Analyses of Wild Ungulates Mitogenome

Toni TEŠIJA[1] (✉)
Toni SAFNER[1,2]

Summary

In the last two decades, mitochondrial DNA (mtDNA) has become one of the most used genetic markers for studying evolution and phylogeny. Understanding mitochondrial genome (mitogenome) structure, gene arrangements, base compositions and noncoding regions can reveal valuable information on various organisms, including wild ungulates. A huge number of sequenced mammalian mitogenomes are available on GenBank, including 8074 mitogenome sequences of wild ungulates. The Next-Generation Sequencing (NGS) approach can reveal mitogenome regions that can explain the variation in populations. Such variations could show ungulates' resistance to pathogens or processes of their adaptation. Despite the extensive number of new sequenced mitogenomes, there is still a lot of ambiguity around the mitogenome architecture. Methods for NGS, genome assembly and annotation are still in the process of development, aiming to make these tools powerful enough to reveal enormous knowledge crucial for assessing wild ungulate species' conservation status, and their ecological status. In this review paper, we describe the methods and principles derived from mtDNA studies on wild ungulate species. We outline basic sequence preparation methods and specialized software for mitogenome assembly and annotation. Additionally, we present several different approaches of mitogenome comparison.

Key words

mitochondrial DNA, mitogenome, mtDNA assembly, mtDNA annotation, wild ungulates, phylogeny

[1] University of Zagreb, Faculty of Agriculture, Department of Plant Breeding, Genetics and Biometrics, Svetošimunska cesta 25, 10000 Zagreb, Croatia

[2] Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE CroP-BioDiv), Svetošimunska cesta 25, 10000 Zagreb, Croatia

✉ Corresponding author: ttesija@agr.hr

Received: May 28, 2020 | Accepted: September 23, 2020

INTRODUCTION

The Structure of Mitochondrial DNA

Mitochondria are organelles in eukaryotic cells responsible for producing energy, which play a pivotal role in mammals' temperature regulation (Boore, 1999; Wallace, 2007). During the oxidative phosphorylation (OXPHOS) process (that results in adenosine triphosphate ATP synthesis), mitochondria oxidize metabolic substrates in order to generate energy and water. The metabolic processes differ between taxa (da Fonseca et al., 2008), essentially depending on species' metabolic requirements (e.g. body size, diet, high altitudes). Mammal mtDNA is a small (~16 kb) circular molecule with two strands (Heavy or H strand and Light or L strand), divided in two parts: coding and non-coding. The coding part contains 13 protein-coding genes (PCGs) for OXPHOS, the 22 transfer RNA genes (tRNA) for protein translation and 2 ribosomal RNA genes (rRNA) for the mitochondrial ribosome. This part is highly conserved for mammalian mitogenomes and has a low rate of gene rearrangements compared to the nuclear genome (Wolstenholme, 1992). The H strand encodes 12 PCGs, 2 rRNA genes and 14 tRNA genes and it is rich in guanin. The L strand encodes one subunit (ND6) and 8 tRNA genes (Barshad et al., 2018; Barchiesi and Vascotto, 2019). The non-coding parts of mtDNA comprise three parts: the control region (CR, Displacement loop, D-loop), the O_L region (the origin of L-strand replication), and the intergenic regions. The D-loop is long about 700-1,300 bp in length while the O_L is long about 30 bp. The intergenic regions (spacers) are located within some genes. MtDNA genes do not have introns, except for the intergenic spacers that are limited to a few bases, or, in some cases, completely absent from the mitogenome (Taanman, 1999). Dromedary camel (*C. dromedaries*) has a total of 116 bp of intergenic regions (22 different locations) in its mitogenome (Manee et al., 2019), which is identical or very similar for species within the same genus. DNA and RNA synthesis start in D-loop regions (Saccone et al., 2000), which are well conserved (Garesse and Vallejo, 2001). However, other parts of the D-loop regions present the most polymorphic region of mtDNA (Upholt and Dawid, 1977; Shadel and Clayton, 1997). Those variable parts of the D-loop evolve four to five times faster in comparison to other regions of mitogenome (Fumagalli et al., 1996).

The replication of mammalian mitogenome is asymmetric since the two strands are synthesized from two distinct replication sites. The H strand replication site (O_H) is located at the D-loop region while L strand is in the O_L region located between the genes encoding tRNAs asparagine and cystine (Hixson et al., 1986). During replication, the H strand is replicated first, and the parental H strand remains single until it is paired with newly synthesized L strand. Before pairing, the parental H strand is exposed to oxidative damage because it is only partially protected by proteins. This process is one of the causes of large number of mutations in mtDNA. In addition, nuclear proteins that are transported from cytosol into mitochondria are directly involved in mitochondrial function and can cause mutations in mitochondrial genes (Patrushev et al., 2014). The rate of the mutations caused by those reasons is the highest in the D-loop region of mtDNA (Brown and Simpson, 1982; Reyes et al., 1998). Mutations accumulating in D-loop region are not lethal, and they can be passed to progeny. On the other hand, the number of such mutations accumulating in the coding region is small, since they might be lethal, and only lesser number of them is passed down to future generations (Gupta et al., 2015).

Since the D-loop region is the most rapidly evolving part of the mitogenome (Upholt and Dawid, 1977; Walberg and Clayton, 1981), where the largest number of mutations is accumulated, it is commonly used as a molecular marker for finding genetic differentiation between animal species, as well as between individuals of the same species (Gupta et al., 2015).

Another mtDNA region with high diversity is cytochrome c oxidase subunit 1 (COI), which is often used as animal DNA barcoding system. Animal barcoding systems are used to provide additional information for species that are not correctly identified by typical techniques based on morphology (Naseem et al., 2020). Hebert et al. (2003) proposed an approach where, based on the nucleotide sequence diversity in the COI gene, it is possible to differentiate species from diverse taxonomic groups.

D-loop Region Structure and Repeating Sequences

Most of the mammals have similar D-loop location and organization. The D-loop is located between the genes encoding tRNAs Pro and Phe and it consists of three main domains. The first domain is called the right domain (or conserved sequence blocks, CBS) and is adjacent to the tRNA gene encoding Phe. This region contains crucial regulatory parts for replication, transcription of both strands and the segment with conserved sequence blocks (CBSs) involved in the processing of RNA primers for synthesis of H-strand. The second is the left domain (or the extended terminal associated sequenced domain, ETAS), which is adjacent to the tRNA gene encoding Pro. In this region synthesis of the heavy strand is paused. The third part is the central conserved domain. It is presumed to be the site of origin for replication and the most highly conserved region of the D-loop (Saccone et al., 1991; Sbisà et al., 1997; Shadel and Clayton, 1997; Boore, 1999).

As mentioned before, the main factors defining the variability of the D-loop region are processes during replication, but there are other contributing factors such as the different numbers of repeating sequences (RS). These regions are common in the ETAS and CSB domains (Polziehn and Strobeck, 2002; An et al., 2010; Meadows et al., 2011). Furthermore, the highest rate of sequence variation between individuals is found in hypervariable region 1 and hypervariable region 2 (Miller et al., 1996; Jazin et al., 1998). The D-loop is responsible for length variations of mitogenomes due to the number of RS across different taxa (Brown et al., 1996; Xu et al., 2005; Meadows et al., 2011; Cui et al., 2012). In mammalian D-loop regions, the RS have been found in five positions (RS 1-5). RS1 and RS2 are found in the ETAS domain of the D-loop, where the H strand replication pauses; RS3, RS4, and RS5 are found in the CBS domain, upstream of the site for the H strand origin of replication (Hoelzel et al., 1994).

Mitogenome Sequence Variations

In the coding sequence, the first and second positions of the codon are often stable, but the third position mutates fast (Kimura, 1968). The mtDNA does not have the same repair system as the

nuclear DNA, which means it is exposed to rapid molecular evolution (Matosiuk et al., 2014). The changes in base composition might lead to modifications in amino acid composition (Foster et al., 1997). The changes in genome sequence are divided into four different types: sequence rearrangements, additions, deletions and substitutions (Brown, 1985). The most common type of mutations in mitogenome are nucleotide substitutions. The substitution rate has been estimated to be about 5-10 times greater than in the nuclear DNA, although rates vary between different parts of the mitogenome (Brown et al., 1979). The substitutions can be detected as synonymous (i.e. silent) or nonsynonymous (Jukes and Bhurshan, 1986). If a synonymous substitution occurs, the corresponding protein sequence will be identical, but if there is a nonsynonymous substitution, it will affect the amino acid, making changes in the protein sequence (Foster et al., 1997).

The genetic code of the mammalian mitogenome is universal, albeit with few exceptions. For example, the UGA codes for tryptophan rather than for STOP; AGA and AGG, normally coded for arginine, codes for STOP codons; AUA codes for methionine and not isoleucine; and ubiquitous AUG start codon is sometimes replaced by AUA or AUU in mitochondrial genes; AGR (R=A, G) specifies a STOP codon in vertebrate mtDNA, codes for serine in mtDNA of echinoderms and codes for arginine in yeast mtDNA, as in the standard genetic code (Xiufeng and Árnason, 1994; Taanman, 1999; Gupta et al., 2015).

The substitution rate in synonymous and nonsynonymous positions varies between genes. A higher mutation rate may be associated with the damage caused by free radicals generated during the OXPHOS process (Reyes et al., 1998). The ETAS domain has high substitution rate in most mammals. Korean gorals (*N. caudatus, N. goral*), Korean native goats (*C. hircus*) and Japanese serows (*C.swinhoei, C. crispus*) have the highest number of substitutions in ETAS domain (An et al., 2010).

The most remarkable feature of mammalian mitogenome is the distribution of G and C bases between the two strands (Reyes et al., 1998). GC-content is the percentage (or ratio) of nitrogenous bases in DNA or RNA sequences molecule that are either guanine or cytosine. The coding regions have shown that the length of the coding sequence is directly proportional to GC-content. Furthermore, the stop codon has a bias toward T and A nucleotides, thus the shorter the sequence the higher the AT bias (Wuitschick and Karrer, 1999; Pozzoli et al., 2008; Romiguier et al., 2010; Manee et al., 2019).

