Genetic structure and phylogenetic relationships of *Phyllidiella pustulosa* species from Seribu Islands, North Sulawesi, Halmahera, and West Papua

N O Yonatika*, N Widiasih, M Hamidah, M D Nurhakim, H Budiarto, D M C Bintang, L M I Sani, D F Lestari, WA Setyaningsih, B Subhan and H Madduppa

Department of Marine Science and Technology, Faculty of Fisheries and Marine Science, IPB University, Dramaga, Bogor 16680, West Java, Indonesia

*E-mail: Niken_oktv@apps.ipb.ac.id

Abstract. *Phyllidiella pustulosa* are brightly coloured gastropod molluscs frequently found in coral reefs of the tropical Indo-Pacific. *Phyllidiella pustulosa* is widely distributed in Indonesia, such as Seribu Island, North Sulawesi, West Papua, and Halmahera. Based on the genetic characteristics of an individual's DNA sequence, differences between species can be identified. In this paper, we would like to provide the molecular analysis and phylogenetic relationship among nudibranchs from Indonesian waters. Identification was made by measuring the genetic distance between species. The phylogenetic tree reconstruction was made using the Kimura 2-parameter model with 1000 times bootstrap with neighbor-joining and maximum likelihood method. There is 46 DNA Sequence obtained from 4 different regions (Seribu Island, Halmahera, North Sulawesi, and West Papua). The genetic distance of West Papua and Halmahera has the smallest value among other populations, which is between 0.0051-1.4629, compared to the population in Halmahera. The phylogenetic tree also shows populations from West Papua and Halmahera are on the same lineage, indicating that the population in West Papua and Halmahera had the closest relation. The study suggested that North Sulawesi, Halmahera and West Papua have genetic mixing of the same region, which is distinctive from Seribu Island.

Keywords: DNA barcoding, evolutionary relationship, genetic characteristic, nudibranch, phylogenetic tree

1. Introduction

Phyllidiid nudibranchs are brightly coloured gastropods molluscs, frequently found in coral reefs of the tropical Indo-Pacific. *Phyllidiella pustulosa*, one of the most common nudibranchs in tropical to temperate water of the Indo-Pacific [1]. Based on molecular analyses, *Phyllidiella pustulosa*, described with a high intraspecific variation and cryptic speciation [2]. Characteristics for morphological identification include elongated, ovate body shape and pink clustered tubercles on a black dorsum [3]. They are well-known for their ability to ingest toxic nitrogenous sesquiterpenoids from their diets and use either these metabolites themselves or their biosynthetically transformed derivatives as a weapon for chemical defence [4]. *Phyllidiella pustulosa* are distributed in Indonesian waters, such as Seribu Island, North Sulawesi, West Papua, and Halmahera.

Nudibranch species are species that are vulnerable to predators because their shells are completely absent. This Mollusc class develops chemical compounds for protection from predators. Nudibranch has two methods of selecting bioactive molecules to either defend defensive chemicals or *de novo*
biosynthesis, they have an extraordinary collection of active compounds. This compound's benefits can be used for a variety of antibacterial, antifungal, anticancer, anti-inflammatory and antifeedant activities. Therefore, this mollusc class is a valuable model for studying the biological role of molluscs in the marine environment and their secondary metabolites. In addition, most of the approaches reported by many nudibranch molluscs are based on morphology and anatomy due to some difficulties in DNA isolation [5].

Based on genetic characteristics, every species is different from each other. This difference can occur in many ways. The determination of the species of sea slugs (Nudibranchs) is still mostly done by morphological observations, whereas in the same nudibranch species can alone have some distinct genetic diversity. The use of DNA as a feature of a species has several advantages, namely it is more thermostable than protein, more sensitive, not influenced by the environment and growth factors, and almost all tissues can be used as a source of genetic material [6]. DNA barcoding is a method that uses mitochondrial DNA with one strand of sufficient DNA to compare between animal species [7]. DNA barcodes can identify an organism down to the species level.

