Computational Molecular Evolution Project
Name: Drashti Mehta; SID: 801262877

Interspecies Comparison of the Primates by Understanding the Evolution of GULO Gene

Introduction:

Primate history and evolution has always been the primary debate for the evolution biologists. The evolutionary biologists study different physiological or molecular changes in different primate species which show the evidence of the evolution. Primates are remarkably recent animals. The **first primate-like animals** appeared around 65 million years ago (Mesozoic Era). Reptiles were being replaced by mammals during this Era, thus giving rise to the **proto-primates** or the primate-like mammals in the Paleocene Epoch at the beginning of the Cenozoic Era. The **Promimians** began to arise in the Eocene Epoch and the **Anthropoidea** arose in early Oligocene Epoch. This is a highly debatable data as different evolution biologists propose different theory for primate evolution.

To study the evolution of the primates, **GULO** (L-gulonolactone oxidase) gene was taken into consideration. The GULO enzyme produces vitamin C. It catalyzes the reaction of L-gulono-1,4-lactone with oxygen to L-xylo-hex-3-gulonolactone and hydrogen peroxide using FAD as a cofactor. L-xylo-hex-3-gulonolactone then converts to hexuronic acid (ascorbic acid AKA Vitamin C) (Fig. 6). If there is a deficiency of this enzyme, the production of vitamin C is halted. In primates, the loss of activity of GULO occurred around 63 million years ago at the time of split of the suborders **Haplorrhini** and **Strepsirrhini**. The Haplorrhine primates with simple nose do not possess the ability to synthesize the vitamin C but the Strepsirrhine primates which are wet nosed (including lemurs) make their own vitamin C enzymatically. Some of the Haplorrhines contain the GULO gene in an inactive or pseudo form called "the non-functional gulonolactone oxidase pseudogene: **GULOP**".

The possible reason for the loss of this enzyme is the presence of **Glut1** protein found on the membranes of the cells throughout the bodies. This Glut1 protein (on red blood cells) transports DHA and glucose both (strongly favoring DHA over glucose), with it, the presence of another membrane protein called **stomatin** is crucial for glucose transporters to switch to DHA. When scientists studied the RBC of organisms containing GULO in active form, they found that they contained Glut4 instead of Glut1. Thus, it was clear that due to presence of Glut1, the inability to generate the reduced version of DHA, vitamin C, from glucose could be linked to human red blood cells.

The **CFTR** (Cystic Fibrosis Transmembrane Conductance Receptor) gene found in primates also plays a role in defining the evolution. For the current study, it is taken for understanding the evolution of GULO gene with the help of the Ancestral State Reconstruction (ASR).

Method and Materials:

The data collected for this paper was simulated. The NCBI (National Center for Biotechnology Information) website was used to access the data. The CDS regions of the CFTR gene of 18 primates and 1 outgroup were downloaded in the FASTA format (Table S1). The sequences for the CFTR gene from all the 19 organisms were taken into one single FASTA file for MSA (Multiple Sequence Alignment). The MSA method used here was **Clustal Omega** in R using the "msa" package. The nucleotide substitution model with lowest BIC score was "GTR+G" and highest loglikelihood was "GTR+G+I" found using "phangorn" package in R. The output file was stored in NEXUS format.

This NEXUS file was uploaded on the LINUX cluster for performing **Bayesian Interference** using **MrBayes** for the **MCMC (Markov Chain Monte Carlo) method**. The MCMC process is a method for getting random samples from a distribution in order to be able to determine the properties of that distribution. Slurm scripts were used for carrying out the operation. The GTR+G model was chosen, with the outgroup as "*Mus_musculus*" and 100,000 iterations were carried out with the sample frequency to be 100. The *nex.con.tre output file from the Bayesian interference was shifted to remote device from server to carry out further steps. Bayesian output trees are obtained (2 from this data) and first one is chosen to be plotted using "phytools" package in R with stating the outgroup again as "*Mus_musculus*" (Fig. 2A).

The **Ancestral State Reconstruction (ASR)** with maximum parsimony was carried out for the GULO gene using R. The NEXUS tree output file from the Bayesian interference was used to plot the tree (Fig. 3) and the numeric states were assigned to all the species as per the active or inactive state of the GULO gene, 0 being the active state and 2 being the inactive state (Table S1). Custom matrix was implemented and the transition cost thus assigned was different for 4 different the transitions: a) from active GULO to the same = 0; b) from active GULO to inactive GULO = 1x; c) from inactive GULO to active GULO = 10x; and d) from inactive GULO to the same = 0. The pie matrix was created with the labels showing the ASR and thus plotted using "castor" package in R. (Fig. 5)

Results and supplementary materials:

Supplementary Table S1.

Summary of CFTR secuences analyzed

Sequences were downloaded from Genbank (www.ncbi.nlm.nih.gov).

