ISSN: 0028-0712

THE COMPLETE MITOCHONDRIAL GENOME OF THE VIETNAMESE *MACACA MULATTA*

Tram Thi Thuy Nguyen[1,2], Chi Nguyen Quynh Ho[2,3], Cuong Phan Minh Le[3], Nhan Lu Chinh Phan[3], Loan Tung Dang[4], Quan Ke Thai[5], Mai Thi Phuong Nguyen[6], Phuc Bui Tran[7], Long Thanh Le[2,3,*], Vu Quang-Manh[8,*]

[1] Biotechnology Department, Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi 100000, Vietnam, tramttn@upes.edu.vn; orcid: 0009-0007-3733-744X

[2] Department of General and Biomedical Science, Ho Chi Minh City University of Physical Education and Sports, Ho Chi Minh 700000, Vietnam, quynhchihonguyen@gmail.com; orcid: 0000-0002-3267-243X

[3] Animal Biotechnology Department, Institute of Tropical Biology, Vietnam Academy of Science and Technology, Ho Chi Minh 700000, Vietnam, lephanminhcuong@gmail.com; plcnhan@itb.ac.vn; orcid: 0000-0002-7029-1232; longlt@itb.ac.vn; orcid: 0000-0003-4732-3643

[4] University of Science, Vietnam National University, Ho Chi Minh, 700000, Vietnam, dttloan@hcmus.edu.vn, orcid: 0000-0001-7729-1553

[5] Faculty of Natural Sciences Education, Saigon University, Ho Chi Minh 700000, Vietnam, quan.tk@cb.sgu.edu.vn; orcid: 0000-0002-2264-2579

[6]Tay Nguyen Institute for Scientific Research, Academy of Science and Technology, Dalat City 670000, Vietnam, phuongmaipvsh@yahoo.com

[7]Nguyen Tat Thanh University, Ho Chi Minh, 700000, Vietnam, tbphuc@ntt.edu.vn

[8]Hoa Binh University (HBU): Nam Tu Liem, Hanoi, 100000, Vietnam, vqmanh@gmail.com, orcid: 0000-0002-6561-8392.

[*] Authors to whom correspondence should be addressed.

Abstract:

This study investigates the genetic divergence between Vietnamese *Macaca mulatta* and Myanmar *Macaca mulatta*, revealing a total of 704 nucleotide differences. Notably, the protein-coding gene nad1 exhibits 38 significant variations, primarily characterized by the substitution of cytosine (C) in the Vietnamese population with thymine (T) in the Myanmar population. Additional differences are noted across several other genes, including *nad2* (40 differences), *cox1* (59 differences), and *nad4* (76 differences), among others. Importantly, there are no nucleotide loss mutations, indicating that all differences are substitutions. The substantial genetic variation underscores the evolutionary adaptations of these populations to their respective environments. Furthermore, 60 differing positions in the control region highlight potential regulatory changes in mitochondrial gene expression, suggesting that environmental factors have significantly influenced the genetic makeup of these primate populations.

Keywords: *Macaca mulatta*; Vietnam; mitogenome; phylogenetic

Body text:

ISSN: 0028-0712

Introduction:

The genus *Macaca* encompasses a diverse group of primates widely distributed across Asia and North Africa, with significant ecological and evolutionary importance. Among them, the Vietnamese *Macaca mulatta*, commonly known as yellow-haired monkeys or Rhesus Macaque, is notable for its adaptability to various habitats, including urban environments(Roos & Zinner 2015). Understanding the mitochondrial genome of this species can provide insights into its evolutionary history, population genetics, and adaptations to environmental changes. Mitochondrial DNA (mtDNA) is a valuable tool for phylogenetic studies due to its maternal inheritance, relatively rapid mutation rate, and compact structure, which includes 37 genes essential for mitochondrial function (Spuhler 1988). The complete sequencing of mtDNA has become increasingly achievable with advancements in sequencing technologies, allowing for comprehensive analyses of genetic variation across species (Smith 2016). Previous studies on the mitochondrial genomes of various *Macaca* species have revealed significant phylogenetic relationships within the genus, highlighting the evolutionary dynamics that have shaped their diversification (Li *et al.* 2011). However, the complete mitochondrial genome of the Vietnamese *Macaca mulatta* has not been fully characterized, leaving a gap in our understanding of its genetic makeup and evolutionary trajectory. This study aims to present the complete mitochondrial genome of the Vietnamese *Macaca mulatta*, focusing on its structure, gene content, and nucleotide composition. By comparing these findings with other *Macaca* species, we seek to elucidate phylogenetic relationships and explore potential adaptations within this species. Our results will not only contribute to the existing body of knowledge on primate genetics but also inform conservation strategies for the long-tailed macaque in Vietnam.

Methods:

DNA Extraction and sequencing

Tissue samples of Vietnamese *Macaca mulatta* were obtained from the Biological Museum at the Tay Nguyen Institute for Scientific Research, part of the Vietnam Academy of Science and Technology. The mitogenome was extracted utilising the QIAamp DNA Mini kit (Qiagen, USA) in accordance with the manufacturer's guidelines. Assessment of DNA sample integrity: The concentration of DNA was assessed utilising the fluorometric method (Qubit). The OD260/OD280 ratio was assessed through absorbance measurement. The size of the DNA was ascertained through agarose gel electrophoresis. Samples that met the quality criteria of concentration $\geq$ 2 ng/μL, amount $\geq$ 90 ng, and OD260/OD280 $\geq$ 1.70 were deemed appropriate for subsequent experimental procedures. Samples exhibiting DNA sizes less than 1,000 base pairs will be flagged.

Whole genome sequencing libraries were constructed utilising the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, USA) in accordance with the manufacturer's guidelines. The library concentration was quantified using a fluorometric technique, while the average library size was assessed with a Bioanalyzer (Agilent) in accordance with Illumina's library evaluation protocols. Samples were deemed appropriate for sequencing when the concentration was $\geq$ 0.50 ng/μL for genome sizes less than 1 Gb, or $\geq$ 2 ng/μL for genome sizes over 1 Gb. The libraries were subsequently sequenced with the 150PE Next Generation Sequencing technique on a MiniSeq, MiSeq, or NovaSeq equipment (Illumina).

Evaluating interpretation of data quality

ISSN: 0028-0712

The raw sequencing data, following optimisation, de novo assembly, and genome annotation, were recorded in the FASTQ file format, which includes read and sequence information along with the relevant quality scores.

De novo assembly

De novo assembly refers to the process of constructing genome sequences from short or long reads independently of a reference genome sequence. The GetOrganelle (v1.7.7.0) pipeline was employed with optimal parameters for the de novo assembly of the sample genome sequence. The assembly results are detailed in Table S4, indicating a contig length of 16,558 bp, aligning with the mitochondrial genome size of approximately 16 kb.

Bioinformatics analysis

The raw sequencing data underwent purification utilising fastp tool v0.23.1 (Chen *et al.* 2018). Nucleotides characterised by poor solution quality or deemed unreliable or unknown (type N nucleotides) were excluded based on the Phred-score values assigned to each nucleotide (Gurevich *et al.* 2013). Following purification, the reads were assembled de novo utilising the GetOrganelle tool (v1.7.7.0) (Jin *et al.* 2020). The quality of the de novo assembly was evaluated using the Quast v5.2.0 tool (Gurevich *et al.* 2013) and through local alignment of reads to the assembled contigs. This facilitated the identification of areas exhibiting significantly low depth coverage values in comparison to adjacent regions, as recorded in the assembly results.

