Investigation of Red Fox (Vulpes vulpes) Complete Miochondrial Genome and Phylogenetic Analysis with Canidae and Felidae Family

Name: Xiang He
Student Number: 1723303
Candidate Number: 48461

May 4, 2018

Abstract

To verify the phylogenetic relationship between the red fox and some other canids, two protein-coding genes (CYTB and CYTC subunit 1) were used to construct phylogenetic trees by the Jukes-Cantor model and Kimura two-parameter model, respectively. The results showed that the red fox and the arctic fox are in the same groups. They all belong to the red fox branch in the canine family, while the golden wolf and domestic dog belong to the wolf branch, which is consistent with the existing phylogenetic study results. Besides, the genetic distance between fox branch and cat branch are also measured to show that foxes belong to the Canine family rather than Felidae family.

1 Introduction

The red fox is the largest kind of the true foxes and one of the most widely distributed species of the order Carnivora, being present across the entire Northern Hemisphere from the Arctic Circle to North Africa, North America and Eurasia. It belongs to the Canine family. Two interesting questions will be explored in this report:

1. What is the genetic relationship between the red fox and some other canine species including golden wolf and dog?
2. Which family is more similar to foxes on the genetic distance level, Canidae family or Felidae family?

This report will do comprehensive research on the questions above with relative methods.

Mitochondrial DNA is selected as the primary data source since mitochondria are only passed down through the mother. Thus every individual will just have one version of mtDNA [3]. The complete mitochondrial genome sequence of the red fox (Vulpes vulpes) is available in NCBI [1]. Comparing complete animal mitochondrial genome sequences is now common for phylogenetic reconstruction and as a model for genome evolution. Like most mammals, the entire mitochondrial genome of the red fox contains 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes, and one control region. In addition to the unusual ATT for the ND3 gene start codon, the mitochondrial protein codes of the red fox and arctic fox, wolf, domestic

dog, and coyote follow the same pattern. A more extended AC-rich random repeat was found between conserved sequence segments 1 and 2 in the control region. [20] And the basic information of red fox mtDNA is shown below:

Species: Vulpes vulpes (red fox)
Accession Number: NC_008434 [1]
Base Counts: A: 5257 (31.3%), C: 4378 (26.1%), G: 2502 (14.8 %), T: 4676 (27.8%)
Content Plot: shown in Figure 1

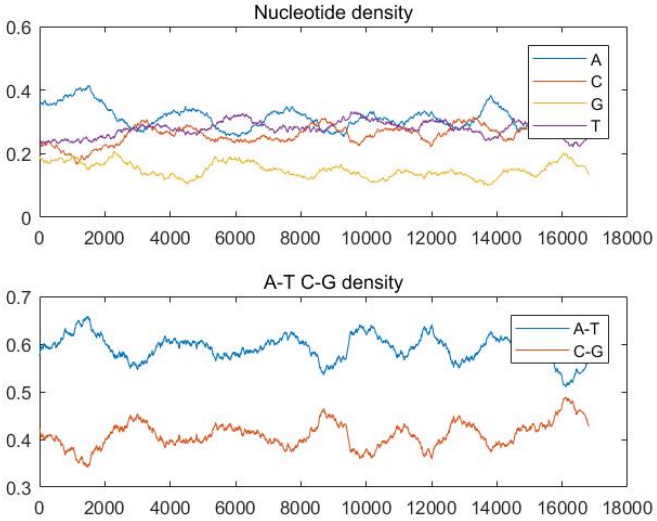


Figure 1: GC-content of Vulpes vulpes mitochondrion genome

To perform a phylogenetic analysis, cytochrome b (CYTB) and cytochrome c oxidase subunit 1 (COX1) are selected to perform a specific evaluation. Their amino acid sequences (first 50) are given below:

CYTB Protein: "MTNIRKTHPLAKIVNDSFIDLPAPSNISAWWNFGSLLGVCLILQI ATGLF"
COX1 Protein: "MFINRWLFSTNHKDIGTLYLLFGAWAGMVGTALSLLIRAELGQPG TLLGD"
This report will use the mtDNA of red foxes to discuss the red fox mitochondrial genome structure and evolution. A phylogenetic tree of some related species based on the Basic Local Alignment Search Tool (BLAST) [5] results of the CYTB and COX1 proteins will be reconstructed to confirm the evolution position of red fox and its relationship with some Felidae species.