Sr. No.	Organism	Common name	GULO presence	Accession Number (CFTR)	
				Protein	Nucleotide
1	*Aotus nancymaae*	*Nancy Ma's night m.*	No	*ABJ08890.1*	*DP000197.1*
2	*Ateles geoffroyi*	*Black-handed spider m.*	No	*ABI75275.1*	*DP000177.1*
3	*Callicebus moloch*	*Red-bellied titi*	No	*ABB89795.1*	*DP000019.1*

4	*Callithrix jacchus*	*white tufted ear marmoset*	No	*ABA90397.1*	*DP000014.2*
5	*Cercopithecus aethiops*	*Vervet monkey (Grivet)*	No	*ABC87491.1*	*DP000029.1*
6	*Colobus guereza*	*Black and white colobus or guezera*	No	*ABJ08857.1*	*DP000193.1*
7	*Gorilla gorilla*	*Gorilla*	No	*ABC87455.1*	*DP000025.1*
8	*Homo sapiens*	*Humans*	No	*NP_000483.3*	*NM_000492.3*
9	*Lemur catta*	*Ring tailed lemur*	Yes	*AAY88996.1*	*DP000004.1*
10	*Macaca fascicularis*	*Crab Eating Macaque*	No	*AAF80467.1*	*AH009552.1*
11	*Macaca mulatta*	*rhesus monkey*	No	*AAC14011.1*	*AF013753.1*
12	*Macaca nemestrina*	*Pig tailed Macaque*	No	*AAD46905.1*	*AH008050.1*
13	*Microcebus murinus*	*gray mouse lemur*	Yes	*ABB89826.1*	*DP000022.2*
14	*Mus musculus*	*Mouse*	Yes	*NP_066388.1*	*NC_000072.7*
15	*Nomascus leucogenys*	*White Cheeked Gibbon*	No	*ABJ08867.1*	*DP000194.1*
16	*Otolemur garnettii*	*Galago or bushbaby*	Yes	*ABA90408.1*	*DP000013.2*
17	*Pan troglodytes*	*Chimpanzee*	No	*NP_001073386.1*	*NM_001079917.1*
18	*Pongo abelii*	*Sumatran orangutan*	No	*ABC87466.1*	*DP000026.2*
19	*Saimiri boliviensis*	*Black capped squirrel m. or bolivian squirrel m.*	No	*ABI75310.1*	*DP000180.1*

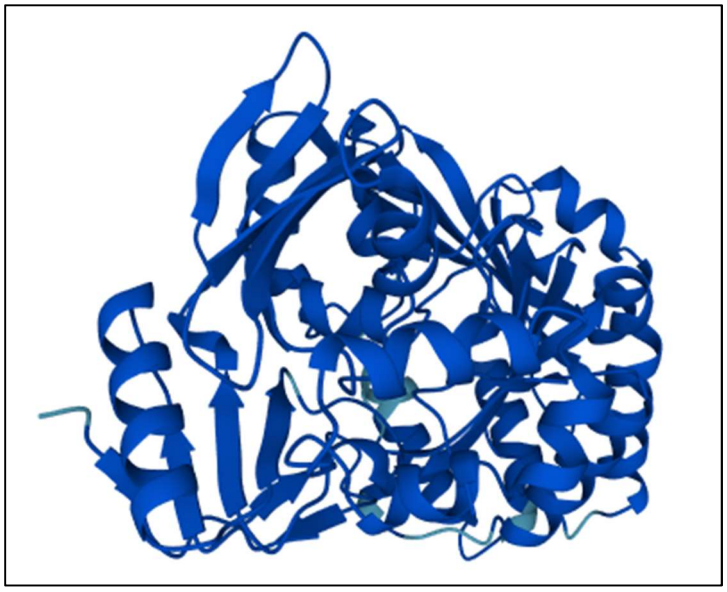

Fig. 1: structure of GULO for *Mus musculus* obtained through UniProt.