The mitogenome was annotated utilising Galaxy (The Galaxy Community *et al.* 2024) and Mito Fish (Iwasaki *et al.* 2013)**.** Mito Annotator (Iwasaki *et al.* 2013) was utilised to construct a genetic map of the complete *Macaca mulatta* mitogenome. Amino acid and nucleotide compositions of all 57 primate mitogenomes were assessed and compared using MEGA 11, with AT-skew calculated as (A - T)/ (A + T) and GC-skew as (G - C)/ (G + C). The tRNA sequences were aligned with homologous sequences from related species. The A+T rich regions, protein-coding genes, and ribosomal RNAs were identified through comparison with other Passeriformes mitogenomes. The Relative Synonymous Codon Usage (RSCU) values for the complete mitogenome of *Macaca mulatta* were calculated utilising MEGA 11. The predictions of the secondary structures of transfer RNA (t-RNA) were conducted using tRNAscan-SE software.

Phylogenetic tree

To ascertain molecular based phylogenetic position of *Macaca Mulatta* and its relationship with other families in order Primate, 13 PCGs of 57 spiecies' sequence alignments was used to ananlyze with the Maximum Likelihood method (ML) and Tamura-Nei model conducted in MEGA11 (Tamura *et al.* 2021)**.**

Results:

Complete Mitochondrial Genome Analysis

ISSN: 0028-0712

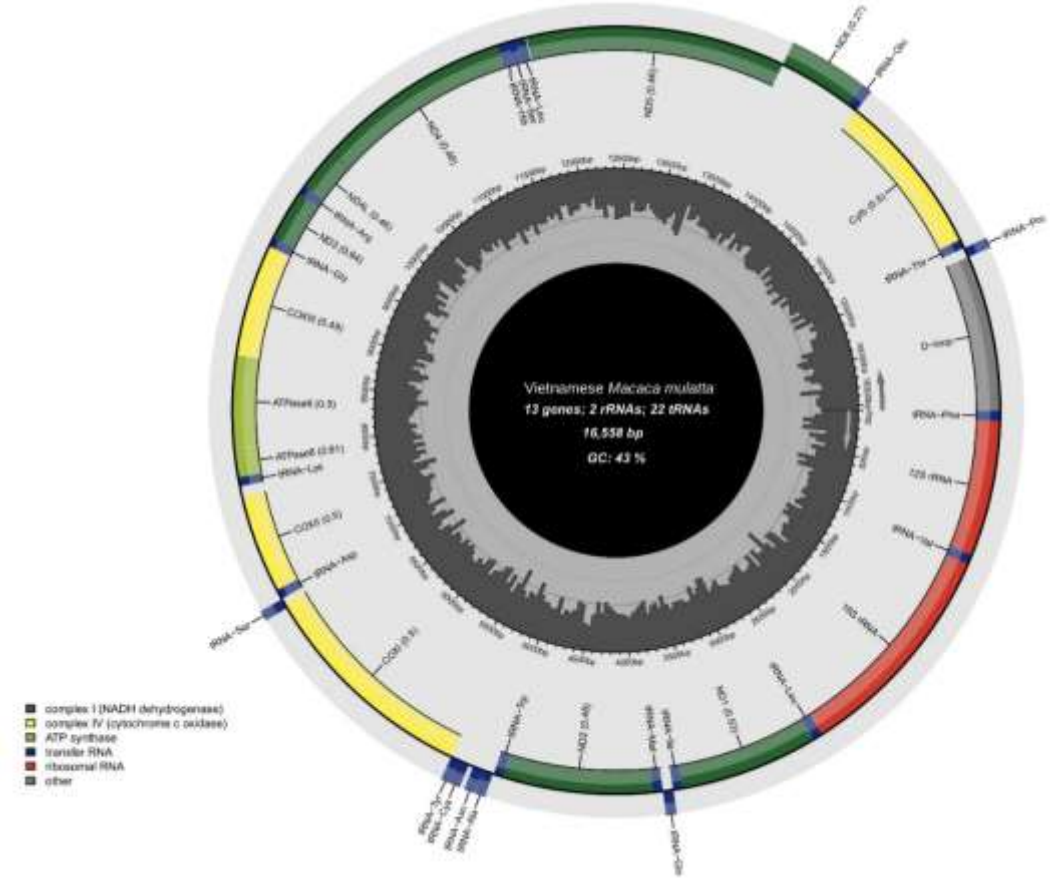

Figure 1. The complete mitochondrial genome structure of *Macaca mulatta* is illustrated as a circular diagram, with genes represented by colored blocks. The control region appears in grey, while the *16S* (large rRNA) and *12S* (small rRNA) genes are indicated in red. The 22 transfer RNA (*tRNA*) genes are marked in dark blue, and the 13 protein-coding genes (PCGs) are shown in green and yellow.

The arrangement of genes within the mitochondrial genome is illustrated in **Figure 1**, revealing a circular structure consisting of 16,566 base pairs. This genome contains a total of 37 genes, which include 13 protein-coding genes (PCGs), 22 transfer RNA (*tRNA*) genes, and 2 ribosomal RNA (rRNA) genes. Genes on the H strand are displayed on the inner circle, and those on the L strand are found on the outer circle.

Table 1. The complete mitochondrial genome of *Macaca mulatta* contains 37 genes, detailing their positions and lengths.

Name	Start	Stop	Strand	Length	Intergenic Nucleotides
trnF(ttc)	248	319	H	72	0
rrnS	320	1266	H	947	0
trnV(gta)	1267	1335	H	69	0
rrnL	1334	2894	H	1561	0
trnL2(tta)	2895	2969	H	75	2
nad1	2972	3922	H	951	4
trnI(atc)	3927	3995	H	69	0
trnQ(caa)	3993	4064	L	72	1
trnM(atg)	4066	4133	H	68	0
nad2	4134	5165	H	1032	10
trnW(tga)	5176	5242	H	67	7
tRNA(gca)	5250	5318	L	69	1
trnN(aac)	5320	5392	L	73	32
trnC(tgc)	5425	5493	L	69	0
trnY(tac)	5494	5556	L	63	4
cox1	5561	7099	H	1539	2
trnS2(tca)	7102	7170	L	69	3
trnD(gac)	7174	7241	H	68	1
cox2	7243	7902	H	660	97
trnK(aaa)	8000	8063	H	64	1
atp8	8065	8259	H	195	0
atp6	8226	8900	H	675	5
cox3	8906	9688	H	783	1
trnG(gga)	9690	9757	H	68	0
nad3	9758	10102	H	345	1
trnR(cga)	10104	10168	H	65	0
nad4l	10169	10462	H	294	0
nad4	10459	11829	H	1371	7
trnH(cac)	11837	11905	H	69	0
trnS1(agc)	11906	11964	H	59	0
trnL1(cta)	11965	12035	H	71	6
nad5	12042	13841	H	1800	12
nad6	13854	14375	L	522	0
trnE(gaa)	14376	14444	L	69	4
Cob	14449	15585	H	1137	4
trnT(aca)	15590	15653	H	64	1
trnP(cca)	15655	15722	L	68	206

Table 1 presents the gene structure of the complete mitochondrial genome of *Macaca mulatta*, consisting of 37 genes organized into three categories: 13 protein-coding genes (PCGs), 22 transfer RNA (*tRNA*) genes, and 2 ribosomal RNA (rRNA) genes. This arrangement is typical for mitochondrial genomes and reflects a compact organization that enhances transcription and replication efficiency. The genes are predominantly located on the heavy (H) strand, with varying lengths; for instance, *nad5* is the longest at 1800 base pairs, while *trnK* is the shortest at 64 base pairs. This variability in gene length suggests different functional roles and evolutionary pressures, particularly in energy metabolism and protein synthesis, which are crucial for the species' adaptability to diverse environments (Nabholz *et al.* 2012).

