The complete mitochondrial genome of the Indochinese jackal (*Canis aureus cruesemanni*) and its relationship to other subspecies of golden jackal

Medhini S. Sosale[1,2]*, Nucharin Songsasen[3], Osman İbiş [4,5,6], Cody Edwards[2,7], Henrique V. Figueiró[2], Klaus-Peter Koepfli[2,3]

[1]Department of Bioengineering, Volgenau School of Engineering, George Mason University, Fairfax, Virginia, USA

[2]Smithsonian-Mason School of Conservation, George Mason University, Front Royal, Virginia, USA

[3]Center for Species Survival, Smithsonian's National Zoo and Conservation Biology Institute, Front Royal, Virginia, USA

[4]Department of Agricultural Biotechnology, Faculty of Agriculture, Erciyes University, Kayseri, Turkey

[5]Genome and Stem Cell Center (GENKOK), Erciyes University, Kayseri, Turkey

[6]Vectors and Vector-Borne Diseases Implementation and Research Center, Erciyes University, Kayseri, Turkey

[7]Department of Biology, George Mason University, Fairfax, Virginia, USA

*Corresponding author

medhini.sosale@gmail.com

22026 Sunstone Ct, Broadlands, VA 20148

Abstract

The Indochinese jackal (*Canis aureus cruesemanni*) is a subspecies of the golden jackal (*Canis aureus*) found in Southeast Asia. While this species has been genetically studied in Europe, the Middle East, and India, current research is lacking on the population(s) in Southeast Asia. Using a genome skimming approach, we assembled the first complete mitochondrial genome for an Indochinese jackal from Thailand. The mitogenome contained 37 annotated genes and is 16,729 bps in length. Phylogenetic analysis with 21 additional canid mitogenomes, along with analyses of a cytochrome *b* gene-only data set, supports the Indochinese jackal as a distinct lineage, and therefore subspecies, among golden jackals.

Keywords: *Canis aureus*; subspecies; phylogeny; mitochondrial genome; golden jackal

1. Introduction

The golden jackal, *Canis aureus* (Linnaeus, 1758), a member of the family Canidae, is found across southern Eurasia and is generally thriving (Krofel et al., 2017). It has been designated as a species of Least Concern by the IUCN Red List and has received CITES Appendix III protection in India (Hoffmann et al., 2018). The taxonomy of this species has been in flux in recent years due to genetic and genomic evidence supporting the recognition of golden jackal populations in Africa as a distinct species, the African wolf, *Canis lupaster* (Hemprich and Ehrenberg, 1832; Koepfli et al., 2015; Krofel et al., 2022). Within Eurasia, up to eight subspecies have been described, although their delimitations are not well defined (Moehlman and Hayssen, 2018). Among these is the Indochinese jackal, *Canis aureus cruesemanni* (Matschie, 1900), first described by German zoologist Paul Matschie from living specimens in the Berlin Zoological Garden that came from southwest Siam (now Thailand). In Thailand, this jackal inhabits dry

dipterocarp forests and other open landscapes, and like other golden jackals, it is omnivorous and active at twilight and night (Parr, 2003).

Recent studies on golden jackals employing microsatellites and/or partial mitochondrial DNA sequences have focused on analyzing genetic diversity and structure within regional populations (e.g., India, Yumnam et al., 2015; Iran, Yusefi et al., 2021) or across multiple populations, often with the aim to understand the origin of the population(s) that have been rapidly expanding in western Europe (Fabbri et al., 2014; Rutkowski et al., 2015; Spassov and Acosta-Pankov, 2019). Phylogeographic analyses have generally revealed only little to moderate genetic structuring, depending on the scale of the geographic sampling. However, none of these studies have so far included samples from the part of the species' range in Southeast Asia. The Indochinese jackal likely represents the easternmost population of the golden jackal (Moehlman and Hayssen, 2018). Therefore, we predicted that it may be genetically differentiated from western populations in the species' range. To test this, we report the sequencing, assembly, and annotation of the first complete mitochondrial genome of an Indochinese jackal from Thailand.

2. Materials and Methods

2.1 Indochinese jackal (Thailand) sample collection and preparation

As part of a field expedition investigating the ecology of wild canid species in Thailand, a male Indochinese jackal was live-trapped in the Salakpra Wildlife Sanctuary, Kanchanaburi Province, Thailand (latitude = 14.309912, longitude = 99.256454). A capture permit was given by the Department of National Park, Wildlife and Plant Conservation (permit#0907.4/17810). The jackal was estimated to be 1-2 years old based on dental evaluation. A 3 ml whole blood sample

was collected and stored at -80°C (NZCBI acuc # 14-01) in the personal collection of N. Songsasen (SongsasenN@si.edu) at the Smithsonian's National Zoo and Conservation Biology Institute (https://nationalzoo.si.edu/center-for-species-survival). An aliquot of whole blood was delivered to Psomagen, Inc. (Rockville, MD) for DNA extraction, library preparation, and sequencing. Genomic DNA was extracted using the Mag-Bind Blood and Tissue Kit (Omega Bio-Tek Inc., Norcross, GA) and evaluated for quality and concentration with a Picogreen and Victor X2 fluorometry assay (Life Technologies, Carlsbad, CA), an Agilent 4200 Tapestation (Agilent Technologies, Santa Clara, CA), and 1% gel electrophoresis. DNA was sheared into 350 bp fragments with a Covaris S220 ultrasonicator (Woburn, MA) and used to prepare a genomic library with the TruSeq DNA PCR-free library kit (Illumina, San Diego, CA). The library was quality checked on an Agilent 4200 Tapestation, quantitated via quantitative PCR using a Lightcycler (Roche Life Science, St. Louis, MO), and then paired-end sequenced (2 x 150 bp) on an Illumina NovaSeq 6000 instrument to a depth of 20x. A total of 374,172,520 reads were generated, 92.2% of which had a ≥Q30 score.