When measuring this difference between species, genetic distance can be identified as one of the critical criteria [8]. The phylogenetic tree or evolutionary tree is a branching diagram that shows the evolutionary relationship (phylogeny) between various biological species or other entities based on the similarities and differences in their physical or genetic characteristics. A phylogenetic tree, also called phylogeny, is a diagram describing the evolutionary lineage of genes of different species, organisms, or common ancestors. Phylogeny is very useful for organizing biodiversity knowledge, constructing taxonomic structures, and in-depth understanding of events that occurred during evolution [9]. Phylogenetic tree reconstruction analysis is based on isolated domains to understand the process of protein differentiation, primarily through partial repetitions (domains) or entire genes (proteins). This analysis helps to view individual proximity and protein functions in gene preparation [10]. In this paper, we would like to provide the molecular analysis and phylogenetic relationship among nudibranchs from Indonesian waters.

2. Materials and methods

2.1. Time and location
This research was performed using primary data from Marine Biosystem and Biodiversity Laboratorium Marine Science and Technology, Faculty of Fisheries and Marine Science, IPB University, and secondary data obtained from the BOLD System website. Primary data were collected from Nudibranchia samples in Indonesia waters, particularly from Seribu Island and secondary data from North Sulawesi, Halmahera Island, and West Papua. Each location/station provide a different number of samples, which are 7 samples in the Seribu Islands, 14 samples in North Sulawesi, 2 samples in Halmahera Island, and 15 samples in West Papua. Secondary data were taken online from April to May 2021.

2.2. Tools and materials
Tools and materials used in this research are DNA Sequence from the BOLD SYSTEM website, some software, i.e. Mega-X, DnaSP v6, Arlequin 3.5.2.2, Network 10.2, Notepad ++. Whilst, tools and materials used in tissue sampling were tissue from Phyllidiella pustulosa species, tubes containing 96% ethanol, and labels; DNA Extraction using 10% chelox solution made from a mixture of chelate-resin and H2O solutions, centrifuge machines, vortex mixers, and heat block; electrophoresis using 50 ml TBE buffer and 0.5 g agarose, a coloring agent, microwave, electrophoresis machine, ultraviolet transilluminator light; DNA amplification and sequencing using 3 µL DNA templates, 12.5 µL My taq 2x HS Red Mix, 1.25 µL forward and reverse primers, and 9 µL ddH2O, Thermocycler machine [11].
2.3. Sample collection
Collecting samples using the random sampling method is done by taking individual nudibranch found during the dive. Seven nudibranch samples were obtained from Seribu Island in Indonesia. Later, the sample was inserted into a 2 mL tube containing 96% ethanol and labelled on each tube for sample identification. Substrate identification is made visually by taking and feeling the texture then putting it on slate paper to be photographed. Following procedure to analyze nudibranch is through photos that have been taken and adjusted to the substrate type records. The entire samples are a collection of Marine Biosystems and Biodiversity Laboratorium, Marine Science and Technology, Faculty of Fisheries and Marine Science, IPB University.