2 Methodology

First of all, to explore the relationship between red foxes and some other canine species, CYTB and COX1proteins BLAST [5] is performed to search a few related species. To make the species more various, the selected species are not all on the top of the results. The CYTB BLAST

results are given in Table 1. And to explore the polymorphic sites that exist in these two proteins, the multiple alignments of the selected species are shown in Figure 2 and Figure 3.

Table 1: BLAST Results of CYTB Protein

Accession	Taxonomic Name	Common Name	Protein Name	Scores	Identities
NC_026529 [16]	Vulpes lagopus	Arctic fox	Cytochrome b	742	98%
NC_027956 [2]	Canis anthus	African golden wolf	Cytochrome b	730	96%
HM106324 [17]	Martes americana	American marten	Cytochrome b	704	92%
MG256392 [18]	Nyctereutes procyonoides	Raccoon dog	Cytochrome b	700	94%
None	Herpestes ichneumon	Egyptian mongoose	Cytochrome b	694	90%
NC_028305 [9]	Prionailurus viverrinus	Fishing cat	Cytochrome b	693	90%
NC_002391 [13]	Talpa europaea	European mole	Cytochrome b	692	89%

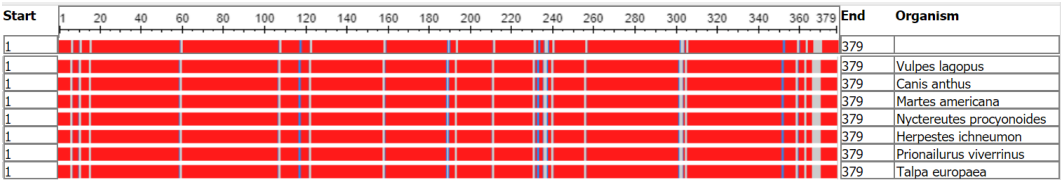


Figure 2: Multiple Alignment Results of CYTB Protein

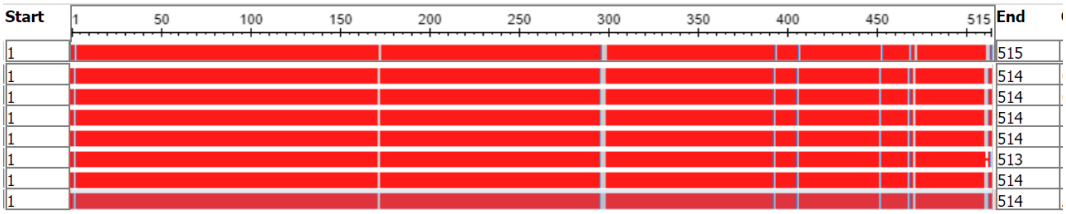
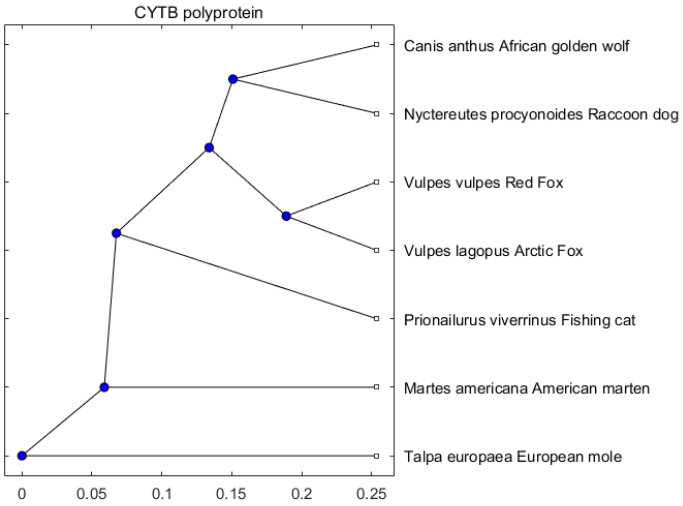


Figure 3: Multiple Alignment Results of COX1 Protein



Except for Egyptian mongoose, the complete mtDNA genome of all other species are available and can be downloaded from GenBank. The rooted phylogenetic tree [10] based on CYTB protein sequence for these species is generated and shown in the left. This genetic tree is generated by Jukes-Cantor [6] model. In [20], Zhong *et al.* conducted a phylogenetic analyses for red foxes by using neighbour-joining (NJ) [14] and maximum parsimony (MP) [4].