Except the GC-content, there are two more measures: the GC-skew and the AT-skew, calculated as (G – C)/(G+C) and (A – T)/(A + T), respectively. All of the mentioned measures are used to understand various mutational patterns between the genomes, and as indicators of difference between the two strands (Gibson et al., 2005). For example, the negative GC-skew represents the cytosine richness over guanine. AT- and GT-skewness can be calculated either for the entire coding sequence, only the D-loop, or only for the PCGs (Manee et al., 2019). According to An et al. (2010) research of Caprine species' D-loop, A+T<G+C is present in all the domains of all five species. The ETAS and the central domain showed a base content of G>C>A>T and C>A>G>T, respectively. In all species, thymine (T) is the least displayed across all domains of the control region. Reyes et al. (1998) in their research analyzed 25 complete mammalian mitogenomes and investigated the relationship between the compositional features of all three positions in amino acids and third positions of fourfold degenerate codons. They concluded that transcribed H strand is richer in G base compared to L strand (L strand mRNAs are very poor in G, particularly in the third position), and that A nucleotide is more represented on L strand, while C and T differ among species.

Why Study Mitochondrial DNA?

MtDNA studies are a valuable tool for understanding evolutionary relationships, especially for the non-model species for which the nuclear genomes have not yet been assembled. Mitogenome sequences evolve rapidly, but gene regions of PCGs often stay unchanged over long periods of evolutionary time. For that reason, mitogenomes could give us valuable information about evolution of different taxa (Mereu et al., 2008; Manee et al., 2019; Prada and Boore, 2019).

Mitogenomes have several beneficial features. MtDNA is maternally inherited for most species since the paternal mitochondria are destroyed in the process of fertilization. Therefore, mtDNA lacks recombination (Clayton, 1992), so the haplotypes are shared between all individuals within a maternal line (Hutchinson et al., 1974; Gupta et al., 2015). Finally, mtDNA has well conserved protein-coding regions with little noncoding intergenic regions and with some overlapping between genes (Cantatore and Saccome, 1987; Taanman, 1999) and multicopy status in the cell (Robin and Wong 1988).

Using mtDNA in Ungulate Phylogeny

Ungulates present a large, diverse group of herbivores of large body-size. Based on the structure of their toes, they are divided into Cetartiodactyla, or the even-toed ungulates and Perissodactyla, or the odd-toed ungulates. Ungulates range from tropical forests to tundra regions, with between 250 and 450 species currently described (Groves and Grubb, 2011). The great diversity of such a large group makes it a good model for evolutionary paths studies, as well as for developing new methods of phylogenetic analysis. Many proposals have been made to resolve ungulate relationships, but there has been a lack of consensus between the results of morphological and molecular studies (Irwin et al., 1991; Allard et al., 1992; Hassanin and Douzery, 1999). Different phylogeny methods combined with analyses of nucleotide sequences (mostly mtDNA and its parts) have been used in the last 30 years to resolve their taxonomy. Phylogenetic tree constructed from the mtDNA is one of the major tools used to answer evolutionary questions (Chikuni et al., 1995; Budowle et al., 2003; Song et al., 2016). Gene arrangements are relatively similar within major groups of ungulates with few exceptions, and the differences between them could resolve some of the deepest branches of metazoan phylogeny. Even small changes in gene arrangements allow the estimation of relatedness and divergence times via calibrated molecular clocks (Tobe et al., 2010).

SCOPE OF THE REVIEWED LITERATURE

In order to find and survey the literature for our topic, we searched for titles, abstracts, and keywords in the Web of Science and Scopus database. Our search criteria were the

terms "mitochondrial", "mitogenome", "mtDNA", "mitogenome assembly", "mitogenome annotation" in combination with "ungulates", "wild ungulates", "even-toed ungulates", "Bovidae", "Caprine", "Artiodactyla", "Cetartiodactyla", Ruminantia", "Cervidae". We examined all publications that were likely to include information about mitogenome assembly and annotation, phylogenetic information and relationship between wild ungulates.

METHODS FOR mtDNA ASSEMBLY AND ANNOTATION

Next-generation Sequencing Technology

Next-generation sequencing technology (NGS) has revolutionized the field of genomics. Using NGS, an entire mitogenome can be sequenced and assembled in a very short period. NGS sequencers are unable to read the entire DNA strand at once, so the raw results are numerous readings of only short fragments of the entire sequence. On their own, such fragments offer very little information since every additional analysis requires a complete sequence of genes or mitogenomes. In order to reconstruct an entire sequence, reads must be arranged in the correct order. It is necessary to assemble those reads by overlapping or by using reference sequence (these can comprise a single gene, a group of genes or even an entire mitogenome). There are several NGS platforms with different approaches and properties (e.g. Roche, Solid). Compared to the Sanger method where reads are 650-800 bp long, Roche's 454 sequencer produces reads between 205-400 bp, while Solexa/Solid reads are generally within 100 bp.

Most of NGS platforms use short reads technology. On the other hand, the third-generation sequencing (or long-read sequencing) methods use a different approach, producing long reads (length between 2-5k bp). It is assumed that longer reads give more information and better results. However, long-read sequencing platforms give reads with high rates of random short indels and single nucleotide errors (Alkanaq et al., 2019). Moreover, production of long reads is expensive due to an increased cost in reagents and an increased running time of instruments, as well as a computer memory limitation (Ji et al., 2011). The most common platform for mitogenome assembly is Illumina that gives short reads within 25 or 150 bp, providing a large data output (15-1,800 Gbp). Illumina has a low sequencing price, and uses a different approach depending on research field (Song et al., 2016).

Mitogenome Assembly

Genome assembly is a process in which longer sequence is reconstructed from a collection of randomly sampled fragments using specific software. The aim is to create a mitogenome assembly with the longest possible assembled sequence and with the smallest number of mis-assemblies (Dominguez Del Angel et al., 2018). Before the development of the NGS technology, different methods were used for mitogenome assembly and the most popular ones were based on primer walking where mitochondria were isolated and amplified with PCR primers from closely related organisms (Bignell et al., 1996; Al-Nakeeb et al., 2017). However, the number of mitochondrial sequences in NGS data generated from whole-genome sequencing (WGS) is higher compared to nuclear sequences due to the high copy number of mitochondria per cell, what makes NGS data a valuable resource for extracting and assembling mitogenomes without isolation. For the same reason, the mitochondrial reads from WGS data will have a higher read depth compared to the nuclear genome which can be used to isolate the reads belong to mitogenome and use them for the assembly (Al-Nakeeb et al., 2017).

NGS results are stored in FASTQ files which contain small DNA fragments called reads that have limited information and because of that, reads should be assembled in contiguous sequences (or contigs) using dedicated software (Ye et al., 2011; Song et al., 2016). Contigs are assembled fragments of DNA constructed from smaller, overlapping fragments. Assembly of mitogenomes is easier compared to the whole genome assembly, due to its length. For reference, the length of the human genome is around 3 Gb, while the length of human mitogenome is about 16 kb (Gupta et al., 2015).

Before starting the assembly process, raw NGS data should be assessed and checked for quality (number of reads, GC content, number of duplicated reads, etc.). This step also includes cleaning, trimming, and correcting sequences that are too short or incomplete (Ekblom and Wolf, 2014). The most popular tool for assessing the quality of reads is FastQC (Andrews, 2010), which produces graphs and crucial statistics that show the average read quality. It is generally recommended to check the assembler documentation before starting the analyses to determine the input requirements (Dominguez Del Angel et al., 2018).

There are two main approaches to genome assembly. The first one is a genome assembly in which reads are mapped against the reference sequence from a related species. This approach generally requires less computational memory and time (Dominguez Del Angel et al., 2018). If the genomic sequence of related species is previously known, the assembly becomes less complex (Gordon et al., 1998). If there is no reference genome, the new genome is assembled using a *de novo* approach, whereby reads are compared and gathered into a longer sequence by overlapping them (Song et al., 2016; Paszkiewicz and Studholme, 2010; Dierckxsens et al., 2017).

When standard, whole genome assemblers are used for mtDNA assembly they usually do not result in a high-quality mitogenome sequence. This is because they have been programmed for assembly of nuclear genomes and the problems occur when those programs find the parts of sequence with extremely high depth (Meng et al., 2019). Furthermore, these assemblers have not been designed for generating circular genomes, so they often fail to recognize RS in mitogenomes of some species (mostly invertebrate species) (Dierckxsens et al., 2017; Meng et al., 2019).

The fact that ungulates have a very similar mitogenome structure makes most of the assemblers that are specialized for mitogenome assembly successful in mitogenome reconstruction. Several tools have been created specifically for mitogenome assembly and most of them are using Illumina reads.

Some of the most often used tools for *de novo* assembly of the mtDNA include MitoZ (Meng et al., 2019), MITObim (Hahn et al., 2013), NOVOPlasty (Dierckxsens et al., 2017), and Norgal (Al-Nakeeb et al., 2017). NOVOPlasty and MITObim are based on subtype of a *de novo* approach: "seed-and-extend", where a small

part of mitochondrial sequence (CYTB or COI gene) is used as a starting position from a related or distant single seed sequence (Dierckxsens et al., 2017; Alqahtani and Mandoiu, 2020). The seed sequence can also be a complete mitogenome from a very close species or subspecies. It is used to start an assembly process iteratively by scanning the start and stop positions. This algorithm uses the hash table that is a type of data structure for indexing genomes providing a list of genomic position for each possible read (Wu, 2016). After scanning the start and stop positions, the hash table is created, and similar reads are grouped together creating a circular sequence (Dierckxsens et al., 2017). The sequence will be complete when both ends overlap by at least 200 bp and when a consensus sequence is produced (Dierckxsens et al., 2017).