2.4. Lab processing
2.4.1. DNA extraction
The extraction phase is done using two methods, chelex 10% and extraction kit Geneaid (blood and tissues). Extraction method using chelex 10% means a method used to store DNA in raisin, while extraction kit Geneaid (blood and tissues) method is defined as a method used to filter DNA using several kinds of buffer. Extraction with chelex 10% method is done by a 105℃ heating method for 60 minutes. This phase must be sterile to prevent contamination on the sample. In the extraction phase using kit, the first thing to do is take a little bit of the nudibranch stomach using a pinset, and then the sample is carved to make the cells easy to analyze. Stomach issues were chosen because this nudibranch moves using the stomach, so there are many muscles in it and many mitochondrial cells. After the sample was carved, the sample was inserted into a 1.5 mL tube, then add 200 μL GST Buffer and 20 μL Proteinase-K. The mixture is then incubated using a heating block at 60℃ for an hour or until the sample is lysis. After the lysis sample, centrifuge for 2 minutes at 14,000 rpm, move the supernatant to the new 1.5 mL tube. Add 200 μL GSB Buffer and vortex for 10 seconds. We added 200 μL of absolute ethanol, immediately vortex for 10 seconds. Put GD column in 2 ml Collection Tube, and move all the mixtures into GD column. Centrifuge for a minute at 14,000 rpm. After that, a 2 mL Collection Tube with all liquid inside was thrown away. GD column inserted in the new 2 mL Collection Tube. About 400 μLW1 buffer were added to the GD column then centrifuge for 30 seconds at 14,000 rpm. Throw The liquid in the Tube Collection discarded and put it back in the GD column. Add 600 μL Wash Buffer to GD column centrifuge for 30 seconds at 14,000 rpm. Throw the liquid on the Tube Collection and put it back. Centrifuge for 3 minutes at 14,000 rpm to dry the GD column. Transferred the dry GD column to a 1.5 mL tube. Add 100 μL Elution Buffer and 100 μL ddH₂O, then let it be for 3 minutes. Then centrifuge for 30 seconds at 14,000 rpm for pure DNA elixir.

2.4.2. DNA amplification
The PCR (Polymerase Chain Reaction) is a technique or method of multiplication (replication) of DNA enzymatically without using the organisms. With this technique, DNA could be produced in large numbers with a relatively short amount of time, making it easier to other techniques that use DNA. This research uses primer with a target locus of COI that are LCO: (5’ GGT CAA ATC AAG AAG TTG G 3’) and HCO: (5’ NO ACT TCA GGG TO CCA AAA AAT CA 3’). The PCR method that is used is the hotstart method using two master mixes, Kapa master mix and 2x Taq Master Mix. PCR process is done using the thermocycler, which is already programmed for 35 cycles. Every cycle consists of double-dip process (pre denatured) at 95℃ for 3 minutes, then denatured at 94℃ for 45 seconds and annealing at 45℃ for 45 seconds, and extension at 72℃ for 2 minutes. Later it is continued with the final elongated at 72℃ for 10 minutes.

2.4.3. Electrophoresis
Electrophoresis is a method for separating chemical compounds based on the rate of movement of molecules in an electric current [12]. This electrophoresis aims to determine the presence or absence of DNA in PCR products. Electrophoresis medium was 1% agarose (agarose 0.5 g and 50 mL TAE Buffer) with 4 L Ethidium Bromide (EtBr) as a dye. The next step is to mix 3 L of PCR results with 1 L of
loading dye, the mixture is then inserted into the agarose well. Electrophoresis using an electrophoresis machine with a voltage of 220 V and a current of 400 mA with a time of 25 minutes. The length of the DNA base strand can be measured by using a 4 μL Low mass ladder which is inserted into the first well at agarose. The results of electrophoresis in the form of bands can be seen by using a UV transilluminator.

2.4.4. DNA sequencing
The nucleotide Sequencing Cycle (DNA Sequencing) is a method for determining the sequence of nucleotides in DNA. The principle of sequencing using the PCR method as step it. The results are DNA fragments with varying lengths. The PCR products were sent to the UC Berkeley sequencing facility, Dept. of Molecular and Cell Biology Sequencing Facility, USA, using the ABI sequencer machine 1377. Lab processing steps can briefly be seen in Figure 1.

![Flowchart](image)

**Figure 1.** Sample collection and lab processing flowchart.

2.5. Data collecting method
A total of 31 DNA sequences were obtained from the BOLD SYSTEM website by accessing https://v3.boldsystems.org/. Meanwhile, 7 samples of the Seribu Islands were analyzed at the Laboratory of Marine Biodiversity and Biosystematics of IPB.

