www.als-journal.com/ ISSN 2310-5380/ December2023

Full Length Research Article
Advancements in Life Sciences – International Quarterly Journal of Biological Sciences

Open Access

The genetic diversity of *Drosophila* flies based on cytochrome-c oxidase subunit I gene from North Sulawesi

Yermia Semuel Mokosuli*, Herry Maurits Sumampouw

ARTICLE INFO

Date Received: 22/06/2023;
Date Revised: 18/08/2023;
Date Published Online: 31/12/2023;
Date Updated: 09/09/2025

Authors' Affiliation:
Universitas Negeri Manado - Indonesia

***Corresponding Author:**
Yermia Semuel Mokosuli
Email:
yermiamokosuli@unima.ac.id

How to Cite:
Mokosuli YS, Sumampouw HM (2023). The genetic diversity of Drosophila flies based on cytochrome-c oxidase subunit I gene from North Sulawesi. Adv. Life Sci. 10(4): 563-570.

Keywords:
Genetic diversity; phylogeny; CO1; *Drosophila*; Sulawesi Utara

Editorial Note:
You are viewing the latest version of this article having language corrections.

Abstract

Background: *Drosophila* sp. is a fruit fly species that can spread quickly worldwide. A short life cycle, small body size, and fast adaptation to new habitats allow fruit flies to live in various parts of the world. This study characterized the genetic variation of the cytochrome oxidase c subunit I (COI) gene in Drosophila species from different North Sulawesi regions.

Methods: Fruit fly samples were collected from six districts: Central Minahasa, Southeast Minahasa, South Minahasa, North Minahasa, Bolaang Mongondow, and Sitaro. DNA extracted from thoracic tissue The COI gene was amplified by polymerase chain reaction (PCR) and sequenced by the Sanger method. Sequence characterization using BioEdit and MEGA XI programs.

Result: The consensus COI gene sequence length ranged from 688 bp to 700 bp. Disparity, genetic distance, and phylogenetic analyses consistently revealed that the sample from Bolaang was the most genetically distinct, forming a separate branch from the other five monophyletic North Sulawesi samples. Genetic distance analysis showed that Bolaang's gene sequence has the farthest genetic distance. COI gene consensus alignment analysis with ClustalW showed high genetic variation. The phylogenetic construction showed that COI Bolaang had the most significant differences in sequence characteristics from the other five sequences in a monophyletic group with different nodes. Phylogenic reconstruction with the 21 most similar sequences from BLAST showed similarities in the four *Drosophila* species, namely *D. atriplex*, *D. melanogaster*, *D. lacteicornis,* and *D. pandora.*

Conclusion: Based on the COI gene, there are significant variations in fruit flies in North Sulawesi.

Introduction

The genus *Drosophila* is widely used in the study of biological phenomena, including developmental biology and genomics [1,2], microbiome [3], human disease mechanisms [4, 5], and genetic studies [6, 7]. Understanding intraspecies and interspecies phylogenetic relationships within this family greatly influences studying these biological phenomena. However, studies on the genetic diversity of fruit flies are still underreported, primarily based on their habitat origin and biogeographic regions. There are still no research reports on the genetic diversity of fruit flies in Indonesia, more specifically in the biogeographical region of Wallacea. Fruit flies are found in tropical to subtropical areas. This causes high intraspecific genetic variation in fruit flies around the world. Furthermore, the short life cycle, small body size, and high reproduction speed reinforce the process of divergent evolution in fruit flies [8-10].

North Sulawesi Province is the northernmost part of Sulawesi Island, Indonesia. However, Sulawesi is included in the Wallacea zone, which has specific flora and fauna biodiversity, in contrast to the western part of Indonesia, which has oriental biogeographical characteristics, and the eastern part of Indonesia, which has Australian biogeographical characteristics [11, 12]. The species endemicity in Sulawesi is relatively high. Many endemic fruit plant species are also food sources for fruit flies. Some endemic plant species include Pakoba (*Syzygium luzonense* (Merr.), Nutmeg, Beringin Minahasa (*Ficus minahassae*), Langsat forest (*Lansium domesticum* L.), Forest Mangosteen (*Garcinia mangostana* L.), tome (*Flacourtia inermis* L.), etc. Insects can coevolve with plant food sources, thereby affecting their genetic diversity. It is suspected that this process took place in *Drosophila* sp. Research on the genetic diversity of fruit flies is essential in North Sulawesi, a region that produces various types of local fruits.

Morphological characteristics have historically been used to identify taxonomy and genetic diversity. In the last two decades, molecular markers for animal identification have been widely used, namely the Cytochrome oxidase subunit I (COI) gene. DNA barcoding data can be used for more than just taxonomic studies. It will have significant effects on a wide range of biological disciplines, such as ecology (quick examination of the food chain and biodiversity), conservation biology (surveillance of protected species), biosafety (identifying invasive pest species early), pharmaceuticals (identifying medically necessary infections and their carriers), and pharmacology (the detection of active substances). The COI gene has been widely used as a molecular marker in the *Drosophila* family, so its accuracy is believed [13-15]. Based on previous morphological studies, there were differences in morphology, especially eye color, abdomen color, wing length, and average body length of *Drosophila* sp. from island locations and the Sulawesi mainland. The difference in the average body length of fruit flies affects the shape and size of the reproductive organs. This can lead to reproductive isolation [16]. Geographical isolation has caused morphological variations of Drosophila sp. Demographic history, geographic isolation, and environmental factors together form the genetic structure of the population of *D. melanogaster* [17, 18]. *Drosophila* sp. can coevolve with food source plants in chemoreceptors [19].

However, little is known about research on the genetic diversity of fruit flies in the Sulawesi (Wallacea) biogeographical zone. Therefore, a study was conducted to characterize the COI gene of the North Sulawesi fruit fly from various locations. The findings of this study provide excellent material for evolutionary studies and phylogenetic relationships, particularly at the species level of *Drosophila* sp. in specific areas.

Methods

Collection of samples *Drosophila* sp.

Adult *Drosophila* sp. samples were obtained directly from several locations in North Sulawesi (Figure 1). At each location, ten adult individuals of *Drosophila* sp. were isolated and preserved in 95% alcohol for 24 hours. After 24 hours, the fruit flies were transferred to a new bottle containing 95% alcohol. Sample collection was carried out from March to October 2022.