The analysis of intergenic regions in the complete mitochondrial genome of *Macaca mulatta* reveals a total of 366 intergenic nucleotides across 36 intergenic spaces. Notably, the longest intergenic region is between genes *trnF* and *trnP* measuring 206 nucleotides, while the shortest is between *trnL2* and *nad1*, at just 2 nucleotides. These findings suggest variability in the regulatory potential of intergenic regions, with longer spaces possibly indicating areas for regulatory elements, while shorter spaces reflect a more compact genomic organization. The presence of intergenic nucleotides is crucial for understanding gene regulation and mitochondrial function, providing insights into the evolutionary adaptations of this species (Zardoya 2020).

Nucleotide Composition

Table 2. Nucleotide composition metrics in different areas of the mitogenomes of twenty-three representative primate species.

	Accession Number	Whole		Protein-Coding Genes (PCGs)		Large Ribosomal RNA (*rrnL*)		Small Ribosomal RNA (*rrnS*)	
		Length	AT (%)	Length (bp)	AT (%)	Length (bp)	AT (%)	Length (bp)	AT (%)
Vietnamese *Macaca mulatta*		16566	56.9	11304	57	1561	57.4	947	55.2
Myanamar Macaca mulatta	JQ821843	16,564	56.8	11,310	56	1560	57.4	947	55.6
Macaca leonina	KP330231	17,050	57.2	11,310	57	1562	57.8	947	54.6
Macaca arctoides	KM360179	16,559	56.7	11,313	56	1566	57.3	951	55.8
Pygathrix cinerea	JQ821842	16,535	61.1	11,292	61	1562	59.9	948	59.4
Pygathrix nemaeus	JF293096	15,467	61.3	11,292	61	1564	59.7	948	59.4
Pygathrix nigripes	MH064177	16,534	61.4	11,304	62	1563	59.7	949	58.8
Presbytis femoralis	KU899140	16,548	61.9	11,277	62	1563	61	935	58.3
Nasalis larvatus	DQ355298	16,570	60.9	11,298	61	1568	59.9	949	57.5
Rhinopithecus roxellana	KM504390	16,552	61.5	11,292	62	722	59.6	949	58.2
Rhinopithecus brelichi	JQ821836	16,553	61.6	11,295	62	1571	60	949	58.3
Rhinopithecus bieti	JQ821839	16,550	61.5	11,298	62	1570	59.7	949	58
Rhinopithecus avunculus	JF293093	16,552	61.6	11,289	62	1572	60.2	949	58.5
Papio hamadryas	NC001992	16,521	56.3	11,292	56	1572	57.7	947	55.1
Callithrix jacchus	KM588314	16,499	59.7	11,280	59	1555	60.5	953	56.6
Homo sapiens neanderthalensis	NC011137	16,565	55.6	11,283	55	1560	57.2	954	54.4
Homo sapiens neanderthalens	OM062614	16,565	55.7	11,283	55	1560	57.3	954	54.4
Cebus albifrons	NC002763	16,554	60.9	11,268	61	1556	60.2	958	57.6
Moschiola indica	NC037993	16,444	61.4	11,310	61	1576	61.6	958	56.4

Table 2 presents nucleotide composition metrics for the mitochondrial genomes of various primate species, highlighting similarities and differences with the Vietnamese *Macaca mulatta.* The nucleotide composition of the mitochondrial genome of the Vietnamese *Macaca mulatta* reveals an AT content of 56.9%. The complete mitochondrial genome spans 16,566 base pairs, with the protein-coding region comprising 11,304 base pairs. Within this genome, the AT

content among the protein-coding genes (PCGs) is 57%, while the large ribosomal RNA (*rrnL*) and small ribosomal RNA (*rrnS*) display AT contents of 57.4% and 55.2%, respectively.

Notably, the Vietnamese *Macaca mulatta* and Myanmar *Macaca mulatta* exhibit very similar mitochondrial genome characteristics, with only minor differences. The Vietnamese population has a genome length of 16,566 bp and an AT content of 56.9%, while the Myanmar population is slightly shorter at 16,564 bp, with an AT content of 56.8 and other *Macaca* species *Macaca arctoides* is also slightly shorter at 56.7%. The AT content is critical for understanding evolutionary dynamics as it can impact the stability and function of the mitochondrial DNA.

The analysis of nucleotide composition reveals that protein-coding genes (PCGs) in *Macaca mulatta* have a slightly elevated AT content of 57%. In contrast, the large ribosomal RNA (*rrnL*) and small ribosomal RNA (*rrnS*) genes show AT contents of 57.4% and 55.2%, respectively. This variation in AT content across different genomic regions emphasizes the functional significance of nucleotide composition in relation to gene expression and metabolic processes. Comparative studies with other primate species, particularly *Pygathrix* and *Presbytis*, indicate a notable increase in AT content, highlighting the genetic diversity among primates and offering insights into the evolutionary adaptations that shape mitochondrial genomes across different lineages (Bergey *et al.* 2020)

Both the Vietnamese *Macaca mulatta* and Myanmar *Macaca mulatta* share comparable lengths for their protein-coding sequences, with the Vietnamese variant measuring 11,304 bp and the Myanmar variant at 11,310 bp. The lengths of the ribosomal RNA genes are identical in both populations, with *rrnL* at 1,561 bp and *rrnS* at 947 bp. These similarities suggest a close genetic relationship between the two populations, likely indicative of recent common ancestry and similar evolutionary pressures.

Table 3. GC-skew and AT-skew in PCGs.

	Accession Number	T(U)	C	A	G	Total	AT-Skew	GC-Skew
Vietnamese *Macaca mulatta*	PP623106	30.9	26.1	30.5	12.5	868.6	−0.00605	−0.35278
Myanmar *Macaca mulatta*	JQ821843	26.6	31	29.8	12.6	870	0.056444	−0.42052
Macaca arctoides	KM360179	26.4	31	30	12.5	870.2	0.063577	−0.4244
Papio hamadryas	NC001992	26.2	31.2	29.7	12.9	868.6	0.061768	−0.41663
Callithrix jacchus	KM588314	28.3	27.8	31.1	12.9	867.7	0.046296	−0.36736
Homo sapiens neanderthalensis	OM062614	26.2	31.9	28.8	13.1	867.8	0.046422	−0.41631
Cebus albifrons	NC002763	29.9	26.7	31.3	12.1	866.8	0.02306	−0.37663

Table 3 presents the GC-skew and AT-skew values of protein-coding genes (PCGs) across different primate species, including the Vietnamese Macaca mulatta. The GC skew is calculated as (G - C) / (G + C), while the AT skew is calculated as (A - T) / (A + T). These

metrics provide insights into the nucleotide composition and potential evolutionary dynamics of the mitochondrial genomes.

