IBJ
Indonesian Biodiversity Journal
http://ejurnal.unima.ac.id/index.php/ibj

indonesian.biodivers.j.
Vol. 4, No. 1, April 2023
ISSN : 2722-2659

BASE SUBSTITUTION PATTERNS IN PARTIAL OF THE CYTOCHROME C OXIDASE UNIT I (COI) mtDNA GENES IN THE RED JUNGLEFOWL (*Gallus sp.*) of NORTH SULAWESI AND SOME *Gallus sp. ACCESSIONS*

Decky D.W. Kamagi

[1]Biology Departement, Faculty of Matehematics and Natural Science, Universitas Negeri Manado, Indonesia

**Corresponding author: deckykamagi@unima.ac.id*

Received: February 26, 2023 *Accepted: April 10, 2023*

Abstract

Partial of the cytochrome C oxidase unit I (COI) mtDNA genes of the red jungle fowl (*Gallus sp.*) were isolated and sequenced to examine the base substitution pattern compared to several accessions taken from GenBank. Samples were collected from several areas in North Sulawesi in the form of muscle tissue and prepared with 95% alcohol, and stored at temperatures below 500 C. Total DNA was isolated using the innuPREP DNA micro kit with a modified protocol. The target gene amplification used the primers BirdR1 and BirdF1, and the amplicons were sequenced at the Singapore FirstBase laboratory. Target gene sequences have been blasted through the nucleotide blast program provided by the National Center for Biotechnology Information (NCBI)—sequence analysis using MEGA5 and DnaSPv5 software. Target gene sequences were aligned using the Clustal-W program on MEGA5. The amplified target gene sequence length was 490 bp. The blast results showed that the target gene sequence was 98% identical to the COI gene sequence from several Gallus-gallus accessions from GenBank. The results of multiple alignments between the sequences of some of the genes examined showed the presence of polymorphic sites (S). The number of polymorphic sites (S) is 11, while the conserved sequence (C) is 386/397 = 0.977. The number of haplotypes (h) is 5, and the diversity of haplotypes (Hd) is 0.703. The overall genetic distance average was 0.008. The polymorphic sites and base substitution events occurred at sites number 145 (A↔G), 241 (G↔C), 268 (C↔A), 295 (C↔A), 348 (T↔C), 381(A) ↔T), 384 (T↔C), 390 (T↔A), 393 (C↔T), 394 (T↔C) and 395 (C↔T). The substitution pattern at these polymorphic sites consists of substitution transitions and transversions. The ratio of transition substitution and transversion events (Ts/Tv) = 1.4056. The substitution events at these sites are located at the bases in the first and third positions of the codon. Substitution events at these polymorphic sites partially cause the exchange of amino acids. The exchange of amino acids occurs at codons 49, 81, 90, 99 and 130.

Keywords: COI gene, Red jungle fowl, North Sulawesi

INTRODUCTION

Chicken (*Gallus gallus*) has the greatest variety of species and morphology among the species on earth. Chickens are widespread and easy to find in several areas of Indonesia. Chickens of various breeds and their advantages can be found almost everywhere. Broiler chickens, laying hens, free-range chickens,

indonesian.biodivers.j.
Vol. 4, No. 1, April 2023
ISSN : 2722-2659

and dexterity-fighting chickens have distinctive characteristics due to domestication, crossing and breeding, and the development of genetic diversity. The processes of domestication, crossing, and breeding, as well as diversity development, began with the main genetic source, namely jungle fowl found in several biogeographical regions. The red jungle fowl is the ancestor of domestic chickens and is spread across Southeast Asia: Sumatra, Java to Bali, Sulawesi, the Philippines, Malaysia, India, Pakistan, and Thailand. The red jungle fowl is the main ancestor of all domestic chickens, supported by several pieces of evidence, including archaeological findings in the Indus Valley, Hebei, China, which are thought to have lived as early as 5400 BC (West & Zhou, 1988), molecular evidence in the form of mitochondrial control region sequences (Fumihito et al., 1996) and microsatellite data from various existing chicken populations.

