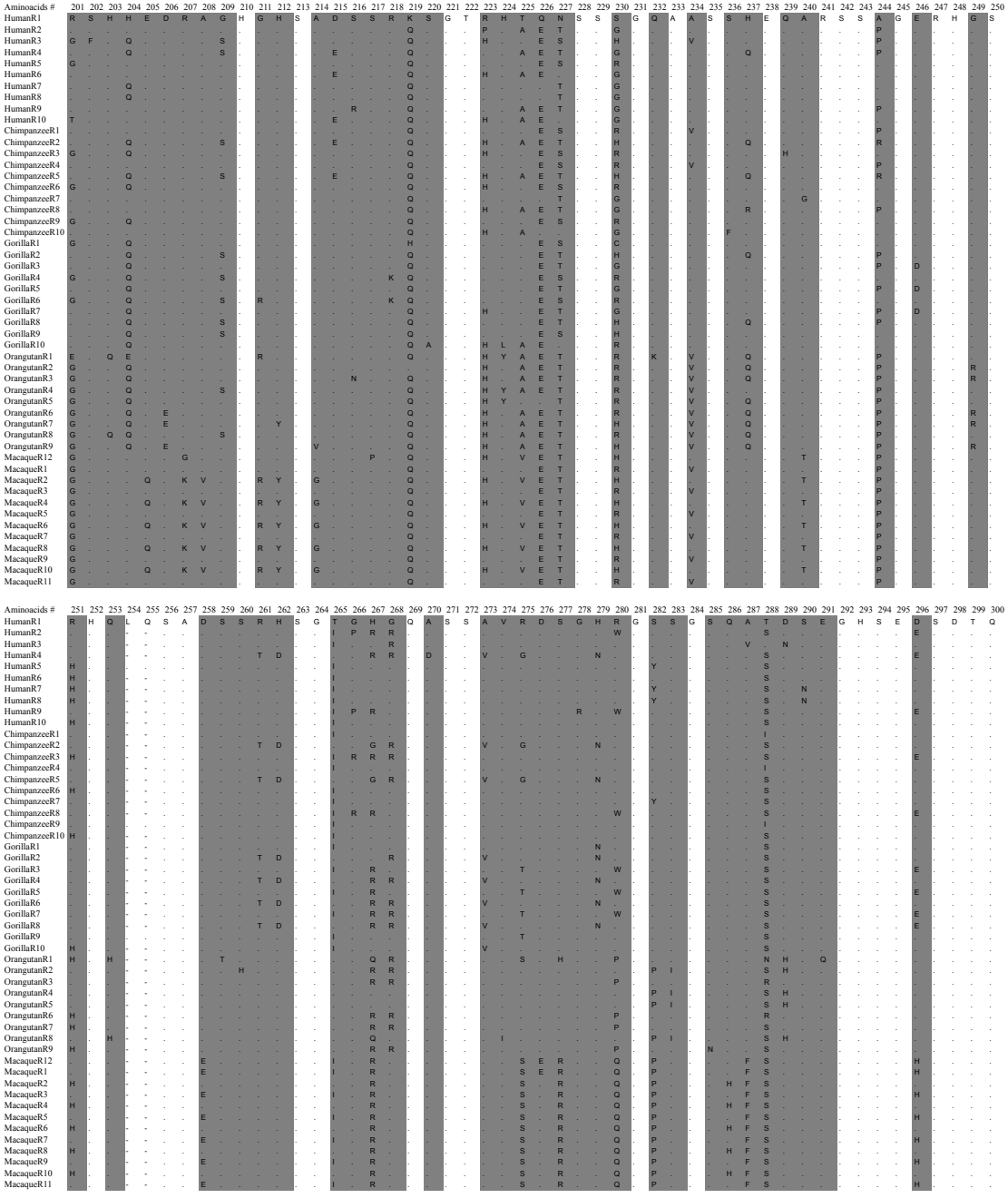




B. Aminoacid alignment of all complete repeat sequences to Human R1 performed by ClustalW.




Aminoacids # 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
HumanR1 Q S G S R S A S R Q T R N E E Q S G D G T R H S G S R H E A - S Q A D S S R H S Q V G Q G Q S S G
HumanR2 R S T K D K S A I A E E
HumanR3 K D K S W E
HumanR4 P S R P T R G E P A T
HumanR5 V S K R
HumanR6 K Y D K S H W L
HumanR7 H D S R E
HumanR8 T S H I A V E
HumanR9 T S V S T H I A V E
HumanR10 D S W H S A
ChimpanzeeR1 S
ChimpanzeeR2 P S P T R S P E A V
ChimpanzeeR3 S R
ChimpanzeeR4 N S R E
ChimpanzeeR5 P S P T R S P E A V
ChimpanzeeR6 S R
ChimpanzeeR7 H D F H R E
ChimpanzeeR8 T D V S H I A E
ChimpanzeeR9 D S H S E A
ChimpanzeeR10 H D K S W A E
GorillaR1 D S L H T R S V A
GorillaR2 T S T H A E
GorillaR3 S R
GorillaR4 D S H T R S A
GorillaR5 S R T R S A
GorillaR6 D S H T R S A
GorillaR7 S R T R L A
GorillaR8 T K S H T H A E
GorillaR9 T S L R
GorillaR10 D S L H T R S A
OrangutanR1 R D S P S T R S K A A
OrangutanR2 T P H S W H T R S E V A
OrangutanR3 K S H Q T F S E V A
OrangutanR4 T H S T R L E V A
OrangutanR5 R H S T W L E V A
OrangutanR6 T S S Q T H T R S G V A
OrangutanR7 H A V L E V A
OrangutanR8 K H G Y T R A E V A
OrangutanR9 H S V T K S E V A
MacaqueR12 E A H S N R A S A
MacaqueR1 E A H S N R A S A A
MacaqueR2 E A H S P S N R A S S A A
MacaqueR3 E A H S N R A S A A
MacaqueR4 E A H S N R A S A A
MacaqueR5 E A H S N R A S A A
MacaqueR6 E A H S N R A S A A
MacaqueR7 E A H S N R A S A A
MacaqueR8 E A H S N R A S A A
MacaqueR9 E A H S N R A S A A
MacaqueR10 E A H S N R A S A A
MacaqueR11 E A H S N R A S A A
Aminoacids # 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
HumanR1 P R T S R N Q G S S V S Q D S D S Q G H S E D S E R W S G S A S R N H H G S A Q E Q S R D G S R H P
HumanR2 S R E R H
HumanR3 S H E R D L S R
HumanR4 S K R E R E R Y R H
HumanR5 L W
HumanR6 P R E R
HumanR7 W F L
HumanR8 S R E L
HumanR9 S R R E R V H
HumanR10 R
ChimpanzeeR1 S H E R
ChimpanzeeR2 S K R R E Q E Y R H N
ChimpanzeeR3 W
ChimpanzeeR4 S H E R
ChimpanzeeR5 S K R R E Q E Y R H N
ChimpanzeeR6 W
ChimpanzeeR7 H R E E R
ChimpanzeeR8 S R E R
ChimpanzeeR9 S R R E Q R
ChimpanzeeR10 S R F E R R
GorillaR1 H E R
GorillaR2 S R Q E Y R H N
GorillaR3 N R H
GorillaR4 G H R H
GorillaR5 N R H
GorillaR6 G H R H
GorillaR7 N R H
GorillaR8 S R Y R E Q E T Y R H N
GorillaR9 N R Q D R Y R H N
GorillaR10 R H N
OrangutanR1 S R H P E E Q P K H T
OrangutanR2 S R T P N E P R Y S R H
OrangutanR3 S R Y E T R P Y H
OrangutanR4 S R E T P Y G R A H
OrangutanR5 S R E T R P Y S R A H
OrangutanR6 S H G E N E Y R H
OrangutanR7 S H P E K R P Y R A H
OrangutanR8 S H P E T R P Y H
OrangutanR9 S K W E Y S R H
MacaqueR12 F T R R E Q E V R H Y
MacaqueR1 F T R R E Q E V R H
MacaqueR2 F T R R E Q E V R H G S
MacaqueR3 F T R R E Q E V R H
MacaqueR4 F T R R E Q E V R H G S
MacaqueR5 F T R R E Q E V R H
MacaqueR6 F T R R E Q E V R H G S
MacaqueR7 F T R R E Q E V R H
MacaqueR8 F T R R E Q E V R H G S
MacaqueR9 F T R R E Q E V R H
MacaqueR10 F T R R E Q E V R H G S
MacaqueR11 F T R R E Q E V R H


Aminoacids # 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250
HumanR1 R S H H E D R A G H G H S A D S S R K S G T R H T Q N S S S G Q A A S S H E Q A R S S A G E R H G S
HumanR2 P A E T G P
HumanR3 G F Q S H Q E S G H E S G H V P
HumanR4 Q E Q S A E T G Q P
HumanR5 G Q E S R Q P
HumanR6 Q E H A E G Q P
HumanR7 Q T G Q P
HumanR8 Q T G Q P
HumanR9 R Q A E T G Q P
HumanR10 T Q E H A E G Q P
ChimpanzeeR1 Q E S R V Q P
ChimpanzeeR2 Q S E H A E T H Q R
ChimpanzeeR3 G Q Q E S R H Q H
ChimpanzeeR4 Q E S R V Q P
ChimpanzeeR5 Q S E H A E T H Q R
ChimpanzeeR6 G Q Q E S R H Q P
ChimpanzeeR7 Q T G Q G
ChimpanzeeR8 Q H A E T G Q P
ChimpanzeeR9 G Q Q E S R Q P
ChimpanzeeR10 Q H A G F Q P
GorillaR1 G Q E S C H Q P
GorillaR2 Q S E T H Q P
GorillaR3 Q E T G Q P D
GorillaR4 G Q S K Q E S R Q P
GorillaR5 Q E T G Q P D
GorillaR6 G Q S R K Q E S R Q P
GorillaR7 Q H E T G Q P D
GorillaR8 Q S E T H Q P
GorillaR9 Q S E S H Q P
GorillaR10 Q H L A E R Q P
OrangutanR1 E Q E R Q H Y A E T R K V Q P
OrangutanR2 G Q H A E T R V Q P R
OrangutanR3 G Q N H A E T R V Q P R
OrangutanR4 G Q S H Y A E T R V Q P
OrangutanR5 G Q H Y T R V Q P
OrangutanR6 G Q E H A E T R V Q P R
OrangutanR7 G Q Y H A E T H V Q P R
OrangutanR8 G Q Q S H A E T R V Q P
OrangutanR9 G Q E V H A E T H V Q P R
MacaqueR12 G P Q H V E T H P T
MacaqueR1 G Q E T R V Q P
MacaqueR2 G Q R K V R Y G Q H V E T H P T
MacaqueR3 G Q E T R V Q P
MacaqueR4 G Q R K V R Y G Q H V E T H P T
MacaqueR5 G Q E T R V Q P
MacaqueR6 G Q R K V R Y G Q H V E T H P T
MacaqueR7 G Q E T R V Q P
MacaqueR8 G Q R K V R Y G Q H V E T H P T
MacaqueR9 G Q E T R V Q P
MacaqueR10 G Q R K V R Y G Q H V E T H P T
MacaqueR11 G Q E T R V Q P
Aminoacids # 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
HumanR1 R H Q L Q S A D S S R H S G T G H G Q A S S A V R D S G H R G S Q A T D S E G H S E D D S T Q
HumanR2 I P R W S E
HumanR3 I R V N
HumanR4 T D R R D V G N S T E
HumanR5 H I Y S
HumanR6 H I S
HumanR7 H I Y S N
HumanR8 H I Y S N
HumanR9 I P R R W S E
HumanR10 H I S
ChimpanzeeR1 I I S
ChimpanzeeR2 T D G R V G N S
ChimpanzeeR3 H I R R R S E
ChimpanzeeR4 I I S
ChimpanzeeR5 T D G R V G N S
ChimpanzeeR6 H I S
ChimpanzeeR7 I Y S
ChimpanzeeR8 I R R W S E
ChimpanzeeR9 I I S
ChimpanzeeR10 H I S
GorillaR1 I N S
GorillaR2 T D R V N S
GorillaR3 I R T W S E
GorillaR4 T D R R V N S
GorillaR5 I R T W S E
GorillaR6 T D R R V N S
GorillaR7 I R T W S E
GorillaR8 T D R R V N S E
GorillaR9 I T S
GorillaR10 H I V S
OrangutanR1 H H T Q R S H P N H Q
OrangutanR2 H R R P I S H
OrangutanR3 R R P R
OrangutanR4 P I S
OrangutanR5 P I S H
OrangutanR6 H R R P R
OrangutanR7 H R R P S
OrangutanR8 H Q I P I S H
OrangutanR9 H R R P N S
MacaqueR12 E I R Q P F S H
MacaqueR1 E I R S E R Q P F S H
MacaqueR2 H R S R Q P H F S
MacaqueR3 E I R S R Q P F S H
MacaqueR4 H R S R Q P H F S
MacaqueR5 E I R S R Q P F S H
MacaqueR6 H R S R Q P H F S
MacaqueR7 E I R S R Q P F S H
MacaqueR8 H R S R Q P H F S
MacaqueR9 E I R S R Q P F S H
MacaqueR10 H R S R Q P H F S
MacaqueR11 E I R S R Q P F S H

Aminoacids #	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325
HumanR1	S	V	S	G	H	G	Q	A	G	H	H	Q	Q	S	H	Q	E	S	A	R	D	R	S	G	E
HumanR2	.	.	.	.	.	.	.	D	.	P	.	.	.	.	.	.	.	.	.	.	.	W	.	.	G
HumanR3	.	.	.	A	.	.	.	.	.	P	.	.	.	.	.	K	.	.	.	.	G	Q	.	.	.
HumanR4	.	.	.	A	.	.	.	.	.	P	.	.	.	.	.	.	.	.	T	.	G	Q	.	.	.
HumanR5	.	.	.	A	Q	.	K	.	.	P	.	.	.	.	.	K	.	.	.	.	G	Q	.	.	.
HumanR6	.	.	.	A	.	.	.	.	.	P	.	.	.	.	.	.	.	.	T	.	G	.	.	A	G
HumanR7	.	.	.	A	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
HumanR8	.	.	.	A	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
HumanR9	.	.	.	.	.	.	.	.	.	P	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
HumanR10	.	.	.	A	.	.	.	.	.	P	.	.	.	.	.	.	.	.	T	.	G	.	.	A	G
ChimpanzeeR1	.	.	.	S	.	.	.	.	.	P	.	.	.	.	.	K	.	.	.	.	G	Q	.	.	.
ChimpanzeeR2	.	.	.	A	.	.	.	.	.	P	R	.	.	.	.	.	.	.	T	.	G	Q	.	.	.
ChimpanzeeR3	.	.	.	.	.	.	.	D	.	P	.	.	.	.	.	.	.	.	.	.	.	W	.	.	G
ChimpanzeeR4	.	.	.	S	.	.	.	.	.	P	.	.	.	.	.	K	.	.	.	.	G	Q	.	.	.
ChimpanzeeR5	.	.	.	A	.	.	.	.	.	P	R	.	.	.	.	.	.	.	T	.	G	Q	.	.	.
ChimpanzeeR6	.	L	.	A	.	.	.	.	.	P	.	.	.	.	.	.	.	.	T	.	G	.	.	A	.
ChimpanzeeR7	.	.	.	A	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
ChimpanzeeR8	.	.	.	.	.	.	.	D	.	P	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
ChimpanzeeR9	.	.	.	A	Q	.	.	.	.	P	.	.	R	.	.	K	.	.	.	.	G	Q	.	.	.
ChimpanzeeR10	.	.	.	A	.	.	.	.	.	P	.	.	.	.	.	.	.	.	T	.	G	.	.	A	.
GorillaR1	.	.	.	A	.	.	.	.	.	P	R	.	.	.	.	.	.	.	P	H	G	.	.	.	.
GorillaR2	.	.	.	A	.	.	.	.	.	P	R	.	.	.	.	.	.	.	T	.	A	Q	.	.	.
GorillaR3	.	.	.	.	.	.	.	.	.	P	.	.	.	.	.	.	.	.	T	.	.	.	.	.	G
GorillaR4	.	.	.	A	.	.	.	.	.	P	R	.	.	.	.	.	.	.	T	.	A	Q	.	.	.
GorillaR5	.	.	.	.	.	.	.	.	.	P	.	.	.	.	.	.	.	.	T	.	.	.	.	.	G
GorillaR6	.	.	.	A	.	.	.	.	.	P	R	.	.	.	.	.	.	.	T	.	A	Q	.	.	.
GorillaR7	.	.	.	.	.	.	.	.	.	P	.	.	.	.	.	.	.	.	T	.	.	.	.	.	G
GorillaR8	.	.	.	A	.	.	.	.	.	P	R	.	.	.	.	.	.	.	T	.	A	Q	.	R	.
GorillaR9	.	.	.	A	.	.	.	.	.	P	R	.	.	.	.	.	.	.	T	.	G	.	.	A	.
GorillaR10	.	.	.	A	.	.	.	.	.	P	.	.	.	.	.	.	.	.	T	.	G	Q	.	A	.
OrangutanR1	.	.	.	P	.	.	.	.	.	A	.	R	.	.	.	.	.	.	.	H	G	.	.	.	.
OrangutanR2	.	.	.	A	.	.	.	P	R	P	.	.	.	.	.	.	.	.	.	.	G	Q	.	.	.
OrangutanR3	.	.	.	A	.	.	.	.	.	P	.	.	.	.	.	K	.	.	.	H	G	.	.	.	.
OrangutanR4	.	.	.	A	.	.	.	.	.	P	.	.	.	.	.	K	.	.	.	H	G	.	.	.	.
OrangutanR5	.	.	.	A	.	.	.	.	.	P	.	.	.	R	.	.	.	.	.	.	G	W	.	.	.
OrangutanR6	.	.	.	A	.	.	.	D	.	P	.	.	.	.	.	.	.	.	.	H	G	W	.	.	.
OrangutanR7	.	.	.	A	.	.	.	.	.	P	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
OrangutanR8	.	.	.	P	.	.	.	P	R	P	.	.	.	.	.	.	.	.	.	.	G	Q	.	.	.
OrangutanR9	.	.	.	A	.	E	.	.	.	P	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR12	.	.	.	.	.	.	.	.	.	P	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR1	.	.	.	.	.	.	.	.	.	P	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR2	.	.	.	.	Q	.	.	.	.	R	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR3	.	.	.	.	.	.	.	.	.	P	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR4	.	.	.	.	Q	.	.	.	.	R	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR5	.	.	.	.	.	.	.	.	.	P	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR6	.	.	.	.	Q	.	.	.	.	R	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR7	.	.	.	.	.	.	.	.	.	P	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR8	.	.	.	.	Q	.	.	.	.	R	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR9	.	.	.	.	.	.	.	.	.	P	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR10	.	.	.	.	Q	.	.	.	.	R	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.
MacaqueR11	.	.	.	.	.	.	.	.	.	P	.	P	H	.	.	.	.	.	.	.	G	.	.	.	.

Table 4. Percentage of similarity between repeats, total numbers of pairs analyzed and numbers of pairs with higher synonymous and non-synonymous variations for filaggrin gene

Gene	Species	Percentage of similarirty Minimum	Percentage of similarity Maximum	Total number of pairs analyzed	Pairs higher synonymous variations	Pairs higher non-synonymous variations
Filaggrin	Human	89.81	96.19	45	43	2
	Pan troglodytes	89.3	100	45	43	2
	Gorilla gorilla	90.64	99.9	45	37	8
	Pongo abelii	88.68	96.91	36	36	0
	Macaque	93.52	100	66	66	0

Table 5. Nucleotide variation, average synonymous variations, average non-synonymous variations and average Ka/Ks from all repeats within a species for filaggrin gene

	Filaggrin			
	Nucleotide variation	**Ks**	**Ka**	**Average (1-Ka/Ks)*100**
Human	0.77	0.11	0.07	32.23
Pan troglodytes	0.75	0.12	0.07	38.14
Gorilla gorilla	0.61	0.08	0.06	23.47
Pongo abelii	0.77	0.11	0.07	35.49
Macaque	0.31	0.07	0.02	71.26
Papio anubis	-	-	-	-
Callithrix jacchus	-	-	-	-

[a]*K*s refers to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon.

[b]*K*a refers to the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon.

[c]*K*a/*K*s > 1 (bold) suggests positive selection.

[d]Purifying selection is measured by (1 − *K*a/*K*s) × 100.

X= sequences not available in NCBI database
-= sequences not analyzed

Table 6. Ka/Ks ratio >1 for pairs of repeats within a species for filaggrin gene

Filaggrin								
Repeat 1	**Repeat 2**	**SynDif**	**SynPos**	**Ks**	**NSynDif**	**NSynPos**	**Ka**	**Ka/Ks**
Gorilla-repeat 2	Gorilla-repeat 8	2	228.330	0.009	16	743.670	0.022	2.477
Gorilla-repeat 2	Gorilla-repeat 9	14	230.670	0.063	50	741.330	0.071	1.117
Gorilla-repeat 8	Gorilla-repeat 9	16	231.170	0.073	56	740.830	0.080	1.098
Gorilla-repeat 7	Gorilla-repeat 8	19.5	229.580	0.090	67.5	742.420	0.097	1.075
Gorilla-repeat 3	Gorilla-repeat 8	19.5	230.170	0.090	66.5	741.830	0.096	1.062
Gorilla-repeat 1	Gorilla-repeat 8	19	228.750	0.088	65	743.250	0.093	1.057
Gorilla-repeat 5	Gorilla-repeat 8	20.5	230.170	0.095	66.5	741.830	0.096	1.007
Gorilla-repeat 2	Gorilla-repeat 7	19.5	229.080	0.090	63.5	742.920	0.091	1.003
Chimpanzee-repeat 3	Chimpanzee-repeat 6	1	229.420	0.004	13	742.580	0.018	4.023
Chimpanzee-repeat 10	Chimpanzee-repeat 3	17	228.000	0.079	56	744.000	0.079	1.010
Human-repeat 10	Human-repeat 3	14.83	229.670	0.068	50.17	742.330	0.071	1.049
Human-repeat 3	Human-repeat 7	15.5	229.250	0.071	52.5	742.750	0.074	1.047

[a]*K*s refers to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon.

[b]*K*a refers to the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon.

[c]*K*a/*K*s > 1 (bold) suggests positive selection.

[d]Purifying selection is measured by (1 − *K*a/*K*s) × 100.

Table 7. Detected positively selected codons and branches, and p-values for filaggrin gene.

SITE-BASED			
Gene	**Models compared**	**p-value**	**Positive selected codons**
FLG	M1a vs M2a	1.40 E-42	21,24,26,75,99,110,135,144,150,157,178,187,190,191,224,226,228,231,252,268,269,305,323
	M7 vs M8	1.05 E-50	4, 21, 24, 26, 75, 99, 110, 127, 135, 144, 150, 152, 157, 178, 187, 190, 191, 205, 224, 226, 228, 231, 245, 252, 268, 269, 305, 309, 320, 323

BRANCH-BASED TESTS			
Gene	**Models compared**	**p-value**	**Branches under independent or unique ω/various branches under same ω**
FLG	M0 vs free ratio	0.13	Unique ω/various branches under same ω

BRANCH-SITE MODEL			
Gene	**M0N0 vs M2N2**	**p-value**	**Positive selected codons**
FLG	Macaque-cluster	1.27 E-60	NO SITES
	Orangutan-cluster	2.54 E-70	26, 119, 160, 177, 179, 207, 250, 284
	Gorilla/Chimpanzee/Human cluster	1.80 E-93	24, 26, 78, 99, 110, 114, 134, 135, 144, 150, 152, 157, 169, 178, 187, 191, 224, 228, 231, 245, 268, 320, 323

Table 8. Comparison between filaggrin repeat units and filaggrin-like repeat units from macaque.

Species	No. of repeats	Total no. variation sites	Average (1 − Ka/Ks) x 100[a]	Nucleotide variation
CRAB-EATING MACAQUE				
Crab-eating macaque FLG	12	69	70.96	0.032
Crab-eating macaque FLG-like	5	73	71.33	0.039

[a] (1-Ka/Ks) x 100 measures the percentage of purifying selection.

Table 9. The *K*a/*K*s ratio and (1 − *K*a/*K*s) × 100 for pairs of repeats of the filaggrin-like gene within macaque.

Crab-eating macaque FLG-like

Repeat 1	Repeat 2	Ks	Ka	Ka/Ks	(1 − *K*a/*K*s) x 100[d]
Macaque-like1	Macaque-like2	0.031	0.016	0.53	47
Macaque-like1	Macaque-like3	0.144	0.033	0.23	77
Macaque-like1	Macaque-like4	0.108	0.024	0.22	78
Macaque-like1	Macaque-like5	0.085	0.016	0.19	81
Macaque-like2	Macaque-like3	0.17	0.044	0.26	74
Macaque-like2	Macaque-like4	0.133	0.038	0.29	71
Macaque-like2	Macaque-like5	0.099	0.030	0.30	70
Macaque-like3	Macaque-like4	0.035	0.011	0.31	69
Macaque-like3	Macaque-like5	0.076	0.019	0.25	75
Macaque-like4	Macaque-like5	0.048	0.014	0.29	71

[a]*K*s refers to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon.

[b]*K*a refers to the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon.

[c]*K*a/*K*s > 1 (bold) suggests positive selection.

[d]Purifying selection is measured by (1 − *K*a/*K*s) × 100.

Table 10. Positively selected codons of the filaggrin-like in macaque.

Crab-eating macaque FLG-like		
Sites-based test		
Models compared	***P* value**	**Positively selected codons**
M1a vs. M2a[a]	0.99	No sites
M7 vs M8 [a]	0.99	No sites

[a]The site-based test compared the M1a(nearly neutral) and M2a (positive selection) models and the M7 (Beta) and M8 (Beta and ω)models.

Table 11. Recombination and gene conversion events in the filaggrin gene in primates, as detected by GeneConv.

	Breakpoint positions				Detection methods					
Recombination event number	**Begin**	**End**	**Recombinant sequence**	**Major parental sequence**	**RDP**	**GENECONV**	**Boots can**	**Max chi**	**Chim aera**	**SiSs can**
Crab-eating macaque										
1	656	735	Crab-eating macaque-R1	Crab-eating macaque-R12	NS	7.75E-03	NS	2.63 E-04	7.46E-04	NS
2	1	266	Crab-eating macaque-R2	Crab-eating macaque-R11	NS	6.05E-03	4.08E-03	NS	NS	NS
Orangutan										
1	651	972	Orangutan-R7	Orangutan-R4	8.95 E-05	2.36E-04	1.27E-04	1.32 E-04	3.66E-04	1.04 E-04
Gorilla										
1	575	978	Gorilla-R8	Gorilla-R10	NS	NS	NS	1.69 E-04	1.25E-03	1.72 E-05
2	245	594	Gorilla-R9	Gorilla-R4	NS	NS	6.75E-08	3.85 E-04	NS	4.92 E-08
Chimpanzee										
1	172	975	Chimpanzee-R6	Chimpanze-R3	NS	NS	4.16E-03	1.70 E-03	2.93E-03	NS
Human										
1	534	975	Human-R8	Human-R9	NS	5.85E-09	5.25E-11	3.59 E-09	4.89E-10	8.55 E-12
2	504	975	Human-R10	Human-R3	NS	2.66E-03	5.50E-08	3.39 E-06	2.62E-07	9.72 E-06

Hornerin

Table 12. Percentage of similarity between repeats, total numbers of pairs analyzed and numbers of pairs with higher synonymous and non-synonymous variations for hornerin gene

Species	Percentage of similarity Minimum	Percentage of similarity Maximum	Total number of pairs analyzed	Pairs higher synonymous variations	Pairs higher non-synonymous variations
Human	79.79	96.81	15	15	0
Pan troglodytes	78.95	94.8	15	15	0
Gorilla gorilla	78.96	86.71	6	6	0
Pongo abelii	84.94	98.3	15	12	3
Macaque	75.76	90.99	15	15	0

Table 13. Nucleotide variation, average synonymous variations, average non-synonymous variations and average Ka/Ks from all repeats within a species

	Nucleotide variation	Average Ks	Average Ka	Average Ka/Ks	Average (1-Ka/Ks)*100
Human	0.13	0.22	0.12	0.54	46.3
Pan troglodytes	0.14	0.23	0.14	0.63	36.84
Gorilla gorilla	0.16	0.28	0.16	0.56	44.16
Pongo abelii	0.14	0.2	0.11	0.71	29.05
Macaque	0.17	0.32	0.26	0.51	49.47

[a]*K*s refers to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon.

[b]*K*a refers to the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon.

[c]*K*a/*K*s > 1 (bold) suggests positive selection.

[d]Purifying selection is measured by (1 − *K*a/*K*s) × 100.

Table 14. Ka/Ks ratio >1 for pairs of repeats within a species

Orangutan								
Longest-1	**Longest- 2**	**SynDif**	**SynPos**	**Ks**	**NSynDif**	**NSynPos**	**Ka**	**Ka/Ks**
Orangutan-3	Orangutan-4	11.00	248.50	0.05	44.00	678.50	0.07	1.49
Orangutan-3	Orangutan-5	12.00	250.17	0.05	39.00	676.83	0.06	1.21
Orangutan-2	Orangutan-3	16.33	249.00	0.07	45.67	678.00	0.07	1.03

[a]*K*s refers to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon.

[b]*K*a refers to the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon.

[c]*K*a/*K*s > 1 (bold) suggests positive selection.

[d]Purifying selection is measured by (1 − *K*a/*K*s) × 100.

Table 15. Best-fit model comparison for hornerin

Hornerin			
Model	**#Param**	**AICc**	**lnL**
K2+G	155	20390.10	-10039.36
K2+G+I	156	20392.12	-10039.36
T92+G	156	20415.01	-10050.81
T92+G+I	157	20417.00	-10050.80
GTR+G	162	20395.02	-10034.76
TN93+G	159	20421.87	-10051.22
GTR+G+I	163	20396.83	-10034.66
TN93+G+I	160	20423.87	-10051.21
HKY+G	158	20459.31	-10070.94
HKY+G+I	159	20461.24	-10070.90
K2+I	155	20629.16	-10158.89
T92+I	156	20638.59	-10162.60
GTR+I	162	20614.42	-10144.46
TN93+I	159	20641.64	-10161.10
K2	154	20706.31	-10198.48
HKY+I	158	20684.82	-10183.70
T92	155	20716.23	-10202.43
TN93	158	20711.21	-10196.90
GTR	161	20688.12	-10182.32
HKY	157	20767.31	-10225.95
JC+G	154	21076.95	-10383.80
JC+G+I	155	21078.97	-10383.80
JC+I	154	21303.88	-10497.27
JC	153	21378.46	-10535.56
INCLUDING MOUSE			
Hornerin			
Model	**#Param**	**AICc**	**lnL**
K2+G	79	10199.61	-5020.48
T92+G	80	10200.29	-5019.82
K2+G+I	80	10201.63	-5020.48
T92+G+I	81	10202.31	-5019.82
TN93+G	83	10204.55	-5018.92
HKY+G	82	10216.19	-5025.75
GTR+G	86	10190.77	-5009.00
TN93+G+I	84	10206.57	-5018.92
HKY+G+I	83	10218.21	-5025.75
GTR+G+I	87	10192.79	-5009.00
K2+I	79	10281.81	-5061.59
T92+I	80	10282.22	-5060.78
HKY+I	82	10298.85	-5067.08
TN93+I	83	10294.72	-5064.01
GTR+I	86	10283.04	-5055.14

K2	78	10389.39	-5116.38
T92	79	10389.60	-5115.48
TN93	82	10392.25	-5113.78
JC+G	78	10429.57	-5136.47
HKY	81	10408.07	-5122.70
GTR	85	10379.82	-5104.54
JC+G+I	79	10431.58	-5136.47
JC+I	78	10502.37	-5172.87
JC	77	10603.69	-5224.54

Table 16. Detected positively selected codons and branches, and p-values.

SITE-BASED			
Gene	**Models compared**	**p-value**	**Positive selected codons**
HRNR	M1a vs M2a	2.31E-18	37, 51, 57, 116, 149, 231, 334, 353, 428
	M7 vs M8	5.93E-23	37, 51, 57, 107, 116, 128, 136, 149, 231, 322, 334, 340, 353, 393, 414, 428, 445

BRANCH-BASED TESTS			
Gene	**Models compared**	**p-value**	**Branches under independent or unique ω/various branches under same ω**
HRNR	M0 vs free ratio	0.0009	Independent ω

BRANCH-SITE MODEL			
Gene	**M0N0 vs M2N2**	**p-value**	**Positive selected codons**
HRNR	Macaque	8.59E-64	37, 51, 116, 149, 231, 334, 353, 393, 428
	Orangutan	1.27E-45	NO SITES
	Repeat 1	4.14E-49	228

Filaggrin 2

Table 17. Percentage of similarity between repeats, total numbers of pairs analyzed and numbers of pairs with higher synonymous and non-synonymous variations for filaggrin-2 gene

Gene	Species	Percentage of similarirty Minimum	Percentage of similarity Maximum	Total number of pairs analyzed	Pairs higher synonymous variations	Pairs higher non-synonymous variations
FLG-2-FLG-like	Human	74.89	91.71	91	54	37
	Pan troglodytes	73.18	90.78	78	45	33
	Macaque	74.86	98.22	91	72	19
	Papio anubis	90.97	99.11	231	200	31
	Callithrix jacchus	79.11	99.11	28	24	4
FLG-2-HRNR-like	Human	80.52	96	36	36	0
	Pan troglodytes	80.51	95.56	36	36	0
	Macaque	83.12	93.07	21	21	0
	Papio anubis	76.19	92.64	21	20	1
	Callithrix jacchus	79.74	91.15	10	10	0

Table 18. Nucleotide variation, average synonymous variations, average non-synonymous variations and average Ka/Ks from all repeats within a species

FLG2-FLG-like				
	Nucleotide variation	**Ks**	**Ka**	**Average (1-Ka/Ks)*100**
Human	0.14	0.17	0.16	-16.33
Pan troglodytes	0.15	0.17	0.16	-11.15
Gorilla gorilla	X	X	X	X
Pongo abelii	X	X	X	X
Macaque	0.12	0.19	0.12	28.47
Papio anubis	0.15	0.27	0.17	34.07
Callithrix jacchus	0.15	0.21	0.17	22.52
FLG2-HRNR-like				
	Nucleotide variation	**Ks**	**Ka**	**Average (1-Ka/Ks)*100**
Human	0.11	0.24	0.09	61.23
Pan troglodytes	0.11	0.24	0.08	58.37
Gorilla gorilla	X	X	X	X
Pongo abelii	X	X	X	X
Macaque	0.12	0.24	0.1	51.4
Papio anubis	0.15	0.32	0.12	53.56
Callithrix jacchus	0.12	0.34	0.08	69.35

[a]*K*s refers to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon.

[b]*K*a refers to the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon.

[c]*K*a/*K*s > 1 (bold) suggests positive selection.

[d]Purifying selection is measured by $(1 - Ka/Ks) \times 100$.

X= sequences not available in NCBI database
-= sequences not analyzed

Table 19. Ka/Ks ratio >1 for pairs of repeats within a species

Filaggrin2-filaggrin-like								
Repeat 1	Repeat 2	SynDif	SynPos	Ks	NSynDif	NSynPos	Ka	Ka/Ks
Baboon-repeat 11	Baboon-repeat 14	0.5	48.67	0.01	3.5	149.33	0.024	2.311
Baboon-repeat 13	Baboon-repeat 15	4	47	0.09	13	151	0.091	1.012
Baboon-repeat 1	Baboon-repeat 16	6.83	46.33	0.164	26.17	151.67	0.196	1.194
Baboon-repeat 15	Baboon-repeat 17	2	47.58	0.043	11	150.42	0.077	1.776
Baboon-repeat 15	Baboon-repeat 18	1	48.67	0.021	5	149.33	0.034	1.649
Baboon-repeat 17	Baboon-repeat 18	1	47.58	0.021	8	150.42	0.055	2.592
Baboon-repeat 13	Baboon-repeat 19	1	46.83	0.022	10	151.17	0.069	3.194
Baboon-repeat 13	Baboon-repeat 21	2	47.75	0.043	12	150.25	0.084	1.958
Baboon-repeat 13	Baboon-repeat 22	3.5	46.67	0.079	17.5	151.33	0.126	1.59
Baboon-repeat 19	Baboon-repeat 22	2.5	48.17	0.054	9.5	149.83	0.066	1.23
Baboon-repeat 1	Baboon-repeat 3	8.33	47.08	0.202	30.67	150.92	0.237	1.174
Baboon-repeat 16	Baboon-repeat 3	2	47.75	0.043	16	150.25	0.115	2.666
Baboon-repeat 22	Baboon-repeat 3	6	48.25	0.136	23	149.75	0.172	1.264
Baboon-repeat 1	Baboon-repeat 4	9.33	46.75	0.232	37.67	151.25	0.303	1.304
Baboon-repeat 12	Baboon-repeat 4	26.33	48	0.986	84.67	150	1.048	1.062
Baboon-repeat 16	Baboon-repeat 4	4	47.42	0.09	22	150.58	0.163	1.816
Baboon-repeat 19	Baboon-repeat 4	7.5	48.08	0.175	23.5	149.92	0.176	1.005
Baboon-repeat 2	Baboon-repeat 4	9	47	0.221	31	151	0.24	1.085
Baboon-repeat 21	Baboon-repeat 4	8	49	0.184	26	149	0.199	1.079
Baboon-repeat 22	Baboon-repeat 4	8	47.92	0.189	26	150.08	0.197	1.043
Baboon-repeat 3	Baboon-repeat 4	8.5	48.17	0.201	26.5	149.83	0.202	1.002
Baboon-repeat 2	Baboon-repeat 5	6	47.5	0.138	24	150.5	0.179	1.296
Baboon-repeat 4	Baboon-repeat 5	5.5	48.33	0.123	19.5	149.67	0.143	1.16

Baboon-repeat 13	Baboon-repeat 6	3	46.92	0.067	12	151.08	0.084	1.257
Baboon-repeat 2	Baboon-repeat 7	4	47.25	0.09	21	150.75	0.154	1.716
Baboon-repeat 3	Baboon-repeat 7	5	48.42	0.111	18	149.58	0.131	1.181
Baboon-repeat 4	Baboon-repeat 7	5.5	48.08	0.124	23.5	149.92	0.176	1.417
Baboon-repeat 1	Baboon-repeat 8	7.83	46.25	0.192	29.17	151.75	0.222	1.157
Baboon-repeat 16	Baboon-repeat 8	1	46.92	0.022	5	151.08	0.034	1.565
Baboon-repeat 3	Baboon-repeat 8	4	47.67	0.089	18	150.33	0.13	1.465
Baboon-repeat 4	Baboon-repeat 8	5	47.33	0.114	18	150.67	0.13	1.142
Chimpanzee-repeat 1	Chimpanzee-repeat 10r	9.83	53.75	0.21	32.17	168.25	0.221	1.052
Chimpanzee-repeat 10r	Chimpanzee-repeat 12	5.5	54.58	0.108	21.5	167.42	0.141	1.302
Chimpanzee-repeat 1	Chimpanzee-repeat 13	7.33	54	0.15	34.67	168	0.241	1.611
Chimpanzee-repeat 10r	Chimpanzee-repeat 13	2.5	54.58	0.047	19.5	167.42	0.127	2.677
Chimpanzee-repeat 12	Chimpanzee-repeat 13	6.5	54.83	0.129	24.5	167.17	0.163	1.264
Chimpanzee-repeat 1	Chimpanzee-repeat 2r	9.5	52.75	0.206	37.5	169.25	0.263	1.275
Chimpanzee-repeat 13	Chimpanzee-repeat 2r	8.5	53.58	0.178	34.5	168.42	0.239	1.343
Chimpanzee-repeat 1	Chimpanzee-repeat 3r	10.83	54	0.234	36.17	168	0.254	1.087
Chimpanzee-repeat 10r	Chimpanzee-repeat 3r	1.5	54.58	0.028	16.5	167.42	0.106	3.775
Chimpanzee-repeat 12	Chimpanzee-repeat 3r	7.5	54.83	0.151	23.5	167.17	0.156	1.031
Chimpanzee-repeat 13	Chimpanzee-repeat 3r	4	54.83	0.077	22	167.17	0.145	1.887
Chimpanzee-repeat 1	Chimpanzee-repeat 4	6.67	52.5	0.139	40.33	169.5	0.286	2.058
Chimpanzee-repeat 10r	Chimpanzee-repeat 4	4	53.08	0.079	18	168.92	0.115	1.447
Chimpanzee-repeat 12	Chimpanzee-repeat 4	8	53.33	0.167	26	168.67	0.173	1.031
Chimpanzee-repeat 13	Chimpanzee-repeat 4	2	53.33	0.039	21	168.67	0.136	3.535
Chimpanzee-repeat 2r	Chimpanzee-repeat 4	8	52.08	0.172	31	169.92	0.209	1.216
Chimpanzee-repeat 3r	Chimpanzee-repeat 4	6	53.33	0.122	22	168.67	0.143	1.176

Chimpanzee-repeat 1	Chimpanzee-repeat 5	9.83	52.83	0.214	40.17	169.17	0.286	1.335
Chimpanzee-repeat 10r	Chimpanzee-repeat 5	5	53.42	0.1	28	168.58	0.188	1.877
Chimpanzee-repeat 12	Chimpanzee-repeat 5	8	53.67	0.166	28	168.33	0.188	1.132
Chimpanzee-repeat 13	Chimpanzee-repeat 5	4.5	53.67	0.089	31.5	168.33	0.215	2.422
Chimpanzee-repeat 3r	Chimpanzee-repeat 5	4.5	53.67	0.089	23.5	168.33	0.155	1.738
Chimpanzee-repeat 4	Chimpanzee-repeat 5	5	52.17	0.103	30	169.83	0.201	1.965
Chimpanzee-repeat 1	Chimpanzee-repeat 6	7.83	53.42	0.163	32.17	168.58	0.22	1.349
Chimpanzee-repeat 10r	Chimpanzee-repeat 6	3.5	54	0.068	15.5	168	0.098	1.451
Chimpanzee-repeat 13	Chimpanzee-repeat 6	3	54.25	0.057	21	167.75	0.137	2.387
Chimpanzee-repeat 3r	Chimpanzee-repeat 6	5	54.25	0.098	20	167.75	0.13	1.32
Chimpanzee-repeat 4	Chimpanzee-repeat 6	5.5	52.75	0.112	19.5	169.25	0.125	1.114
Chimpanzee-repeat 13	Chimpanzee-repeat 7	6	54.25	0.12	26	167.75	0.174	1.45
Chimpanzee-repeat 3r	Chimpanzee-repeat 7	7	54.25	0.142	27	167.75	0.181	1.28
Chimpanzee-repeat 5	Chimpanzee-repeat 7	8.5	53.08	0.18	30.5	168.92	0.207	1.147
Chimpanzee-repeat 13	Chimpanzee-repeat 8	7	54.42	0.141	25	167.58	0.166	1.179
Chimpanzee-repeat 5	Chimpanzee-repeat 8	6.5	53.25	0.133	21.5	168.75	0.14	1.048
Human-repeat 1	Human-repeat 10	6.83	54.25	0.138	32.17	167.75	0.222	1.606
Human-repeat 1	Human-repeat 11	8.83	54.25	0.184	31.17	167.75	0.214	1.163
Human-repeat 11	Human-repeat 13	5.5	54.75	0.108	18.5	167.25	0.12	1.109
Human-repeat 1	Human-repeat 14	6.33	54.33	0.127	33.67	167.67	0.234	1.845
Human-repeat 10	Human-repeat 14	3	54.75	0.057	29	167.25	0.197	3.466
Human-repeat 11	Human-repeat 14	2.5	54.75	0.047	19.5	167.25	0.127	2.69
Human-repeat 13	Human-repeat 14	6.5	54.83	0.129	23.5	167.17	0.156	1.207
Human-repeat 1	Human-repeat 2	10.5	53.08	0.23	33.5	168.92	0.23	1.003
Human-repeat 14	Human-repeat 2	8	53.58	0.167	29	168.42	0.196	1.175

Human-repeat 1	Human-repeat 3	9.33	54.58	0.194	33.67	167.42	0.234	1.206
Human-repeat 10	Human-repeat 3	6.5	55	0.129	23.5	167	0.156	1.212
Human-repeat 11	Human-repeat 3	1.5	55	0.028	14.5	167	0.092	3.32
Human-repeat 13	Human-repeat 3	7	55.08	0.139	22	166.92	0.145	1.041
Human-repeat 14	Human-repeat 3	4	55.08	0.076	20	166.92	0.131	1.709
Human-repeat 1	Human-repeat 4	6	52.5	0.124	43	169.5	0.31	2.498
Human-repeat 10	Human-repeat 4	3	52.92	0.059	30	169.08	0.203	3.432
Human-repeat 11	Human-repeat 4	4	52.92	0.08	20	169.08	0.129	1.615
Human-repeat 13	Human-repeat 4	8	53	0.169	29	169	0.195	1.157
Human-repeat 14	Human-repeat 4	2	53	0.039	23	169	0.15	3.881
Human-repeat 2	Human-repeat 4	8.5	51.75	0.185	29.5	170.25	0.197	1.063
Human-repeat 3	Human-repeat 4	6	53.25	0.122	22	168.75	0.143	1.173
Human-repeat 1	Human-repeat 5	6.83	53.17	0.141	36.17	168.83	0.252	1.789
Human-repeat 10	Human-repeat 5	4.5	53.58	0.089	30.5	168.42	0.207	2.327
Human-repeat 11	Human-repeat 5	3.5	53.58	0.068	24.5	168.42	0.162	2.367
Human-repeat 13	Human-repeat 5	6	53.67	0.121	25	168.33	0.166	1.367
Human-repeat 14	Human-repeat 5	3	53.67	0.058	28	168.33	0.188	3.238
Human-repeat 3	Human-repeat 5	4	53.92	0.078	25	168.08	0.166	2.123
Human-repeat 4	Human-repeat 5	3.5	51.83	0.071	28.5	170.17	0.19	2.677
Human-repeat 1	Human-repeat 6	6.83	53.58	0.14	32.17	168.42	0.22	1.577
Human-repeat 10	Human-repeat 6	4.5	54	0.088	25.5	168	0.17	1.921
Human-repeat 11	Human-repeat 6	3.5	54	0.068	16.5	168	0.105	1.553
Human-repeat 14	Human-repeat 6	3	54.08	0.058	22	167.92	0.144	2.5
Human-repeat 3	Human-repeat 6	5	54.33	0.098	18	167.67	0.116	1.18
Human-repeat 4	Human-repeat 6	5.5	52.25	0.113	20.5	169.75	0.132	1.161

Human-repeat 14	Human-repeat 7	7.5	54.25	0.153	24.5	167.75	0.162	1.063
Human-repeat 1	Human-repeat 8	9.83	53.42	0.211	33.17	168.58	0.228	1.08
Human-repeat 14	Human-repeat 8	6.5	53.92	0.131	24.5	168.08	0.162	1.234
Macaque-repeat 10	Macaque-repeat 12	1	51.75	0.02	7	161.25	0.045	2.281
Macaque-repeat 1	Macaque-repeat 13	8.83	51.17	0.196	29.17	161.83	0.206	1.05
Macaque-repeat 10	Macaque-repeat 14	4	51.25	0.082	15	161.75	0.099	1.201
Macaque-repeat 12	Macaque-repeat 14	3	51.83	0.06	10	161.17	0.065	1.076
Macaque-repeat 13	Macaque-repeat 2	7.5	50.83	0.164	24.5	162.17	0.169	1.027
Macaque-repeat 1	Macaque-repeat 3	10.83	50.67	0.252	35.17	162.33	0.256	1.015
Macaque-repeat 13	Macaque-repeat 3	5	52	0.103	18	161	0.121	1.177
Macaque-repeat 1	Macaque-repeat 4	9.33	50.58	0.212	44.67	162.42	0.343	1.618
Macaque-repeat 13	Macaque-repeat 4	4.5	51.92	0.092	23.5	161.08	0.162	1.761
Macaque-repeat 2	Macaque-repeat 4	9.5	50.25	0.218	32.5	162.75	0.232	1.066
Macaque-repeat 1	Macaque-repeat 5	10.83	50.25	0.254	36.17	162.75	0.264	1.037
Macaque-repeat 3	Macaque-repeat 5	5	51.08	0.105	18	161.92	0.12	1.147
Macaque-repeat 1	Macaque-repeat 6	9.83	50.58	0.225	32.17	162.42	0.23	1.022
Macaque-repeat 3	Macaque-repeat 6	3	51.42	0.061	16	161.58	0.106	1.75
Macaque-repeat 14	Macaque-repeat 7	3	51.83	0.06	10	161.17	0.065	1.076
Macaque-repeat 10	Macaque-repeat 8	1	51.83	0.02	4	161.17	0.025	1.292
Macaque-repeat 14	Macaque-repeat 8	3	51.92	0.06	13	161.08	0.085	1.421
Macaque-repeat 3	Macaque-repeat 9	2	51	0.04	20	162	0.135	3.347
Macaque-repeat 4	Macaque-repeat 9	5.67	50.92	0.121	20.33	162.08	0.137	1.139
Marmoset-repeat 1	Marmoset-repeat 5r	5.67	51	0.12	19.33	162	0.13	1.081
Marmoset-repeat 3	Marmoset-repeat 5r	6.17	50.67	0.133	19.83	162.33	0.133	1.005
Marmoset-repeat 5r	Marmoset-repeat 6	6.17	51.67	0.13	20.83	161.33	0.142	1.09

Marmoset-repeat 5r	Marmoset-repeat 8	6.17	51.67	0.13	20.83	161.33	0.142	1.09

Filaggrin2-Hornerin-like								
Repeat 1	**Repeat 2**	**SynDif**	**SynPos**	**Ks**	**NSynDif**	**NSynPos**	**Ka**	**Ka/Ks**
Baboon-repeat 5	Baboon-repeat 7	7	56.83	0.13	22	168.17	0.14	1.07

[a]*K*s refers to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon.

[b]*K*a refers to the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon.

[c]*K*a/*K*s > 1 (bold) suggests positive selection.

[d]Purifying selection is measured by (1 − *K*a/*K*s) × 100.

Table 20. Best-fit model comparison for filaggrin-2

FLG2-Filaggrin-like			
Model	**#Param**	**AICc**	**lnL**
GTR+G	148	5980.14	-2840.64
GTR+G+I	149	5982.18	-2840.64
TN93+G	145	5982.91	-2845.09
K2+G	141	5983.28	-2849.34
TN93+G+I	146	5984.95	-2845.09
K2+G+I	142	5985.32	-2849.34
T92+G	142	5986.36	-2849.87
T92+G+I	143	5988.40	-2849.87
HKY+G	144	6005.85	-2857.57
HKY+G+I	145	6007.89	-2857.57
GTR+I	148	6078.68	-2889.91
TN93+I	145	6082.93	-2895.09
K2+I	141	6083.29	-2899.35
JC+G	140	6085.39	-2901.42
T92+I	142	6087.11	-2900.24
JC+G+I	141	6087.43	-2901.42
HKY+I	144	6104.88	-2907.09
GTR	147	6141.34	-2922.26
K2	140	6144.58	-2931.01
TN93	144	6146.45	-2927.87
T92	141	6150.86	-2933.13
HKY	143	6166.87	-2939.10
JC+I	140	6184.41	-2950.93
JC	139	6245.04	-2982.26
FLG2-Hornerin-like			
Model	**#Param**	**AICc**	**lnL**
GTR+G	80	3259.43	-1548.86
K2+G	73	3261.38	-1556.97
GTR+G+I	81	3261.46	-1548.85
K2+G+I	74	3263.41	-1556.97
T92+G	74	3263.58	-1557.05
T92+G+I	75	3265.61	-1557.05
HKY+G	76	3267.13	-1556.79
TN93+G	77	3267.20	-1555.80
HKY+G+I	77	3269.16	-1556.78
TN93+G+I	78	3269.24	-1555.80
GTR+I	80	3279.61	-1558.95
K2+I	73	3281.05	-1566.81
T92+I	74	3283.25	-1566.89
HKY+I	76	3286.28	-1566.37
TN93+I	77	3287.57	-1565.99
GTR	79	3308.93	-1574.63
K2	72	3315.03	-1584.82

T92	73	3317.27	-1584.92
HKY	75	3319.64	-1584.07
TN93	76	3321.38	-1583.91
JC+G	72	3340.73	-1597.67
JC+G+I	73	3342.76	-1597.66
JC+I	72	3359.91	-1607.26
JC	71	3393.06	-1624.85

INCLUDING MOUSE REPEATS

FLG2-Filaggrin-like

Model	#Param	AICc	lnL
K2+G	169	7319.35	-3489.07
K2+G+I	170	7321.32	-3489.04
GTR+G	176	7321.89	-3483.20
GTR+G+I	177	7323.87	-3483.17
T92+G	170	7325.87	-3491.31
T92+G+I	171	7327.81	-3491.26
TN93+G	173	7329.76	-3490.19
TN93+G+I	174	7331.73	-3490.16
HKY+G+I	173	7346.12	-3498.38
HKY+G	172	7351.94	-3502.31
K2+I	169	7396.78	-3527.78
GTR+I	176	7400.58	-3522.55
T92+I	170	7404.93	-3530.84
TN93+I	173	7406.66	-3528.65
HKY+I	172	7421.23	-3536.95
JC+G	168	7427.79	-3544.31
JC+G+I	169	7429.76	-3544.27
GTR	175	7442.52	-3544.54
K2	168	7445.42	-3553.12
TN93	172	7452.64	-3552.66
T92	169	7456.41	-3557.60
HKY	171	7467.50	-3561.11
JC+I	168	7503.78	-3582.30
JC	167	7551.06	-3606.96

Table 21. Detected positively selected codons and branches, and p-values.

SITE-BASED			
Gene	**Models compared**	**p-value**	**Positive selected codons**
FLG2/FLG	M1a vs M2a	1.10E-14	5, 35, 44, 71, 74, 76
	M7 vs M8	2.57E-18	5, 35, 40, 44, 71, 74, 76
FLG2/HRNR	M1a vs M2a	-	-
	M7 vs M8	-	-
BRANCH-BASED TESTS			
Gene	**Models compared**	**p-value**	**Branches under independent or unique ω/various branches under same ω**
FLG2/FLG	M0 vs free ratio	0.01	Independent ω
FLG2/HRNR	M0 vs free ratio	0.26	Unique ω/various branches under same ω
BRANCH-SITE MODEL			
Gene	**M0N0 vs M2N2**	**p-value**	**Positive selected codons**
FLG2/FLG	Marmoset	2.96E-32	24, 30
	Repeat 1	2.65E-46	5, 6, 35, 40, 44, 71, 74, 76
	Repeat 2	2.22E-29	5, 27
	Repeat 3	3.47E-28	NO SITES
	Repeat 4	8.38E-28	NO SITES
FLG2/HRNR	Repeat 1	3.11E-06	NO SITES

Repetin

Table 22. Percentage of similarity between repeats, total numbers of pairs analyzed and numbers of pairs with higher synonymous and non-synonymous variations for repetin gene

Gene	Species	Percentage of similarirty Minimum	Percentage of similarity Maximum	Total number of pairs analyzed	Pairs higher synonymous variations	Pairs higher non-synonymous variations
Repetin	Human	36.67	100	378	361	17
	Pan troglodytes	36.67	97.22	171	159	12
	Gorilla gorilla	40	100	351	329	22
	Pongo abelii	36.67	100	120	115	5
	Macaque	36.67	100	378	357	21

Table 23. Nucleotide variation, average synonymous variations, average non-synonymous variations and average Ka/Ks from all repeats within a species

	Repetin			
	Nucleotide variation	**Ks**	**Ka**	**Average (1-Ka/Ks)*100**
Human	0.21	0.49	0.2	58.48
Pan troglodytes	0.26	0.61	0.26	56.56
Gorilla gorilla	0.21	0.45	0.21	52.84
Pongo abelii	0.26	0.57	0.28	50.89
Macaque	0.19	0.39	0.19	50.2
Papio anubis	-	-	-	-
Callithrix jacchus	-	-	-	-

[a]*K*s refers to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon.