Actually, the NJ analysis was performed by utilising Kimura 2-parameter [7] as the nucleotide substitution model, which is based on more than just the one parameter, α, indicating the general probability of substitution [3]. On the contrary, the Jukes-Cantor model assumes that all of the substitutions are evenly likely, which is unrealistic for real-world applications. It is worth to mention that the Kimura model is only defined for nucleotide sequence and that is the reason I chose Jukes-Cantor here. And the MP analysis is an optimal criterion under which the phylogenetic tree will prefer to minimise the total number of character-state changes. Under the MP criterion, the optimal tree will minimise the amount of homoplasy eventually. Although this method is simple, no existed algorithm can generate the most-parsimonious tree. Usually, a heuristic search is conducted, which is also used in [20]. Since the phylogenetic tree that needed here is simple, this method is unnecessary.

There are some other popular distance metrics that have been widely applied in the gene distance measurement. One of them is Tajima-Nei distance [19]. In real data, nucleotide frequencies often deviate substantially from 0.25. In this case, the Tajima-Nei distance gives a better estimate of the number of nucleotide substitutions than the Jukes-Cantor distance. Note that this assumes equality of substitution rates among sites and between transitional and transversional substitutions [1]. For this task, the Tajima-Nei model gives us the same result of the Kimura model. Thus the result is not provided here.

To explore the relationships between red fox and Felidae family, some Felidae species including Felis catus (Domestic cat, `NC_001700` [11]), Leopardus pardalis (Ocelot, `NC_028315` [9]) and Panthera tigris (Tiger, `EF551003` [8]) are added into the phylogenetic tree. And the crocodile species is selected as an outgroup to make a comparison. Besides, only the amino acid sequence of CYTB protein is not sufficient to generate a meaningful genetic tree. Thus, another protein: COX1 is extracted and applied to every species. Also, to analyse the characteristics from genetic level, nucleotide sequence and amino acid sequence are both used to generate the corresponding phylogenetic trees. The distance of nucleotide sequences are measured by Kimura 2-parameter model and the unweighted pair group method using arithmetic averages, or 'UPGMA' method, is used for the hierarchical clustering. The distance of amino acid sequences are evaluated by Jukes-Cator model and the weighted pair group method average, or 'WPGMA' is employed for

[1] https://www.megasoftware.net/web_help_7/hc_tajima_nei_distance.htm

the corresponding hierarchical clustering. The produced trees are provided in Figure 4 and 5. Some of the distance values are provided in Table 2. In [20], the consensus phylogenetic tree is evaluated with 12 concatenated heavy-strand protein-coding genes, which is much more reliable than using only two proteins. However, since time is limited, more complete research will be performed in the future.

Table 2: COX1 protein Sequence Distances

Species	Red Fox	Arctic Fox	Golden wolf	Domestic cat	River crocodile
Red Fox	0	0.0137	0.0376	0.0519	0.1743
Arctic Fox	0.0137	0	0.0356	0.0622	0.1696
Golden wolf	0.0376	0.0356	0	0.0539	0.1837
Domestic cat	0.0519	0.0622	0.0539	0	0.1864
River crocodile	0.1743	0.1696	0.1837	0.1864	0

Considering that four phylogenetic trees give us different results on the genetic relationships of selected species, a consensus phylogenetic tree is created for better analysis. The various trees are weighted by their total distance values.