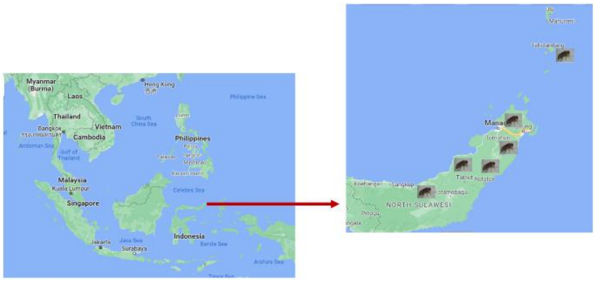


Figure 1: Map of Indonesia, the location of the *Drosophila* sp. collection in North Sulawesi Province is labeled with a picture of a *Drosophila* sp.

Extraction, Purification of DNA, PCR, and Sequencing.

Fruit fly genomic DNA was extracted from the thoracic tissue (Fig. 2). The tissues were homogenized (SPEX Sample Prep 1600 Mini G) before being digested overnight at 56°C. A Quick-DNATM Miniprep Kit from Zymo Research USA was used to extract DNA according to the manufacturer's instructions. Extracted total DNA was stored at -20 °C for use in the COI gene amplification stage. The 700 bp COI gene fragment was amplified using the primer pairs listed below: HCO, 5'-TAAACTTCAGGGTGACCAAAAAATCA-3' and LCO, 5'-

GGTCAACAAATCATAAAAGATATTGG-3' (Folmer et al., 1994) [20]. The PCR process used 2x MyTaq HS Red Mix Bioline (USA), DNA template, ddH₂O, and COI primer. The PCR component was 2x MyTaq HS Red Mix Bioline (USA) 25 µl; primer forward 1 µl; primer reverse 1 µl; DNA template 2 µl, ddH_2O 21 µl. PCR conditions were denaturation at 94 °C (60 seconds), annealing at 50 °C (30 seconds), Extension at 72 °C (30 seconds), and Final Extension 72^0C (60 seconds) [21].

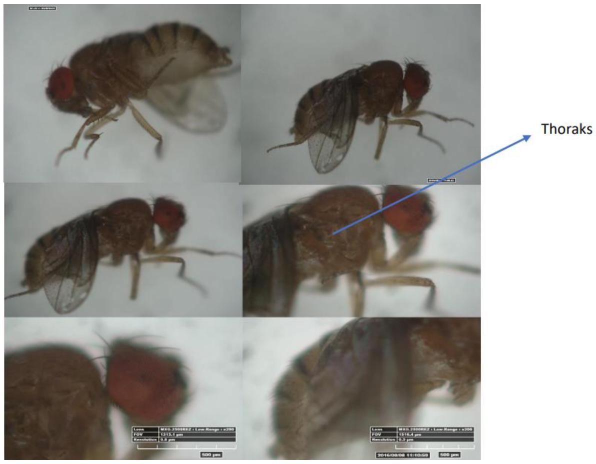


Figure 2: *Drosophila* sp. samples used for DNA extraction, 200x with Hirox KH8700 stereomicroscope.

Amplicon Visualization and Sequencing

The amplicon of the COI gene of Drosophila produced by PCR was visualized using conventional electrophoresis (1% TBE agarose—M, 100bp ladder (loaded 2.5 L). Sequencing uses the ABI PRISM 3730xl Genetic Analyzer developed by Applied Biosystems, USA, through SingaporeFIRST BASE sequencing services. The sequencing output in a seq file was analyzed using MEGA XI and BioEdit.

Data analysis

Sequence Analysis

Partial sequences of the COI gene, 600-700 bp long, were collected from 6 North Sulawesi, Indonesia districts. Forward and reverse sequences were analyzed using BioEdit to obtain consensus sequences. Each sequence is first translated into an amino acid sequence to detect and remove any stop codons in the middle of the sequence. In addition to careful manual checking, sequence checking, and editing were done using the BioEdit software (Hall, 1999). Sequence characteristics were analyzed for the COI gene, and the amino acids were analyzed using BioEdit. Each sample sequence was validated using the GenBank NCBI's online Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi). For phylogenetic tree reconstruction, the 21 accessions from GenBank NCBI were selected as in-group and out-group (Table 1). In MEGA XI software version 10.2.6, multiple sequences were reassigned using the ClustalW Algorithm [22].

No	Accession Number Species	Origin
1	Select seq MK659835.1 *Drosophila pandora*	China
2	Select seq OK037194.1 *Drosophila atriplex*	Malaysia
3	Select seq MK659836.1 *Drosophila parapallidosa*	China
4	Select seq MK659805.1 *Drosophila ananassae*	USA
5	Select seq MK659811.1 *Drosophila bipectinata*	China
6	Select seq MK659826.1 *Drosophila malerkotliana*	France
7	Select seq AY757282.1 *Drosophila parabipectinata*	Canada
8	Select seq OK175852.1 *Drosophila ananassae*	Khandesh
9	Select seq AY757280.1 *Drosophila pseudoananassae*	Canada
10	Select seq MK455900.1 *Drosophila* sp.	India
11	Select seq AB907180.1 *Drosophila melanogaster*	India
12	Select seq AB830535.1 *Drosophila lacteicornis*	Japan
13	Select seq MK659817.1 *Drosophila ficusphila*	China
14	Select seq AB669729.1 *Drosophila bocki*	Japan
15	Select seq MK659806.1 *Drosophila anomalata*	China
16	Select seq AB669702.1 *Drosophila neoasahinai*	Japan
17	Select seq AB669709.1 *Drosophila tani*	Japan
18	Select seq AB669704.1 *Drosophila rufa*	Japan
19	Select seq AF200849.1 *Drosophila simulans*	USA
20	Select seq AB669699.1 *Drosophila lacteicornis*	Japan

Table 1: BLAST results for phylogeny construction with *Drosophila* sp. COI sequences from North Sulawesi.