[b]*K*a refers to the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon.

[c]*K*a/*K*s > 1 (bold) suggests positive selection.

[d]Purifying selection is measured by $(1 - Ka/Ks) \times 100$.

X= sequences not available in NCBI database
-= sequences not analyzed

Table 24. Ka/Ks ratio >1 for pairs of repeats within a species

Repetin								
Repeat 1	**Repeat 2**	**SynDif**	**SynPos**	**Ks**	**NSynDif**	**NSynPos**	**Ka**	**Ka/Ks**
Human - Repeat 25	Human - Repeat 28	0.5	7.17	0.073	3.5	28.83	0.132	1.809
Human - Repeat 1	Human - Repeat 2	3.83	8.08	0.75	16.17	27.92	1.109	1.478
Human - Repeat 12	Human - Repeat 3	2.67	7.92	0.447	12.33	28.08	0.661	1.477
Human - Repeat 10	Human - Repeat 17	1	7.83	0.14	5	28.17	0.203	1.448
Human - Repeat 21	Human - Repeat 9	1	6.58	0.17	6	29.42	0.238	1.402
Human - Repeat 10	Human - Repeat 20	1	7.58	0.145	5	28.42	0.201	1.383
Human - Repeat 17	Human - Repeat 25	0.67	7.33	0.097	3.33	28.67	0.126	1.303
Human - Repeat 20	Human - Repeat 25	0.67	7.08	0.101	3.33	28.92	0.125	1.245
Human - Repeat 21	Human - Repeat 5	0.5	7	0.075	2.5	29	0.092	1.22
Human - Repeat 23	Human - Repeat 25	0.5	7	0.075	2.5	29	0.092	1.22
Human - Repeat 3	Human - Repeat 8	2.67	8.17	0.429	10.33	27.83	0.512	1.195
Human -	Human -	1	6.75	0.165	5	29.25	0.194	1.176

Repeat 5	Repeat 9							
Human - Repeat 22	Human - Repeat 28	1	7.75	0.142	4	28.25	0.157	1.108
Human - Repeat 10	Human - Repeat 23	1	7.5	0.147	4	28.5	0.155	1.059
Human - Repeat 14	Human - Repeat 17	1	7.5	0.147	4	28.5	0.155	1.059
Human - Repeat 25	Human - Repeat 3	3.17	7.67	0.6	11.83	28.33	0.61	1.017
Human - Repeat 14	Human - Repeat 20	1	7.25	0.152	4	28.75	0.154	1.01
Chimp anzee-repeat16	Chimp anzee-repeat19	0.5	7.17	0.073	3.5	28.83	0.132	1.809
Chimp anzee-repeat10	Chimp anzee-repeat7	0.5	7	0.075	3.5	29	0.132	1.752
Chimp anzee-repeat14	Chimp anzee-repeat16	0.5	6.67	0.079	3.5	29.33	0.13	1.646
Chimp anzee-repeat12	Chimp anzee-repeat3	2.67	8	0.441	12.33	28	0.664	1.506
Chimp anzee-repeat1	Chimp anzee-repeat2	3.83	8.08	0.75	16.17	27.92	1.109	1.478
Chimp anzee-repeat3	Chimp anzee-repeat8	2.67	7.92	0.447	11.33	28.08	0.579	1.295
Chimp anzee-repeat15	Chimp anzee-repeat7	0.67	7	0.102	3.33	29	0.125	1.225
Chimp anzee-	Chimp anzee-	0.5	7	0.075	2.5	29	0.092	1.22

repeat4	repeat7							
Chimpanzee-repeat5	Chimpanzee-repeat9	1	6.75	0.165	5	29.25	0.194	1.176
Chimpanzee-repeat10	Chimpanzee-repeat17	1	7.83	0.14	4	28.17	0.157	1.125
Chimpanzee-repeat16	Chimpanzee-repeat3	3.17	7.42	0.632	12.83	28.58	0.685	1.084
Chimpanzee-repeat18	Chimpanzee-repeat2	3.33	8.08	0.599	11.67	27.92	0.611	1.021
Gorilla-repeat13	Gorilla-repeat14	0.5	7.33	0.072	4.5	28.67	0.176	2.463
Gorilla-repeat14	Gorilla-repeat6	0.5	7.33	0.072	4.5	28.67	0.176	2.463
Gorilla-repeat24	Gorilla-repeat27	0.5	7.17	0.073	3.5	28.83	0.132	1.809
Gorilla-repeat1	Gorilla-repeat2	3.83	8.08	0.75	16.17	27.92	1.109	1.478
Gorilla-repeat11	Gorilla-repeat3	2.67	8.25	0.423	11.33	27.75	0.59	1.395
Gorilla-repeat23	Gorilla-repeat24	0.67	7.67	0.092	3.33	28.33	0.128	1.385
Gorilla-repeat14	Gorilla-repeat25	0.5	7.67	0.068	2.5	28.33	0.094	1.377
Gorilla-repeat14	Gorilla-repeat4	0.5	7.67	0.068	2.5	28.33	0.094	1.377
Gorilla-repeat14	Gorilla-repeat9	0.5	7.67	0.068	2.5	28.33	0.094	1.377
Gorilla-repeat16	Gorilla-repeat24	0.67	7.33	0.097	3.33	28.67	0.126	1.303
Gorilla-repeat22	Gorilla-repeat24	0.5	7	0.075	2.5	29	0.092	1.22
Gorilla	Gorilla	1	8.17	0.134	4	27.83	0.16	1.193

-repeat 23	-repeat 9							
Gorilla -repeat 18	Gorilla -repeat 23	1	8.08	0.135	4	27.92	0.159	1.176
Gorilla -repeat 13	Gorilla -repeat 23	1	7.83	0.14	4	28.17	0.157	1.125
Gorilla -repeat 16	Gorilla -repeat 9	1	7.83	0.14	4	28.17	0.157	1.125
Gorilla -repeat 21	Gorilla -repeat 27	1	7.83	0.14	4	28.17	0.157	1.125
Gorilla -repeat 23	Gorilla -repeat 6	1	7.83	0.14	4	28.17	0.157	1.125
Gorilla -repeat 16	Gorilla -repeat 18	1	7.75	0.142	4	28.25	0.157	1.108
Gorilla -repeat 14	Gorilla -repeat 18	1	7.58	0.145	4	28.42	0.156	1.075
Gorilla -repeat 13	Gorilla -repeat 16	1	7.5	0.147	4	28.5	0.155	1.059
Gorilla -repeat 16	Gorilla -repeat 6	1	7.5	0.147	4	28.5	0.155	1.059
Gorilla -repeat 24	Gorilla -repeat 3	3.17	7.67	0.6	11.83	28.33	0.61	1.017
Orangu tan-repeat 1	Orangu tan-repeat 2	3.83	8.42	0.701	17.17	27.58	1.328	1.895
Orangu tan-repeat 10	Orangu tan-repeat 16	1	8.08	0.135	6	27.92	0.253	1.874
Orangu tan-repeat 11	Orangu tan-repeat 16	0.5	7.17	0.073	3.5	28.83	0.132	1.809
Orangu tan-repeat 12	Orangu tan-repeat 16	0.5	7.17	0.073	3.5	28.83	0.132	1.809
Orangu tan-	Orangu tan-	0.5	6.33	0.083	2.5	29.67	0.089	1.072

repeat 5	repeat 9							
Macaque-Repeat 15	Macaque-Repeat 28	0.5	7.42	0.071	4.5	28.58	0.177	2.503
Macaque-Repeat 28	Macaque-Repeat 8	0.5	7.42	0.071	4.5	28.58	0.177	2.503
Macaque-Repeat 17	Macaque-Repeat 28	1	7.58	0.145	5	28.42	0.201	1.383
Macaque-Repeat 1	Macaque-Repeat 2	3.83	7.42	0.876	17.17	28.58	1.21	1.381
Macaque-Repeat 15	Macaque-Repeat 9	1	7.42	0.149	5	28.58	0.199	1.341
Macaque-Repeat 8	Macaque-Repeat 9	1	7.42	0.149	5	28.58	0.199	1.341
Macaque-Repeat 28	Macaque-Repeat 5	1	7.25	0.152	5	28.75	0.198	1.299
Macaque-Repeat 12	Macaque-Repeat 15	0.5	7.33	0.072	2.5	28.67	0.093	1.297
Macaque-Repeat 12	Macaque-Repeat 8	0.5	7.33	0.072	2.5	28.67	0.093	1.297
Macaque-Repeat 11	Macaque-Repeat 15	0.67	7	0.102	3.33	29	0.125	1.225
Macaque-Repeat 11	Macaque-Repeat 8	0.67	7	0.102	3.33	29	0.125	1.225
Macaque-Repeat 14	Macaque-Repeat 15	0.5	7	0.075	2.5	29	0.092	1.22

Macaque-Repeat 14	Macaque-Repeat 8	0.5	7	0.075	2.5	29	0.092	1.22
Macaque-Repeat 6	Macaque-Repeat 7	0.5	6.67	0.079	2.5	29.33	0.091	1.146
Macaque-Repeat 18	Macaque-Repeat 28	1	7.92	0.138	4	28.08	0.158	1.142
Macaque-Repeat 26	Macaque-Repeat 28	1	7.67	0.143	4	28.33	0.156	1.091
Macaque-Repeat 10	Macaque-Repeat 28	1	7.58	0.145	4	28.42	0.156	1.075
Macaque-Repeat 24	Macaque-Repeat 28	1	7.58	0.145	4	28.42	0.156	1.075
Macaque-Repeat 28	Macaque-Repeat 7	1	7.58	0.145	4	28.42	0.156	1.075
Macaque-Repeat 5	Macaque-Repeat 6	0.5	6.33	0.083	2.5	29.67	0.089	1.072
Macaque-Repeat 25	Macaque-Repeat 9	1	7.42	0.149	4	28.58	0.155	1.042

[a]*K*s refers to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon.

[b]*K*a refers to the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon.

[c]*K*a/*K*s > 1 (bold) suggests positive selection.

[d]Purifying selection is measured by (1 − *K*a/*K*s) × 100.

Table 25. Best-fit model comparison for repetin

Repetin			
Model	**#Param**	**AICc**	**lnL**
K2+G	235	1736.23	-619.29
K2+G+I	236	1737.28	-618.69
T92+I	236	1737.68	-618.90
K2	234	1737.88	-621.24
T92+G	236	1737.89	-619.00
K2+I	235	1738.24	-620.30
T92	235	1738.48	-620.42
T92+G+I	237	1739.31	-618.59
HKY+I	238	1740.68	-618.15
HKY+G	238	1741.30	-618.46
HKY	237	1741.54	-619.70
HKY+G+I	239	1741.56	-617.47
TN93+I	239	1742.37	-617.88
TN93+G	239	1744.26	-618.82
TN93	238	1744.46	-620.04
TN93+G+I	240	1744.47	-617.80
GTR+I	242	1748.38	-617.51
GTR	241	1749.68	-619.28
GTR+G	242	1750.19	-618.41
GTR+G+I	243	1750.49	-617.43
JC+I	234	1805.86	-655.23
JC	233	1805.97	-656.40
JC+G	234	1806.00	-655.30
JC+G+I	235	1807.90	-655.13

INCLUDING MOUSE REPEATS

Repetin			
Model	**#Param**	**AICc**	**lnL**
HKY+G	334	2811.94	-1052.03
TN93+G	335	2812.75	-1051.31
HKY+G+I	335	2813.65	-1051.76
TN93+G+I	336	2814.12	-1050.87
K2+G	331	2814.32	-1056.58
K2+G+I	332	2815.03	-1055.82
HKY+I	334	2815.50	-1053.81
HKY	333	2815.71	-1055.04
GTR+G	338	2815.79	-1049.46
TN93+I	335	2816.26	-1053.07
K2+I	331	2816.50	-1057.67
TN93	334	2817.14	-1054.63
T92+G	332	2817.38	-1056.99
GTR+G+I	339	2817.71	-1049.29
T92+G+I	333	2818.16	-1056.26

K2	330	2818.83	-1059.96
T92+I	332	2819.77	-1058.19
GTR+I	338	2821.29	-1052.21
GTR	337	2821.57	-1053.47
T92	331	2822.02	-1060.43
JC+G	330	2894.89	-1097.99
JC+G+I	331	2895.69	-1097.27
JC+I	330	2896.99	-1099.04
JC	329	2899.07	-1101.20

Table 26. Detected positively selected codons and branches, and p-values.

SITE-BASED			
Gene	**Models compared**	**p-value**	**Positive selected codons**
RPTN	M1a vs M2a	2.01E-11	2
	M7 vs M8	2.25E-12	2
BRANCH-BASED TESTS			
Gene	**Models compared**	**p-value**	**Branches under independent or unique ω/various branches under same ω**
RPTN	M0 vs free ratio	0.0003	Independent ω
BRANCH-SITE MODEL			
Gene	**M0N0 vs M2N2**	**p-value**	**Positive selected codons**
RPTN	Repeat 1	0.6	-
	Repeat 2	0.15	-
	Repeat 3	2.93E-12	2

Cornulin

Table 27. Percentage of similarity between repeats, total numbers of pairs analyzed and numbers of pairs with higher synonymous and non-synonymous variations for cornulin gene

Gene	Species	Percentage of similarirty Minimum	Percentage of similarity Maximum	Total number of pairs analyzed	Pairs higher synonymous variations	Pairs higher non-synonymous variations
Cornulin	Human	91.67		1	1	0
	Pan troglodytes	91.11	98.33	6	6	0
	Gorilla gorilla	-		x	x	x
	Pongo abelii	89.44		1	1	0
	Macaque	90.96		1	1	0

Table 28. Nucleotide variation, average synonymous variations, average non-synonymous variations and average Ka/Ks from all repeats within a species

	Cornulin			
	Nucleotide variation	**Ks**	**Ka**	**Average (1-Ka/Ks)*100**
Human	0.08	0.12	0.08	35.35
Pan troglodytes	0.05	0.08	0.05	43.37
Gorilla gorilla	0	0	0	0
Pongo abelii	0.11	0.14	0.11	22.39
Macaque	0.08	0.11	0.08	22.24
Papio anubis	-	-	-	-
Callithrix jacchus	-	-	-	-

[a]*K*s refers to the number of synonymous nucleotide substitutions per total number of synonymous sites for each codon.

[b]*K*a refers to the number of nonsynonymous nucleotide substitutions per total number of nonsynonymous sites for each codon.

[c]*K*a/*K*s > 1 (bold) suggests positive selection.

[d]Purifying selection is measured by (1 − *K*a/*K*s) × 100.

X= sequences not available in NCBI database
-= sequences not analyzed

Table 29. Best-fit model comparison for cornulin

Cornulin			
Model	**#Param**	**AICc**	**lnL**
TN93+G	25	976.85	-463.09
HKY+G	24	977.89	-464.64
TN93+G+I	26	978.91	-463.09
TN93	24	979.74	-465.56
HKY+G+I	25	979.95	-464.64
GTR+I	28	980.76	-461.96
GTR+G	28	981.28	-462.22
TN93+I	25	981.83	-465.58
HKY	23	982.37	-467.90
GTR+G+I	29	983.34	-462.22
GTR	27	984.28	-464.75
HKY+I	24	984.42	-467.90
T92+I	22	1004.92	-480.20
T92+G	22	1005.96	-480.72
T92+G+I	23	1008.01	-480.72
T92	21	1010.77	-484.14
K2+G	21	1011.02	-484.27
K2+G+I	22	1013.06	-484.27
K2	20	1015.42	-487.49
K2+I	21	1017.44	-487.48
JC+G	20	1029.10	-494.33
JC+G+I	21	1031.14	-494.33
JC+I	20	1031.51	-495.53
JC	19	1033.08	-497.34

INCLUDING MOUSE REPEATS

Cornulin			
Model	**#Param**	**AICc**	**lnL**
GTR+G	32	1589.56	-762.31
GTR	31	1590.59	-763.86
TN93+G	29	1591.33	-766.28
GTR+G+I	33	1591.62	-762.31
GTR+I	32	1591.82	-763.44
TN93+I	29	1592.50	-766.87
TN93+G+I	30	1593.39	-766.28
TN93	28	1594.83	-769.06
HKY+G	28	1598.15	-770.72
HKY+I	28	1598.51	-770.90
K2+G	25	1601.29	-775.36
HKY+G+I	29	1602.68	-771.96
K2+G+I	26	1603.34	-775.36
HKY	27	1604.07	-774.70
T92+G	26	1605.65	-776.52

T92+G+I	27	1607.70	-776.52
K2	24	1609.05	-780.26
K2+I	25	1609.23	-779.33
T92	25	1612.75	-781.09
T92+I	26	1615.62	-781.50
JC+G	24	1642.43	-796.95
JC+G+I	25	1644.44	-796.94
JC	23	1647.16	-800.34
JC+I	24	1648.32	-799.90

Table 30. Detected positively selected codons and branches, and p-values.

SITE-BASED			
Gene	**Models compared**	**p-value**	**Positive selected codons**
CRNN	M1a vs M2a	-	-
	M7 vs M8	-	-
BRANCH-BASED TESTS			
Gene	**Models compared**	**p-value**	**Branches under independent or unique ω/various branches under same ω**
CRNN	M0 vs free ratio	0.29	Unique ω/various branches under same ω
BRANCH-SITE MODEL			
Gene	**M0N0 vs M2N2**	**p-value**	**Positive selected codons**
CRNN	Repeat 1	0.11	-
	Repeat 2	0.25	-

Table 31. Copy-number variation of FLG in patients with atopic dermatitis and controls. **(A)** Repeat-allele variation **(B)** Repeat-number variation

A.

Repeat-allele	Control	Percentage (%)	Cases	Percentage (%)	OR	p-value	CI	
10	50	16.23	59	16.67	0.91	0.74	0.58	1.45
11	124	40.26	122	34.46	0.76	0.12	0.54	1.08
12	134	43.51	173	48.87	1			
TOTAL	308		354					

B.

Repeat-allele	Control	Percentage (%)	Cases	Percentage (%)	OR	p-value	CI	
10	50	16.23	59	16.67	0.91	0.74	0.58	1.45
11	124	40.26	122	34.46	0.76	0.12	0.54	1.08
12	134	43.51	173	48.87	1			
TOTAL	308		354					

OR= odds ratio
CI= confidence interval

Table 32. FLG copy-number variation comparison with SCORAD scale in patients with AD

A.

Repeat-number	Mild	Percentage (%)	Moderate	Percentage (%)	Severe	Percentage (%)
20	0	0	4	2.29	0	0
21	6	3.43	14	8.00	2	1.14
22	18	10.29	27	15.43	8	4.57
23	16	9.14	25	14.29	8	4.57
24	13	7.43	23	13.14	11	6.29
TOTAL	53		93		29	

B.

Mild vs Severe allele comparison								
Allele-number	**Mild**	**Severe**	**OR**	**p-val**	**CI**		**chi-square**	**p-value**
10	19	6	1.95	0.24	0.66	6.56	1.16	0.28
12	55	34						
11	32	18	1.10	0.86	0.51	2.42	0.01	0.94
12	55	34						
10	19	6	1.77	0.43	0.55	6.42	0.62	0.43
11	32	18						

Moderate vs Severe allele comparison								
Allele-number	**Moderate**	**Severe**	**OR**	**p-val**	**CI**		**chi-square**	**p-value**
10	34	6	2.31	0.09	0.85	7.35	2.41	0.12
12	83	34						
11	69	18	1.57	0.20	0.78	3.22	1.43	0.23
12	83	34						
10	34	6	1.47	0.63	0.50	4.96	0.27	0.61
11	69	18						

OR= odds ratio
CI= confidence interval

Table 33. Genotype among patients and controls

Polymorphism	Genotype	Cases	Controls	OR	p-value	CI	
Common FLG-mutations in Europeans							
R501X	A	328	266				
	a	2	0				
c2882del4	A	330	266				
	a	0	0				
R2447X	A	330	266				
	a	0	0				
3702delG	A	330	266				
	a	0	0				
S3247X	A	330	266				
	a	0	0				
Mutations in Ecuadorian							
E2250Q	A	279	246	0.45	0.003	0.24	0.79
	a	51	20				
E2652D	A	242	218	0.61	0.014	0.4	0.92
	a	88	48				

A= Wild-type allele
a= Mutated allele
OR= odds ratio
CI= confidence interval

Table 34. Allele frequencies among cases and controls

Polymorphism	Genotype	Frequency cases	Frequency controls	Polyphen	Ex_Freq	Ex_AMR	1000G_all	1000_amr
E2250Q	A	0.85	0.92	0.81				
	a	0.15	0.08	Possibly damaging	0.02	0.02	0	0
E2652D	A	0.73	0.82	0.672				
	a	0.27	0.18	Possibly damaging	0.34	0.26	0.19	0.17

Figures

Figure 1. Schematic representation of divergent, concerted and birth-and-death evolutionary models

Figure 2. Structure of SFTPs and organization in the EDC.

Figure 3. Schematic representation of the SFTPs and phylogenic origin. (A) Schematic representation of the SFTPs Dark gray= A domain corresponds to the N-terminal domain containing the EF-hands. Light gray= B domain in profilaggrin. Black= spacer or linker sequences. Dash black and white= C-terminal domains. White= repeated region. Interrupted lines in TCHH represent the different domains. (B) Phylogenic origin of the SFTPs based on the S100-domain. Cornulin ancestor can be first found in amniotes, sauropsids have a similar TCHH-like protein (SCFN) and filaggrin-type appeared in mammalians. CRNN was lost in the lizard branch.

Figure 4. Diagram of filaggrin gene structure and repeated region. Human filaggrin gene includes 3 exons and 2 introns, with the repeat region being found on the third. The rest of the human filaggrin gene comprises exon 1 with 54bp, intron 1 with 9613bp, exon 2 with 159bp, an intron 2 with 570bp and, in exon 3 a Domain A, which contains two calcium-binding domain, and Domain B, which facilitates the translocation in terminally differentiating keratinocytes, with 879bp, for a total of 11275bp which we further refer this region as the "non-repeated region". Neither exons other than exon 3 nor introns have filaggrin repeated units. The filaggrin repeat region consists of complete repeats that are flanked by two partial repeats. Black = partial repeats and white = complete repeats. UTR = Untranslated region.

Figure 5. Dot-plot matrices between the human repeat region (*x*-axis) and primate exon 3 (*y*-axis) sequences of the filaggrin gene downloaded from the National Center for Biotechnology Information database. (**A**) Dot-plot matrix between the human repeat region and orangutan exon 3 sequences, compared using the nucleotide method with a 5-unit size filter option; misalignment regions were found between 5,200 and 9,100 bp, as shown in Figure 1. (**B**) Dot-plot matrix between the human repeat region and gorilla exon 3 sequences, using the nucleotide method with a 5-unit size filter

option; misalignment regions (shown as red lines) were found at the end of the gorilla sequence after 9,000 bp, as shown in Figure 1. (**C**) Dot-plot matrix between the chimpanzee FLG-like sequence and chimpanzee-repeat 1, using the nucleotide method with a 5-unit size filter option; filaggrin repeat units are shown as yellow lines. Filaggrin-like gene in chimpanzee includes 1 complete repeat between two incomplete repeats. (**D**) Dot-plot matrix between the crab-eating macaque FLG-like sequence located in chromosome 1 and macaque-repeat 1, using the nucleotide method with a 5-unit size filter option; filaggrin repeat units are shown as purple lines. Filaggrin-like gene in macaque located in chromosome 1 includes 5 complete filaggrin repeats. (**E**) Dot-plot matrix between the crab-eating macaque FLG-like sequence located in chromosome 3 and macaque-repeat 1, using the nucleotide method with a 5-unit size filter option; none filaggrin repeat units were found.

Figure 6. Partial and complete repeat sequences of *FLG* in human, chimpanzee, gorilla, orangutan, and crab-eating macaque. Left shows the repeat order of *FLG* in these primates based on the sequences from the National Center for Biotechnology Information database; the gorilla sequence had an unframed repeat at the end of the sequence, and the orangutan sequence had unframed repeats in the middle. Right shows the repeat order of *FLG* sequences acquired through the use of both PacBio RSII and MiSeq in this study. Black = partial repeats, white = complete repeats, and gray = regions with gaps. The nucleotide length of each repeat is shown beneath each repeat. The number of repeats varies between species.

Figure 7. Maximum likelihood tree reconstruction using all complete repeats for filaggrin

Figure 8. Neighbor-joining tree reconstruction using all complete repeats for filaggrin

Figure 9. "Reconciled" trees using complete repeats of *FLG* in human, chimpanzee, gorilla, orangutan, and crab-eating macaque. (**A**) "Reconciled" gene tree indicating duplication (red squares) and loss events (light gray italics) from the most common ancestor of these primates: the crab-eating macaque repeats duplicated 11 times, orangutan repeats duplicated 8 times, and gorilla/chimpanzee/human repeats

duplicated 18 times, and in human, the counterpart to chimpanzee-repeat 9 was lost, while in chimpanzee, the counterparts to human-repeats 10 and 8 were lost, and in gorilla, the counterpart to human-repeat 2, 9 and 8 and chimpanzee-repeat 8 was lost. (**B**) "Reconciled" species tree indicating duplication and loss events in each species from their most common ancestor: crab-eating macaque repeats duplicated 11 times, orangutan repeats duplicated 8 times, gorilla repeats duplicated 6 times, and chimpanzee repeats duplicated 3 times, while the most common ancestor between human and chimpanzee duplicated 6 times, and the most common ancestor between human, chimpanzee, and gorilla duplicated 4 times; and 1 repeat was lost in gorilla, 2 repeats were lost in chimpanzee, and 1 repeat was lost in human. The number of repeats found in each species is provided in parentheses.

Figure 10. Phylogenetic tree reconstructions using all complete repeats of the filaggrin gene in human, chimpanzee, gorilla, orangutan, crab-eating macaque, dog and mouse. (**A**) Maximum likelihood tree reconstruction using all complete repeats in human, chimpanzee, gorilla, orangutan, crab-eating macaque, dog, and mouse and the following parameters: partial deletion option, Tamura-Nei model with gamma distribution and invariable sites, nearest-neighbor-interchange heuristic method, 1,000 bootstrap resampling, and a cutoff value of 50%. (**B**) Neighbor-joining-tree reconstruction using all complete repeats of *FLG* in human, chimpanzee, gorilla, orangutan, crab-eating macaque, dog, and mouse. The following parameters were used: pairwise deletion, proportional nucleotide differences, 1,000 bootstrap resampling, and a cutoff value of 50%. Bootstrap values are shown at the beginning of each branch. (**C**) A "Reconciled" tree using the complete repeats of the filaggrin gene in human, chimpanzee, gorilla, orangutan, crab-eating macaque, dog, and mouse. This **"**reconciled" tree was able to detect duplication (red squares) and loss events (light gray italics) from the most common ancestor of these species. Mouse repeats duplicated 15 times, dog repeats duplicated 3 times, crab-eating macaque repeats duplicated 11 times, orangutan repeats duplicated 8 times, and gorilla/chimpanzee/human repeats duplicated 18 times. In human, the counterpart to chimpanzee-repeat 9 has been lost, while in chimpanzee; the counterparts to human-repeats 10 and 8 have been lost.

Figure 11. "Divergence" tree reconstruction using all complete repeats of *FLG* in five primate species. The following parameters were used: the site model TN93, a substitution rate with gamma distribution = 4, a log-normal relaxed clock, and the birth-and-death model for all three nucleotide positions. The *x*-axis scale is time in Mya. The crab-eating macaque ancestor repeat diverged around 28 Mya, while the orangutan ancestor repeat diverged around 21 Mya. The gorilla/chimpanzee/human repeats duplicated during the last 9 Mya, while the crab-eating macaque repeats duplicated in the last 5.5 Mya.

Figure 12. Repeat order of gorilla, chimpanzee, and human *FLG* repeats, as inferred by phylogenetic and "reconciled" tree. Gorilla duplicated repeats "C," "D," and "G," chimpanzee and human duplicated repeat "A," and human duplicated repeat "B" are shown. In chimpanzee, the counterpart to human-repeats B-2 has been lost, while in human, the counterpart to chimpanzee-repeat "I" has been lost, and in chimpanzee, the counterparts to human and gorilla-repeat H have been lost. In gorilla, the counterpart to human and chimpanzee repeats B has been lost.

Figure 13. (A) Gel electrophoresis picture of filaggrin repeated regions of all primates. Primers used are described in Table S1. Ladder used is λ DNA *Hin*d *III*. Human repeated region vary from 12000 to 14000 bp depending on the number of repeats. Repeat variation is also seen in chimpanzee and macaque. **Upper** Gel electrophoresis sample order: Chimpanzee GAIN ID: 0143, 0212, 0158, 0170, 0204, 0211, 0131, 0279, 0169, 0276, 0159 0345, and Primate ID: 954 and 956. **Middle** Gel electrophoresis sample order: Gorilla Primate ID: 1943 and 3846, and GAIN ID:0080, Orangutan Primate ID: 2541 and GAIN ID: 0091, 0031, 0110 and 0010, *Macacca fascicularis* ID:181 and 246 and *Macacca mulatta* ID: 725 and 970. **Lower** Gel electrophoresis sample order: First four samples include PCR product for human repeated region: homozygote for 12 repeats, heterozygote for 11 and 12 repeats, heterozygote for 10 and 11 repeats, heterozygote for 10 and 12 repeats. Final four samples shows the repeat variation region: homozygote for 12 repeats, heterozygote for 11 and 12 repeats, heterozygote for 10 and 11 repeats, heterozygote for 10 and 12 repeats.

(B) Gel electrophoresis picture of non-repeated of all primates. Primers used are described in additional table 2. Ladder used is λ DNA *Hin*d *III*. In human, the

reported length is around 11000 bp, which is shown here to be the same for chimpanzee, gorilla, orangutan and macaque. **Upper** Gel electrophoresis sample order: Human, Chimpanzee GAIN ID: 0143, 0212, 0158, 0170, 0204, 0211, 0131, 0279, 0169, 0276, 0159 0345, and Primate ID: 954 and 956. Lower gel sample order: Gorilla Primate ID: 1943 and 3846, and GAIN ID:0080, Orangutan Primate ID: 2541 and GAIN ID: 0091, 0031, 0110 and 0010, *Macacca fascicularis* ID:181 and 246 and *Macacca mulatta* ID: 725 and 970.

Figure 14. Neighbor-joining tree reconstruction using complete repeats of the filaggrin gene in five primate species without possible recombinant repeats. Repeats that were found to have a possible recombinant region or gene conversion were excluded. The following parameters were used: pairwise deletion, proportional nucleotide differences, 1,000 bootstrap resampling, and a cutoff value of 50%.

Figure 15. Phylogenetic tree reconstructions using filaggrin repeats of human, chimpanzee, gorilla, orangutan, macaque, baboon, and marmoset (A) Maximum likelihood tree reconstruction (B) Neighbor-joining tree reconstruction

Hornerin

Figure 16. Schematic representation of hornerin repeated region formation.

Figure 17. Longest, second-longest, quartic, tertiary and secondary units of hornerin in human, chimpanzee, gorilla, orangutan, and crab-eating macaque. The nucleotide length of each repeat is shown beneath each repeat. The number of repeats varies between species.

Figure 18. Macaque and mouse, and human initial-117 bp unit dot-matrix alignment. (A) Dot-matrix comparison of human initial 117 bp units from longest repeat-4 with the longest unit of macaque repeat-4. The 117 bp units blasted in various regions of the macaque repeat-4 with a low similarity, therefore macaque initial 117 bp units analysis was excluded. (B) Mouse unit 1 A and B aligned to human longest unit 1. Mouse unit 1 A aligned in three regions the first from 590 to 704 bp, the second from 625 to 662 bp and the third from 1333 to 1369, and mouse unit B did not show any

alignment. (C) Mouse unit 1 A and B aligned to human initial-117 bp unit with a low alignment threshold. There are various shorter than 117 bp regions matches.

Figure 19. Human initial-117 bp units and primates longest repeat graphic alignment. Graphic alignment comparison of human initial-117 bp units from longest repeat-4 with the longest unit of human repeat-4, chimpanzee repeat-4, gorilla repeat-2 and orangutan repeat-4. Repeat-4 of human was used, as it is the only complete unit described by Takaishi et al. (2005).

Figure 20. Phylogenetic tree analyses for hornerin longest (A) Maximum likelihood reconstruction of longest repeats, (B) Neighbor-joining tree reconstruction of longest repeats, (C) Maximum likelihood reconstruction of longest repeats adding mouse repeats, (D) "Reconciled" trees using the longest repeats.

Figure 21. Maximum likelihood reconstruction of second-longest repeats. The second-longest units gathered by subunit type.

Figure 22. Phylogenetic tree analyses for hornerin initial-117 bp units (A) Maximum likelihood reconstruction of initial-117 bp unit in human, (B) Maximum likelihood reconstruction of initial-117 bp unit in chimpanzee, (C) Maximum likelihood reconstruction of initial-117 bp unit in gorilla, (D) Maximum likelihood reconstruction of initial-117 bp unit in orangutan, (E) Maximum likelihood reconstruction of initial-117 bp in human, chimpanzee, gorilla and orangutan. Boxed initial-117 bp units are either human or clustered on a different type.

Figure 23. Quartic, tertiary, secondary and primary units reconstructed by using the clusters from the 39 bp units' phylogenetic tree analysis. (A) Human, chimpanzee, gorilla and orangutan graphic representation of the types of quartic, tertiary, secondary and primary units. White color = conserved. Light grey = partially displaced unit, for example a v01n01 secondary unit which is placed in the cluster of n01 type but is misplaced in the v02 cluster. Dark grey = misplaced unit, for example a v01n01 secondary unit which is misplaced in the cluster of n02 type and v02 type cluster or a v01 tertiary unit misplaced in the v02 cluster. Black = non-existent unit (B) Crab-eating macaque graphic representation of the types of quartic and primary

units. Color gradient from white to grey for the primary unit order found in the primary phylogenetic tree analysis. Crab-eating macaque does not duplicate in ninths but has unique duplications and losses.

Figure 24. Maximum-likelihood tree analyses for hornerin primary units in primates. (A) Human, (B) Chimpanzee, (C) Gorilla, (D) Orangutan, (E) Crab-eating macaque, (F) Maximum-likelihood tree for all primary units of primates without a 50% cut-off value for bootstrap and collapsed branches showing the type of primary unit by order from 1 to 9. (G) Maximum-likelihood tree for all primary units of primates without a 50% cut-off value for bootstrap. Dots mark misplaced units.

Figure 25. Maximum-likelihood tree analyses for hornerin secondary units in primates. Boxed secondary units are either human or clustered on a different type.

Figure 26. Maximum-likelihood tree analyses for hornerin tertiary units in primates Boxed tertiary units are either human or clustered on a different type.

Filaggrin-2

Figure 27. Repeat sequences of filaggrin-2-FLG-like and filaggrin-2-HRNR-like in human, chimpanzee, macaque, baboon and marmoset. The nucleotide length of each repeat is shown beneath each repeat. The number of repeats varies between species.

Figure 28. Phylogenetic tree analyses for filaggrin-2-FLG-like and filaggrin-2-HRNR-like (A) Maximum likelihood reconstruction of filaggrin-2-FLG-like repeats, (B) Maximum likelihood reconstruction of filaggrin-2-HRNR-like repeats, (C) Neighbor-joining tree reconstruction of filaggrin-2-FLG-like repeats, (D) Neighbor-joining tree reconstruction of filaggrin-2-HRNR-like repeats, (E) “Reconciled” trees using filaggrin-2-FLG-like repeats, (F) “Reconciled” trees using filaggrin-2-HRNR-like repeats, (G) Maximum likelihood reconstruction of filaggrin-2-FLG-like repeats adding mouse repeats.

Repetin

Figure 29. Repeat sequences of repetin in human, chimpanzee, gorilla, orangutan, and crab-eating macaque. The nucleotide length of each repeat is shown beneath each repeat. The number of repeats varies between species.

Figure 30. Phylogenetic tree analyses for repetin (A) Maximum likelihood reconstruction of repetin repeats, (B) Neighbor-joining tree reconstruction of repetin repeats, (C) "Reconciled" trees using repetin repeats, (D) Maximum likelihood reconstruction of repetin repeats adding mouse repeats.

Cornulin

Figure 31. Repeat sequences of cornulin in human, chimpanzee, gorilla, orangutan, and crab-eating macaque. The nucleotide length of each repeat is shown beneath each repeat. The number of repeats varies between species.

Figure 32. Phylogenetic tree analyses for cornulin (A) Maximum likelihood reconstruction of cornulin repeats, (B) Neighbor-joining tree reconstruction of repetin repeats, (C) "Reconciled" trees using cornulin repeats, (D) Maximum likelihood reconstruction of cornulin repeats adding mouse repeats.

Figure 33. Protein alignment of different descriptions of cornulin repeats. Little et al description is not alignment with the previously described cornulin repeated sequence by Xu et al.

Figure 34. Repeat order of chimpanzee, and human for all SFTPs, as inferred by phylogenetic and "reconciled" tree.

Figure 35. Repeat sequences of SFTPs in human and mouse. The nucleotide length of each repeat is shown beneath each repeat. The number of repeats varies between species.

Figure 1.

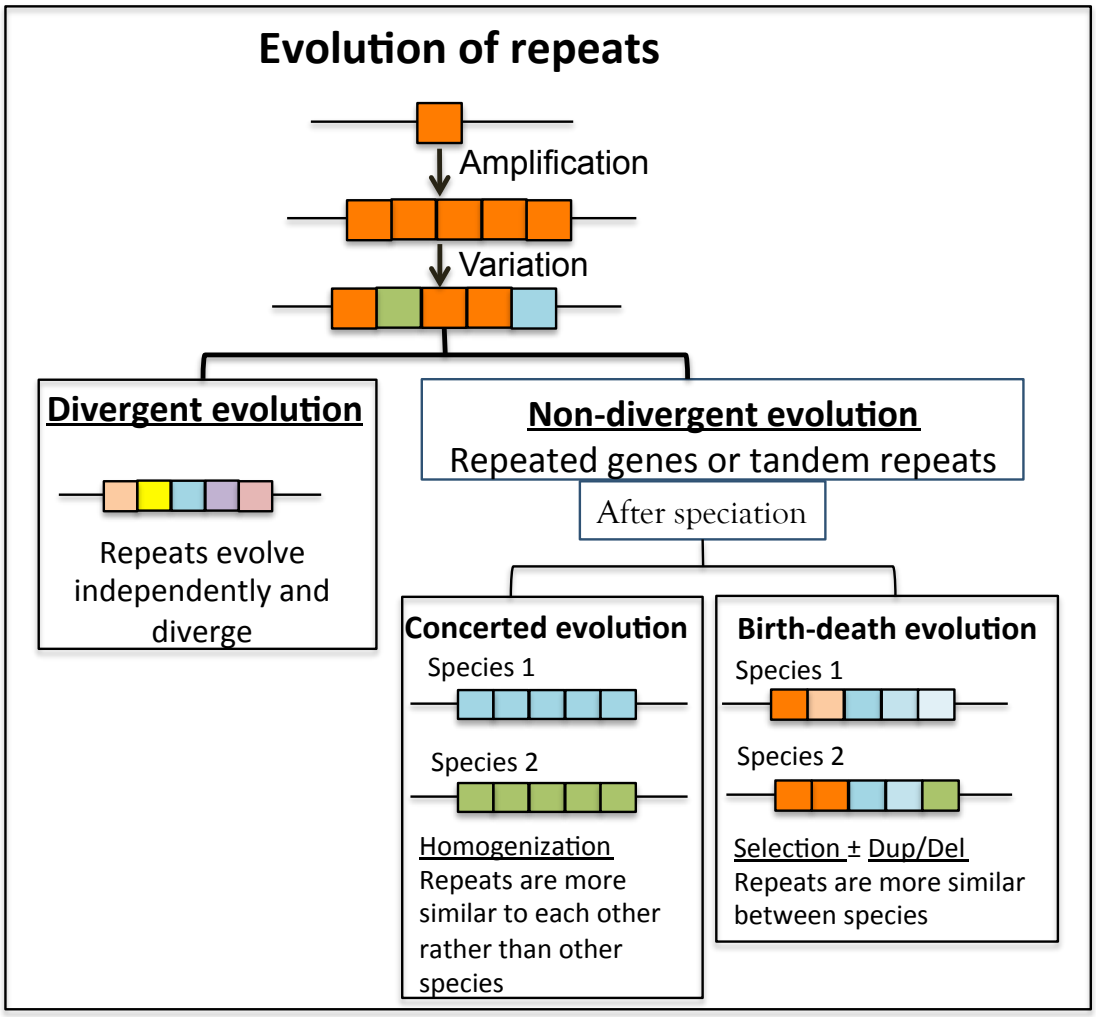

Evolution of repeats
Amplification
Variation
Divergent evolution
Repeats evolve independently and diverge
Non-divergent evolution
Repeated genes or tandem repeats
After speciation
Concerted evolution
Species 1
Species 2
Homogenization
Repeats are more similar to each other rather than other species
Birth-death evolution
Species 1
Species 2
Selection ± Dup/Del
Repeats are more similar between species

Figure 2.

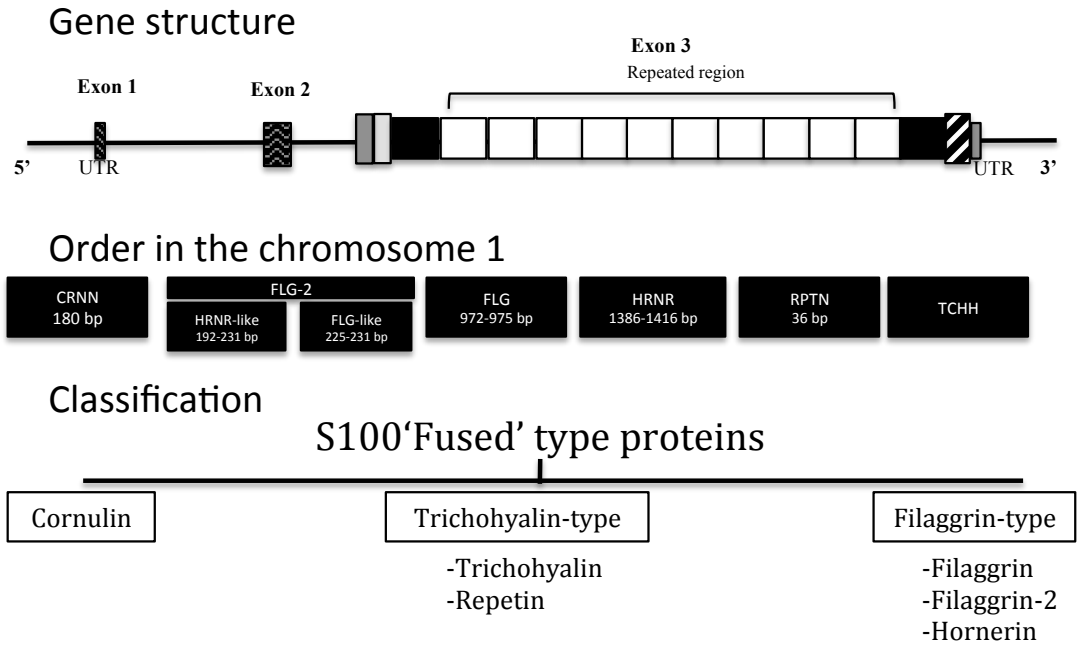
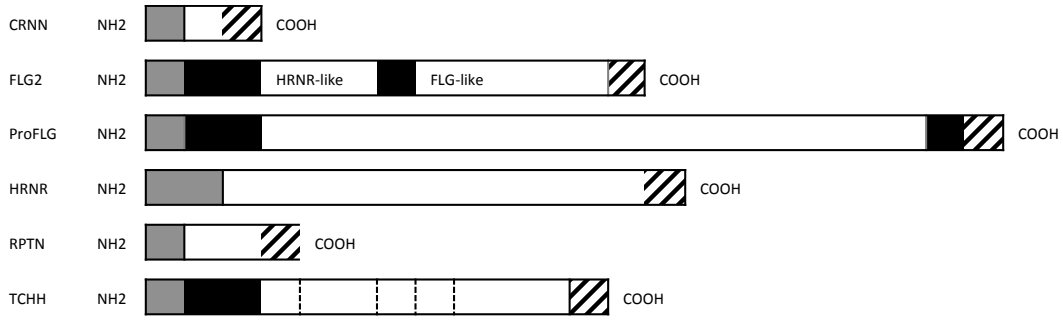

Gene structure
Exon 1
Exon 2
Exon 3
Repeated region
5'
UTR
UTR
3'
Order in the chromosome 1
CRNN
180 bp
FLG-2
HRNR-like
192-231 bp
FLG-like
225-231 bp
FLG
972-975 bp
HRNR
1386-1416 bp
RPTN
36 bp
TCHH
Classification
S100'Fused' type proteins
Cornulin
Trichohyalin-type
-Trichohyalin
-Repetin
Filaggrin-type
-Filaggrin
-Filaggrin-2
-Hornerin

Figure 3.

A.


CRNN
NH2
COOH
FLG2
NH2
HRNR-like
FLG-like
COOH
ProFLG
NH2
COOH
HRNR
NH2
COOH
RPTN
NH2
COOH
TCHH
NH2
COOH

B.

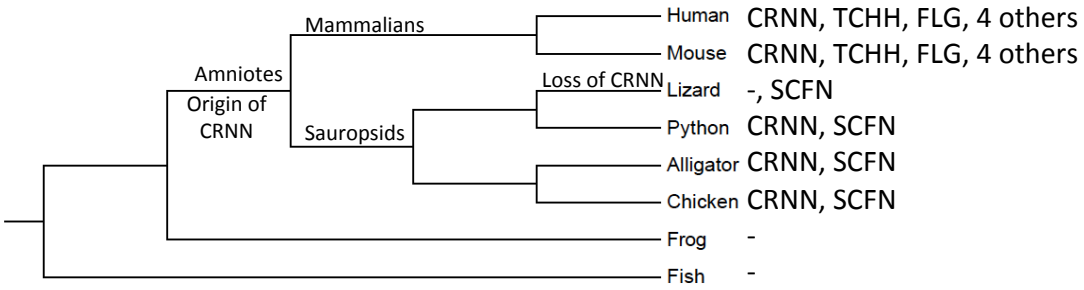

Mammalians
Amniotes
Origin of
CRNN
Loss of CRNN
Sauropsids
Human
CRNN, TCHH, FLG, 4 others
Mouse
CRNN, TCHH, FLG, 4 others
Lizard
-, SCFN
Python
CRNN, SCFN
Alligator
CRNN, SCFN
Chicken
CRNN, SCFN
Frog
-
Fish
-

Figure 4.

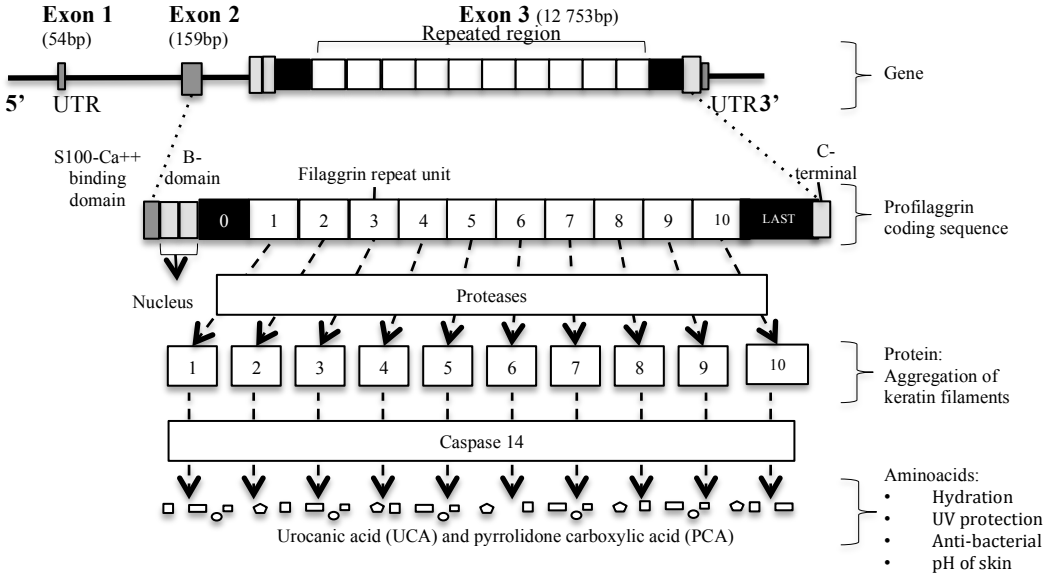
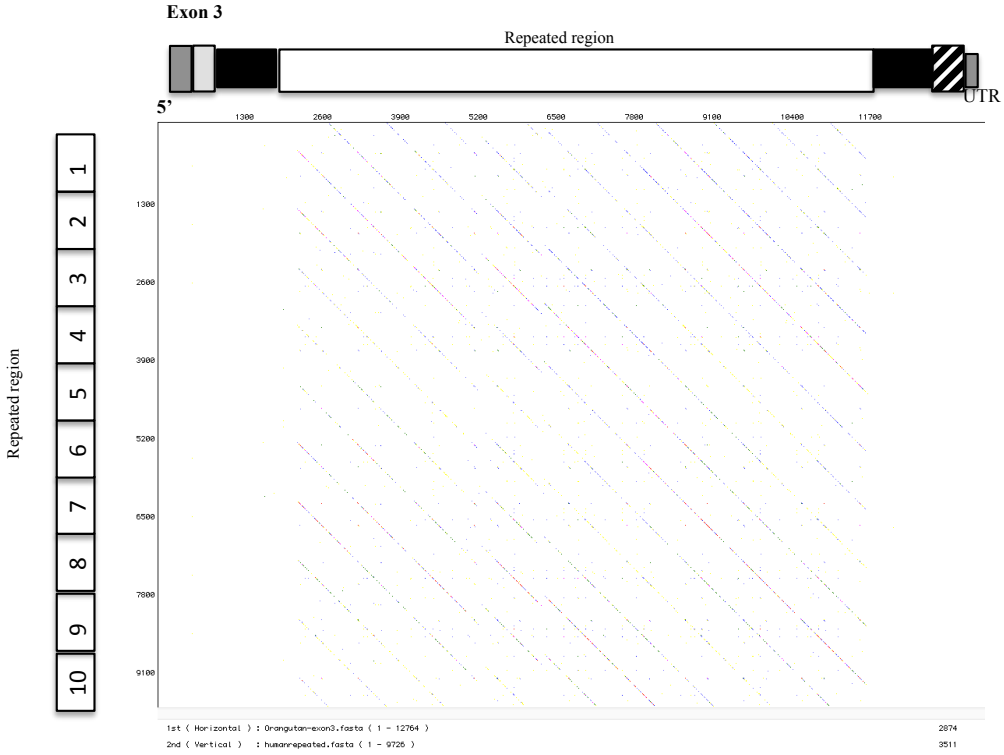
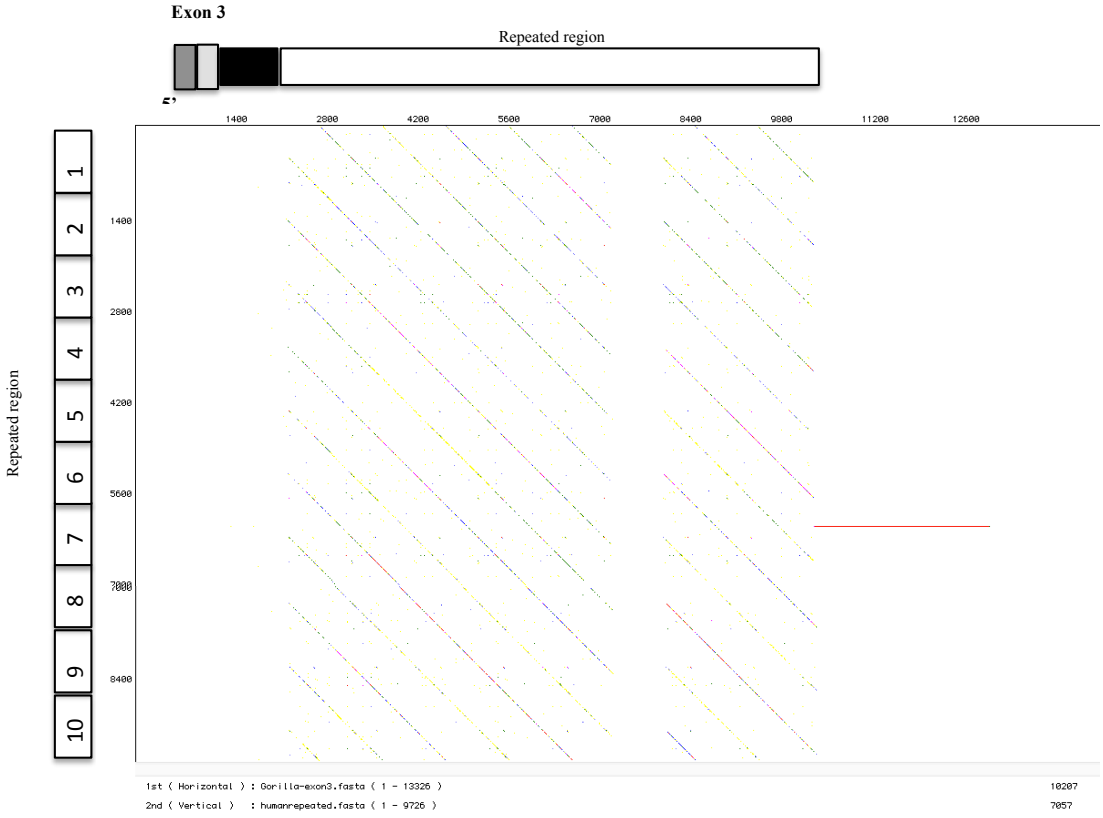

Exon 1
(54bp)
Exon 2
(159bp)
Exon 3 (12 753bp)
Repeated region
5' UTR
UTR 3'
Gene
S100-Ca++ binding domain
B-domain
Filaggrin repeat unit
C-terminal
0
1
2
3
4
5
6
7
8
9
10
LAST
Profilaggrin coding sequence
Nucleus
Proteases
1
2
3
4
5
6
7
8
9
10
Protein: Aggregation of keratin filaments
Caspase 14
Aminoacids:
• Hydration
• UV protection
• Anti-bacterial
• pH of skin
Urocanic acid (UCA) and pyrrolidone carboxylic acid (PCA)

Figure 5.