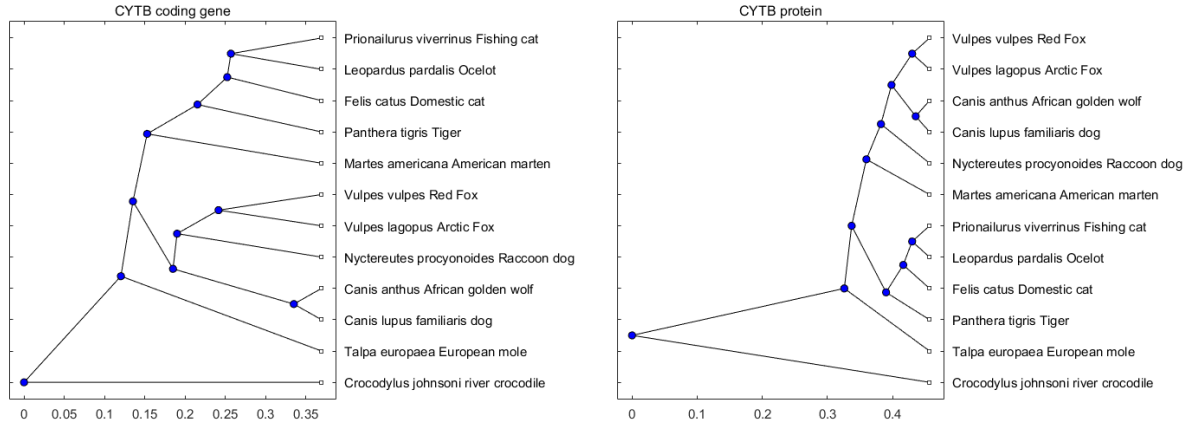


Figure 4: CYTB NT and AA Phylogenetic Tree

3 Results Analysis

First, the mitochondrial genome sequence of red fox shows that it is an A-T rich (nearly 60%) sequence. According to [20], A-T rich was found not just in the whole genome level, but also in rRNA and tRNA genes. Some other canine family including golden wolf and dog also reveal a high-proportion A-T base.

For further investigations of red foxes and other Canidae animals, we can see from the multiple alignments results that sequence polymorphism are common in the gene. Usually, the point

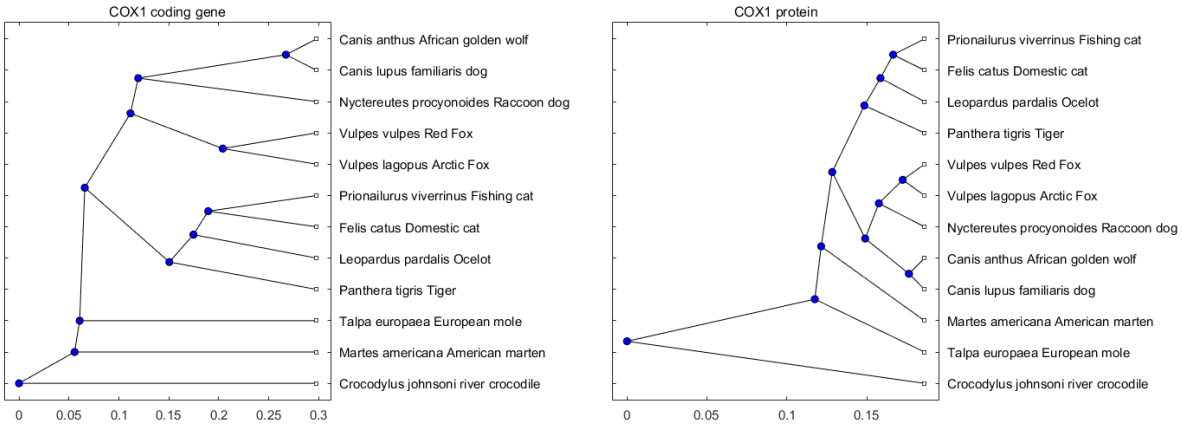


Figure 5: COX1 NT and AA Phylogenetic Tree

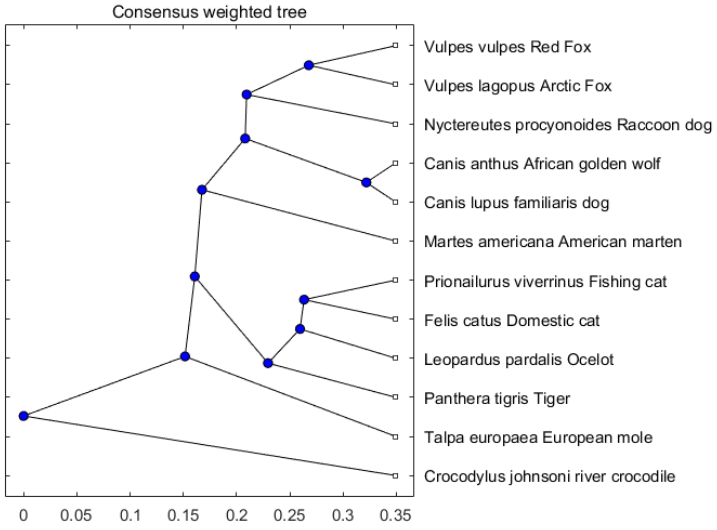


Figure 6: Consensus Phylogenetic Tree

mutations occur in some specific locations (SNP) and lead to different translated amino acids. The polymorphism can be divided into transitions and transversions, which are quite essential to build a distance measurement model. Sequences differences between species are referred as substitutions. Species of Canidae animals always show same transitions, which indicates that they are very close to the gene level. Species among different branches can give various transitions and transversions on the SNPs. To measure their distance between each other, a more specific measurement is essential.