Sequence characteristic analysis of wild chicken's (*Gallus gallus*) mitochondrial genes in North Sulawesi remains limited. Isolation, amplification, and sequencing of target genes, namely mtDNA genes in chickens, will pave the way for character analysis of target gene sequences. Finding the right method of isolating mtDNA will make it easier to obtain isolates and can then be followed up through gene amplification via PCR. Finally, sequence profiles can be revealed to analyze genetic variation. As a genetic resource (germ plasma), the jungle fowl population must be preserved so that the genetic resources are not extinct. Currently, populations of several wild species are threatened with extinction due to illegal hunting, natural enemies, and human-caused destruction of natural habitats and climate change. Efforts to conserve genetic resources (germ plasma) will be more focused and effective if the characteristics and diversity of the population are known with certainty. Information on the diversity and kinship of a population can be traced and obtained based on morphological (phenotypic), behavioral, and genetic variations of existing species. Disclosure of genetic diversity has been carried out through studies at the DNA level. Until now, many studies of genetic diversity and population biology have been carried out using mitochondrial DNA (mtDNA) genes (Avise et al., 1979; Brown et al., 1982).

The characteristics of the mtDNA genome of North Sulawesi wild-type chickens are still limited, so it is necessary to carry out basic research and studies in order to obtain some information that can be used to reveal the genetic diversity of each species, which will be useful for captive breeding strategies and programs. Or domestication. Genetic diversity disclosure has been widely carried out through protein and isoenzyme studies, but the genetic diversity that can be uncovered is not optimal due to low protein polymorphism. To overcome this problem, many researchers have turned their attention to the study of genetic diversity through DNA studies. DNA studies can produce results that can reveal more precise differences in distinguishing between intra and interspecies regarding the structure, composition, and organization at the DNA level. The DNA material used is generally nuclear DNA or total DNA. Until now, many studies of genetic diversity and population biology have been carried out using genes or mitochondrial DNA (mtDNA) (Avise et al., 1979; Brown et al., 1982). Interspecific genetic diversity studies based on mtDNA differences and similarities can produce phylogenetic reconstructions of several species that are close together. Variation patterns in mtDNA can be used to manage species or to investigate endangered species (Moritz et al., 1987). The results of the analysis of mtDNA gene sequences can be

indonesian.biodivers.j.
Vol. 4, No. 1, April 2023
ISSN : 2722-2659

placed in two distinct areas of interest, namely: a) gene conservation, in the form of identification and management of genetic diversity, and b) gene ecology, in which mtDNA variations are used as guides and aids for demographic studies of populations. Gene conservation can be carried out based on phylogenetic information and is generally very relevant for long-term planning. In contrast, gene ecology, which suggests allele frequencies, will provide useful information for population management in the short term.

The complete genome of *Gallus gallus* mitochondrial genes has been mapped (Kornegay et al., 1993) so that some of these mitochondrial genes can be used as genetic markers and can be used as study material to reveal the evolutionary history of sequences, kinship relationships, and biogeographical distribution. Various mtDNA primers for PCR-based DNA amplification, especially in groups of birds and other vertebrates, are available (Sorenson et al., 1999), so now the synthesis of various primers for PCR amplification of almost all mtDNA genes in avian groups can easily be carried out. It can even be designed with various interests, as described above. The mtDNA sequence of chicken Gallus gallus is generally known (Desjardins & Morais, 1990). However, this sequence in wild-type chickens (partridge) in North Sulawesi is still very limited and may need to be identified. The preserved mtDNA sequences from chickens, in general, can be used as a comparison to identify and trace the evolutionary history of the types of chickens living in North Sulawesi. Likewise, based on the sequence of several genes from mtDNA from chickens, it can be considered in phylogenetic reconstruction to reveal the kinship of chickens in North Sulawesi and several accessions of chickens obtained from GenBank.

RESEARCH METHODS

Sampling and Sample Preparation

Samples were collected from several muscle tissues, prepared with 95% alcohol treatment, and stored at below 50^0C.

DNA extraction

Total DNA was extracted using the innuPREPDNA Micro Kit. The isolation protocol follows the manual kit protocol with certain modifications.