Phylogenetic tree reconstruction

The Maximum Composite Likelihood model was used to compute the COI gene disparity matrix of the Sulawesi fruit fly. The North Sulawesi fruit fly COI gene sequence was aligned using the ClustalW model in the MEGA X program version 11.2.6 and online at the European Bioinformatics Institute (https://www.ebi.ac.uk). On 27 sequences, a phylogenetic tree reconstruction based on partial COI gene sequences was performed (in-group and outgroup accessions obtained from GenBank). MEGA X version 11.2.6 was used to reconstruct the phylogenetic tree. Meanwhile, the phylogenetic tree reconstruction methods used are Neighbor-Joining (N.J.) and Maximum Likelihood (ML). The NJ reconstruction was computed using the Kimura 2-Parameter substitution model (K2P) [22]. Bootstrap 1000 replicates were used to evaluate phylogenetic tree reconstruction.

Results

Characterization of Sequences Obtained

The amplicons of the sequenced results were classified as good, as evidenced by the chromatogram of the sequenced results, which showed that the bands of the nitrogenous base types did not coincide much (Fig. 3). Sequencing of the COI gene sequence of fruit flies revealed an average length of 600 bp – 700 bp (Table 2). Characteristics of *Drosophila* sp. DNA COI gene has a G+C ratio of 30.20% to 32.82% (Appendix 1). At the same time, the A+T ratio is between 67.01% and 65.78%. The consensus COI gene molecular length is between 688 bp and 716 bp. All the characteristics of the Drosophila sp. COI gene from various regions in North Sulawesi, analyzed with the BioEdit Program, is shown in Appendix 1.

No.	Samples	Sequence length (bp)		Consensus sequence length (Bio Edit v.7.2.5 & MEGA 11)
		LCO 1490	HCO 2198	
1	L1 (Minahasa)	683	700	716
2	L2 (Minteng)	683	696	688
3	L3 (Minut)	682	684	692
4	L4 (Minsel)	685	684	713
5	L5 (Sitaro)	685	684	697
6	L6 (Bolaang)	684	687	685

Table 2: The sequence length of the *Drosophila* sp COI gene from North Sulawesi.

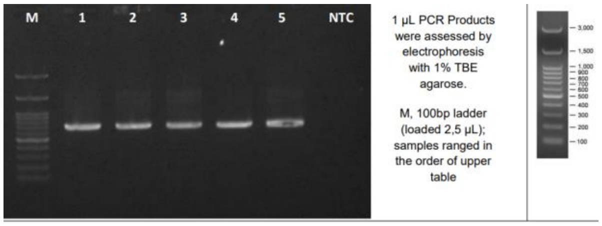
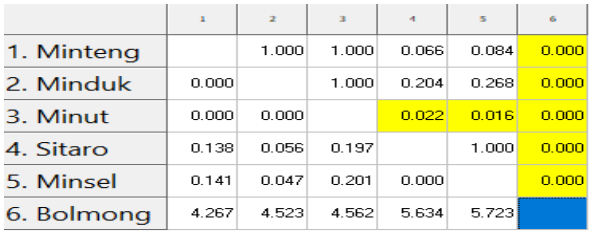


Figure 3: *Drosophila* sp. from North Sulawesi COI DNA profile amplified by primers LCO 1490 and HCO 2198. Electrophoresis of 1 µL PCR products with 1% TBE agarose M, 100bp ladder (loaded 2.5 L) was performed; 1). 1: L1 (Minahasa), 2: L2 (Southeast Minahasa), 3: L3 (North Minahasa), 4: L4 (South Minahasa), 5: L5 (Siau Tangulandang Biaro), 6: L6 (Bolaang Mongondow).

Use of the sequence analysis model

Genetic Disparity and Distance Matrix

The disparity matrix analysis was used to estimate divergent evolution between sequences. The number of base substitutions per site between sequences is displayed. The Maximum Composite Likelihood model was used in the analysis. This investigation involves five nucleotide sequences. The included codon positions are 1st+2nd+3rd+Noncoding. According to the disparity matrix analysis results, the longest distance was found between the Bolaang COI gene sequence and the other five sequences. In contrast, the closest distances are COI Minsel and COI Sitaro (Table 3).



	1	2	3	4	5	6
1. Minteng		1.000	1.000	0.066	0.084	0.000
2. Minduk	0.000		1.000	0.204	0.268	0.000
3. Minut	0.000	0.000		0.022	0.016	0.000
4. Sitaro	0.138	0.056	0.197		1.000	0.000
5. Minsel	0.141	0.047	0.201	0.000		0.000
6. Bolmong	4.267	4.523	4.562	5.634	5.723	

Table 3: Disparity Matrix of COI Gene DNA, *Drosophila* sp. from North Sulawesi.

The ratio of genetic differences between species or populations is depicted by genetic distance. The Bolaang Mongondow COI gene sequence had the most significant genetic distance between the COI genes of the Sulawesi fruit fly. The Minsel COI, Sitaro COI, and Main Minahasa COI genes followed this. COI Sitaro and COI Minsel have the closest genetic distance (Table 4).

	Minteng	Minduk	Minut	Sitaro	Minel	Bolaang
Minteng	0.0000	0.0029	0.0227	0.1140	0.1105	1.2270
Minduk	0.0029	0.0000	0.0335	0.1297	0.1357	1.2377
Minut	0.0227	0.0335	0.0000	0.0924	0.0889	1.2208
Sitaro	0.1140	0.1297	0.0924	0.0000	0.0000	1.2680
Minsel	0.1105	0.1357	0.0889	0.0000	0.0000	1.2617
Bolaang	1.2270	0.1237	1.2208	0.1268	0.1261	0.0000

Table 4: The genetic distance of the *Drosophila* sp COI gene from North Sulawesi.