2.2 Indochinese jackal (Canis aureus cruesemanni) mitogenome assembly

Raw reads were evaluated using FastQC (Andrews, 2010) and then subsampled to 40 million reads using BBMap version 38.96 (Bushnell, 2014). Subsampled reads were trimmed and filtered using AdapterRemoval (Lindgreen, 2012) within PALEOMIX version 1.3.6 (Schubert et al, 2014). The read set was then mapped to the reference mitochondrial genome of the gray wolf (*Canis lupus*, Björnerfeldt et al., 2006; GenBank: DQ480505) with the Geneious mapper using medium-low sensitivity and five iterations of fine-tuning in Geneious Prime version 2022.0.2 (https://www.geneious.com). Annotation of the mitogenome assembly was performed using the

MITOS2 webserver (Donath et al., 2019) and yielded 22 tRNA regions, 2 rRNA regions, 13 protein-coding gene regions, and a control region containing the D-loop (Figure 1). The Codon Usage webserver (https://www.bioinformatics.org/sms/index.html) was used to calculate the number and frequency of each codon type for the 13 protein-coding genes in the Indochinese jackal in comparison to the gray wolf reference mitogenome. Specifically, protein coding gene sequences were extracted from mitogenomes using Geneious Prime, edited to remove incomplete stop codons, and then concatenated before being analyzed in the webserver.

*2.3 Collection and sequencing of golden jackal (*Canis aureus*) sample from Turkey*

We also sequenced the complete mitochondrial genome of a female young adult golden jackal from Dagbeli, Antalya, Turkey (N 37°15’26.04’’ – E 30° 29’ 36.89’’, altitude 785 meters), representing the subspecies *C. a. moreotica*. This sample was obtained via muscle tissue collected from a road-killed specimen on 26 July 2021 (Ericeyes University collection number 1886). Since the specimen was road-killed, ethical approval is not necessarily required, but the sample was still approved by the Local Ethical Committee of Laboratory Animal Experimentation at Erciyes University (Protocol Nr.: 14/126, Date: September 10, 2014). The mitogenome was generated following the same methods described in İbiş et al. (2020) using mitogenome-specific primers. A total of 542,318 raw reads with 88.1% having a score over Q30 were generated. Read processing, mapping, and assembly were performed using the same methods described above for the Indochinese jackal, except 25 iterations of fine-tuning were used during mapping in Geneious Prime.

2.4 Multiple sequence alignment

Mitogenomes of 18 other canid species and two golden jackals putatively representing *Canis aureus syriacus* (Israel) and *Canis aureus indicus* (India) were downloaded from GenBank (Table 1) and imported into Geneious Prime 2022.0.2, along with the mitogenomes of the Indochinese jackal and golden jackal from Turkey. Putative subspecies designations of golden jackals used in this study followed Moehlman and Hayssen (2018). We generated a multiple sequence alignment using MAFFT version 7.450 (Katoh and Standley, 2013) with default options (algorithm = AUTO, scoring matrix = 200 PAM/k=2, gap open penalty = 1.53, offset value = 0.123). Due to poor alignment in the repetitive regions, we trimmed the control region from the alignment. The resulting 15,958 bp alignment was used to construct a maximum-likelihood phylogeny with RAxML version 8.2.11 (Stamatakis, 2014) using the rapid hill-climbing algorithm and GTR+GAMMA model of substitution. One hundred (100) bootstrap replicates were subsequently employed to calculate node support (bootstrapping using rapid hill-climbing, random seed = 1 setting in Geneious Prime).

2.5 CYTB sequence alignment and analysis

To place the Indochinese jackal into a wider phylogeographic context, we also analyzed a data set comprised of complete cytochrome *b* (*CYTB*) gene sequences. A total of 18 *CYTB* sequences were downloaded from GenBank (Table 2) and imported into Geneious Prime. The final taxon set included *Canis aureus* (n=13), with sequences representing animals sampled from Afghanistan, Egypt, India, Israel, and Serbia, plus the *CYTB* sequences from the jackals from Thailand and Turkey; *Canis lupaster* (n=4), and *Canis lupus* (n=3). As with the mitogenome sequences, the *CYTB* sequences were aligned with MAFFT version 7.450 (Katoh and Standley,

2013) using default settings. We employed RAxML version 8.2.11 (Stamatakis, 2014) using the rapid hill-climbing algorithm and GTR+GAMMA model of substitution to estimate the maximum-likelihood phylogeny from the 1,140 bp alignment. Node support was calculated using 100 bootstrap replicates as described above for the mitogenome data set. Finally, we used the *CYTB* sequences to construct a haplotype network using TCS with the program PopART (Leigh and Bryant, 2015).

3. Results and Discussion

3.1 Mitogenome information

For the mitogenome of the Indochinese jackal from Thailand, we extracted a 16,729 bp consensus sequence, which had an average coverage of 231x. For the golden jackal from Turkey, a 16,669 bp consensus sequence was extracted with an average coverage of 4,125x. The difference in length between the two mitogenomes is accounted for by differing numbers of repetitive elements in the control region, a common feature in all vertebrate mitogenomes (Formenti et al., 2021). Annotation of both mitogenome assemblies resulted in 13 protein-coding genes, 2 rRNAs, 22 tRNAs, and the control region, which is the standard composition for mammalian mitogenomes (Gibson et al., 2005). Table 3 further specifies each annotation shown in Figure 1 by providing information regarding annotation length, start and stop nucleotide position, and replication strand. The majority of the genes are replicated on the plus (heavy) strand, including the protein-coding genes, with the exception of *ND6*, which is replicated on the minus (light) strand. Table 4 expands upon this by providing the codon usage of protein coding genes in the *Canis aureus cruesemanni* sample as well as the *Canis lupus* mitogenome used for reference. The two mitogenomes are mostly similar in codon usage, as the number of each codon

only differed by a maximum of 10. The fractions were also very similar, differing by a maximum of 0.1. Interestingly, only the *Canis lupus* sequence included a GTG start codon.