2.6. Data analysis
All DNA sequences were arranged and edited using MEGA 6.0 (Molecular Evolutionary Genetics Analysis) application. The results obtained from the nucleotide data were then matched with existing data on GenBank at NCBI (National Center for Biotechnology Information) using BLAST (Basic Local Alignment Search Tool) (http://blast.ncbi.nlm.nih.gov). After the BLAST results were obtained, then the process of making phylogeny trees was reconstructed using the neighbor-joining and Maximum Likelihood Tree methods using the bootstrap number 1000 with the Kimura-2 model parameter. This phylogeny tree was created using 45 Phyllidiella pustulosa sequences in addition to 12 sequences from GenBank. A bootstrap value equal to 95% will indicate that the branching is accurate, consistent, and will not change if other phylogenetic tree methods are used. If the bootstrap value is higher than 70%, then the branching is permanent or significant. Biodiversity analysis was observed and analyzed using DNAsp v6 application, which was analyzed using Arlequin 3.5, which included the values of haplotype
diversity ($H_d$) and nucleotide diversity ($\pi$) for each population. All sequences have been uploaded to the Barcode of Life Informasi System (BOLD). Figure 2 shows a flow chart of the data analysis process:

**Figure 2.** Data analysis flowchart.

3. Results and discussion

3.1. Genetic distance

Genetic distance was originally devised to estimate the degree of genetic differentiation between populations. Genetic distance is the extent of gene differences between populations or species that is measured by some numerical quantity [13].

Table 1 shows the genetic distances between populations of *Phyllidiella pustulosa* from the Seribu Islands, North Sulawesi, West Papua, and Halmahera. Figures in the table show the genetic distances of species populations between regions where the smaller the numbers indicate the closer the relationships between species are.

**Table 1.** The genetic distance of *Phyllidiella pustulosa* population in Indonesia.

|     | 1     | 2     | 3     | 4     |
|-----|-------|-------|-------|-------|
| 1   |       |       |       |       |
| 2   | 1.4629| 0.0107| 0.0108|
| 3   | 1.4456| 0.1170| 0.0051|
| 4   | 1.4583| 0.1047| 0.0388|

The population in West Papua has a very adjacency relation with the population in Halmahera with genetic distances of 0.0051-0.388. This matter could happen because the West Papua genetic material possibly came from Halmahera based on the geographic territory. There's also an adjacency relation
between the population in North Sulawesi with West Papua and Halmahera. Moreover, the result of genetic distance from the four regions populations of Phyllidiella pustulosa showed the value between 0.0051-1.4629. The genetic distance population in West Papua is between 0.0051-1.4629 compared to the population in Halmahera. This genetic distance of these two regions' population has the smallest value among other populations which indicates that the population in West Papua has the closest relation with the population in Halmahera, the smaller value of genetic distance between two populations means the closer relation between the two populations [14]. The close relation is also shown between the population in North Sulawesi with the population in West Papua and Halmahera in the range 0.0107-0.0108. This is suggested that the genetic material from the three regions was from the same stock population. Moreover, the three regions are located in eastern Indonesia with very complex conditions caused by the interaction of the three major world plates, The Pacific Ocean Plate, The Australian Indian Plate, and The Eurasian Continental Plate, known as the Triple Junction [15]. Geodynamic conditions are very active that cause faults and subduction, so the gene flow can possibly happen [14]. In contrast to the Seribu Islands, the genetic distance population relation is the most distant. This is because the Seribu Islands are far from the other three regions. Seribu Islands are in the western part of Indonesia to the north of Java Island. Other locations are in the eastern part of Indonesia and separated by other large islands. Genetic diversity is closely related to geographical conditions or positively correlated with geographical distance. Geographical locations that are relatively close to support the genetic locations of the population will have locations and are in one group and vice versa [16].

3.2. Evolutionary relationships of taxa
Genetic characteristics based on ancestry and evolution can be identified through the reconstruction of the phylogenetic tree. Phylogenetic tree reconstruction was carried out using two methods, namely neighbor-joining and maximum likelihood. Bootstrap values (expressed as percentages of 1,000 replications) are shown at major branching points, Himantuta leoparda JX2634 was an out-group. Figure 3 shows a phylogeny tree made using the neighbor-joining method.