The most popular commercial tools for genome assembly are Geneious Prime ® (Kearse et al., 2012) and CLC (CLCbio, Aarhus, Denmark) assembler where one can easily manage to get a new mitogenome using both approaches, reference-based or *de novo*. Miller et al. (2012) set a good example by using CLC software to construct bighorn sheep (*O. canadensis*) mitogenome where they imported NGS reads (around 312 million short reads) into CLC and aligned them against a reference mitogenome of a domestic sheep (*O. aries*). Of the 312 million short reads, about 470,000 reads were mapped to reference genome, creating a new consensus sequence. This small number of mapped reads is usual for mitogenome construction since the mitogenome has 16 kb. The rest of the reads that are not mapped mostly belong to the nuclear sequence. The CLC approach was used for *de novo* assembly of mitogenome of Indian hog deer (*A. porcinus*) (Hill et al., 2017), while Caparroz et al. (2015) used de novo approach in Geneious Prime in order to assemble a new mitogenome of brown brocket deer (*M. gouazoubira*).

After the assembly, it is important to evaluate the completeness and contiguity of assembled mitogenome, since errors could occur for many reasons (e.g. RS, wrong orientation etc.). The assembled sequence may contain not only the information about the wanted mitochondrial DNA, but also misassembled bases originating from other sequences (i.e. nuclear, organelle or even contaminating DNAs), with common errors such as duplications and deletions of some regions. Therefore, the first step is to compare the size of consensus genome with the genome of a related species. Sequence homology check by BLAST algorithm will confirm completion of the target genome. The second step is read-mapping to the consensus sequence. This step will show the parts of genome that need more sequence data to ensure the accuracy. For circular genomes, connection of both ends needs to be checked by additional read-mapping. If both ends are not connected, this gap must be filled using Sanger sequencing or an additional NGS run. Read-mapping can be performed using any of the currently available programs (e.g. CLC, Geneious, BWA (Li and Durbin, 2009) and Bowtie (Langmead, 2011)) which include an aligning function (Yandell and Ence, 2012). Mohandesan et al. (2017) used a similar approach for camel mitogenome. They first used BWA for mapping NGS reads against reference genome and then, they imported all mapped reads into CLC to create a consensus sequence.

It is understandable why mitogenome sequencing became relatively easy and cost-effective. This resulted in a huge amount of sequenced mitogenomes in online databases, and at the same time, in the development of programs and open source tools for genome assembly and annotation.

To confirm this, in April of 2020, we searched GenBank data for complete mitogenome sequences of even-toed ungulates. From given results we excluded Cetacea and Delphinidae records. We got 8074 sequences, both the complete mitogenomes sequences and sequences that present only small parts of mitogenomes (D-loop, PCGs, etc.). In order to get only complete sequences, we included option "Sequence length" and specified mitogenome length from 16000 to 17000 bp that is characteristic for mammalian mitogenome. This resulted in 2652 mitogenome sequences. The most common sequences in this search belong to cattle (560), pig (299) and goat (256).

Mitogenome Prediction and Annotation

After completing the draft genome construction, it is necessary to identify and annotate the genes. Genome annotation is the process where biological information is attached to genome sequence. It is performed by first analyzing the draft genome sequence structure and composition, and then comparing it to a known genome sequence of a closely related species (Dominguez Del Angel et al., 2018). Usually, the quality of mitogenome assembly (estimated as similarity with a reference sequence) is between 90 and 100%, which represents a very good result. This measure is very important for the annotation step, since annotation strongly depends on the quality of the assembled genome. The mitogenomes that have an above 90 % completeness typically yield satisfying annotation (Ekblom and Wolf, 2014). The annotation process can be done manually or through an automated computer analysis. The very process can be divided into two main steps: 1) structural annotation (gene prediction) that consists of the identification of genomic elements (open reading frames, coding regions, gene structure), and 2) functional annotation that consists of attaching biological information to genomic elements (biological function, expression). The first step is gene prediction, the process of identifying parts of encoded genes that are likely to occur in the sequence (Wang et al., 2004). This process determines genes' location and their structure in the genome (Dominguez Del Angel et al., 2018). The procedure involves translating the nucleotide sequence and finding open reading frames. There are two methods for gene prediction procedures: the similarity-based and the *ab initio* method. The first approach tries to find similarities in the gene region between the sequences and it is based on the assumption that exons are more conserved evolutionary than the other nonfunctional regions. On the other hand, *ab initio* uses gene structure as a template to detect genes using two types of sequence information: signal sensors (refers to short sequence motifs such as splice site, start and stop codons, etc.) and content sensors, used for exon detection in a way that allows coding sequences to be distinguished from the surrounding non-coding regions (Wang et al., 2004). *Ab initio* gene predictor, using mathematical models, identifies genes together with their intron-exon parts and uses genomic traits such as codon frequencies and lengths of intron-exon regions for the purpose of distinguishing genes from the intergenic regions (Korf, 2004). The functional annotation process assigns biologically relevant information to the predicted proteins and to the features they derive from (e.g. gene, mRNA). The annotation of nuclear genome is usually specific since every

genome is different, and it is necessary to rebuild and retrain genome regions for each new species. Contrarily, the mitogenome annotation is easier since most vertebrate mitochondria are very similar. The function of predicted proteins can be defined by comparison of a given sequence and the sequences from different public repositories. In the last 15 years, several mitogenome annotation programs have been developed in order to fulfill mitogenome annotation step automatically. The most popular tools for annotation are DOGMA (Wyman et al., 2004), MITOS (Bernt et al., 2013), GeSeq (Tillich et al., 2017) and MitoZ (Meng et al., 2019). The DOGMA, MITOS and GeSeq are web applications for annotation, while MitoZ is an open source software. Most of the tools require a reference sequence from a related species in FASTA format. GeSeq can read GenBank format files, with additional information about the annotated regions. MitoZ can perform the annotation step without a reference sequence, but it is highly recommended to use a reference, since every tool uses GenBank data for comparison. The fastest software is GeSeq, which also has the option to annotate multiple sequences. Usually, tools for annotation use the BLAST method, while GeSeq uses the BLAT method to identify the coding and the non-coding parts by searching for open reading frames from similar sequences in Genbank. The tRNA scan-SE software (Lowe and Eddy, 1997; Lowe and Chan, 2016) is used for annotation of tRNA genes and it is implemented into most of the annotation software (Cui et al., 2007; Jiang et al., 2013; Matosiuk et al., 2014; Zhou et al., 2019). Identifications of the D-loop elements are usually based on previous reference data from closely related species. MitoZ uses MiTFi (Jühling et al., 2012) to annotate tRNA genes. Overall, the fastest tools are GeSeq and MITOS.

After the annotation step, it is necessary to check whether the lengths of the annotated genes are similar to the reference. It is also very important to make sure that the stop codons are used correctly, since indels can cause the presence of stop codons in the middle of gene (Bernt et al., 2013). Analyzing those elements can provide information about specific genome properties and similarities compared to the closely related species. It also includes an additional quality check for the predicted gene set where it is possible to identify problematic regions by the presence of specific domain, ortholog assignment, or similar. The output of a genome annotation is most often in a GFF format and it includes structural and functional features of the mitogenome, but not the actual sequence. Other output formats are GTF, BED, Genbank and EMBL, of which the last two include both sequence and annotation information (Domiguez Del Angel et al., 2018)

USING mtDNA IN UNGULATE PHYLOGENY

Problems with Mitogenomes

The number of mtDNA sequences in GenBank increases rapidly. According to the currently available literature, mitogenomes are considered well researched (Boore et al., 2005). However, it is necessary to be careful when using reference sequences from GenBank because of the possibility that the sequences are incomplete, incorrect, or contain pseudogenes. Such problems could have consequences for the future genomic studies of mtDNA and could cause misleading interpretations (Hassanin et al., 2010). Prada and Boore (2019) investigated the mitogenome annotations of the 304 complete sequences representing 29 taxonomic orders available on GenBank. They were comparing nucleotide sequences of orthologous regions of evolutionarily close species and found a significant number of rearrangements confirming errors in annotation. These include false inversion of genes encoding tRNA and partial or complete deletions of genes encoding tRNA and the D-loop regions. This problem was also highlighted by Hassanin et al. (2010), who pointed out the errors in a domestic goat (*C. hircus*) reference sequenced by Parma et al. (2003). To re-examine their findings, they independently assembled the goat mitogenome and compared it with mitogenomes available for the goat (six sequences) and four Caprine species. Their phylogenetic analyses discovered that Parma et al. (2003) sequenced only 44.5% of the total mtDNA and that five of the six goat mitogenomes available on Genbank were contaminated by the NuMt (nuclear sequences of mitochondrial origin) fragments. Furthermore, Prada and Boore (2019) indicated that errors in mitogenome annotation were probably present in many deposited mitogenomes, which should be corrected by the curators of the NCBI. On the other hand, many authors, when exploring and assembling new mitogenomes, sparsely explain their methods or hypotheses they tested (Hu et al., 2015; Hu et al., 2016; Hill et al., 2017; Mao et al., 2017; Nguyen et al., 2017). They usually give information about the program they used and how many genes the new mitogenome contains. This problem was addressed by Smith (2016), with the author concluding that mitochondrial research has become repetitive and suffers from a general lack of hypothesis testing. Besides that, it is essential to use proper mitogenome annotation and correctly deposit it to the GenBank.