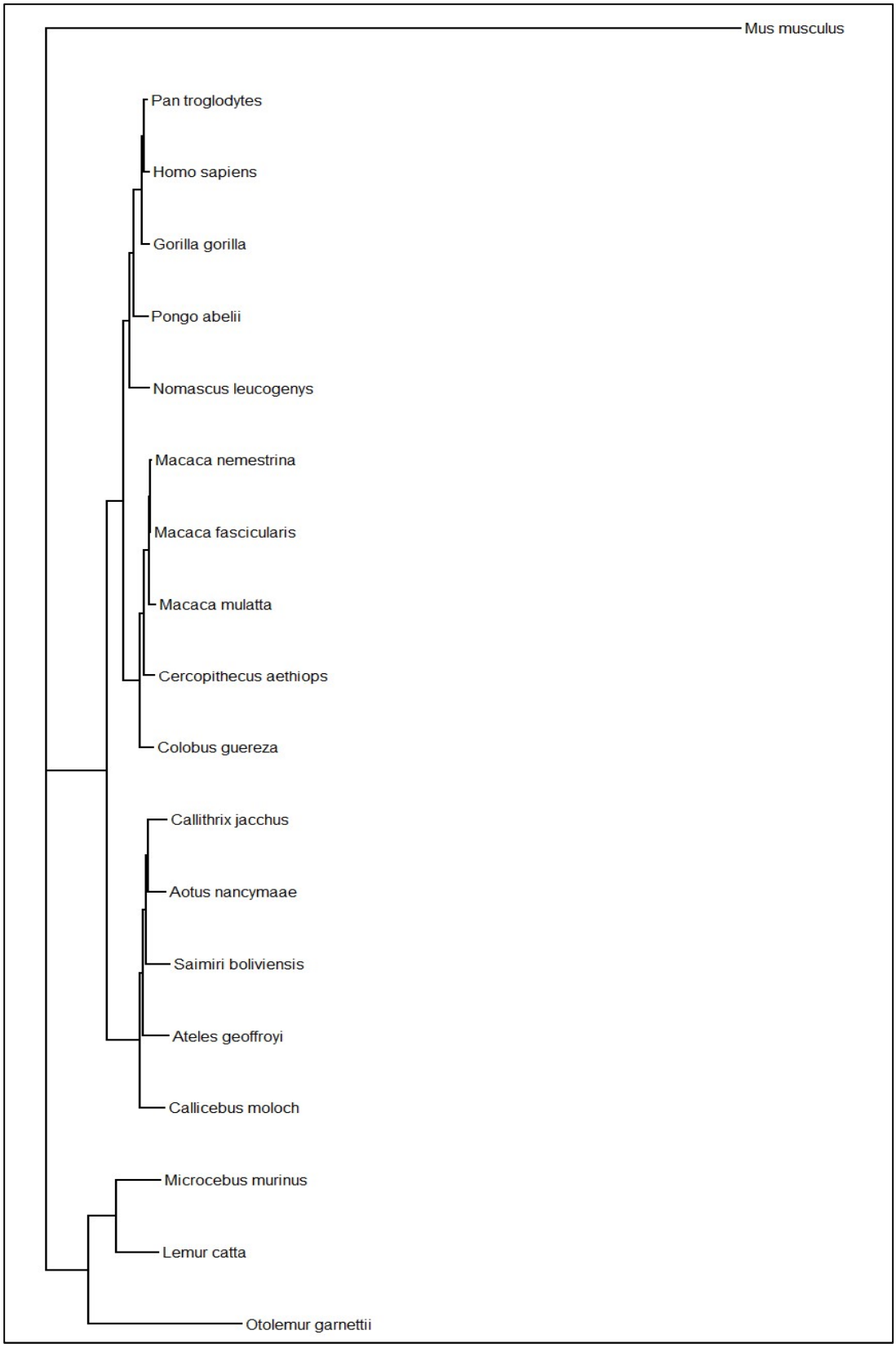

Fig. 2A: Bayesian Phylogenetic Tree

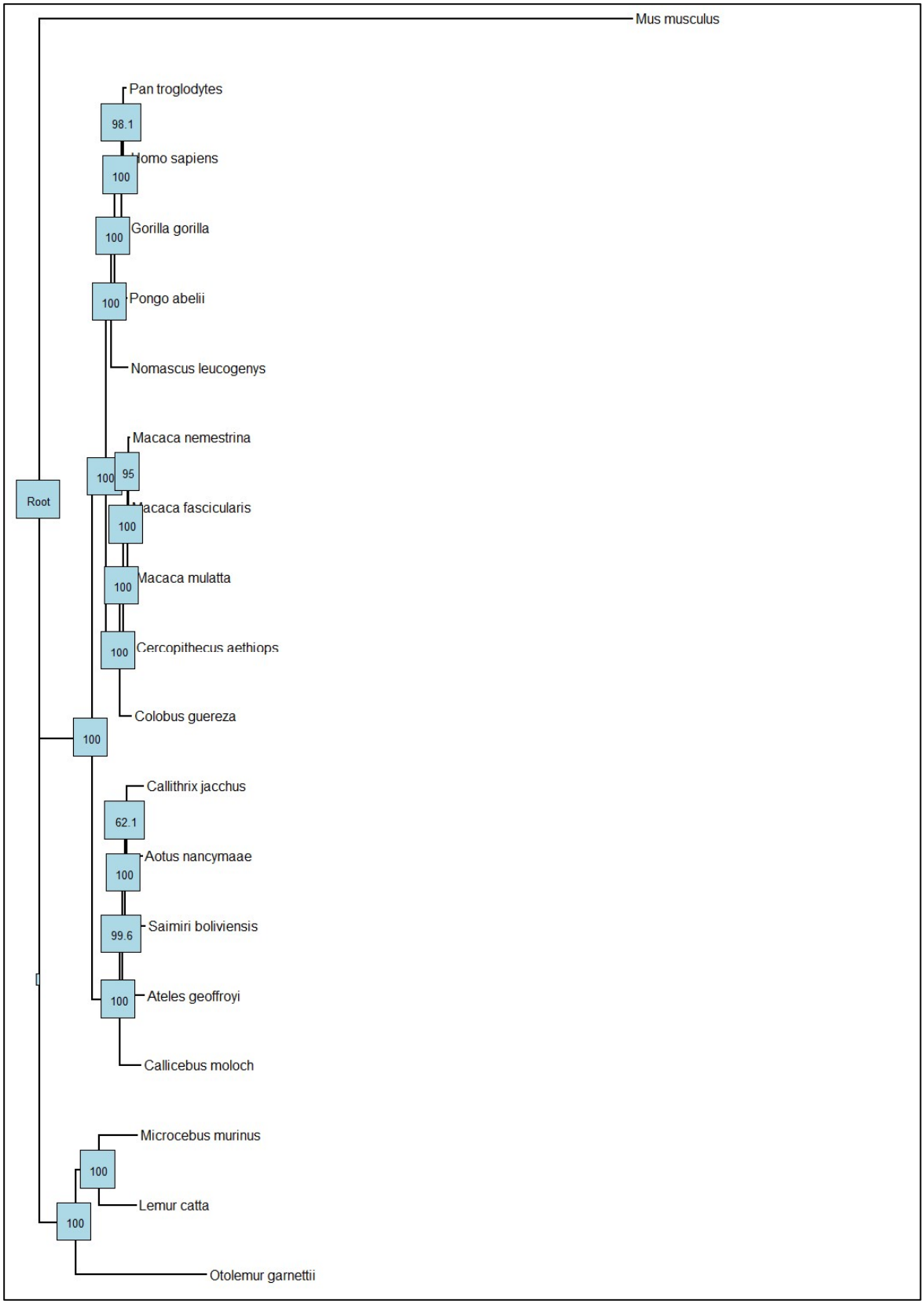

Fig 2B: Bayesian Phylogenetic Tree with the Bootstrap values

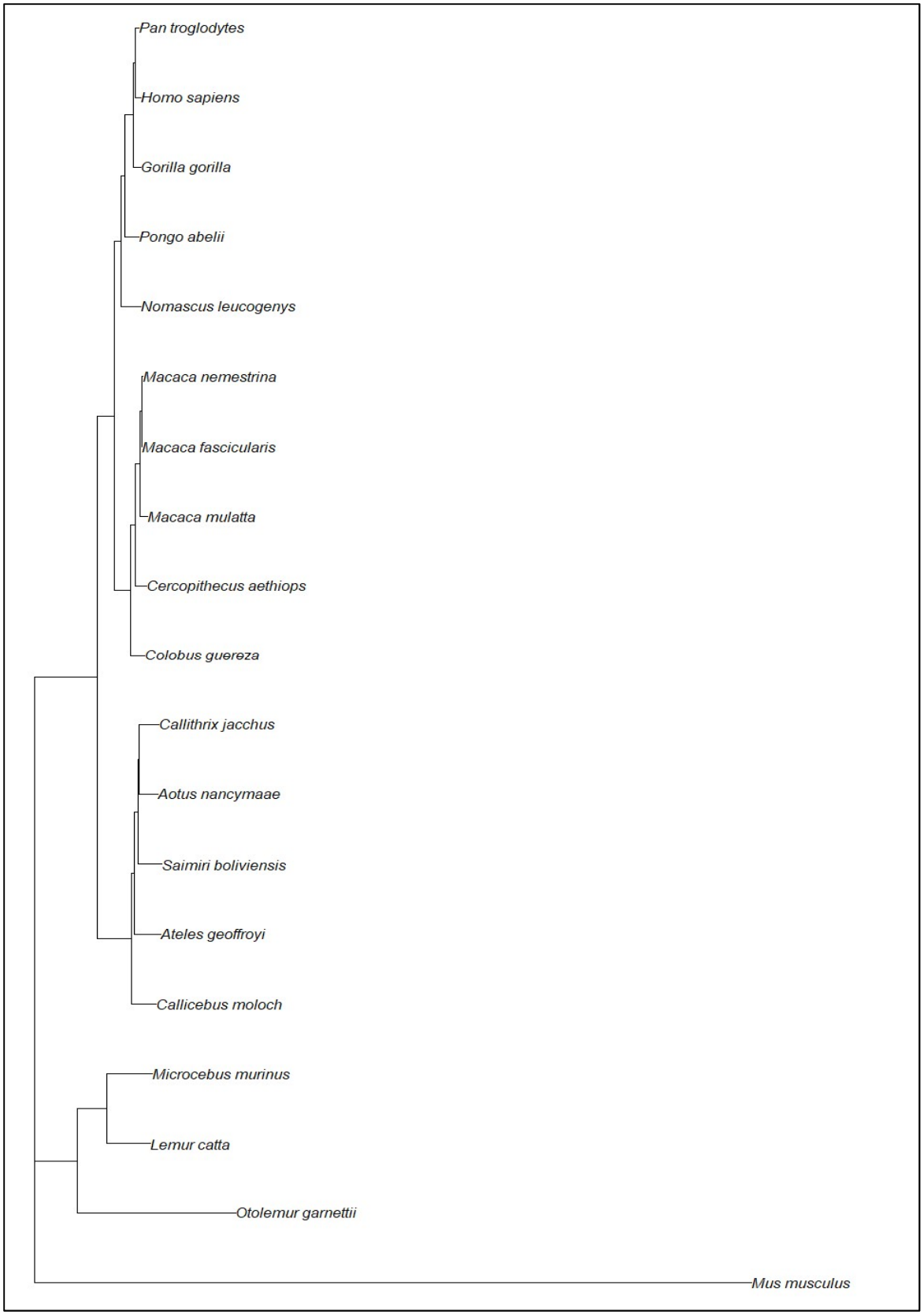

Fig. 3: Maximum Likelihood Tree of primates

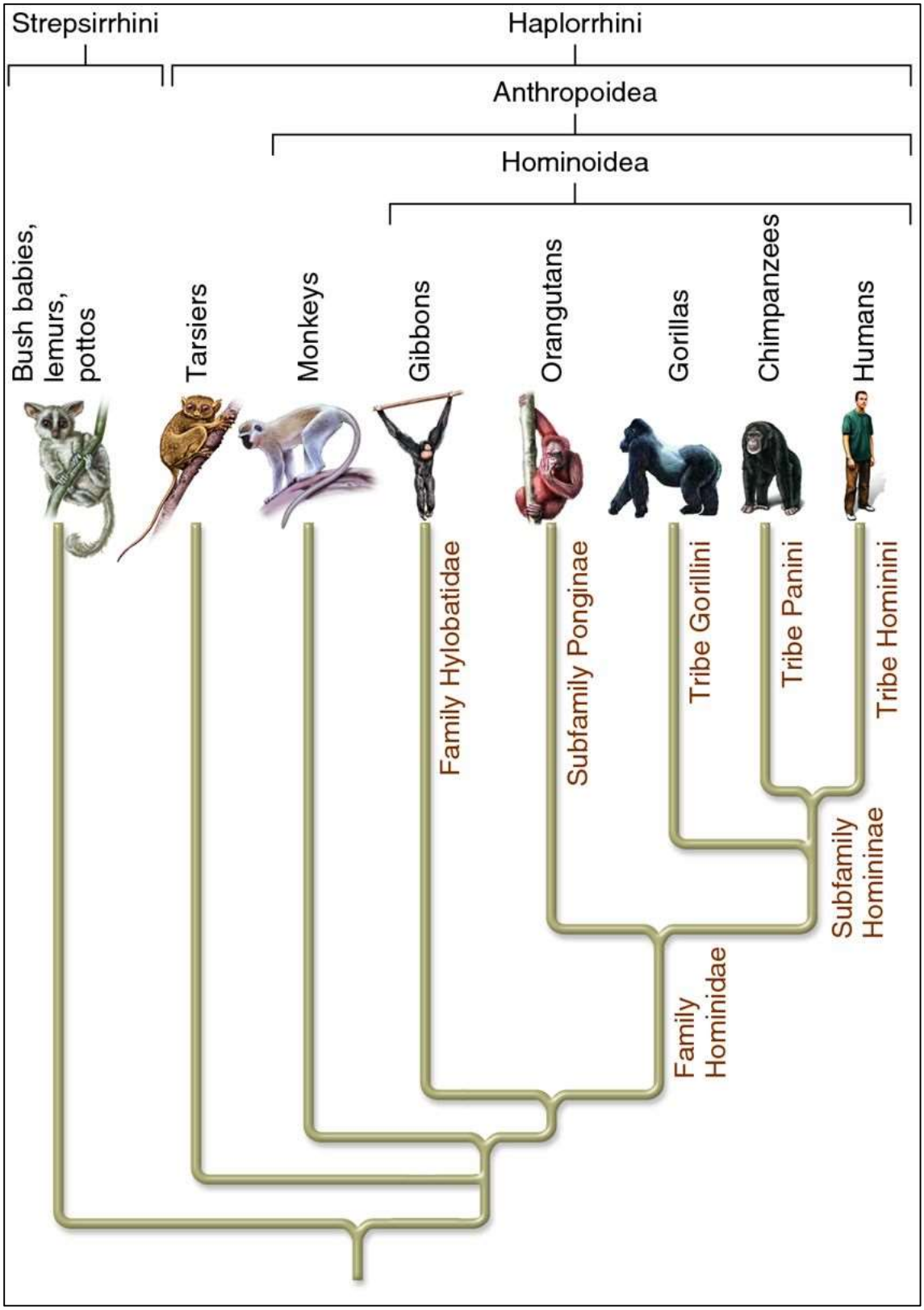