The closest species based on the phylogeny tree are species in North Sulawesi with West Papua, North Sulawesi with Halmahera, and West Papua with Halmahera. The farthest related species are the Seribu Islands. Diagram for Evolutionary Network is used to infer the phylogenetic relationships among the species or genes, including morphological, biological, and bionomic characters, allozyme, RFLP data have been extensively used to infer the evolutionary relationship among the species during the pre-genomic era [9]. The evolutionary history was inferred using the Neighbor-Joining method [16]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [8]. The evolutionary distances were computed using the Kimura 2-parameter method [17] and are in the units of the number of base substitutions per site. This analysis involved 46 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 620 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. The Neighbor-Join method is very similar to The Fitch and Margoliash methods, except that the differences algorithm determines the sequence for pairs. The neighbor-join method is very suitable when the mean evolution of the lineage splits is under different considerations. When long branches of a tree of known topology change through stimulating levels from evolutionary change, neighbor method joins are the most suitable for predicting trees properly [18].

The relationship of Phyllidiella pustulosa from various regions can be seen from the reconstruction of the phylogeny tree in Figure 4. The results of data analysis showed that samples of nudibranch (Phyllidiella pustulosa) were united in the same clade, indicating that all samples were of the same species. The species of sea rabbit from West Papua and Halmahera are on the same lineage; it shows that the species from both areas share a more recent common ancestor with each other. This follows the results of genetic distance analysis which shows that the population in West Papua has a really close relation with the population in Halmahera. Some populations from the North Sulawesi region are also closely related to the species population from the West Papua region, while the populations from the
Seribu Islands region are on a different lineage from other regions. The most common reason for the genetic divergence is genetic isolation [19]. Seribu Island is located far from the other three regions, so there is little chance of genetic mixing. There is no gene flow between Seribu Island and the other island. Subsequently, populations on the island diverge from those of the same species on the mainland and other islands.

Figure 3. Phylogenetic tree using the neighbor-joining method.
Reconstruction of the phylogeny tree using the maximum likelihood method is particularly accurate for establishing molecular phylogeny. The maximum likelihood procedure can more frequently recover the correct tree from the simulated data set than other methods. Another important advantage of the maximum likelihood method is comparing different trees and evolutionary models within a statistical framework [20]. The evolutionary history was inferred using the Maximum Likelihood method and the Kimura 2-parameter model [17]. The tree with the highest log likelihood (-5003.40) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. This analysis involved 46 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 620 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [21].

3.3. Haplotype connectivity
The haplotype network construction shown in Figure 5 shows 38 haplotypes, from 46 sequences of *Phyllidiella pustulosa* found in 4 different locations. A haplotype is defined as the combination of alleles for different polymorphisms that occur on the same chromosome [22]. Haplotype network construction is a widely used approach for analyzing and visualizing the relationships among DNA sequences within
a population or species [23]. From this haplotype connectivity, the shape of the haplotype connection of *Phyllidiella pustulosa* found in North Sulawesi, Seribu Island, West Papua, and Halmahera.

Based on Figure 5, the highest haplotype frequency (H) from the population of *Phyllidiella pustulosa* was H_32 (8.89%) found in North Sulawesi, followed by H_9 (6.67%) that was also found in North Sulawesi. Based on the data analysis it was found that the haplotype diversity (hd) value was 0.9879, this score shows that the haplotype diversity was very high since according to [24] the haplotype diversity value which is considered to be high was ranged from 0.5 to 1. The high value of haplotype diversity indicates that the locations where samples were taken from were far apart and that the genetic diversity of *Phyllidiella pustulosa* is high, the higher the haplotype diversity value means the more various haplotypes from the two areas hence the level of genetic diversity will be higher and *vice versa* [24].

![Figure 5. Haplotype connectivity of *Phyllidiella pustulosa* population in Indonesia.](image)

3.4. Structure and genetic diversity

The value of genetic diversity provides information for adapting to environmental and climate change and disease [6]. High genetic diversity in populations is based on the assumption that the rate of evolution in a changing environment is limited by genetic variation [18]. Genetic diversity is important for species’ survival because species that have modest genetic diversity may be more susceptible to disease or the effects of environmental changes. The increase of genetic diversity is implicated in the offspring production with various characteristics that can survive in various environmental conditions and prevent the possibility of gene damage in the population [25].