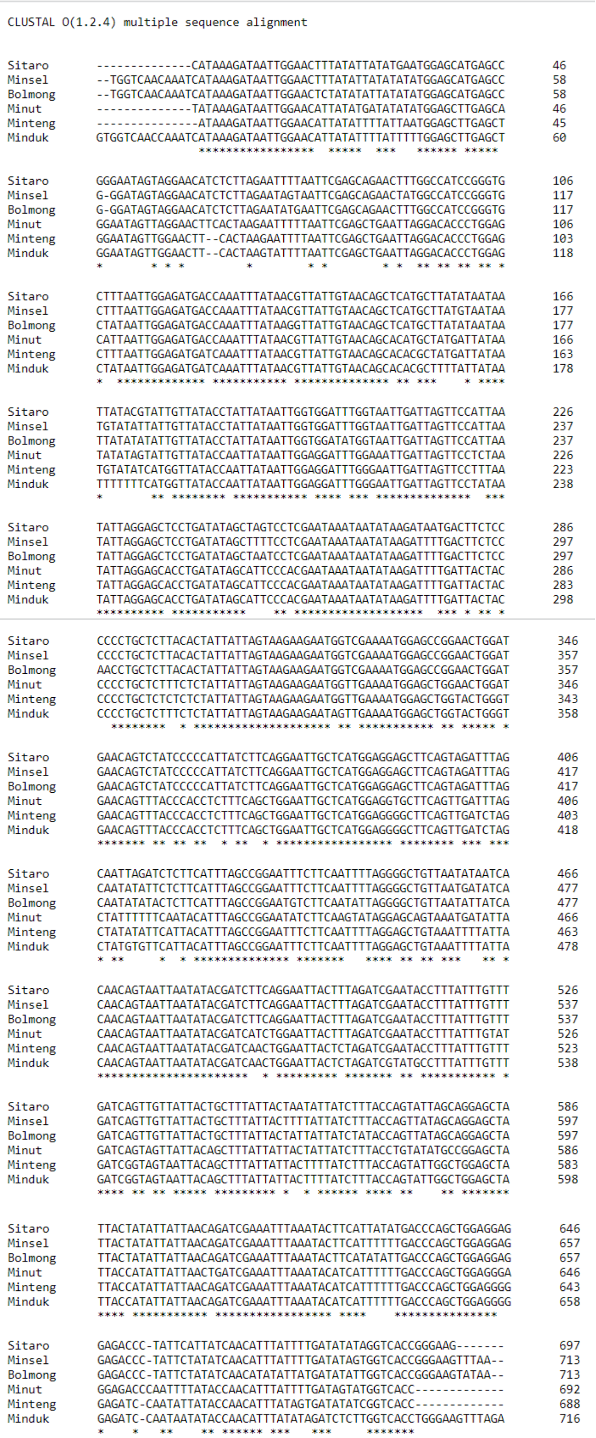
Furthermore, Maximum Likelihood (ML) analysis was used to estimate nucleotide substitution. Each entry represents the likelihood of substitution (r) from one base (row) to another base (column). The Tamura-Nei model was used to estimate turnover patterns and rates. Transitional substitution levels are bold, and transversion substitutions are in italics. When evaluating r, the relative instantaneous value must be taken into account. For clarity, the sum of the r values was set to 100. The nucleotide frequencies were A = 30.43%, T/U = 37.03%, C = 15.57%, and G = 16.97%. (Table 5). To estimate the ML value, the tree topology is calculated automatically. The maximum possible Log for this calculation is -2222.697. Six nucleotide sequences are used in this analysis. The included codon positions are 1st+2nd+3rd+Noncoding. The final dataset contains 736 positions in total. MEGA X1 was used for evolution analysis.

Base	A	T/U	C	G
A	-	*8.97*	*3.77*	6.88
T/U	*7.37*	-	9.58	*4.11*
C	*7.37*	22.78	-	*4.11*
G	12.34	*8.97*	*3.77*	-