A Brief Historical Summary of the Usage of Mitochondrial Segments

Within the ungulate clade, the Ruminantia suborder is the most diverse group among the currently extant species. Their proposed taxonomy is based on the analyses of different parts of their mitochondrial DNA. The segments of mtDNA that were most often used were two rRNA genes (16S and 12S) (Miyamoto et al., 1989; Allard et al., 1992; Gatesy et al., 1992), CYTB (Irwin et al., 1991; Chikuni et al., 1995; Groves and Shields, 1996; Groves and Shields, 1997; Hassanin et al., 1998; Randi et al., 1998; Hassanin and Douzery, 1999;) and the D-loop (Douzery and Randi, 1997; Polziehn and Strobeck, 2002). Molecular analyses based on the single-gene sequences placed most of Ruminantia species into specific genera or tribes, but the relationships among some of the groups were not determined (Gentry, 1990). For some species, those analyses were not conclusive. Good examples are the Saiga antelope (*S. tatarica*) and the chiru (*P. hodgsonii*). Their phylogenetic status has been debated for years and remains unresolved. Grubb (1993) and McKenna and Bell (1997) proposed that those two species should be included in the Antilopinae subfamily of Bovidae (bovids), while Gentry (1992) recommended their incorporation into Caprinae. To resolve the relationship between the two species and other Caprinae, Hassanin et al. (1998) conducted a complete CYTB sequence for 18 species of Caprinae. The authors concluded that most of the analyses based on a single gene (in this case, the CYTB gene) led to limited conclusions and that other parts of mitogenome and larger taxa samples should be included in further analyses in order to confirm or refute the given hypothesis. However, their analyses

contradicted the classical systematics of Caprinae. They proposed that the Saiga should be excluded from the Caprinae (the same results were given by Chikuni et al. (1995)) while the results placed the chiru as a sister taxon to Caprinae. Irwin et al. (1991) used CYTB sequences to construct the phylogeny tree of bovids and Cervidae (cervids), where the analysis placed both families in one group. To confirm this result, Chikuni et al. (1995) sequenced the complete CYTB of the lesser mouse deer (*T. javanicus*), sika deer (*C. nippon*), water buffalo (*B. bubalis*), and Japanese serow and preformed the analysis together with the CYTB sequences published in the previous reports (Anderson et al., 1981; Irwin et al., 1991). The northern giraffe (*G. camelopardalis*), pronghorn (*A. americana*), and fallow deer (*D. dama*) were separated from bovids, but the sika deer and black-tailed deer (*O. hemionus*) were placed together with bovids with low bootstrap confidence levels. They concluded that given results might be different because of the rapid evolution in synonymous sites and the biased transition/transversion ratio in mtDNA and because the saturation of nucleotide substitutions could affect the evolutionary analyses. Another reason includes an introgressive hybridization between ancestral species of bovids and cervids. The cervid evolution is also complex, and Grubb (1993) has proposed the identification system for cervids that classifies deer genera into four subfamilies. Randi et al. (1998) provided new phylogenetic information on relationship among cervids using CYTB sequences, while Kuwayama and Ozawa (2000) used the same approach on the European red deer (*C. elaphus*), wapiti (*C. canadensis*) and sika deer to resolve their relationships, since the previous studies showed that those three species were monophyletic. Their analysis showed that the wapiti was more closely related to the sika deer than to the European red deer. Those conclusions were conflicted with traditional taxonomy results based on morphology, which suggested a close relationship between the wapiti and the European red deer. However, the new analyses of the Cetacea additionally complicated the relationships within ungulates, putting the Cetacea as the sister to the Hippopotamidae. This suggested an inclusion of Artiodactyla and Cetacea into a single order named Cetartiodactyla (Montgelard et al., 1997; Ursing and Arnason, 1998; Nikaido et al., 1999; Ursing et al., 2000). Besides that, placing the Moschidae family into Ruminantia was also questionable since the Moschidae are placed between cervids and bovids. The next step in studying evolutionary relationship was to use a combination of the mentioned mtDNA fragments with sequences of nuclear genes such as κ-casein (Chikuni et al., 1995; Cronin et al., 1996), β-casein (Gatesy et al., 1996), γ- fibrinogen and other genetic loci (Gatesy, 1997; Hassanin and Douzery, 1999; Gatesy and Arctander, 2000; Matthee and Davis, 2001). For all these studies, Hassanin and Douzery (2003) concluded that the taxonomic samples were not appropriately determined, since the representatives of several groups were not included. For that reason, the same authors sequenced molecular markers that could show different directions of cetartiodactyl evolution. They analyzed a large dataset consisting of 23 species, which consisted of CYTB gene and two rRNA gene sequences and nuclear markers that revealed new insights, such as that Mochidae were closer to bovids rather than to cervids.

Analyses of the D-loop region have been conducted for the purpose of establishing intra-specific and inter-specific relations, determining maternal contributions, and tracing the origin of modern and ancient animals (Gupta et al., 2015). The first D-loop comprehensive study was performed by Sbisà et al. (1997), including D-loop sequences from ten different mammalian orders. Furthermore, RS in the D-loop have been used for phylogenetic comparisons among cervids due to the absence of variation within individuals (Cook et al., 1999). Polziehn and Strobeck (2002) were using the D-loop and CYTB to test the hypothesis that red deer was a distinct species from wapiti. They highlighted that number of RS within cervids could be used to infer relatedness. However, using RS of the D-loop is limited to studying relations among different species, and it depends on number of RS for given taxa. Ursing et al. (2000) showed that the RS of alpaca (*V. pacos*), sheep (*O. aries*), and pig (*S. scrofa domesticus*) were profoundly different and that studying their RS relations provided limited information. Additionally, many studies failed to resolve the basal artiodactyl divergences because their analyses were based on short sequences on several genes (Ursing et al., 2000)

Using Whole Mitogenomes in Phylogenetics

Mitogenome sequence comparisons are much more informative than single gene comparisons (Ingman et al., 2000; Ursing et al., 2000; Boore et al., 2005; Miller et al., 2012; Kim et al., 2014) because they contain more polymorphic sites and they can provide significant insights into the evolution of organisms and mitogenomes (Boore, 1999). Ursing and Arnason (1998) and Ursing et al. (2000) were among the first authors to present the usage of complete mitogenomes for phylogeny analyses, when they used the sequences of 12 PCGs with the combination of the D-loop. Phylogenetic studies are usually performed by comparing data sets that contain the whole mitogenome data or data that contain 12 (+1) concatenated sequences of PCGs. The ND6 is usually excluded because it is encoded by the L strand which has a different base composition from the H strand (Gibson et al., 2005; Mereu et al., 2008; Miller et al., 2012; Douglas et al., 2011; Jiang et al., 2013; Świsłocka et al., 2020). The rationale of using only PCGs is that they are more conserved, and therefore less prone to assembly errors. On the other hand, this approach reduces the number of polymorphic sites, leading to lower resolution power in phylogenies. Zhou et al. (2019) used sequences that contained 13 PCGs and 2 rRNA genes but without any initiation and termination codons while Hassanin et al. (2009), Matosiuk et al. (2014), and Mohandesan et al. (2017) created data for every gene family separately (ND, ATP, COI, CYTB) in order to explore comparing rates of synonymous and nonsynonymous substitutions (ω=dN/dS). Hassanin et al. (2009) concluded that the largest differences had been found in the ATP gene family while the other three complexes were more conserved. Matosiuk et al. (2014) concluded that the largest differences had been found in the ND gene family while Mohandesan et al. (2017) found the highest differences in the CYTB gene.

All approaches for mitogenome analyses may reveal differences such as gene gains, duplications, rearrangements, or inversions of gene fragments. Zurano et al. (2019) used a different approach to resolve phylogenetic relationships of Cetartiodactyla, including two datasets. The first set contained mitogenome sequences of 225 species, and the second set contained data from 93 species that had at least one mitochondrial gene available. Both datasets gave nearly identical topologies that had small differences in divergence

time estimates, which meant that using smaller sets of data (i.e. a single gene, CYTB, or the D-loop) could have confirmed the given results from the complete mitogenome analyses. A similar approach has been used by Meadows et al. (2011) and Jiang et al. (2013), who examined four datasets to resolve the phylogeny of wild sheep species (*O. musimon, O.vignei,* and *O. ammon hodgsoni*) and create phylogenetic trees. Their approach was based on 1) whole mitogenomes; 2) concatenated sequence from smaller regions (D-loop and protein-coding regions); 3) CYTB regions, and 4) D-loop regions. Jiang et al. (2013) concluded that all the analyses showed identical topology with high support by posterior probability and bootstrap values. Meadows et al. (2011) concluded that using fragments like CYTB or some other PCGs would not give full information about relationships and that using the whole mitogenome sequence might successfully resolve phylogenetic relationships between animals. Similar explanation to Meadows et al. (2011) has been given by Arif et al. (2012) who used 12 representative Bovidae species to address the question whether single genes (two rRNA gens, COI, CYTB and D-loop) can provide the same phylogenetic information as compared to complete mitochondrial sequences. The phylogenetic trees constructed from a single gene sequences were not identical with the given results of whole mitogenome sequences (the trees obtained by analysis of the CYTB gene showed differences compared to other trees from a single gene). They concluded that the use of complete mitogenome sequences should be preferred over individual genes. However, Naseem et al. (2020) used the COI gene for the identification of wild ungulates from Pakistan. They analyzed 86 specimens of 19 wild ungulates species. With intraspecific and interspecific distances and with neighbour-joining tree, they were able to discriminate all species into their respective clades. Their conclusion, however, was similar to those of previous research, suggesting that the COI gene can be an efficient marker for species identification. However, it is recommended to include a larger taxon sampling and more mitochondrial genes in order to resolve taxonomic questions. Phylogenetic relationships cannot be fully described for all ungulate clades or species using only mtDNA since some linages have limited mitogenomic diversity. Świsłocka et al. (2020) compared mitogenomic diversity of the European moose (*A. alces*) and other cervids by comparing whole mitogenome sequences. They found only between 13 and 18 fixed nucleotide substitutions in the European moose mtDNA, which indicated a lower diversity of the whole mitogenome sequences than in the other hoofed mammals.