3.2 Phylogenetic and haplotype network analysis

The sample from Thailand yielded approximately 89.8% to 97.3% similarity when compared to the other canid samples included in the 15,958 bp multiple sequence alignment. The golden jackal sample from Turkey yielded approximately 89.1% to 99.9% similarity when compared with the same canid samples, with its highest similarity corresponding to a *Canis aureus* sample from India. The phylogenetic tree in Figure 2 based on the mitogenome alignment shows that the Indochinese jackal is grouped with other golden jackals with 100% bootstrap support. The four golden jackal sequences comprise four unique haplotypes. The Indochinese jackal forms a distinct lineage and shows a maximum of 2.8% sequence divergence from other putative subspecies of golden jackal, supporting the former's distinctiveness. The golden jackal from Turkey is shown to be most closely related to other *Canis aureus* samples, specifically those from India and Israel. The phylogenetic tree in Figure 3 based on the *CYTB* alignment shows a similar pattern of relationships despite the larger number of sequences and wider geographic representation, with the Indochinese golden jackal forming the earliest branching lineage and showing a 4.1% sequence divergence from other golden jackals, including the sample from Turkey which is on its own branch between the sample from India and a cluster of several others from Israel and other countries. The grouping of sequences from India and more western regions is consistent with the results reported by Yumnam et al. (2015). Interestingly, the topology within golden jackals shows a pattern of branching from east to west, suggesting that Southeast Asia may be the region of origin for this species. However, additional studies using more

samples and data from the nuclear genome are needed to test this hypothesis. The haplotype network analysis of the *CYTB* gene sequences in Figure 4 places the Indochinese jackal sample separately from the remaining golden jackal samples by 46 substitutions, which is only slightly less that the number of substitutions (57) separating golden jackals from the African wolf. The figure also demonstrates that the Indochinese jackal shares a common ancestor with other populations of golden jackal, further supporting the information obtained from the phylogenetic trees. Furthermore, this figure places the Turkey sample amongst other *Canis aureus* samples, confirming that it is indicative of a golden jackal that is phylogenetically separate from *Canis aureus cruesemanni*.

4. Conclusions

Southeast Asia is home to some of the most threatened biodiversity in the world, and this includes populations such as the Indochinese jackal. Our research produced the first-ever complete mitogenome assembly for this animal. It also demonstrated through phylogenetic analysis and comparison with other golden jackals and Canidae species that it may be classified as a distinct subspecies of the golden jackal (*Canis aureus*). Ultimately, our results provide a foundation for further studies of the Indochinese jackal, which will help inform the evolutionary history and conservation status of this little-known subspecies.

Acknowledgments

The authors would like to thank Dr. Anneke DeLuycker and Dr. Joshua Davis at the Smithsonian-Mason School of Conservation in Front Royal, VA for their support of this research project through the Spring 2022 Endangered Species Conservation course. The authors would also like to thank the Faculty of Forestry, Kasetsart University, the Department of National

Parks, Wildlife and Plant Conservation, Thailand, and the Zoological Park Organization of Thailand for their support in sample collection.

Declaration of Competing Interest

The content of this paper was developed solely by the authors. We declare no competing interest.

Data Availability Statement

The raw read data for the Indochinese jackal from Thailand were deposited in the NCBI Short Read Archive under a BioProject (accession: PRJNA847318) with a BioSample ID of SAMN29334461 and an SRA accession number of SRX15907383. The new mitogenomes sequenced and assembled from golden jackals from Thailand and Turkey were deposited into NCBI's GenBank under the accession numbers ON986207 and OP345200, respectively.

Credit authorship contribution statement

Medhini Sosale: assembled and annotated the mitogenome of the golden jackal from Thailand, analyzed the data, interpreted the data, and drafted the initial manuscript. **Nucharin Songsasen:** performed the field work and collected the sample of the golden jackal from Thailand. **Osman İbiş:** collected the sample of the golden jackal from Turkey, performed the experiments to generate the mitogenome sequences, assembly, and annotation of this sample. **Henrique V. Figueiró** and **Cody W. Edwards**: conceived and designed the study and assisted in the interpretation of the data. **Klaus-Peter Koepfli:** conceived and designed the study, assisted with data interpretation, and revised the initial manuscript draft. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

An, J. et al. *Canis lupus mitochondrion, complete genome* (966798673). (2016). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/KT901460.1

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Bioinformatics. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Björnerfeldt, S., Webster, M. T., & Vilà, C. (2006). Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Research*, *16*(8), 990–994. https://doi.org/10.1101/gr.5117706

Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner. No. LBNL-7065E. Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA.

Chen, L. & Zhang, H. H. (2009). *Cuon alpinus mitochondrion, complete genome* (262188097). (2009). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/NC_013445.1

da Silva Coelho, F. A., Gill, S., Tomlin, C. M., Heaton, T. H., & Lindqvist, C. (2021). An early dog from southeast Alaska supports a coastal route for the first dog migration into the Americas. *Proceedings. Biological Sciences*, *288*(1945), 20203103. https://doi.org/10.1098/rspb.2020.3103

Donath, A., Jühling, F., Al-Arab, M., Bernhart, S. H., Reinhardt, F., Stadler, P. F., Middendorf,

M., & Bernt, M. (2019). Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Research, 47*(20), 10543–10552. https://doi.org/10.1093/nar/gkz833

Fabbri, E., Caniglia, R., Galov, A., Arbanasić, H., Lapini, L., Bošković, I., Florijančić, T., Vlasseva, A., Ahmed, A., Mirchev, R. L., & Randi, E. (2014). Genetic structure and expansion of golden jackals (*Canis aureus*) in the north-western distribution range (Croatia and eastern Italian Alps). *Conservation Genetics*, *15*(1), 187–199. https://doi.org/10.1007/s10592-013-0530-7

Formenti, G., Rhie, A., Balacco, J., Haase, B., Mountcastle, J., Fedrigo, O., Brown, S., Capodiferro, M. R., Al-Ajli, F. O., Ambrosini, R., Houde, P., Koren, S., Oliver, K., Smith, M., Skelton, J., Betteridge, E., Dolucan, J., Corton, C., Bista, I., … The Vertebrate Genomes Project Consortium. (2021). Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. *Genome Biology*, *22*(1), 120. https://doi.org/10.1186/s13059-021-02336-9

Gibson, A., Gowri-Shankar, V., Higgs, P. G., & Rattray, M. (2005). A Comprehensive Analysis of Mammalian Mitochondrial Genome Base Composition and Improved Phylogenetic Methods. *Molecular Biology and Evolution*, *22*(2), 251–264. https://doi.org/10.1093/molbev/msi012

Hemprich, F.G., & Ehrenberg, C.G. (1832). *Symbolae Physicaequae ex Itinere Africam Borealem er Asoam OccidentalemDecas Secunda*. ExOfficina Academica.