Target Gene Amplification

Amplification of each target gene using primers and Bird R1 and BirdF1 for the COI gene. PCR components and conditions were prepared based on the general PCR components and conditions with modifications at certain stages.

Sequencing

The FirstBase Singapore laboratory carried out target gene sequencing.

Alignment of Sequences (Sequences Alignment)

The homologous sequences of each target gene sequence were aligned using the Clustal-W program assisted by the MEGA5.2 program.

Sequence Character Analysis

indonesian.biodivers.j.
Vol. 4, No. 1, April 2023
ISSN : 2722-2659

Character profile analysis of each target gene sequence using MEGA5.2 and DnaSPv5 software. Sequence profile character analysis of each target gene includes examining pattern and rate of substitution, genetic distance, and phylogenetic tree reconstruction.

RESULTS AND DISCUSSION

Electrophoresis

The results of the amplicon measurements of each target gene via 1.5% agarose gel electrophoresis are shown in Figure 1.

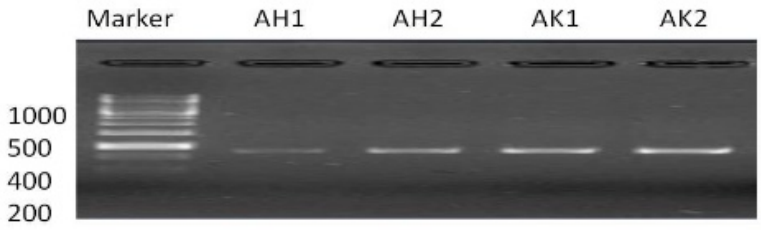

Figure 1. Separation of COI gene amplicon on 1.5% agarose gel
AH1 and AH2 = South Minahasa jungle fowl, AK1 and AK2 = North Minahasa jungle fowl

Sequencing

The amplified target gene sequence length was 490 bp. The complete sequence of some COI genes needs to be included. The proportions of target sequence bases are shown in Table 1. The proportions of bases successively starting from the largest are C=33.4%, T=26.4%, A=25.1%, and G=15, 1%. The bases in the sequences of all target gene samples showed no substitution events. Thus the target genes were 100% identical.

Tabel 1. Proporsi of COI Gene Base (%)

Domain	T	C	A	G
Red Junglefowl AH1	26,4	33,4	25,1	15,1
Red Junglefowl AH2	26,4	33,4	25,1	15,1
Red Junglefowl AK1	26,4	33,4	25,1	15,1
Red Junglefowl AK2	26,4	33,4	25,1	15,1

Multiple Sequence Alignment

A fragment of the target gene sequence with a length of 395nt was aligned to the corresponding fragment of sequences from several Gallus gallus accessions retrieved from GenBank. Duplicate align data is not included. The alignment results revealed that there were 11 polymorphic sites or various sites. The number of haplotypes (h) was 5, and the diversity of haplotypes (Hd) was 0.703. The mean overall genetic distance was 0.008. The polymorphic sites and base substitution events are site numbers 145 (A↔G), 141 (G↔C), 268 (C↔A), 295 (C↔A), 348 (T↔C), 381(A↔ T), 384 (T↔C), 390 (T↔A), 393 (C↔T), 394 (T↔C)and 395 (C↔T). Substitution at these polymorphic sites consists of transitional

indonesian.biodivers.j.
Vol. 4, No. 1, April 2023
ISSN : 2722-2659

substitution and transversion. The ratio of transition substitution and transversion events (Ts/Tv) = 1.4056. Some of these substitutions occur in the bases in the first and third positions of the codon. Substitution events at these polymorphic sites partially cause the exchange of amino acids during translation. The exchange of amino acids occurs at codons 49, 81, 90, 99, and 130.

Phylogenic Tree Reconstruction

Based on the estimation of the substitution pattern, the HKY model (Hasegawa-Kishino-Yano) is the best substitution pattern for the occurrence of substitution in the COI gene sequence. Based on the HKY substitution pattern, the Maximum Likelihood method, and the 1000x Bootstrapping test, the phylogenetic tree reconstruction is shown in Figure 2. The tree topology shows a tree without roots, where all red junglefowl samples cluster with *Gallus gallus gallus* and is separated from other accession groups.