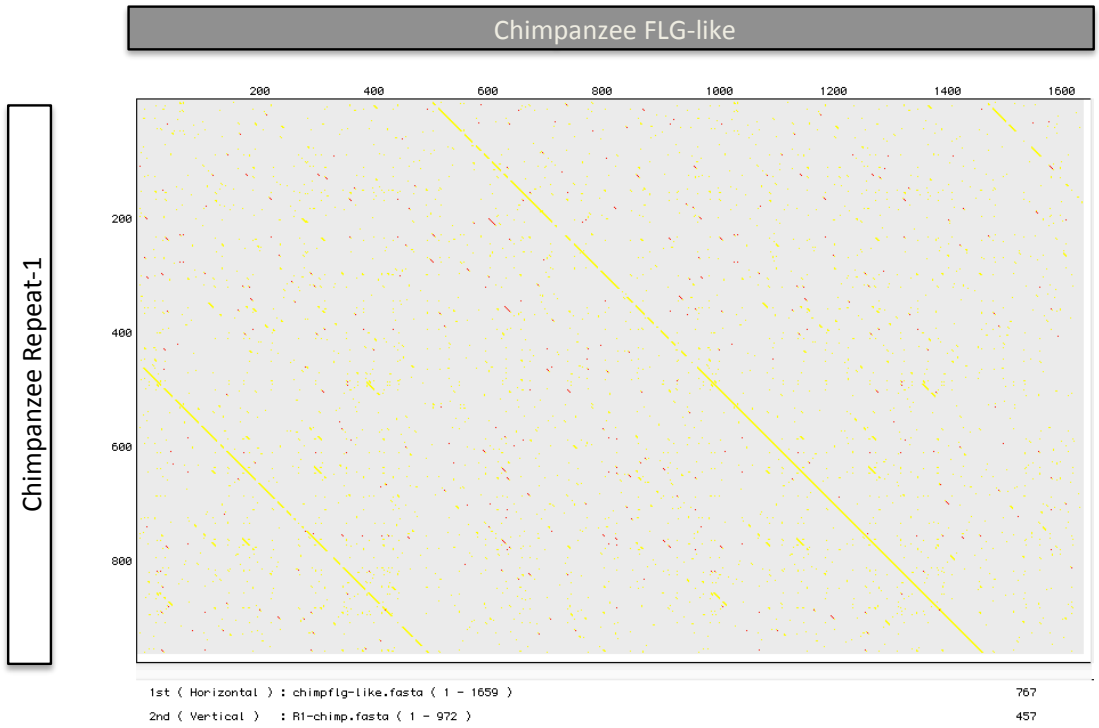
A.


Exon 3
Repeated region
5'
UTR
1300
2600
3900
5200
6500
7800
9100
10400
11700
Repeated region
1
2
3
4
5
6
7
8
9
10
1st (Horizontal) : Orangutan-exon3.fasta (1 - 12764)
2874
2nd (Vertical) : humanrepeated.fasta (1 - 9726)
3511

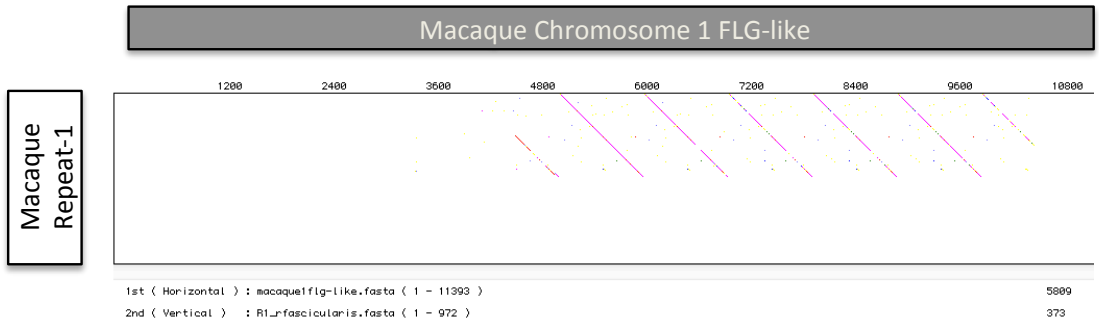
B


Exon 3
Repeated region
5'
1400
2800
4200
5600
7000
8400
9800
11200
12600
Repeated region
1
2
3
4
5
6
7
8
9
10
1st (Horizontal) : Gorilla-exon3.fasta (1 - 13326)
10207
2nd (Vertical) : humanrepeated.fasta (1 - 9726)
7057

C


Chimpanzee FLG-like
Chimpanzee Repeat-1
200
400
600
800
1000
1200
1400
1600
200
400
600
800
1st (Horizontal) : chimpflg-like.fasta (1 - 1659)
767
2nd (Vertical) : R1-chimp.fasta (1 - 972)
457

D


Macaque Chromosome 1 FLG-like
Macaque Repeat-1
1200
2400
3600
4800
6000
7200
8400
9600
10800
1st (Horizontal) : macaquelflg-like.fasta (1 - 11393)
5809
2nd (Vertical) : R1_rfascicularis.fasta (1 - 972)
373

E

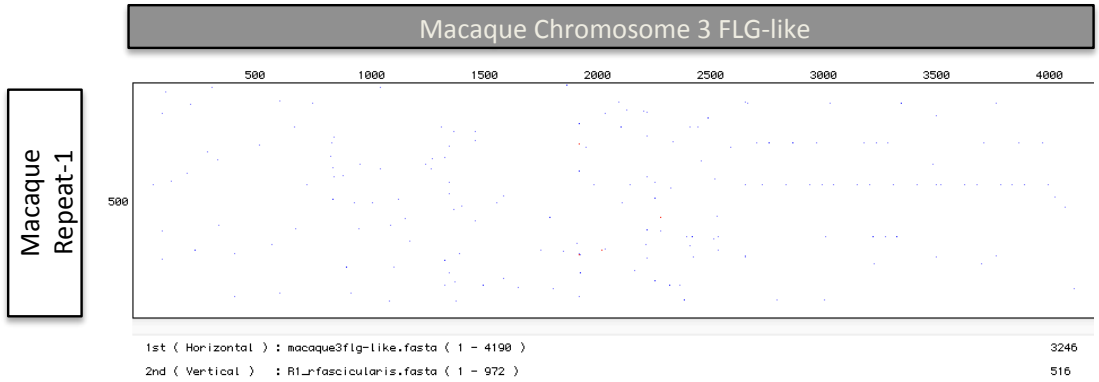

Macaque Chromosome 3 FLG-like
500
1000
1500
2000
2500
3000
3500
4000
Macaque Repeat-1
500
1st (Horizontal) : macaque3flg-like.fasta (1 - 4190)
3246
2nd (Vertical) : R1_rfascicularis.fasta (1 - 972)
516

Figure 6.

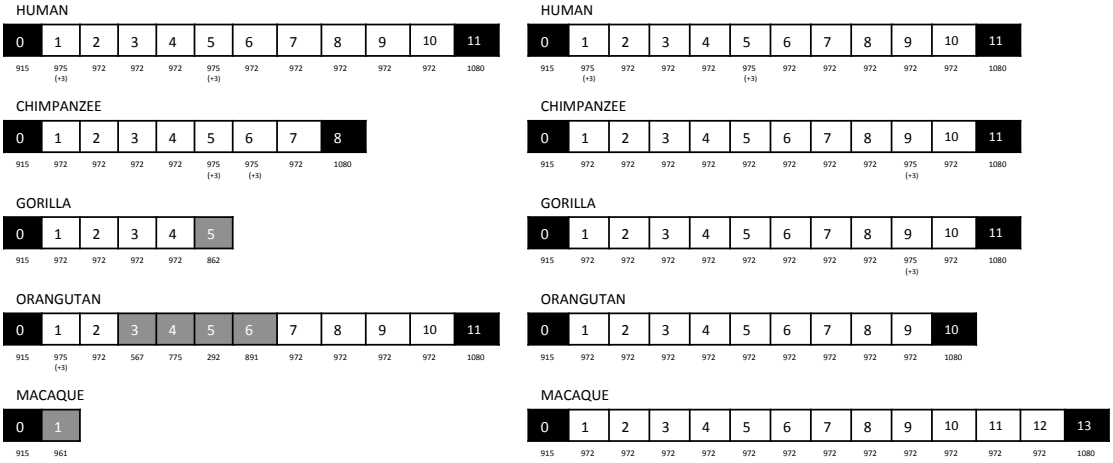

HUMAN
0 1 2 3 4 5 6 7 8 9 10 11
915 975 (+3) 972 972 972 975 (+3) 972 972 972 972 972 1080
CHIMPANZEE
0 1 2 3 4 5 6 7 8
915 972 972 972 972 975 (+3) 975 (+3) 972 1080
GORILLA
0 1 2 3 4 5
915 972 972 972 972 862
ORANGUTAN
0 1 2 3 4 5 6 7 8 9 10 11
915 975 (+3) 972 567 775 292 891 972 972 972 972 1080
MACAQUE
0 1
915 961
HUMAN
0 1 2 3 4 5 6 7 8 9 10 11
915 975 (+3) 972 972 972 975 (+3) 972 972 972 972 972 1080
CHIMPANZEE
0 1 2 3 4 5 6 7 8 9 10 11
915 972 972 972 972 972 972 972 972 975 (+3) 972 1080
GORILLA
0 1 2 3 4 5 6 7 8 9 10 11
915 972 972 972 972 972 972 972 972 975 (+3) 972 1080
ORANGUTAN
0 1 2 3 4 5 6 7 8 9 10
915 972 972 972 972 972 972 972 972 972 1080
MACAQUE
0 1 2 3 4 5 6 7 8 9 10 11 12 13
915 972 972 972 972 972 972 972 972 972 972 972 972 1080

Figure 7.

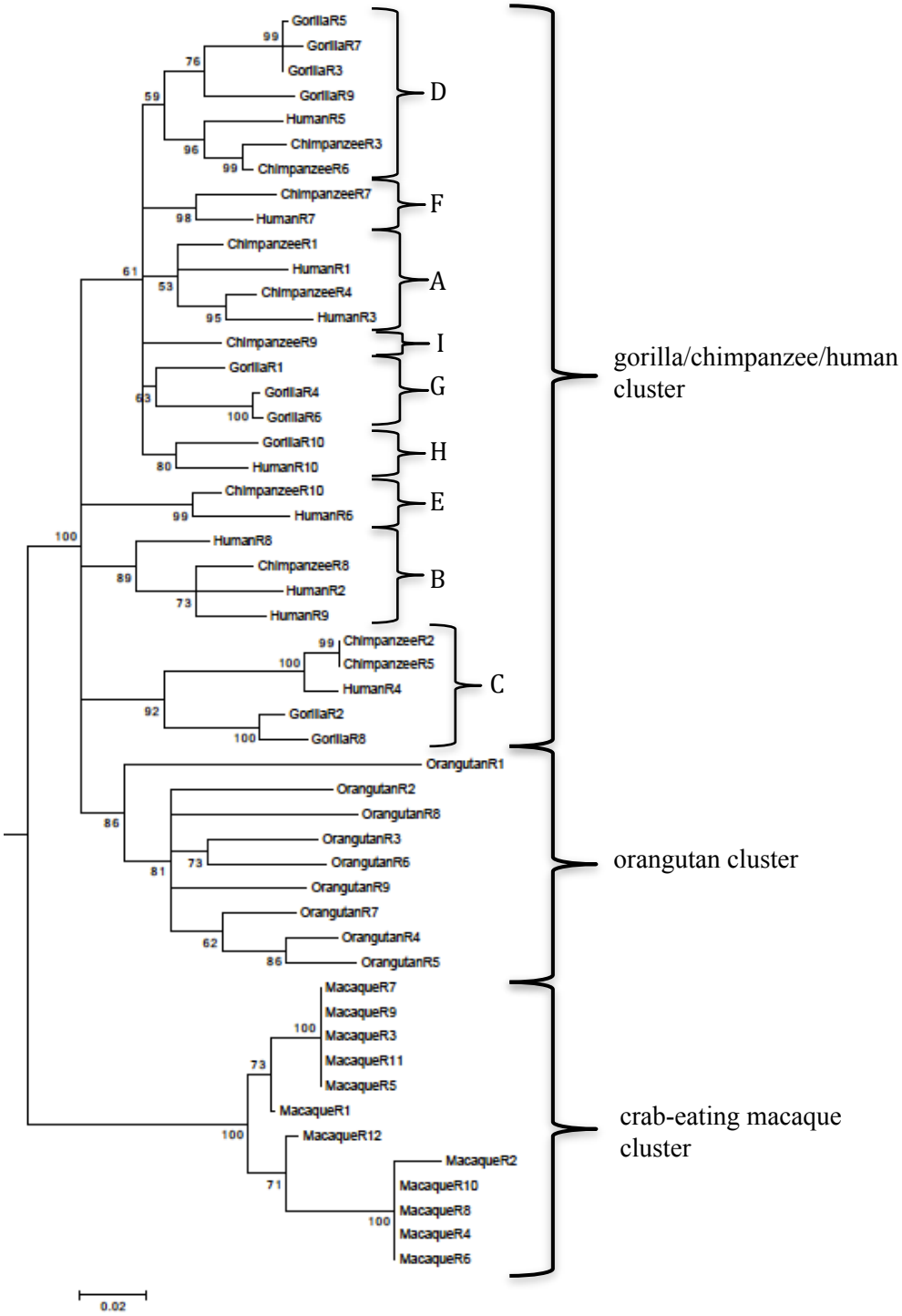

GorillaR5
GorillaR7
GorillaR3
GorillaR9
HumanR5
ChimpanzeeR3
ChimpanzeeR6
ChimpanzeeR7
HumanR7
ChimpanzeeR1
HumanR1
ChimpanzeeR4
HumanR3
ChimpanzeeR9
GorillaR1
GorillaR4
GorillaR6
GorillaR10
HumanR10
ChimpanzeeR10
HumanR6
HumanR8
ChimpanzeeR8
HumanR2
HumanR9
ChimpanzeeR2
ChimpanzeeR5
HumanR4
GorillaR2
GorillaR8
OrangutanR1
OrangutanR2
OrangutanR8
OrangutanR3
OrangutanR6
OrangutanR9
OrangutanR7
OrangutanR4
OrangutanR5
MacaqueR7
MacaqueR9
MacaqueR3
MacaqueR11
MacaqueR5
MacaqueR1
MacaqueR12
MacaqueR2
MacaqueR10
MacaqueR8
MacaqueR4
MacaqueR6
0.02
D
F
A
I
G
H
E
B
C
gorilla/chimpanzee/human cluster
orangutan cluster
crab-eating macaque cluster

Figure 8.

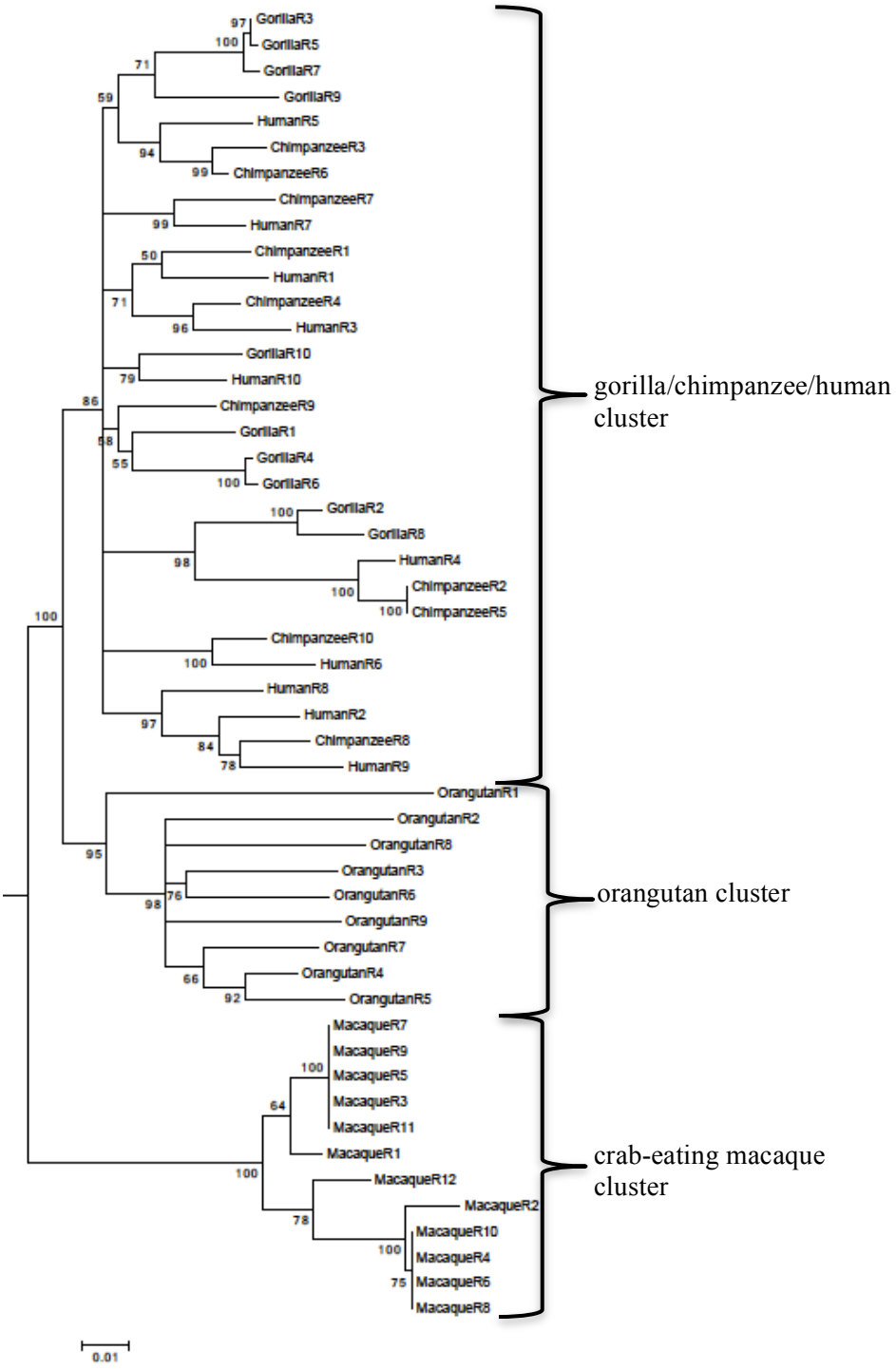
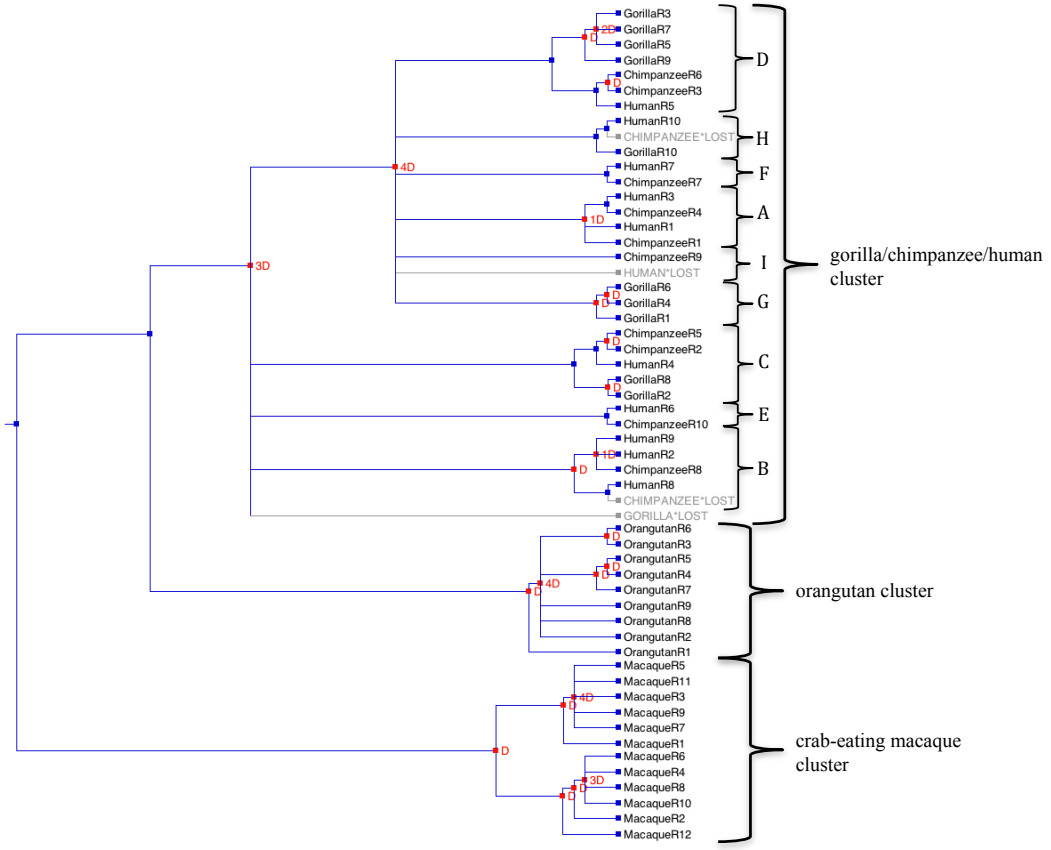

GorillaR3
GorillaR5
GorillaR7
GorillaR9
HumanR5
ChimpanzeeR3
ChimpanzeeR6
ChimpanzeeR7
HumanR7
ChimpanzeeR1
HumanR1
ChimpanzeeR4
HumanR3
GorillaR10
HumanR10
ChimpanzeeR9
GorillaR1
GorillaR4
GorillaR6
GorillaR2
GorillaR8
HumanR4
ChimpanzeeR2
ChimpanzeeR5
ChimpanzeeR10
HumanR6
HumanR8
HumanR2
ChimpanzeeR8
HumanR9
OrangutanR1
OrangutanR2
OrangutanR8
OrangutanR3
OrangutanR6
OrangutanR9
OrangutanR7
OrangutanR4
OrangutanR5
MacaqueR7
MacaqueR9
MacaqueR5
MacaqueR3
MacaqueR11
MacaqueR1
MacaqueR12
MacaqueR2
MacaqueR10
MacaqueR4
MacaqueR6
MacaqueR8
0.01
gorilla/chimpanzee/human cluster
orangutan cluster
crab-eating macaque cluster

Figure 9.

A.


GorillaR3
GorillaR7
GorillaR5
GorillaR9
ChimpanzeeR6
ChimpanzeeR3
HumanR5
HumanR10
CHIMPANZEE*LOST
GorillaR10
HumanR7
ChimpanzeeR7
HumanR3
ChimpanzeeR4
HumanR1
ChimpanzeeR1
ChimpanzeeR9
HUMAN*LOST
GorillaR6
GorillaR4
GorillaR1
ChimpanzeeR5
ChimpanzeeR2
HumanR4
GorillaR8
GorillaR2
HumanR6
ChimpanzeeR10
HumanR9
HumanR2
ChimpanzeeR8
HumanR8
CHIMPANZEE*LOST
GORILLA*LOST
OrangutanR6
OrangutanR3
OrangutanR5
OrangutanR4
OrangutanR7
OrangutanR9
OrangutanR8
OrangutanR2
OrangutanR1
MacaqueR5
MacaqueR11
MacaqueR3
MacaqueR9
MacaqueR7
MacaqueR1
MacaqueR6
MacaqueR4
MacaqueR8
MacaqueR10
MacaqueR2
MacaqueR12
D
H
F
A
I
G
C
E
B
gorilla/chimpanzee/human cluster
orangutan cluster
crab-eating macaque cluster

B.

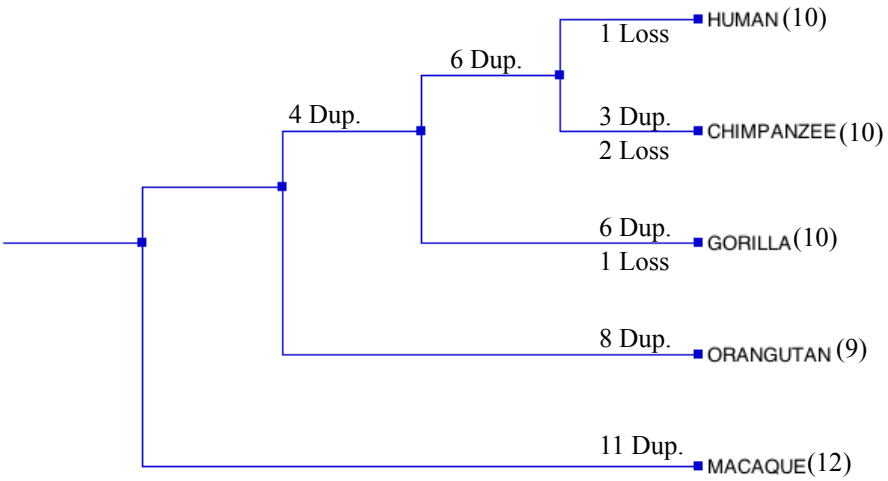
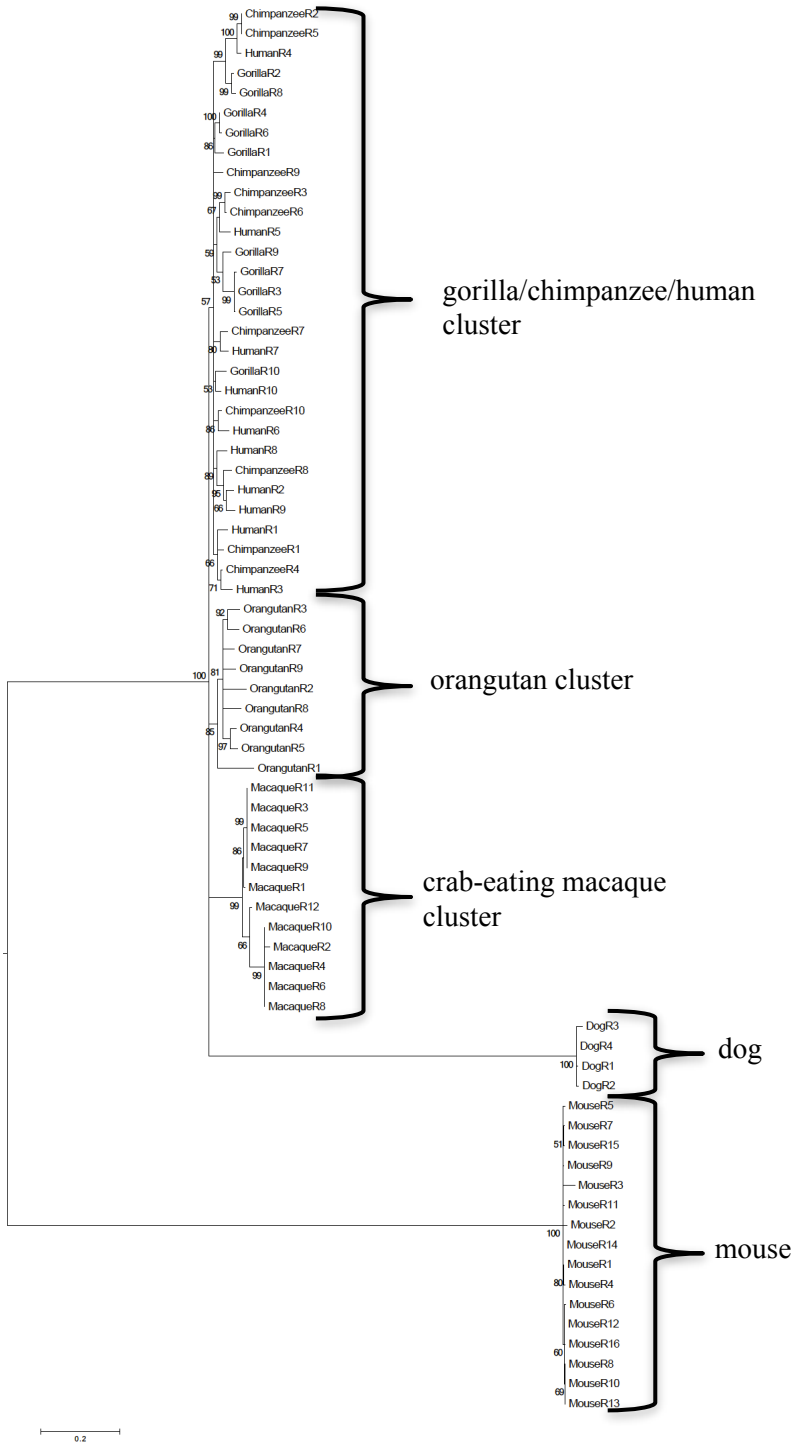
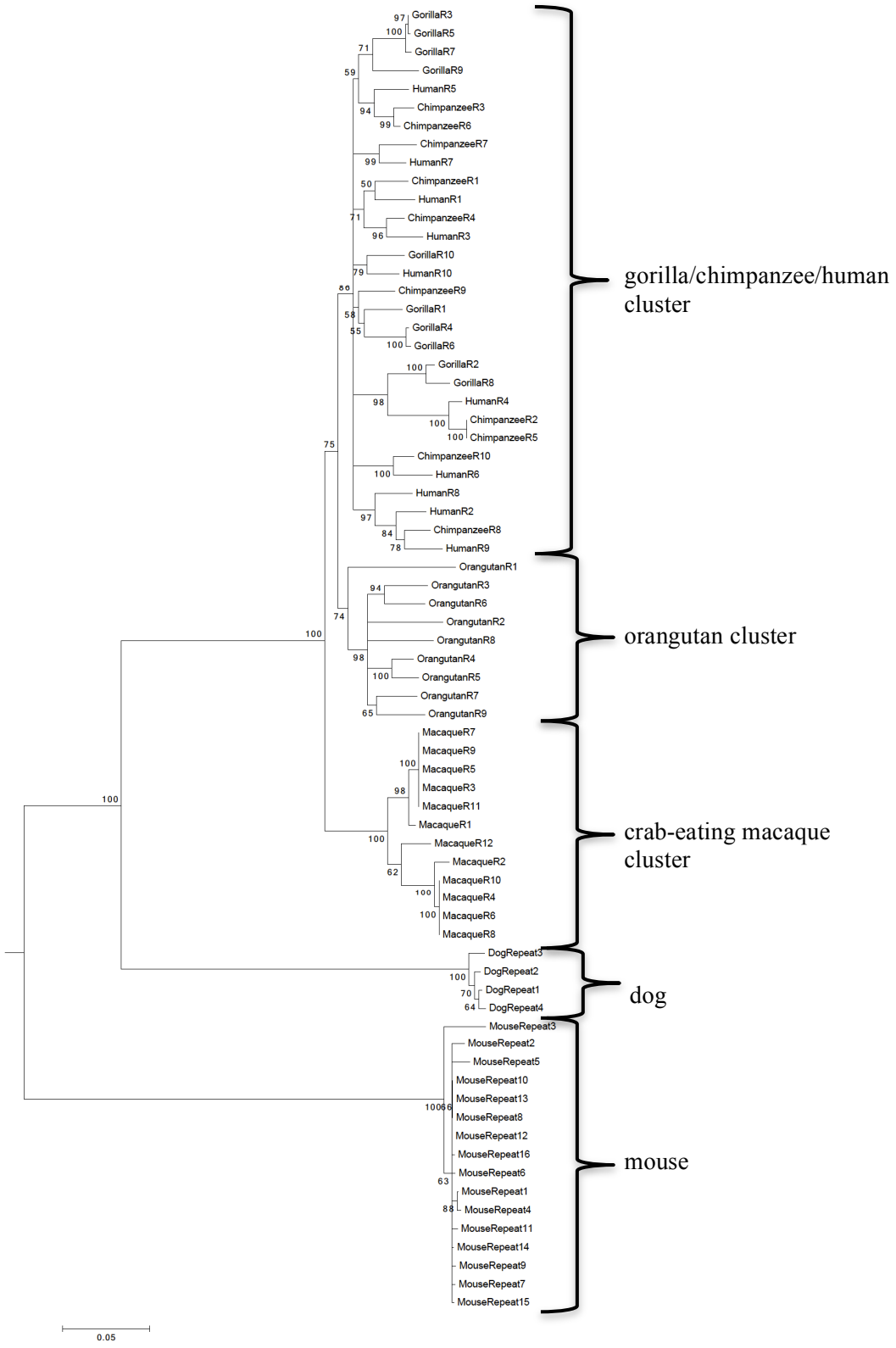

HUMAN (10)
1 Loss
6 Dup.
4 Dup.
3 Dup.
CHIMPANZEE (10)
2 Loss
6 Dup.
GORILLA (10)
1 Loss
8 Dup.
ORANGUTAN (9)
11 Dup.
MACAQUE (12)

Figure 10.

A.


ChimpanzeeR2
ChimpanzeeR5
HumanR4
GorillaR2
GorillaR8
GorillaR4
GorillaR6
GorillaR1
ChimpanzeeR9
ChimpanzeeR3
ChimpanzeeR6
HumanR5
GorillaR9
GorillaR7
GorillaR3
GorillaR5
ChimpanzeeR7
HumanR7
GorillaR10
HumanR10
ChimpanzeeR10
HumanR6
HumanR8
ChimpanzeeR8
HumanR2
HumanR9
HumanR1
ChimpanzeeR1
ChimpanzeeR4
HumanR3
gorilla/chimpanzee/human cluster
OrangutanR3
OrangutanR6
OrangutanR7
OrangutanR9
OrangutanR2
OrangutanR8
OrangutanR4
OrangutanR5
OrangutanR1
orangutan cluster
MacaqueR11
MacaqueR3
MacaqueR5
MacaqueR7
MacaqueR9
MacaqueR1
MacaqueR12
MacaqueR10
MacaqueR2
MacaqueR4
MacaqueR6
MacaqueR8
crab-eating macaque cluster
DogR3
DogR4
DogR1
DogR2
dog
MouseR5
MouseR7
MouseR15
MouseR9
MouseR3
MouseR11
MouseR2
MouseR14
MouseR1
MouseR4
MouseR6
MouseR12
MouseR16
MouseR8
MouseR10
MouseR13
mouse
0.2

B.


GorillaR3
GorillaR5
GorillaR7
GorillaR9
HumanR5
ChimpanzeeR3
ChimpanzeeR6
ChimpanzeeR7
HumanR7
ChimpanzeeR1
HumanR1
ChimpanzeeR4
HumanR3
GorillaR10
HumanR10
ChimpanzeeR9
GorillaR1
GorillaR4
GorillaR6
GorillaR2
GorillaR8
HumanR4
ChimpanzeeR2
ChimpanzeeR5
ChimpanzeeR10
HumanR6
HumanR8
HumanR2
ChimpanzeeR8
HumanR9
OrangutanR1
OrangutanR3
OrangutanR6
OrangutanR2
OrangutanR8
OrangutanR4
OrangutanR5
OrangutanR7
OrangutanR9
MacaqueR7
MacaqueR9
MacaqueR5
MacaqueR3
MacaqueR11
MacaqueR1
MacaqueR12
MacaqueR2
MacaqueR10
MacaqueR4
MacaqueR6
MacaqueR8
DogRepeat3
DogRepeat2
DogRepeat1
DogRepeat4
MouseRepeat3
MouseRepeat2
MouseRepeat5
MouseRepeat10
MouseRepeat13
MouseRepeat8
MouseRepeat12
MouseRepeat16
MouseRepeat6
MouseRepeat1
MouseRepeat4
MouseRepeat11
MouseRepeat14
MouseRepeat9
MouseRepeat7
MouseRepeat15
0.05
gorilla/chimpanzee/human cluster
orangutan cluster
crab-eating macaque cluster
dog
mouse

C.

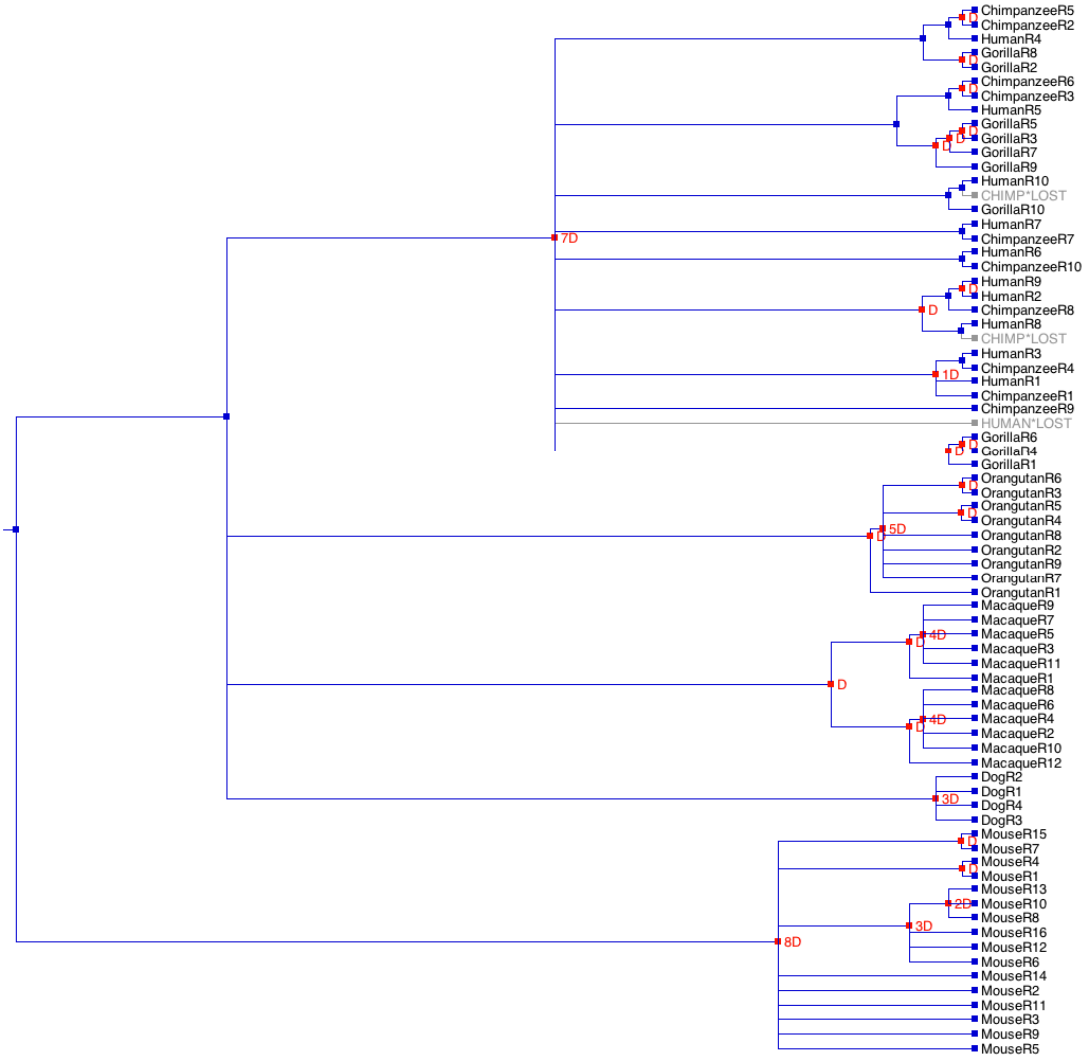

ChimpanzeeR5
ChimpanzeeR2
HumanR4
GorillaR8
GorillaR2
ChimpanzeeR6
ChimpanzeeR3
HumanR5
GorillaR5
GorillaR3
GorillaR7
GorillaR9
HumanR10
CHIMP*LOST
GorillaR10
HumanR7
ChimpanzeeR7
HumanR6
ChimpanzeeR10
HumanR9
HumanR2
ChimpanzeeR8
HumanR8
CHIMP*LOST
HumanR3
ChimpanzeeR4
HumanR1
ChimpanzeeR1
ChimpanzeeR9
HUMAN*LOST
GorillaR6
GorillaR4
GorillaR1
OrangutanR6
OrangutanR3
OrangutanR5
OrangutanR4
OrangutanR8
OrangutanR2
OrangutanR9
OrangutanR7
OrangutanR1
MacaqueR9
MacaqueR7
MacaqueR5
MacaqueR3
MacaqueR11
MacaqueR1
MacaqueR8
MacaqueR6
MacaqueR4
MacaqueR2
MacaqueR10
MacaqueR12
DogR2
DogR1
DogR4
DogR3
MouseR15
MouseR7
MouseR4
MouseR1
MouseR13
MouseR10
MouseR8
MouseR16
MouseR12
MouseR6
MouseR14
MouseR2
MouseR11
MouseR3
MouseR9
MouseR5
7D
5D
4D
4D
3D
2D
3D
8D
1D
D

Figure 11.

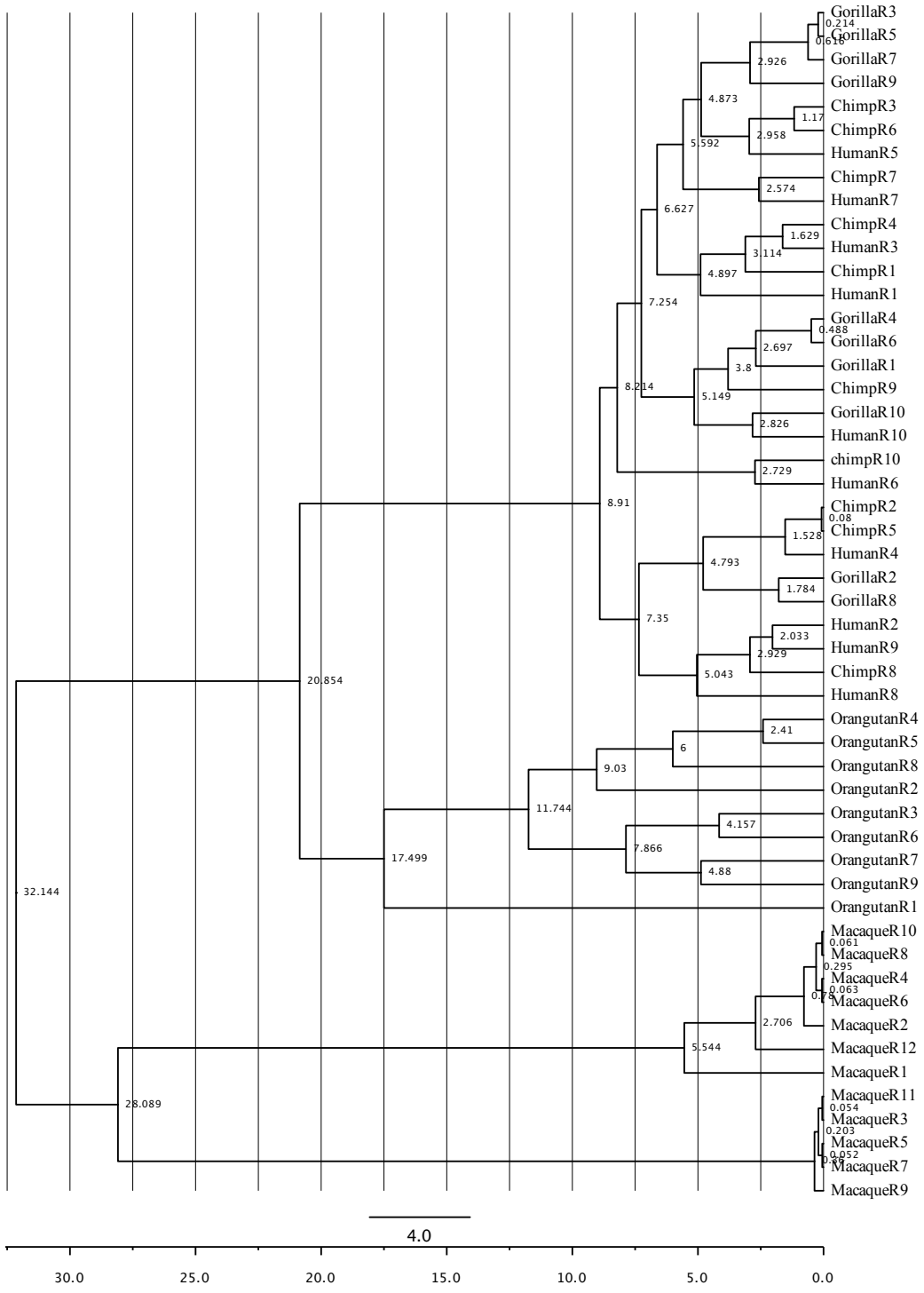

GorillaR3
GorillaR5
GorillaR7
GorillaR9
ChimpR3
ChimpR6
HumanR5
ChimpR7
HumanR7
ChimpR4
HumanR3
ChimpR1
HumanR1
GorillaR4
GorillaR6
GorillaR1
ChimpR9
GorillaR10
HumanR10
chimpR10
HumanR6
ChimpR2
ChimpR5
HumanR4
GorillaR2
GorillaR8
HumanR2
HumanR9
ChimpR8
HumanR8
OrangutanR4
OrangutanR5
OrangutanR8
OrangutanR2
OrangutanR3
OrangutanR6
OrangutanR7
OrangutanR9
OrangutanR1
MacaqueR10
MacaqueR8
MacaqueR4
MacaqueR6
MacaqueR2
MacaqueR12
MacaqueR1
MacaqueR11
MacaqueR3
MacaqueR5
MacaqueR7
MacaqueR9
0.214
0.616
2.926
4.873
1.17
2.958
5.592
2.574
6.627
1.629
3.114
4.897
7.254
0.488
2.697
3.8
5.149
8.214
2.826
2.729
8.91
0.08
1.528
4.793
1.784
7.35
2.033
2.929
5.043
20.854
2.41
6
9.03
11.744
4.157
7.866
4.88
17.499
32.144
0.061
0.295
0.063
2.706
5.544
0.054
0.203
0.052
28.089
4.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0

Figure 12.

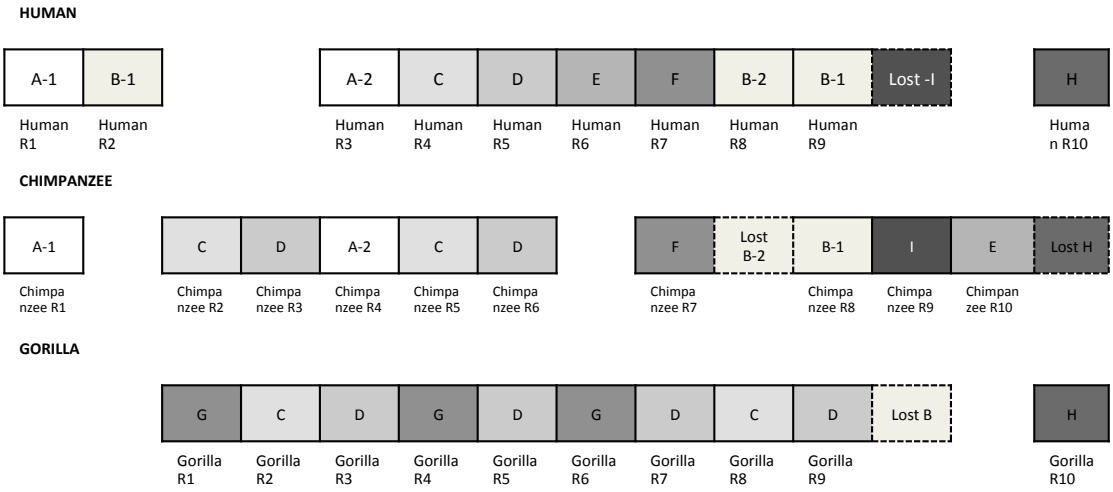
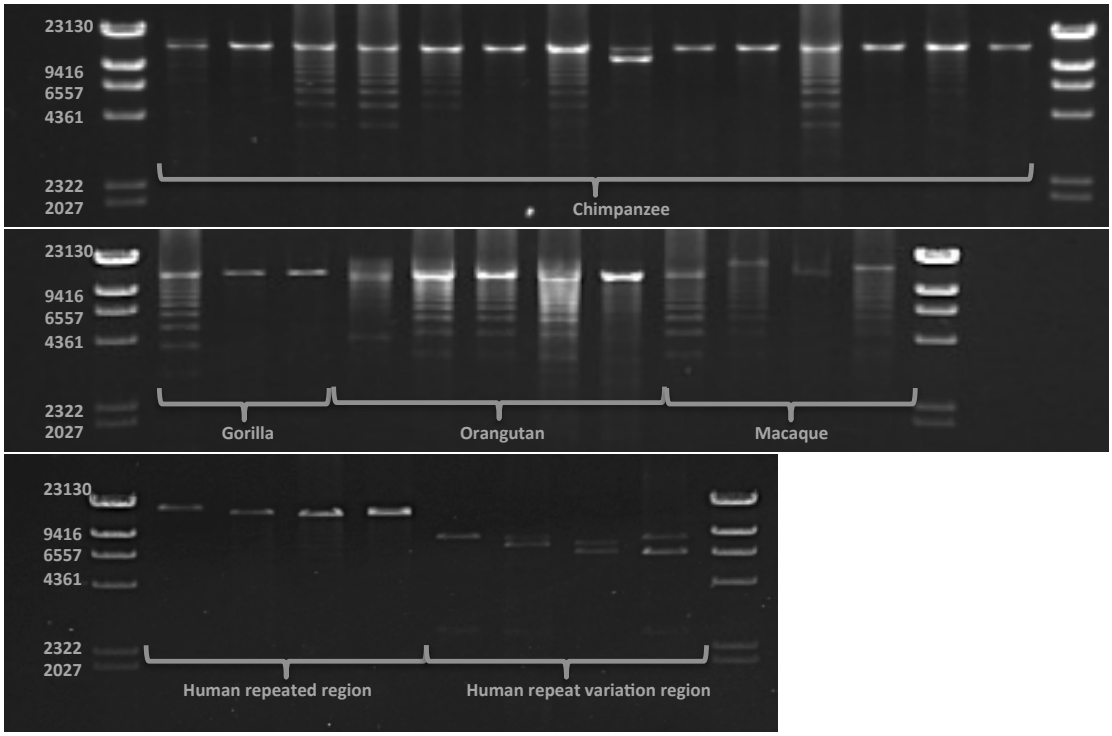

HUMAN
A-1
B-1
A-2
C
D
E
F
B-2
B-1
Lost -I
H
Human R1
Human R2
Human R3
Human R4
Human R5
Human R6
Human R7
Human R8
Human R9
Huma n R10
CHIMPANZEE
A-1
C
D
A-2
C
D
F
Lost B-2
B-1
I
E
Lost H
Chimpa nzee R1
Chimpa nzee R2
Chimpa nzee R3
Chimpa nzee R4
Chimpa nzee R5
Chimpa nzee R6
Chimpa nzee R7
Chimpa nzee R8
Chimpa nzee R9
Chimpan zee R10
GORILLA
G
C
D
G
D
G
D
C
D
Lost B
H
Gorilla R1
Gorilla R2
Gorilla R3
Gorilla R4
Gorilla R5
Gorilla R6
Gorilla R7
Gorilla R8
Gorilla R9
Gorilla R10

Figure 13.

A.


23130
9416
6557
4361
2322
2027
Chimpanzee
23130
9416
6557
4361
2322
2027
Gorilla
Orangutan
Macaque
23130
9416
6557
4361
2322
2027
Human repeated region
Human repeat variation region

B.

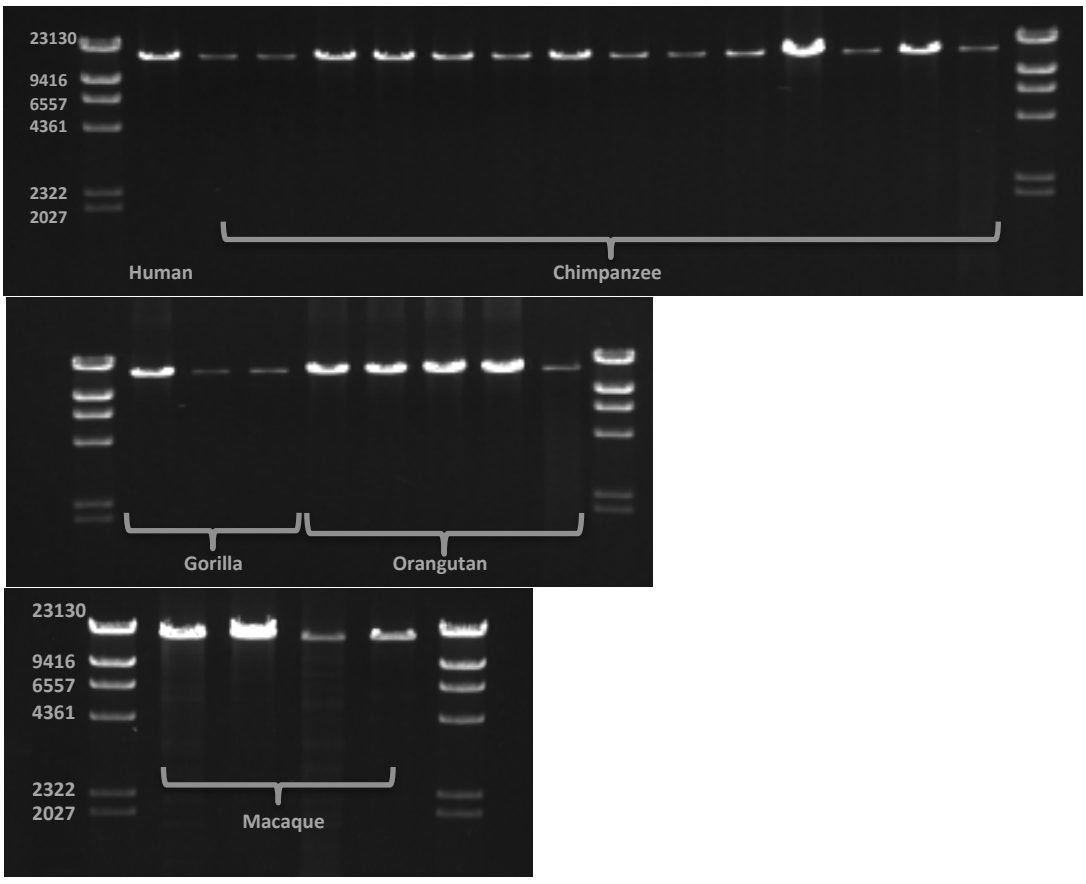

23130
9416
6557
4361
2322
2027
Human
Chimpanzee
Gorilla
Orangutan
23130
9416
6557
4361
2322
2027
Macaque

Figure 14.