Table 5: Substitution Matrix Maximum Likelihood Estimate

Alignment Analysis with ClustalW

The COI gene sequences of the North Sulawesi fruit fly were aligned using ClustalW. ClustalW is a standard method for aligning multiple nucleotide sequences or homologous proteins. ClustalW uses the progressive alignment method for multi-sequence alignment. In this case, the sequence with the highest alignment score is aligned first. Then, progressively more distant sequence groups are aligned until a global alignment is collected. The alignment results indicated the sites of nucleotide differences between the 6 COI gene sequences of fruit flies from North Sulawesi. The Bolaang COI gene sequence showed the highest nucleotide variation. The Minteng COI, Minduk COI, and Minut COI gene sequences followed this. The COI Minsel and COI Sitaro gene sequences show many nucleotide similarities (Figure 4).



```
CLUSTAL O(1.2.4) multiple sequence alignment


Sitaro          --------------CATAAAGATAATTGGAACTTTATATTATATGAATGGAGCATGAGCC      46
Minsel          --TGGTCAACAAATCATAAAGATAATTGGAACTTTATATTATATATATGGAGCATGAGCC      58
Bolmong         --TGGTCAACAAATCATAAAGATAATTGGAACTCTATATATTATATATGGAGCATGAGCC      58
Minut           --------------TATAAAGATAATTGGAACATTATATGATATATATGGAGCTTGAGCA      46
Minteng         ---------------ATAAAGATAATTGGAACATTATATTTTATTAATGGAGCTTGAGCT      45
Minduk          GTGGTCAACCAAATCATAAAGATAATTGGAACATTATATTTTATTTTTGGAGCTTGAGCT      60
                               *****************  *****  ***   ****** *****

Sitaro          GGGAATAGTAGGAACATCTCTTAGAATTTTAATTCGAGCAGAACTTTGGCCATCCGGGTG      106
Minsel          G-GGATAGTAGGAACATCTCTTAGAATAGTAATTCGAGCAGAACTATGGCCATCCGGGTG      117
Bolmong         G-GGATAGTAGGAACATCTCTTAGAATATGAATTCGAGCAGAACTTTGGCCATCCGGGTG      117
Minut           GGAATAGTTAGGAACTTCACTAAGAATTTTTAATTCGAGCTGAATTAGGACACCCTGGAG      106
Minteng         GGAATAGTTGGAACTT--CACTAAGAATTTTAATTCGAGCTGAATTAGGACACCCTGGAG      103
Minduk          GGAATAGTTGGAACTT--CACTAAGTATTTTAATTCGAGCTGAATTAGGACACCCTGGAG      118
                *       * * *         *        * *        * *  ** ** ** ** *

Sitaro          CTTTAATTGGAGATGACCAAATTTATAACGTTATTGTAACAGCTCATGCTTATATAATAA      166
Minsel          CTTTAATTGGAGATGACCAAATTTATAACGTTATTGTAACAGCTCATGCTTATGTAATAA      177
Bolmong         CTATAATTGGAGATGACCAAATTTATAAGGTTATTGTAACAGCTCATGCTTATATAATAA      177
Minut           CATTAATTGGAGATGACCAAATTTATAACGTTATTGTAACAGCACATGCTATGATTATAA      166
Minteng         CTTTAATTGGAGATGATCAAATTTATAACGTTATTGTAACAGCACACGCTATGATTATAA      163
Minduk          CTATAATTGGAGATGATCAAATTTATAACGTTATTGTAACAGCACACGCTTTTATTATAA      178
                *  ************* *********** ************** ** ***    * ****

Sitaro          TTATACGTATTGTTATACCTATTATAATTGGTGGATTTGGTAATTGATTAGTTCCATTAA      226
Minsel          TGTATATTATTGTTATACCTATTATAATTGGTGGATTTGGTAATTGATTAGTTCCATTAA      237
Bolmong         TTATATATATTGTTATACCTATTATAATTGGTGGATATGGTAATTGATTAGTTCCATTAA      237
Minut           TATATAGTATTGTTATACCAATTATAATTGGAGGATTTGGAAATTGATTAGTTCCTCTAA      226
Minteng         TGTATATCATGGTTATACCAATTATAATTGGAGGATTTGGGAATTGATTAGTTCCTTTAA      223
Minduk          TTTTTTTCATGGTTATACCAATTATAATTGGAGGATTTGGGAATTGATTAGTTCCTATAA      238
                *       ** ******** *********** **** *** **************  ***

Sitaro          TATTAGGAGCTCCTGATATAGCTAGTCCTCGAATAAATAATATAAGATAATGACTTCTCC      286
Minsel          TATTAGGAGCTCCTGATATAGCTTTTCCTCGAATAAATAATATAAGATTTTGACTTCTCC      297
Bolmong         TATTAGGAGCTCCTGATATAGCTAATCCTCGAATAAATAATATAAGATTTTGACTTCTCC      297
Minut           TATTAGGAGCACCTGATATAGCATTCCCACGAATAAATAATATAAGATTTTGATTACTAC      286
Minteng         TATTAGGAGCACCTGATATAGCATTCCCACGAATAAATAATATAAGATTTTGATTACTAC      283
Minduk          TATTAGGAGCACCTGATATAGCATTCCCACGAATAAATAATATAAGATTTTGATTACTAC      298
                ********** ***********    ** *******************  *** * ** *

Sitaro          CCCCTGCTCTTACACTATTATTAGTAAGAAGAATGGTCGAAAATGGAGCCGGAACTGGAT      346
Minsel          CCCCTGCTCTTACACTATTATTAGTAAGAAGAATGGTCGAAAATGGAGCCGGAACTGGAT      357
Bolmong         AACCTGCTCTTACACTATTATTAGTAAGAAGAATGGTCGAAAATGGAGCCGGAACTGGAT      357
Minut           CCCCTGCTCTTTCTCTATTATTAGTAAGAAGAATGGTTGAAAATGGAGCTGGAACTGGAT      346
Minteng         CCCCTGCTCTCTCTCTATTATTAGTAAGAAGAATGGTTGAAAATGGAGCTGGTACTGGGT      343
Minduk          CCCCTGCTCTTTCTCTATTATTAGTAAGAAGAATAGTTGAAAATGGAGCTGGTACTGGGT      358
                  ********  * ******************** ** *********** ** ***** *

Sitaro          GAACAGTCTATCCCCCATTATCTTCAGGAATTGCTCATGGAGGAGCTTCAGTAGATTTAG      406
Minsel          GAACAGTCTATCCCCCATTATCTTCAGGAATTGCTCATGGAGGAGCTTCAGTAGATTTAG      417
Bolmong         GAACAGTCTATCCCCCATTATCTTCAGGAATTGCTCATGGAGGAGCTTCAGTAGATTTAG      417
Minut           GAACAGTTTACCCACCTCTTTCAGCTGGAATTGCTCATGGAGGTGCTTCAGTTGATTTAG      406
Minteng         GAACAGTTTACCCACCTCTTTCAGCTGGAATTGCTCATGGAGGGGCTTCAGTTGATCTAG      403
Minduk          GAACAGTTTACCCACCTCTTTCAGCTGGAATTGCTCATGGAGGGGCTTCAGTTGATCTAG      418
                ******* ** ** **  * **  * ***************** ******** *** ***

Sitaro          CAATTAGATCTCTTCATTTAGCCGGAATTTCTTCAATTTTAGGGGCTGTTAATATAATCA      466
Minsel          CAATATATTCTCTTCATTTAGCCGGAATTTCTTCAATTTTAGGGGCTGTTAATGATATCA      477
Bolmong         CAATATATACTCTTCATTTAGCCGGAATGTCTTCAATATTAGGGGCTGTTAATATTATCA      477
Minut           CTATTTTTTCAATACATTTAGCCGGAATATCTTCAAGTATAGGAGCAGTAAATGATATTA      466
Minteng         CTATATATTCATTACATTTAGCCGGAATTTCTTCAATTTTAGGAGCTGTAAATTTTATTA      463
Minduk          CTATGTGTTCATTACATTTAGCCGGAATTTCTTCAATTTTAGGAGCTGTAAATTTTATTA      478
                * **     *  * ************** *******   **** ** ** ***   ** *

Sitaro          CAACAGTAATTAATATACGATCTTCAGGAATTACTTTAGATCGAATACCTTTATTTGTTT      526
Minsel          CAACAGTAATTAATATACGATCTTCAGGAATTACTTTAGATCGAATACCTTTATTTGTTT      537
Bolmong         CAACAGTAATTAATATACGATCTTCAGGAATTACTTTAGATCGAATACCTTTATTTGTTT      537
Minut           CAACAGTAATTAATATACGATCATCTGGAATTACTTTAGATCGAATACCTTTATTTGTAT      526
Minteng         CAACAGTAATTAATATACGATCAACTGGAATTACTCTAGATCGAATACCTTTATTTGTTT      523
Minduk          CAACAGTAATTAATATACGATCAACTGGAATTACTCTAGATCGTATGCCTTTATTTGTTT      538
                **********************  * ********* ******* ** *********** *

Sitaro          GATCAGTTGTTATTACTGCTTTATTACTAATATTATCTTTACCAGTATTAGCAGGAGCTA      586
Minsel          GATCAGTTGTTATTACTGCTTTATTACTTTTATTATCTTTACCAGTTATAGCAGGAGCTA      597
Bolmong         GATCAGTTGTTATTACTGCTTTATTACTATTATTATCTATACCAGTTATAGCAGGAGCTA      597
Minut           GATCAGTAGTTATTACAGCTTTATTATTACTATTATCTTTACCTGTATATGCCGGAGCTA      586
Minteng         GATCGGTAGTAATTACAGCTTTATTATTACTTTTATCTTTACCAGTATTGGCTGGAGCTA      583
Minduk          GATCGGTAGTAATTACAGCTTTATTATTACTTTTATCTTTACCAGTATTGGCTGGAGCTA      598
                **** ** ** ***** ********* *  * ****** **** **    ** *******

Sitaro          TTACTATATTATTAACAGATCGAAATTTAAATACTTCATTATATGACCCAGCTGGAGGAG      646
Minsel          TTACTATATTATTAACAGATCGAAATTTAAATACTTCATTTTTTGACCCAGCTGGAGGAG      657
Bolmong         TTACTATATTATTAACAGATCGAAATTTAAATACTTCATATATTGACCCAGCTGGAGGAG      657
Minut           TTACCATATTATTAACTGATCGAAATTTAAATACATCATTTTTTGACCCAGCTGGAGGGA      646
Minteng         TTACCATATTATTAACAGATCGAAATTTAAATACATCATTTTTTGACCCAGCTGGAGGGG      643
Minduk          TTACCATATTATTAACAGATCGAAATTTAAATACATCATTTTTTGACCCAGCTGGAGGGG      658
                **** *********** ***************** ****    ***************

Sitaro          GAGACCC-TATTCATTATCAACATTTATTTTGATATATAGGTCACCGGGAAG-------      697
Minsel          GAGACCC-TATTCTATATCAACATTTATTTTGATATAGTGGTCACCGGGAAGTTTAA--      713
Bolmong         GAGACCC-TATTCTATATCAACATATATTATGATATATTGGTCACCGGGAAGTATAA--      713
Minut           GGAGACCCAATTTTATACCAACATTTATTTTGATAGTATGGTCACC-------------      692
Minteng         GAGATC-CAATATTATACCAACATTTATAGTGATATATCGGTCACC-------------      688
Minduk          GAGATC-CAATAATATACCAACATTTATATAGATCTCTTGGTCACCTGGGAAGTTTAGA      716
                *    *   **    ** ****** ***   ***     *******
```