Adaptive Evolution of Mitochondria

Variation of protein-coding regions of a mitochondrial genome can directly influence metabolic performances and seriously affect chemical processes in the mitochondria (Blier et al., 2001; Ballard and Whitlock, 2004; Mishmar et al., 2003; Mohandesan et al., 2017). However, changes in amino acids may also improve the aerobic capacity and adaptation to new environments since the metabolic function varies widely among mammalian species (Suarez et al., 2004; Mohandesan et al., 2017), and since variation in the OXPHOS regions has been connected to different life history traits and environment adaptations (McNab, 2000). Finding these differences could give us useful information about the adaptive evolution of the mitochondria within species (Blier et al., 2001). A good example is research from da Fonseca et al. (2008), who explored how the mitochondrial genetic variation might be connected to the diverse metabolic patterns of 41 mammalian species. They combined molecular analyses with secondary structure prediction analyses, and used estimates of metabolic rates based on oxygen consumption under aerobic conditions for physiological differences among species. The genes that had the largest changes in the amino acid site were ND and ATP genes, and the ones that had the smallest changes were CYTB and CO genes. Although variations in the CYTB region are rare and usually hard to detect, several variations were found in CYTB regions (changes per site in elephants, cetaceans, seals, foxes, and bats). Usually, these changes were not crucial but may reflect better animal adaptation to environment. Hassanin et al. (2009) have studied and compared sequences of 17 Caprine and 18 Bovine species that are well adapted to life at high altitudes. They calculated the ratio of natural selection operating at the amino acid sequence level by comparing rates of synonymous and nonsynonymous substitutions (ω=dN/dS). These rates give a measure of selection at the protein level, where $\omega > 1$ indicates positive selection, while $\omega < 1$ indicates negative selection. Their analyses showed the highest ω values were found for ATPase and CYTB, and the lowest for ND and CO regions. Analyses revealed that ω has increased during the Caprinae evolution, suggesting higher levels of selective pressures in this taxonomic group. Usually, the mtDNA rate of ω is below 1, which means that most of nonsynonymous mutations are eliminated by purifying selection (Bazin et al., 2006; Meiklejohn et al., 2007). Xu et al. (2005) investigated genetic mechanisms of adaptations to the high altitude of native mammals on the Tibetan Plateau. Evidence from the synonymous and nonsynonymous ratio showed that the COI gene had more functional mutations in the yak (*B. grunniens*) and Tibetan antelope (chiru) compared to other mammals. Therefore, they concluded that changes in the COI gene were under a positive selection for Tibetan antelope and yak because of their adaptation to the high altitude unique to the Tibetan Plateau. Similar results have been shown by Mohandesan et al. (2017) about the genus *Camelus* that is well-adapted to environments of varying altitude and temperature. The same authors found a high variation in the CYTB that could be explained by the more specialized metabolic requirements of camels (adaptation to a low-energy diet or to living at higher altitudes).

CONCLUSION

Mitochondrial DNA is most often used to examine the taxonomic and phylogenetic relationships of various taxonomic groups of organisms, including ungulates. Initially, only parts of mtDNA (single genes or regions) were used to address phylogenetic relationships between genera and families. However, some relationships could not be resolved because of the low number of informative (variable) sites in used sequences, so new methods were developed that used whole mitogenomes.

The process of assembly and annotation of ungulate mitogenomes is automated, and many research groups have access to comprehensive sets of data, as well as specialized software. Based on the literature review, for mitogenome assembly we recommend using specialized software that use reference sequences such as NOVOPlasty or MITObim. Before it is used, the reference sequence should be checked for quality. For the annotation step, we recommend GeSeq, MitoZ and tRNA scan-

SE software. Furthermore, to avoid ambiguity of the assembled sequence, we recommend using at least two programs for assembly and annotation and then comparing the results. Most research papers describing new mitogenomes lack detailed descriptions of the assembly and annotation methods, making it hard to replicate the experiments. This means that it would be extremely useful to find a unique approach for analyzing mitogenomes.

Besides the constant increase and improvement of available methods for mtDNA sequencing and analysis, there are some persistent challenges still limiting the use of mtDNA or influencing the results of the analyses. Overview of the literature revealed examples of errors in assembly and annotation of deposited mtDNA sequences, insufficient descriptions of methodologies used in published papers, inconsistent results depending on the different methods applied to the same sets of sequences. There seems to be, however, lack of interest in systematic approach to identify and solve these problems. Most of the reviewed literature is focused on utilization of mtDNA in solving taxonomical problems using previously described methodology, without consideration for potential sources of errors (that can occur during sequencing, assembly, annotation or any other step in the analysis).

ACKNOWLEDGMENTS

This research was funded by the Croatian Science Foundation (CSF) within the project "Young Researchers' Career Development Project – Training of Doctoral Students", co-financed by the European Union, under the OP "Efficient Human Resources 2014-2020" from the ESF funds. This research was also founded by projects "DNA as evidence of the distribution and vitality of endangered Balkan chamois", "The role of hunting related activities in the range expansion of recently established wild ungulate populations in the Mediterranean" founded by the Croatian Science Foundation, and the RESBIOS European Union's Horizon 2020 Research and Innovation Program (No. 872146).

REFERENCES

Alkanaq A.N., Hamanaka K., Sekiguchi F., Taguri M., Takata A., Miyake N., Miyatake S., Mizuguchi T., Matsumoto N. (2019). Comparison of Mitochondrial DNA Variants Detection Using Short- and Long-read Sequencing. J Hum Genet 64 (11): 1107–1116. doi:10.1038/s10038-019-0654-9

Allard M.W., Miyamoto M.M., Jarecki L., Kraus F., Tennant M.R. (1992). DNA Systematics and Evolution of the Artiodactyl Family Bovidae. Proc Natl Acad Sci U S A 89 (9): 3972–3976. doi:10.1073/pnas.89.9.3972

Al-Nakeeb K., Petersen T.N., Sicheritz-Pontén T. (2017). Norgal: Extraction and *de novo* Assembly of Mitochondrial DNA from Whole-genome Sequencing Data. BMC Bioinformatics 18 (1): 1–7. doi:10.1186/s12859-017-1927-y

Alqahtani F., Măndoiu I.I. (2020). Statistical Mitogenome Assembly with RepeaTs. J Comput Biol 27 (0): 1–15. doi:10.1089/cmb.2019.0505

An J., Okumura H., Lee Y.S., Kim K.S., Min M.S., Lee H. (2010). Organization and Variation of the Mitochondrial DNA Control Region in Five Caprinae Species. Genes and Genomics 32 (4): 335–344. doi:10.1007/s13258-010-0023-8

Anderson S., Bankier A.T., Barrell B.G., De Bruijn M.H.L., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F., Schreier P.H., Smith A.J.H., Staden R., Young I.G. (1981). Sequence and Organization of the Human Mitochondrial Genome. Nature 290 (5806): 457–465. doi:10.1038/290457a0

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

Arif I.A., Bakir M.A., Khan H.A. (2012). Inferring the Phylogeny of Bovidae Using Mitochondrial DNA Sequences: Resolving Power of Individual Genes Relative to Complete Genomes. Evol Bioinforma 2012 (8): 139–150. doi:10.4137/EBO.S8897

Ballard J.W.O., Whitlock M.C. (2004). The Incomplete Natural History of Mitochondria. Mol Ecol 13 (4): 729–744. doi:10.1046/j.1365-294X.2003.02063.x

Barchiesi A., Vascotto C. (2019). Transcription, Processing, and Decay of Mitochondrial RNA in Health and Disease. Int J Mol Sci 20 (9): 2221. doi:10.3390/ijms20092221

Barshad G., Marom S., Cohen T., Mishmar D. (2018). Mitochondrial DNA Transcription and Its Regulation: An Evolutionary Perspective. Trends Genet 34 (9): 682–692. doi:10.1016/j.tig.2018.05.009

Bazin E., Glémin S., Galtier N. (2006). Population Size Does not Influence Mitochondrial Genetic Diversity in Animals. Science 312 (5773): 570–572. doi:10.1126/science.1122033

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M., Stadler P.F. (2013). MITOS: Improved *de novo* Metazoan Mitochondrial Genome Annotation. Mol Phylogenet Evol 69 (2): 313–319. doi:10.1016/j.ympev.2012.08.023

Bignell G.R., Miller A.R., Evans I.H. (1996). Isolation of Mitochondrial DNA. Methods Mol Biol 53 (332258): 109–116. doi:10.1385/0-89603-319-8:109

Blier P.U., Dufresne F., Burton R.S. (2001). Natural Selection and the Evolution of mtDNA-encoded Peptides: Evidence for Intergenomic Co-Adaptation. Trends Genet 17 (7): 400–406. doi:10.1016/S0168-9525(01)02338-1

Boore J. (1999). Animal Mitochondrial Genomes. Nucleic Acids Res 27 (8): 1767–1780. doi:10.1093/nar/27.8.1767

Boore J.L., Macey J.R., Medina M. (2005). Sequencing and Comparing Whole Mitochondrial Genomes of Animals. Methods Enzymol 395: 311–348. doi:10.1016/S0076-6879(05)95019-2

Brown G.G., Simpson M. V. (1982). Novel Features of Animal mtDNA Evolution as Shown by Sequences of Two Rat Cytochrome Oxidase Subunit II Genes. Proc Natl Acad Sci U S A 79 (10 I): 3246–3250. doi:10.1073/pnas.79.10.3246

Brown J.R., Beckenbach K., Beckenbach A.T., Smith M.J. (1996). Length Variation, Heteroplasmy and Sequence Divergence in the Mitochondrial DNA of Four Species of Sturgeon (Acipenser). Genetics 142 (2): 525–535

Brown W.M., George M., Wilson A.C. (1979). Rapid Evolution of Animal Mitochondrial DNA. Proc Nat Acad Sci USA, 76:1967-1971.