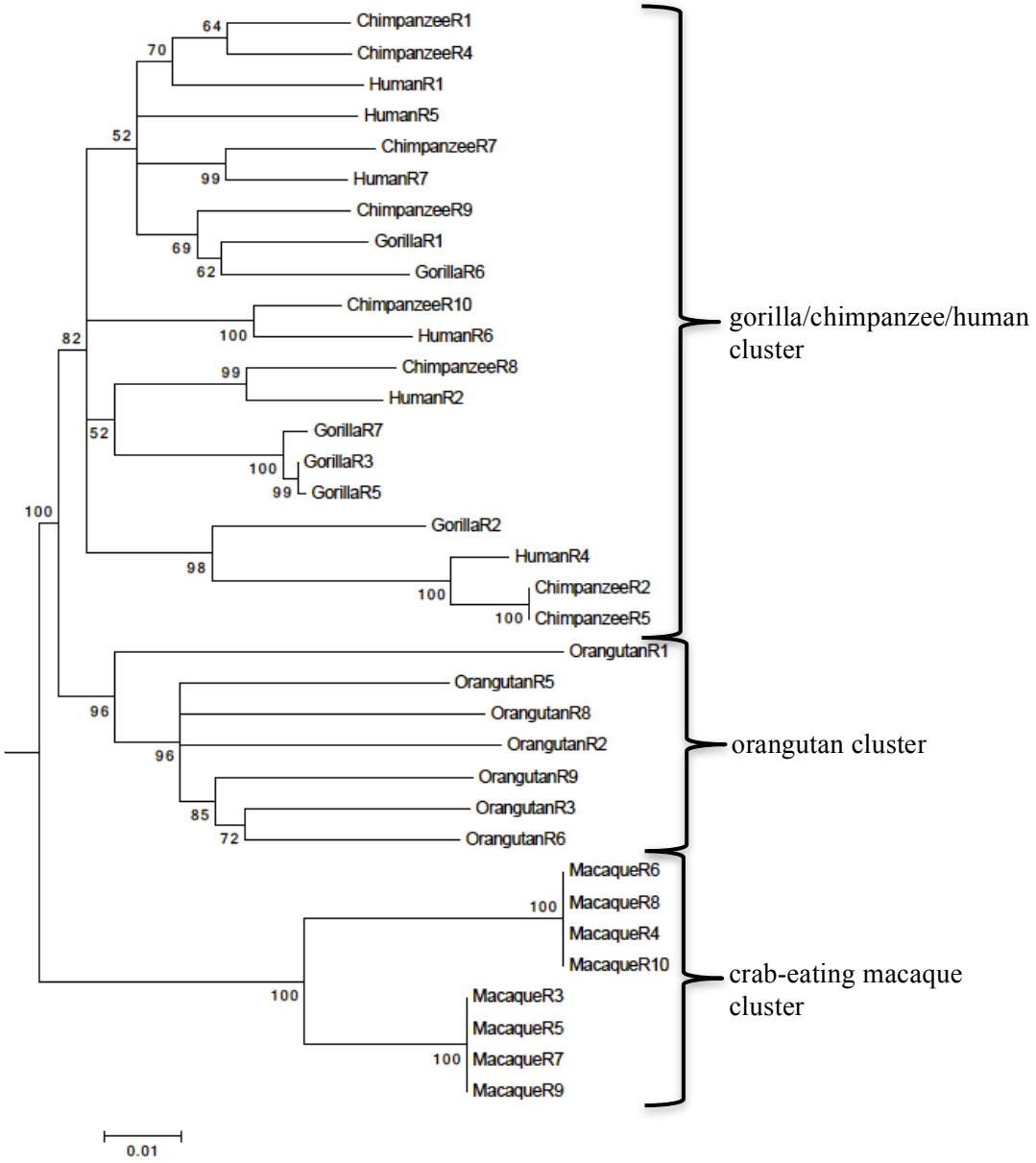
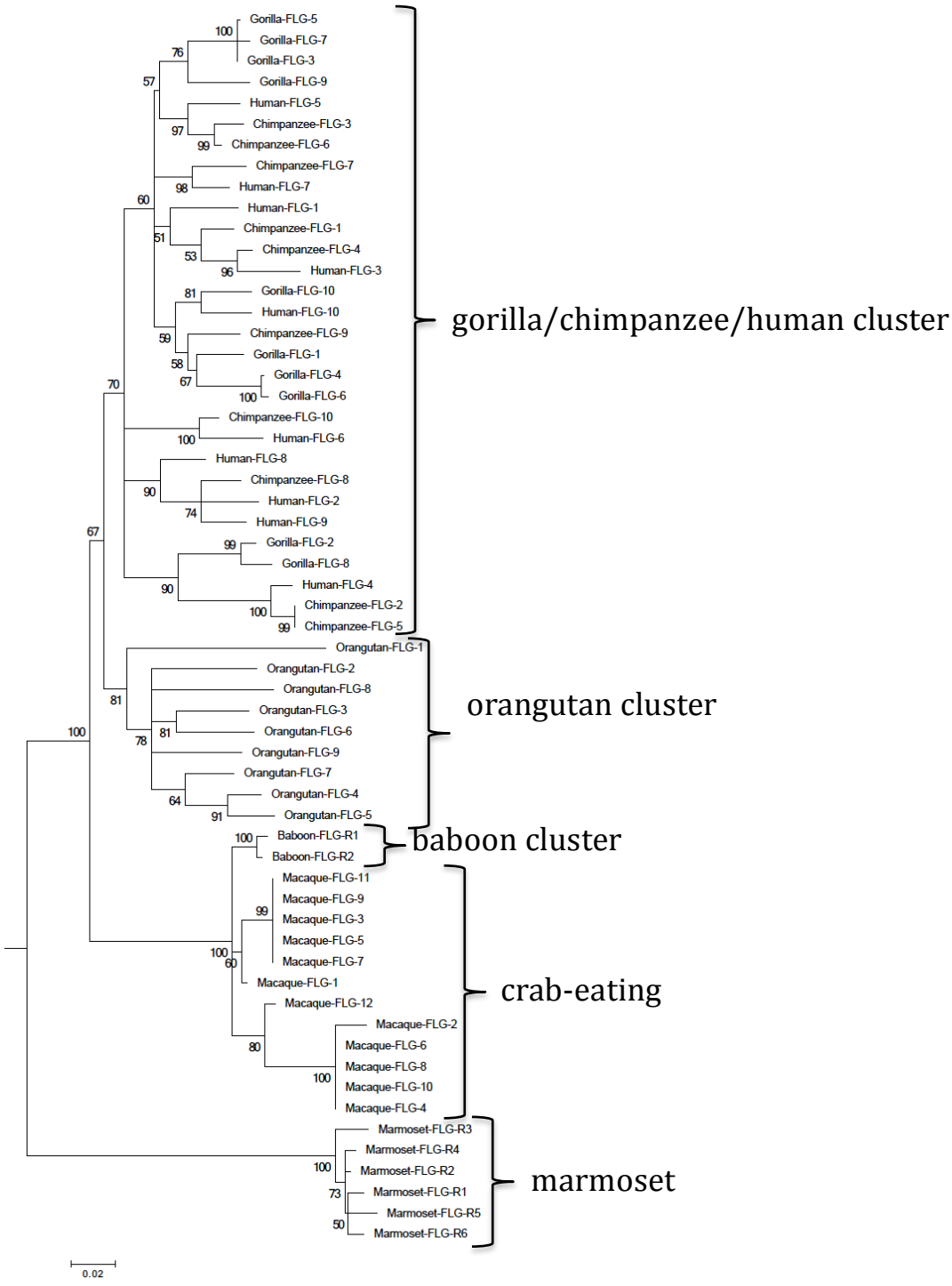
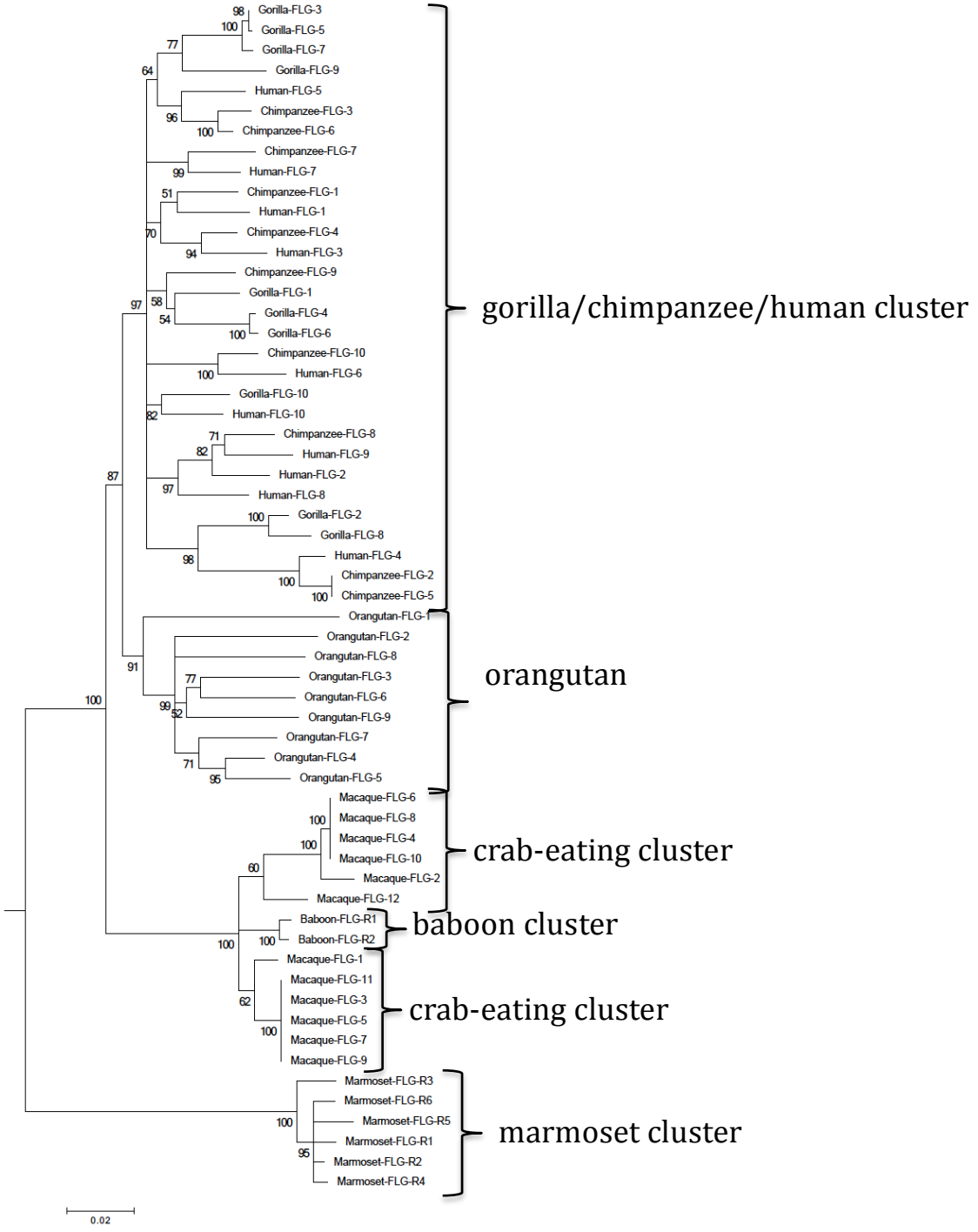

ChimpanzeeR1
ChimpanzeeR4
HumanR1
HumanR5
ChimpanzeeR7
HumanR7
ChimpanzeeR9
GorillaR1
GorillaR6
ChimpanzeeR10
HumanR6
ChimpanzeeR8
HumanR2
GorillaR7
GorillaR3
GorillaR5
GorillaR2
HumanR4
ChimpanzeeR2
ChimpanzeeR5
OrangutanR1
OrangutanR5
OrangutanR8
OrangutanR2
OrangutanR9
OrangutanR3
OrangutanR6
MacaqueR6
MacaqueR8
MacaqueR4
MacaqueR10
MacaqueR3
MacaqueR5
MacaqueR7
MacaqueR9
0.01
gorilla/chimpanzee/human cluster
orangutan cluster
crab-eating macaque cluster

Figure 15.

A.


Gorilla-FLG-5
Gorilla-FLG-7
Gorilla-FLG-3
Gorilla-FLG-9
Human-FLG-5
Chimpanzee-FLG-3
Chimpanzee-FLG-6
Chimpanzee-FLG-7
Human-FLG-7
Human-FLG-1
Chimpanzee-FLG-1
Chimpanzee-FLG-4
Human-FLG-3
Gorilla-FLG-10
Human-FLG-10
Chimpanzee-FLG-9
Gorilla-FLG-1
Gorilla-FLG-4
Gorilla-FLG-6
Chimpanzee-FLG-10
Human-FLG-6
Human-FLG-8
Chimpanzee-FLG-8
Human-FLG-2
Human-FLG-9
Gorilla-FLG-2
Gorilla-FLG-8
Human-FLG-4
Chimpanzee-FLG-2
Chimpanzee-FLG-5
Orangutan-FLG-1
Orangutan-FLG-2
Orangutan-FLG-8
Orangutan-FLG-3
Orangutan-FLG-6
Orangutan-FLG-9
Orangutan-FLG-7
Orangutan-FLG-4
Orangutan-FLG-5
Baboon-FLG-R1
Baboon-FLG-R2
Macaque-FLG-11
Macaque-FLG-9
Macaque-FLG-3
Macaque-FLG-5
Macaque-FLG-7
Macaque-FLG-1
Macaque-FLG-12
Macaque-FLG-2
Macaque-FLG-6
Macaque-FLG-8
Macaque-FLG-10
Macaque-FLG-4
Marmoset-FLG-R3
Marmoset-FLG-R4
Marmoset-FLG-R2
Marmoset-FLG-R1
Marmoset-FLG-R5
Marmoset-FLG-R6
0.02
gorilla/chimpanzee/human cluster
orangutan cluster
baboon cluster
crab-eating
marmoset

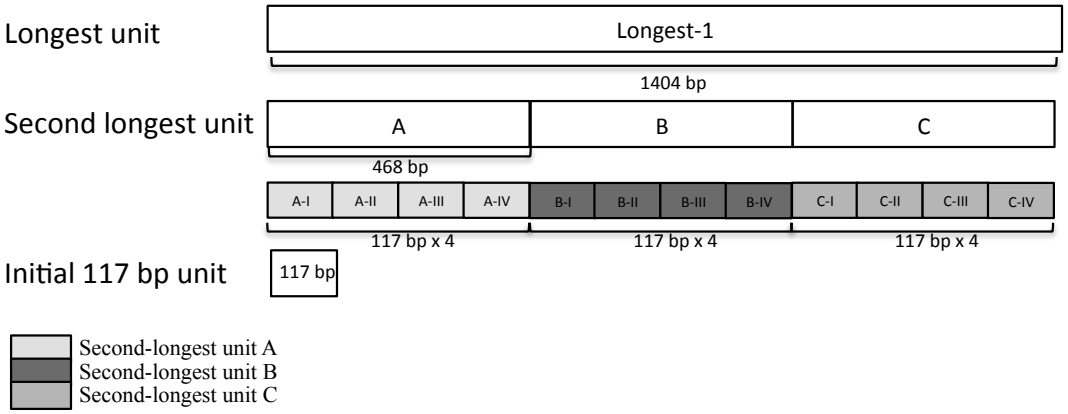
B.


Gorilla-FLG-3
Gorilla-FLG-5
Gorilla-FLG-7
Gorilla-FLG-9
Human-FLG-5
Chimpanzee-FLG-3
Chimpanzee-FLG-6
Chimpanzee-FLG-7
Human-FLG-7
Chimpanzee-FLG-1
Human-FLG-1
Chimpanzee-FLG-4
Human-FLG-3
Chimpanzee-FLG-9
Gorilla-FLG-1
Gorilla-FLG-4
Gorilla-FLG-6
Chimpanzee-FLG-10
Human-FLG-6
Gorilla-FLG-10
Human-FLG-10
Chimpanzee-FLG-8
Human-FLG-9
Human-FLG-2
Human-FLG-8
Gorilla-FLG-2
Gorilla-FLG-8
Human-FLG-4
Chimpanzee-FLG-2
Chimpanzee-FLG-5
Orangutan-FLG-1
Orangutan-FLG-2
Orangutan-FLG-8
Orangutan-FLG-3
Orangutan-FLG-6
Orangutan-FLG-9
Orangutan-FLG-7
Orangutan-FLG-4
Orangutan-FLG-5
Macaque-FLG-6
Macaque-FLG-8
Macaque-FLG-4
Macaque-FLG-10
Macaque-FLG-2
Macaque-FLG-12
Baboon-FLG-R1
Baboon-FLG-R2
Macaque-FLG-1
Macaque-FLG-11
Macaque-FLG-3
Macaque-FLG-5
Macaque-FLG-7
Macaque-FLG-9
Marmoset-FLG-R3
Marmoset-FLG-R6
Marmoset-FLG-R5
Marmoset-FLG-R1
Marmoset-FLG-R2
Marmoset-FLG-R4
0.02
gorilla/chimpanzee/human cluster
orangutan
crab-eating cluster
baboon cluster
crab-eating cluster
marmoset cluster

Hornerin
Figure 16

A)

Takaishi et al.



B)

Paar et al.

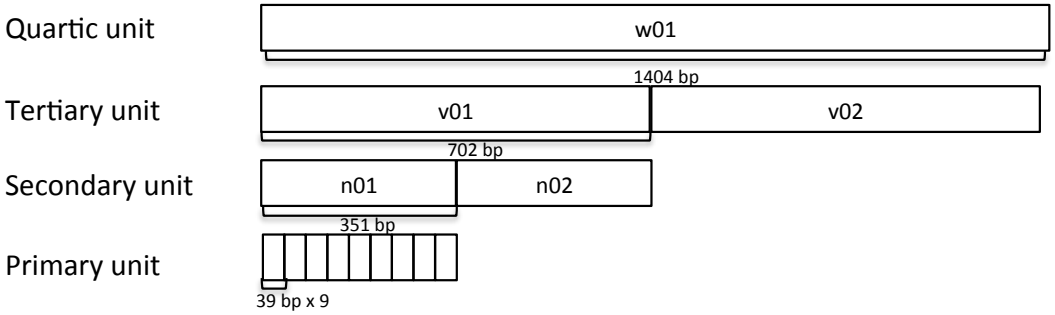


Figure 17


Takaishi et al.
Longest units
Human
1 2 3 4 5 6
1410 1398 1410 1410 1410 936
Chimpanzee
1 2 3 4 5 6
1410 1404 1416 1407 1404 942
Gorilla
1 2 3 4
1398 1410 1326 942
Orangutan
1 2 3 4 5 6
1407 1410 1443 1410 1410 945
Macaque
1 2 3 4 5 6
981 1398 1233 1416 1419 939
Second-longest units or subunits
Human
1A 1B 1C 2A 2B 2C 3A 3B 3C 4A 4B 4C 5A 5B 5C 6A 6B
468 468 474 465 459 474 468 468 474 468 468 474 468 468 474 468 468
Chimpanzee
1A 1B 1C 2A 2B 2C 3A 3B 3C 4A 4B 4C 5A 5B 5C 6A 6B
468 468 474 465 465 474 474 468 474 468 465 474 468 462 474 474 468
Gorilla
1A 1B 1C 2A 2B 2C 3A 3B 3C 4A 4B
468 468 462 474 468 468 390 462 474 474 468
Orangutan
1A 1B 1C 2A 2B 2C 3A 3B 3C 4A 4B 4C 5A 5B 5C 6A 6B
468 465 474 468 468 474 468 501 474 468 468 474 468 468 474 477 468
Macaque
1A 1C 2A 2B 2C 3A 3B 3C 4A 4B 4C 5A 5B 5C 6A 6B
465 516 465 480 453 390 375 468 468 480 468 471 480 468 471 468
Paar et al.
Quartic units
Human
w01 w02 w03 w04 w05 w06
1410 1398 1410 1410 1410 936
Chimpanzee
w01 w02 w03 w04 w05 w06
1410 1404 1416 1407 1404 942
Gorilla
w01 w02 w03 w04
1398 1410 1326 942
Orangutan
w01 w02 w03 w04 w05 w06
1407 1410 1443 1410 1410 945
Tertiary unit
Human
1_v01 1_v02 2_v01 2_v02 3_v01 3_v02 4_v01 4_v02 5_v01 5_v02 6_v01 6_v02
702 708 693 705 702 708 702 708 702 708 702 234
Chimpanzee
1_v01 1_v02 2_v01 2_v02 3_v01 3_v02 4_v01 4_v02 5_v01 5_v02 6_v01 6_v02
702 708 699 705 708 708 702 705 699 705 708 234
Gorilla
1_v01 1_v02 2_v01 2_v02 3_v01 3_v02 4_v01 4_v02
702 696 708 702 621 705 708 234
Orangutan
1_v01 1_v02 2_v01 2_v02 3_v01 3_v02 4_v01 4_v02 5_v01 5_v02 6_v01 6_v02
699 708 702 708 735 708 702 708 702 708 711 234
Secondary unit
Human
1_n01 1_n02 1_n01 1_n02 2_n01 2_n02 2_n01 2_n02 3_n01 3_n02 3_n01 3_n02 4_n01 4_n02 4_n01 4_n02 5_n01 5_n02 5_n01 5_n02 6_n01 6_n02 6_n01
351 351 351 357 348 345 348 357 351 351 351 357 351 351 351 357 351 351 351 357 351 351 234
Chimpanzee
1_n01 1_n02 1_n01 1_n02 2_n01 2_n02 2_n01 2_n02 3_n01 3_n02 3_n01 3_n02 4_n01 4_n02 4_n01 4_n02 5_n01 5_n02 5_n01 5_n02 6_n01 6_n02 6_n01
351 351 351 357 348 351 348 357 351 357 351 357 351 351 348 357 351 348 348 357 357 351 234
Gorilla
1_n01 1_n02 1_n01 1_n02 2_n01 2_n02 2_n01 2_n02 3_n01 3_n02 3_n01 3_n02 4_n01 4_n02 4_n01
351 351 351 345 354 354 351 351 273 348 348 357 357 351 234
Orangutan
1_n01 1_n02 1_n01 1_n02 2_n01 2_n02 2_n01 2_n02 3_n01 3_n02 3_n01 3_n02 4_n01 4_n02 4_n01 4_n02 5_n01 5_n02 5_n01 5_n02 6_n01 6_n02 6_n01
351 348 351 357 351 351 357 351 351 384 351 357 351 351 351 357 351 351 351 357 360 351 234

Figure 18
A.


Macaque-HRNR-4
Human-HRNR-A-I
Human-HRNR-A-II
Human-HRNR-A-III
Human-HRNR-A-IV
Human-HRNR-B-I
Human-HRNR-B-II
Human-HRNR-B-III
Human-HRNR-B-IV
Human-HRNR-C-I
Human-HRNR-C-II
Human-HRNR-C-III
Human-HRNR-C-IV

B.


Human-HRNR-1
No match identifiable
Mouse-HRNR-1-A
Mouse-HRNR-1-B

C.

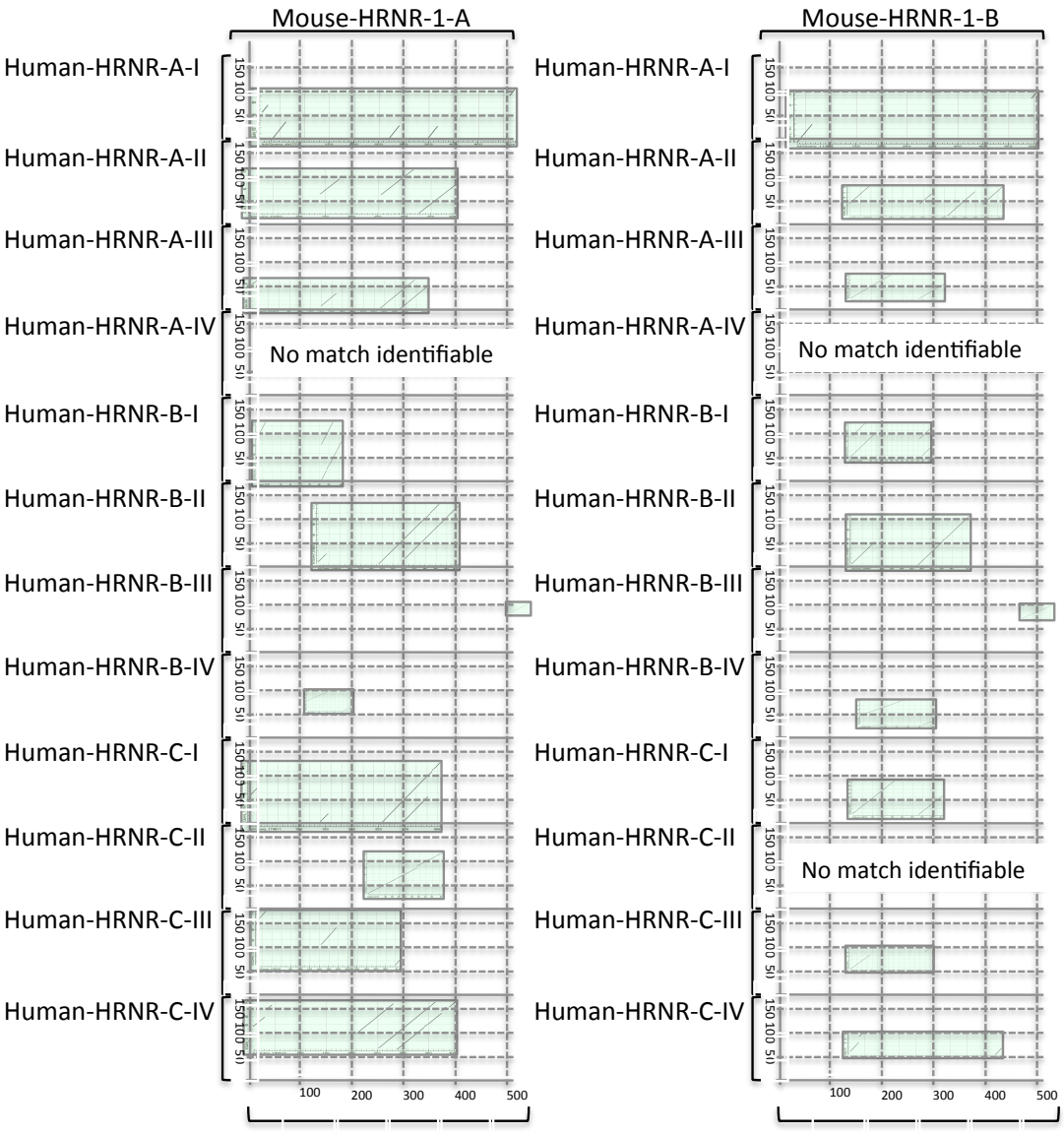

Mouse-HRNR-1-A
Mouse-HRNR-1-B
Human-HRNR-A-I
Human-HRNR-A-II
Human-HRNR-A-III
Human-HRNR-A-IV
No match identifiable
Human-HRNR-B-I
Human-HRNR-B-II
Human-HRNR-B-III
Human-HRNR-B-IV
Human-HRNR-C-I
Human-HRNR-C-II
Human-HRNR-C-III
Human-HRNR-C-IV
No match identifiable
150 100 50
100 200 300 400 500

Figure 19

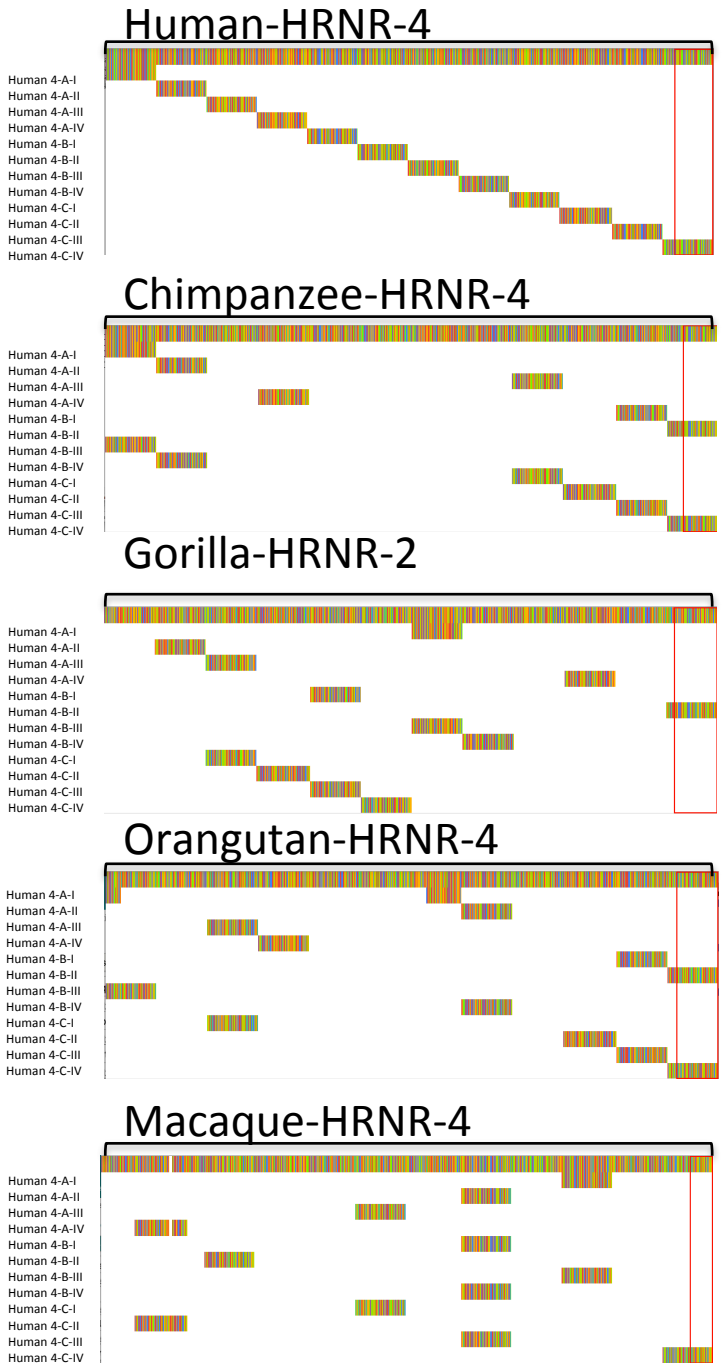
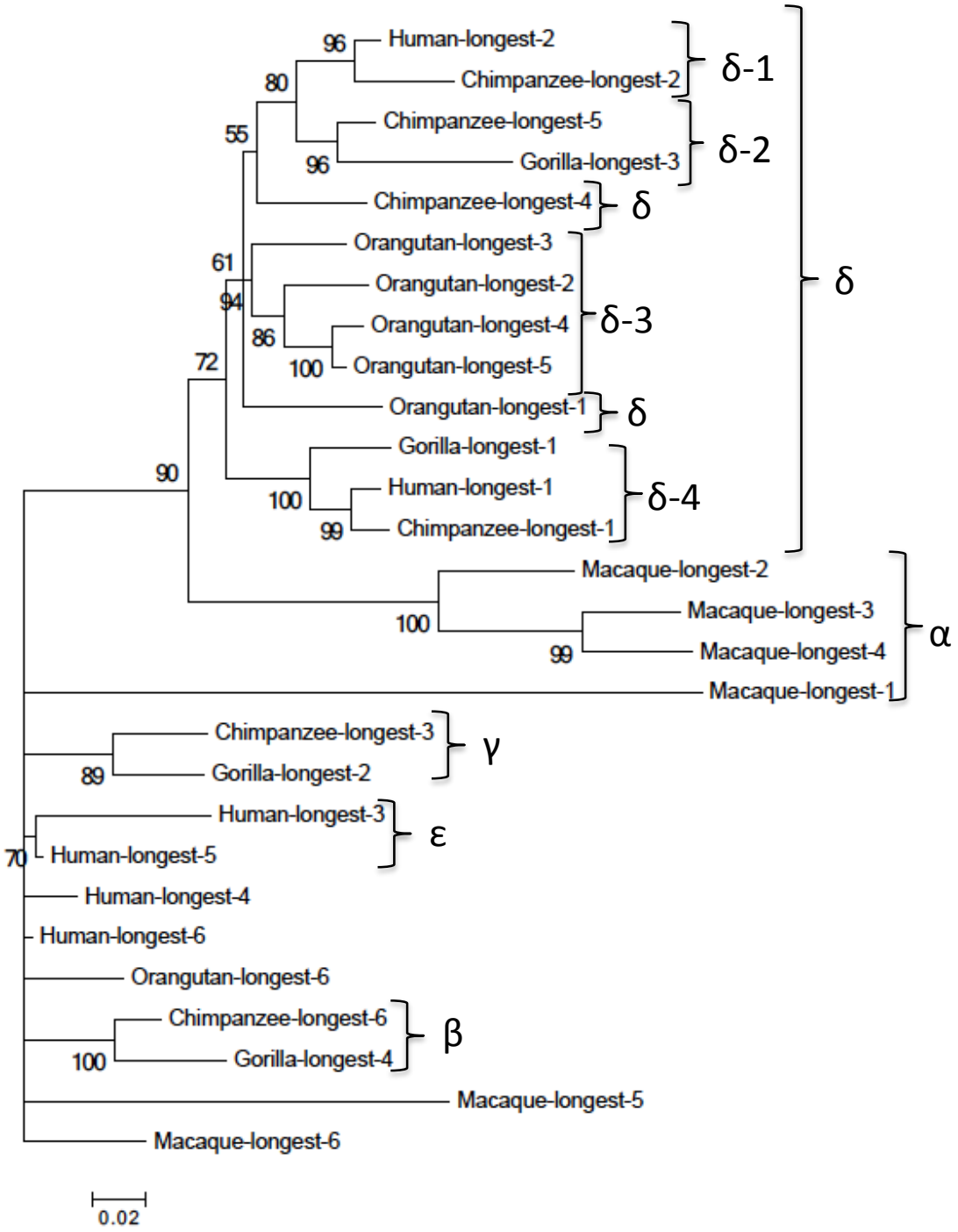
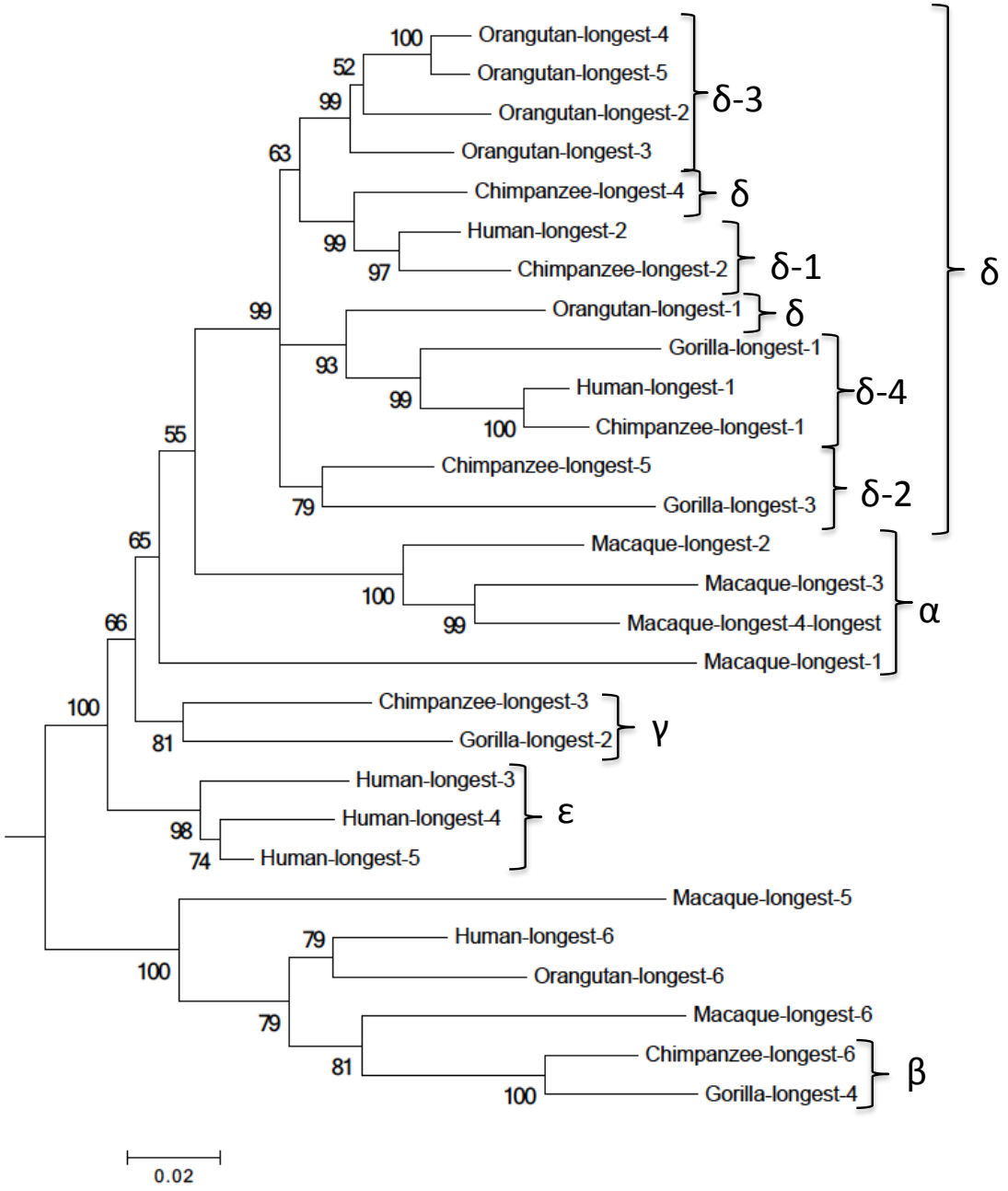

Human-HRNR-4
Human 4-A-I
Human 4-A-II
Human 4-A-III
Human 4-A-IV
Human 4-B-I
Human 4-B-II
Human 4-B-III
Human 4-B-IV
Human 4-C-I
Human 4-C-II
Human 4-C-III
Human 4-C-IV
Chimpanzee-HRNR-4
Human 4-A-I
Human 4-A-II
Human 4-A-III
Human 4-A-IV
Human 4-B-I
Human 4-B-II
Human 4-B-III
Human 4-B-IV
Human 4-C-I
Human 4-C-II
Human 4-C-III
Human 4-C-IV
Gorilla-HRNR-2
Human 4-A-I
Human 4-A-II
Human 4-A-III
Human 4-A-IV
Human 4-B-I
Human 4-B-II
Human 4-B-III
Human 4-B-IV
Human 4-C-I
Human 4-C-II
Human 4-C-III
Human 4-C-IV
Orangutan-HRNR-4
Human 4-A-I
Human 4-A-II
Human 4-A-III
Human 4-A-IV
Human 4-B-I
Human 4-B-II
Human 4-B-III
Human 4-B-IV
Human 4-C-I
Human 4-C-II
Human 4-C-III
Human 4-C-IV
Macaque-HRNR-4
Human 4-A-I
Human 4-A-II
Human 4-A-III
Human 4-A-IV
Human 4-B-I
Human 4-B-II
Human 4-B-III
Human 4-B-IV
Human 4-C-I
Human 4-C-II
Human 4-C-III
Human 4-C-IV

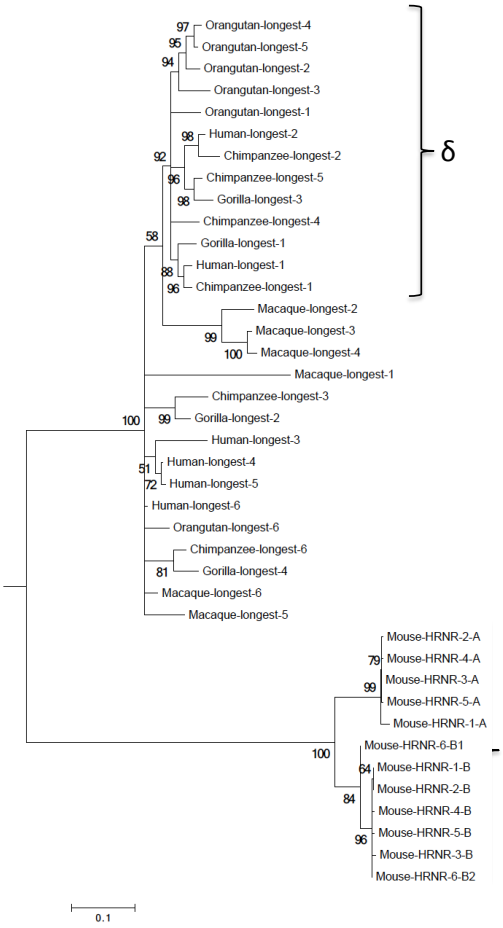
Figure 20
A.


Human-longest-2
Chimpanzee-longest-2
Chimpanzee-longest-5
Gorilla-longest-3
Chimpanzee-longest-4
Orangutan-longest-3
Orangutan-longest-2
Orangutan-longest-4
Orangutan-longest-5
Orangutan-longest-1
Gorilla-longest-1
Human-longest-1
Chimpanzee-longest-1
Macaque-longest-2
Macaque-longest-3
Macaque-longest-4
Macaque-longest-1
Chimpanzee-longest-3
Gorilla-longest-2
Human-longest-3
Human-longest-5
Human-longest-4
Human-longest-6
Orangutan-longest-6
Chimpanzee-longest-6
Gorilla-longest-4
Macaque-longest-5
Macaque-longest-6
96
80
55
96
61
94
86
100
72
90
100
99
100
99
89
70
100
δ-1
δ-2
δ
δ-3
δ
δ-4
δ
α
γ
ε
β
0.02

B.


Orangutan-longest-4
Orangutan-longest-5
Orangutan-longest-2
Orangutan-longest-3
Chimpanzee-longest-4
Human-longest-2
Chimpanzee-longest-2
Orangutan-longest-1
Gorilla-longest-1
Human-longest-1
Chimpanzee-longest-1
Chimpanzee-longest-5
Gorilla-longest-3
Macaque-longest-2
Macaque-longest-3
Macaque-longest-4-longest
Macaque-longest-1
Chimpanzee-longest-3
Gorilla-longest-2
Human-longest-3
Human-longest-4
Human-longest-5
Macaque-longest-5
Human-longest-6
Orangutan-longest-6
Macaque-longest-6
Chimpanzee-longest-6
Gorilla-longest-4
δ-3
δ
δ-1
δ
δ-4
δ-2
δ
α
γ
ε
β
100
52
99
63
99
97
99
93
99
100
55
79
65
100
99
66
100
81
98
74
79
100
79
81
100
0.02

C.



Mouse hornerin units described by Makino et al.

1A	1B	2A	2B	3A	3B	4A	4B	5B1	5B2	6A	6B
534	513	531	513	531	513	531	513	513	513	531	513

D.

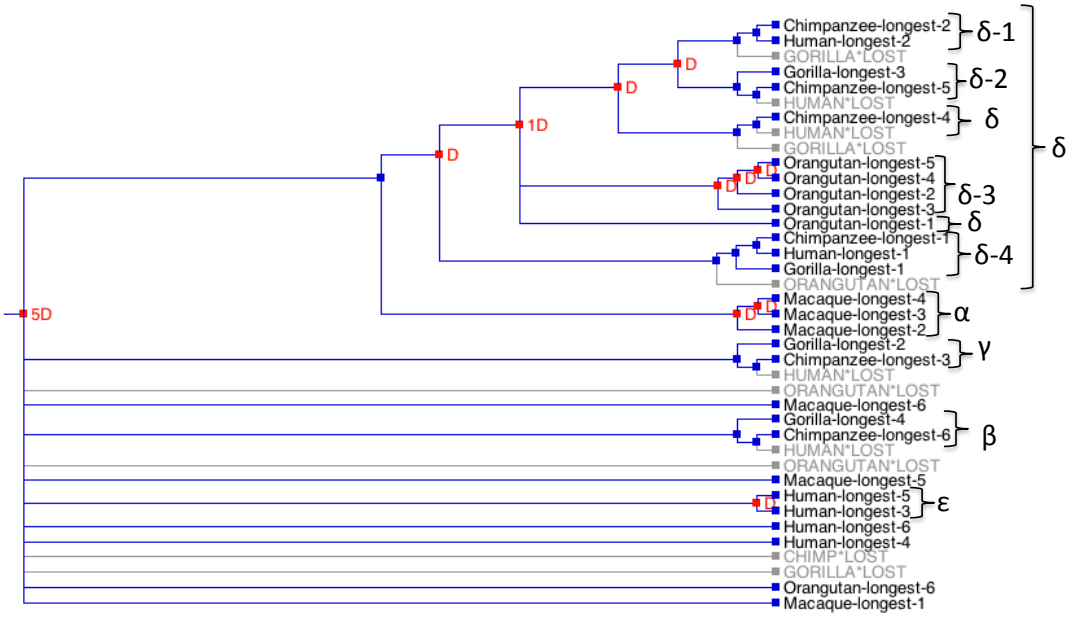

Chimpanzee-longest-2
Human-longest-2
GORILLA*LOST
Gorilla-longest-3
Chimpanzee-longest-5
HUMAN*LOST
Chimpanzee-longest-4
HUMAN*LOST
GORILLA*LOST
Orangutan-longest-5
Orangutan-longest-4
Orangutan-longest-2
Orangutan-longest-3
Orangutan-longest-1
Chimpanzee-longest-1
Human-longest-1
Gorilla-longest-1
ORANGUTAN*LOST
Macaque-longest-4
Macaque-longest-3
Macaque-longest-2
Gorilla-longest-2
Chimpanzee-longest-3
HUMAN*LOST
ORANGUTAN*LOST
Macaque-longest-6
Gorilla-longest-4
Chimpanzee-longest-6
HUMAN*LOST
ORANGUTAN*LOST
Macaque-longest-5
Human-longest-5
Human-longest-3
Human-longest-6
Human-longest-4
CHIMP*LOST
GORILLA*LOST
Orangutan-longest-6
Macaque-longest-1
δ-1
δ-2
δ
δ-3
δ
δ-4
δ
α
γ
β
ε
5D
D
1D

Figure 21

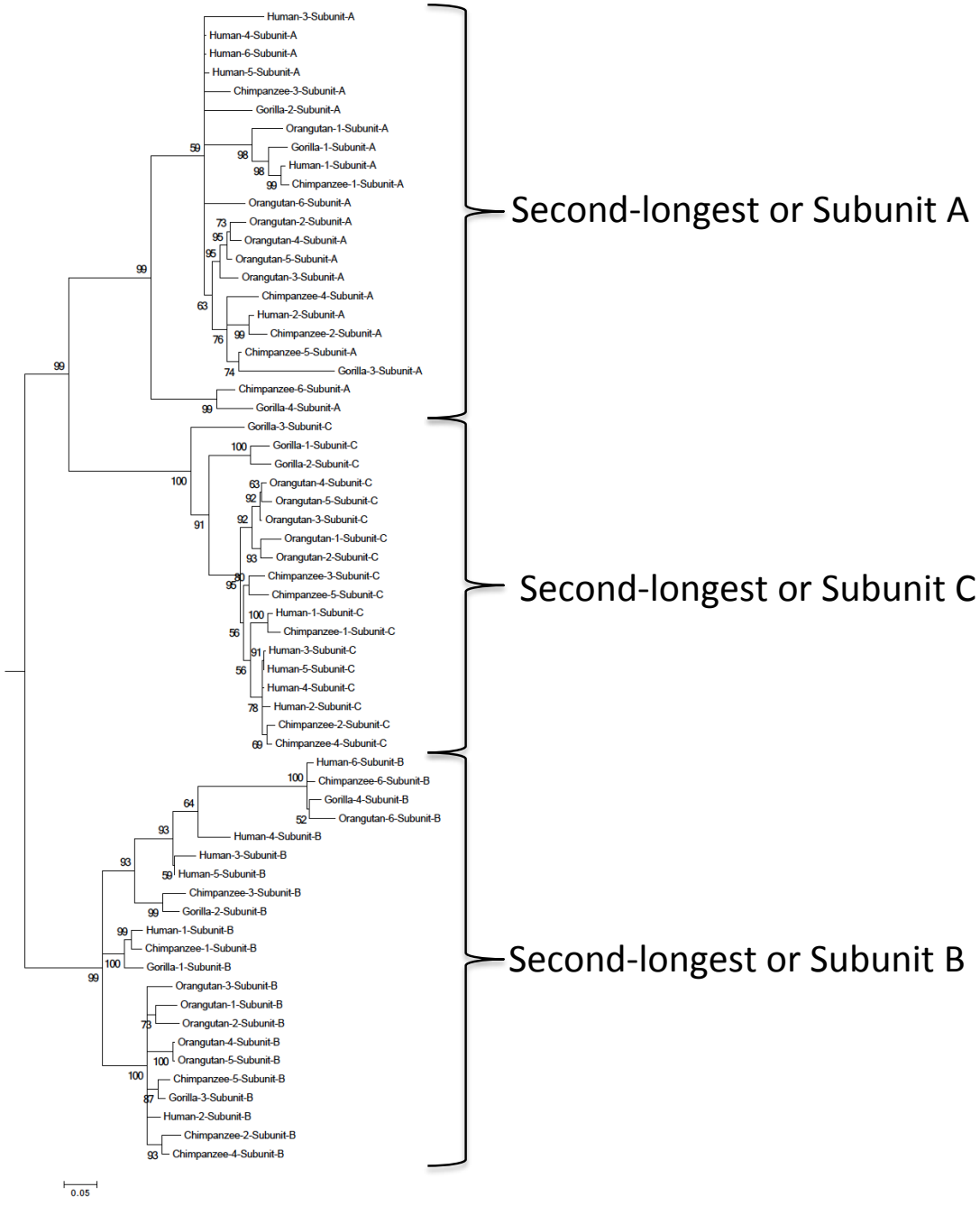
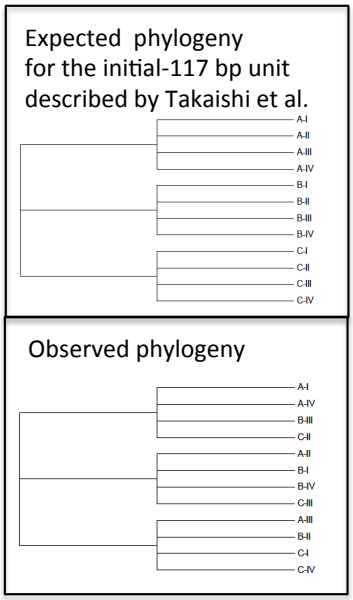
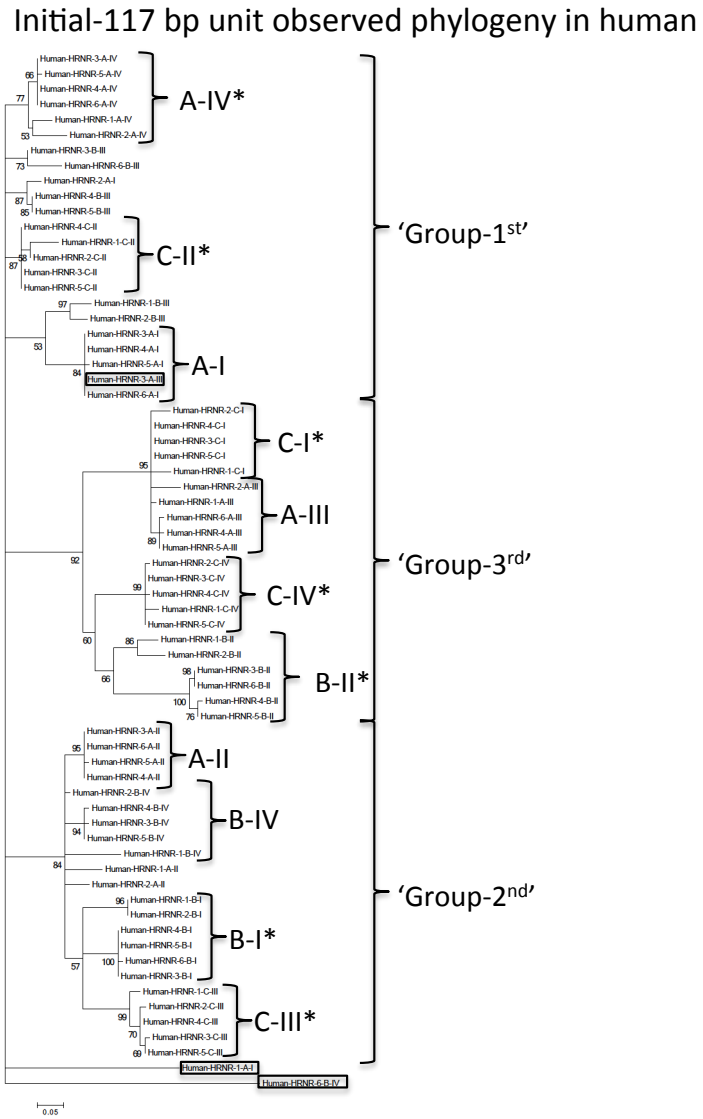

Human-3-Subunit-A
Human-4-Subunit-A
Human-6-Subunit-A
Human-5-Subunit-A
Chimpanzee-3-Subunit-A
Gorilla-2-Subunit-A
Orangutan-1-Subunit-A
Gorilla-1-Subunit-A
Human-1-Subunit-A
Chimpanzee-1-Subunit-A
Orangutan-6-Subunit-A
Orangutan-2-Subunit-A
Orangutan-4-Subunit-A
Orangutan-5-Subunit-A
Orangutan-3-Subunit-A
Chimpanzee-4-Subunit-A
Human-2-Subunit-A
Chimpanzee-2-Subunit-A
Chimpanzee-5-Subunit-A
Gorilla-3-Subunit-A
Chimpanzee-6-Subunit-A
Gorilla-4-Subunit-A
Gorilla-3-Subunit-C
Gorilla-1-Subunit-C
Gorilla-2-Subunit-C
Orangutan-4-Subunit-C
Orangutan-5-Subunit-C
Orangutan-3-Subunit-C
Orangutan-1-Subunit-C
Orangutan-2-Subunit-C
Chimpanzee-3-Subunit-C
Chimpanzee-5-Subunit-C
Human-1-Subunit-C
Chimpanzee-1-Subunit-C
Human-3-Subunit-C
Human-5-Subunit-C
Human-4-Subunit-C
Human-2-Subunit-C
Chimpanzee-2-Subunit-C
Chimpanzee-4-Subunit-C
Human-6-Subunit-B
Chimpanzee-6-Subunit-B
Gorilla-4-Subunit-B
Orangutan-6-Subunit-B
Human-4-Subunit-B
Human-3-Subunit-B
Human-5-Subunit-B
Chimpanzee-3-Subunit-B
Gorilla-2-Subunit-B
Human-1-Subunit-B
Chimpanzee-1-Subunit-B
Gorilla-1-Subunit-B
Orangutan-3-Subunit-B
Orangutan-1-Subunit-B
Orangutan-2-Subunit-B
Orangutan-4-Subunit-B
Orangutan-5-Subunit-B
Chimpanzee-5-Subunit-B
Gorilla-3-Subunit-B
Human-2-Subunit-B
Chimpanzee-2-Subunit-B
Chimpanzee-4-Subunit-B
0.05
Second-longest or Subunit A
Second-longest or Subunit C
Second-longest or Subunit B

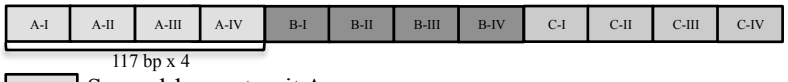
Figure 22

A.





Expected formation for the initial-117 bp units described by Takaishi et al.