Figure 4: Alignment of COI gene sequences of North Sulawesi Fruit fly Figure 4. Alignment of *Drosophila* sp. COI gene sequences from North Sulawesi.

Characteristics of the amino acid sequences that make up the COI protein

The characteristics of the COI gene protein have a molecular weight between 58939.43 Daltons to 62173.10 Daltons. The highest molecular weight consensus sequence is the Mansell COI gene, while the lowest is the Mitra COI gene. Complete characteristics of COI amino acids from North Sulawesi fruit flies based on analysis with the BioEdit Program are shown in Appendix 2.

COI Protein Genetic Distance Matrix

The genetic distance of the amino acid protein COI *Drosophila* sp. from Sulawesi shows that Bolaang COI has the most significant genetic distance among the other five COIs. Then COI Minahasa Tenggara, COI Minahasa and COI South Minahasa. The closest genetic distance is directed by South Minahasa COI and Siau Tagulandang Biaro COI (Table 6).

	Minteng	Minduk	Minut	Sitaro	Minsel	Bolaang
Minteng	0.0000	0.0221	0.0936	0.2318	0.2267	1.2678
Minduk	0.0221	0.0000	0.1032	0.2451	0.2534	1.2586
Minut	0.0936	0.1032	0.0000	0.2206	0.2155	1.3104
Sitaro	0.2318	0.2451	0.2206	0.0000	0.0049	1.3327
Minsel	0.2267	0.2534	0.2155	0.0049	0.0000	1.3297
Bolaang	1.2678	1.2386	1.3104	1.3327	1.3297	0.0000

Table 6: Genetic Distance of COI amino acids *Drosophila* sp from North Sulawesi.

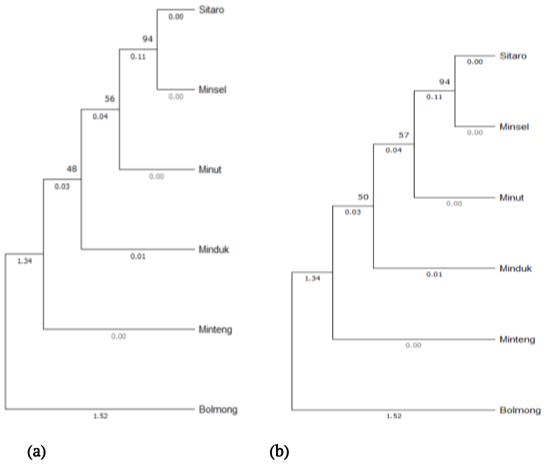


Figure 5: North Sulawesi Drosophila sp. phylogeny construction (a) Neighbor-Joining method, (b). Minimum Evolution; boosted 1000x, 2-Parameter Kimura model

Construction of Phylogeny

The North Sulawesi fruit fly COI gene phylogeny was constructed using the neighbor-joining method, the minimum evolution method, and the maximum composite likelihood substitution model. The phylogenetic tree topology constructed using two methods, the neighbor-joining and minimum evolution methods, was not significantly different. The phylogenetic construction showed *Drosophila* sp. from Bolaang as an outgroup. *Drosophila* sp. Sitaro, Minsel, Minut, Minduk, and Minteng were in one monophyletic group. *Drosophila* sp. sitaro forms a node with minsel. *Drosophila* sp. minut is in a monophyletic group with sitaro and minsel but at different nodes, so are *Drosophila* sp. minduk and minteng. Based on the phylogenetic tree formed, *Drosophila* sp. minsel has the closest phylogenetic relationship with *Drosophila* sp. Sitaro (Fig. 5).