The vertical order of taxa in the phylogenetic tree is meaningless, as only the branch path between them reflects their degree of similarity. According to the BLAST results of CYTB protein, it is easy to conclude that CYTB is common in vertebrates. Both Canidae and Felidae are founded in the BLAST results, and their similarity scores are all in a high-level. And the phylogenetic tree of these species indicates that their relationships with each other are quite

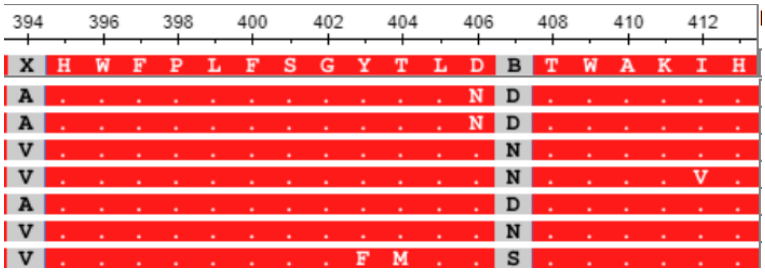


Figure 7: Sequence Polymorphism

close. It is not surprising that arctic fox is the sister group of red fox and they both belong to the same clade. Golden wolf and raccoon dog fall into another group which is close to the fox group. Fishing cat has a notable distance with all those Canidae species. And European is the most dissimilar species with Canids, which shows the same result with the BLAST results (89% identity).

In reference to the BLAST [5] results of red fox CYTB protein, fishing cat is less similar to the red fox species than other Canid species such as the golden wolf. It is evident from all of the produced trees that foxes are more identical to Canidae family rather than Felidae animals. Official taxonomy shows that fox belongs to family Canidae, which is same as this investigation result.

Based on the produced consensus phylogenetic tree, again, red fox and arctic fox fall into the same branch, and golden wolf and dog fall into another branch. This result is consistent with the conclusion of [20] (see Figure 8). The arrangement of multiple genes in red fox is in line with most vertebrates, and the overall organisation of the canids genome is nearly the same, which demonstrates that the mitochondrial genome is evolutionary highly conserved. The branch point (also called an internal node) represents a divergence event or splitting apart of a single group into two descendant groups. Raccoon dog is regarded as a basal Canid species, resembling ancestral forms of the Canidae family. And that is reason raccoon dog is neither in the wolf clade or in the fox clade, which is different from the first basic tree but shows a similar result with Sillero *et al.* ś work [15]. The result of this investigation gives exactly a same result with the official taxonomy, which is provided in the left.

Although the consensus tree is a stable result, there are several differences between the four phylogenetic trees. In CYTB gene tree and COX1 protein tree, Canidae species are more similar to the European mole. But in another two trees, Felidae animals are closer to it. Also, the

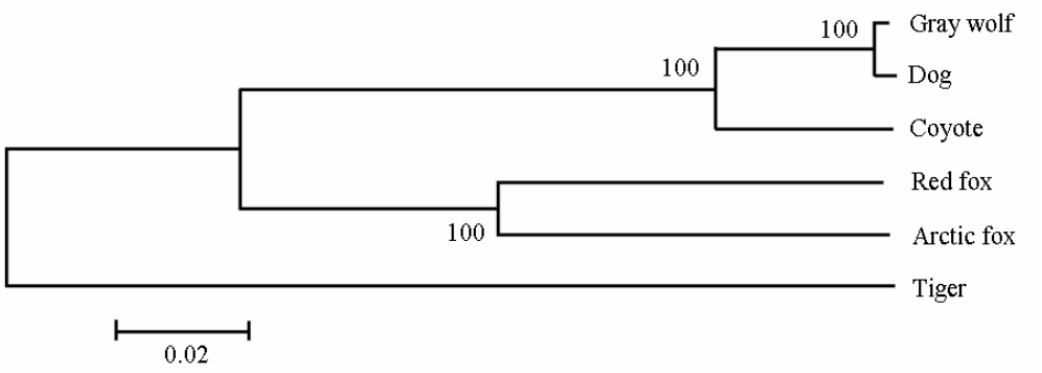


Figure 8: Phylogenetic Tree from [20]

positions of American marten are different. The remaining part of the trees is nearly the same. Considering that these four trees are generated from different kinds of sequences and different distance model, it is normal to obtain different results. Besides, the outgroup in the trees all shows a big difference from other species, which indicates that the results are reliable to some extent.