Hennelly, L. M., Habib, B., Modi, S., Rueness, E. K., Gaubert, P., & Sacks, B. N. (2021). Ancient divergence of Indian and Tibetan wolves revealed by recombination-aware phylogenomics. *Molecular Ecology*, *30*(24), 6687–6700.

https://doi.org/10.1111/mec.16127

Hoffmann, M., Arnold, J., Duckworth, J.W., Jhala, Y., Kamler, J.F. & Krofel, M. 2018. *Canis aureus* (errata version published in 2020). *The IUCN Red List of Threatened Species* 2018: e.T118264161A163507876

Hwang, K-.C. et al. *Lycaon pictus mitochondrion, complete genome* (944203056). (2015a). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/KT598692.1

Hwang, K-.C. et al. *Lycaon pictus mitochondrion, complete genome* (953244877). (2015b). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/NC_028427.1

İbiş, O., Selçuk, A. Y., Sacks, B. N., Yıldız, B., Özcan, S., Kefelioğlu, H., & Tez, C. (2020). Whole mitochondrial genome of long-clawed mole vole (*Prometheomys schaposchnikowi*) from Turkey, with its phylogenetic relationships. *Genomics*, *112*(5), 3247-3255. https://doi.org/10.1016/j.ygeno.2020.06.008

Jia, Q.-H., Liu, H.-B., Gu, X.-L., Li, J.-F., & Liu, Y.-L. (2016). Complete mitochondrial genome of a German Shepherd (Canis lupus familiaris breed German Shepherd) provides insights into genome-wide sequence variations. *Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis*, *27*(1), 118–119. https://doi.org/10.3109/19401736.2013.873930

Jie, Z. *Canis simensis mitochondrion, complete genome* (2220274589). (2022a). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/MT793779.1

Jie, Z. *Canis simensis mitochondrion, complete genome* (2228964576). (2022b). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/NC_062616.1

Katoh, K. & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093%2Fmolbev%2Fmst010

Koepfli, K.-P., Pollinger, J., Godinho, R., Robinson, J., Lea, A., Hendricks, S., Schweizer, R. M., Thalmann, O., Silva, P., Fan, Z., Yurchenko, A. A., Dobrynin, P., Makunin, A., Cahill, J. A., Shapiro, B., Álvares, F., Brito, J. C., Geffen, E., Leonard, J. A., … Wayne, R. K. (2015). Genome-wide Evidence Reveals that African and Eurasian Golden Jackals Are Distinct Species. *Current Biology: CB*, *25*(16), 2158–2165. https://doi.org/10.1016/j.cub.2015.06.060

Krofel, M., Giannatos, G., Ćirovič, D., Stoyanov, S., & Newsome, T. M. (2017). Golden jackal expansion in Europe: A case of mesopredator release triggered by continent-wide wolf persecution? *Hystrix, the Italian Journal of Mammalogy*, *28*(1), 9–15. https://doi.org/10.4404/hystrix-28.1-11819

Krofel, M., Hatlauf, J., Bogdanowicz, W., Campbell, L. a. D., Godinho, R., Jhala, Y. V., Kitchener, A. C., Koepfli, K.-P., Moehlman, P., Senn, H., Sillero-Zubiri, C., Viranta, S., Werhahn, G., & Alvares, F. (2022). Towards resolving taxonomic uncertainties in wolf, dog and jackal lineages of Africa, Eurasia and Australasia. *Journal of Zoology*, *316*(3), 155–168. https://doi.org/10.1111/jzo.12946

Leigh, J.W. & Bryant, D. (2015). PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Lindgreen, S. (2012). AdapterRemoval: Easy Cleaning of Next Generation Sequencing Reads, *BMC Research Notes*, *5*, 337. https://doi.org/10.1186%2F1756-0500-5-337

Linnaeus, C. 1758. *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. Editio decima, reformata. Volume 1. Laurentii Salvii, Stockholm, Sweden.

Loog, L., Thalmann, O., Sinding, M.-H. S., Schuenemann, V. J., Perri, A., Germonpré, M., Bocherens, H., Witt, K. E., Samaniego Castruita, J. A., Velasco, M. S., Lundstrøm, I. K. C., Wales, N., Sonet, G., Frantz, L., Schroeder, H., Budd, J., Jimenez, E.-L., Fedorov, S., Gasparyan, B., … Manica, A. (2020). Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. *Molecular Ecology, 29*(9), 1596–1610. https://doi.org/10.1111/mec.15329

Matschie, P. 1900. Herr Matschie sprach über den Schakal des Menam-Gebietes in Siam. *Sitzungs-Berichte der Gesellschaft der Naturforschender Freunde zu Berlin 3,* 144–145.

Meachen, J. et al. *Canis lupus mitochondrion, complete genome* (1951773529). (2021). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/MT796489.1

Moehlman, P. D., & Hayssen, V. (2018). Canis aureus (Carnivore: Canidae). *Mammalian Species*, *50*(957), 14–25. https://doi.org/10.1093/mspecies/sey002

Parr, J. K. W. 2003. *A Guide to the Large Mammals of Thailand*. Sarakadee Press, Bangkok, Thailand. 206 pp.

Rutkowski, R., Krofel, M., Giannatos, G., Ćirović, D., Männil, P., Volokh, A. M., Lanszki, J., Heltai, M., Szabó, L., Banea, O. C., Yavruyan, E., Hayrapetyan, V., Kopaliani, N., Miliou, A., Tryfonopoulos, G. A., Lymberakis, P., Penezić, A., Pakeltytė, G., Suchecka, E., & Bogdanowicz, W. (2015). A European Concern? Genetic Structure and Expansion of Golden Jackals (Canis aureus) in Europe and the Caucasus. *PLOS ONE, 10*(11), e0141236. https://doi.org/10.1371/journal.pone.0141236

Scheible, M. K. R., Straughan, D. J., Burnham-Curtis, M. K., & Meiklejohn, K. A. (2021). Using hybridization capture to obtain mitochondrial genomes from forensically relevant North American canids: Assessing sequence variation for species identification. *Forensic Science International: Animals and Environments*, *1*, 100018. https://doi.org/10.1016/j.fsiae.2021.100018

Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., & Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, *9*(5), 1056–1082. https://doi.org/10.1038/nprot.2014.063

Spassov, N., & Acosta-Pankov, I. (2019). Dispersal history of the golden jackal (Canis aureus moreoticus Geoffroy, 1835) in Europe and possible causes of its recent population explosion. *Biodiversity Data Journal*, *7*, e34825. https://doi.org/10.3897/BDJ.7.e34825

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Urios, V., Donat, P., Ramírez, C., Monroy-vilchis, O., & Hamid, R. I. (2015). El análisis del genoma mitocondrial del cánido estudiado en Marruecos manifiesta que no es ni lobo (Canis lupus) ni chacal euroasiático (Canis aureus). *Altoter0*.