For *Macaca mulatta*, the GC skew is reported at -0.35278, indicating a predominance of A and T over G and C in the mitochondrial genome. This negative GC skew suggests a bias that may influence the replication and repair processes of mitochondrial DNA, as well as its overall stability (Sahyoun *et al.* 2014). In contrast, the AT skew of -0.00605 is close to zero, suggesting a balanced usage of A and T nucleotides, which may reflect a functional adaptation in the context of energy metabolism and mitochondrial efficiency.

Comparative data from other *Macaca* species, such as the Myanmar *Macaca mulatta* and *Macaca arctoides*, show similar trends in nucleotide skew values, reinforcing the idea of conserved evolutionary patterns within the genus. These skew analyses are crucial for understanding the evolutionary pressures acting on mitochondrial genomes, as they can reveal insights into gene function, replication dynamics, and potential adaptive strategies across different environments.

Nucleotide Composition Pattern

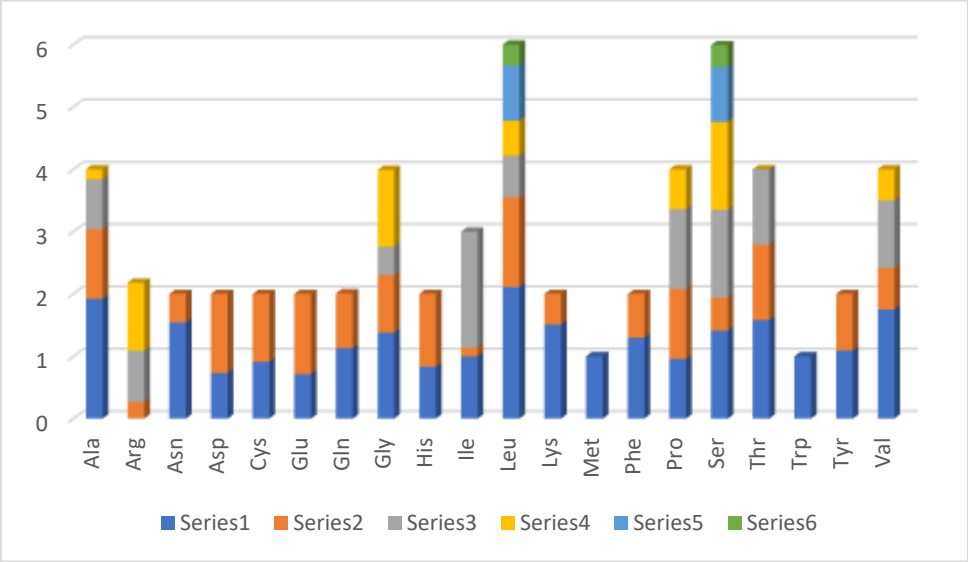

Figure 2. Relative Synonymous Codon Usage (RSCU) of the mitochondrial protein-coding genes in Vietnamese *Macaca mulatta*.

The analysis of Figure 2, which illustrates the Relative Synonymous Codon Usage (RSCU) of mitochondrial protein-coding genes in Vietnamese *Macaca mulatta*, provides valuable insights into the species' genetic adaptations and evolutionary dynamics. The RSCU values represent the frequency of usage of different codons for encoding specific amino acids, with values greater than 1 indicating a strong preference for those codons. In this context, amino acids *leucine (Leu)* and *serine (Ser)* often show elevated RSCU values, suggesting that the corresponding codons are favored in the mitochondrial genes of this species. This preference for certain codons can be attributed to various factors, including the availability of corresponding *tRNAs* and the overall efficiency of the translational machinery, which are critical for the effective synthesis of mitochondrial proteins.

The observed codon bias is significant as it can influence the efficiency of gene expression and protein synthesis, thereby impacting mitochondrial function and energy metabolism. For

instance, a high frequency of specific codons may facilitate faster translation rates, which is particularly advantageous for organisms like *Macaca mulatta* that inhabit diverse environments, including urban areas where metabolic demands can vary dramatically. Understanding these patterns of codon usage is essential, as they not only reflect the evolutionary pressures that have shaped the mitochondrial genome but also provide insights into how these primates have adapted to their ecological niches (Roller *et al.* 2013).

Comparative analyses of RSCU patterns across different *Macaca* species can further elucidate evolutionary relationships and functional adaptations. By examining how codon usage varies among related species, researchers can identify conserved and divergent evolutionary trends that highlight the impact of specific environmental pressures on mitochondrial gene structure and function (Parvathy *et al.* 2022). Such comparisons may reveal that certain codons are universally preferred across species, suggesting strong selective pressures, while others may exhibit variability, indicating adaptations to specific ecological contexts.

Moreover, the implications of these findings extend to conservation strategies. The identification of codon usage patterns in the mitochondrial genome underscores the importance of maintaining genetic diversity within *Macaca mulatta* populations. Genetic diversity is crucial for the adaptability and resilience of species, particularly in the face of environmental changes and habitat loss (Da Silva *et al.* 2012; Wang *et al.* 2019). By understanding the genetic makeup and evolutionary adaptations of *Macaca mulatta*, conservation efforts can be better tailored to preserve not only the species but also its genetic variability, which is essential for its long-term survival.

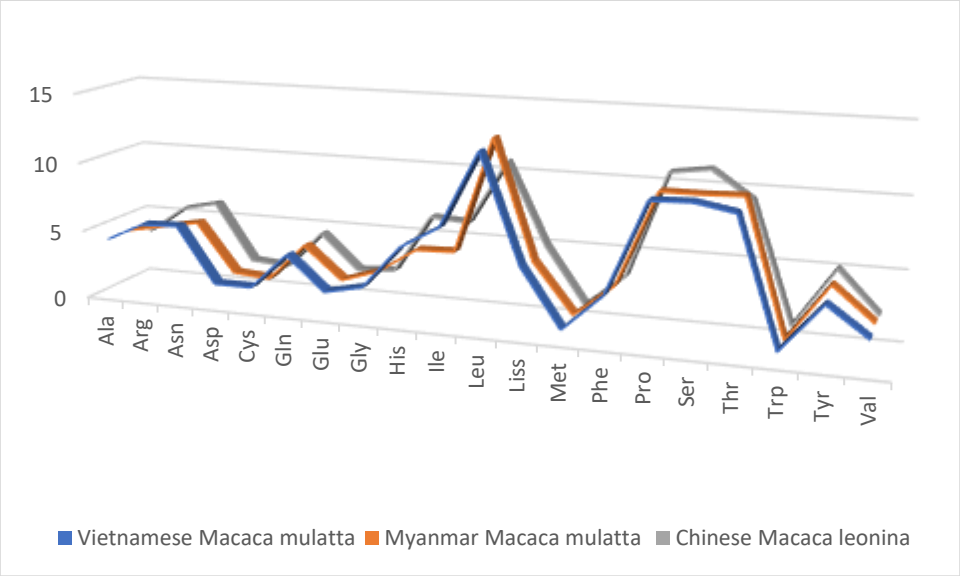

Figure 3. Percentage frequency of amino acid composition in the whole mitogenome of *Macaca* species: Vietnamese *Macaca mulatta;* Myanmar *Macaca mulatta;* Chinese *Macaca leonia*