Brown WM. (1985). The Mitochondrial Genome of Animals; in Macintyre RJ, Molecular Evolutionary Genetics. Plenum, New York

Budowle B., Allard M.W., Wilson M.R., Chakraborty R. (2003). Forensics and Mitochondrial DNA: Applications, Debates, and Foundations. Annu Rev Genomics Hum Genet 4 (1): 119–141. doi:10.1146/annurev.genom.4.070802.110352

Cantatore P., Saccone C. (1987). Organization, Structure and Evolution of Mammalian Mitochondrial Genes. Int Rev Cytol 108:149-208. doi: 10.1016/s0074-7696(08)61438-2

Caparroz R., Mantellatto A.M.B., Bertioli D.J., Figueiredo M.G., Duarte J.M.B. (2015). Characterization of the Complete Mitochondrial Genome and a Set of Polymorphic Microsatellite Markers through Next-Generation Sequencing for the Brown Brocket Deer Mazama gouazoubira. Genet Mol Biol 38 (3): 338–345. doi:10.1590/S1415-475738320140344

Chikuni K., Mori Y., Tabata T., Saito M., Monma M., Kosugiyama M. (1995). Molecular Phylogeny Based on the κ-casein and Cytochrome b sequences in the Mammalian Suborder Ruminantia. J Mol Evol 41 (6): 859–866. doi:10.1007/BF00173165

Clayton D.A. (1992). Structure and Function of the Mitochondrial Genome. J Inherit Metab Dis 15 (4): 439–447. doi:10.1007/BF01799602

Cook C.E., Wang Y., Sensabaugh G. (1999). A Mitochondrial Control

Region and Cytochrome b Phylogeny of Sika Deer (Cervus nippon) and Report of Tandem Repeats in the Control Region. Mol Phylogenet Evol 12 (1): 47–56. doi:10.1006/mpev.1998.0593

Cronin M.A., Stuart R., Pierson B.J., Patton J.C. (1996). K-casein Gene Phylogeny of Higher Ruminants (Pecora, Artiodactyla). Mol Phylogenet Evol 6 (2): 295–311. doi:10.1006/mpev.1996.0078

Cui H., Kong Y., Zhang H. (2012). Oxidative Stress, Mitochondrial Disfunction, and Aging. J Signal Transduct 2012: 1–13. doi:10.1155/2012/646354

Cui P., Ji R., Ding F., Qi D., Gao H., Meng H., Yu J., Hu S., Zhang H. (2007). A Complete Mitochondrial Genome Sequence of the Wild Two-Humped Camel (*Camelus bactrianus ferus*): an Evolutionary History of Camelidae. BMC Genomics 8: 241. doi:10.1186/1471-2164-8-241

da Fonseca R.R., Johnson W.E., O'Brien S.J., Ramos M.J., Antunes A. (2008). The Adaptive Evolution of the Mammalian Mitochondrial Genome. BMC Genomics 9: 1–22. doi:10.1186/1471-2164-9-119

Dierckxsens N., Mardulyn P., Smits G. (2017). NOVOPlasty: *De novo* Assembly of Organelle Genomes from Whole Genome Data. Nucleic Acids Res 45 (4). doi:10.1093/nar/gkw955

Dominguez Del Angel V., Lantz H., Hjerde E., Sterck L., Capella-Gutierrez S., Notredame C., Vinnere Pettersson O., Amselem J., Bouri L., Bocs S., Klopp C., Gibrat J.F., Vlasova A., Leskosek B.L., Soler L., Binzer-Panchal M. (2018). Ten Steps to Get Started in Genome Assembly and Annotation. F1000Research 7. doi:10.12688/f1000research.13598.1

Douglas K.C., Halbert N.D., Kolenda C., Childers C., Hunter D.L., Derr J.N. (2011). Complete Mitochondrial DNA Sequence Analysis of Bison Bison and Bison-Cattle Hybrids: Function and Phylogeny. Mitochondrion 11 (1): 166–175. doi:10.1016/j.mito.2010.09.005

Douzery E., Randi E. (1997). The Mitochondrial Control Region of Cervidae: Evolutionary Patterns and Phylogenetic Content. Mol Biol Evol 14 (11): 1154–1166. doi:10.1093/oxfordjournals.molbev.a025725

Ekblom R., Wolf J.B.W. (2014). A Field Guide to Whole-Genome Sequencing, Assembly and Annotation. Evol Appl 7 (9): 1026–1042. doi:10.1111/eva.12178

Foster P.G., Jermiin L.S., Hickey D.A. (1997). Nucleotide Composition Bias Affects Amino Acid Content in Proteins Coded by Animal Mitochondria. J Mol Evol 44 (3): 282–288.doi:10.1007/PL00006145

Fumagalli L., Taberlet P., Favre L., Hausser J. (1996). Origin and Evolution of Homologous Repeated Sequences in the Mitochondrial DNA Control Region of Shrews. Mol Biol Evol 13 (1): 31–46. doi:10.1093/oxfordjournals.molbev.a025568

Garesse R., Vallejo C.G. (2001). Animal Mitochondrial Biogenesis and Function: A Regulatory Cross-Talk between Two Genomes. Gene 263 (1–2): 1–16. doi:10.1016/S0378-1119(00)00582-5

Gatesy J. (1997). More DNA Support for a cetacea/hippopotamidae Clade: The Blood-Clotting Protein Gene γ-fibrinogen. Mol Biol Evol 14 (5): 537–543. doi:10.1093/oxfordjournals.molbev.a025790

Gatesy J., Arctander P. (2000). Hidden Morphological Support for the Phylogenetic Placement of *Pseudoryx nghetinhensis* with Bovine Bovids: A Combined Analysis of Gross Anatomical Evidence and DNA Sequences from Five Genes. Syst Biol 49 (3): 515–538. doi:10.1080/10635159950127376

Gatesy J., Hayashi C., Cronin M.A., Arctander P. (1996). Evidence from Milk Casein Genes that Cetaceans Are Close Relatives of Hippopotamid Artiodactyls. Mol Biol Evol 13 (7): 954–963. doi:10.1093/oxfordjournals.molbev.a025663

Gatesy J., Yelon D., Des R., Vrba E.S. (1992). Phylogeny of the Bovidae (Artiodactyla, Mammalia), Based on Mitochondrial Ribosomal DNA Sequences. Mol Biol Evol 9(3): 433–446

Gentry A.W. (1992). The Subfamilies and Tribes of the Family Bovidae. Mamm Rev 22 (1): 1–32. doi:10.1111/j.1365-2907.1992.tb00116.x

Gentry, A. W. (1990). Evolution and Dispersal of African Bovidae. In: Horns, Pronghorns, and Antlers (Bubenik, G.A., Bubenik, A.B., eds), SpringerLink, New York, United States of America, pp. 195-227.

Gibson A., Gowri-Shankar V., Higgs P.G., Rattray M. (2005). A Comprehensive Analysis of Mammalian Mitochondrial Genome Base Composition and Improved Phylogenetic Methods. Mol Biol Evol 22 (2): 251–264. doi:10.1093/molbev/msi012

Gordon D., Abajian C., Green P. (1998). Consed: A Graphical Tool for Sequence Finishing. Genome Res 8 (3): 195–202. doi:10.1101/gr.8.3.195

Groves C., Grubb P. (2011). Ungulate Taxonomy. Baltimore, MD: Johns Hopkins University Press.

Groves P., Shields G.F. (1996). Phylogenetics of the Caprinae Based on Cytochrome B Sequence. Mol Phylogenet Evol 5 (3): 467–476. doi:10.1006/mpev.1996.0043

Groves P., Shields G.F. (1997). Cytochrome B Sequences Suggest Convergent Evolution of the Asian Takin and Arctic Muskox. Mol Phylogenet Evol 8 (3): 363–374. doi:10.1006/mpev.1997.0423

Grubb, P. (1993). Family Bovidae. In: Mammal Species of the World. A Taxonomic and Geographic Reference, D. E. Wilson and D. M. Reeder, eds., pp. 393-414, Smithsonian Institution Press, Washington, DC.

Gupta A., Bhardwaj A., Sharma S.P., Pal Y., Kumar M., Kumar S. (2015). Mitochondrial DNA- A Tool for Phylogenetic and Biodiversity Search in Equines. J Biodivers Endanger Species 01 (s1). doi:10.4172/2332-2543.s1-006

Hahn C., Bachmann L., Chevreux B. (2013). Reconstructing Mitochondrial Genomes Directly from Genomic Next-Generation Sequencing Reads - A Baiting and Iterative Mapping Approach. Nucleic Acids Res 41 (13). doi:10.1093/nar/gkt371

Hassanin A., Bonillo C., Nguyen B.X., Cruaud C. (2010). Comparisons between Mitochondrial Genomes of Domestic Goat (Capra hircus) Reveal the Presence of Numts and Multiple Sequencing Errors. Mitochondrial DNA 21 (3–4): 68–76. doi:10.3109/19401736.2010.490583

Hassanin A., Douzery E.J.P. (1999). Evolutionary Affinities of the Enigmatic Saola (Pseudoryx nghetinhensis) in the Context of the Molecular Phylogeny of Bovidae. Proc R Soc B Biol Sci 266 (1422): 893–900. doi:10.1098/rspb.1999.0720

Hassanin A., Douzery E.J.P. (2003). Molecular and Morphological Phylogenies of Ruminantia and the Alternative Position of the Moschidae. Syst Biol 52 (2): 206–228. doi:10.1080/10635150390192726

Hassanin A., Pasquet E., Vigne J.D. (1998). Molecular Systematics of the Subfamily Caprinae (artiodactyla, bovidae) as Determined from Cytochrome B Sequences. J Mamm Evol 5 (3): 217–236. doi:10.1023/A:1020560412929

Hassanin A., Ropiquet A., Couloux A., Cruaud C. (2009). Evolution of the Mitochondrial Genome in Mammals Living at High Altitude: New Insights from a Study of the Tribe Caprini (Bovidae, Antilopinae). J Mol Evol 68 (4): 293–310. doi:10.1007/s00239-009-9208-7

Hebert P.D.N., Ratnasingham S., DeWaard J.R. (2003). Barcoding Animal Life: Cytochrome C Oxidase Subunit 1 Divergences among Closely Related Species. Proc R Soc B Biol Sci 270 (SUPPL. 1): 96–99. doi:10.1098/rsbl.2003.0025

Hill E., Linacre A.M.T., Toop S., Murphy N.P., Strugnell J.M. (2017). The Complete Mitochondrial Genome of Axis Porcinus (Mammalia: Cervidae) from Victoria, Australia, Using MiSeq Sequencing. Mitochondrial DNA Part B Resour 2 (2): 453–454. doi:10.1080/23802359.2017.1357451

Hixson J.E., Wong T.W., Clayton D.A. (1986). Both the Conserved Stem-Loop and Divergent 5'-Flanking Sequences Are Required for Initiation at the Human Mitochondrial Origin of Light-Strand DNA Replication. J Biol Chem 261 (5): 2384–2390