Fig. 4: Pictorial representation of the primate evolution tree

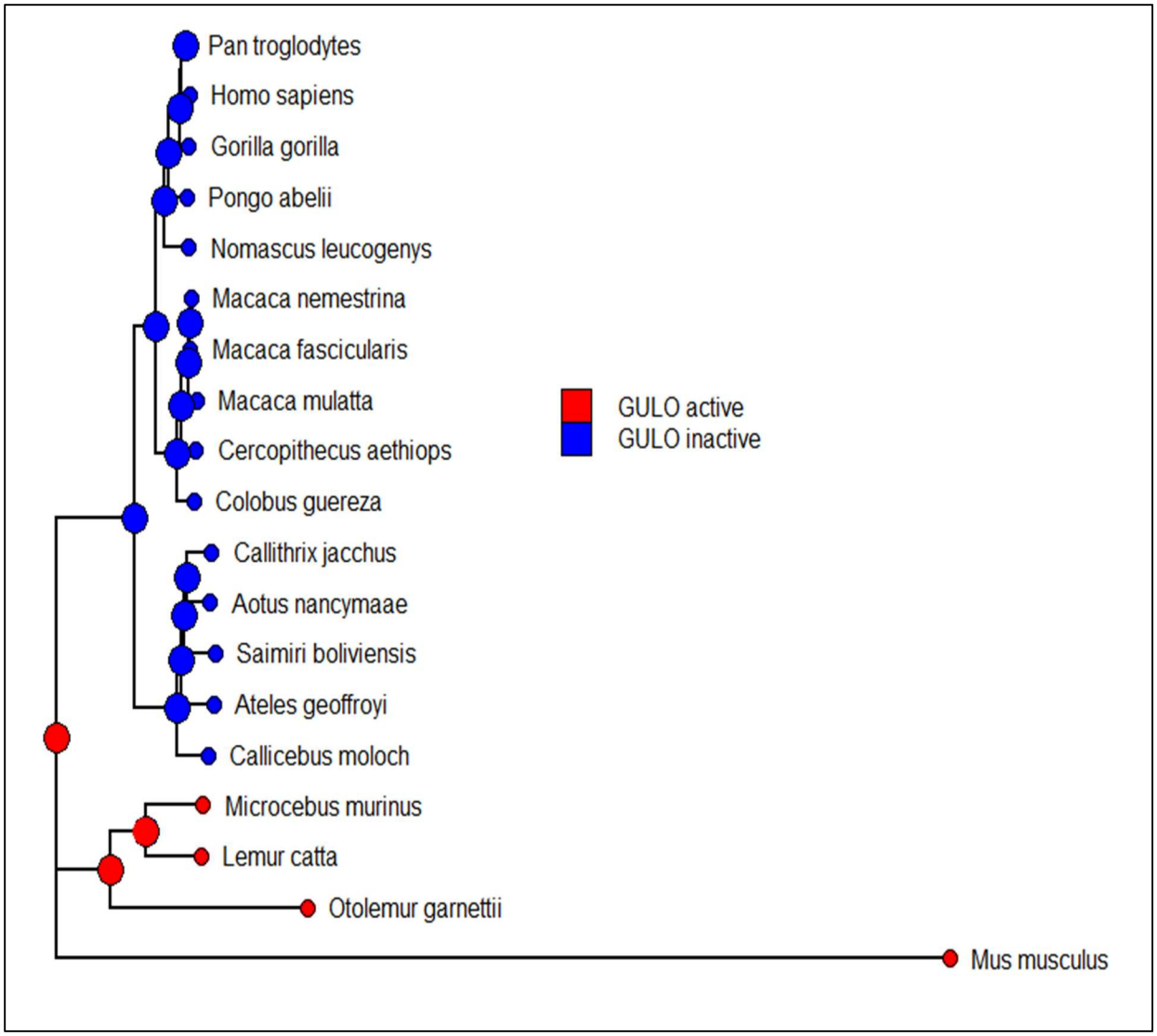

Fig. 5: Ancestral State Reconstruction for GULO gene

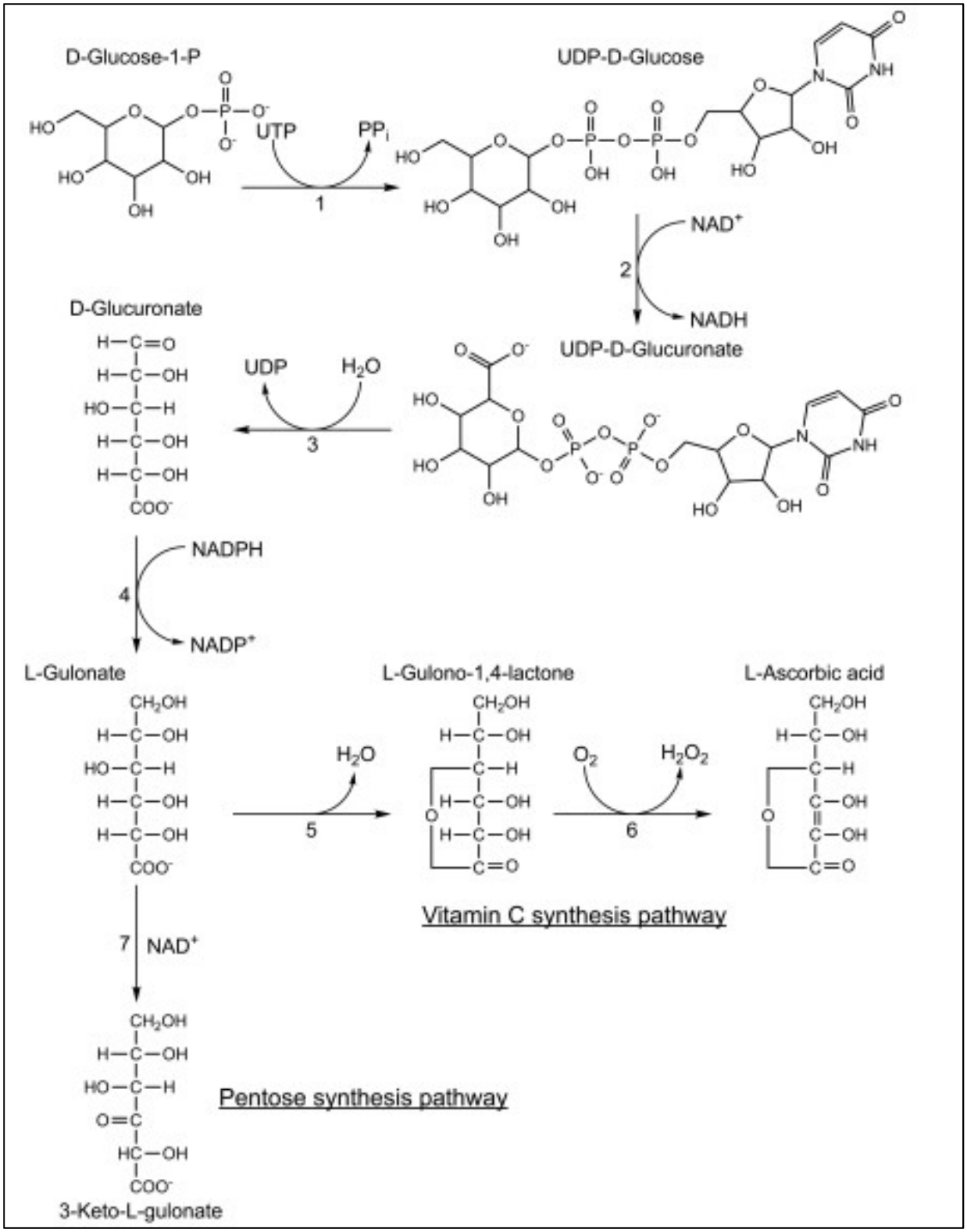

Fig. 6: Vitamin C synthesis pathway

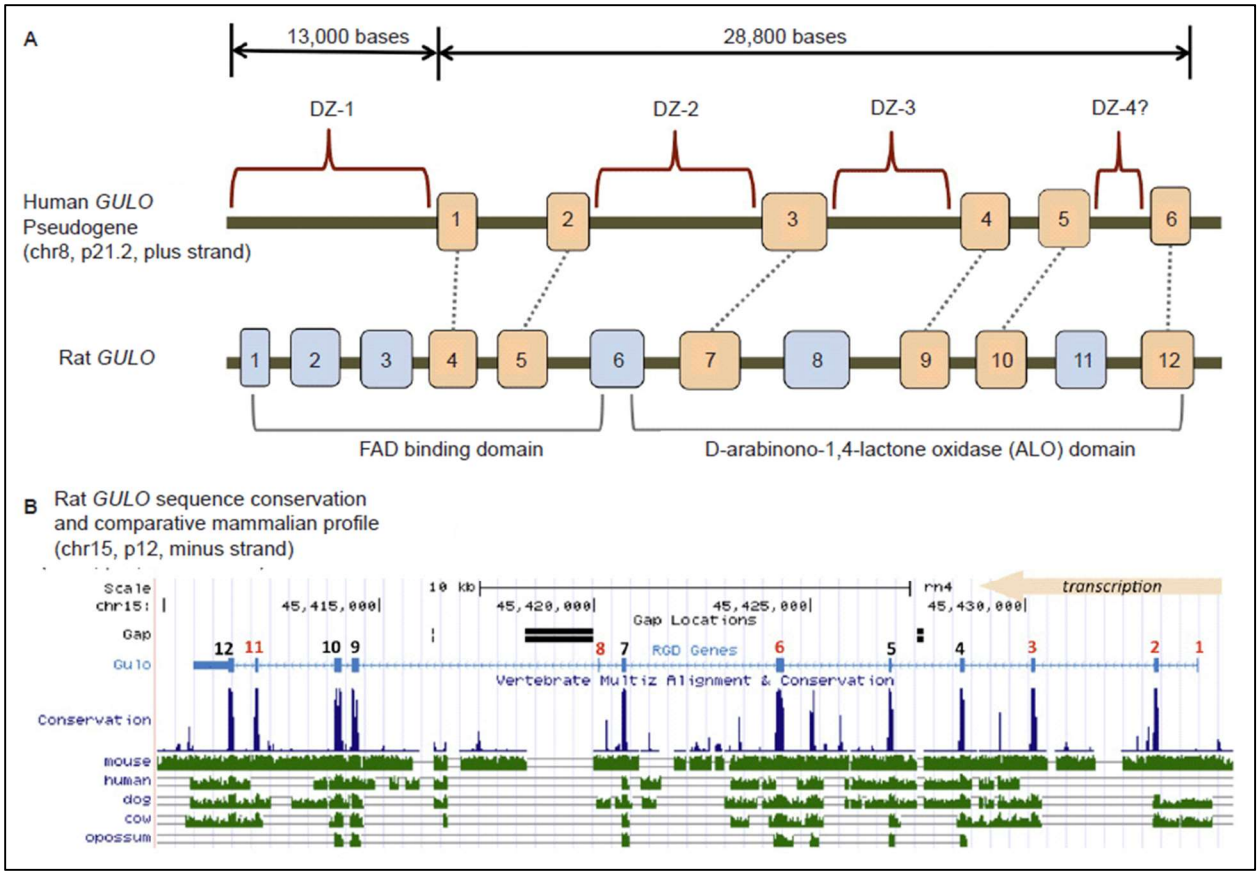

Fig. 7: Structure of the human *GULO* pseudogene. (A) Comparison of the human *GULO* pseudogene region to the rat functional *GULO* sequence (adapted in part from data in Yang (2013) and the UCSC genome browser). Exon 1 of the rat *GULO* gene is missing in other non-rodent mammals. (B) Comparison of several vertebrate *GULO* gene tracks to the functional rat *GULO* sequence showing the highly conserved nature of the exons (data extracted from the UCSC genome browser).

Discussions:

From the result of Bayesian phylogenetics, phylogenetic tree with maximum likelihood was obtained. (Fig 2A). The bootstrap values can be observed in the Fig. 2B, which indicate **how many times out of 100** the same branch was observed when repeating the phylogenetic reconstruction on a re-sampled set of the data. With high bootstrap values, we can see that the tree thus obtained is having higher confidence level of the clade.