Verscheure, S., Backeljau, T., & Desmyter, S. (2014). Dog mitochondrial genome sequencing to enhance dog mtDNA discrimination power in forensic casework. *Forensic Science International. Genetics*, *12*, 60–68. https://doi.org/10.1016/j.fsigen.2014.05.001

Yumnam, B., Negi, T., Maldonado, J. E., Fleischer, R. C., & Jhala, Y. V. (2015).

Phylogeography of the Golden Jackal (Canis aureus) in India. *PLOS ONE*, *10*(9), e0138497. https://doi.org/10.1371/journal.pone.0138497

Yusefi, G. H., Godinho, R., Khalatbari, L., Broomand, S., Fahimi, H., Martínez-Freiría, F., & Alvares, F. (2021). Habitat use and population genetics of golden jackals in Iran: Insights from a generalist species in a highly heterogeneous landscape. *Journal of Zoological Systematics and Evolutionary Research*, *59*(7), 1503–1515. https://doi.org/10.1111/jzs.12519

Figures

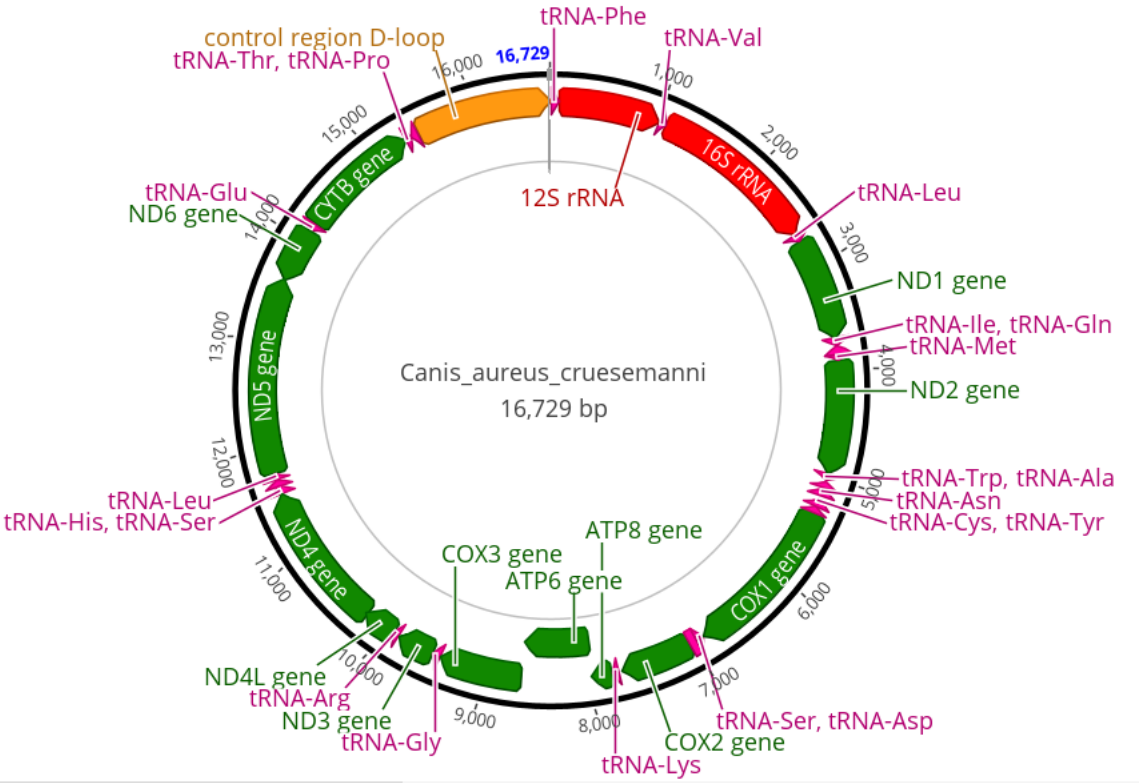


Figure 1: Complete annotated mitogenome of the Indochinese jackal (*Canis aureus cruesemanni*). Colored bars correspond to the different classes of genes: green = protein-coding genes (CDS), red = rRNA genes, magenta = tRNA genes, and orange = the control region. Gene arrows indicate transcription on the plus (right direction) or minus (left direction) strand. Black outer ring shows the relative nucleotide position of the different genes of the mitogenome.