Phylogenetic construction based on amino acids, COI minut was in the outgroup, while COI. COI Minteng, COI Minduk, COI Misel, and COI Sitaro formed a monophyletic group with different nodes. Minteng, Minduk, Minsel, and Sitaro were in the same node. COI gene DNA phylogeny construction differed from the construction based on amino acids (Fig. 6).

Figure 6: Phylogeny based on COI Neighbor-Joining/UPGMA amino acids using BioEdit.

Figure 7: Position of the Sulawesi *Drosophila* sp. based on the COI gene and NCBI BLAST results.

Discussion

The visualization of the amplicon with a stable band on the electrogram supports amplification of the *Drosophila sp.* COI gene from various regions in Sulawesi was successfully carried out using thoracic tissue by modifying proteinase K soaking time, which effectively produces high purity and concentration in insects [23, 24]. Nucleotides sorted by Sequencing produce Forward and Reverse sequences, which are well proven by the chromatograms of each type of nitrogenous bases that are not overlapping, visualized using the BioEdit and MEGA X programs. The consensus sequences are at a length of 685 to 716 bp, indicating the length of the COI gene sequence [25, 26, 27]. The amplification and sequencing results were confirmed as the COI gene. Consensus sequences are the most frequently occurring subsequences in a DNA sequence [28]. This sequence helps determine the location of the protein. Consensus sequences were determined using the BioEdit Program. Consensus sequences were verified for accuracy by BLAST on the NCBI website. The BLAST results showed that the most similar sequence was *Drosophila* sp. Alignment of the six sequences of fruit flies from North Sulawesi using the Clustal W method showed polymorphic sites. The most polymorphic sites were found in the Bolaang fruit fly compared to the COI sequences of other North Sulawesi fruit flies. The genetic range of COI amino acids in fruit flies from North Sulawesi is consistent with the genetic range of COI DNA sequences, where Drosophila sp. from Bolaang showed the farthest genetic distance among five other *Drosophila* sp. However, Drosophila sp. Sitaro and *Drosophila* sp. Minsel had the closest phylogenetic relationship based on amino acids encoded by the COI gene.

Phylogenetic reconstruction based on the sixth COI gene of *Drosophila* sp. from North Sulawesi was built using two methods, namely Neighbor-Joining and Minimum Evolution, to compare and ascertain the topography of the phylogenetic tree formed [29, 30]. The research results proved that the phylogenetic tree's topography was not significantly different between the two methods. This proves that using the COI gene accurately determines the position of the species *Drosophila* sp. Thus, the position and phylogenetic relationship of each *Drosophila* sp. from North Sulawesi based on the COI gene were ascertained. *Drosophila sp.* from Bolaang showed the highest variation in the COI gene among other sequences, so it occupies the outgroup on both phylogeny trees.

Interestingly, this study found that *Drosophila sp.* Sitaro and *Drosophila sp.* Minsel are in the same node. Geographically, Sitaro and Minsel are in different areas. Thus, it or both have the closest phylogenetic relationship compared to another *Drosophila sp.* in North Sulawesi. *Drosophila* sp. Sitaro is in the islands, while *Drosophila* sp. Minsel is on the mainland of North Sulawesi. This indicates that it is necessary to use other marker genes to analyze the phylogeny of *Drosophila* sp.

The phylogenetic construction based on the amino acids of the six COI genes of fruit flies from North Sulawesi is not much different from the phylogenetic construction based on DNA sequences. This shows the consistency of the phylogeny formed based on DNA and the amino acids encoded by the COI gene. Both phylogenetic constructs based on the COI gene DNA and amino acids are consistent with the results of genetic distance analysis.

Reconstruction of the fruit fly phylogeny from North Sulawesi, with 21 sequences from the BLAST results, showed variations in species similarity. *Drosophila* sp. Minteng forms a node with *D. atriplex* (OK037194.1);

thus, *Drosophila* sp. Minteng, based on the COI gene, has the closest resemblance to *D. atriplex*. *Drosophila sp.* minut and Sitaro from the same node. *Drosophila* sp. Minahasa forms a node with *D. melanogaster* [AB907180.1]. Based on the COI gene, *Drosophila sp.* Minahasa has the closest resemblance to D. melanogaster. *Drosophila* sp. Bolaang forms a node with *D. lavteicomis* [MK659823]. Thus, based on the COI gene, the Bolaang fruit fly has the closest resemblance to *D. lavteicomis* [MK659823.1]. At the same time, *Drosophila sp.* from Minsel forms a node with *D. pandora* [MK659835.1] and *D. parapallidosa* [MK659836.1] (Figure 7), which means that the COI gene of the Minsel fruit fly was similar, closest to the two species.

This study found that the intraspecific variation of the COI gene in fruit flies in North Sulawesi was high. Future research is needed using multi-gene barcodes to confirm the status of fruit fly species from North Sulawesi. This study demonstrates significant genetic divergence among *Drosophila* populations from different geographical locations in North Sulawesi, suggesting that the COI gene is effective at revealing intraspecific variation driven by geography. This is the first report on genetic variation and evolutionary relationship reconstruction based on the COI gene of *Drosophila* sp. from North Sulawesi. Although they have high morphological similarities, the genetic variation based on the COI gene of fruit flies from North Sulawesi demonstrates this point in this paper. The findings of this research reveal that a molecular approach based on the partial sequence of the COI gene supports morphological identification results in *Drosophila* sp. found in North Sulawesi. Future studies can use a morphological approach to strengthen the identification of *Drosophila* sp.

Acknowledgments

Thank you to the Directorate of Research, Technology and Community Service, Ministry of Education, Culture, Research and Technology, Republic of Indonesia, for funding this research through the Higher Education Excellence Applied Research Scheme (PTUPT), Universitas Negeri Manado, in 2022-2023 (Contract number: 2235/UN41/PP/2022). Thank you, leaders and staff, Bioactivity and Molecular Biology Laboratory, Universitas Negeri Manado.

Author Contributions

Herry Maurits Sumampouw: constructed the experiments, interpreted the results, and reviewed the article; Mokosuli Yermia Semuel: constructed the experiments, produced the report, statistically analyzed the data, generated illustrations, and wrote the article

Conflicts of interest

The authors declare no conflict of interest.