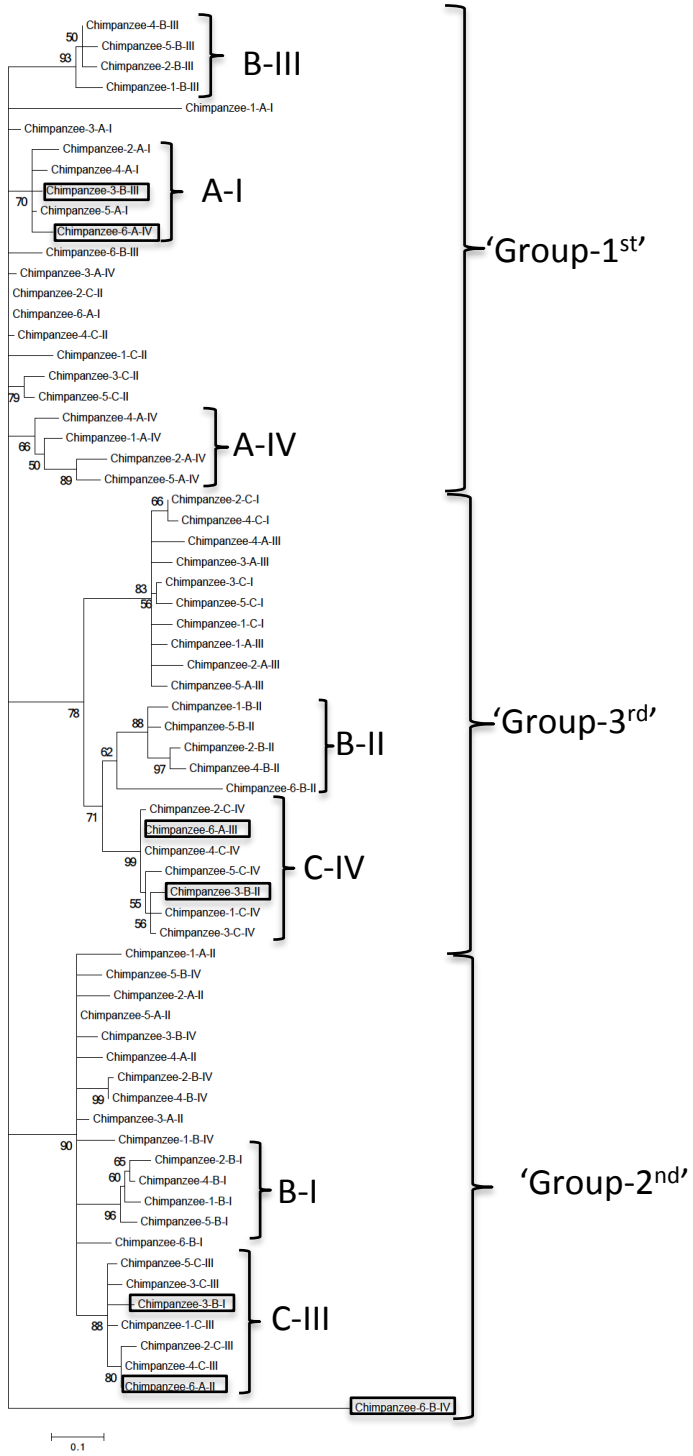


Second-longest unit A
Second-longest unit B
Second-longest unit C

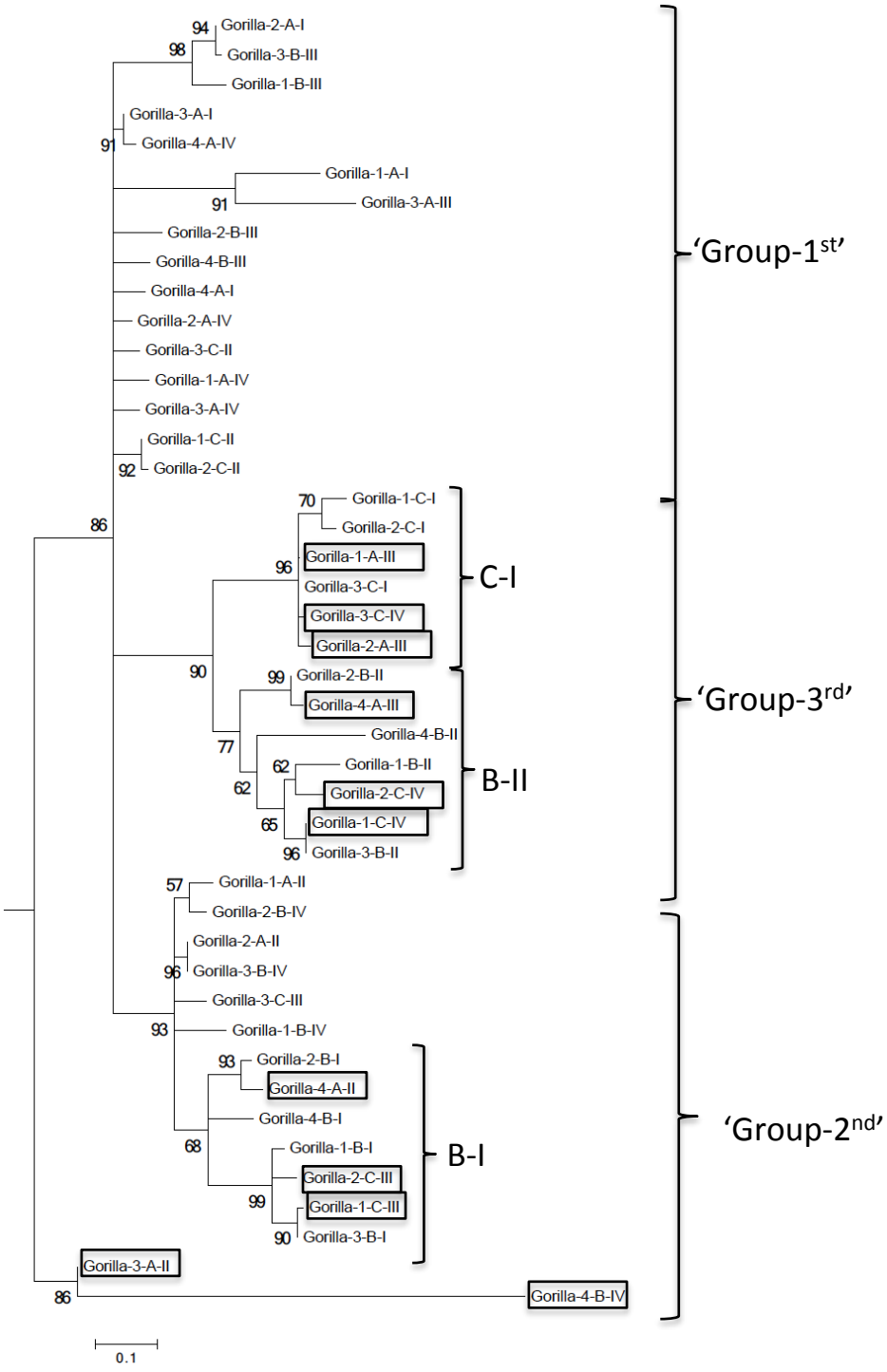
Observed formation for the initial-117 bp units

'Group-1st': A-I, A-IV, B-III and C-II
'Group-2nd': A-II, B-I, B-IV and C-III
'Group-3rd': A-III, B-II, C-I and C-IV

B.


Chimpanzee-4-B-III
Chimpanzee-5-B-III
Chimpanzee-2-B-III
Chimpanzee-1-B-III
B-III
Chimpanzee-1-A-I
Chimpanzee-3-A-I
Chimpanzee-2-A-I
Chimpanzee-4-A-I
Chimpanzee-3-B-III
Chimpanzee-5-A-I
Chimpanzee-6-A-IV
A-I
Chimpanzee-6-B-III
Chimpanzee-3-A-IV
Chimpanzee-2-C-II
Chimpanzee-6-A-I
Chimpanzee-4-C-II
Chimpanzee-1-C-II
Chimpanzee-3-C-II
Chimpanzee-5-C-II
Chimpanzee-4-A-IV
Chimpanzee-1-A-IV
Chimpanzee-2-A-IV
Chimpanzee-5-A-IV
A-IV
'Group-1st'
Chimpanzee-2-C-I
Chimpanzee-4-C-I
Chimpanzee-4-A-III
Chimpanzee-3-A-III
Chimpanzee-3-C-I
Chimpanzee-5-C-I
Chimpanzee-1-C-I
Chimpanzee-1-A-III
Chimpanzee-2-A-III
Chimpanzee-5-A-III
Chimpanzee-1-B-II
Chimpanzee-5-B-II
Chimpanzee-2-B-II
Chimpanzee-4-B-II
Chimpanzee-6-B-II
B-II
Chimpanzee-2-C-IV
Chimpanzee-6-A-III
Chimpanzee-4-C-IV
Chimpanzee-5-C-IV
Chimpanzee-3-B-II
Chimpanzee-1-C-IV
Chimpanzee-3-C-IV
C-IV
'Group-3rd'
Chimpanzee-1-A-II
Chimpanzee-5-B-IV
Chimpanzee-2-A-II
Chimpanzee-5-A-II
Chimpanzee-3-B-IV
Chimpanzee-4-A-II
Chimpanzee-2-B-IV
Chimpanzee-4-B-IV
Chimpanzee-3-A-II
Chimpanzee-1-B-IV
Chimpanzee-2-B-I
Chimpanzee-4-B-I
Chimpanzee-1-B-I
Chimpanzee-5-B-I
B-I
Chimpanzee-6-B-I
Chimpanzee-5-C-III
Chimpanzee-3-C-III
Chimpanzee-3-B-I
Chimpanzee-1-C-III
Chimpanzee-2-C-III
Chimpanzee-4-C-III
Chimpanzee-6-A-II
C-III
'Group-2nd'
Chimpanzee-6-B-IV
0.1

C.


Gorilla-2-A-I
Gorilla-3-B-III
Gorilla-1-B-III
Gorilla-3-A-I
Gorilla-4-A-IV
Gorilla-1-A-I
Gorilla-3-A-III
Gorilla-2-B-III
Gorilla-4-B-III
Gorilla-4-A-I
Gorilla-2-A-IV
Gorilla-3-C-II
Gorilla-1-A-IV
Gorilla-3-A-IV
Gorilla-1-C-II
Gorilla-2-C-II
Gorilla-1-C-I
Gorilla-2-C-I
Gorilla-1-A-III
Gorilla-3-C-I
Gorilla-3-C-IV
Gorilla-2-A-III
Gorilla-2-B-II
Gorilla-4-A-III
Gorilla-4-B-II
Gorilla-1-B-II
Gorilla-2-C-IV
Gorilla-1-C-IV
Gorilla-3-B-II
Gorilla-1-A-II
Gorilla-2-B-IV
Gorilla-2-A-II
Gorilla-3-B-IV
Gorilla-3-C-III
Gorilla-1-B-IV
Gorilla-2-B-I
Gorilla-4-A-II
Gorilla-4-B-I
Gorilla-1-B-I
Gorilla-2-C-III
Gorilla-1-C-III
Gorilla-3-B-I
Gorilla-3-A-II
Gorilla-4-B-IV
0.1
C-I
B-II
B-I
'Group-1st'
'Group-3rd'
'Group-2nd'

D.

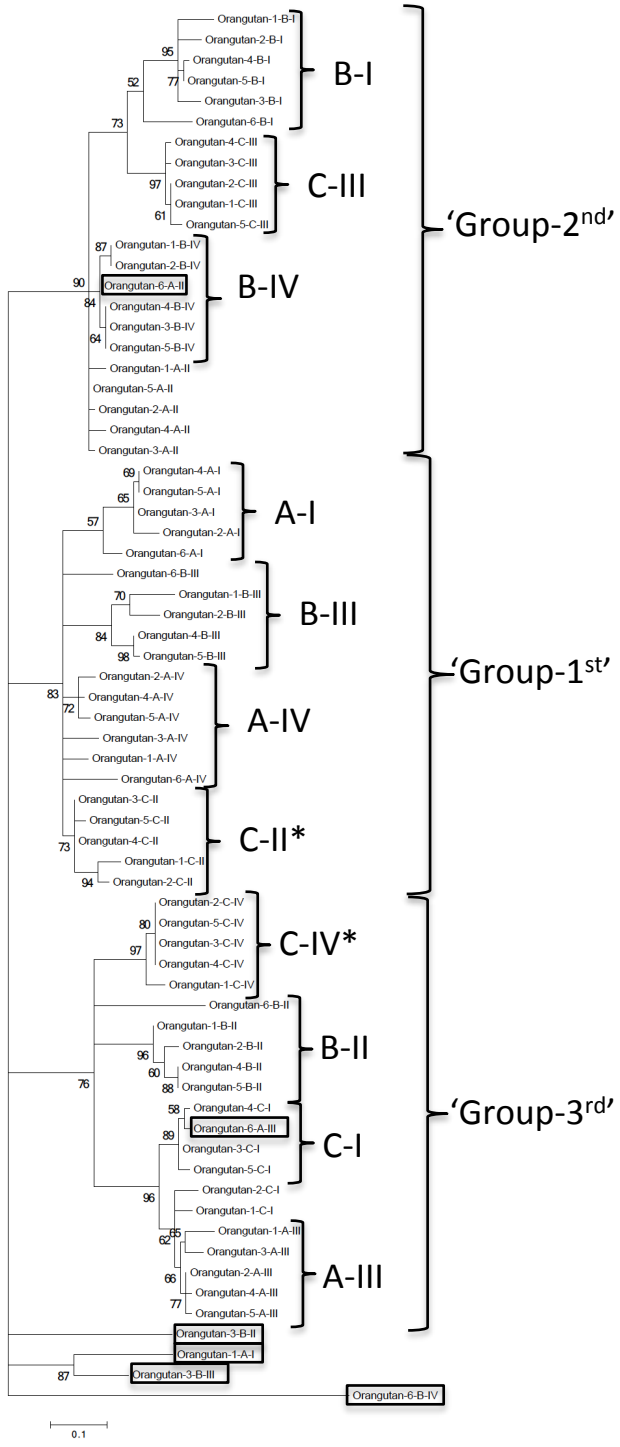
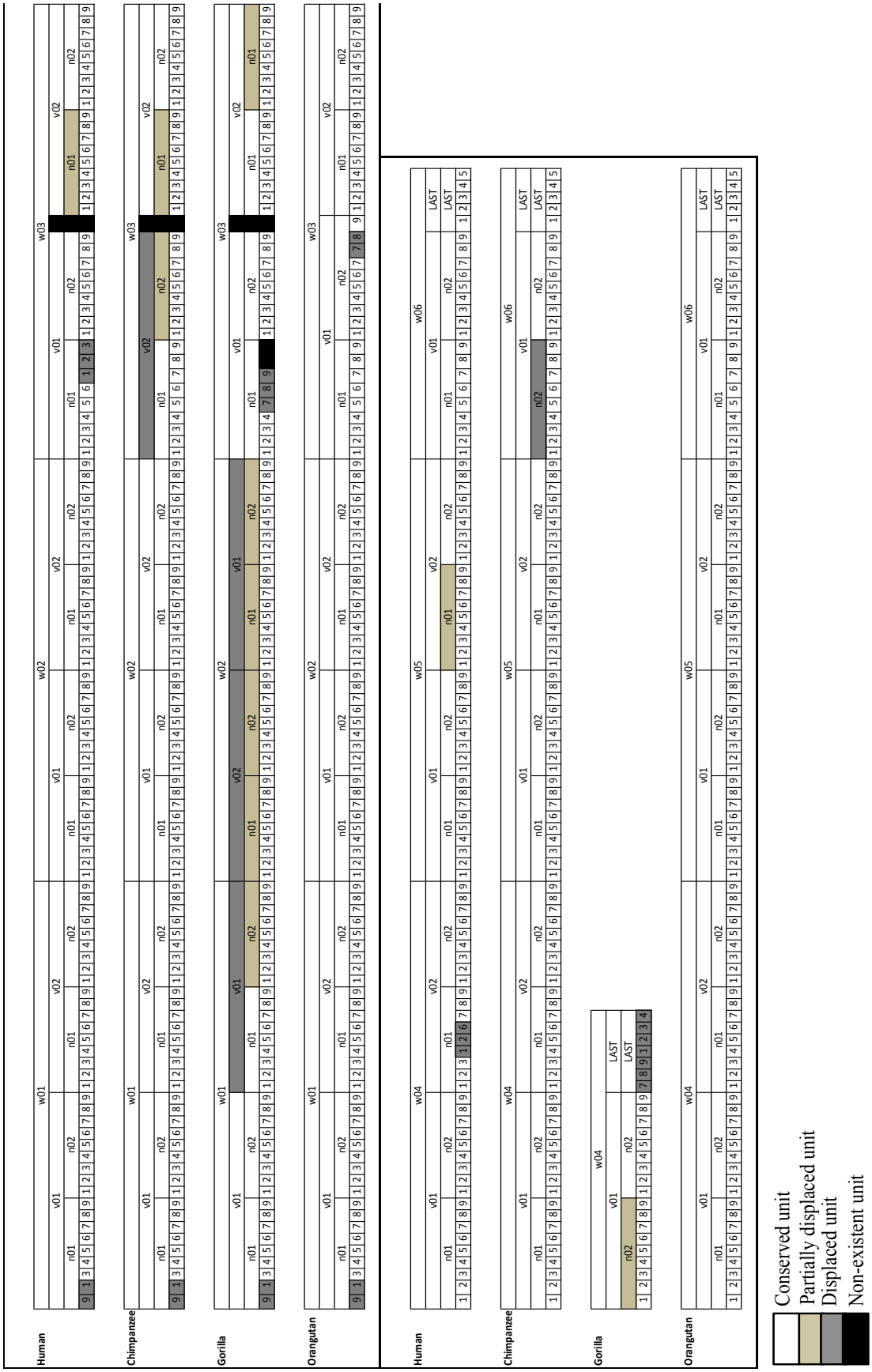
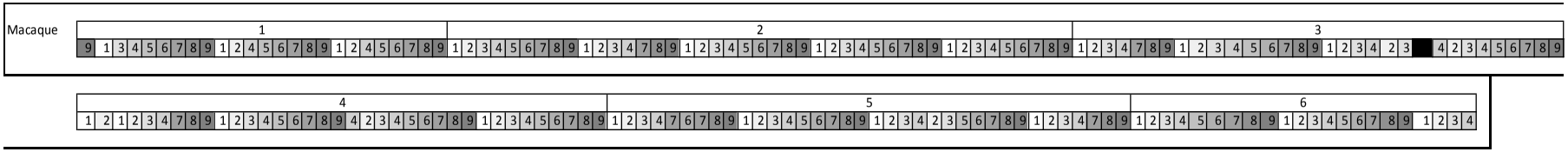

Orangutan-1-B-I
Orangutan-2-B-I
Orangutan-4-B-I
Orangutan-5-B-I
Orangutan-3-B-I
Orangutan-6-B-I
B-I
Orangutan-4-C-III
Orangutan-3-C-III
Orangutan-2-C-III
Orangutan-1-C-III
Orangutan-5-C-III
C-III
Orangutan-1-B-IV
Orangutan-2-B-IV
Orangutan-6-A-II
Orangutan-4-B-IV
Orangutan-3-B-IV
Orangutan-5-B-IV
B-IV
Orangutan-1-A-II
Orangutan-5-A-II
Orangutan-2-A-II
Orangutan-4-A-II
Orangutan-3-A-II
'Group-2nd'
Orangutan-4-A-I
Orangutan-5-A-I
Orangutan-3-A-I
Orangutan-2-A-I
Orangutan-6-A-I
A-I
Orangutan-6-B-III
Orangutan-1-B-III
Orangutan-2-B-III
Orangutan-4-B-III
Orangutan-5-B-III
B-III
Orangutan-2-A-IV
Orangutan-4-A-IV
Orangutan-5-A-IV
Orangutan-3-A-IV
Orangutan-1-A-IV
Orangutan-6-A-IV
A-IV
Orangutan-3-C-II
Orangutan-5-C-II
Orangutan-4-C-II
Orangutan-1-C-II
Orangutan-2-C-II
C-II*
'Group-1st'
Orangutan-2-C-IV
Orangutan-5-C-IV
Orangutan-3-C-IV
Orangutan-4-C-IV
Orangutan-1-C-IV
C-IV*
Orangutan-6-B-II
Orangutan-1-B-II
Orangutan-2-B-II
Orangutan-4-B-II
Orangutan-5-B-II
B-II
Orangutan-4-C-I
Orangutan-6-A-III
Orangutan-3-C-I
Orangutan-5-C-I
C-I
Orangutan-2-C-I
Orangutan-1-C-I
Orangutan-1-A-III
Orangutan-3-A-III
Orangutan-2-A-III
Orangutan-4-A-III
Orangutan-5-A-III
A-III
'Group-3rd'
Orangutan-3-B-II
Orangutan-1-A-I
Orangutan-3-B-III
Orangutan-6-B-IV
0.1

Figure 23

A.

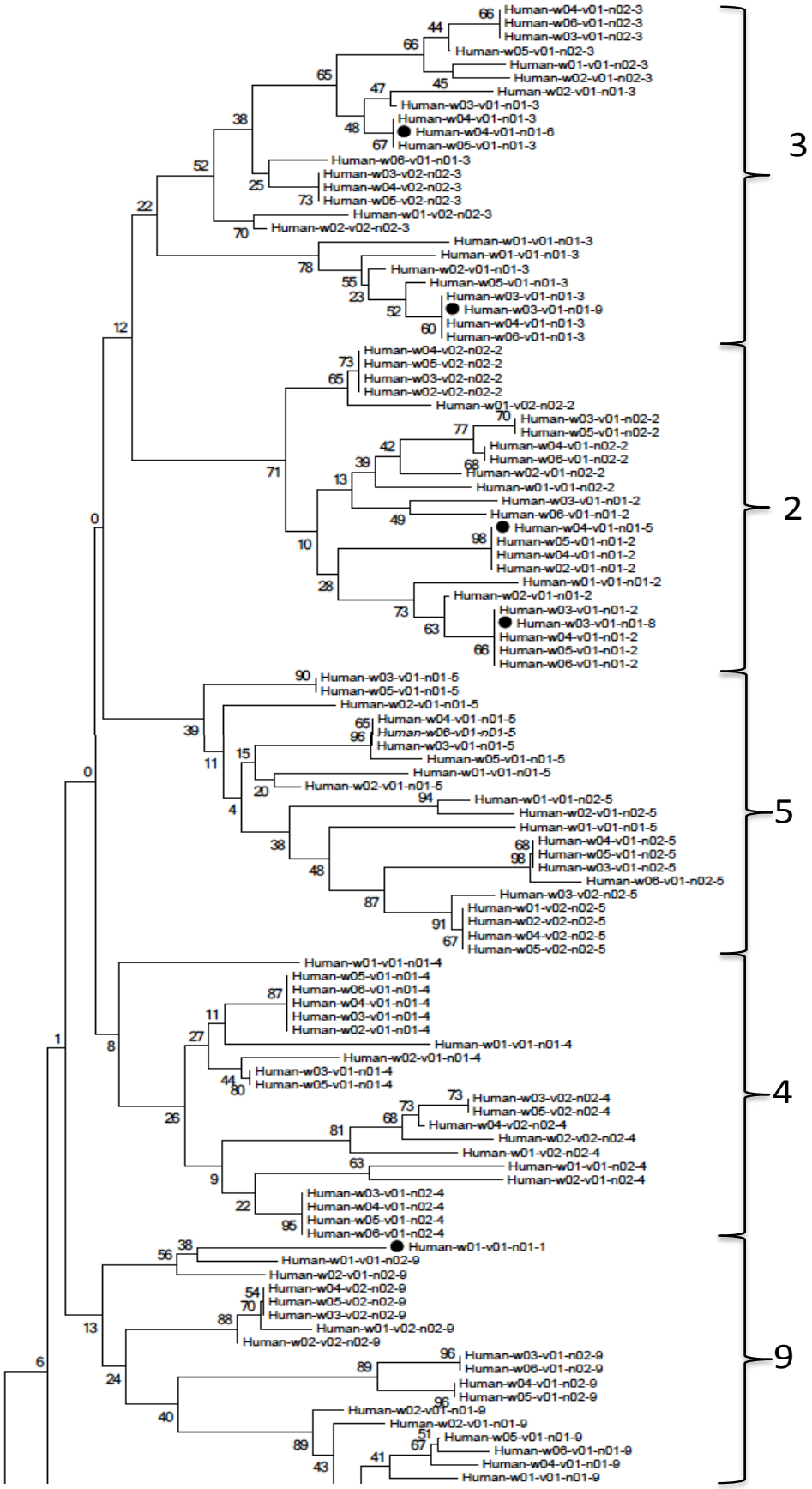

Human
Chimpanzee
Gorilla
Orangutan
Conserved unit
Partially displaced unit
Displaced unit
Non-existent unit

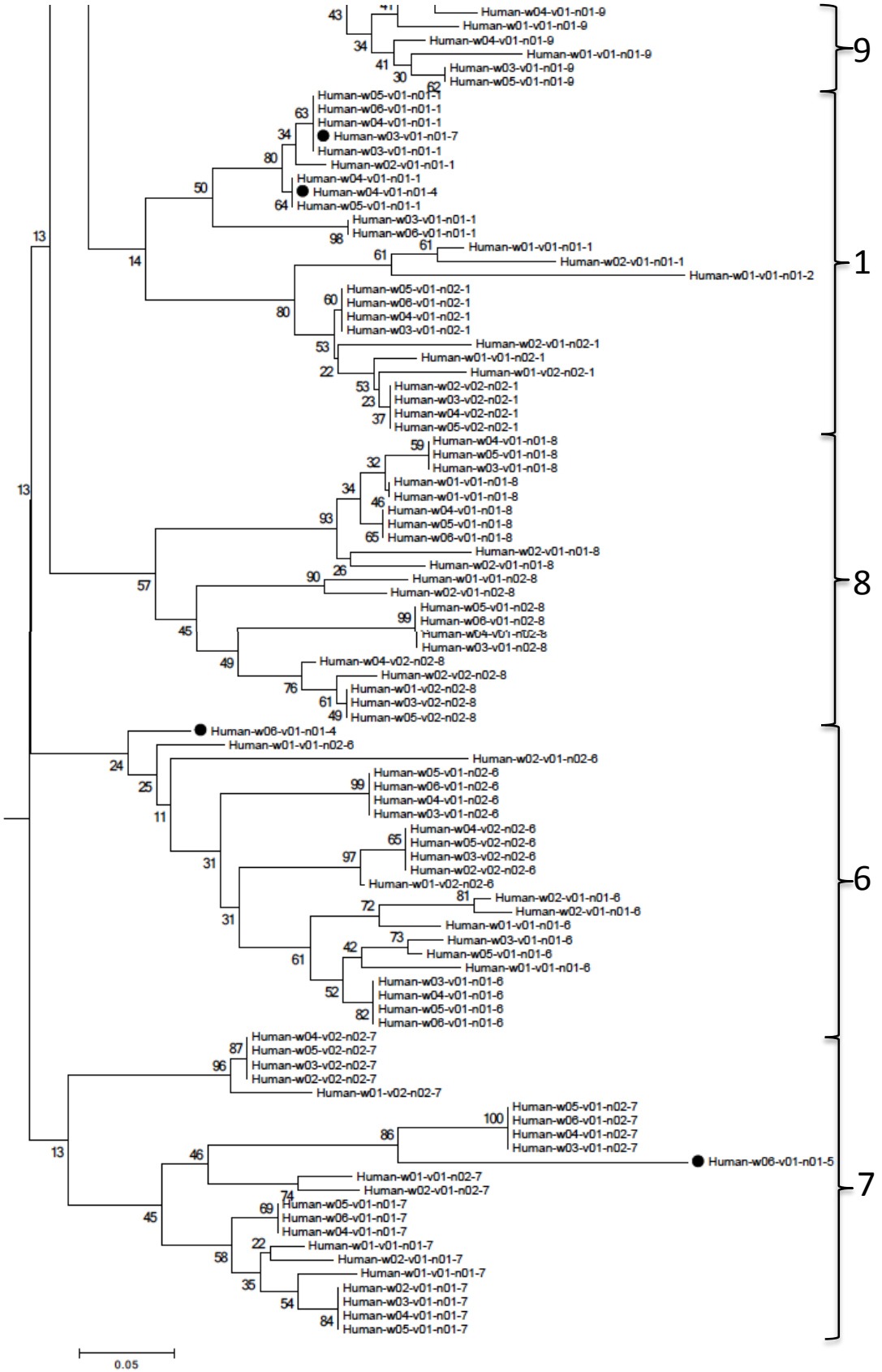
B.



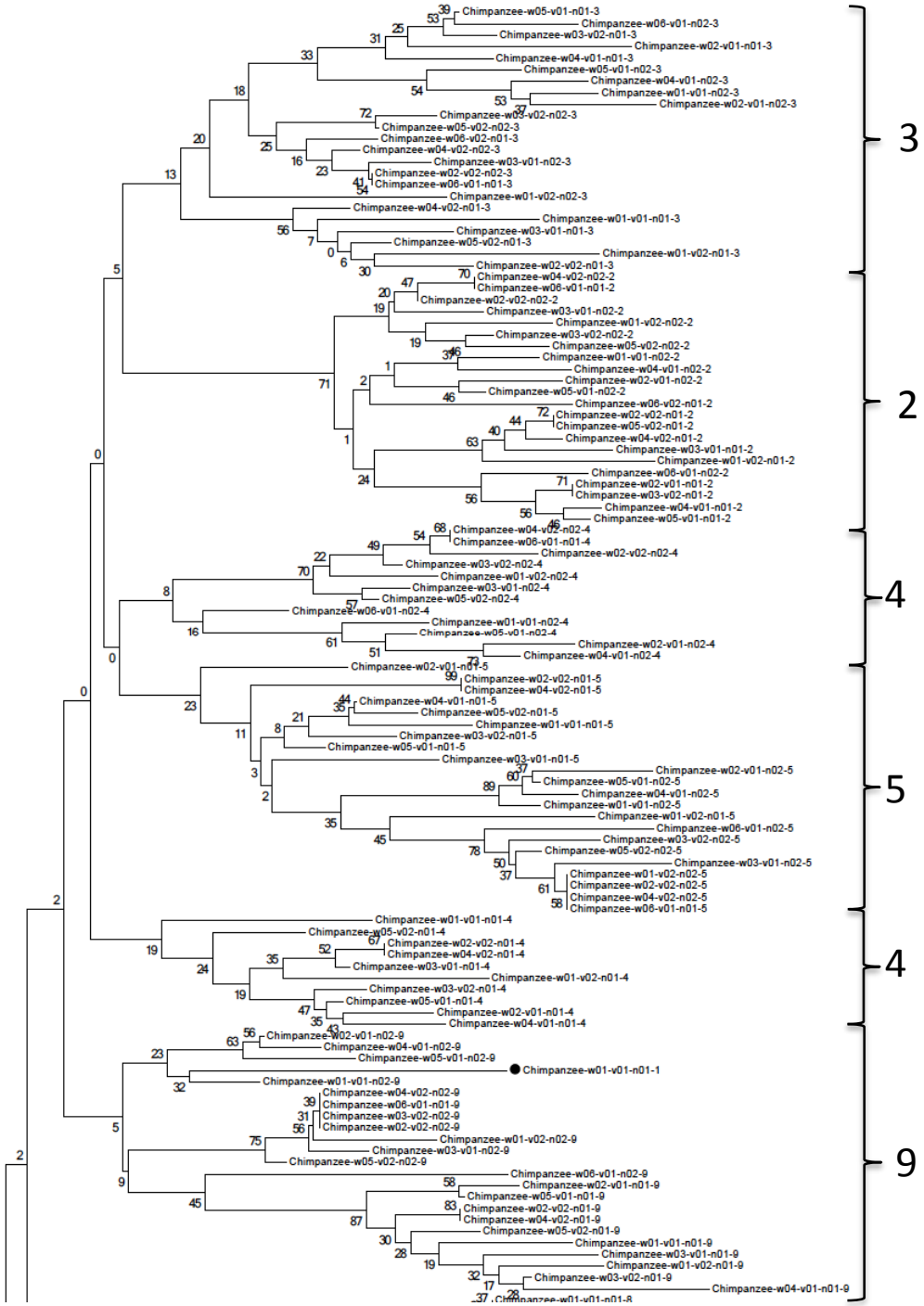
Color gradient for the primary unit order

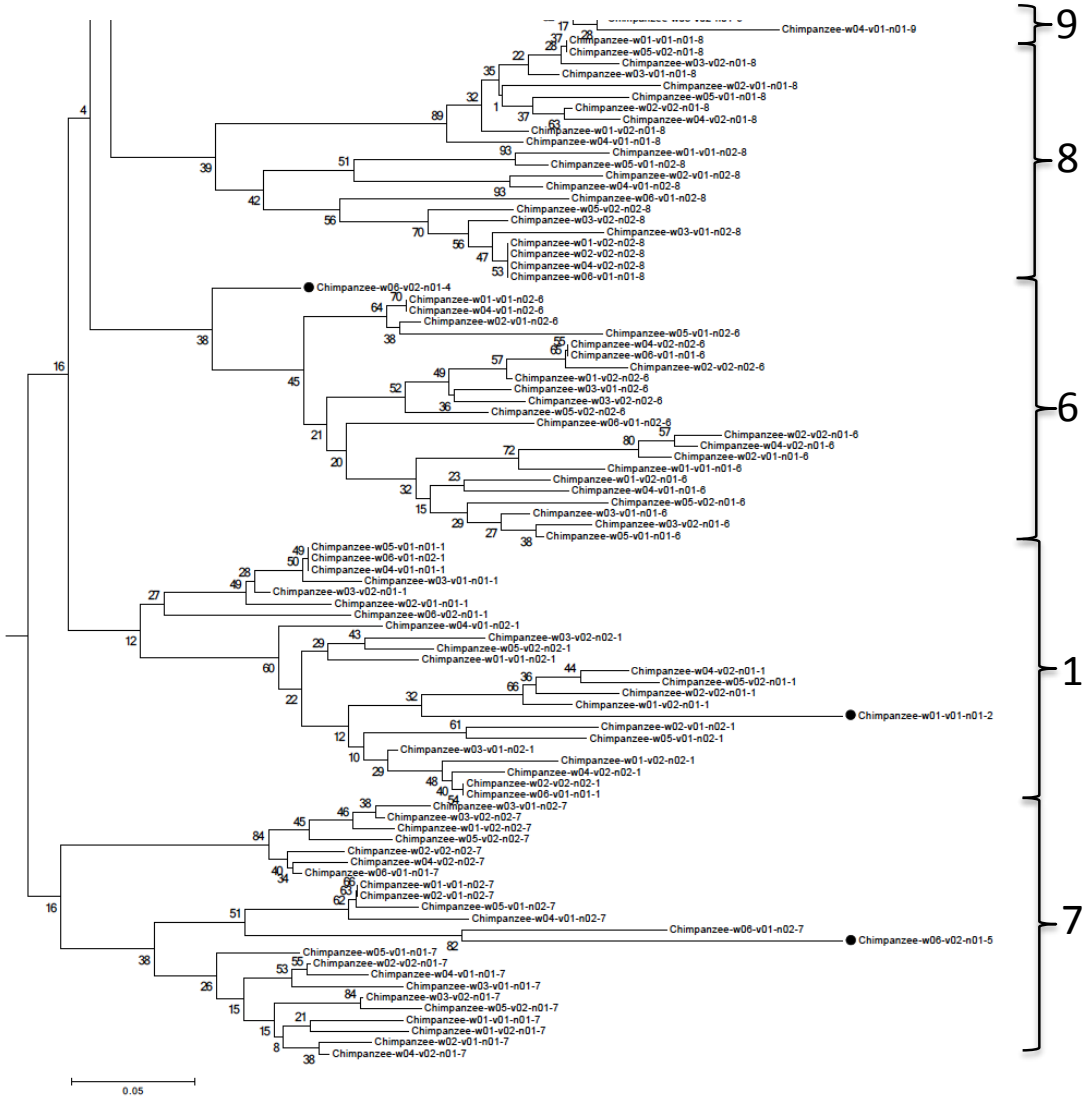
Figure 24
A.


Human-w04-v01-n02-3
Human-w06-v01-n02-3
Human-w03-v01-n02-3
Human-w05-v01-n02-3
Human-w01-v01-n02-3
Human-w02-v01-n02-3
Human-w02-v01-n01-3
Human-w03-v01-n01-3
Human-w04-v01-n01-3
Human-w04-v01-n01-6
Human-w05-v01-n01-3
Human-w06-v01-n01-3
Human-w03-v02-n02-3
Human-w04-v02-n02-3
Human-w05-v02-n02-3
Human-w01-v02-n02-3
Human-w02-v02-n02-3
Human-w01-v01-n01-3
Human-w01-v01-n01-3
Human-w02-v01-n01-3
Human-w05-v01-n01-3
Human-w03-v01-n01-3
Human-w03-v01-n01-9
Human-w04-v01-n01-3
Human-w06-v01-n01-3
Human-w04-v02-n02-2
Human-w05-v02-n02-2
Human-w03-v02-n02-2
Human-w02-v02-n02-2
Human-w01-v02-n02-2
Human-w03-v01-n02-2
Human-w05-v01-n02-2
Human-w04-v01-n02-2
Human-w06-v01-n02-2
Human-w02-v01-n02-2
Human-w01-v01-n02-2
Human-w03-v01-n01-2
Human-w06-v01-n01-2
Human-w04-v01-n01-5
Human-w05-v01-n01-2
Human-w04-v01-n01-2
Human-w02-v01-n01-2
Human-w01-v01-n01-2
Human-w02-v01-n01-2
Human-w03-v01-n01-2
Human-w03-v01-n01-8
Human-w04-v01-n01-2
Human-w05-v01-n01-2
Human-w06-v01-n01-2
Human-w03-v01-n01-5
Human-w05-v01-n01-5
Human-w02-v01-n01-5
Human-w04-v01-n01-5
Human-w06-v01-n01-5
Human-w03-v01-n01-5
Human-w05-v01-n01-5
Human-w01-v01-n01-5
Human-w02-v01-n01-5
Human-w01-v01-n02-5
Human-w02-v01-n02-5
Human-w01-v01-n01-5
Human-w04-v01-n02-5
Human-w05-v01-n02-5
Human-w03-v01-n02-5
Human-w06-v01-n02-5
Human-w03-v02-n02-5
Human-w01-v02-n02-5
Human-w02-v02-n02-5
Human-w04-v02-n02-5
Human-w05-v02-n02-5
Human-w01-v01-n01-4
Human-w05-v01-n01-4
Human-w06-v01-n01-4
Human-w04-v01-n01-4
Human-w03-v01-n01-4
Human-w02-v01-n01-4
Human-w01-v01-n01-4
Human-w02-v01-n01-4
Human-w03-v01-n01-4
Human-w05-v01-n01-4
Human-w03-v02-n02-4
Human-w05-v02-n02-4
Human-w04-v02-n02-4
Human-w02-v02-n02-4
Human-w01-v02-n02-4
Human-w01-v01-n02-4
Human-w02-v01-n02-4
Human-w03-v01-n02-4
Human-w04-v01-n02-4
Human-w05-v01-n02-4
Human-w06-v01-n02-4
Human-w01-v01-n01-1
Human-w01-v01-n02-9
Human-w02-v01-n02-9
Human-w04-v02-n02-9
Human-w05-v02-n02-9
Human-w03-v02-n02-9
Human-w01-v02-n02-9
Human-w02-v02-n02-9
Human-w03-v01-n02-9
Human-w06-v01-n02-9
Human-w04-v01-n02-9
Human-w05-v01-n02-9
Human-w02-v01-n01-9
Human-w02-v01-n01-9
Human-w05-v01-n01-9
Human-w06-v01-n01-9
Human-w04-v01-n01-9
Human-w01-v01-n01-9
3
2
5
4
9

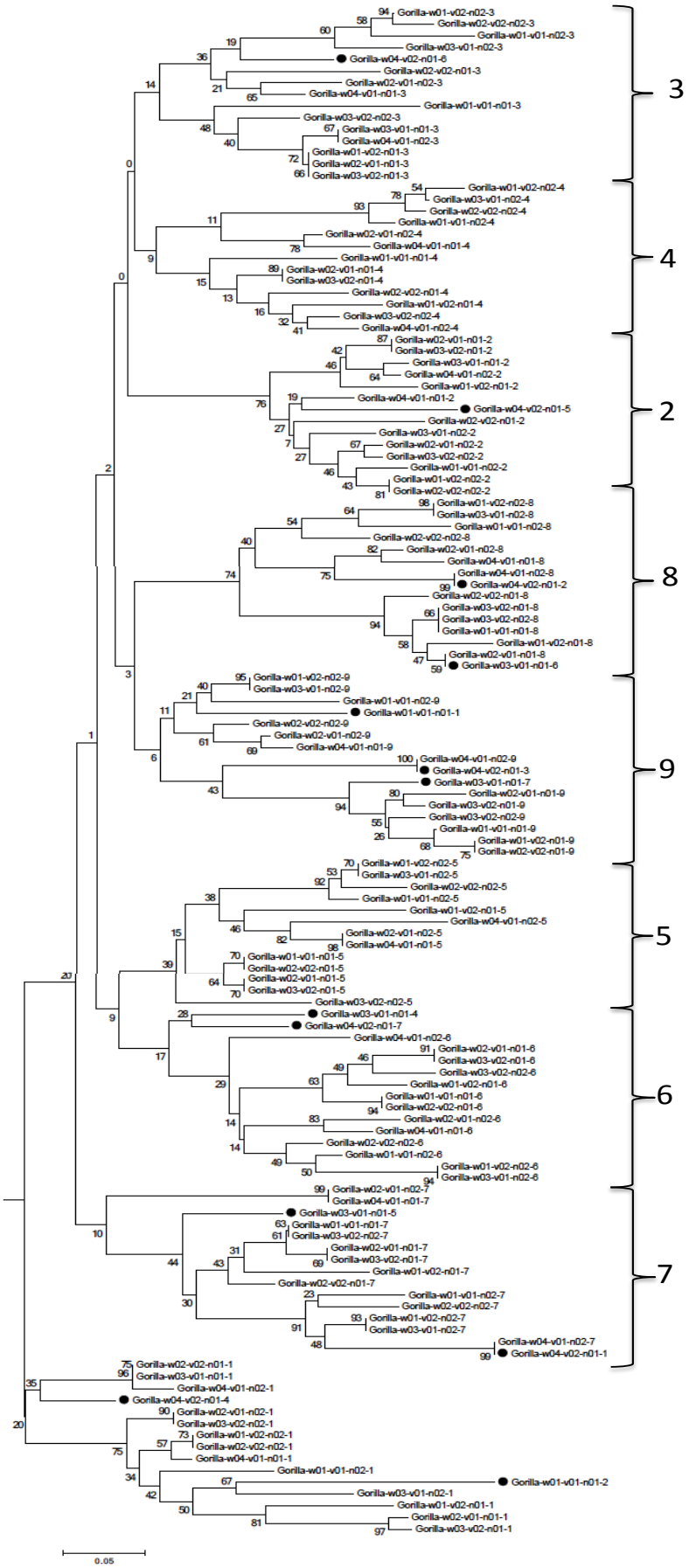

Human-w04-v01-n01-9
Human-w01-v01-n01-9
Human-w04-v01-n01-9
Human-w01-v01-n01-9
Human-w03-v01-n01-9
Human-w05-v01-n01-9
Human-w05-v01-n01-1
Human-w06-v01-n01-1
Human-w04-v01-n01-1
Human-w03-v01-n01-7
Human-w03-v01-n01-1
Human-w02-v01-n01-1
Human-w04-v01-n01-1
Human-w04-v01-n01-4
Human-w05-v01-n01-1
Human-w03-v01-n01-1
Human-w06-v01-n01-1
Human-w01-v01-n01-1
Human-w02-v01-n01-1
Human-w01-v01-n01-2
Human-w05-v01-n02-1
Human-w06-v01-n02-1
Human-w04-v01-n02-1
Human-w03-v01-n02-1
Human-w02-v01-n02-1
Human-w01-v01-n02-1
Human-w01-v02-n02-1
Human-w02-v02-n02-1
Human-w03-v02-n02-1
Human-w04-v02-n02-1
Human-w05-v02-n02-1
Human-w04-v01-n01-8
Human-w05-v01-n01-8
Human-w03-v01-n01-8
Human-w01-v01-n01-8
Human-w01-v01-n01-8
Human-w04-v01-n01-8
Human-w05-v01-n01-8
Human-w06-v01-n01-8
Human-w02-v01-n01-8
Human-w02-v01-n01-8
Human-w01-v01-n02-8
Human-w02-v01-n02-8
Human-w05-v01-n02-8
Human-w06-v01-n02-8
Human-w03-v01-n02-8
Human-w04-v02-n02-8
Human-w02-v02-n02-8
Human-w01-v02-n02-8
Human-w03-v02-n02-8
Human-w05-v02-n02-8
Human-w06-v01-n01-4
Human-w01-v01-n02-6
Human-w02-v01-n02-6
Human-w05-v01-n02-6
Human-w06-v01-n02-6
Human-w04-v01-n02-6
Human-w03-v01-n02-6
Human-w04-v02-n02-6
Human-w05-v02-n02-6
Human-w03-v02-n02-6
Human-w02-v02-n02-6
Human-w01-v02-n02-6
Human-w02-v01-n01-6
Human-w02-v01-n01-6
Human-w01-v01-n01-6
Human-w03-v01-n01-6
Human-w05-v01-n01-6
Human-w01-v01-n01-6
Human-w03-v01-n01-6
Human-w04-v01-n01-6
Human-w05-v01-n01-6
Human-w06-v01-n01-6
Human-w04-v02-n02-7
Human-w05-v02-n02-7
Human-w03-v02-n02-7
Human-w02-v02-n02-7
Human-w01-v02-n02-7
Human-w05-v01-n02-7
Human-w06-v01-n02-7
Human-w04-v01-n02-7
Human-w03-v01-n02-7
Human-w06-v01-n01-5
Human-w01-v01-n02-7
Human-w02-v01-n02-7
Human-w05-v01-n01-7
Human-w06-v01-n01-7
Human-w04-v01-n01-7
Human-w01-v01-n01-7
Human-w02-v01-n01-7
Human-w01-v01-n01-7
Human-w02-v01-n01-7
Human-w03-v01-n01-7
Human-w04-v01-n01-7
Human-w05-v01-n01-7
0.05
9
1
8
6
7

B.

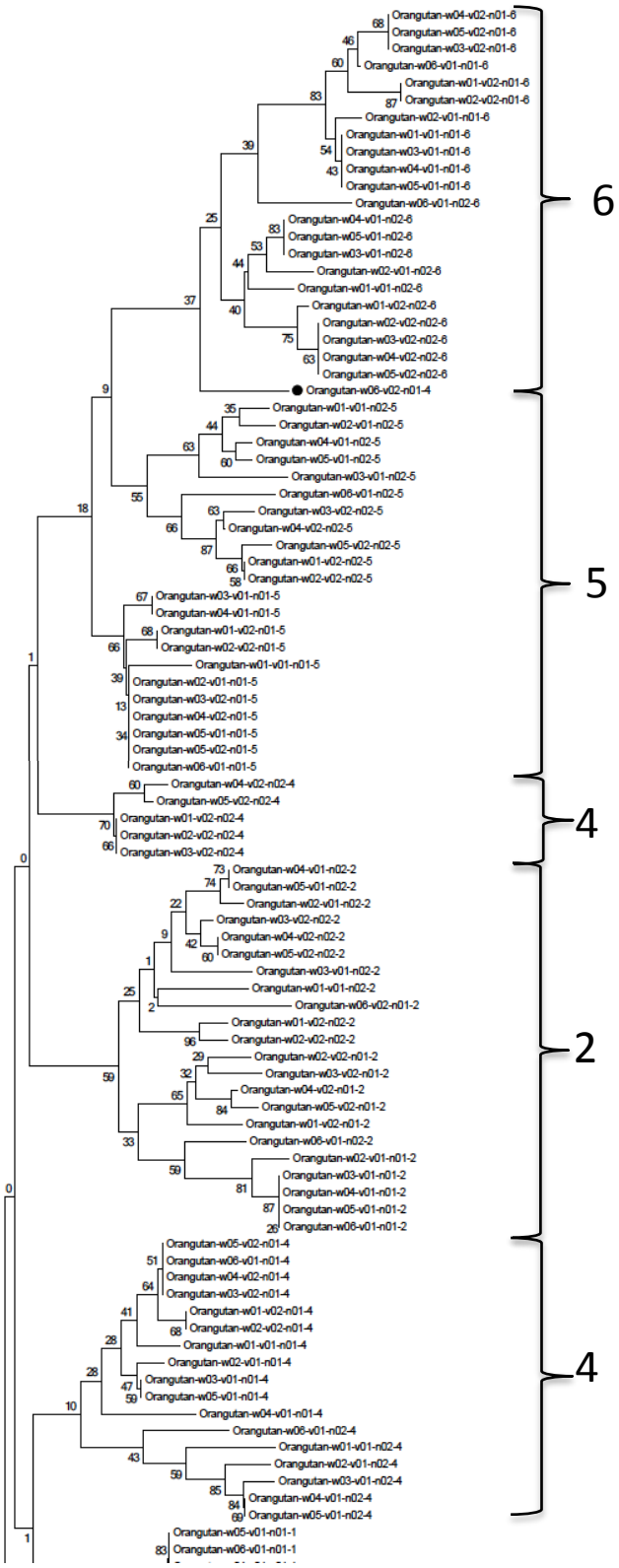

Chimpanzee-w05-v01-n01-3
Chimpanzee-w06-v01-n02-3
Chimpanzee-w03-v02-n01-3
Chimpanzee-w02-v01-n01-3
Chimpanzee-w04-v01-n01-3
Chimpanzee-w05-v01-n02-3
Chimpanzee-w04-v01-n02-3
Chimpanzee-w01-v01-n02-3
Chimpanzee-w02-v01-n02-3
Chimpanzee-w05-v02-n02-3
Chimpanzee-w06-v02-n01-3
Chimpanzee-w04-v02-n02-3
Chimpanzee-w03-v01-n02-3
Chimpanzee-w02-v02-n02-3
Chimpanzee-w06-v01-n01-3
Chimpanzee-w01-v02-n02-3
Chimpanzee-w04-v02-n01-3
Chimpanzee-w01-v01-n01-3
Chimpanzee-w03-v01-n01-3
Chimpanzee-w05-v02-n01-3
Chimpanzee-w01-v02-n01-3
Chimpanzee-w02-v02-n01-3
Chimpanzee-w04-v02-n02-2
Chimpanzee-w06-v01-n01-2
Chimpanzee-w02-v02-n02-2
Chimpanzee-w03-v01-n02-2
Chimpanzee-w01-v02-n02-2
Chimpanzee-w03-v02-n02-2
Chimpanzee-w05-v02-n02-2
Chimpanzee-w01-v01-n02-2
Chimpanzee-w04-v01-n02-2
Chimpanzee-w02-v01-n02-2
Chimpanzee-w05-v01-n02-2
Chimpanzee-w06-v02-n01-2
Chimpanzee-w02-v02-n01-2
Chimpanzee-w05-v02-n01-2
Chimpanzee-w04-v02-n01-2
Chimpanzee-w03-v01-n01-2
Chimpanzee-w01-v02-n01-2
Chimpanzee-w06-v01-n02-2
Chimpanzee-w02-v01-n01-2
Chimpanzee-w03-v02-n01-2
Chimpanzee-w04-v01-n01-2
Chimpanzee-w05-v01-n01-2
Chimpanzee-w04-v02-n02-4
Chimpanzee-w06-v01-n01-4
Chimpanzee-w02-v02-n02-4
Chimpanzee-w03-v02-n02-4
Chimpanzee-w01-v02-n02-4
Chimpanzee-w03-v01-n02-4
Chimpanzee-w05-v02-n02-4
Chimpanzee-w06-v01-n02-4
Chimpanzee-w01-v01-n02-4
Chimpanzee-w05-v01-n02-4
Chimpanzee-w02-v01-n02-4
Chimpanzee-w04-v01-n02-4
Chimpanzee-w02-v01-n01-5
Chimpanzee-w02-v02-n01-5
Chimpanzee-w04-v02-n01-5
Chimpanzee-w04-v01-n01-5
Chimpanzee-w05-v02-n01-5
Chimpanzee-w01-v01-n01-5
Chimpanzee-w03-v02-n01-5
Chimpanzee-w05-v01-n01-5
Chimpanzee-w03-v01-n01-5
Chimpanzee-w02-v01-n02-5
Chimpanzee-w05-v01-n02-5
Chimpanzee-w04-v01-n02-5
Chimpanzee-w01-v01-n02-5
Chimpanzee-w01-v02-n01-5
Chimpanzee-w06-v01-n02-5
Chimpanzee-w03-v02-n02-5
Chimpanzee-w05-v02-n02-5
Chimpanzee-w03-v01-n02-5
Chimpanzee-w01-v02-n02-5
Chimpanzee-w02-v02-n02-5
Chimpanzee-w04-v02-n02-5
Chimpanzee-w06-v01-n01-5
Chimpanzee-w01-v01-n01-4
Chimpanzee-w05-v02-n01-4
Chimpanzee-w02-v02-n01-4
Chimpanzee-w04-v02-n01-4
Chimpanzee-w03-v01-n01-4
Chimpanzee-w01-v02-n01-4
Chimpanzee-w03-v02-n01-4
Chimpanzee-w05-v01-n01-4
Chimpanzee-w02-v01-n01-4
Chimpanzee-w04-v01-n01-4
Chimpanzee-w02-v01-n02-9
Chimpanzee-w04-v01-n02-9
Chimpanzee-w05-v01-n02-9
Chimpanzee-w01-v01-n01-1
Chimpanzee-w01-v01-n02-9
Chimpanzee-w04-v02-n02-9
Chimpanzee-w06-v01-n01-9
Chimpanzee-w03-v02-n02-9
Chimpanzee-w02-v02-n02-9
Chimpanzee-w01-v02-n02-9
Chimpanzee-w03-v01-n02-9
Chimpanzee-w05-v02-n02-9
Chimpanzee-w06-v01-n02-9
Chimpanzee-w02-v01-n01-9
Chimpanzee-w05-v01-n01-9
Chimpanzee-w02-v02-n01-9
Chimpanzee-w04-v02-n01-9
Chimpanzee-w05-v02-n01-9
Chimpanzee-w01-v01-n01-9
Chimpanzee-w03-v01-n01-9
Chimpanzee-w01-v02-n01-9
Chimpanzee-w03-v02-n01-9
Chimpanzee-w04-v01-n01-9
3
2
4
5
4
9

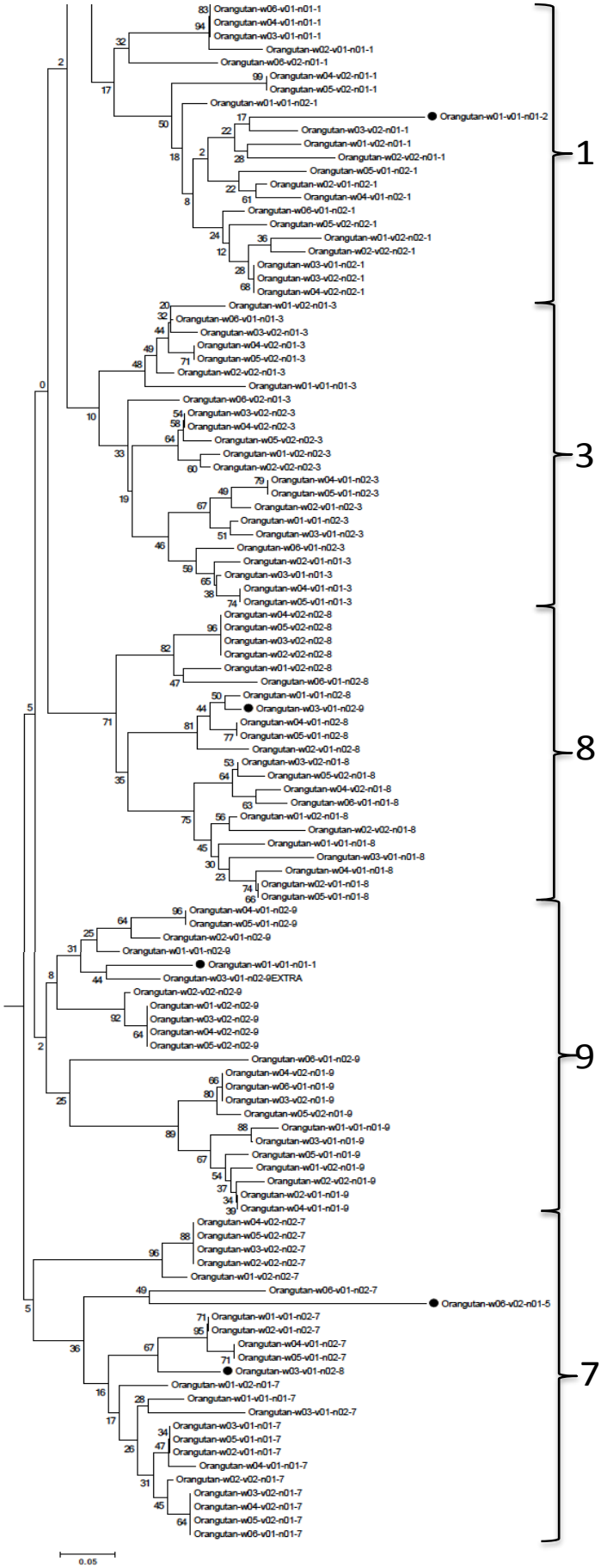

Chimpanzee-w04-v01-n01-9
Chimpanzee-w01-v01-n01-8
Chimpanzee-w05-v02-n01-8
Chimpanzee-w03-v02-n01-8
Chimpanzee-w03-v01-n01-8
Chimpanzee-w02-v01-n01-8
Chimpanzee-w05-v01-n01-8
Chimpanzee-w02-v02-n01-8
Chimpanzee-w04-v02-n01-8
Chimpanzee-w01-v02-n01-8
Chimpanzee-w04-v01-n01-8
Chimpanzee-w01-v01-n02-8
Chimpanzee-w05-v01-n02-8
Chimpanzee-w02-v01-n02-8
Chimpanzee-w04-v01-n02-8
Chimpanzee-w06-v01-n02-8
Chimpanzee-w05-v02-n02-8
Chimpanzee-w03-v02-n02-8
Chimpanzee-w03-v01-n02-8
Chimpanzee-w01-v02-n02-8
Chimpanzee-w02-v02-n02-8
Chimpanzee-w04-v02-n02-8
Chimpanzee-w06-v01-n01-8
Chimpanzee-w06-v02-n01-4
Chimpanzee-w01-v01-n02-6
Chimpanzee-w04-v01-n02-6
Chimpanzee-w02-v01-n02-6
Chimpanzee-w05-v01-n02-6
Chimpanzee-w04-v02-n02-6
Chimpanzee-w06-v01-n01-6
Chimpanzee-w02-v02-n02-6
Chimpanzee-w01-v02-n02-6
Chimpanzee-w03-v01-n02-6
Chimpanzee-w03-v02-n02-6
Chimpanzee-w05-v02-n02-6
Chimpanzee-w06-v01-n02-6
Chimpanzee-w02-v02-n01-6
Chimpanzee-w04-v02-n01-6
Chimpanzee-w02-v01-n01-6
Chimpanzee-w01-v01-n01-6
Chimpanzee-w01-v02-n01-6
Chimpanzee-w04-v01-n01-6
Chimpanzee-w05-v02-n01-6
Chimpanzee-w03-v01-n01-6
Chimpanzee-w03-v02-n01-6
Chimpanzee-w05-v01-n01-6
Chimpanzee-w05-v01-n01-1
Chimpanzee-w06-v01-n02-1
Chimpanzee-w04-v01-n01-1
Chimpanzee-w03-v01-n01-1
Chimpanzee-w03-v02-n01-1
Chimpanzee-w02-v01-n01-1
Chimpanzee-w06-v02-n01-1
Chimpanzee-w04-v01-n02-1
Chimpanzee-w03-v02-n02-1
Chimpanzee-w05-v02-n02-1
Chimpanzee-w01-v01-n02-1
Chimpanzee-w04-v02-n01-1
Chimpanzee-w05-v02-n01-1
Chimpanzee-w02-v02-n01-1
Chimpanzee-w01-v02-n01-1
Chimpanzee-w01-v01-n01-2
Chimpanzee-w02-v01-n02-1
Chimpanzee-w05-v01-n02-1
Chimpanzee-w03-v01-n02-1
Chimpanzee-w01-v02-n02-1
Chimpanzee-w04-v02-n02-1
Chimpanzee-w02-v02-n02-1
Chimpanzee-w06-v01-n01-1
Chimpanzee-w03-v01-n02-7
Chimpanzee-w03-v02-n02-7
Chimpanzee-w01-v02-n02-7
Chimpanzee-w05-v02-n02-7
Chimpanzee-w02-v02-n02-7
Chimpanzee-w04-v02-n02-7
Chimpanzee-w06-v01-n01-7
Chimpanzee-w01-v01-n02-7
Chimpanzee-w02-v01-n02-7
Chimpanzee-w05-v01-n02-7
Chimpanzee-w04-v01-n02-7
Chimpanzee-w06-v01-n02-7
Chimpanzee-w06-v02-n01-5
Chimpanzee-w05-v01-n01-7
Chimpanzee-w02-v02-n01-7
Chimpanzee-w04-v01-n01-7
Chimpanzee-w03-v01-n01-7
Chimpanzee-w03-v02-n01-7
Chimpanzee-w05-v02-n01-7
Chimpanzee-w01-v01-n01-7
Chimpanzee-w01-v02-n01-7
Chimpanzee-w02-v01-n01-7
Chimpanzee-w04-v02-n01-7
0.05
9
8
6
1
7

C.