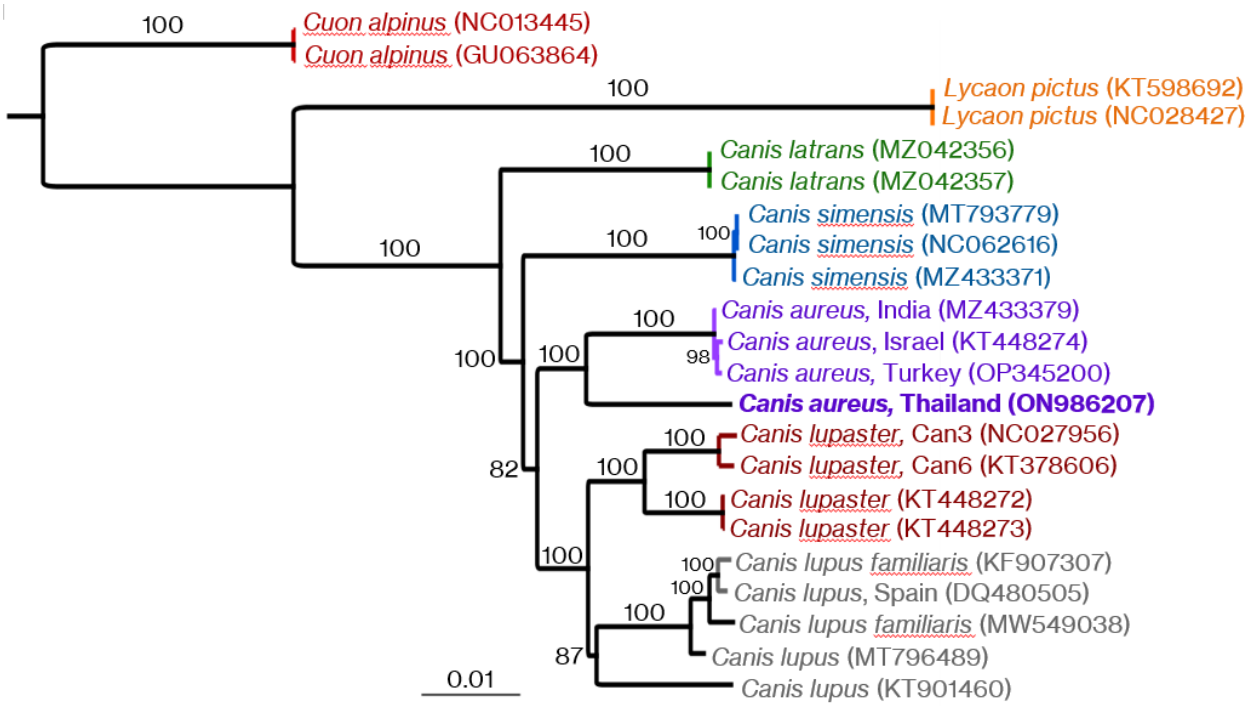
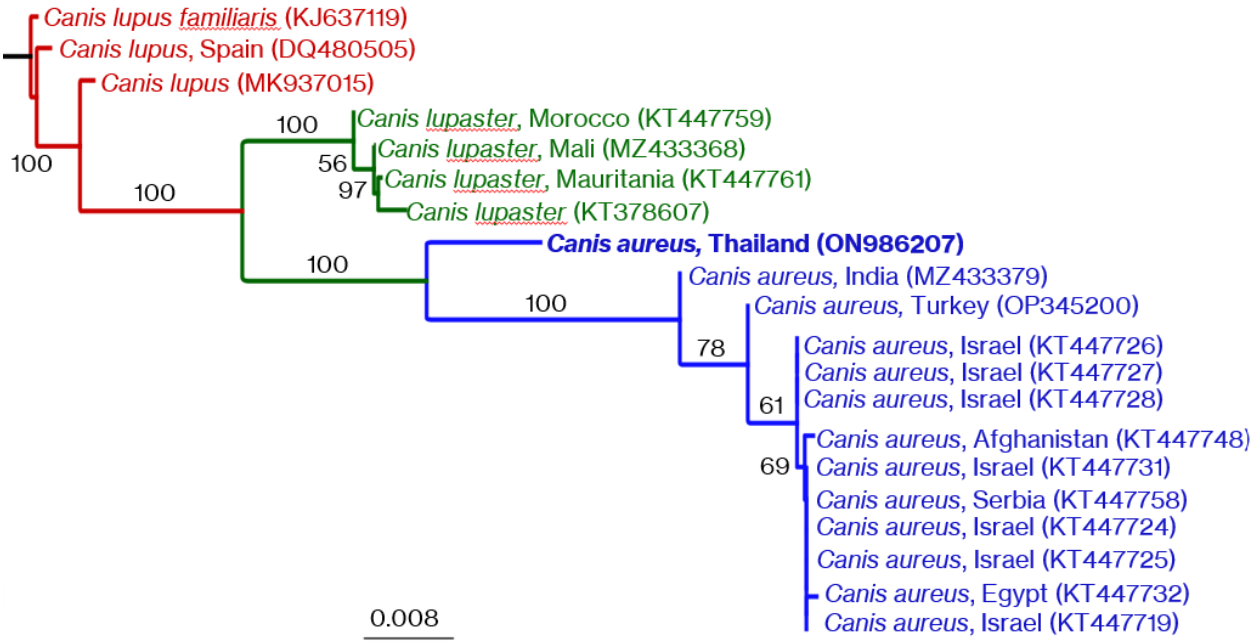


Figure 2: Maximum-likelihood phylogenetic tree based on the 15,958 bp mitogenome alignment showing the relationship of the Indochinese jackal (*Canis aureus*, Thailand) to other golden jackals and species within the genus *Canis*. African painted dog (*Lycaon pictus*) and dhole (*Cuon alpinus*) were used to root the tree. Numbers represent bootstrap values (%) and the GenBank accession number for each sequence is shown in parentheses



396

397 **Figure 3**: Maximum-likelihood phylogenetic tree based on the 1,140 bp *CYTB* alignment. The

398 Indochinese jackal from Thailand is shown in bold blue font. Gray wolf (*Canis lupus*) and

399 domestic dog (*Canis lupus familiaris*) samples were used to root the tree. Numbers represent

400 bootstrap values (%) and the GenBank accession number for each file is shown in parentheses.