References

1. Kumar S, Swaminathan A, Abdel-Daim MM, Mohideen S. A systematic review on the effects of acrylamide and bisphenol A on the development of Drosophila melanogaster. Molecular Biology Reports, (2022); 49(11): 10703-10713.
2. Chiang MH, Ho SM, Wu HY, Lin YC, et al.. Drosophila model for studying gut microbiota in behaviors and neurodegenerative diseases. Biomedicines, (2022); 10(3): 596.
3. Ludington WB, Ja WW. Drosophila as a model for the gut microbiome. PLoS Pathogens, (2020); 16(4): e1008398.
4. Verheyen, EM. The power of Drosophila in modeling human disease mechanisms. Disease Models & Mechanisms, (2020); 15(3): dmm049549.
5. Nitta Y, Sugie A. Studies of neurodegenerative diseases using Drosophila and the development of novel approaches for their analysis. Fly, (2022); 16(1): 275-298.
6. Pal S, Oliver B, Przytycka TM. Stochastic Modeling of Gene Expression Evolution Uncovers Tissue-and Sex-Specific Properties of Expression Evolution in the Drosophila Genus. Journal of Computational Biology, (2023); 30(1): 21-40.
7. Yamaguchi M, Yoshida H. Drosophila as a model organism. Drosophila models for human diseases, (2018); 1-10.
8. Sato K, Tanaka R, Ishikawa Y, Yamamoto D. Behavioral evolution of Drosophila: unraveling the circuit basis. Genes, (2020); 11(2): 157.
9. Shahandeh MP, Turner TL. The complex genetic architecture of male mate choice evolution between Drosophila species. Heredity, (2020); 124(6): 737-750.
10. Hill T, Unckless RL. The dynamic evolution of Drosophila innubila Nudivirus. Infection, Genetics and Evolution, (2018); 57: 151-157.
11. Yuni LPEK, Yuda IP. The island biogeography of Wallacea and Krakatoa Island (pp. 217-229). (2020); Elsevier, Oxford.
12. Struebig MJ, Aninta SG, Beger M, Bani A, Barus H, Brace S. et al. Safeguarding imperiled biodiversity and evolutionary processes in the Wallacea center of endemism. BioScience, (2022); 72(11): 1118-1130.
13. Roman BE, Santana DJ, Prediger C, Madi-Ravazzi L. Phylogeny of Drosophila saltans group (Diptera: Drosophilidae) based on morphological and molecular evidence. Plos one, (2022); 17(4): e0266710.
14. Wang A, Du Z, Luo X, Zhang F, et al. The conserved mitochondrial genomes of Drosophila mercatorum (Diptera: Drosophilidae) with different reproductive modes and phylogenetic implications. International journal of biological macromolecules, (2019); 138, 912-918.
15. Andrews KJ, Bester R, Manrakhan A, Maree HJ. Utilisation of a mitochondrial intergenic region for species differentiation of fruit flies (Diptera: Tephritidae) in South Africa. BMC genomics, (2022); 23(1): 793.
16. Frazee SR, Harper AR, Afkhami M, Wood ML, McCrory JC, et al. Interspecific introgression reveals a role of male genital morphology during the evolution of reproductive isolation in Drosophila. Evolution, (2021); 75(5): 989-1002.
17. Yue L, Cao LJ, Chen JC, Gong YJ, Lin YH, et al. Low levels of genetic differentiation with isolation by geography and environment in populations of Drosophila melanogaster from across China. Heredity, (2021); 126(6); 942-954.
18. Machida WS, Tidon R, Klaczko J. Wing plastic response to temperature variation in two distantly related Neotropical Drosophila species (Diptera, Drosophilidae). Canadian Journal of Zoology, (2022); 100(2); 82-89.
19. Anholt, RR. Chemosensation and evolution of Drosophila host plant selection. IScience, (2020); 23(1): 100799.

20. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 1994; 3(5):294-299
21. Manuahe C, Mokosuli YS, Roring VIY. Optimization of DNA extraction and the position of mosquito Species from southeast minahasa in North sulawesi using NADH dehydrogenase Gene and Cytochrome oxidase Sub Unit 1 Gene. J Entomol Zool Stud 2016;4(4):498-508.
22. Tamura K, Stecher G, Kumar S. MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution (2021); https://doi.org/10.1093/molbev/msab120
23. Suddin S, Mokosuli YS, Marcelina W, Naharia O, Kapagang A. Molecular barcoding based 16S rRNA gene of Thermophilic bacteria from vulcanic sites, Linow Lake, Tomohon. In Materials Science Forum (2019) ; 967, pp. 83-92.
24. Kanan M, Salaki C, Mokosuli YS. Molecular Identification of Bacterial species from Musca domesfica L. and Chrysomya megachepala L. Luwuk City, Central Sulawesi, Indonesia. J Pure Appl Microbiol, (2020); 14(2): 1595-1607.
25. Xu M, Zhou S, Wan X. Phylogenetic implication of large intergenic spacers: insights from a Mitogenomic comparison of Prosopocoilus Stag Beetles (Coleoptera: Lucanidae). Animals, (2022); 12(13): 1595.
26. Naserzadeh Y, Mahmoudi N, Pakina E, Zargar M. Molecular identification and primer design for spotted wing drosophila (Drosophila suzukii). Research on Crops, (2020); 21(2): 364-369.
27. Madi-Ravazzi L, Segala LF, Roman BE, Alevi KCC, Prediger C, et al. Integrative taxonomy and a new species description in the sturtevanti subgroup of the Drosophila saltans group (Diptera: Drosophilidae). Zootaxa, (2021); 4980(2); 269292.
28. Bevers RP, Litovchenko M, Kapopoulou A, Braman VS, Robinson MR, et al. Mitochondrial haplotypes affect metabolic phenotypes in the Drosophila Genetic Reference Panel. Nature Metabolism, (2019); 1(12): 1226-1242.
29. Xu D, Sun M, Gao Z, Zhou Y, Wang Q, et al. Comparison and Phylogenetic Analysis of Mitochondrial Genomes of Talpidae Animals. Animals, (2023); 13(2): 186.
30. Chen W, Wang H. Phylogenetic Tree Selection by Testing Substitution Number in Clade. Diversity, (2022); 14(7): 543.