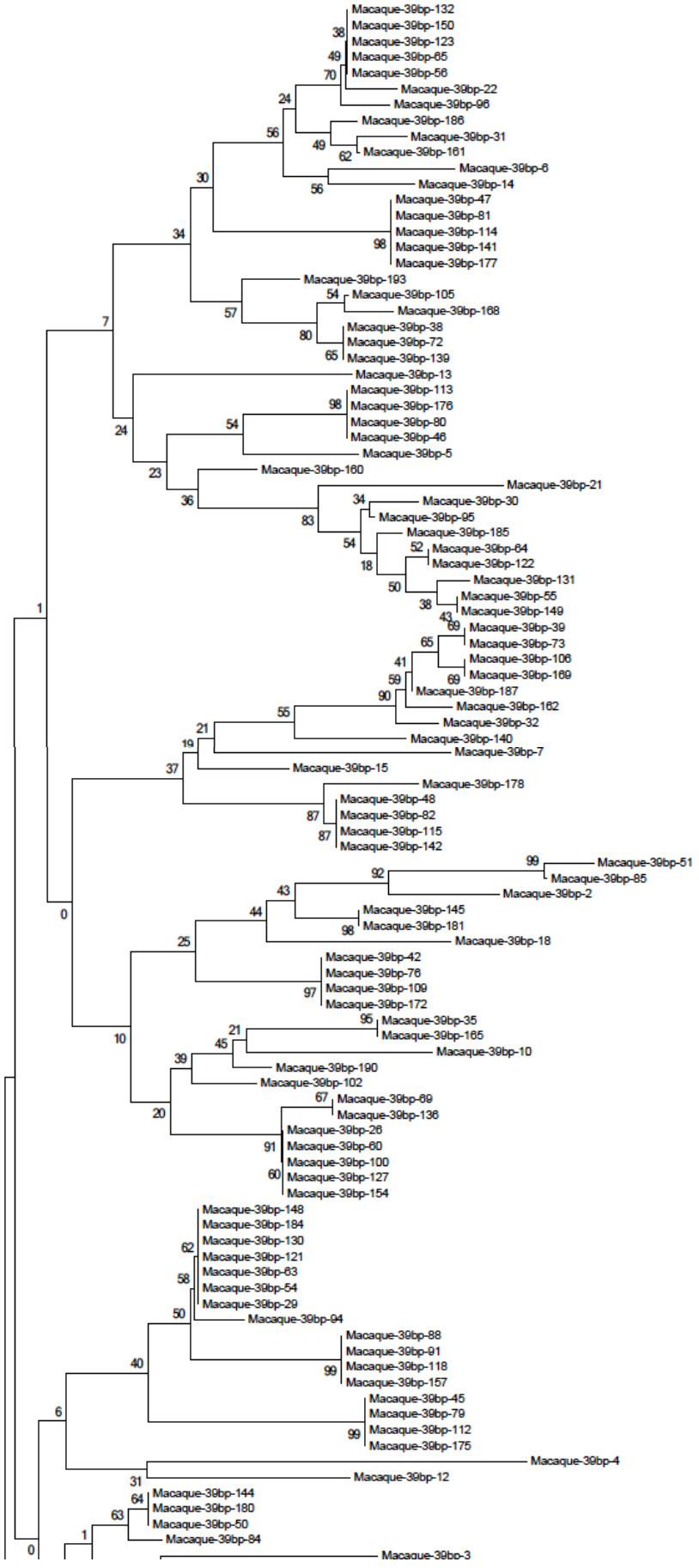

Gorilla-w01-v02-n02-3
Gorilla-w02-v02-n02-3
Gorilla-w01-v01-n02-3
Gorilla-w03-v01-n02-3
Gorilla-w04-v02-n01-6
Gorilla-w02-v02-n01-3
Gorilla-w02-v01-n02-3
Gorilla-w04-v01-n01-3
Gorilla-w01-v01-n01-3
Gorilla-w03-v02-n02-3
Gorilla-w03-v01-n01-3
Gorilla-w04-v01-n02-3
Gorilla-w01-v02-n01-3
Gorilla-w02-v01-n01-3
Gorilla-w03-v02-n01-3
Gorilla-w01-v02-n02-4
Gorilla-w03-v01-n02-4
Gorilla-w02-v02-n02-4
Gorilla-w01-v01-n02-4
Gorilla-w02-v01-n02-4
Gorilla-w04-v01-n01-4
Gorilla-w01-v01-n01-4
Gorilla-w02-v01-n01-4
Gorilla-w03-v02-n01-4
Gorilla-w02-v02-n01-4
Gorilla-w01-v02-n01-4
Gorilla-w03-v02-n02-4
Gorilla-w04-v01-n02-4
Gorilla-w02-v01-n01-2
Gorilla-w03-v02-n01-2
Gorilla-w03-v01-n01-2
Gorilla-w04-v01-n02-2
Gorilla-w01-v02-n01-2
Gorilla-w04-v01-n01-2
Gorilla-w04-v02-n01-5
Gorilla-w02-v02-n01-2
Gorilla-w03-v01-n02-2
Gorilla-w02-v01-n02-2
Gorilla-w03-v02-n02-2
Gorilla-w01-v01-n02-2
Gorilla-w01-v02-n02-2
Gorilla-w02-v02-n02-2
Gorilla-w01-v02-n02-8
Gorilla-w03-v01-n02-8
Gorilla-w01-v01-n02-8
Gorilla-w02-v02-n02-8
Gorilla-w02-v01-n02-8
Gorilla-w04-v01-n01-8
Gorilla-w04-v01-n02-8
Gorilla-w04-v02-n01-2
Gorilla-w02-v02-n01-8
Gorilla-w03-v02-n01-8
Gorilla-w03-v02-n02-8
Gorilla-w01-v01-n01-8
Gorilla-w01-v02-n01-8
Gorilla-w02-v01-n01-8
Gorilla-w03-v01-n01-6
Gorilla-w01-v02-n02-9
Gorilla-w03-v01-n02-9
Gorilla-w01-v01-n02-9
Gorilla-w01-v01-n01-1
Gorilla-w02-v02-n02-9
Gorilla-w02-v01-n02-9
Gorilla-w04-v01-n01-9
Gorilla-w04-v01-n02-9
Gorilla-w04-v02-n01-3
Gorilla-w03-v01-n01-7
Gorilla-w02-v01-n01-9
Gorilla-w03-v02-n01-9
Gorilla-w03-v02-n02-9
Gorilla-w01-v01-n01-9
Gorilla-w01-v02-n01-9
Gorilla-w02-v02-n01-9
Gorilla-w01-v02-n02-5
Gorilla-w03-v01-n02-5
Gorilla-w02-v02-n02-5
Gorilla-w01-v01-n02-5
Gorilla-w01-v02-n01-5
Gorilla-w04-v01-n02-5
Gorilla-w02-v01-n02-5
Gorilla-w04-v01-n01-5
Gorilla-w01-v01-n01-5
Gorilla-w02-v02-n01-5
Gorilla-w02-v01-n01-5
Gorilla-w03-v02-n01-5
Gorilla-w03-v02-n02-5
Gorilla-w03-v01-n01-4
Gorilla-w04-v02-n01-7
Gorilla-w04-v01-n02-6
Gorilla-w02-v01-n01-6
Gorilla-w03-v02-n01-6
Gorilla-w03-v02-n02-6
Gorilla-w01-v02-n01-6
Gorilla-w01-v01-n01-6
Gorilla-w02-v02-n01-6
Gorilla-w02-v01-n02-6
Gorilla-w04-v01-n01-6
Gorilla-w02-v02-n02-6
Gorilla-w01-v01-n02-6
Gorilla-w01-v02-n02-6
Gorilla-w03-v01-n02-6
Gorilla-w02-v01-n02-7
Gorilla-w04-v01-n01-7
Gorilla-w03-v01-n01-5
Gorilla-w01-v01-n01-7
Gorilla-w03-v02-n02-7
Gorilla-w02-v01-n01-7
Gorilla-w03-v02-n01-7
Gorilla-w01-v02-n01-7
Gorilla-w02-v02-n01-7
Gorilla-w01-v01-n02-7
Gorilla-w02-v02-n02-7
Gorilla-w01-v02-n02-7
Gorilla-w03-v01-n02-7
Gorilla-w04-v01-n02-7
Gorilla-w04-v02-n01-1
Gorilla-w02-v02-n01-1
Gorilla-w03-v01-n01-1
Gorilla-w04-v01-n02-1
Gorilla-w04-v02-n01-4
Gorilla-w02-v01-n02-1
Gorilla-w03-v02-n02-1
Gorilla-w01-v02-n02-1
Gorilla-w02-v02-n02-1
Gorilla-w04-v01-n01-1
Gorilla-w01-v01-n02-1
Gorilla-w01-v01-n01-2
Gorilla-w03-v01-n02-1
Gorilla-w01-v02-n01-1
Gorilla-w02-v01-n01-1
Gorilla-w03-v02-n01-1
3
4
2
8
9
5
6
7
0.05

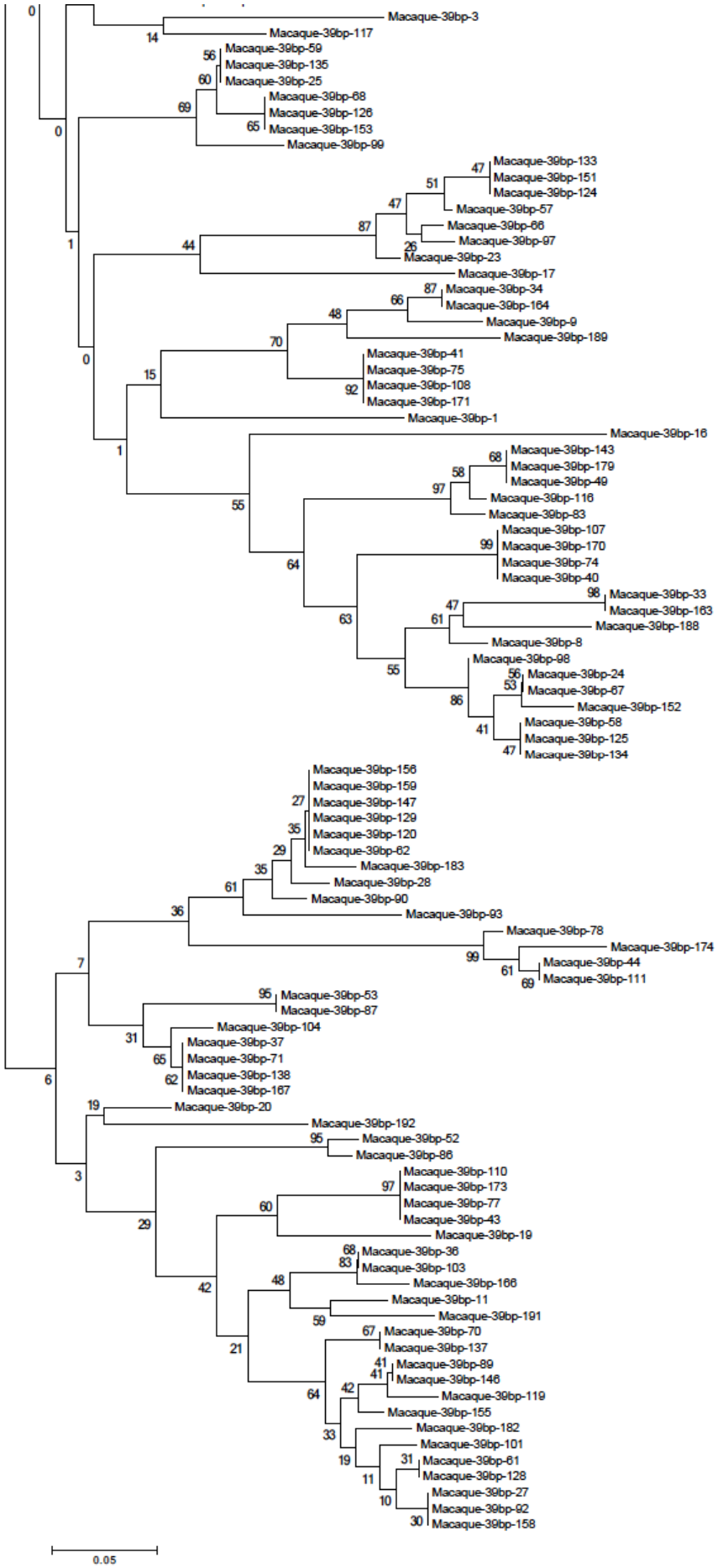
D.


Orangutan-w04-v02-n01-6
Orangutan-w05-v02-n01-6
Orangutan-w03-v02-n01-6
Orangutan-w06-v01-n01-6
Orangutan-w01-v02-n01-6
Orangutan-w02-v02-n01-6
Orangutan-w02-v01-n01-6
Orangutan-w01-v01-n01-6
Orangutan-w03-v01-n01-6
Orangutan-w04-v01-n01-6
Orangutan-w05-v01-n01-6
Orangutan-w06-v01-n02-6
Orangutan-w04-v01-n02-6
Orangutan-w05-v01-n02-6
Orangutan-w03-v01-n02-6
Orangutan-w02-v01-n02-6
Orangutan-w01-v01-n02-6
Orangutan-w01-v02-n02-6
Orangutan-w02-v02-n02-6
Orangutan-w03-v02-n02-6
Orangutan-w04-v02-n02-6
Orangutan-w05-v02-n02-6
Orangutan-w06-v02-n01-4
Orangutan-w01-v01-n02-5
Orangutan-w02-v01-n02-5
Orangutan-w04-v01-n02-5
Orangutan-w05-v01-n02-5
Orangutan-w03-v01-n02-5
Orangutan-w06-v01-n02-5
Orangutan-w03-v02-n02-5
Orangutan-w04-v02-n02-5
Orangutan-w05-v02-n02-5
Orangutan-w01-v02-n02-5
Orangutan-w02-v02-n02-5
Orangutan-w03-v01-n01-5
Orangutan-w04-v01-n01-5
Orangutan-w01-v02-n01-5
Orangutan-w02-v02-n01-5
Orangutan-w01-v01-n01-5
Orangutan-w02-v01-n01-5
Orangutan-w03-v02-n01-5
Orangutan-w04-v02-n01-5
Orangutan-w05-v01-n01-5
Orangutan-w05-v02-n01-5
Orangutan-w06-v01-n01-5
Orangutan-w04-v02-n02-4
Orangutan-w05-v02-n02-4
Orangutan-w01-v02-n02-4
Orangutan-w02-v02-n02-4
Orangutan-w03-v02-n02-4
Orangutan-w04-v01-n02-2
Orangutan-w05-v01-n02-2
Orangutan-w02-v01-n02-2
Orangutan-w03-v02-n02-2
Orangutan-w04-v02-n02-2
Orangutan-w05-v02-n02-2
Orangutan-w03-v01-n02-2
Orangutan-w01-v01-n02-2
Orangutan-w06-v02-n01-2
Orangutan-w01-v02-n02-2
Orangutan-w02-v02-n02-2
Orangutan-w02-v02-n01-2
Orangutan-w03-v02-n01-2
Orangutan-w04-v02-n01-2
Orangutan-w05-v02-n01-2
Orangutan-w01-v02-n01-2
Orangutan-w06-v01-n02-2
Orangutan-w02-v01-n01-2
Orangutan-w03-v01-n01-2
Orangutan-w04-v01-n01-2
Orangutan-w05-v01-n01-2
Orangutan-w06-v01-n01-2
Orangutan-w05-v02-n01-4
Orangutan-w06-v01-n01-4
Orangutan-w04-v02-n01-4
Orangutan-w03-v02-n01-4
Orangutan-w01-v02-n01-4
Orangutan-w02-v02-n01-4
Orangutan-w01-v01-n01-4
Orangutan-w02-v01-n01-4
Orangutan-w03-v01-n01-4
Orangutan-w05-v01-n01-4
Orangutan-w04-v01-n01-4
Orangutan-w06-v01-n02-4
Orangutan-w01-v01-n02-4
Orangutan-w02-v01-n02-4
Orangutan-w03-v01-n02-4
Orangutan-w04-v01-n02-4
Orangutan-w05-v01-n02-4
Orangutan-w05-v01-n01-1
Orangutan-w06-v01-n01-1
6
5
4
2
4

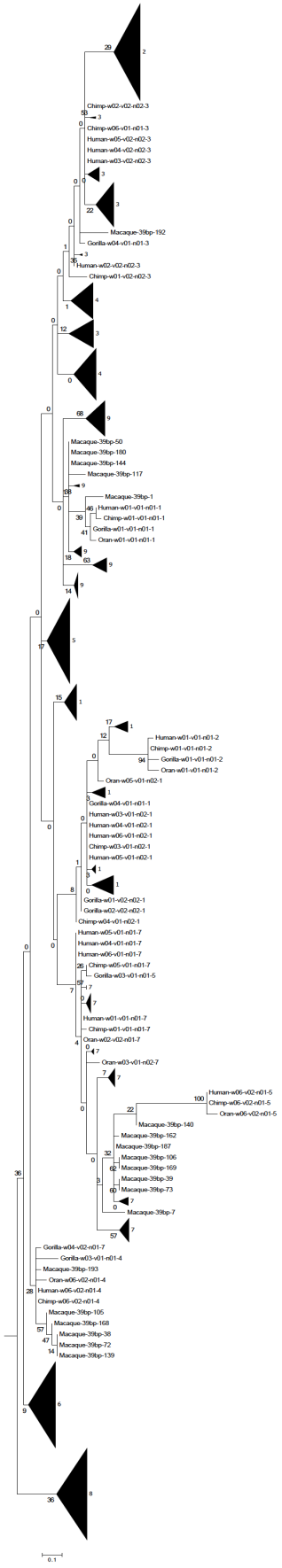

Orangutan-w06-v01-n01-1
Orangutan-w04-v01-n01-1
Orangutan-w03-v01-n01-1
Orangutan-w02-v01-n01-1
Orangutan-w06-v02-n01-1
Orangutan-w04-v02-n01-1
Orangutan-w05-v02-n01-1
Orangutan-w01-v01-n02-1
Orangutan-w01-v01-n01-2
Orangutan-w03-v02-n01-1
Orangutan-w01-v02-n01-1
Orangutan-w02-v02-n01-1
Orangutan-w05-v01-n02-1
Orangutan-w02-v01-n02-1
Orangutan-w04-v01-n02-1
Orangutan-w06-v01-n02-1
Orangutan-w05-v02-n02-1
Orangutan-w01-v02-n02-1
Orangutan-w02-v02-n02-1
Orangutan-w03-v01-n02-1
Orangutan-w03-v02-n02-1
Orangutan-w04-v02-n02-1
Orangutan-w01-v02-n01-3
Orangutan-w06-v01-n01-3
Orangutan-w03-v02-n01-3
Orangutan-w04-v02-n01-3
Orangutan-w05-v02-n01-3
Orangutan-w02-v02-n01-3
Orangutan-w01-v01-n01-3
Orangutan-w06-v02-n01-3
Orangutan-w03-v02-n02-3
Orangutan-w04-v02-n02-3
Orangutan-w05-v02-n02-3
Orangutan-w01-v02-n02-3
Orangutan-w02-v02-n02-3
Orangutan-w04-v01-n02-3
Orangutan-w05-v01-n02-3
Orangutan-w02-v01-n02-3
Orangutan-w01-v01-n02-3
Orangutan-w03-v01-n02-3
Orangutan-w06-v01-n02-3
Orangutan-w02-v01-n01-3
Orangutan-w03-v01-n01-3
Orangutan-w04-v01-n01-3
Orangutan-w05-v01-n01-3
Orangutan-w04-v02-n02-8
Orangutan-w05-v02-n02-8
Orangutan-w03-v02-n02-8
Orangutan-w02-v02-n02-8
Orangutan-w01-v02-n02-8
Orangutan-w06-v01-n02-8
Orangutan-w01-v01-n02-8
Orangutan-w03-v01-n02-9
Orangutan-w04-v01-n02-8
Orangutan-w05-v01-n02-8
Orangutan-w02-v01-n02-8
Orangutan-w03-v02-n01-8
Orangutan-w05-v02-n01-8
Orangutan-w04-v02-n01-8
Orangutan-w06-v01-n01-8
Orangutan-w01-v02-n01-8
Orangutan-w02-v02-n01-8
Orangutan-w01-v01-n01-8
Orangutan-w03-v01-n01-8
Orangutan-w04-v01-n01-8
Orangutan-w02-v01-n01-8
Orangutan-w05-v01-n01-8
Orangutan-w04-v01-n02-9
Orangutan-w05-v01-n02-9
Orangutan-w02-v01-n02-9
Orangutan-w01-v01-n02-9
Orangutan-w01-v01-n01-1
Orangutan-w03-v01-n02-9EXTRA
Orangutan-w02-v02-n02-9
Orangutan-w01-v02-n02-9
Orangutan-w03-v02-n02-9
Orangutan-w04-v02-n02-9
Orangutan-w05-v02-n02-9
Orangutan-w06-v01-n02-9
Orangutan-w04-v02-n01-9
Orangutan-w06-v01-n01-9
Orangutan-w03-v02-n01-9
Orangutan-w05-v02-n01-9
Orangutan-w01-v01-n01-9
Orangutan-w03-v01-n01-9
Orangutan-w05-v01-n01-9
Orangutan-w01-v02-n01-9
Orangutan-w02-v02-n01-9
Orangutan-w02-v01-n01-9
Orangutan-w04-v01-n01-9
Orangutan-w04-v02-n02-7
Orangutan-w05-v02-n02-7
Orangutan-w03-v02-n02-7
Orangutan-w02-v02-n02-7
Orangutan-w01-v02-n02-7
Orangutan-w06-v01-n02-7
Orangutan-w06-v02-n01-5
Orangutan-w01-v01-n02-7
Orangutan-w02-v01-n02-7
Orangutan-w04-v01-n02-7
Orangutan-w05-v01-n02-7
Orangutan-w03-v01-n02-8
Orangutan-w01-v02-n01-7
Orangutan-w01-v01-n01-7
Orangutan-w03-v01-n02-7
Orangutan-w03-v01-n01-7
Orangutan-w05-v01-n01-7
Orangutan-w02-v01-n01-7
Orangutan-w04-v01-n01-7
Orangutan-w02-v02-n01-7
Orangutan-w03-v02-n01-7
Orangutan-w04-v02-n01-7
Orangutan-w05-v02-n01-7
Orangutan-w06-v01-n01-7
1
3
8
9
7
0.05

E.

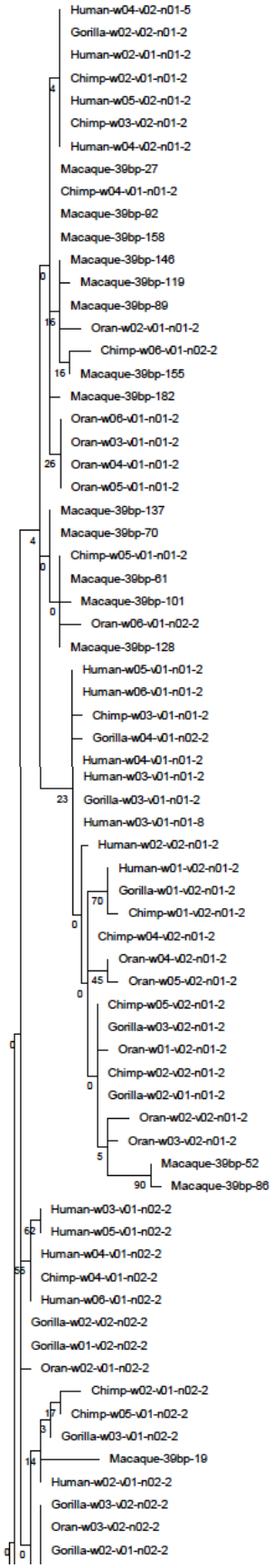

Macaque-39bp-132
Macaque-39bp-150
Macaque-39bp-123
Macaque-39bp-65
Macaque-39bp-56
Macaque-39bp-22
Macaque-39bp-96
Macaque-39bp-186
Macaque-39bp-31
Macaque-39bp-161
Macaque-39bp-6
Macaque-39bp-14
Macaque-39bp-47
Macaque-39bp-81
Macaque-39bp-114
Macaque-39bp-141
Macaque-39bp-177
Macaque-39bp-193
Macaque-39bp-105
Macaque-39bp-168
Macaque-39bp-38
Macaque-39bp-72
Macaque-39bp-139
Macaque-39bp-13
Macaque-39bp-113
Macaque-39bp-176
Macaque-39bp-80
Macaque-39bp-46
Macaque-39bp-5
Macaque-39bp-160
Macaque-39bp-21
Macaque-39bp-30
Macaque-39bp-95
Macaque-39bp-185
Macaque-39bp-64
Macaque-39bp-122
Macaque-39bp-131
Macaque-39bp-55
Macaque-39bp-149
Macaque-39bp-39
Macaque-39bp-73
Macaque-39bp-106
Macaque-39bp-169
Macaque-39bp-187
Macaque-39bp-162
Macaque-39bp-32
Macaque-39bp-140
Macaque-39bp-7
Macaque-39bp-15
Macaque-39bp-178
Macaque-39bp-48
Macaque-39bp-82
Macaque-39bp-115
Macaque-39bp-142
Macaque-39bp-51
Macaque-39bp-85
Macaque-39bp-2
Macaque-39bp-145
Macaque-39bp-181
Macaque-39bp-18
Macaque-39bp-42
Macaque-39bp-76
Macaque-39bp-109
Macaque-39bp-172
Macaque-39bp-35
Macaque-39bp-165
Macaque-39bp-10
Macaque-39bp-190
Macaque-39bp-102
Macaque-39bp-69
Macaque-39bp-136
Macaque-39bp-26
Macaque-39bp-60
Macaque-39bp-100
Macaque-39bp-127
Macaque-39bp-154
Macaque-39bp-148
Macaque-39bp-184
Macaque-39bp-130
Macaque-39bp-121
Macaque-39bp-63
Macaque-39bp-54
Macaque-39bp-29
Macaque-39bp-94
Macaque-39bp-88
Macaque-39bp-91
Macaque-39bp-118
Macaque-39bp-157
Macaque-39bp-45
Macaque-39bp-79
Macaque-39bp-112
Macaque-39bp-175
Macaque-39bp-4
Macaque-39bp-12
Macaque-39bp-144
Macaque-39bp-180
Macaque-39bp-50
Macaque-39bp-84
Macaque-39bp-3

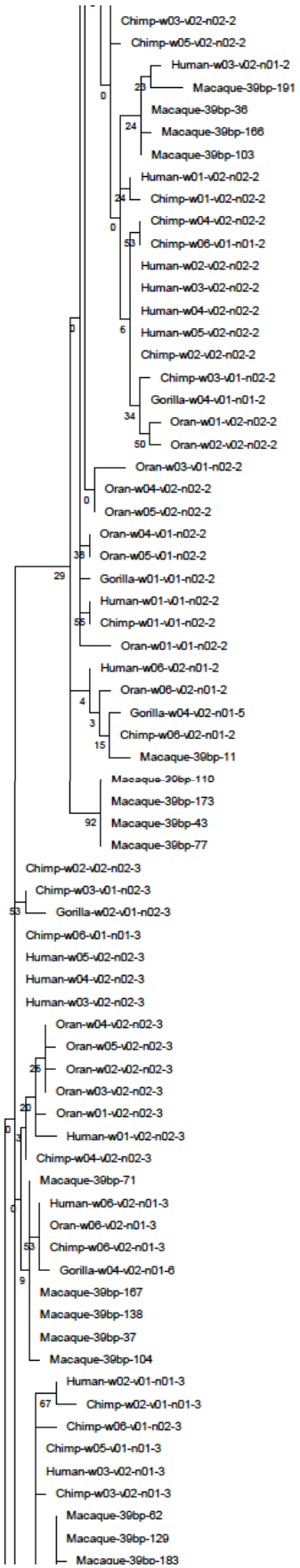

Macaque-39bp-3
Macaque-39bp-117
Macaque-39bp-59
Macaque-39bp-135
Macaque-39bp-25
Macaque-39bp-68
Macaque-39bp-126
Macaque-39bp-153
Macaque-39bp-99
Macaque-39bp-133
Macaque-39bp-151
Macaque-39bp-124
Macaque-39bp-57
Macaque-39bp-66
Macaque-39bp-97
Macaque-39bp-23
Macaque-39bp-17
Macaque-39bp-34
Macaque-39bp-164
Macaque-39bp-9
Macaque-39bp-189
Macaque-39bp-41
Macaque-39bp-75
Macaque-39bp-108
Macaque-39bp-171
Macaque-39bp-1
Macaque-39bp-16
Macaque-39bp-143
Macaque-39bp-179
Macaque-39bp-49
Macaque-39bp-116
Macaque-39bp-83
Macaque-39bp-107
Macaque-39bp-170
Macaque-39bp-74
Macaque-39bp-40
Macaque-39bp-33
Macaque-39bp-163
Macaque-39bp-188
Macaque-39bp-8
Macaque-39bp-98
Macaque-39bp-24
Macaque-39bp-67
Macaque-39bp-152
Macaque-39bp-58
Macaque-39bp-125
Macaque-39bp-134
Macaque-39bp-156
Macaque-39bp-159
Macaque-39bp-147
Macaque-39bp-129
Macaque-39bp-120
Macaque-39bp-62
Macaque-39bp-183
Macaque-39bp-28
Macaque-39bp-90
Macaque-39bp-93
Macaque-39bp-78
Macaque-39bp-174
Macaque-39bp-44
Macaque-39bp-111
Macaque-39bp-53
Macaque-39bp-87
Macaque-39bp-104
Macaque-39bp-37
Macaque-39bp-71
Macaque-39bp-138
Macaque-39bp-167
Macaque-39bp-20
Macaque-39bp-192
Macaque-39bp-52
Macaque-39bp-86
Macaque-39bp-110
Macaque-39bp-173
Macaque-39bp-77
Macaque-39bp-43
Macaque-39bp-19
Macaque-39bp-36
Macaque-39bp-103
Macaque-39bp-166
Macaque-39bp-11
Macaque-39bp-191
Macaque-39bp-70
Macaque-39bp-137
Macaque-39bp-89
Macaque-39bp-146
Macaque-39bp-119
Macaque-39bp-155
Macaque-39bp-182
Macaque-39bp-101
Macaque-39bp-61
Macaque-39bp-128
Macaque-39bp-27
Macaque-39bp-92
Macaque-39bp-158
0.05

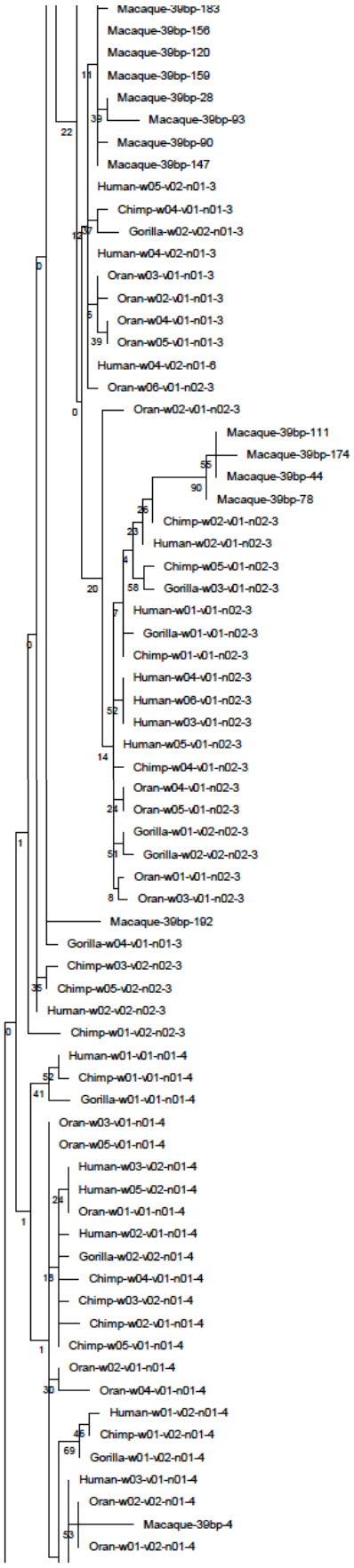
F.

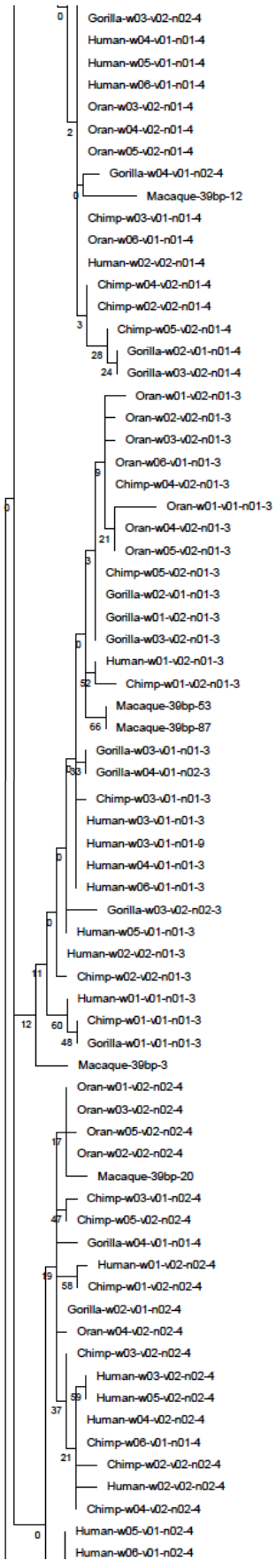

Chimp-w02-v02-n02-3
Chimp-w06-v01-n01-3
Human-w05-v02-n02-3
Human-w04-v02-n02-3
Human-w03-v02-n02-3
Macaque-39bp-192
Gorilla-w04-v01-n01-3
Human-w02-v02-n02-3
Chimp-w01-v02-n02-3
Macaque-39bp-50
Macaque-39bp-180
Macaque-39bp-144
Macaque-39bp-117
Macaque-39bp-1
Human-w01-v01-n01-1
Chimp-w01-v01-n01-1
Gorilla-w01-v01-n01-1
Oran-w01-v01-n01-1
Human-w01-v01-n01-2
Chimp-w01-v01-n01-2
Gorilla-w01-v01-n01-2
Oran-w01-v01-n01-2
Oran-w05-v01-n02-1
Gorilla-w04-v01-n01-1
Human-w03-v01-n02-1
Human-w04-v01-n02-1
Human-w06-v01-n02-1
Chimp-w03-v01-n02-1
Human-w05-v01-n02-1
Gorilla-w01-v02-n02-1
Gorilla-w02-v02-n02-1
Chimp-w04-v01-n02-1
Human-w05-v01-n01-7
Human-w04-v01-n01-7
Human-w06-v01-n01-7
Chimp-w05-v01-n01-7
Gorilla-w03-v01-n01-5
Human-w01-v01-n01-7
Chimp-w01-v01-n01-7
Oran-w02-v02-n01-7
Oran-w03-v01-n02-7
Human-w06-v02-n01-5
Chimp-w06-v02-n01-5
Oran-w06-v02-n01-5
Macaque-39bp-140
Macaque-39bp-162
Macaque-39bp-187
Macaque-39bp-106
Macaque-39bp-169
Macaque-39bp-39
Macaque-39bp-73
Macaque-39bp-7
Gorilla-w04-v02-n01-7
Gorilla-w03-v01-n01-4
Macaque-39bp-193
Oran-w06-v02-n01-4
Human-w06-v02-n01-4
Chimp-w06-v02-n01-4
Macaque-39bp-105
Macaque-39bp-168
Macaque-39bp-38
Macaque-39bp-72
Macaque-39bp-139
0.1

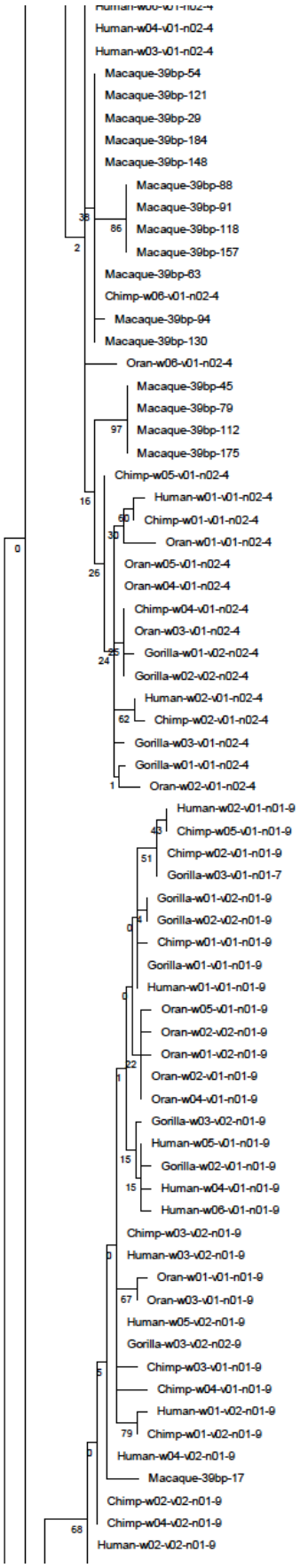
G.

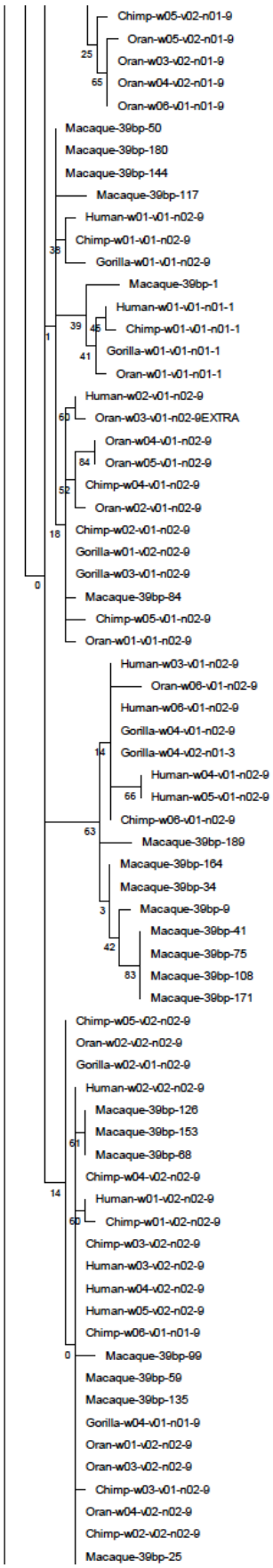

Human-w04-v02-n01-5
Gorilla-w02-v02-n01-2
Human-w02-v01-n01-2
Chimp-w02-v01-n01-2
Human-w05-v02-n01-2
Chimp-w03-v02-n01-2
Human-w04-v02-n01-2
Macaque-39bp-27
Chimp-w04-v01-n01-2
Macaque-39bp-92
Macaque-39bp-158
Macaque-39bp-146
Macaque-39bp-119
Macaque-39bp-89
Oran-w02-v01-n01-2
Chimp-w06-v01-n02-2
Macaque-39bp-155
Macaque-39bp-182
Oran-w06-v01-n01-2
Oran-w03-v01-n01-2
Oran-w04-v01-n01-2
Oran-w05-v01-n01-2
Macaque-39bp-137
Macaque-39bp-70
Chimp-w05-v01-n01-2
Macaque-39bp-61
Macaque-39bp-101
Oran-w06-v01-n02-2
Macaque-39bp-128
Human-w05-v01-n01-2
Human-w06-v01-n01-2
Chimp-w03-v01-n01-2
Gorilla-w04-v01-n02-2
Human-w04-v01-n01-2
Human-w03-v01-n01-2
Gorilla-w03-v01-n01-2
Human-w03-v01-n01-8
Human-w02-v02-n01-2
Human-w01-v02-n01-2
Gorilla-w01-v02-n01-2
Chimp-w01-v02-n01-2
Chimp-w04-v02-n01-2
Oran-w04-v02-n01-2
Oran-w05-v02-n01-2
Chimp-w05-v02-n01-2
Gorilla-w03-v02-n01-2
Oran-w01-v02-n01-2
Chimp-w02-v02-n01-2
Gorilla-w02-v01-n01-2
Oran-w02-v02-n01-2
Oran-w03-v02-n01-2
Macaque-39bp-52
Macaque-39bp-86
Human-w03-v01-n02-2
Human-w05-v01-n02-2
Human-w04-v01-n02-2
Chimp-w04-v01-n02-2
Human-w06-v01-n02-2
Gorilla-w02-v02-n02-2
Gorilla-w01-v02-n02-2
Oran-w02-v01-n02-2
Chimp-w02-v01-n02-2
Chimp-w05-v01-n02-2
Gorilla-w03-v01-n02-2
Macaque-39bp-19
Human-w02-v01-n02-2
Gorilla-w03-v02-n02-2
Oran-w03-v02-n02-2
Gorilla-w02-v01-n02-2

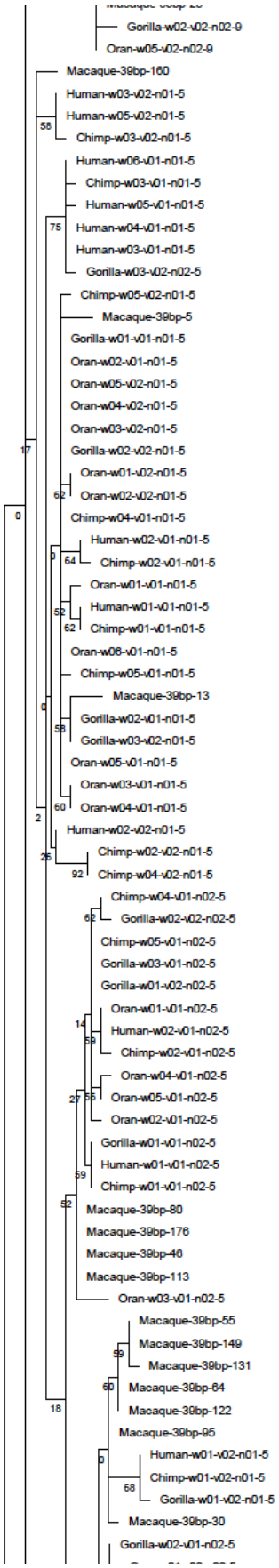

Chimp-w03-v02-n02-2
Chimp-w05-v02-n02-2
Human-w03-v02-n01-2
Macaque-39bp-191
Macaque-39bp-36
Macaque-39bp-166
Macaque-39bp-103
Human-w01-v02-n02-2
Chimp-w01-v02-n02-2
Chimp-w04-v02-n02-2
Chimp-w06-v01-n01-2
Human-w02-v02-n02-2
Human-w03-v02-n02-2
Human-w04-v02-n02-2
Human-w05-v02-n02-2
Chimp-w02-v02-n02-2
Chimp-w03-v01-n02-2
Gorilla-w04-v01-n01-2
Oran-w01-v02-n02-2
Oran-w02-v02-n02-2
Oran-w03-v01-n02-2
Oran-w04-v02-n02-2
Oran-w05-v02-n02-2
Oran-w04-v01-n02-2
Oran-w05-v01-n02-2
Gorilla-w01-v01-n02-2
Human-w01-v01-n02-2
Chimp-w01-v01-n02-2
Oran-w01-v01-n02-2
Human-w06-v02-n01-2
Oran-w06-v02-n01-2
Gorilla-w04-v02-n01-5
Chimp-w06-v02-n01-2
Macaque-39bp-11
Macaque-39bp-110
Macaque-39bp-173
Macaque-39bp-43
Macaque-39bp-77
Chimp-w02-v02-n02-3
Chimp-w03-v01-n02-3
Gorilla-w02-v01-n02-3
Chimp-w06-v01-n01-3
Human-w05-v02-n02-3
Human-w04-v02-n02-3
Human-w03-v02-n02-3
Oran-w04-v02-n02-3
Oran-w05-v02-n02-3
Oran-w02-v02-n02-3
Oran-w03-v02-n02-3
Oran-w01-v02-n02-3
Human-w01-v02-n02-3
Chimp-w04-v02-n02-3
Macaque-39bp-71
Human-w06-v02-n01-3
Oran-w06-v02-n01-3
Chimp-w06-v02-n01-3
Gorilla-w04-v02-n01-6
Macaque-39bp-167
Macaque-39bp-138
Macaque-39bp-37
Macaque-39bp-104
Human-w02-v01-n01-3
Chimp-w02-v01-n01-3
Chimp-w06-v01-n02-3
Chimp-w05-v01-n01-3
Human-w03-v02-n01-3
Chimp-w03-v02-n01-3
Macaque-39bp-62
Macaque-39bp-129
Macaque-39bp-183

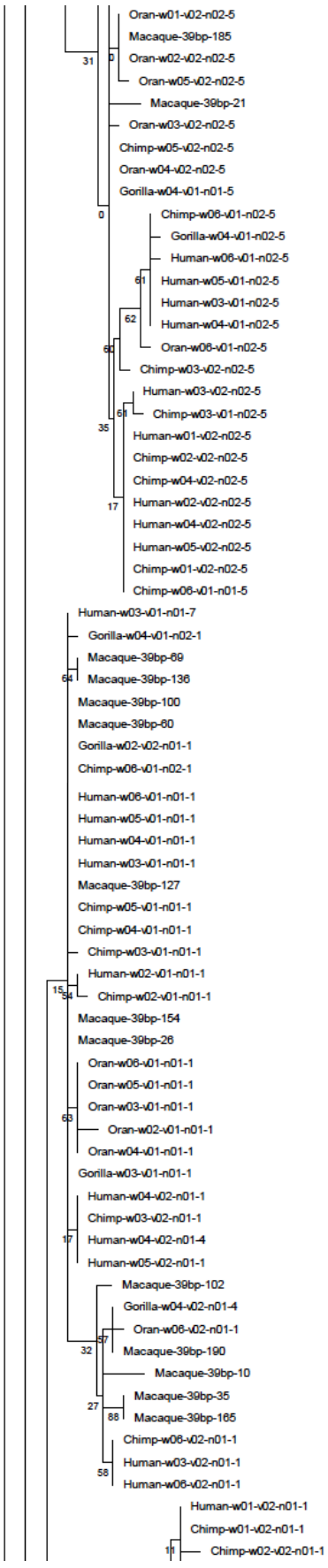

Macaque-39bp-183
Macaque-39bp-156
Macaque-39bp-120
Macaque-39bp-159
Macaque-39bp-28
Macaque-39bp-93
Macaque-39bp-90
Macaque-39bp-147
Human-w05-v02-n01-3
Chimp-w04-v01-n01-3
Gorilla-w02-v02-n01-3
Human-w04-v02-n01-3
Oran-w03-v01-n01-3
Oran-w02-v01-n01-3
Oran-w04-v01-n01-3
Oran-w05-v01-n01-3
Human-w04-v02-n01-6
Oran-w06-v01-n02-3
Oran-w02-v01-n02-3
Macaque-39bp-111
Macaque-39bp-174
Macaque-39bp-44
Macaque-39bp-78
Chimp-w02-v01-n02-3
Human-w02-v01-n02-3
Chimp-w05-v01-n02-3
Gorilla-w03-v01-n02-3
Human-w01-v01-n02-3
Gorilla-w01-v01-n02-3
Chimp-w01-v01-n02-3
Human-w04-v01-n02-3
Human-w06-v01-n02-3
Human-w03-v01-n02-3
Human-w05-v01-n02-3
Chimp-w04-v01-n02-3
Oran-w04-v01-n02-3
Oran-w05-v01-n02-3
Gorilla-w01-v02-n02-3
Gorilla-w02-v02-n02-3
Oran-w01-v01-n02-3
Oran-w03-v01-n02-3
Macaque-39bp-192
Gorilla-w04-v01-n01-3
Chimp-w03-v02-n02-3
Chimp-w05-v02-n02-3
Human-w02-v02-n02-3
Chimp-w01-v02-n02-3
Human-w01-v01-n01-4
Chimp-w01-v01-n01-4
Gorilla-w01-v01-n01-4
Oran-w03-v01-n01-4
Oran-w05-v01-n01-4
Human-w03-v02-n01-4
Human-w05-v02-n01-4
Oran-w01-v01-n01-4
Human-w02-v01-n01-4
Gorilla-w02-v02-n01-4
Chimp-w04-v01-n01-4
Chimp-w03-v02-n01-4
Chimp-w02-v01-n01-4
Chimp-w05-v01-n01-4
Oran-w02-v01-n01-4
Oran-w04-v01-n01-4
Human-w01-v02-n01-4
Chimp-w01-v02-n01-4
Gorilla-w01-v02-n01-4
Human-w03-v01-n01-4
Oran-w02-v02-n01-4
Macaque-39bp-4
Oran-w01-v02-n01-4

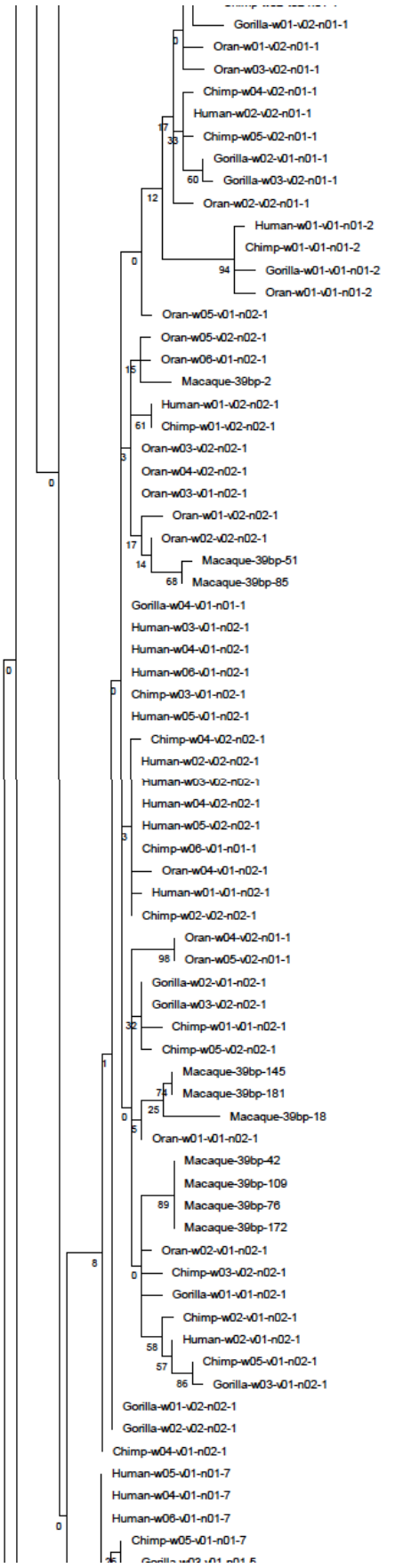

Gorilla-w03-v02-n02-4
Human-w04-v01-n01-4
Human-w05-v01-n01-4
Human-w06-v01-n01-4
Oran-w03-v02-n01-4
Oran-w04-v02-n01-4
Oran-w05-v02-n01-4
Gorilla-w04-v01-n02-4
Macaque-39bp-12
Chimp-w03-v01-n01-4
Oran-w06-v01-n01-4
Human-w02-v02-n01-4
Chimp-w04-v02-n01-4
Chimp-w02-v02-n01-4
Chimp-w05-v02-n01-4
Gorilla-w02-v01-n01-4
Gorilla-w03-v02-n01-4
Oran-w01-v02-n01-3
Oran-w02-v02-n01-3
Oran-w03-v02-n01-3
Oran-w06-v01-n01-3
Chimp-w04-v02-n01-3
Oran-w01-v01-n01-3
Oran-w04-v02-n01-3
Oran-w05-v02-n01-3
Chimp-w05-v02-n01-3
Gorilla-w02-v01-n01-3
Gorilla-w01-v02-n01-3
Gorilla-w03-v02-n01-3
Human-w01-v02-n01-3
Chimp-w01-v02-n01-3
Macaque-39bp-53
Macaque-39bp-87
Gorilla-w03-v01-n01-3
Gorilla-w04-v01-n02-3
Chimp-w03-v01-n01-3
Human-w03-v01-n01-3
Human-w03-v01-n01-9
Human-w04-v01-n01-3
Human-w06-v01-n01-3
Gorilla-w03-v02-n02-3
Human-w05-v01-n01-3
Human-w02-v02-n01-3
Chimp-w02-v02-n01-3
Human-w01-v01-n01-3
Chimp-w01-v01-n01-3
Gorilla-w01-v01-n01-3
Macaque-39bp-3
Oran-w01-v02-n02-4
Oran-w03-v02-n02-4
Oran-w05-v02-n02-4
Oran-w02-v02-n02-4
Macaque-39bp-20
Chimp-w03-v01-n02-4
Chimp-w05-v02-n02-4
Gorilla-w04-v01-n01-4
Human-w01-v02-n02-4
Chimp-w01-v02-n02-4
Gorilla-w02-v01-n02-4
Oran-w04-v02-n02-4
Chimp-w03-v02-n02-4
Human-w03-v02-n02-4
Human-w05-v02-n02-4
Human-w04-v02-n02-4
Chimp-w06-v01-n01-4
Chimp-w02-v02-n02-4
Human-w02-v02-n02-4
Chimp-w04-v02-n02-4
Human-w05-v01-n02-4
Human-w06-v01-n02-4