401

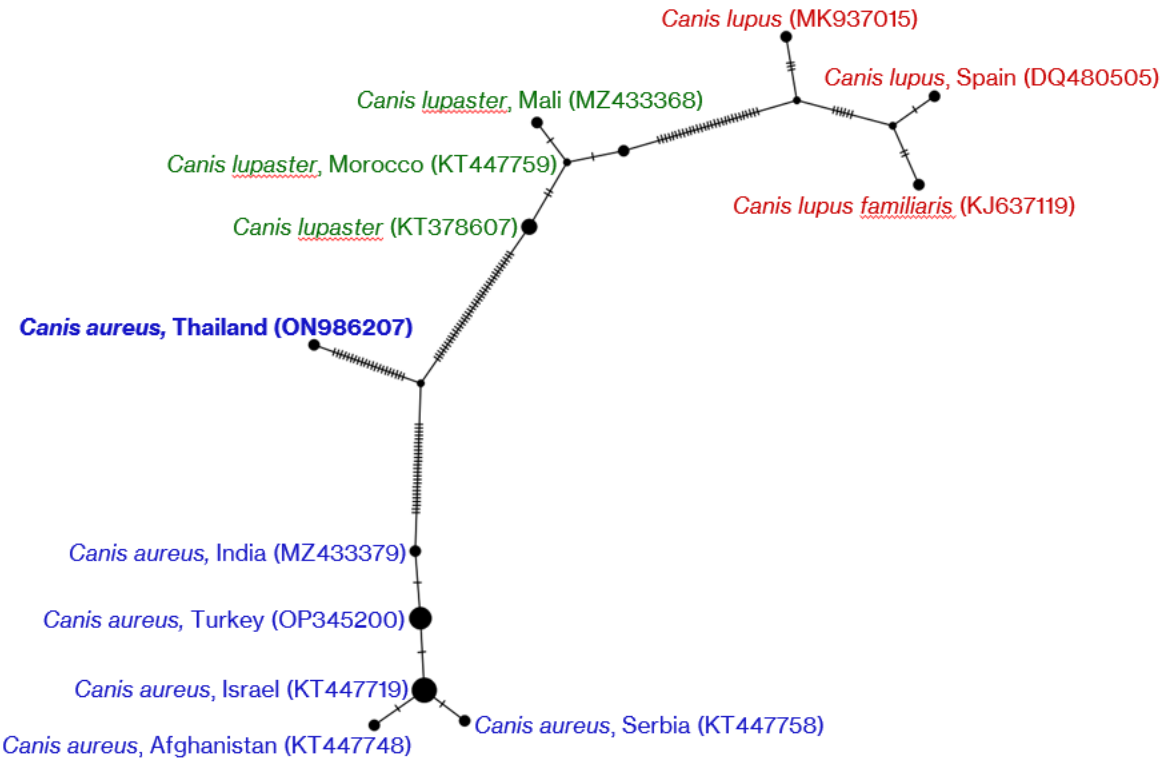


Figure 4: Haplotype network based on the 1,140 bp *CYTB* alignment. The Indochinese jackal from Thailand is shown in bold blue font. Each individual haplotype and/or ancestor is represented by a black dot, and each dash in between the dots corresponds to a substitution. Dot size corresponds to the frequency of a particular haplotype.

Data Tables

Table 1: List of species and GenBank accession numbers used in the mitogenome phylogenetic analysis (Figure 2).

Species Name	GenBank Accession Number	Reference
Cuon alpinus	NC013445	(Chen, L. & Zhang, H. H., 2009)
Cuon alpinus	GU063864	(Chen, L. & Zhang, H. H., 2009)
Lycaon pictus	KT598692	(Hwang, K-.C. et al., 2015)
Lycaon pictus	NC028427	(Hwang, K-.C. et al., 2015)
Canis latrans	MZ042356	(Scheible, M. K. et al., 2021)
Canis latrans	MZ042357	(Scheible, M. K. et al., 2021)
Canis simensis	MT793779	(Jie, Z., 2022)
Canis simensis	NC062616	(Jie, Z., 2022)
Canis simensis	MZ433371	(Hennelly, L. M. et al., 2021)
Canis aureus	MZ433379	(Hennelly, L. M. et al., 2021)
Canis aureus	KT448274	(Koepfli, K-.P. et al., 2015)
Canis lupaster	NC027956	(Urios, V. et al., 2015)
Canis lupaster	KT378606	(Urios, V. et al., 2015)
Canis lupaster	KT448272	(Koepfli, K-.P. et al., 2015)
Canis lupaster	KT448273	(Koepfli, K-.P. et al., 2015)
Canis lupus familiaris	KF907307	(Jia, Q. H. et al., 2016)
Canis lupus	DQ480505	(Bjornerfeldt, S., Webster, M. T., & Vila, C., 2006)
Canis lupus familiaris	MW549038	(da Silva Coehlo, F. A. et al., 2021)
Canis lupus	MT796489	(Meachen, J. et al., 2021)
Canis lupus	KT901460	(An, J., 2016)

Table 2: List of species and GenBank accession numbers used in the *CYTB* analyses (Figures 3 and 4).

Species Name	GenBank Accession Number	Reference
Canis lupus familiaris	KJ637119	(Verscheure, S., Backeljau, T., & Desmyter, S., 2014)
Canis lupus	DQ480505	(Bjornerfeldt, S., Webster, M. T., & Vila, C., 2006)
Canis lupus	MK937015	(Loog, L. et al., 2020)
Canis lupaster	KT447759	(Koepfli, K-.P. et al., 2015)
Canis lupaster	MZ433368	(Hennelly, L. M. et al., 2021)
Canis lupaster	KT447761	(Koepfli, K-.P. et al., 2015)
Canis lupaster	KT378607	(Donat-Torres, M. P. et al., 2015)
Canis aureus	MZ433379	(Hennelly, L. M. et al., 2021)
Canis aureus	KT447726	(Koepfli, K-.P. et al., 2015)
Canis aureus	KT447727	(Koepfli, K-.P. et al., 2015)
Canis aureus	KT447728	(Koepfli, K-.P. et al., 2015)
Canis aureus	KT447748	(Koepfli, K-.P. et al., 2015)
Canis aureus	KT447731	(Koepfli, K-.P. et al., 2015)
Canis aureus	KT447758	(Koepfli, K-.P. et al., 2015)
Canis aureus	KT447724	(Koepfli, K-.P. et al., 2015)
Canis aureus	KT447725	(Koepfli, K-.P. et al., 2015)
Canis aureus	KT447732	(Koepfli, K-.P. et al., 2015)
Canis aureus	KT447719	(Koepfli, K-.P. et al., 2015)

Table 3: Annotation and arrangement of genes in the mitochondrial genome of *Canis aureus cruesemanni*.