Figure 3 presents the percentage frequency of amino acid composition in the mitochondrial genome of Vietnamese *Macaca mulatta*. This analysis provides insights into the distribution of amino acids within the mitochondrial proteins, which are crucial for various metabolic functions and energy production. Amino acids such as methionine and tryptophan are represented in low frequencies, which may influence the overall functional dynamics of the mitochondrial proteins. Conversely, the results indicate that certain amino acids dominate the

composition of the mitochondrial proteins encoded in the genome. Notably, leucine and serine emerge as the most abundant amino acids, reflecting their essential roles in mitochondrial function and protein synthesis. These findings align with previous research highlighting the prevalence of these amino acids in mitochondrial proteins across different species (Smith & McDonough 2005; Nguyen *et al.* 2024)

The amino acid composition analysis not only provides a comprehensive overview of the protein-coding capabilities of the mitochondrial genome in *Macaca mulatta* but also allows for comparisons with other primate species. Such comparisons can shed light on evolutionary adaptations and the functional significance of mitochondrial proteins across different lineages (Spuhler 1988). This detailed examination enhances our understanding of the evolutionary dynamics and biochemical roles of mitochondrial proteins in Vietnamese *Macaca mulatta*.

Protein-coding gene

Figure 4 provides an analysis of the AT- and GC-skew values for the 13 protein-coding genes (PCGs) in the mitochondrial genome of Vietnamese *Macaca mulatta*. Most PCGs, such as *nad1*, *nad2*, *cox1*, and *atp6*, exhibit a positive AT-skew, indicating a relative abundance of adenine (A) compared to thymine (T). This trend suggests an optimization for transcription efficiency, which is crucial for the high metabolic demands of mitochondrial function. In contrast, the consistent negative GC-skew across these genes indicates a predominance of cytosine (C) over guanine (G), which may enhance the stability of the mitochondrial DNA and influence gene expression dynamics.

The skew values observed in Figure 4 reflect the evolutionary adaptations of these genes to their roles in energy metabolism and respiratory efficiency. For instance, genes like *nad5* and *cox2* also show positive AT-skew, reinforcing their functional significance in the electron transport chain and ATP synthesis. The presence of negative GC-skew across the PCGs suggests a conserved genomic structure that supports efficient mitochondrial function. Overall, these patterns not only highlight the intricate evolutionary dynamics within *Macaca mulatta* but also underscore the importance of these genes in maintaining metabolic adaptability in diverse environments**.**

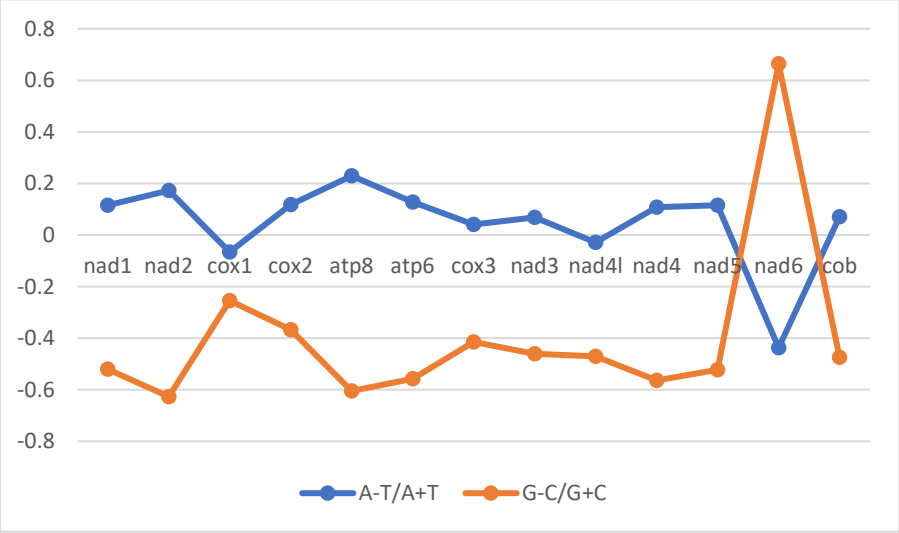

Figure 4. Graphical illustration of AT- and GC-skew in all 13 protein-coding genes of 3.5. tRNA and rRNA Genes

ISSN: 0028-0712

Table 5 outlines the mismatched base pairs found in the *tRNA* genes of the Vietnamese *Macaca mulatta*. *tRNA*: The specific *tRNA* gene and its corresponding amino acid; mismatched Base Pairs: the type of mismatched base pairs identified in the *tRNA* structure, indicating potential variations from standard pairing; Stem: specifies whether the mismatch occurs in the amino acid acceptor (AA) or T-arm region of the *tRNA* and frequency: the number of occurrences of each type of mismatch within the *tRNA* genes.

The table reveals several mismatched base pairs that are crucial for understanding the structural integrity and functional efficiency of *tRNA* molecules. For example, mismatches such as A-C in the *tRNA* for phenylalanine and tryptophan, as well as A-G in methionine, indicate deviations from the canonical Watson-Crick base pairing. These mismatches may have implications for the stability and folding of the *tRNA*, potentially affecting its ability to accurately deliver amino acids during protein synthesis.

The presence of mismatched base pairs in the *tRNA*'s stem regions, especially in the AA and T-arm regions, suggests evolutionary adaptations that might enhance the flexibility or functional capabilities of these *tRNA*s in the mitochondrial context. Understanding these variations provides insights into the evolutionary pressures that have shaped the mitochondrial genome of *Macaca mulatta*, highlighting how these primates have adapted their *tRNA* structures to optimize mitochondrial function and protein synthesis efficiency.

Table 5: The mismatched *tRNA* base pairs from *Macaca mulatta* include AA—amino acid acceptor, T-arm, and AC—anticodon.

tRNA	Mismatched Base Pairs	Stem	Frequency
Phenylalanine GAA	A-C	AC	1
Tryptophan TCA	A-C	AC	1
Methionine CAT	A-G	AA	1
	U-U	T-arm	2
Arginine TCG	A-A	AA	1
Histidine	A-C	T-arm	1
Threonine TGT	C-U	AA	1
Threonine TGT	C-U	AA	1

Figure 5 presents a graphical representation of the secondary structure of the 22 transfer RNA (*tRNA*) genes found in the mitochondrial genome of Vietnamese *Macaca mulatta*. The *tRNA* molecules are depicted with their respective stems and loops, showcasing the intricate folding that is essential for their function. In the figure, connections are color-coded: GC pairs are indicated in red, while AU pairs are shown in blue. This color coding highlights the base-pairing interactions critical for the stability and functionality of each *tRNA*.

The secondary structure of *tRNA* genes is vital for their role in translation, as these molecules are responsible for bringing amino acids to the ribosome during protein synthesis. The presence of GC and AU base pairs influences the stability of the *tRNA* structure, with GC pairs generally providing stronger bonding due to three hydrogen bonds compared to the two hydrogen bonds

ISSN: 0028-0712

formed by AU pairs. The observed patterns in Figure 5 suggest a well-conserved structure among the *tRNA* genes, which is crucial for maintaining the efficiency of mitochondrial protein synthesis in *Macaca mulatta*.