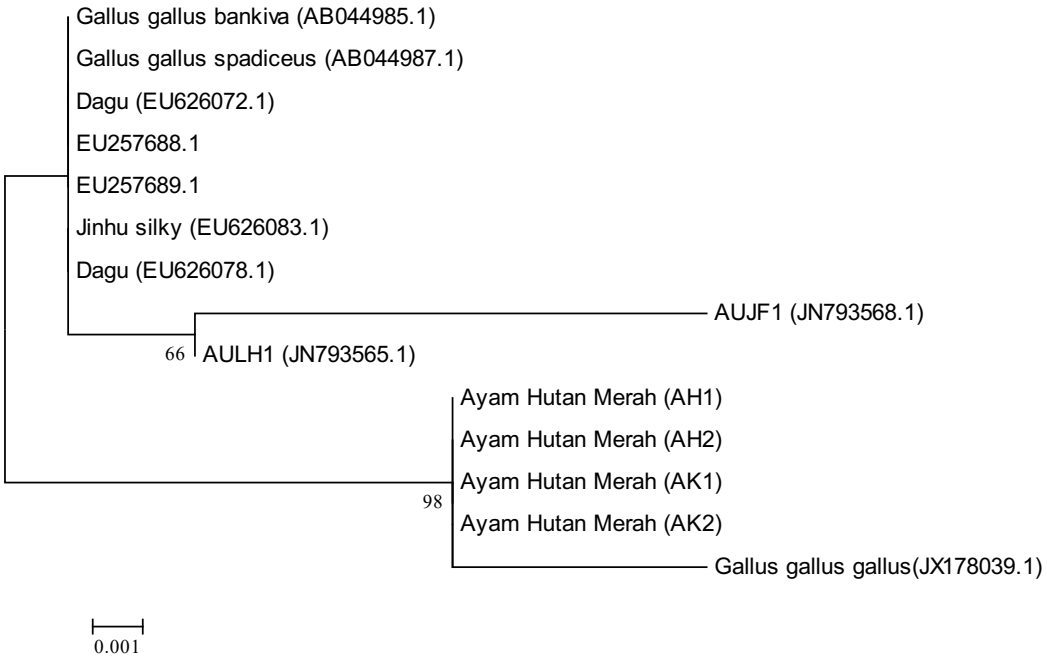

Figure 2. Reconstruction of the red jungle fowl phylogeny tree and several Gallus gallus accessions taken from GenBank based on the COI gene, the Maximum Likelihood method, and the 1000x Bootstrapping test.

DISCUSSION

Sequence profile

The COI gene (Cytochrome C Oxidase unit I) is one of the protein-coding genes in mtDNA. The product of the COI gene is the cytochrome oxidase subunit I. In the chicken mitochondrial genome, the COI gene is located between sequences 6651 to 8201 with a length of 1550 bp, flanked by the tRNA-Tyr gene and the tRNA-Ser gene (Nishibori et al., 2005). The COI gene has a start codon GTG and a stop codon AAG. The COI gene contains a group of discrete characters (each codon position) that describes the mutation rate so that it can be used as a phylogenetic marker or as a barcode (Desjardins and Morais.

indonesian.biodivers.j.
Vol. 4, No. 1, April 2023
ISSN : 2722-2659

1990).

It is well known that individual genes and the whole genome vary significantly in nucleotide composition. The diversity of nucleotide composition is widely found in nuclear and extrachromosomal DNA, such as mtDNA. The diversity of nucleotide composition has been shown to be significantly related to the composition of amino acids in proteins (Singer & Hickey, 2002). The diversity of the nucleotide composition of a gene or genome is reflected in the sequence of nitrogenous bases contained in the two nucleotide strands.