Name	Type	Start	Stop	Strand	Length
tRNA-Phe	tRNA	1	69	+	69
12S rRNA	rRNA	70	1023	+	954
tRNA-Val	tRNA	1024	1090	+	67
16S rRNA	rRNA	1091	2670	+	1580
tRNA-Leu	tRNA	2671	2745	+	75
ND1	Gene	2748	3704	+	957
tRNA-Ile	tRNA	3704	3772	+	69
tRNA-Gln	tRNA	3769	3843	-	75
tRNA-Met	tRNA	3845	3914	+	70
ND2	Gene	3915	4958	+	1044
tRNA-Trp	tRNA	4957	5024	+	68
tRNA-Ala	tRNA	5038	5106	-	69
tRNA-Asn	tRNA	5108	5179	-	72
tRNA-Cys	tRNA	5214	5280	-	67
tRNA-Tyr	tRNA	5281	5348	-	68
COX1	Gene	5350	6894	+	1545
tRNA-Ser	tRNA	6892	6962	-	71
tRNA-Asp	tRNA	6967	7034	+	68
COX2	Gene	7035	7718	+	684
tRNA-Lys	tRNA	7736	7802	+	67
ATP8	Gene	7804	8007	+	204
ATP6	Gene	7965	8645	+	681
COX3	Gene	8645	9428	+	784
tRNA-Gly	tRNA	9429	9496	+	68
ND3	Gene	9497	9842	+	346
tRNA-Arg	tRNA	9843	9911	+	69
ND4L	Gene	9914	10210	+	297
ND4	Gene	10204	11581	+	1378
tRNA-His	tRNA	11582	11650	+	69
tRNA-Ser	tRNA	11651	11710	+	60
tRNA-Leu	tRNA	11711	11780	+	70
ND5	Gene	11781	13601	+	1821
ND6	Gene	13585	14112	-	528
tRNA-Glu	tRNA	14113	14181	-	69
CYTB	Gene	14186	15325	+	1140
tRNA-Thr	tRNA	15326	15395	+	70

tRNA-Pro	tRNA	15395	15460	-	66
control region D-loop	D-loop	15464	16729	+	1266

426

427

Table 4: Comparison of codon usage in protein-coding genes of the mitogenomes of *Canis aureus cruesemanni* (this study) and *Canis lupus* (GenBank: DQ480505).

		Canis aureus cruesemanni			***Canis lupus***		
Amino Acid	**Codon**	**Number**	**/1000**	**Fraction**	**Number**	**/1000**	**Fraction**
Ala	GCA	90	23.68	0.36	91	23.93	0.36
	GCC	84	22.11	0.34	87	22.88	0.35
	GCG	13	3.42	0.05	12	3.16	0.05
	GCT	61	16.05	0.25	61	16.04	0.24
Cys	TGC	19	5	0.76	21	5.52	0.84
	TGT	6	1.58	0.24	4	1.05	0.16
Asp	GAC	36	9.47	0.54	40	10.52	0.58
	GAT	31	8.16	0.46	29	7.63	0.42
Glu	GAA	71	18.68	0.76	75	19.73	0.8
	GAG	23	6.05	0.24	19	5	0.2
Phe	TTC	127	33.42	0.54	126	33.14	0.54
	TTT	108	28.42	0.46	109	28.67	0.46
Gly	GGA	95	25	0.44	93	24.46	0.43
	GGC	60	15.79	0.28	53	13.94	0.25
	GGG	26	6.84	0.12	27	7.1	0.13
	GGT	34	8.95	0.16	43	11.31	0.2
His	CAC	61	16.05	0.63	57	14.99	0.6
	CAT	36	9.47	0.37	38	9.99	0.4
Ile	ATC	159	41.84	0.46	168	44.19	0.49
	ATT	185	48.68	0.54	175	46.03	0.51
Lys	AAA	82	21.58	0.82	80	21.04	0.8
	AAG	18	4.74	0.18	20	5.26	0.2
Leu	CTA	231	60.79	0.39	231	60.76	0.39
	CTC	86	22.63	0.15	91	23.93	0.15
	CTG	30	7.89	0.05	33	8.68	0.06
	CTT	98	25.79	0.17	95	24.99	0.16
	TTA	116	30.53	0.2	117	30.77	0.2
	TTG	29	7.63	0.05	27	7.1	0.05
Met	ATA	207	54.47	0.82	202	53.13	0.81
	ATG	44	11.58	0.18	45	11.84	0.18
	GTG	--	--	--	1	0.26	0
Asn	AAC	93	24.47	0.58	92	24.2	0.59
	AAT	67	17.63	0.42	65	17.1	0.41
Pro	CCA	61	16.05	0.31	60	15.78	0.31
	CCC	61	16.05	0.31	62	16.31	0.32

	CCG	9	2.37	0.05	7	1.84	0.04
	CCT	63	16.58	0.32	64	16.83	0.33
Gln	CAA	68	17.89	0.76	69	18.15	0.78
	CAG	21	5.53	0.24	20	5.26	0.22
Arg	CGA	44	11.58	0.68	45	11.84	0.69
	CGC	9	2.37	0.14	10	2.63	0.15
	CGG	5	1.32	0.08	4	1.05	0.06
	CGT	7	1.84	0.11	6	1.58	0.09
Ser	AGC	31	8.16	0.11	33	8.68	0.12
	AGT	18	4.74	0.06	17	4.47	0.06
	TCA	79	20.79	0.28	76	19.99	0.27
	TCC	70	18.42	0.25	75	19.73	0.27
	TCG	13	3.42	0.05	13	3.42	0.05
	TCT	71	18.68	0.25	68	17.89	0.24
Thr	ACA	111	29.21	0.37	120	31.56	0.39
	ACC	95	25	0.31	90	23.67	0.3
	ACG	20	5.26	0.07	18	4.73	0.06
	ACT	78	20.53	0.26	76	19.99	0.25
Val	GTA	93	24.47	0.5	86	22.62	0.47
	GTC	32	8.42	0.17	28	7.36	0.15
	GTG	16	4.21	0.09	23	6.05	0.13
	GTT	45	11.84	0.24	47	12.36	0.26
Trp	TGA	90	23.68	0.87	93	24.46	0.89
	TGG	14	3.68	0.13	11	2.89	0.11
Tyr	TAC	69	18.16	0.49	75	19.73	0.52
	TAT	73	19.21	0.51	69	18.15	0.48
Stop	AGA	1	0.26	0.13	1	0.26	0.1
	AGG	0	0	0	0	0	0
	TAA	7	1.84	0.88	8	2.1	0.8
	TAG	0	0	0	1	0.26	0.1

430