Moreover, the specific mismatched base pairs noted in the *tRNA* genes can offer insights into evolutionary adaptations and functional efficiencies. Mismatches, such as those found in *tRNA*s for phenylalanine and tryptophan, may reflect unique evolutionary pressures that have shaped the mitochondrial genome of this species. Understanding these structural nuances can enhance our knowledge of mitochondrial function and the evolutionary dynamics within the genus Macaca, particularly in terms of adaptation to diverse ecological niches. Overall, Figure 5 not only illustrates the complexity of *tRNA* structures but also emphasizes their importance in the context of mitochondrial gene expression and metabolic processes.

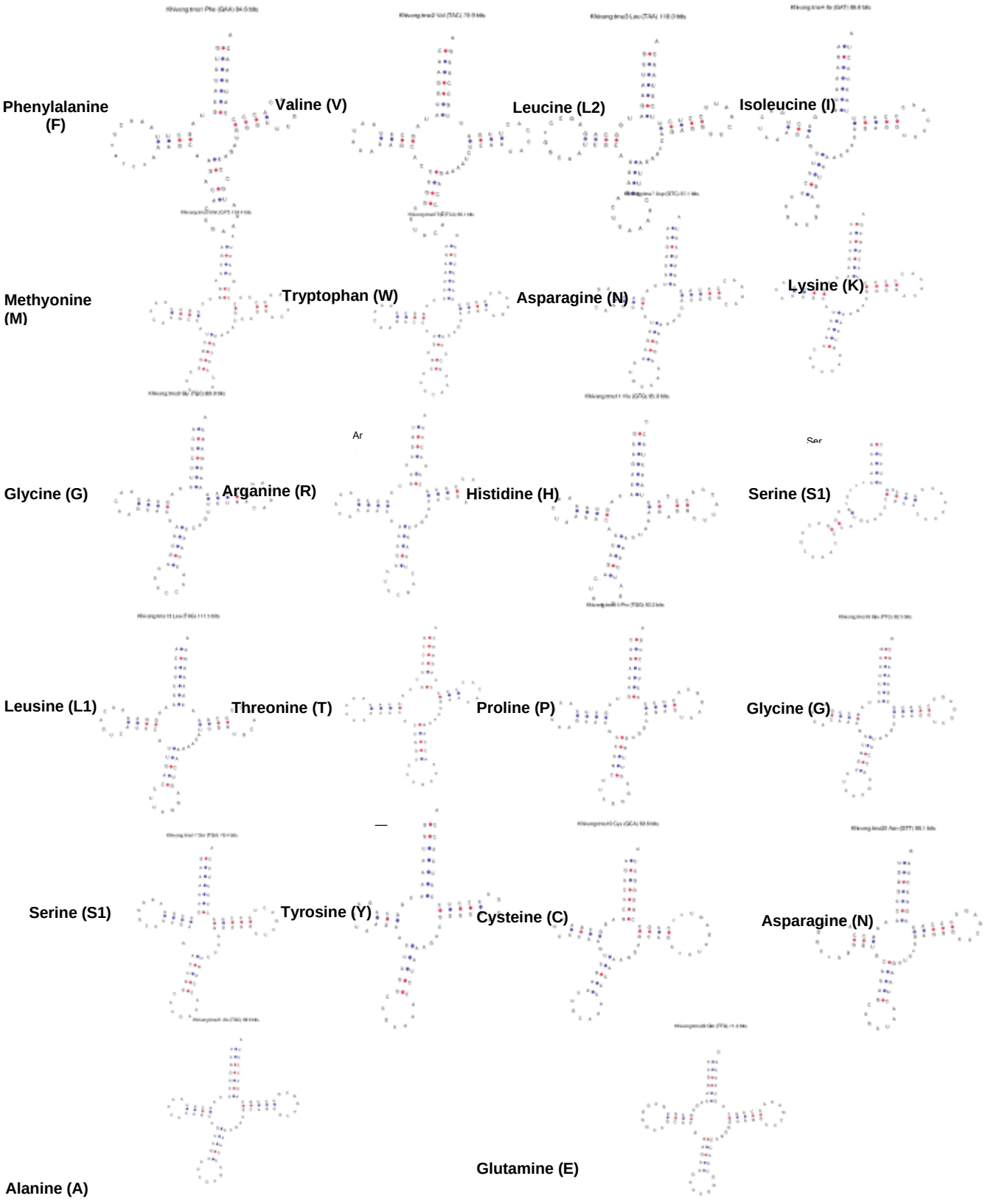

Figure 5. Secondary structure of 22 *tRNA* genes of Vietnamese *Macaca mulatta*. Red color is GC connection, blue color is AU connection.

Differences in nucleotide position between Vietnamese *Macaca mulatta* and Myanmar *Macaca mulatta*

ISSN: 0028-0712

There are 704 different points in Vietnamese Macaca mulatta and Myanmar Macaca mulatta whole genome sequence. Notably, the protein-coding gene *nad1* shows 38 differing positions, primarily where cytosine (C) in the Vietnamese population is replaced by thymine (T) in the Myanmar population. Additional differences are observed in the following genes: *nad2*: 40 differing positions; *cox1*: 59 differing positions; *cox2*: 26 differing positions; *atp8*: 16 differing positions; *atp6*: 37 differing positions; *cox3*: 37 differing positions; *nad3*: 13 differing positions; *nad4l*: 22 differing positions; *nad4*: 76 differing positions; *nad5*: 67 differing positions; *nad6*: 21 differing positions; *cob*: 63 differing positions. Importantly, there are no nucleotide loss mutation; all differences represent nucleotide substitutions between these two populations. The substantial number of differences indicates significant genetic variation, suggesting that these two populations have undergone considerable evolutionary changes to adapt to their respective environments. Of particular interest is the 60 differing positions found in the control region (Table 6). This area is crucial for the regulation of mitochondrial gene expression and may reflect adaptive responses to environmental pressures.

The genetic divergence observed between Vietnamese and Myanmar *Macaca mulatta* populations highlights the evolutionary impact of environmental factors on genetic makeup. The presence of 704 nucleotide differences, particularly in protein-coding genes, suggests that these populations have adapted to their specific ecological niches.

For instance, variations in the *nad1* gene, which is involved in the electron transport chain, may influence metabolic processes essential for survival in different habitats (Zhang *et al.* 2017). Similarly, the differences in the *cob* and *atp* genes, which play roles in energy production, could signify adaptations to varying food resources and energy demands (Wolters 2018).

The significant differences in the control region are particularly noteworthy. This region is known to regulate the transcription of mitochondrial genes and is sensitive to evolutionary pressures (Zhang & Hewitt 1997). The 60 differing positions in this area may indicate a response to local environmental conditions, such as climate or habitat type, which can influence mitochondrial function and efficiency.