The mitochondrial genome of vertebrates and mammals is a closed circular strand and contains 13 protein-coding genes, two rRNA genes, and 22 tRNA genes. The two strands that comprise the genome are the H strand (heavy strand) and the L strand (light strand). Of the 13 coding (coding) genes, 12 are on the H strand, and only one is on the L strand. The non-coding regions are mainly confined to a region called the D-loop, thought to have a functional role in replication and transcription and the initiation of strand replication L (Clayton, 2000). Variation in the base composition of nucleotides is usually most pronounced in the base positions at synonymous codons of genes due to redundancies in the genetic code and variations in DNA content, which may have little effect on the amino acid content of proteins (Singer & Hickey, 2000).

As a result of the double alignment of sample sequences plus sequences taken from GenBank, it was found that several sites were polymorphic as a result of base substitution events, which involved transition and transversion substitution events. Based on the ratio of transitions and transversions Ts/Tv > 1, it turns out that there are more transitional substitutions than transversion substitutions. Comparison of the occurrence of transition substitution, which is much larger than transversion substitution, is in line with Brown et al. al. (1982) that transition substitution is more dominant in mitochondrial genes than transversion substitution. The high occurrence of substitution transitions compared to transversions is common in eutherian mtDNA sequences (Brown et al., 1984). Kocher et al. (1989) suggested that usually, in mtDNA, transition substitution is more dominant than transversion substitution, and T↔C saturation will be much greater than A↔G. Furthermore, Kocher et al. (1989) suggested that most nucleotide substitutions at the species level are transitions, while those at the genus level are transversions. This confirms that the red jungle fowl and several accessions taken from the Gen Bank are closely related because they are still in the species taxa hierarchy.

Brown et al. (1982) suggested that the emergence of high transition events in mtDNA manifested increased mutation pressure. However, the nucleotide substitution rate did not correlate with the rate of structural change in the organelle's genome. In higher animals, mtDNA is indeed fast in nucleotide substitution, but the arrangement and size of the genes of the genome are the same for each species (Castro et al., 1998). Brown et al. (1982) said that the factors responsible for the high mutation rate of mtDNA exceeded that of nuclear DNA, including a) the tendency for damage to the replication system, b) the inefficiency of the editing function, and c) the high rate of exchange (turnover).

The sequence profile, which is characterized by the proportion of each base that is owned by each

indonesian.biodivers.j.
Vol. 4, No. 1, April 2023
ISSN : 2722-2659

sample studied, it turns out that the content of base C is greater than the other bases, while the smallest is base G. Likewise, the proportion of bases owned by each sample was added with several accessions taken from the Gen Bank, the proportion of C base was the largest and G base was the smallest. In this regard, Avise (1994) suggested that the characteristic of protein-coding genes in mitochondria is the composition of low G and high C bases. The findings of Irwin et al. (1991) in several mammalian vertebrates showed that the frequency of G bases at all positions in the codon was low, and in humans, the frequency of G bases was the lowest.

Phylogeny

Domestication of wild chickens started in Southeast Asia, then spread to China, and then to Europe (West & Zhou, 1988). The jungle fowl that live in Southeast Asia consists of *Gallus gallus gallus* (red jungle fowl), Gallus lafayette (La Fayette jungle fowl), and *Gallus varius* (green jungle fowl) (Sawai et al., 2010). The red jungle fowl is markedly sexually dimorphic, with the male having red fleshy wattles, and is distributed in almost all regions. The La Fayette chicken morphologically resembles the red jungle fowl but is only found in Sri Lanka. The green jungle fowl is only found on Java, Bali, and Lombok islands. Two subspecies of the red jungle fowl, namely *Gallus gallus gallus* and Gallus gallus spadieus, are the ancestors of the domestic chicken, while *Gallus gallus bankiva* does not contribute (Fumihito et al., 1996). The results of a phylogenetic tree reconstruction based on the COI mtDNA gene, where the North Sulawesi red jungle fowl is grouped with *Gallus gallus gallus* and separated from other subspecies groups, provide evidence and support that the North Sulawesi red jungle fowl is closely related to *Gallus gallus gallus*, or in other words, chicken the red forest of North Sulawesi is grouped into the subspecies *Gallus gallus gallus*.