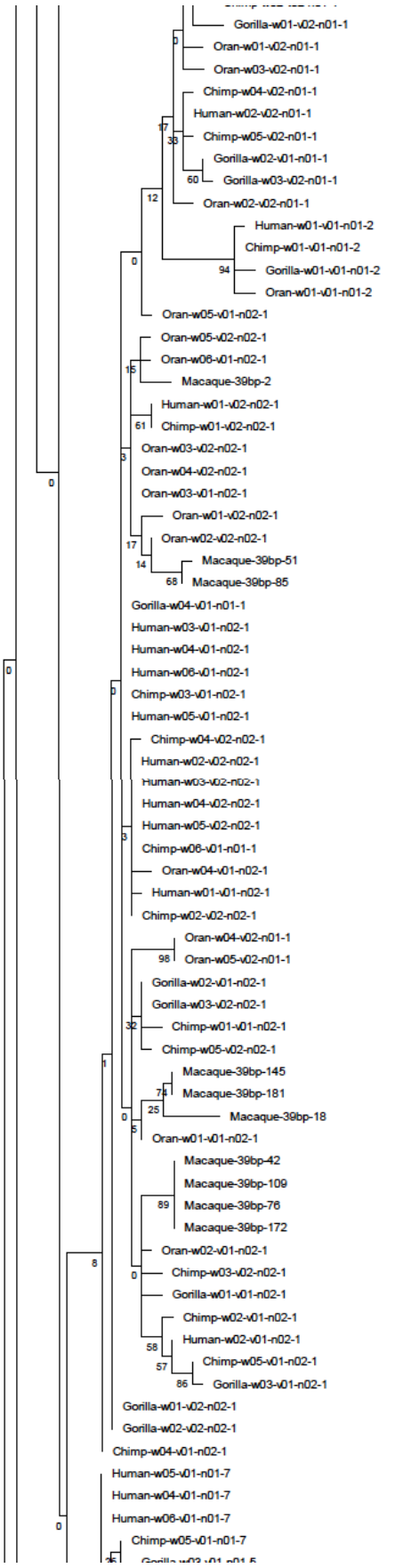

Human-w04-v01-n02-4
Human-w03-v01-n02-4
Macaque-39bp-54
Macaque-39bp-121
Macaque-39bp-29
Macaque-39bp-184
Macaque-39bp-148
Macaque-39bp-88
Macaque-39bp-91
Macaque-39bp-118
Macaque-39bp-157
Macaque-39bp-63
Chimp-w06-v01-n02-4
Macaque-39bp-94
Macaque-39bp-130
Oran-w06-v01-n02-4
Macaque-39bp-45
Macaque-39bp-79
Macaque-39bp-112
Macaque-39bp-175
Chimp-w05-v01-n02-4
Human-w01-v01-n02-4
Chimp-w01-v01-n02-4
Oran-w01-v01-n02-4
Oran-w05-v01-n02-4
Oran-w04-v01-n02-4
Chimp-w04-v01-n02-4
Oran-w03-v01-n02-4
Gorilla-w01-v02-n02-4
Gorilla-w02-v02-n02-4
Human-w02-v01-n02-4
Chimp-w02-v01-n02-4
Gorilla-w03-v01-n02-4
Gorilla-w01-v01-n02-4
Oran-w02-v01-n02-4
Human-w02-v01-n01-9
Chimp-w05-v01-n01-9
Chimp-w02-v01-n01-9
Gorilla-w03-v01-n01-7
Gorilla-w01-v02-n01-9
Gorilla-w02-v02-n01-9
Chimp-w01-v01-n01-9
Gorilla-w01-v01-n01-9
Human-w01-v01-n01-9
Oran-w05-v01-n01-9
Oran-w02-v02-n01-9
Oran-w01-v02-n01-9
Oran-w02-v01-n01-9
Oran-w04-v01-n01-9
Gorilla-w03-v02-n01-9
Human-w05-v01-n01-9
Gorilla-w02-v01-n01-9
Human-w04-v01-n01-9
Human-w06-v01-n01-9
Chimp-w03-v02-n01-9
Human-w03-v02-n01-9
Oran-w01-v01-n01-9
Oran-w03-v01-n01-9
Human-w05-v02-n01-9
Gorilla-w03-v02-n02-9
Chimp-w03-v01-n01-9
Chimp-w04-v01-n01-9
Human-w01-v02-n01-9
Chimp-w01-v02-n01-9
Human-w04-v02-n01-9
Macaque-39bp-17
Chimp-w02-v02-n01-9
Chimp-w04-v02-n01-9
Human-w02-v02-n01-9

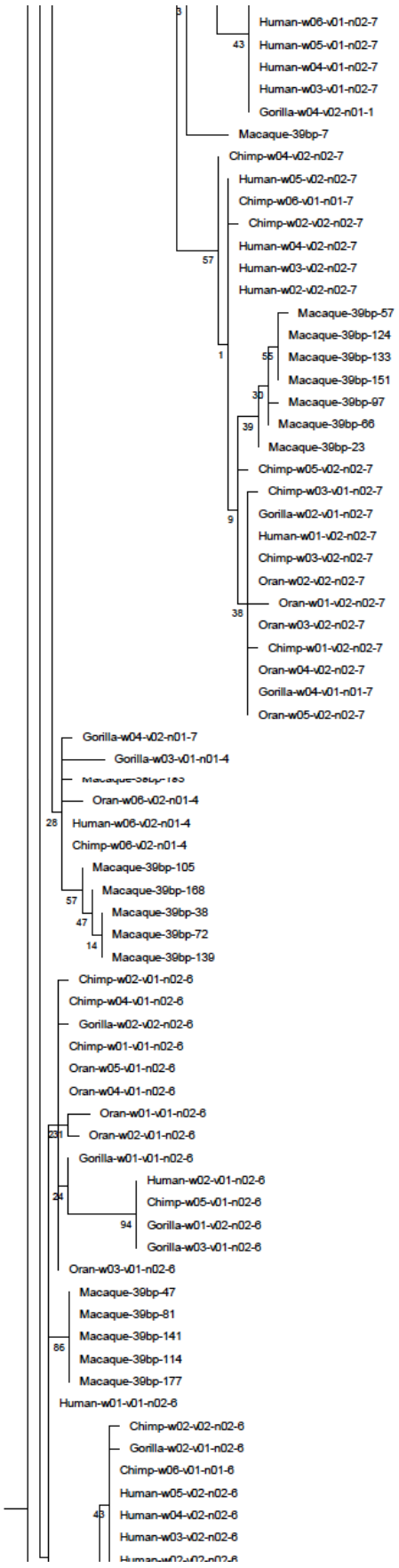

Chimp-w05-v02-n01-9
Oran-w05-v02-n01-9
Oran-w03-v02-n01-9
Oran-w04-v02-n01-9
Oran-w06-v01-n01-9
Macaque-39bp-50
Macaque-39bp-180
Macaque-39bp-144
Macaque-39bp-117
Human-w01-v01-n02-9
Chimp-w01-v01-n02-9
Gorilla-w01-v01-n02-9
Macaque-39bp-1
Human-w01-v01-n01-1
Chimp-w01-v01-n01-1
Gorilla-w01-v01-n01-1
Oran-w01-v01-n01-1
Human-w02-v01-n02-9
Oran-w03-v01-n02-9EXTRA
Oran-w04-v01-n02-9
Oran-w05-v01-n02-9
Chimp-w04-v01-n02-9
Oran-w02-v01-n02-9
Chimp-w02-v01-n02-9
Gorilla-w01-v02-n02-9
Gorilla-w03-v01-n02-9
Macaque-39bp-84
Chimp-w05-v01-n02-9
Oran-w01-v01-n02-9
Human-w03-v01-n02-9
Oran-w06-v01-n02-9
Human-w06-v01-n02-9
Gorilla-w04-v01-n02-9
Gorilla-w04-v02-n01-3
Human-w04-v01-n02-9
Human-w05-v01-n02-9
Chimp-w06-v01-n02-9
Macaque-39bp-189
Macaque-39bp-164
Macaque-39bp-34
Macaque-39bp-9
Macaque-39bp-41
Macaque-39bp-75
Macaque-39bp-108
Macaque-39bp-171
Chimp-w05-v02-n02-9
Oran-w02-v02-n02-9
Gorilla-w02-v01-n02-9
Human-w02-v02-n02-9
Macaque-39bp-126
Macaque-39bp-153
Macaque-39bp-88
Chimp-w04-v02-n02-9
Human-w01-v02-n02-9
Chimp-w01-v02-n02-9
Chimp-w03-v02-n02-9
Human-w03-v02-n02-9
Human-w04-v02-n02-9
Human-w05-v02-n02-9
Chimp-w06-v01-n01-9
Macaque-39bp-99
Macaque-39bp-59
Macaque-39bp-135
Gorilla-w04-v01-n01-9
Oran-w01-v02-n02-9
Oran-w03-v02-n02-9
Chimp-w03-v01-n02-9
Oran-w04-v02-n02-9
Chimp-w02-v02-n02-9
Macaque-39bp-25

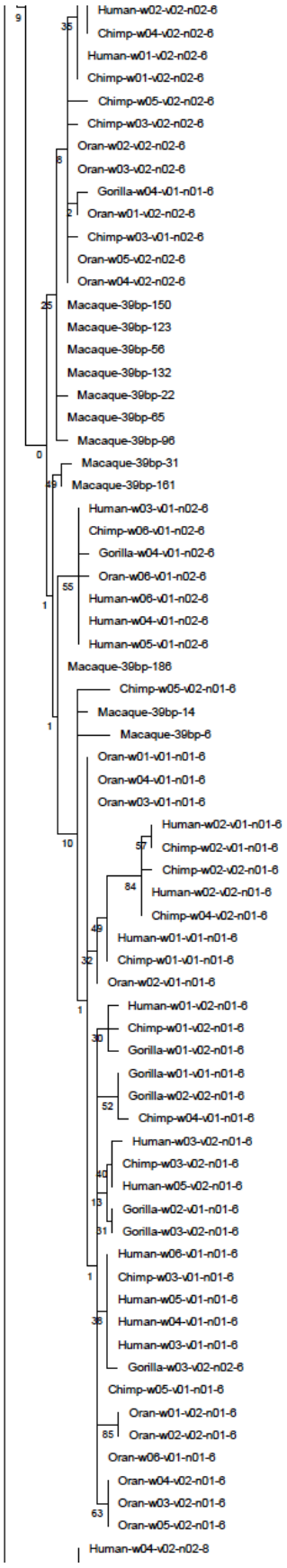

Gorilla-w02-v02-n02-9
Oran-w05-v02-n02-9
Macaque-39bp-160
Human-w03-v02-n01-5
Human-w05-v02-n01-5
Chimp-w03-v02-n01-5
Human-w06-v01-n01-5
Chimp-w03-v01-n01-5
Human-w05-v01-n01-5
Human-w04-v01-n01-5
Human-w03-v01-n01-5
Gorilla-w03-v02-n02-5
Chimp-w05-v02-n01-5
Macaque-39bp-5
Gorilla-w01-v01-n01-5
Oran-w02-v01-n01-5
Oran-w05-v02-n01-5
Oran-w04-v02-n01-5
Oran-w03-v02-n01-5
Gorilla-w02-v02-n01-5
Oran-w01-v02-n01-5
Oran-w02-v02-n01-5
Chimp-w04-v01-n01-5
Human-w02-v01-n01-5
Chimp-w02-v01-n01-5
Oran-w01-v01-n01-5
Human-w01-v01-n01-5
Chimp-w01-v01-n01-5
Oran-w06-v01-n01-5
Chimp-w05-v01-n01-5
Macaque-39bp-13
Gorilla-w02-v01-n01-5
Gorilla-w03-v02-n01-5
Oran-w05-v01-n01-5
Oran-w03-v01-n01-5
Oran-w04-v01-n01-5
Human-w02-v02-n01-5
Chimp-w02-v02-n01-5
Chimp-w04-v02-n01-5
Chimp-w04-v01-n02-5
Gorilla-w02-v02-n02-5
Chimp-w05-v01-n02-5
Gorilla-w03-v01-n02-5
Gorilla-w01-v02-n02-5
Oran-w01-v01-n02-5
Human-w02-v01-n02-5
Chimp-w02-v01-n02-5
Oran-w04-v01-n02-5
Oran-w05-v01-n02-5
Oran-w02-v01-n02-5
Gorilla-w01-v01-n02-5
Human-w01-v01-n02-5
Chimp-w01-v01-n02-5
Macaque-39bp-80
Macaque-39bp-176
Macaque-39bp-46
Macaque-39bp-113
Oran-w03-v01-n02-5
Macaque-39bp-55
Macaque-39bp-149
Macaque-39bp-131
Macaque-39bp-64
Macaque-39bp-122
Macaque-39bp-95
Human-w01-v02-n01-5
Chimp-w01-v02-n01-5
Gorilla-w01-v02-n01-5
Macaque-39bp-30
Gorilla-w02-v01-n02-5

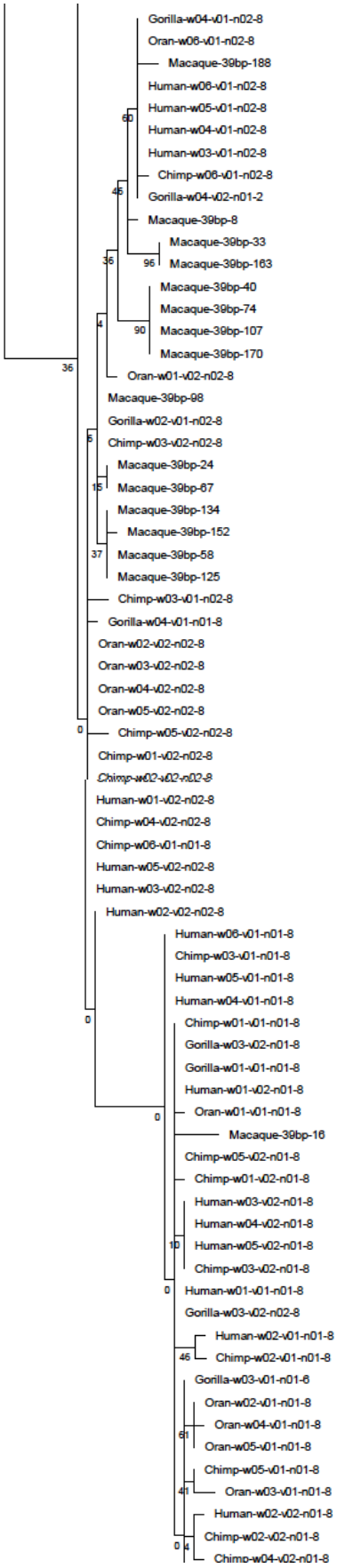

Oran-w01-v02-n02-5
Macaque-39bp-185
Oran-w02-v02-n02-5
Oran-w05-v02-n02-5
Macaque-39bp-21
Oran-w03-v02-n02-5
Chimp-w05-v02-n02-5
Oran-w04-v02-n02-5
Gorilla-w04-v01-n01-5
Chimp-w06-v01-n02-5
Gorilla-w04-v01-n02-5
Human-w06-v01-n02-5
Human-w05-v01-n02-5
Human-w03-v01-n02-5
Human-w04-v01-n02-5
Oran-w06-v01-n02-5
Chimp-w03-v02-n02-5
Human-w03-v02-n02-5
Chimp-w03-v01-n02-5
Human-w01-v02-n02-5
Chimp-w02-v02-n02-5
Chimp-w04-v02-n02-5
Human-w02-v02-n02-5
Human-w04-v02-n02-5
Human-w05-v02-n02-5
Chimp-w01-v02-n02-5
Chimp-w06-v01-n01-5
Human-w03-v01-n01-7
Gorilla-w04-v01-n02-1
Macaque-39bp-69
Macaque-39bp-136
Macaque-39bp-100
Macaque-39bp-60
Gorilla-w02-v02-n01-1
Chimp-w06-v01-n02-1
Human-w06-v01-n01-1
Human-w05-v01-n01-1
Human-w04-v01-n01-1
Human-w03-v01-n01-1
Macaque-39bp-127
Chimp-w05-v01-n01-1
Chimp-w04-v01-n01-1
Chimp-w03-v01-n01-1
Human-w02-v01-n01-1
Chimp-w02-v01-n01-1
Macaque-39bp-154
Macaque-39bp-26
Oran-w06-v01-n01-1
Oran-w05-v01-n01-1
Oran-w03-v01-n01-1
Oran-w02-v01-n01-1
Oran-w04-v01-n01-1
Gorilla-w03-v01-n01-1
Human-w04-v02-n01-1
Chimp-w03-v02-n01-1
Human-w04-v02-n01-4
Human-w05-v02-n01-1
Macaque-39bp-102
Gorilla-w04-v02-n01-4
Oran-w06-v02-n01-1
Macaque-39bp-190
Macaque-39bp-10
Macaque-39bp-35
Macaque-39bp-165
Chimp-w06-v02-n01-1
Human-w03-v02-n01-1
Human-w06-v02-n01-1
Human-w01-v02-n01-1
Chimp-w01-v02-n01-1
Chimp-w02-v02-n01-1


Gorilla-w01-v02-n01-1
Oran-w01-v02-n01-1
Oran-w03-v02-n01-1
Chimp-w04-v02-n01-1
Human-w02-v02-n01-1
Chimp-w05-v02-n01-1
Gorilla-w02-v01-n01-1
Gorilla-w03-v02-n01-1
Oran-w02-v02-n01-1
Human-w01-v01-n01-2
Chimp-w01-v01-n01-2
Gorilla-w01-v01-n01-2
Oran-w01-v01-n01-2
Oran-w05-v01-n02-1
Oran-w05-v02-n02-1
Oran-w06-v01-n02-1
Macaque-39bp-2
Human-w01-v02-n02-1
Chimp-w01-v02-n02-1
Oran-w03-v02-n02-1
Oran-w04-v02-n02-1
Oran-w03-v01-n02-1
Oran-w01-v02-n02-1
Oran-w02-v02-n02-1
Macaque-39bp-51
Macaque-39bp-85
Gorilla-w04-v01-n01-1
Human-w03-v01-n02-1
Human-w04-v01-n02-1
Human-w06-v01-n02-1
Chimp-w03-v01-n02-1
Human-w05-v01-n02-1
Chimp-w04-v02-n02-1
Human-w02-v02-n02-1
Human-w04-v02-n02-1
Human-w05-v02-n02-1
Chimp-w06-v01-n01-1
Oran-w04-v01-n02-1
Human-w01-v01-n02-1
Chimp-w02-v02-n02-1
Oran-w04-v02-n01-1
Oran-w05-v02-n01-1
Gorilla-w02-v01-n02-1
Gorilla-w03-v02-n02-1
Chimp-w01-v01-n02-1
Chimp-w05-v02-n02-1
Macaque-39bp-145
Macaque-39bp-181
Macaque-39bp-18
Oran-w01-v01-n02-1
Macaque-39bp-42
Macaque-39bp-109
Macaque-39bp-76
Macaque-39bp-172
Oran-w02-v01-n02-1
Chimp-w03-v02-n02-1
Gorilla-w01-v01-n02-1
Chimp-w02-v01-n02-1
Human-w02-v01-n02-1
Chimp-w05-v01-n02-1
Gorilla-w03-v01-n02-1
Gorilla-w01-v02-n02-1
Gorilla-w02-v02-n02-1
Chimp-w04-v01-n02-1
Human-w05-v01-n01-7
Human-w04-v01-n01-7
Human-w06-v01-n01-7
Chimp-w05-v01-n01-7


Gorilla-w01-v02-n01-1
Oran-w01-v02-n01-1
Oran-w03-v02-n01-1
Chimp-w04-v02-n01-1
Human-w02-v02-n01-1
Chimp-w05-v02-n01-1
Gorilla-w02-v01-n01-1
Gorilla-w03-v02-n01-1
Oran-w02-v02-n01-1
Human-w01-v01-n01-2
Chimp-w01-v01-n01-2
Gorilla-w01-v01-n01-2
Oran-w01-v01-n01-2
Oran-w05-v01-n02-1
Oran-w05-v02-n02-1
Oran-w06-v01-n02-1
Macaque-39bp-2
Human-w01-v02-n02-1
Chimp-w01-v02-n02-1
Oran-w03-v02-n02-1
Oran-w04-v02-n02-1
Oran-w03-v01-n02-1
Oran-w01-v02-n02-1
Oran-w02-v02-n02-1
Macaque-39bp-51
Macaque-39bp-85
Gorilla-w04-v01-n01-1
Human-w03-v01-n02-1
Human-w04-v01-n02-1
Human-w06-v01-n02-1
Chimp-w03-v01-n02-1
Human-w05-v01-n02-1
Chimp-w04-v02-n02-1
Human-w02-v02-n02-1
Human-w04-v02-n02-1
Human-w05-v02-n02-1
Chimp-w06-v01-n01-1
Oran-w04-v01-n02-1
Human-w01-v01-n02-1
Chimp-w02-v02-n02-1
Oran-w04-v02-n01-1
Oran-w05-v02-n01-1
Gorilla-w02-v01-n02-1
Gorilla-w03-v02-n02-1
Chimp-w01-v01-n02-1
Chimp-w05-v02-n02-1
Macaque-39bp-145
Macaque-39bp-181
Macaque-39bp-18
Oran-w01-v01-n02-1
Macaque-39bp-42
Macaque-39bp-109
Macaque-39bp-76
Macaque-39bp-172
Oran-w02-v01-n02-1
Chimp-w03-v02-n02-1
Gorilla-w01-v01-n02-1
Chimp-w02-v01-n02-1
Human-w02-v01-n02-1
Chimp-w05-v01-n02-1
Gorilla-w03-v01-n02-1
Gorilla-w01-v02-n02-1
Gorilla-w02-v02-n02-1
Chimp-w04-v01-n02-1
Human-w05-v01-n01-7
Human-w04-v01-n01-7
Human-w06-v01-n01-7
Chimp-w05-v01-n01-7


Human-w06-v01-n02-7
43
Human-w05-v01-n02-7
Human-w04-v01-n02-7
Human-w03-v01-n02-7
Gorilla-w04-v02-n01-1
Macaque-39bp-7
Chimp-w04-v02-n02-7
Human-w05-v02-n02-7
Chimp-w06-v01-n01-7
Chimp-w02-v02-n02-7
Human-w04-v02-n02-7
57
Human-w03-v02-n02-7
Human-w02-v02-n02-7
Macaque-39bp-57
Macaque-39bp-124
1
55
Macaque-39bp-133
Macaque-39bp-151
30
Macaque-39bp-97
39
Macaque-39bp-66
Macaque-39bp-23
Chimp-w05-v02-n02-7
Chimp-w03-v01-n02-7
9
Gorilla-w02-v01-n02-7
Human-w01-v02-n02-7
Chimp-w03-v02-n02-7
Oran-w02-v02-n02-7
Oran-w01-v02-n02-7
38
Oran-w03-v02-n02-7
Chimp-w01-v02-n02-7
Oran-w04-v02-n02-7
Gorilla-w04-v01-n01-7
Oran-w05-v02-n02-7
Gorilla-w04-v02-n01-7
Gorilla-w03-v01-n01-4
Oran-w06-v02-n01-4
28
Human-w06-v02-n01-4
Chimp-w06-v02-n01-4
Macaque-39bp-105
Macaque-39bp-168
57
Macaque-39bp-38
47
Macaque-39bp-72
14
Macaque-39bp-139
Chimp-w02-v01-n02-6
Chimp-w04-v01-n02-6
Gorilla-w02-v02-n02-6
Chimp-w01-v01-n02-6
Oran-w05-v01-n02-6
Oran-w04-v01-n02-6
Oran-w01-v01-n02-6
231
Oran-w02-v01-n02-6
Gorilla-w01-v01-n02-6
Human-w02-v01-n02-6
24
Chimp-w05-v01-n02-6
94
Gorilla-w01-v02-n02-6
Gorilla-w03-v01-n02-6
Oran-w03-v01-n02-6
Macaque-39bp-47
Macaque-39bp-81
Macaque-39bp-141
86
Macaque-39bp-114
Macaque-39bp-177
Human-w01-v01-n02-6
Chimp-w02-v02-n02-6
Gorilla-w02-v01-n02-6
Chimp-w06-v01-n01-6
Human-w05-v02-n02-6
43
Human-w04-v02-n02-6
Human-w03-v02-n02-6


Human-w02-v02-n02-6
Chimp-w04-v02-n02-6
Human-w01-v02-n02-6
Chimp-w01-v02-n02-6
Chimp-w05-v02-n02-6
Chimp-w03-v02-n02-6
Oran-w02-v02-n02-6
Oran-w03-v02-n02-6
Gorilla-w04-v01-n01-6
Oran-w01-v02-n02-6
Chimp-w03-v01-n02-6
Oran-w05-v02-n02-6
Oran-w04-v02-n02-6
Macaque-39bp-150
Macaque-39bp-123
Macaque-39bp-56
Macaque-39bp-132
Macaque-39bp-22
Macaque-39bp-65
Macaque-39bp-96
Macaque-39bp-31
Macaque-39bp-161
Human-w03-v01-n02-6
Chimp-w06-v01-n02-6
Gorilla-w04-v01-n02-6
Oran-w06-v01-n02-6
Human-w06-v01-n02-6
Human-w04-v01-n02-6
Human-w05-v01-n02-6
Macaque-39bp-186
Chimp-w05-v02-n01-6
Macaque-39bp-14
Macaque-39bp-6
Oran-w01-v01-n01-6
Oran-w04-v01-n01-6
Oran-w03-v01-n01-6
Human-w02-v01-n01-6
Chimp-w02-v01-n01-6
Chimp-w02-v02-n01-6
Human-w02-v02-n01-6
Chimp-w04-v02-n01-6
Human-w01-v01-n01-6
Chimp-w01-v01-n01-6
Oran-w02-v01-n01-6
Human-w01-v02-n01-6
Chimp-w01-v02-n01-6
Gorilla-w01-v02-n01-6
Gorilla-w01-v01-n01-6
Gorilla-w02-v02-n01-6
Chimp-w04-v01-n01-6
Human-w03-v02-n01-6
Chimp-w03-v02-n01-6
Human-w05-v02-n01-6
Gorilla-w02-v01-n01-6
Gorilla-w03-v02-n01-6
Human-w06-v01-n01-6
Chimp-w03-v01-n01-6
Human-w05-v01-n01-6
Human-w04-v01-n01-6
Human-w03-v01-n01-6
Gorilla-w03-v02-n02-6
Chimp-w05-v01-n01-6
Oran-w01-v02-n01-6
Oran-w02-v02-n01-6
Oran-w06-v01-n01-6
Oran-w04-v02-n01-6
Oran-w03-v02-n01-6
Oran-w05-v02-n01-6
Human-w04-v02-n02-8


Gorilla-w04-v01-n02-8
Oran-w06-v01-n02-8
Macaque-39bp-188
Human-w06-v01-n02-8
Human-w05-v01-n02-8
Human-w04-v01-n02-8
Human-w03-v01-n02-8
Chimp-w06-v01-n02-8
Gorilla-w04-v02-n01-2
Macaque-39bp-8
Macaque-39bp-33
Macaque-39bp-163
Macaque-39bp-40
Macaque-39bp-74
Macaque-39bp-107
Macaque-39bp-170
Oran-w01-v02-n02-8
Macaque-39bp-98
Gorilla-w02-v01-n02-8
Chimp-w03-v02-n02-8
Macaque-39bp-24
Macaque-39bp-67
Macaque-39bp-134
Macaque-39bp-152
Macaque-39bp-58
Macaque-39bp-125
Chimp-w03-v01-n02-8
Gorilla-w04-v01-n01-8
Oran-w02-v02-n02-8
Oran-w03-v02-n02-8
Oran-w04-v02-n02-8
Oran-w05-v02-n02-8
Chimp-w05-v02-n02-8
Chimp-w01-v02-n02-8
Chimp-w02-v02-n02-8
Human-w01-v02-n02-8
Chimp-w04-v02-n02-8
Chimp-w06-v01-n01-8
Human-w05-v02-n02-8
Human-w03-v02-n02-8
Human-w02-v02-n02-8
Human-w06-v01-n01-8
Chimp-w03-v01-n01-8
Human-w05-v01-n01-8
Human-w04-v01-n01-8
Chimp-w01-v01-n01-8
Gorilla-w03-v02-n01-8
Gorilla-w01-v01-n01-8
Human-w01-v02-n01-8
Oran-w01-v01-n01-8
Macaque-39bp-16
Chimp-w05-v02-n01-8
Chimp-w01-v02-n01-8
Human-w03-v02-n01-8
Human-w04-v02-n01-8
Human-w05-v02-n01-8
Chimp-w03-v02-n01-8
Human-w01-v01-n01-8
Gorilla-w03-v02-n02-8
Human-w02-v01-n01-8
Chimp-w02-v01-n01-8
Gorilla-w03-v01-n01-6
Oran-w02-v01-n01-8
Oran-w04-v01-n01-8
Oran-w05-v01-n01-8
Chimp-w05-v01-n01-8
Oran-w03-v01-n01-8
Human-w02-v02-n01-8
Chimp-w02-v02-n01-8
Chimp-w04-v02-n01-8

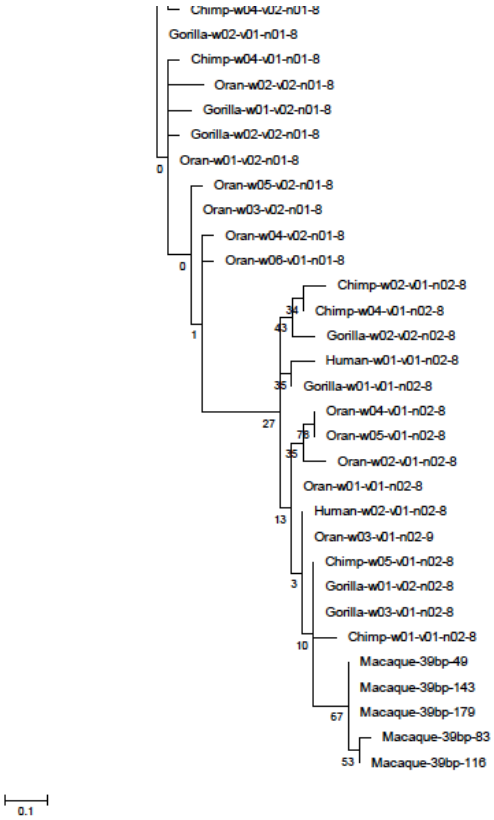

Chimp-w04-v02-n01-8
Gorilla-w02-v01-n01-8
Chimp-w04-v01-n01-8
Oran-w02-v02-n01-8
Gorilla-w01-v02-n01-8
Gorilla-w02-v02-n01-8
Oran-w01-v02-n01-8
Oran-w05-v02-n01-8
Oran-w03-v02-n01-8
Oran-w04-v02-n01-8
Oran-w08-v01-n01-8
Chimp-w02-v01-n02-8
Chimp-w04-v01-n02-8
Gorilla-w02-v02-n02-8
Human-w01-v01-n02-8
Gorilla-w01-v01-n02-8
Oran-w04-v01-n02-8
Oran-w05-v01-n02-8
Oran-w02-v01-n02-8
Oran-w01-v01-n02-8
Human-w02-v01-n02-8
Oran-w03-v01-n02-9
Chimp-w05-v01-n02-8
Gorilla-w01-v02-n02-8
Gorilla-w03-v01-n02-8
Chimp-w01-v01-n02-8
Macaque-39bp-49
Macaque-39bp-143
Macaque-39bp-179
Macaque-39bp-83
Macaque-39bp-116
0
0
1
34
43
35
27
75
35
13
3
10
67
53
0.1

Figure 25.

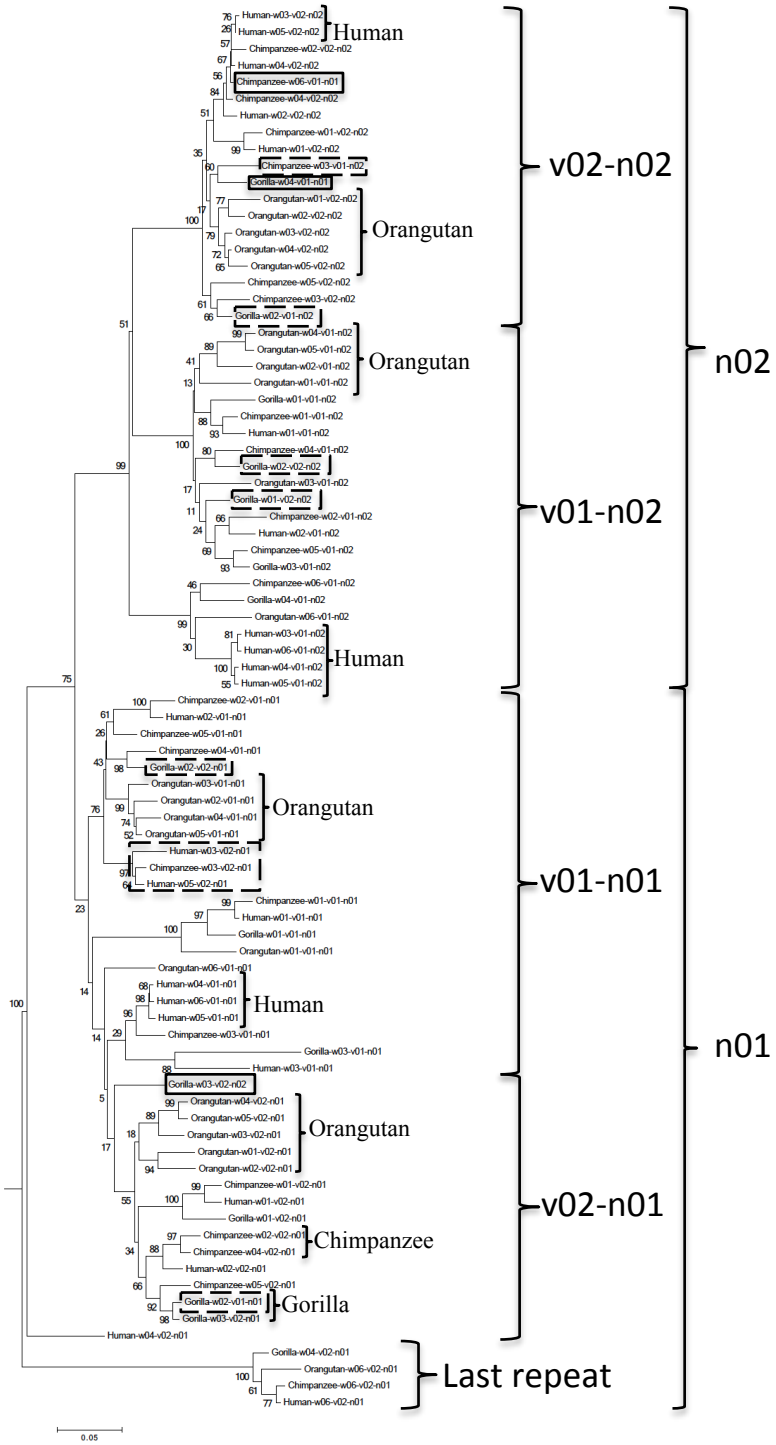
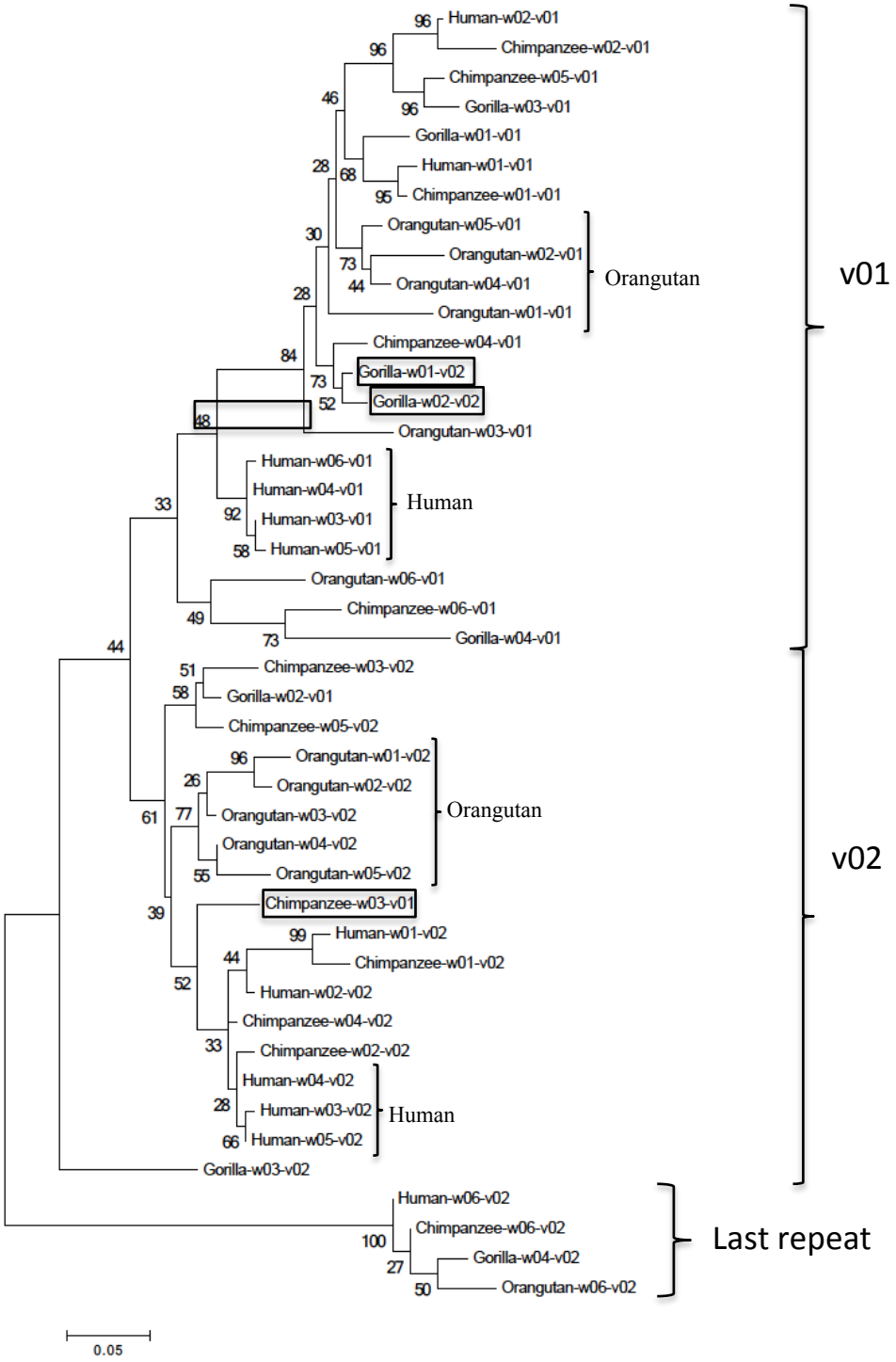

Human-w03-v02-n02
Human-w05-v02-n02
Human
Chimpanzee-w02-v02-n02
Human-w04-v02-n02
Chimpanzee-w06-v01-n01
Chimpanzee-w04-v02-n02
Human-w02-v02-n02
Chimpanzee-w01-v02-n02
Human-w01-v02-n02
Chimpanzee-w03-v01-n02
Gorilla-w04-v01-n01
Orangutan-w01-v02-n02
Orangutan-w02-v02-n02
Orangutan-w03-v02-n02
Orangutan-w04-v02-n02
Orangutan-w05-v02-n02
Orangutan
Chimpanzee-w05-v02-n02
Chimpanzee-w03-v02-n02
Gorilla-w02-v01-n02
v02-n02
Orangutan-w04-v01-n02
Orangutan-w05-v01-n02
Orangutan-w02-v01-n02
Orangutan-w01-v01-n02
Orangutan
Gorilla-w01-v01-n02
Chimpanzee-w01-v01-n02
Human-w01-v01-n02
Chimpanzee-w04-v01-n02
Gorilla-w02-v02-n02
Orangutan-w03-v01-n02
Gorilla-w01-v02-n02
Chimpanzee-w02-v01-n02
Human-w02-v01-n02
Chimpanzee-w05-v01-n02
Gorilla-w03-v01-n02
v01-n02
Chimpanzee-w06-v01-n02
Gorilla-w04-v01-n02
Orangutan-w06-v01-n02
Human-w03-v01-n02
Human-w06-v01-n02
Human-w04-v01-n02
Human-w05-v01-n02
Human
n02
Chimpanzee-w02-v01-n01
Human-w02-v01-n01
Chimpanzee-w05-v01-n01
Chimpanzee-w04-v01-n01
Gorilla-w02-v02-n01
Orangutan-w03-v01-n01
Orangutan-w02-v01-n01
Orangutan-w04-v01-n01
Orangutan-w05-v01-n01
Orangutan
Human-w03-v02-n01
Chimpanzee-w03-v02-n01
Human-w05-v02-n01
v01-n01
Chimpanzee-w01-v01-n01
Human-w01-v01-n01
Gorilla-w01-v01-n01
Orangutan-w01-v01-n01
Orangutan-w06-v01-n01
Human-w04-v01-n01
Human-w06-v01-n01
Human-w05-v01-n01
Human
Chimpanzee-w03-v01-n01
Gorilla-w03-v01-n01
Human-w03-v01-n01
Gorilla-w03-v02-n02
Orangutan-w04-v02-n01
Orangutan-w05-v02-n01
Orangutan-w03-v02-n01
Orangutan-w01-v02-n01
Orangutan-w02-v02-n01
Orangutan
Chimpanzee-w01-v02-n01
Human-w01-v02-n01
Gorilla-w01-v02-n01
Chimpanzee-w02-v02-n01
Chimpanzee-w04-v02-n01
Chimpanzee
Human-w02-v02-n01
Chimpanzee-w05-v02-n01
Gorilla-w02-v01-n01
Gorilla-w03-v02-n01
Gorilla
v02-n01
n01
Human-w04-v02-n01
Gorilla-w04-v02-n01
Orangutan-w06-v02-n01
Chimpanzee-w06-v02-n01
Human-w06-v02-n01
Last repeat
0.05

Figure 26.


Human-w02-v01
Chimpanzee-w02-v01
Chimpanzee-w05-v01
Gorilla-w03-v01
Gorilla-w01-v01
Human-w01-v01
Chimpanzee-w01-v01
Orangutan-w05-v01
Orangutan-w02-v01
Orangutan-w04-v01
Orangutan-w01-v01
Orangutan
Chimpanzee-w04-v01
Gorilla-w01-v02
Gorilla-w02-v02
Orangutan-w03-v01
Human-w06-v01
Human-w04-v01
Human-w03-v01
Human-w05-v01
Human
Orangutan-w06-v01
Chimpanzee-w06-v01
Gorilla-w04-v01
Chimpanzee-w03-v02
Gorilla-w02-v01
Chimpanzee-w05-v02
Orangutan-w01-v02
Orangutan-w02-v02
Orangutan-w03-v02
Orangutan-w04-v02
Orangutan-w05-v02
Orangutan
Chimpanzee-w03-v01
Human-w01-v02
Chimpanzee-w01-v02
Human-w02-v02
Chimpanzee-w04-v02
Chimpanzee-w02-v02
Human-w04-v02
Human-w03-v02
Human-w05-v02
Human
Gorilla-w03-v02
Human-w06-v02
Chimpanzee-w06-v02
Gorilla-w04-v02
Orangutan-w06-v02
Last repeat
v01
v02
0.05

Filaggrin-2
Figure 27.

FLG-2-FLG-like

Human

1	2	3	4	5	6	7	8	9	10	11	12	13	14
225	231	231	225	231	225	225	231	225	225	225	225	225	225

Chimpanzee

1	2	3	4	5	6	7	8	9	10	11	12	13
225	231	231	225	231	225	222	231	225	225	225	225	225

Macaque

1	2	3	4	5	6	7	8	9	10	11	12	13	14
225	231	231	231	228	225	225	225	225	225	225	225	225	225

Baboon

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
225	231	231	222	225	225	225	225	225	225	225	225	225	225	225	225	207	225	225	225	225	225

Marmoset

1	2	3	4	5	6	7	8
225	225	225	233	204	225	225	225

FLG-2-HRNR-like

Human

1	2	3	4	5	6	7	8	9
231	225	225	231	225	231	231	231	231

Chimpanzee

1	2	3	4	5	6	7	8	9
231	225	225	231	225	231	231	231	231

Macaque

1	2	3	4	5	6	7
231	225	225	231	231	231	231

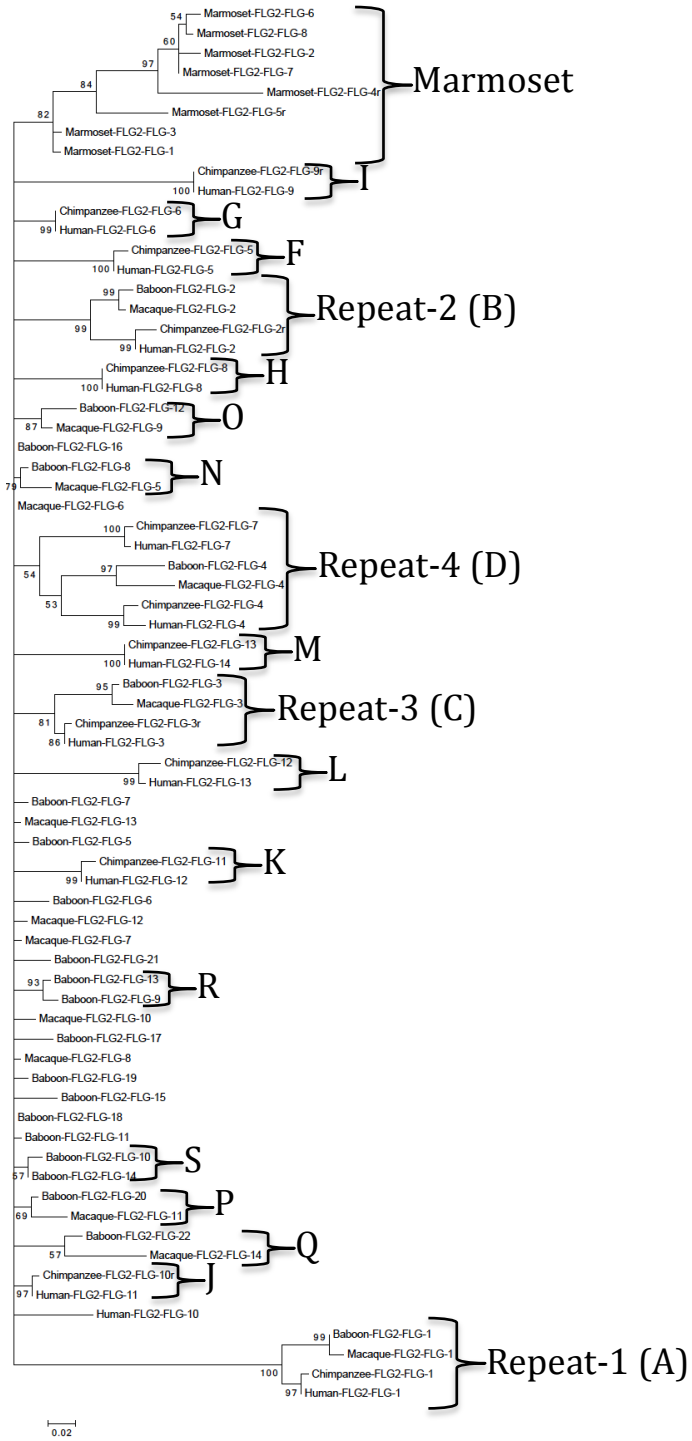
Baboon

1	2	3	4	5	6	7
231	231	225	231	231	231	231

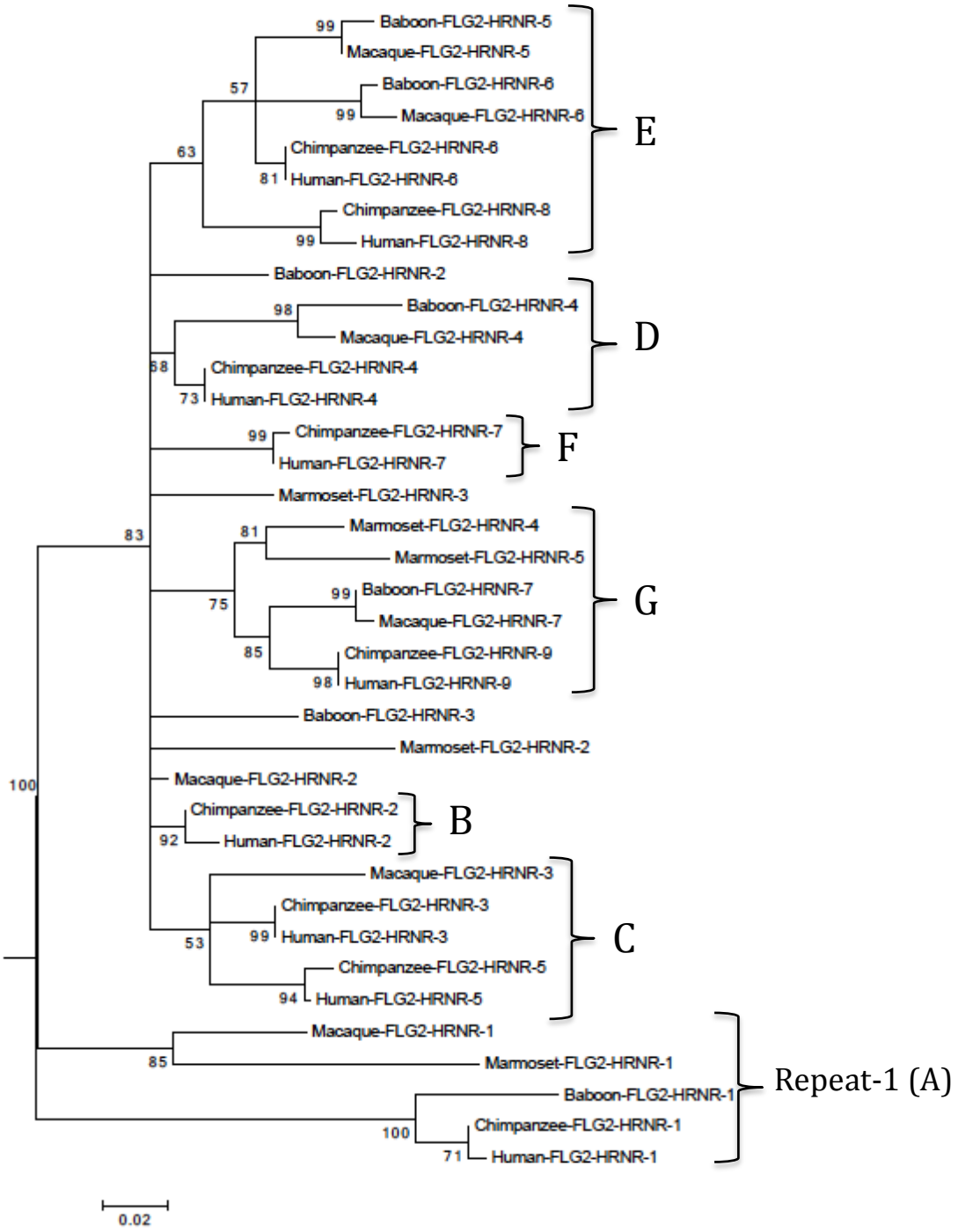
Marmoset

1	2	3	4	5
231	192	231	192	231

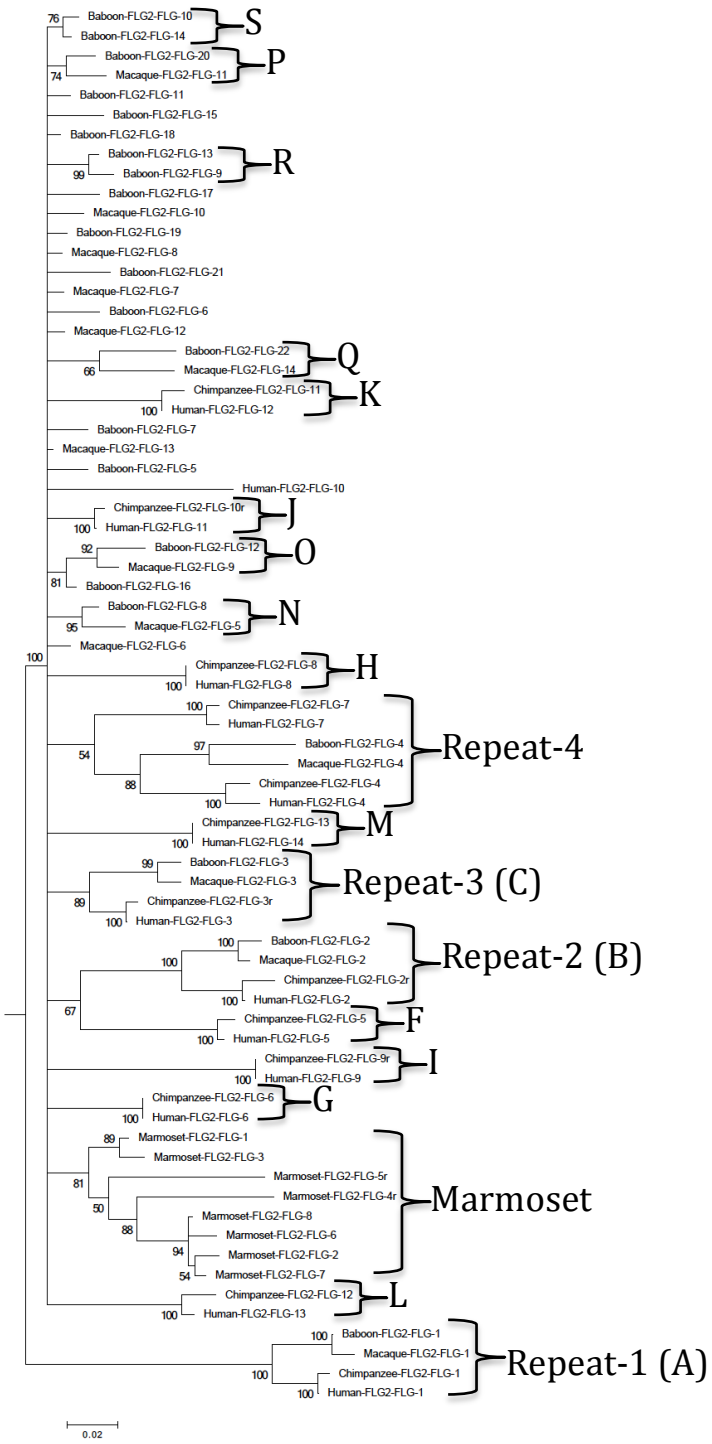
Figure 28.
A.