Hoelzel A.R., Lopez J.V., Dover G.A., O'Brien S.J. (1994). Rapid Evolution of a Heteroplasmic Repetitive Sequence in the Mitochondrial DNA Control Region of Carnivores. J Mol Evol 39 (2): 191–199. doi:10.1007/BF00163807

Hu X. Di, Li K., Gao L.Z. (2016). The Complete Mitochondrial Genome of Celebes Wild Boar, *sus celebensis* (Cetartiodactyla: Suina: Suidae), and Comparative Mitochondrial Genomics of the Sus species. Mitochondrial DNA 27 (2): 1476–1477. doi:10.3109/19401736.2014.953099

Hu X. Di, Yang X.T., En-Yang. (2015). The Complete Mitochondrial

Genome of European Wild Boar, *Sus scrofa scrofa*. Mitochondrial DNA 27 (5): 3244–3245. doi:10.3109/19401736.2015.1007366

Hutchinson C.A., Newbold J.E., Potter S.S., Edgell M.H. (1974). Maternal Inheritance of Mammalian Mitochondrial DNA. Nature 251(5475): 536-538. doi: 10.1038/251536a0

Ingman M., Kaessmann H., Pääbo S., Gyllensten U. (2000). Mitochondrial Genome Variation and the Origin of Modem Humans. Nature 408 (6813): 708–713. doi:10.1038/35047064

Irwin D.M., Kocher T.D., Wilson A.C. (1991). Evolution of the Cytb Gene of Mammals. J Mol Evol 32:128-144 128–144. doi: 10.1007/bf02515385

Jazin E., Soodyall H., Jalonen P., Lindholm E., Stoneking M., Gyllensten U. (1998). Mitochondrial Mutation Rate Revisited: Hot Spot and Polymorphism. Nat Gen, 18:109-110. doi: 10.1038/ng0298-109

Ji Y., Shi Y., Ding G., Li Y. (2011). A New Strategy for Better Genome Assembly from Very Short Reads. BMC Bioinformatics 12. doi:10.1186/1471-2105-12-493

Jiang L., Wang G., Tan S., Gong S., Yang M., Peng Q., Peng R., Zou F. (2013). The Complete Mitochondrial Genome Sequence Analysis of Tibetan Argali (Ovis ammon hodgsoni): Implications of Tibetan Argali and Gansu Argali as the Same Subspecies. Gene 521 (1): 24–31. doi:10.1016/j.gene.2013.03.049

Jühling F., Pütz J., Bernt M., Donath A., Middendorf M., Florentz C., Stadler P.F. (2012). Improved Systematic tRNA Gene Annotation Allows New Insights into the Evolution of Mitochondrial tRNA Structures and into the Mechanisms of Mitochondrial Genome Rearrangements. Nucleic Acids Res 40 (7): 2833–2845. doi:10.1093/nar/gkr1131

Jukes TH and Bhurshan V. (1986). Silent Nucleotide Substitutions and G + C Content of Some Mitochondrial and Bacterial Genes. Mol Biol Evol, 24:39-44. doi:10.1007/BF02099949

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. (2012). Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. Bioinformatics 28 (12): 1647–1649. doi:10.1093/bioinformatics/bts199

Kim K.M., Park J.H., Bhattacharya D., Yoon H.S. (2014). Applications of Next-Generation Sequencing to Unravelling the Evolutionary History of Algae. Int J Syst Evol Microbiol 64 (PART 2): 333–345. doi:10.1099/ijs.0.054221-0

Kimura M. (1968). Evolutionary Rate at the Molecular Level. Nature 217: 624 - 626

Korf I. (2004). Gene Finding in Novel Genomes. BMC Bioinformatics 5: 1–9. doi:10.1186/1471-2105-5-59

Kuwayama R., Ozawa T. (2000). Phylogenetic Relationships among European Red Deer, Wapiti, and Sika Deer Inferred from Mitochondrial DNA Sequences. Mol Phylogenet Evol 15 (1): 115–123. doi:10.1006/mpev.1999.0731

Langmead B. (2011). Alignment with Bowtie. Curr Protoc Bioinforma 1–24. doi:10.1002/0471250953.bi1107s32

Li H., Durbin R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. Bioinformatics 25 (14): 1754–1760. doi:10.1093/bioinformatics/btp324

Lowe T.M., Chan P.P. (2016). tRNAscan-SE On-line: Integrating Search and Context for Analysis of Transfer RNA Genes. Nucleic Acids Res 44 (W1): W54–W57. doi:10.1093/nar/gkw413

Lowe T.M., Eddy S.R. (1997). TRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res 25 (5): 955–964. doi:10.1093/nar/25.5.0955

Manee M.M., Alshehri M.A., Binghadir S.A., Aldhafer S.H., Alswailem R.M., Algarni A.T., AL-Shomrani B.M., AL-Fageeh M.B. (2019). Comparative Analysis of Camelid Mitochondrial Genomes. J Genet 98 (3): 88. doi:10.1007/s12041-019-1134-x

Mao H., Liu H., Ma G., Yang Q., Guo X., Lamaocao Z. (2017). The Complete Mitochondrial Genome of *Ovis ammon darwini* (Artiodactyla: Bovidae). Conserv Genet Resour 9 (1): 59–62. doi:10.1007/s12686-016-0620-1

Matosiuk M., Sheremetyeva I.N., Sheremetyev I.S., Saveljev A.P., Borkowska A. (2014). Evolutionary Neutrality of mtDNA Introgression: Evidence from Complete Mitogenome Analysis in Roe Deer. J Evol Biol 27 (11): 2483–2494. doi:10.1111/jeb.12491

Matthee C.A., Davis S.K. (2001). Molecular Insights into the Evolution of the Family Bovidae: A Nuclear DNA Perspective. Mol Biol Evol 18 (7): 1220–1230. doi:10.1093/oxfordjournals.molbev.a003908

McKenna M.C., Bell S.K. (1997). Classification of Mammals above the Species Level, Columbia University Press, New York.

McNab B.K. (2000). Energy Constraints on Carnivore Diet. Nature 407:531-641.doi: 10.1038/35036695

Meadows J.R.S., Hiendleder S., Kijas J.W. (2011). Haplogroup Relationships between Domestic and Wild Sheep Resolved Using a Mitogenome Panel. Heredity (Edinb) 106 (4): 700–706. doi:10.1038/hdy.2010.122

Meiklejohn C.D., Montooth K.L., Rand D.M. (2007). Positive and Negative Selection on the Mitochondrial Genome. Trends Genet 23 (6): 259–263. doi:10.1016/j.tig.2007.03.008

Meng G., Li Y., Yang C., Liu S. (2019). MitoZ: A Toolkit for Animal Mitochondrial Genome Assembly, Annotation and Visualization. Nucleic Acids Res 47 (11): e63. doi:10.1093/nar/gkz173

Mereu P., Palici Di Suni M., Manca L., Masala B. (2008). Complete Nucleotide mtDNA Sequence of Barbary Sheep (*Ammotragus lervia*). DNA Seq - J DNA Seq Mapp 19 (3): 241–245. doi:10.1080/10425170701550599

Miller J.M., Malenfant R.M., Moore S.S., Coltman D.W. (2012). Short Reads, Circular Genome: Skimming Solid Sequence to Construct the Bighorn Sheep Mitochondrial Genome. J Hered 103 (1): 140–146. doi:10.1093/jhered/esr104

Miller K.W.P., Dawson J.L., Hagelberg E. (1996). A Concordance of Nucleotide Substitutions in the First and Second Hypervariable Segments of the Human mtDNA Control Region. Int J Legal Med 109 (3): 107–113. doi:10.1007/BF01369668

Mishmar D., Ruiz-Pesini E., Golik P., Macaulay V., Clark A.G., Hosseini S., Brandon M., Easleyf K., Chen E., Brown M.D., Sukernik R.I., Olckers A., Wallace D.C. (2003). Natural Selection Shaped Regional mtDNA Variation in Humans. Proc Natl Acad Sci U S A 100 (1): 171–176. doi:10.1073/pnas.0136972100

Miyamoto M.M., Tanhauser S.M., Laipis P.J. (1989). Systematic Relationships in the Artiodactyl Tribe Bovini (Family Bovidae), as Determined from Mitochondrial DNA Sequences. Syst Zool 38 (4): 342. doi:10.2307/2992400

Mohandesan E., Fitak R.R., Corander J., Yadamsuren A., Chuluunbat B., Abdelhadi O., Raziq A., Nagy P., Stalder G., Walzer C., Faye B., Burger P.A. (2017). Mitogenome Sequencing in the Genus Camelus Reveals Evidence for Purifying Selection and Long-term Divergence between Wild and Domestic Bactrian Camels. Sci Rep 7 (1): 1–12. doi:10.1038/s41598-017-08995-8

Montgelard C., Catzeflis F.M., Douzery E. (1997). Phylogenetic Relationships of Artiodactyls and Cetaceans as Deduced from the Comparison of Cytochrome B and 12S rRNA Mitochondrial Sequences. Mol Biol Evol 14 (5): 550–559. doi:10.1093/oxfordjournals.molbev.a025792

Naseem A., Batool S., Abbas F. (2020). Utility of Mitochondrial COI Gene for Identification of Wild uUgulate Species of Conservational Importance from Pakistan. Mitochondrial DNA Part B Resour 5 (2): 1924–1928. doi:10.1080/23802359.2020.1756476

Nguyen H.D., Bui T.A., Nguyen P.T., Kim O.T.P., Vo T.T.B. (2017). The Complete Mitochondrial Genome Sequence of the Indigenous i Pig (Sus scrofa) in Vietnam. Asian-Australasian J Anim Sci 30 (7): 930–937. doi:10.5713/ajas.16.0608

Nikaido M., Rooney A.P., Okada N. (1999). Phylogenetic Relationships among Cetartiodactyls Based on Insertions of Short and Long Interpersed Elements: Hippopotamuses Are the Closest Extant Relatives of Whales. Proc Natl Acad Sci U S A 96 (18): 10261–10266. doi:10.1073/pnas.96.18.10261