4 Conclusions

From this research, we can conclude that red fox and arctic belong to the same branch (fox) of Canidae family, which are also very similar to wolf branch. And compared to some Felidae species, red fox species is significantly closer to Canidae species, which can be demonstrated by the taxonomy. Also, the conclusion of this report is consistent with a relevant article [20]. However, the methods used in this report have some limitations. The number of protein-coding genes used are too small, which may lead to an unreliable result. And genes used in the phylogenetic analysis are all from one mitochondrial genome. This will only provide one independent estimate of the species tree [12]. Thus, further investigations are necessary to obtain more data (e.g. nuclear DNA) to analyse the gene more comprehensively.

References

[1] U. Arnason, A. Gullberg, A. Janke, M. Kullberg, N. Lehman, E. A. Petrov, and R. Väinölä. Pinniped phylogeny and a new hypothesis for their origin and dispersal. *Molecular Phylogenetics and Evolution*, 41(2):345 – 354, 2006.

[2] D. Bertè. Remarks on the skull morphology of canis lupaster hemprich and herenberg, 1832 from the collection of the natural history museum “g. doria” of genoa, italy. 4:19–29, 03 2017.

[3] N. Cristianini and M. W. Hahn. *Introduction to Computational Genomics: A Case Studies Approach*. Cambridge University Press, New York, NY, USA, 2007.

[4] G. Jin, L. Nakhleh, S. Snir, and T. Tuller. Maximum likelihood of phylogenetic networks. *Bioinformatics*, 22(21):2604–2611, 2006.

[5] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, and T. L. Madden. Ncbi blast: a better web interface. *Nucleic Acids Research*, 36(suppl_2):W5–W9, 2008.

[6] T. Jukes and C. Cantor. Evolution of protein molecules. *Munro HN, editor, Mammalian Protein Metabolism*, pages 21–132, 1969.

[7] M. Kimura. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16(2):111–120, 1980.

[8] W. Lei, W. XiaoBing, L. Zhu, and Z. Jiang. Mitogenomic analysis of the genus panthera. *Science China Life Sciences*, 54(10):917–930, Oct 2011.

[9] G. Li, B. W. Davis, E. Eizirik, and W. J. Murphy. Phylogenomic evidence for ancient hybridization in the genomes of living cats (felidae). *Genome research*, 26 1:1–11, 2016.

[10] B. C. Livezey and R. L. Zusi. Higher-order phylogenetics of modern aves based on comparative anatomy. *Netherlands Journal of Zoology*, 51(2):179–205, 2001.

[11] J. V. Lopez, S. Cevario, and S. J. O'Brien. Complete nucleotide sequences of the domestic cat (felis catus) mitochondrial genome and a transposed mtdna tandem repeat (numt) in the nuclear genome. *Genomics*, 33(2):229 – 246, 1996.

[12] W. S. Moore. Inferring phylogenies from mtdna variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, 49(4):718–726, 1995.

[13] S. K. Mouchaty, A. Gullberg, A. Janke, and U. Arnason. The phylogenetic position of the talpidae within eutheria based on analysis of complete mitochondrial sequences. *Molecular Biology and Evolution*, 17(1):60–067, 2000.

[14] N. Saitou and M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4):406–425, 1987.

[15] C. Sillero, M. Hoffmann, and D. Macdonald. Canids: Foxes, wolves, jackals and dogs. status survey and conservation action plan. page 430, 01 2004.

[16] S.-Q. Yan, P.-C. Guo, Y. Yue, W.-H. Li, C.-Y. Bai, Y.-M. Li, J.-H. Sun, and Z.-H. Zhao. The complete sequence of the mitochondrial genome of arctic fox (alopex lagopus). *Mitochondrial DNA Part A*, 27(6):4095–4096, 2016. PMID: 25629488.

[17] L. Yu, D. Peng, J. Liu, P. Luan, L. Liang, H. Lee, M. Lee, O. A. Ryder, and Y. Zhang. On the phylogeny of mustelidae subfamilies: analysis of seventeen nuclear non-coding loci and mitochondrial complete genomes. *BMC Evolutionary Biology*, 11(1):92, Apr 2011.

[18] H. Zhang and L. Chen. The complete mitochondrial genome of the raccoon dog. *Mitochondrial DNA*, 21(3-4):59–61, 2010.

[19] A. Zharkikh. Estimation of evolutionary distances between nucleotide sequences. *Journal of Molecular Evolution*, 39(3):315–329, Sep 1994.

[20] H.-M. Zhong, H.-H. Zhang, W.-L. Sha, C.-D. Zhang, and Y.-C. Chen. Complete mitochondrial genome of the red fox (vuples vuples) and phylogenetic analysis with other canid species. 31:122–30, 04 2010.