Table 6. Different positions in control region of Vietnamese *Macaca mulatta* and Myanmar *Macaca mulatta*

	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	9	9	9	9
	1	1	2	6	8	8	1	3	4	4	6	6	7	7	8	6	7	8	8	9	9	9	2	2	3	4	4	5	5	5
	2	5	4	0	3	4	3	8	2	7	2	6	3	5	6	8	7	4	9	1	3	7	7	8	7	0	6	2	3	5
Vietnamese *Macaca mulatta*	C	A	T	A	T	G	C	G	T	A	C	A	C	T	T	C	T	A	C	T	G	A	G	T	C	T	T	C	C	T
Myanmar *Macaca mulatta*	T	G	C	G	C	A	T	A	C	G	T	G	T	C	C	T	C	G	T	C	A	G	A	C	T	C	A	T	T	C

	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	1	1	3	5	5	5	5	5	5	5	5	5	5	5
	5	5	6	9	9	0	0	0	2	4	5	5	8	8	9	0	0	9	1	1	2	2	3	3	3	3	5	5	6
	6	9	0	3	9	1	3	7	8	6	3	4	3	8	9	6	7	3	6	7	3	7	1	4	6	8	2	5	6
Vietnamese *Macaca mulatta*	T	A	A	G	A	T	A	T	G	T	T	C	C	C	A	G	T	T	T	T	C	T	A	A	T	T	T	C	G
Myanmar *Macaca mulatta*	C	G	G	A	G	C	T	C	A	C	C	T	T	T	G	A	C	C	C	C	T	C	G	G	C	C	C	T	A

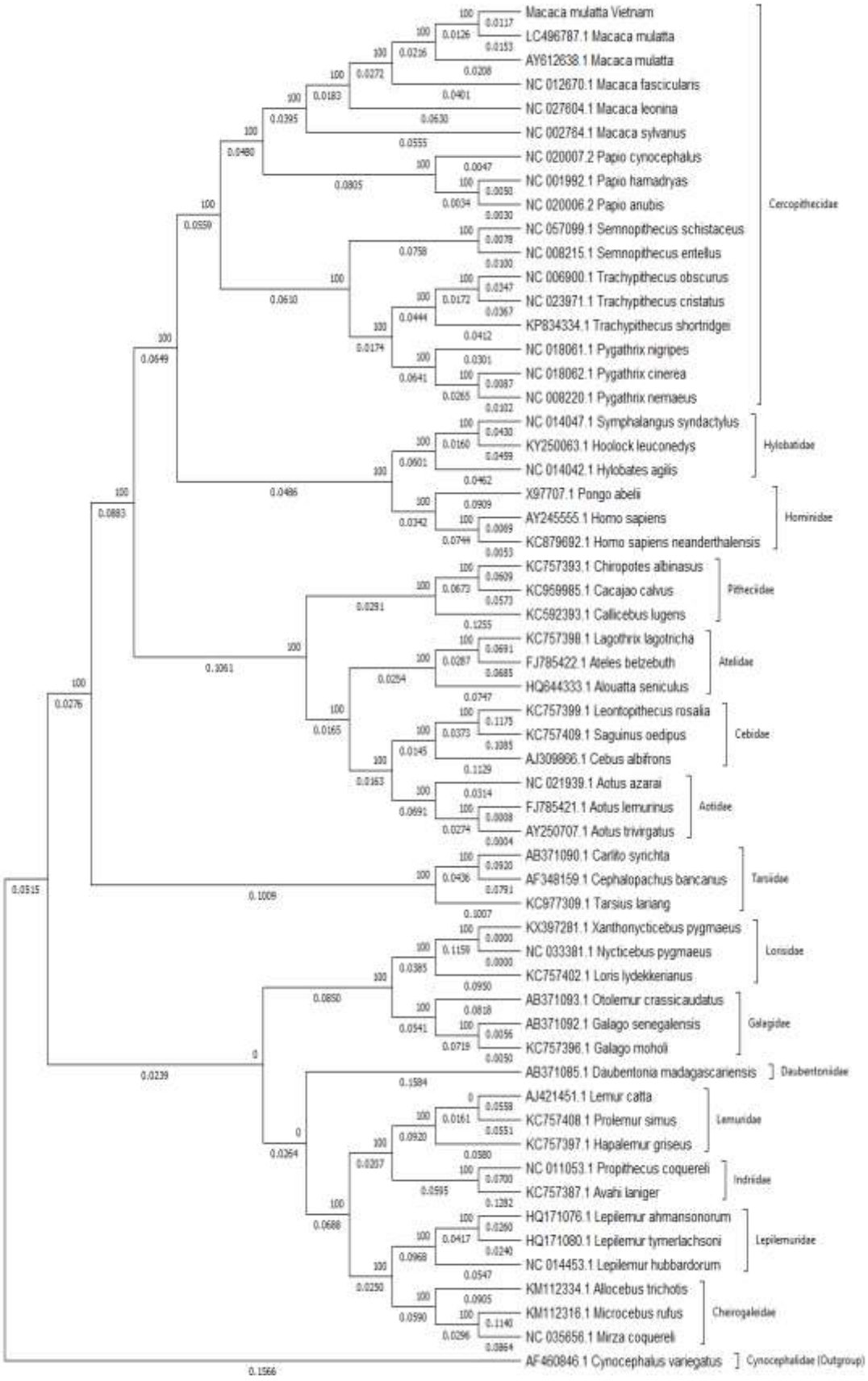

Figure 6. Maximum likelihood phylogeny of 57 whole primate mitogenomes

The phylogenetic tree shows that the Vietnamese *Macaca mulatta* is closely related to other macaque species, such as *Macaca fascicularis* and *Macaca mulatta* from different regions. This close clustering suggests that these species share a recent common ancestor, indicating that they have diverged relatively recently in evolutionary terms. The genetic similarities reflect their adaptations to similar environments and life histories, emphasizing the evolutionary connections within the *Macaca* genus.

In contrast, when comparing Vietnamese *M. mulatta* to more distantly related primates, such as members of the Hominidae family (e.g., humans and chimpanzees), significant genetic divergence is evident. The longer branch lengths leading to these species indicate a longer evolutionary history and greater genetic differentiation. This divergence may be attributed to differences in habitat, behavior, and ecological adaptations that have developed over millions of years. The relationship highlights the evolutionary pathways that have led to the distinct characteristics observed in the *Macaca* genus versus other primate families.

Discussion:

The complete mitochondrial genome of the Vietnamese *Macaca mulatta* provides significant insights into the genetic makeup, evolutionary history, and ecological adaptations of this species. The analysis reveals a circular mitochondrial genome of 16,566 base pairs, containing 37 genes, including 13 protein-coding genes (PCGs), 2 ribosomal RNA genes (*rrnL* and *rrnS*), and 22 transfer RNA genes. This gene arrangement is typical for primate mitochondrial genomes and underscores the compact organization that enhances transcription and replication efficiency (Spuhler 1988).

The study highlights the phylogenetic relationships within the *Macaca* genus. Previous research has demonstrated significant genetic divergence among *Macaca* species, indicating complex evolutionary dynamics shaped by environmental pressures and habitat adaptations. The close genetic relationship between the Vietnamese and Myanmar populations, with only 704 nucleotide differences, suggests recent common ancestry and similar evolutionary pressures. The differences observed in protein-coding genes, particularly in the nad1 gene, which is pivotal for the electron transport chain, may reflect adaptations to specific ecological niches (Zhang *et al.* 2017). Such variations are critical for understanding how these primates have evolved to thrive in diverse environments, including urban areas.

Nucleotide Composition and Codon Usage

The nucleotide composition analysis reveals an AT content of 56.9% in the mitochondrial genome, with the protein-coding regions exhibiting a slightly higher AT content (57%). This nucleotide bias can influence the stability and function of mitochondrial DNA, impacting replication and transcription processes (Sahyoun *et al.* 2014). The Relative Synonymous Codon Usage (RSCU) analysis indicates a preference for certain codons, notably for leucine and serine, which may enhance translational efficiency. This preference is crucial for optimizing protein synthesis in response to varying metabolic demands, particularly in fluctuating urban environments (Roller *et al.* 2013).