Parma P., Feligini M., Greppi G., Enne G. (2003). The Complete Nucleotide Sequence of Goat (Capra hircus) Mitochondrial Genome

Goat Mitochondrial Genome. DNA Seq - J DNA Seq Mapp 14 (3): 199–203. doi:10.1080/1042517031000089487

Paszkiewicz K., Studholme D.J. (2010). *De novo* Assembly of Short Sequence Reads. Brief Bioinform 11 (5): 457–472. doi:10.1093/bib/bbq020

Patrushev M.V., Kamenski P.A., Mazunin I.O. (2014). Mutations in Mitochondrial DNA and Approaches for Their Correction. Biochem 79 (11): 1151–1160. doi:10.1134/S0006297914110029

Polziehn R.O., Strobeck C. (2002). A Phylogenetic Comparison of Red Deer and Wapiti Using Mitochondrial DNA. Mol Phylogenet Evol 22 (3): 342–356. doi:10.1006/mpev.2001.1065

Pozzoli U., Menozzi G., Fumagalli M., Cereda M., Comi G.P., Cagliani R., Bresolin N., Sironi M. (2008). Both Selective and Neutral Processes Drive GC Content Evolution in the Human Genome. BMC Evol Biol 8 (1): 1–12. doi:10.1186/1471-2148-8-99

Prada C.F., Boore J.L. (2019). Gene Annotation Errors Are Common in the Mammalian Mitochondrial Genomes Database. BMC Genomics 20 (1): 1–8. doi:10.1186/s12864-019-5447-1

Randi E., Mucci N., Pierpaoli M., Douzery E. (1998). New Phylogenetic Perspectives on the Cervidae (Artiodactyla) Are Provided by the Mitochondrial Cytochrome B Gene. Proc R Soc Lond 265 : 793–801

Reyes A., Gissi C., Pesole G., Saccone C. (1998). Asymmetrical Directional Mutation Pressure in the Mitochondrial Genome of Mammals. Mol Biol Evol 15 (8): 957–966. doi:10.1093/oxfordjournals.molbev.a026011

Robin E.D., Wong R. (1988). Mitochondrial DNA Molecules and Virtual Number of Mitochondria per Cell in Mammalian Cells. J Cell Physiol 136 (3): 507–513. doi:10.1002/jcp.1041360316

Romiguier J., Ranwez V., Douzery E.J.P., Galtier N. (2010). Contrasting GC-content Dynamics across 33 Mammalian Genomes: Relationship with Life-history Traits and Chromosome Sizes. Genome Res 20 (8): 1001–1009. doi:10.1101/gr.104372.109

Saccone C., Gissi C., Lanave C., Larizza A., Pesole G., Reyes A. (2000). Evolution of the Mitochondrial Genetic System: An Overview. Gene 261 (1): 153–159. doi:10.1016/S0378-1119(00)00484-4

Saccone C., Pesole G., Sbisà E. (1991). The Main Regulatory Region of Mammalian Mitochondrial DNA. J Mol Evol 33: 83–91. doi: 10.1007/BF02100199

Sbisà E., Tanzariello F., Reyes A., Pesole G., Saccone C. (1997). Mammalian Mitochondrial D-loop Region Structural Analysis: Identification of New Conserved Sequences and Their Functional and Evolutionary Implications. Gene 205 (1–2): 125–140. doi:10.1016/S0378-1119(97)00404-6

Shadel G.S., Clayton D.A. (1997). Mitochondrial DNA Maintenance in Vertebrates. Annu Rev Biochem 66 (1): 409–435. doi:10.1146/annurev.biochem.66.1.409

Smith D.R. (2016). The Past, Present and Future of Mitochondrial Genomics: Have We Sequenced Enough mtDNAs? Brief Funct Genomics 15 (1): 47–54. doi:10.1093/bfgp/elv027

Song H.J., Lee J.M., Graf L., Rho M., Qiu H., Bhattacharya D., Yoon H.S. (2016). A Novice's Guide to Analyzing NGS-derived Organelle and Metagenome Data. Algae 31 (2): 137–154. doi:10.4490/algae.2016.31.6.5

Suarez R.K., Darveau C.A., Childress J.J. (2004). Metabolic Scaling: A Many-splendoured Thing. Comp Biochem Physiol - B Biochem Mol Biol 139 (3 SPEC.ISS.): 531–541. doi:10.1016/j.cbpc.2004.05.001

Świsłocka M., Matosiuk M., Ratkiewiez M., Borkowska A., Czajkowska M., Mackiewicz P. (2020). Phylogeny and Diversity of Moose (Alces alces, Cervidae, Mammalia) Revealed by Complete Mitochondrial Genomes. Hystrix, Ital J Mammal 0. doi:10.4404/hystrix-00252-2019

Taanman J.W. (1999). The Mitochondrial Genome: Structure, Transcription, Translation and Replication. Biochim Biophys Acta - Bioenerg 1410 (2): 103–123. doi:10.1016/S0005-2728(98)00161-3

Tillich M., Lehwark P., Pellizzer T., Ulbricht-Jones E.S., Fischer A., Bock R., Greiner S. (2017). GeSeq - Versatile and Accurate Annotation of Organelle Genomes. Nucleic Acids Res 45 (W1): W6–W11. doi:10.1093/nar/gkx391

Tobe S.S., Kitchener A.C., Linacre A.M.T. (2010). Reconstructing Mammalian Phylogenies: A Detailed Comparison of the Cytochrome B and Cytochrome Oxidase Subunit in Mitochondrial Genes. PLoS One 5 (11). doi:10.1371/journal.pone.0014156

Upholt W.B., Dawid I.B. (1977). Mapping of Mitochondrial DNA of Individual Sheep and Goats: Rapid Evolution in the D- loop Region. Cell 11 (3): 571–583. doi:10.1016/0092-8674(77)90075-7

Ursing B.M., Arnason U. (1998). Analyses of Mitochondrial Genomes Strongly Support A Hippopotamus-Whale Clade. Proc R Soc B Biol Sci 265 (1412): 2251–2255. doi:10.1098/rspb.1998.0567

Ursing B.M., Slack K.E., Arnason U. (2000). Subordinal Artiodactyl Relationships in the Light of Phylogenetic Analysis of 12 Mitochondrial Protein-Coding Genes. Zool Scr 29 (2): 83–88. doi:10.1046/j.1463-6409.2000.00037.x

Walberg M.W., Clayton D.A. (1981). Sequence and Properties of the Human KB Cell and Mouse L cell D-loop Regions of Mitochondrial DNA. Nucl Acids Res 9(20):5411-5421. doi: 10.1093/nar/9.20.5411

Wallace D.C. (2007). Why Do We Still Have a Maternally Inherited Mitochondrial DNA? Insights from Evolutionary Medicine. Annu Rev Biochem 76 (1): 781–821. doi:10.1146/annurev.biochem.76.081205.150955

Wang Z., Chen Y., Li Y. (2004). A Brief Review of Computational Gene Prediction Methods. Genomics, Proteomics Bioinforma / Beijing Genomics Inst 2 (4): 216–221. doi:10.1016/S1672-0229(04)02028-5

Wolstenholme D.R. (1992). Animal Mitochondria1 DNA: Structure and Evolution. Intern Rev of Cyt 141:173-216. doi.org/10.1016/S0074-7696(08)62066-5

Wu T.D. (2016). Bitpacking Techniques for Indexing Genomes: I. Hash tables. Algorithms Mol Biol 11 (1): 1-13. doi:10.1186/S13015-016-0069-5

Wuitschick J.D., Karrer K.M. (1999). Analysis of Genomic G + C Content, Codon Usage, Initiator Codon Context and *Translation Termination* Sites in Tetrahymena thermophila. J Eukaryot Microbiol 46 (3): 239–247. doi:10.1111/j.1550-7408.1999.tb05120.x

Wyman S.K., Jansen R.K., Boore J.L. (2004). Automatic Annotation of Organellar Genomes with DOGMA. Bioinformatics 20 (17): 3252–3255. doi:10.1093/bioinformatics/bth352

Xiufeng X., Árnason Ú. (1994). The Complete Mitochondrial DNA Sequence of the Horse, *Equus caballus*: Extensive Heteroplasmy of the Control Region. Gene 148 (2): 357–362. doi:10.1016/0378-1119(94)90713-7

Xu S.Q., Yang Y.Z., Zhou J., Jing G.E., Chen Y.T., Wang Jun, Yang H.M., Wang Jian, Yu J., Zheng X.G., Ge R.L. (2005). A Mitochondrial Genome Sequence of the Tibetan Antelope (*Pantholops hodgsonii*). Genomics, Proteomics Bioinforma 3 (1): 5–17. doi:10.1016/S1672-0229(05)03003-2

Yandell M., Ence D. (2012). A Beginner's Guide to Eukaryotic Genome Annotation. Nat Rev Genet 13 (5): 329–342. doi:10.1038/nrg3174

Ye L., Hillier L.D.W., Minx P., Thane N., Locke D.P., Martin J.C., Chen L., Mitreva M., Miller J.R., Haub K. V., Dooling D.J., Mardis E.R., Wilson R.K., Weinstock G.M., Warren W.C. (2011). A Vertebrate Case Study of the Quality of Assemblies Derived from Next-generation Sequences. Genome Biol 12 (3). doi:10.1186/gb-2011-12-3-r31

Zhou M., Yu J., Li B., Ouyang B., Yang J. (2019). The Complete Mitochondrial Genome of *Budorcas taxicolor tibetana* (Artiodactyla: Bovidae) and Comparison with Other Caprinae species: Insight into the Phylogeny of the *genus Budorcas*. Int J Biol Macromol 121: 223–232. doi:10.1016/j.ijbiomac.2018.10.020

Zurano J.P., Magalhães F.M., Asato A.E., Silva G., Bidau C.J., Mesquita D.O., Costa G.C. (2019). Cetartiodactyla: Updating a Time-calibrated Molecular Phylogeny. Mol Phylogenet Evol 133 (December 2018): 256–262. doi:10.1016/j.ympev.2018.12.015

aCS86_1