The outcome of the Ancestral State Reconstruction for the GULO gene in the primates is shown in Fig. 5. We can understand the biosynthesis of Vitamin C (ascorbic acid) from the Fig. 6. The mutations in the GULO gene can be understood by comparing mouse functional and human non-functional GULO gene (Fig. 7).

The early ancestors possessed the GULO gene. During the split between the Strepsirrhini and the Haplorrhini suborders of the primates, the GULO gene is showing different traits in both these suborders. The Strepsirrhini (shown in red in Fig. 5) have the intact GULO gene while the Haplorrhini suborder (shown in blue in Fig. 5) seems to have lost the activity of GULO gene and thus the ability to synthesize their own Vitamin C.

Conclusion:

From the above result, we understand that the split of suborders Strepsirrhini and Haplorrhini plays a vital role in primate evolution. By studying the mutations in the course of evolution for the GULO gene in the primates, we can understand the evolution of the primates. The lemurs are capable of producing the Vitamin C on their own, even the outgroup, mice can also produce the same but the other primates cannot.

References:

1. Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A., ... & Pecon-Slattery, J. (2011). A molecular phylogeny of living primates. PLoS genetics, 7(3), e1001342.

2. Yang, H. (2013). Conserved or lost: molecular evolution of the key gene GULO in vertebrate vitamin C biosynthesis. Biochemical genetics, 51(5-6), 413-425.

3. Pizzo, L., Iriarte, A., Alvarez-Valin, F., & Marín, M. (2015). Conservation of CFTR codon frequency through primates suggests synonymous mutations could have a functional effect. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 775, 19-25.

4. Finley Sr, R. V. Is there good evidence for biological evolution?

5. Meredith, R. W., Janečka, J. E., Gatesy, J., Ryder, O. A., Fisher, C. A., Teeling, E. C., ... & Murphy, W. J. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. science, 334(6055), 521-524.

6. Kumar, S., & Hedges, S. B. (1998). A molecular timescale for vertebrate evolution. Nature, 392(6679), 917-920.

7. Reis, M. D., Gunnell, G. F., Barba-Montoya, J., Wilkins, A., Yang, Z., & Yoder, A. D. (2018). Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. Systematic Biology, 67(4), 594-615.

8. Perez, S. I., Tejedor, M. F., Novo, N. M., & Aristide, L. (2013). Divergence times and the evolutionary radiation of New World monkeys (Platyrrhini, Primates): an analysis of fossil and molecular data. PloS one, 8(6), e68029.

9. Thierry, B. (2013). Identifying constraints in the evolution of primate societies. Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1618), 20120342.

10. van Tuinen, M., & Hedges, S. B. (2001). Calibration of avian molecular clocks. Molecular Biology and Evolution, 18(2), 206-213.

11. Henriques, S. F., Duque, P., López-Fernández, H., Vázquez, N., Fdez-Riverola, F., Reboiro-Jato, M., ... & Vieira, J. (2019). Multiple independent L-gulonolactone oxidase (GULO) gene losses and vitamin C synthesis reacquisition events in non-Deuterostomian animal species. BMC evolutionary biology, 19(1), 1-12.

12. Zimmermann, E., & Radespiel, U. (2007). Primate life histories.

13. Horvath, J. E., & Willard, H. F. (2007). Primate comparative genomics: lemur biology and evolution. Trends in Genetics, 23(4), 173-182.

14. Sweigart, M. A., Zhu, C. F., Burt, D. M., DeHoll, P. D., Agrawal, C. M., Clanton, T. O., & Athanasiou, K. A. (2004). Intraspecies and interspecies comparison of the compressive properties of the medial meniscus. Annals of biomedical engineering, 32(11), 1569-1579.

15. Britannica, T. Editors of Encyclopaedia (2016, April 7). list of primates. Encyclopedia Britannica. https://www.britannica.com/topic/list-of-primates-2060305

16. Pozzi, L., Hodgson, J. A., Burrell, A. S., Sterner, K. N., Raaum, R. L., & Disotell, T. R. (2014). Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. Molecular phylogenetics and evolution, 75, 165-183.

17. Glazko, G. V., & Nei, M. (2003). Estimation of divergence times for major lineages of primate species. Molecular biology and evolution, 20(3), 424-434.

18. Raaum, R. L. (2015). Molecular evidence on primate origins and evolution. Handbook of paleoanthropology, 1083-1133.

19. Perez, S. I., Tejedor, M. F., Novo, N. M., & Aristide, L. (2013). Divergence times and the evolutionary radiation of New World monkeys (Platyrrhini, Primates): an analysis of fossil and molecular data. PloS one, 8(6), e68029.

20. Drouin, G., Godin, J. R., & Pagé, B. (2011). The genetics of vitamin C loss in vertebrates. Current genomics, 12(5), 371-378.

21. De Tullio, M. C. (2010). The mystery of vitamin C. Nat. Educ, 3(9), 48.

22. Tomkins, J. P. (2014). The Human GULO Pseudogene—Evidence for Evolutionary Discontinuity and Genetic Entropy. Answers Research Journal, 7, 91–101.https://answersresearchjournal.org/human-gulo-pseudogene-discontinuity/.