Marmoset-FLG2-FLG-6
Marmoset-FLG2-FLG-8
Marmoset-FLG2-FLG-2
Marmoset-FLG2-FLG-7
Marmoset-FLG2-FLG-4r
Marmoset-FLG2-FLG-5r
Marmoset-FLG2-FLG-3
Marmoset-FLG2-FLG-1
Marmoset
Chimpanzee-FLG2-FLG-9r
Human-FLG2-FLG-9
I
Chimpanzee-FLG2-FLG-6
Human-FLG2-FLG-6
G
Chimpanzee-FLG2-FLG-5
Human-FLG2-FLG-5
F
Baboon-FLG2-FLG-2
Macaque-FLG2-FLG-2
Chimpanzee-FLG2-FLG-2r
Human-FLG2-FLG-2
Repeat-2 (B)
Chimpanzee-FLG2-FLG-8
Human-FLG2-FLG-8
H
Baboon-FLG2-FLG-12
Macaque-FLG2-FLG-9
O
Baboon-FLG2-FLG-16
Baboon-FLG2-FLG-8
Macaque-FLG2-FLG-5
N
Macaque-FLG2-FLG-6
Chimpanzee-FLG2-FLG-7
Human-FLG2-FLG-7
Baboon-FLG2-FLG-4
Macaque-FLG2-FLG-4
Chimpanzee-FLG2-FLG-4
Human-FLG2-FLG-4
Repeat-4 (D)
Chimpanzee-FLG2-FLG-13
Human-FLG2-FLG-14
M
Baboon-FLG2-FLG-3
Macaque-FLG2-FLG-3
Chimpanzee-FLG2-FLG-3r
Human-FLG2-FLG-3
Repeat-3 (C)
Chimpanzee-FLG2-FLG-12
Human-FLG2-FLG-13
L
Baboon-FLG2-FLG-7
Macaque-FLG2-FLG-13
Baboon-FLG2-FLG-5
Chimpanzee-FLG2-FLG-11
Human-FLG2-FLG-12
K
Baboon-FLG2-FLG-6
Macaque-FLG2-FLG-12
Macaque-FLG2-FLG-7
Baboon-FLG2-FLG-21
Baboon-FLG2-FLG-13
Baboon-FLG2-FLG-9
R
Macaque-FLG2-FLG-10
Baboon-FLG2-FLG-17
Macaque-FLG2-FLG-8
Baboon-FLG2-FLG-19
Baboon-FLG2-FLG-15
Baboon-FLG2-FLG-18
Baboon-FLG2-FLG-11
Baboon-FLG2-FLG-10
Baboon-FLG2-FLG-14
S
Baboon-FLG2-FLG-20
Macaque-FLG2-FLG-11
P
Baboon-FLG2-FLG-22
Macaque-FLG2-FLG-14
Q
Chimpanzee-FLG2-FLG-10r
Human-FLG2-FLG-11
J
Human-FLG2-FLG-10
Baboon-FLG2-FLG-1
Macaque-FLG2-FLG-1
Chimpanzee-FLG2-FLG-1
Human-FLG2-FLG-1
Repeat-1 (A)
0.02

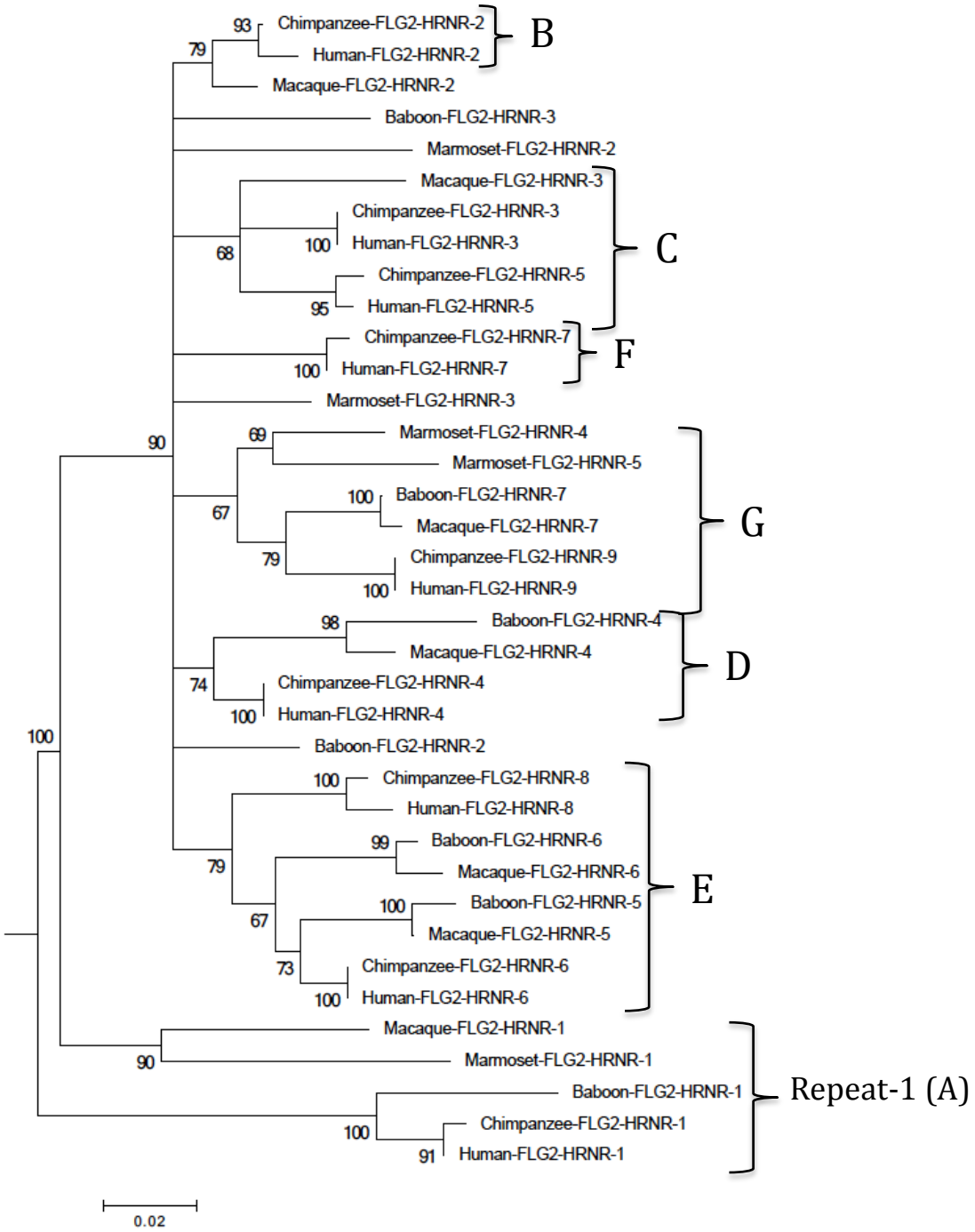
B.


Baboon-FLG2-HRNR-5
Macaque-FLG2-HRNR-5
Baboon-FLG2-HRNR-6
Macaque-FLG2-HRNR-6
Chimpanzee-FLG2-HRNR-6
Human-FLG2-HRNR-6
Chimpanzee-FLG2-HRNR-8
Human-FLG2-HRNR-8
E
Baboon-FLG2-HRNR-2
Baboon-FLG2-HRNR-4
Macaque-FLG2-HRNR-4
Chimpanzee-FLG2-HRNR-4
Human-FLG2-HRNR-4
D
Chimpanzee-FLG2-HRNR-7
Human-FLG2-HRNR-7
F
Marmoset-FLG2-HRNR-3
Marmoset-FLG2-HRNR-4
Marmoset-FLG2-HRNR-5
Baboon-FLG2-HRNR-7
Macaque-FLG2-HRNR-7
Chimpanzee-FLG2-HRNR-9
Human-FLG2-HRNR-9
G
Baboon-FLG2-HRNR-3
Marmoset-FLG2-HRNR-2
Macaque-FLG2-HRNR-2
Chimpanzee-FLG2-HRNR-2
Human-FLG2-HRNR-2
B
Macaque-FLG2-HRNR-3
Chimpanzee-FLG2-HRNR-3
Human-FLG2-HRNR-3
Chimpanzee-FLG2-HRNR-5
Human-FLG2-HRNR-5
C
Macaque-FLG2-HRNR-1
Marmoset-FLG2-HRNR-1
Baboon-FLG2-HRNR-1
Chimpanzee-FLG2-HRNR-1
Human-FLG2-HRNR-1
Repeat-1 (A)
0.02

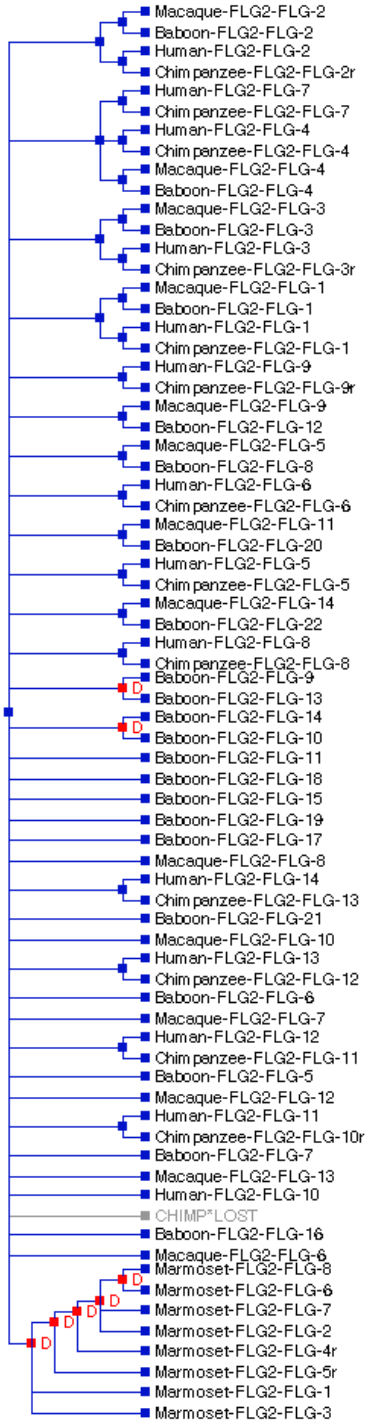
C.


Baboon-FLG2-FLG-10
Baboon-FLG2-FLG-14
S
Baboon-FLG2-FLG-20
Macaque-FLG2-FLG-11
P
Baboon-FLG2-FLG-11
Baboon-FLG2-FLG-15
Baboon-FLG2-FLG-18
Baboon-FLG2-FLG-13
Baboon-FLG2-FLG-9
R
Baboon-FLG2-FLG-17
Macaque-FLG2-FLG-10
Baboon-FLG2-FLG-19
Macaque-FLG2-FLG-8
Baboon-FLG2-FLG-21
Macaque-FLG2-FLG-7
Baboon-FLG2-FLG-6
Macaque-FLG2-FLG-12
Baboon-FLG2-FLG-22
Macaque-FLG2-FLG-14
Q
Chimpanzee-FLG2-FLG-11
Human-FLG2-FLG-12
K
Baboon-FLG2-FLG-7
Macaque-FLG2-FLG-13
Baboon-FLG2-FLG-5
Human-FLG2-FLG-10
Chimpanzee-FLG2-FLG-10r
Human-FLG2-FLG-11
J
Baboon-FLG2-FLG-12
Macaque-FLG2-FLG-9
O
Baboon-FLG2-FLG-16
Baboon-FLG2-FLG-8
Macaque-FLG2-FLG-5
N
Macaque-FLG2-FLG-6
Chimpanzee-FLG2-FLG-8
Human-FLG2-FLG-8
H
Chimpanzee-FLG2-FLG-7
Human-FLG2-FLG-7
Baboon-FLG2-FLG-4
Macaque-FLG2-FLG-4
Chimpanzee-FLG2-FLG-4
Human-FLG2-FLG-4
Repeat-4
Chimpanzee-FLG2-FLG-13
Human-FLG2-FLG-14
M
Baboon-FLG2-FLG-3
Macaque-FLG2-FLG-3
Chimpanzee-FLG2-FLG-3r
Human-FLG2-FLG-3
Repeat-3 (C)
Baboon-FLG2-FLG-2
Macaque-FLG2-FLG-2
Chimpanzee-FLG2-FLG-2r
Human-FLG2-FLG-2
Repeat-2 (B)
Chimpanzee-FLG2-FLG-5
Human-FLG2-FLG-5
F
Chimpanzee-FLG2-FLG-9r
Human-FLG2-FLG-9
I
Chimpanzee-FLG2-FLG-6
Human-FLG2-FLG-6
G
Marmoset-FLG2-FLG-1
Marmoset-FLG2-FLG-3
Marmoset-FLG2-FLG-5r
Marmoset-FLG2-FLG-4r
Marmoset-FLG2-FLG-8
Marmoset-FLG2-FLG-6
Marmoset-FLG2-FLG-2
Marmoset-FLG2-FLG-7
Marmoset
Chimpanzee-FLG2-FLG-12
Human-FLG2-FLG-13
L
Baboon-FLG2-FLG-1
Macaque-FLG2-FLG-1
Chimpanzee-FLG2-FLG-1
Human-FLG2-FLG-1
Repeat-1 (A)
0.02

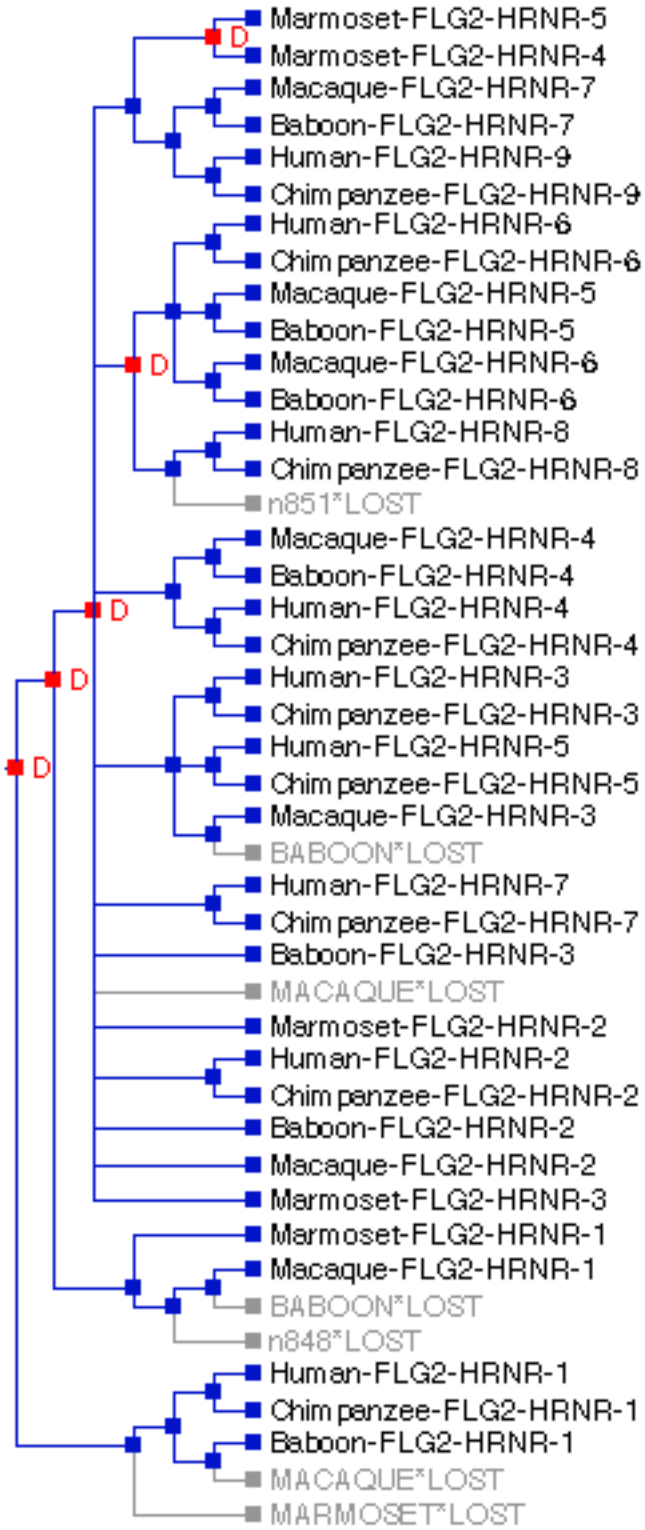
D.


Chimpanzee-FLG2-HRNR-2
Human-FLG2-HRNR-2
Macaque-FLG2-HRNR-2
Baboon-FLG2-HRNR-3
Marmoset-FLG2-HRNR-2
Macaque-FLG2-HRNR-3
Chimpanzee-FLG2-HRNR-3
Human-FLG2-HRNR-3
Chimpanzee-FLG2-HRNR-5
Human-FLG2-HRNR-5
Chimpanzee-FLG2-HRNR-7
Human-FLG2-HRNR-7
Marmoset-FLG2-HRNR-3
Marmoset-FLG2-HRNR-4
Marmoset-FLG2-HRNR-5
Baboon-FLG2-HRNR-7
Macaque-FLG2-HRNR-7
Chimpanzee-FLG2-HRNR-9
Human-FLG2-HRNR-9
Baboon-FLG2-HRNR-4
Macaque-FLG2-HRNR-4
Chimpanzee-FLG2-HRNR-4
Human-FLG2-HRNR-4
Baboon-FLG2-HRNR-2
Chimpanzee-FLG2-HRNR-8
Human-FLG2-HRNR-8
Baboon-FLG2-HRNR-6
Macaque-FLG2-HRNR-6
Baboon-FLG2-HRNR-5
Macaque-FLG2-HRNR-5
Chimpanzee-FLG2-HRNR-6
Human-FLG2-HRNR-6
Macaque-FLG2-HRNR-1
Marmoset-FLG2-HRNR-1
Baboon-FLG2-HRNR-1
Chimpanzee-FLG2-HRNR-1
Human-FLG2-HRNR-1
B
C
F
G
D
E
Repeat-1 (A)
0.02

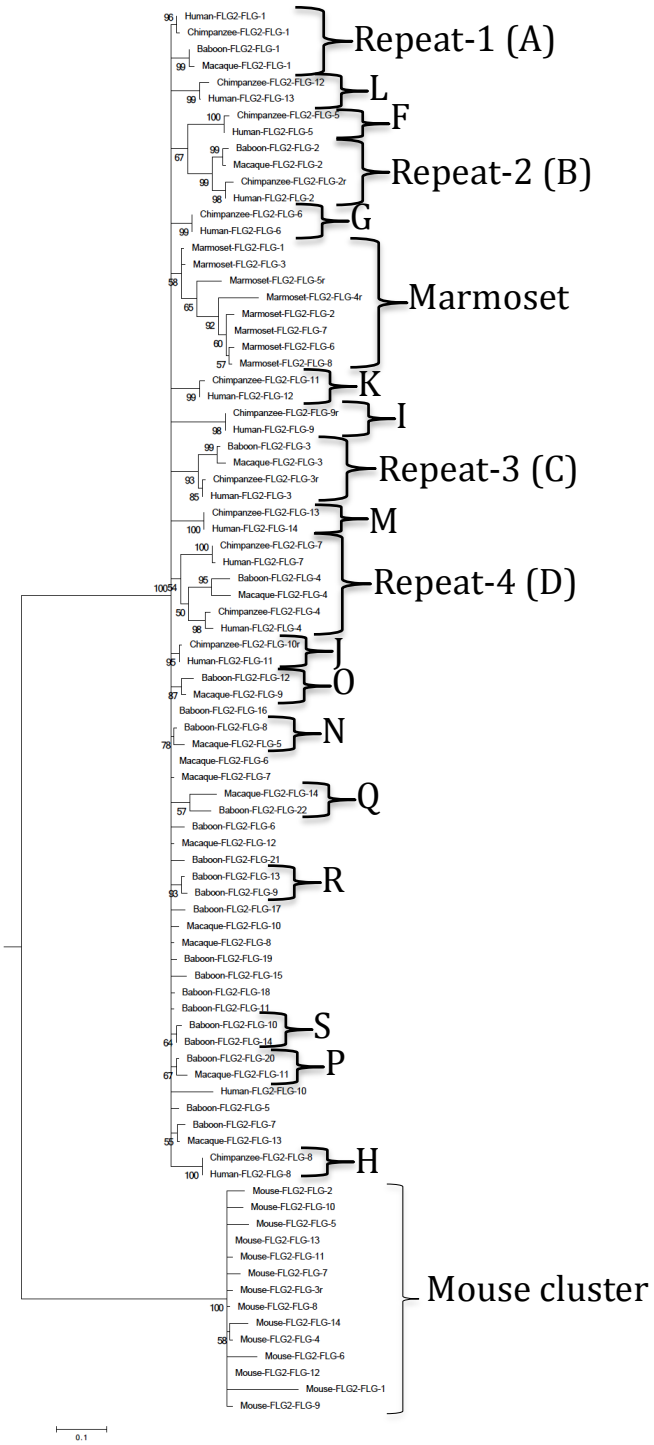
E.


Macaque-FLG2-FLG-2
Baboon-FLG2-FLG-2
Human-FLG2-FLG-2
Chimpanzee-FLG2-FLG-2r
Human-FLG2-FLG-7
Chimpanzee-FLG2-FLG-7
Human-FLG2-FLG-4
Chimpanzee-FLG2-FLG-4
Macaque-FLG2-FLG-4
Baboon-FLG2-FLG-4
Macaque-FLG2-FLG-3
Baboon-FLG2-FLG-3
Human-FLG2-FLG-3
Chimpanzee-FLG2-FLG-3r
Macaque-FLG2-FLG-1
Baboon-FLG2-FLG-1
Human-FLG2-FLG-1
Chimpanzee-FLG2-FLG-1
Human-FLG2-FLG-9
Chimpanzee-FLG2-FLG-9r
Macaque-FLG2-FLG-9
Baboon-FLG2-FLG-12
Macaque-FLG2-FLG-5
Baboon-FLG2-FLG-8
Human-FLG2-FLG-6
Chimpanzee-FLG2-FLG-6
Macaque-FLG2-FLG-11
Baboon-FLG2-FLG-20
Human-FLG2-FLG-5
Chimpanzee-FLG2-FLG-5
Macaque-FLG2-FLG-14
Baboon-FLG2-FLG-22
Human-FLG2-FLG-8
Chimpanzee-FLG2-FLG-8
Baboon-FLG2-FLG-9
D
Baboon-FLG2-FLG-13
Baboon-FLG2-FLG-14
D
Baboon-FLG2-FLG-10
Baboon-FLG2-FLG-11
Baboon-FLG2-FLG-18
Baboon-FLG2-FLG-15
Baboon-FLG2-FLG-19
Baboon-FLG2-FLG-17
Macaque-FLG2-FLG-8
Human-FLG2-FLG-14
Chimpanzee-FLG2-FLG-13
Baboon-FLG2-FLG-21
Macaque-FLG2-FLG-10
Human-FLG2-FLG-13
Chimpanzee-FLG2-FLG-12
Baboon-FLG2-FLG-6
Macaque-FLG2-FLG-7
Human-FLG2-FLG-12
Chimpanzee-FLG2-FLG-11
Baboon-FLG2-FLG-5
Macaque-FLG2-FLG-12
Human-FLG2-FLG-11
Chimpanzee-FLG2-FLG-10r
Baboon-FLG2-FLG-7
Macaque-FLG2-FLG-13
Human-FLG2-FLG-10
CHIMP*LOST
Baboon-FLG2-FLG-16
Macaque-FLG2-FLG-6
Marmoset-FLG2-FLG-8
D
Marmoset-FLG2-FLG-6
D
Marmoset-FLG2-FLG-7
D
Marmoset-FLG2-FLG-2
D
Marmoset-FLG2-FLG-4r
Marmoset-FLG2-FLG-5r
Marmoset-FLG2-FLG-1
Marmoset-FLG2-FLG-3

F.


Marmoset-FLG2-HRNR-5
Marmoset-FLG2-HRNR-4
Macaque-FLG2-HRNR-7
Baboon-FLG2-HRNR-7
Human-FLG2-HRNR-9
Chimpanzee-FLG2-HRNR-9
Human-FLG2-HRNR-6
Chimpanzee-FLG2-HRNR-6
Macaque-FLG2-HRNR-5
Baboon-FLG2-HRNR-5
Macaque-FLG2-HRNR-6
Baboon-FLG2-HRNR-6
Human-FLG2-HRNR-8
Chimpanzee-FLG2-HRNR-8
n851*LOST
Macaque-FLG2-HRNR-4
Baboon-FLG2-HRNR-4
Human-FLG2-HRNR-4
Chimpanzee-FLG2-HRNR-4
Human-FLG2-HRNR-3
Chimpanzee-FLG2-HRNR-3
Human-FLG2-HRNR-5
Chimpanzee-FLG2-HRNR-5
Macaque-FLG2-HRNR-3
BABOON*LOST
Human-FLG2-HRNR-7
Chimpanzee-FLG2-HRNR-7
Baboon-FLG2-HRNR-3
MACAQUE*LOST
Marmoset-FLG2-HRNR-2
Human-FLG2-HRNR-2
Chimpanzee-FLG2-HRNR-2
Baboon-FLG2-HRNR-2
Macaque-FLG2-HRNR-2
Marmoset-FLG2-HRNR-3
Marmoset-FLG2-HRNR-1
Macaque-FLG2-HRNR-1
BABOON*LOST
n848*LOST
Human-FLG2-HRNR-1
Chimpanzee-FLG2-HRNR-1
Baboon-FLG2-HRNR-1
MACAQUE*LOST
MARMOSET*LOST
D
D
D
D
D

G.


Repeat-1 (A)
L
F
Repeat-2 (B)
G
Marmoset
K
I
Repeat-3 (C)
M
Repeat-4 (D)
J
O
N
Q
R
S
P
H
Mouse cluster
Human-FLG2-FLG-1
Chimpanzee-FLG2-FLG-1
Baboon-FLG2-FLG-1
Macaque-FLG2-FLG-1
Chimpanzee-FLG2-FLG-12
Human-FLG2-FLG-13
Chimpanzee-FLG2-FLG-5
Human-FLG2-FLG-5
Baboon-FLG2-FLG-2
Macaque-FLG2-FLG-2
Chimpanzee-FLG2-FLG-2r
Human-FLG2-FLG-2
Chimpanzee-FLG2-FLG-6
Human-FLG2-FLG-6
Marmoset-FLG2-FLG-1
Marmoset-FLG2-FLG-3
Marmoset-FLG2-FLG-5r
Marmoset-FLG2-FLG-4r
Marmoset-FLG2-FLG-2
Marmoset-FLG2-FLG-7
Marmoset-FLG2-FLG-6
Marmoset-FLG2-FLG-8
Chimpanzee-FLG2-FLG-11
Human-FLG2-FLG-12
Chimpanzee-FLG2-FLG-9r
Human-FLG2-FLG-9
Baboon-FLG2-FLG-3
Macaque-FLG2-FLG-3
Chimpanzee-FLG2-FLG-3r
Human-FLG2-FLG-3
Chimpanzee-FLG2-FLG-13
Human-FLG2-FLG-14
Chimpanzee-FLG2-FLG-7
Human-FLG2-FLG-7
Baboon-FLG2-FLG-4
Macaque-FLG2-FLG-4
Chimpanzee-FLG2-FLG-4
Human-FLG2-FLG-4
Chimpanzee-FLG2-FLG-10r
Human-FLG2-FLG-11
Baboon-FLG2-FLG-12
Macaque-FLG2-FLG-9
Baboon-FLG2-FLG-16
Baboon-FLG2-FLG-8
Macaque-FLG2-FLG-5
Macaque-FLG2-FLG-6
Macaque-FLG2-FLG-7
Macaque-FLG2-FLG-14
Baboon-FLG2-FLG-22
Baboon-FLG2-FLG-6
Macaque-FLG2-FLG-12
Baboon-FLG2-FLG-21
Baboon-FLG2-FLG-13
Baboon-FLG2-FLG-9
Baboon-FLG2-FLG-17
Macaque-FLG2-FLG-10
Macaque-FLG2-FLG-8
Baboon-FLG2-FLG-19
Baboon-FLG2-FLG-15
Baboon-FLG2-FLG-18
Baboon-FLG2-FLG-11
Baboon-FLG2-FLG-10
Baboon-FLG2-FLG-14
Baboon-FLG2-FLG-20
Macaque-FLG2-FLG-11
Human-FLG2-FLG-10
Baboon-FLG2-FLG-5
Baboon-FLG2-FLG-7
Macaque-FLG2-FLG-13
Chimpanzee-FLG2-FLG-8
Human-FLG2-FLG-8
Mouse-FLG2-FLG-2
Mouse-FLG2-FLG-10
Mouse-FLG2-FLG-5
Mouse-FLG2-FLG-13
Mouse-FLG2-FLG-11
Mouse-FLG2-FLG-7
Mouse-FLG2-FLG-3r
Mouse-FLG2-FLG-8
Mouse-FLG2-FLG-14
Mouse-FLG2-FLG-4
Mouse-FLG2-FLG-6
Mouse-FLG2-FLG-12
Mouse-FLG2-FLG-1
Mouse-FLG2-FLG-9
0.1

Repetin
Figure 29.

Repetin

Human

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36

Chimpanzee

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36

Gorilla

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36

Orangutan

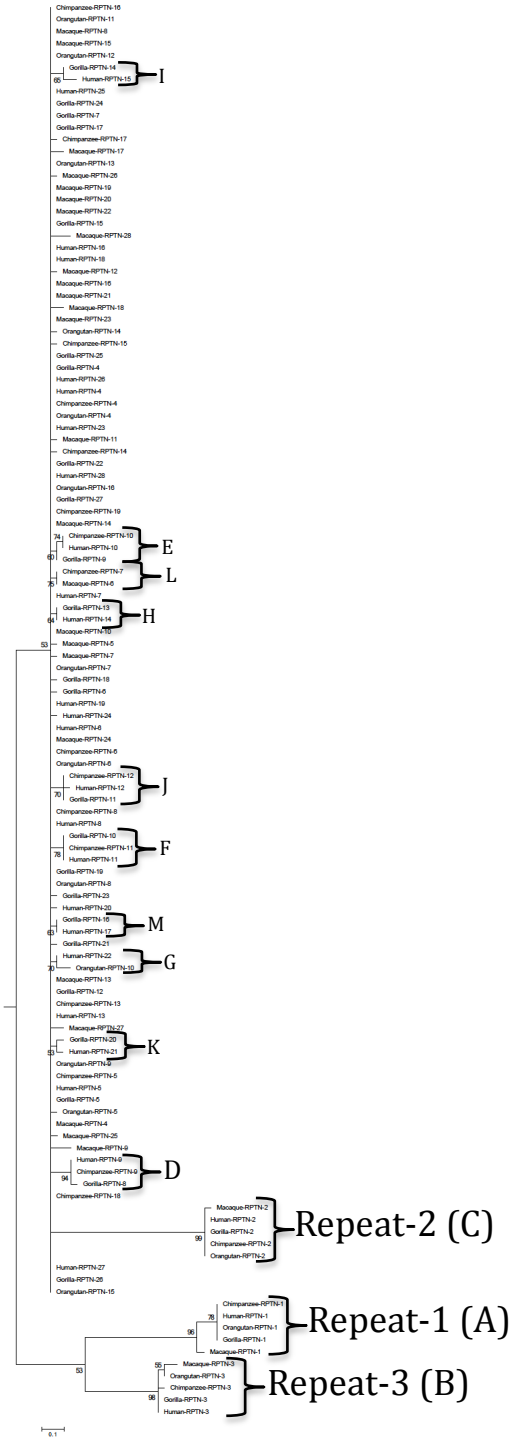
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36

Macaque

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36

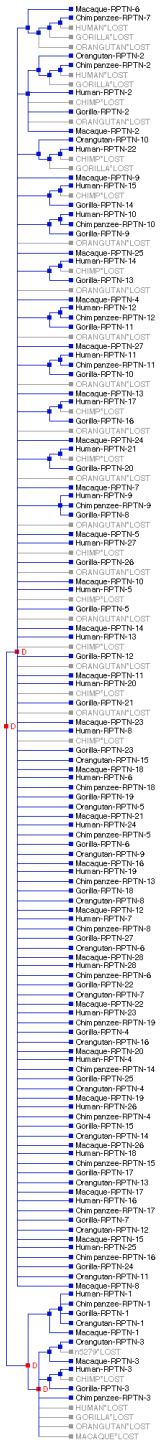
Figure 30.

A.

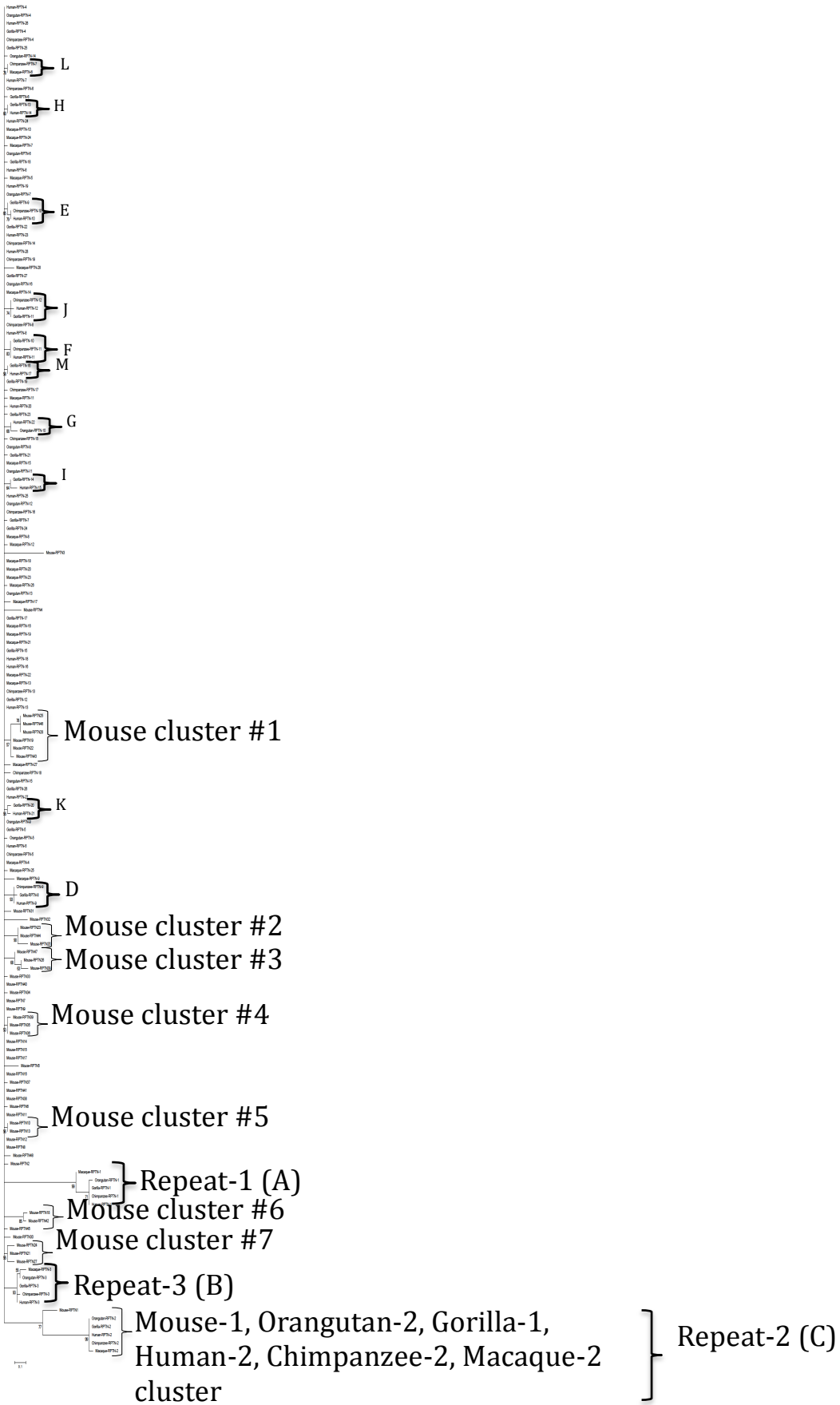


B.

C.


Macaque-RPTN-6
Chim penzee-RPTN-7
HUMAN*LOST
GORILLA*LOST
ORANGUTAN*LOST
Orangutan-RPTN-2
Chim penzee-RPTN-2
HUMAN*LOST
GORILLA*LOST
Human-RPTN-2
CHIMP*LOST
Gorilla-RPTN-2
ORANGUTAN*LOST
Macaque-RPTN-2
Orangutan-RPTN-10
Human-RPTN-22
CHIMP*LOST
GORILLA*LOST
Macaque-RPTN-9
Human-RPTN-15
CHIMP*LOST
Gorilla-RPTN-14
Human-RPTN-10
Chim penzee-RPTN-10
Gorilla-RPTN-9
ORANGUTAN*LOST
Macaque-RPTN-25
Human-RPTN-14
CHIMP*LOST
Gorilla-RPTN-13
ORANGUTAN*LOST
Macaque-RPTN-4
Human-RPTN-12
Chim penzee-RPTN-12
Gorilla-RPTN-11
ORANGUTAN*LOST
Macaque-RPTN-27
Human-RPTN-11
Chim penzee-RPTN-11
Gorilla-RPTN-10
ORANGUTAN*LOST
Macaque-RPTN-13
Human-RPTN-17
CHIMP*LOST
Gorilla-RPTN-16
ORANGUTAN*LOST
Macaque-RPTN-24
Human-RPTN-21
CHIMP*LOST
Gorilla-RPTN-20
ORANGUTAN*LOST
Macaque-RPTN-7
Human-RPTN-9
Chim penzee-RPTN-9
Gorilla-RPTN-8
ORANGUTAN*LOST
Macaque-RPTN-5
Human-RPTN-27
CHIMP*LOST
Gorilla-RPTN-26
ORANGUTAN*LOST
Macaque-RPTN-10
Human-RPTN-5
CHIMP*LOST
Gorilla-RPTN-5
ORANGUTAN*LOST
Macaque-RPTN-14
Human-RPTN-13
CHIMP*LOST
D
Gorilla-RPTN-12
ORANGUTAN*LOST
Macaque-RPTN-11
Human-RPTN-20
CHIMP*LOST
Gorilla-RPTN-21
ORANGUTAN*LOST
Macaque-RPTN-23
Human-RPTN-8
CHIMP*LOST
D
Gorilla-RPTN-23
Orangutan-RPTN-15
Macaque-RPTN-18
Human-RPTN-6
Chim penzee-RPTN-18
Gorilla-RPTN-19
Orangutan-RPTN-5
Macaque-RPTN-21
Human-RPTN-24
Chim penzee-RPTN-5
Gorilla-RPTN-6
Orangutan-RPTN-9
Macaque-RPTN-16
Human-RPTN-19
Chim penzee-RPTN-13
Gorilla-RPTN-18
Orangutan-RPTN-8
Macaque-RPTN-12
Human-RPTN-7
Chim penzee-RPTN-8
Gorilla-RPTN-27
Orangutan-RPTN-6
Macaque-RPTN-28
Human-RPTN-28
Chim penzee-RPTN-6
Gorilla-RPTN-22
Orangutan-RPTN-7
Macaque-RPTN-22
Human-RPTN-23
Chim penzee-RPTN-19
Gorilla-RPTN-4
Orangutan-RPTN-16
Macaque-RPTN-20
Human-RPTN-4
Chim penzee-RPTN-14
Gorilla-RPTN-25
Orangutan-RPTN-4
Macaque-RPTN-19
Human-RPTN-26
Chim penzee-RPTN-4
Gorilla-RPTN-15
Orangutan-RPTN-14
Macaque-RPTN-26
Human-RPTN-18
Chim penzee-RPTN-15
Gorilla-RPTN-17
Orangutan-RPTN-13
Macaque-RPTN-17
Human-RPTN-16
Chim penzee-RPTN-17
Gorilla-RPTN-7
Orangutan-RPTN-12
Macaque-RPTN-15
Human-RPTN-25
Chim penzee-RPTN-16
Gorilla-RPTN-24
Orangutan-RPTN-11
Macaque-RPTN-8
Human-RPTN-1
Chim penzee-RPTN-1
Gorilla-RPTN-1
Orangutan-RPTN-1
Macaque-RPTN-1
Orangutan-RPTN-3
rh5079*LOST
Macaque-RPTN-3
Human-RPTN-3
CHIMP*LOST
D
Gorilla-RPTN-3
Chim penzee-RPTN-3
HUMAN*LOST
GORILLA*LOST
ORANGUTAN*LOST
MACAQUE*LOST

D.


L
H
E
J
F
M
G
I
Mouse cluster #1
K
D
Mouse cluster #2
Mouse cluster #3
Mouse cluster #4
Mouse cluster #5
Repeat-1 (A)
Mouse cluster #6
Mouse cluster #7
Repeat-3 (B)
Mouse-1, Orangutan-2, Gorilla-1, Human-2, Chimpanzee-2, Macaque-2 cluster
Repeat-2 (C)

Cornulin
Figure 31.

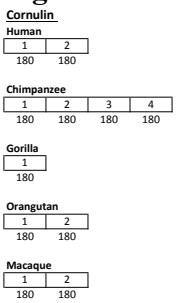
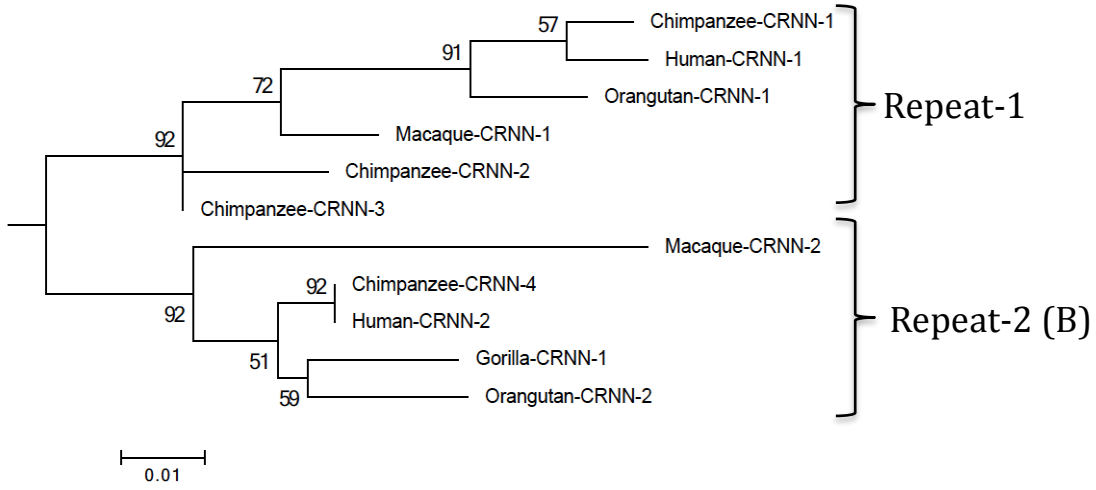
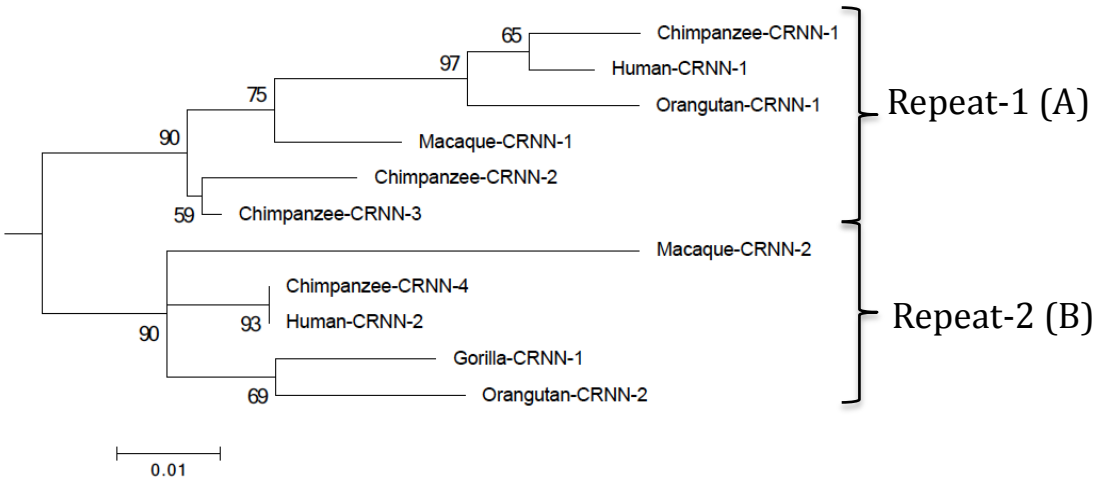


Figure 32

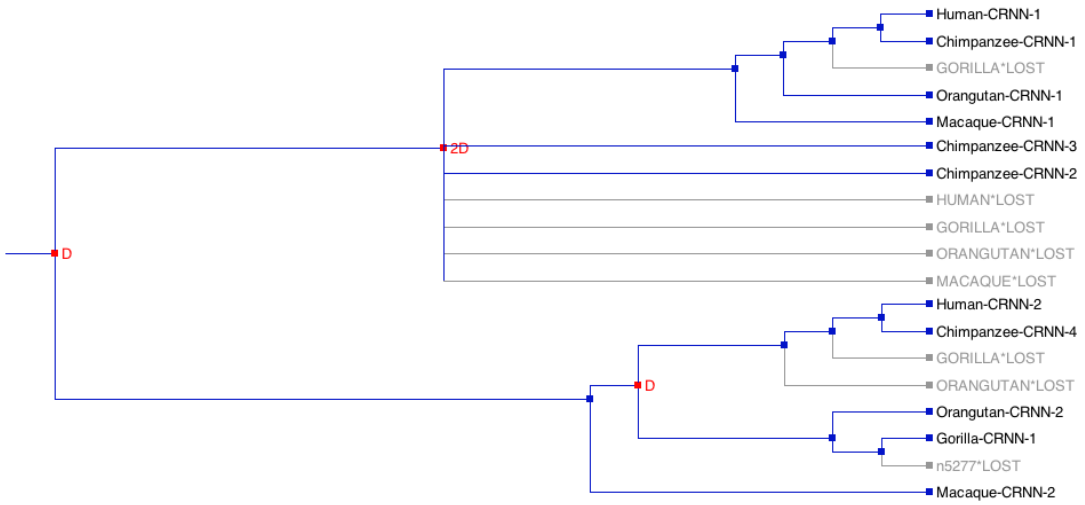
A.



B.



C.



D.

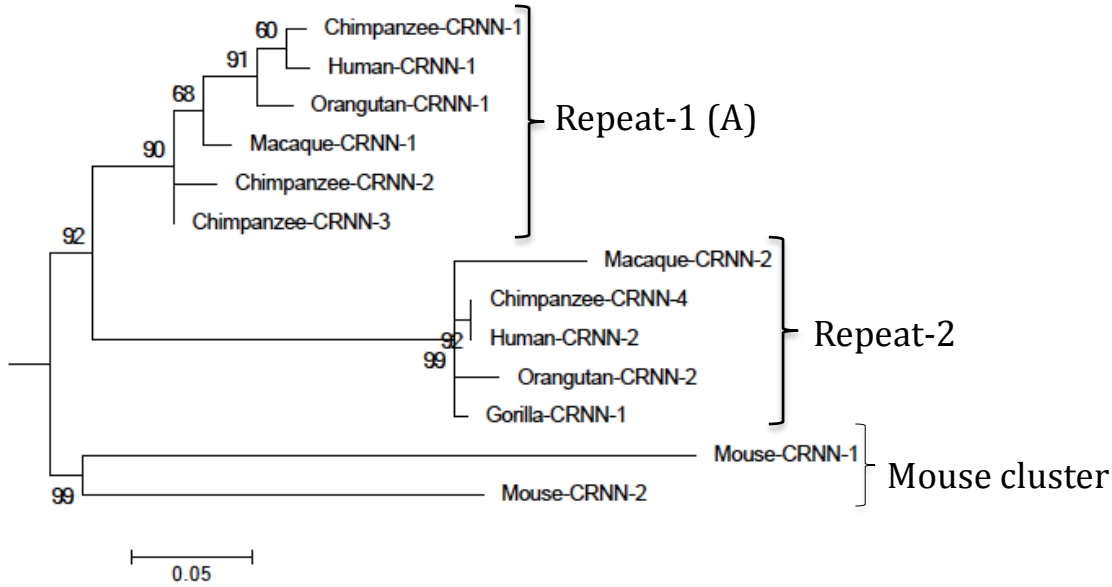

60
91
68
90
92
99
92
99
Chimpanzee-CRNN-1
Human-CRNN-1
Orangutan-CRNN-1
Macaque-CRNN-1
Chimpanzee-CRNN-2
Chimpanzee-CRNN-3
Macaque-CRNN-2
Chimpanzee-CRNN-4
Human-CRNN-2
Orangutan-CRNN-2
Gorilla-CRNN-1
Mouse-CRNN-1
Mouse-CRNN-2
Repeat-1 (A)
Repeat-2
Mouse cluster
0.05

Figure 33.

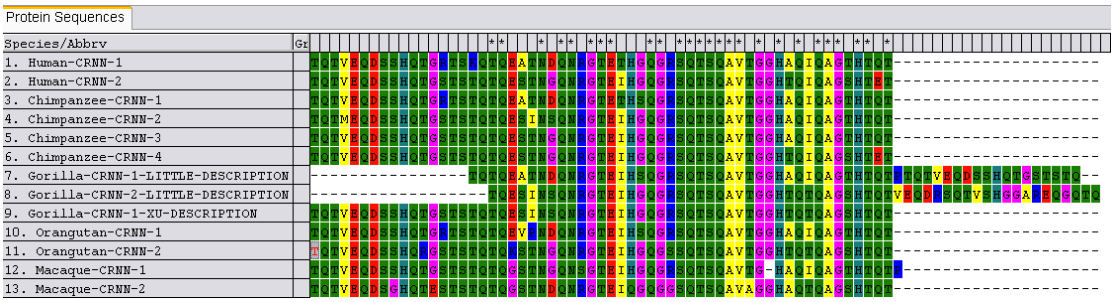

Protein Sequences
Species/Abbrv
1. Human-CRNN-1
2. Human-CRNN-2
3. Chimpanzee-CRNN-1
4. Chimpanzee-CRNN-2
5. Chimpanzee-CRNN-3
6. Chimpanzee-CRNN-4
7. Gorilla-CRNN-1-LITTLE-DESCRIPTION
8. Gorilla-CRNN-2-LITTLE-DESCRIPTION
9. Gorilla-CRNN-1-XU-DESCRIPTION
10. Orangutan-CRNN-1
11. Orangutan-CRNN-2
12. Macaque-CRNN-1
13. Macaque-CRNN-2

Figure 34.

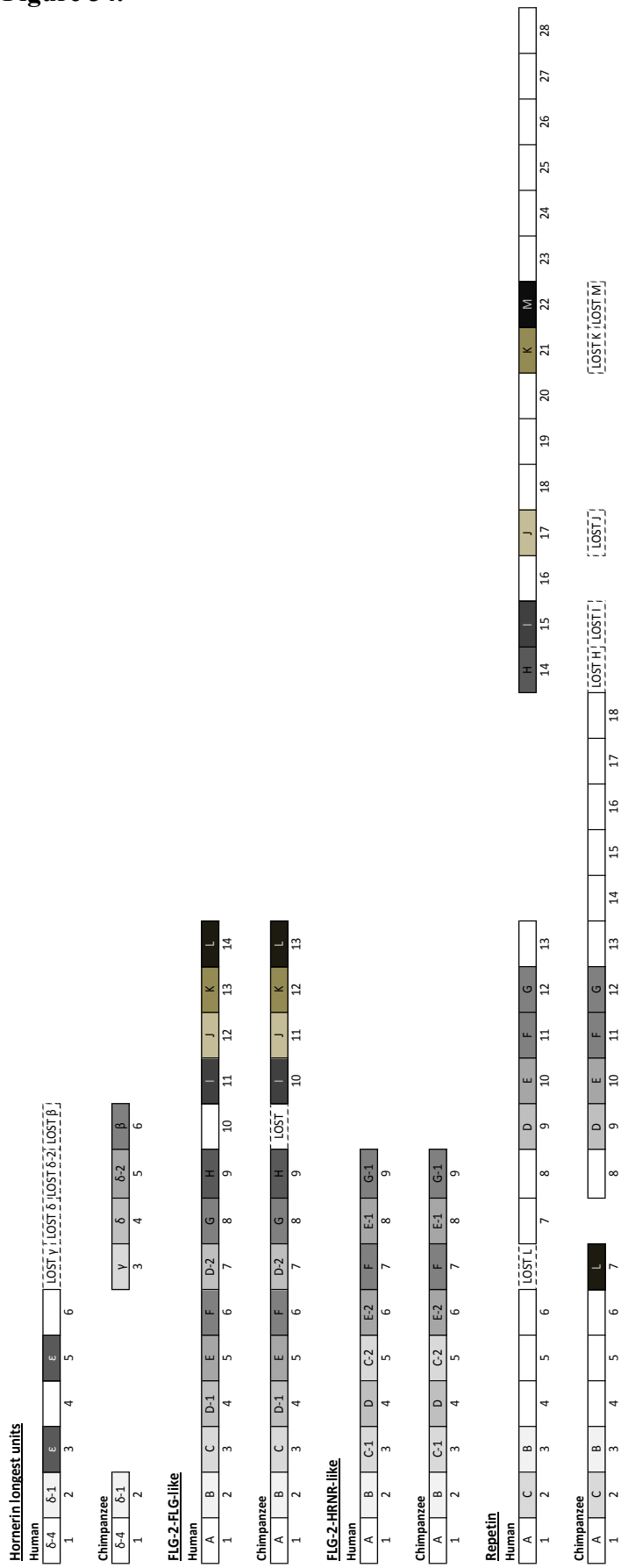

Hornerin longest units
Human
δ-4 δ-1 ε LOST γ LOST δ LOST δ-2 LOST β
1 2 3 4 5 6
Chimpanzee
δ-4 δ-1
1 2
γ δ δ-2 β
3 4 5 6
FLG-2-FLG-like
Human
A B C D-1 E F D-2 G H I J K L
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Chimpanzee
A B C D-1 E F D-2 G H LOST I J K L
1 2 3 4 5 6 7 8 9 10 11 12 13
FLG-2-HRNR-like
Human
A B C-1 D C-2 E-2 F E-1 G-1
1 2 3 4 5 6 7 8 9
Chimpanzee
A B C-1 D C-2 E-2 F E-1 G-1
1 2 3 4 5 6 7 8 9
Repetin
Human
A C B LOST L D E F G H I J K M
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
LOST H LOST I LOST J LOST K LOST M
Chimpanzee
A C B L D E F G
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
LOST H LOST I LOST J LOST K LOST M
Cornulin
Human
A-1 B
1 2
Chimpanzee
A-1 A-2 A-3 B
1 2 3 4

Figure 35.

Filaggrin

Human

1	2	3	4	5	6	7	8	9	10
975	972	972	972	975	972	972	972	972	972

Mouse

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
738	750	750	738	738	759	750	765	750	765	750	765	759	750	750	765

Hornerin

Human

1	2	3	4	5	6
1410	1398	1410	1410	1410	936

Mouse

1A	1B	2A	2B	3A	3B	4A	4B	5B1	5B2	6A	6B
534	513	531	513	531	513	531	513	513	513	531	513

FLG-2-HRNR-like **FLG-2-FLG-like**

Human

1	2	3	4	5	6	7	8	9		1	2	3	4	5	6	7	8	9	10	11	12	13	14
231	225	225	231	225	231	231	231	231		225	231	231	225	231	225	225	231	225	225	225	225	225	225

Mouse

Not found in mouse		1	2	3	4	5	6	7	8	9	10	11	12	13	14
		219	234	234	240	234	228	234	240	228	234	234	234	234	234

Repetin

Human

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36

Mouse

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36

Cornulin

Human

1	2
180	180

Mouse

1	2
177	177