Intergenic Regions and Regulatory Potential

The analysis of intergenic regions suggests variability in regulatory potential, with longer intergenic regions potentially indicating areas for regulatory elements. The observation of 366 intergenic nucleotides emphasizes the importance of these regions in gene regulation and mitochondrial function. This regulatory capacity is particularly relevant for understanding how

*Macaca mulatta*adapts to environmental challenges, as regulatory elements play a crucial role in mitochondrial gene expression and metabolic responses (Zardoya 2020).

Mismatched Base Pairs in tRNA Genes

The identification of mismatched base pairs in tRNA genes reveals structural adaptations that may enhance the flexibility and functional capabilities of tRNAs in mitochondrial contexts. Mismatches, such as A-C and A-G, indicate deviations from standard base pairing, which could affect the stability of tRNA molecules and their ability to accurately deliver amino acids during protein synthesis. These variations suggest evolutionary pressures that have shaped the tRNA structures to optimize mitochondrial function (Nguyen *et al.* 2024).

Conclusions:

This study successfully characterizes the complete mitochondrial genome of the Vietnamese *Macaca mulatta*, filling a significant gap in our understanding of its genetic structure and evolutionary adaptations. The insights gained from the nucleotide composition, phylogenetic relationships, and regulatory mechanisms contribute to the broader knowledge of primate genetics and evolution. Furthermore, this research underscores the importance of genetic diversity in conservation strategies, as maintaining genetic variability is crucial for the adaptability and resilience of *Macaca mulatta* populations in the face of environmental changes. Future research should aim to investigate the functional implications of the observed genetic variations and explore the potential for conservation efforts to bolster genetic diversity within this species.

Funding:

This work was supported by Vietnam Academy of Science and Technology, grant ĐLTE00.08/22-23.

References:

BERGEY C.M., TRUJILLO A.E., DISOTELL T.R., TOSI A.J. & BURRELL A.S. 2020. — Genomics of adaptation, convergence, and evolutionary conservation in haplorrhine primates. *American journal of biological anthropology* 171

CHEN S., ZHOU Y., CHEN Y. & GU J. 2018. — fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34 (17): i884–i890. https://doi.org/10.1093/bioinformatics/bty560

DA SILVA M.J.F., MINHOS T., SÁ R.M. & BRUFORD M. 2012. — Using genetics as a tool in primate conservation. *Nature Education Knowledge* 3 (10): 89

GUREVICH A., SAVELIEV V., VYAHHI N. & TESLER G. 2013. — QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29 (8): 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

IWASAKI W., FUKUNAGA T., ISAGOZAWA R., YAMADA K., MAEDA Y., SATOH T.P., SADO T., MABUCHI K., TAKESHIMA H., MIYA M. & NISHIDA M. 2013. — MitoFish and MitoAnnotator: A Mitochondrial Genome Database of Fish with an Accurate and Automatic Annotation Pipeline. *Molecular Biology and Evolution* 30 (11): 2531–2540. https://doi.org/10.1093/molbev/mst141

JIN J.-J., YU W.-B., YANG J.-B., SONG Y., DEPAMPHILIS C.W., YI T.-S. & LI D.-Z. 2020. — GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21 (1): 241. https://doi.org/10.1186/s13059-020-02154-5

LI D.Y., XU H.L., SMITH D.G., CHENG A.C., TRASK J.S., ZHU Q., YAO Y.F., DU D.D. & NI Q.Y. 2011. — Phylogenetic analysis of chinese rhesus macaques (*Macaca mulatta*) based on mitochondrial control region sequences. *American Journal of Primatology* 73 (9): 883–895. https://doi.org/10.1002/ajp.20956

NABHOLZ B., ELLEGREN H. & WOLF J.B. 2012. — High levels of gene expression explain the strong evolutionary constraint of mitochondrial protein-coding genes. *Molecular biology and evolution* 30 (2): 272–284

NGUYEN M.T.P., NGUYEN T.T.T., HA T.T., HO C.N.Q., LE C.P.M., HOANG H.N.Q., NGUYEN Q.T.N., NGUYEN T.T., LUU D.T. & DANG K.D. 2024. — Characterization of the Complete Mitochondrial Genome of the Central Highland Grey-Shanked Douc Langur (Pygathrix cinerea), a Critically Endangered Species Endemic to Vietnam (Mammalia: Primates). *Current Issues in Molecular Biology* 46 (9): 9928–9947

PARVATHY S.T., UDAYASURIYAN V. & BHADANA V. 2022. — Codon usage bias. *Molecular Biology Reports* 49 (1): 539–565. https://doi.org/10.1007/s11033-021-06749-4

ROLLER M., LUCIĆ V., NAGY I., PERICA T. & VLAHOVIČEK K. 2013. — Environmental shaping of codon usage and functional adaptation across microbial communities. *Nucleic acids research* 41 (19): 8842–8852

ROOS C. & ZINNER D. 2015. — Diversity and evolutionary history of macaques with special focus on Macaca mulatta and Macaca fascicularis, *The nonhuman primate in nonclinical drug development and safety assessment*.Elsevier. p. 3–16.

SAHYOUN A.H., BERNT M., STADLER P.F. & TOUT K. 2014. — GC skew and mitochondrial origins of replication. *Mitochondrion* 17: 56–66

SMITH D.G. & MCDONOUGH J. 2005. — Mitochondrial DNA variation in Chinese and Indian rhesus macaques (Macaca mulatta). *American Journal of Primatology* 65 (1): 1–25. https://doi.org/10.1002/ajp.20094

SMITH D.R. 2016. — The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Briefings in functional genomics* 15 (1): 47–54

SPUHLER J.N. 1988. — Evolution of mitochondrial DNA in monkeys, apes, and humans. *American Journal of Physical Anthropology* 31 (S9): 15–48. https://doi.org/10.1002/ajpa.1330310504

TAMURA K., STECHER G. & KUMAR S. 2021. — MEGA11: Molecular Evolutionary Genetics Analysis Version 11, *in* BATTISTUZZI F.U. (ed.). *Molecular Biology and Evolution* 38 (7): 3022–3027. https://doi.org/10.1093/molbev/msab120

WANG W., ZHENG Y., ZHAO J. & YAO M. 2019. — Low genetic diversity in a critically endangered primate: shallow evolutionary history or recent population bottleneck? *BMC Evolutionary Biology* 19 (1): 134. https://doi.org/10.1186/s12862-019-1451-y

WOLTERS J.F. 2018. — Functional variation in the mitochondrial genome of the yeast Saccharomyces cerevisiae PhD Thesis. *State University of New York at Binghamton*.

ZARDOYA R. 2020. — Recent advances in understanding mitochondrial genome diversity. *F1000Research* 9

ZHANG D.-X. & HEWITT G.M. 1997. — Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology* 25 (2): 99–120

ZHANG Q.-L., ZHANG L., ZHAO T.-X., WANG J., ZHU Q.-H., CHEN J.-Y. & YUAN M.-L. 2017. — Gene sequence variations and expression patterns of mitochondrial genes are associated with the adaptive evolution of two Gynaephora species (Lepidoptera: Lymantriinae) living in different high-elevation environments. *Gene* 610: 148–155