CONCLUSION

Some of the COI gene sequences of the North Sulawesi red jungle fowl show a specific profile. Based on some COI genes, the North Sulawesi red jungle fowl is closely related to Gallus gallus gallus. Further research is needed to obtain complete information regarding the sequence profile and evolutionary history of the North Sulawesi red jungle fowl through disclosure of the entire mtDNA genome or the use of other genetic markers.

REFERENCE

Avise, J. C., R. A., Lansman., & R. O., Shade. (1979). Use endonuclease to Measure Mitochondrial DNA Sequence Relatedness in Natural Population. I. Population Structure and Evolution in Genus Peromyscus. *Genetics, 92,* 279-295.

Brown, W. M., E.M. Prager, A. Wang, & Wilson, A. C. (1982). Mitochondrial DNA sequences of Primates: Tempo and Mode of Evolution. *Journal of Molecular Evolution, 18,* 225-239.

Desjardins, P., & R. Morais. (1990). Sequence and gene organization of the chicken mitochondrial genome. *Journal of Molecular Biology, 212*, 599-634.

Fumihito A, Miyake T, Takada M, Shingu R, Endo T, *et al*. (1996) Monophyletic origin and unique dispersal

indonesian.biodivers.j.
Vol. 4, No. 1, April 2023
ISSN : 2722-2659

patterns of domestic fowls. Proc Natl Acad Sci U S A 93: 6792-6795.

Karanth, K.P., S. Lalji, R. V. Collura., & S.C. Beth. (2008). Molecular phylogeny and biogeography of langurs and leaf monkey of South Asia. *Molecular Phylogenetic Evolution, 46,* 683-694.

Kornegay,J.R., Kocher,T.D., Williams,L.A., & Wilson,A.C. (1993). Pathways of lysozyme evolution inferred from the sequences of cytochrome b in birds. *Journal Molecular Evolution. 37(4),* 367-379.

Lim, L. S., K. C. Ang, M.C. Mahani, A.W. Shahrom., & B.M. Md-Zain. (2010). Mitochondrial DNA polymorphism and phylogenetic relationships of proto malays in peninsular Malaysia. *Journal of Biology Science, 10*, 71-83.

Master, J. C., M. Boniotto, S. Crovella, C. Roos, L. Pozzi., & M. Delpero, (2007). Phylogenetic relationships Among the Lorisoidea As Indicates by Craniodental Morphology and Mitochondrial Sequence data, *American Journal of Primatology, 69,* 6-15.

Moritz, C., T. E. Dowling., &W. M. Brown. (1987). Evolution of animal mitochondrial DNA. *Annual Review of Ecology and Systematics, 18,* 269-292.

Nishibori, M., Shimogiri, T., Hayashi, T., & Yasue, H. (2005). Molecular evidence for hybridization of species in the genus Gallus except for Gallus varius. *Animal Genetics, 36,* 367-375.

Randi, E. (1996). A mitochondrial cytochrome *b* phylogeny of the *Alectoris* partridges. *Mol. Phylogenet. Evol, 6,* 214-227.

Sawai H., Kim H. L., Kuno K., Suzuki S., Gotoh H., Takada M., Takahata N., Satta Y., & Fumihito A. (2010). The Origin and Genetic Variation of Domestic Chickens with Special Reference to Junglefowls *Gallus g. gallus* and *G. varius*. *PlosOne, 5(5).*

Roos, C., T. Nadler, & L. Walter, (2008). Mitochondrial phylogeny, taxonomy and biogeography of silver langur spesies group (*Trachypithecus cristatus*). *Molecular, Phylogenetics Evolution 47,* 629-636.

Sorenson, M.D., J.C. Ast, D.E. Dimcheff, T. Yuri & D.P. Mindell. (1999). Primers for a PCR based approach to mitochondrial genome sequencing in birds and other vertebrates. *Molecular Phylogenetics and Evolution, 12,* 105-114.

West, B., & Zhou, B. (1988) Did chickens go north? New evidence for domestication. *Journal of Archaeological Science, 15,* 515-533.