*Phyllidiella pustulosa* species found in four research locations had different genetic diversity values based on Table 2. *Phyllidiella pustulosa* in the Seribu Islands has a genetic diversity of 0.9643 +/- 0.0772, North Sulawesi has a genetic diversity value of 0.9474 +/- 0.0379, West Papua has a genetic diversity value of 0.9917 +/- 0.0254, and Halmahera has a genetic diversity value of 0.001745 +/- 0.002468. Based on the results of the data analysis, it shows that the high category of genetic diversity is in the regions of North Sulawesi and West Papua. The Seribu Islands region has moderate genetic diversity and Halmahera Island has the lowest diversity value compared to other regions. [6] States that high genetic diversity reflects a large population size, while a decrease in population size will reduce
Low genetic diversity values indicate a small population size and will reduce genetic diversity [26].

Table 2. Genetic diversity of *Phyllidiella pustulosa* based on sample size, haplotype number, genetic diversity, and nucleotide diversity from each of the population sites in Indonesia.

| No | Location      | N  | Nucleotide Diversity | Genetic Diversity | Haplotype Number |
|----|---------------|----|----------------------|-------------------|------------------|
| 1  | Kep. Seribu   | 8  | 0.109948 +/- 0.060663 | 0.9643 +/- 0.0772 | 7                |
| 2  | Sulawesi Utara| 19 | 0.089975 +/- 0.045528 | 0.9474 +/- 0.0379 | 14               |
| 3  | Papua Barat   | 16 | 0.059002 +/- 0.030419 | 0.9917 +/- 0.0254 | 15               |
| 4  | Halmahera     | 2  | 1.0000 +/- 0.5000     | 0.001745 +/- 0.002468 | 2                |

FST: 0.76497

The Function Sequence Table (FST) value is a value that indicates the degree of inbreeding that occurs in the total population. Hartl & Clarke divided the FST value into 4 levels of genetic difference, namely low (<0.05), moderate (0.05-0.15), high (0.15-0.25) and very high (>0.25). Based on these criteria, the FST value of this study is 0.76497, so it can be categorized as having a very high genetic difference. Furthermore, the p-value at 0.00 also shows that there are significant differences in the *Phyllidiella pustulosa* species in the Seribu Islands, North Sulawesi, West Papua and Halmahera. The p-value of the FST also showed that the population of the *Phyllidiella pustulosa* species was significantly different (P value <0.05). This indicates that the genetic structure among the observed population is very high due to the relatively low gene flow [20].

4. Conclusion
The purpose of this paper was to provide phylogenetic analysis based on the molecular features (genetic distance, phylogenetic tree, haplotype connectivity, structure and genetic diversity) of nudibranchs from Indonesian waters. Based on the analysis conveyed, it can be concluded that multiple factors such as geographical distance and population size influence the phylogenetic relationship of nudibranchs in Indonesian waters. Greater distance will lead to greater genetic distance, while greater population size will lead to greater genetic diversity. Future analysis and more DNA sequence data could be useful, so this paper could provide more accurate and credible results.

References
[1] Bogdanov A, Papu A, Kehraus S, Cruesemann M, Wagele H and Konig G M 2020 Metabolome of *Phyllidiela pustulosa* species complex (Nudibranchia, Heterobranchia, Gastropoda) reveals rara dichloroimidic sesquieterpene derivatives from a Phylogenetically distinct and undescribed clade *J. Nat. Prod.* 83 1-9
[2] Stoffels B E M W, Meij S V D, Hoeksema B W, Alphen J V, Alen T V, Munoz M A M, Voogd N J, Tuti Y and Velde GVD 2016 Phylogenetic relationships within the Phyllidiidae (Opisthobranchia, Nudibranchia) *Zookeys* 605 1-35
[3] Brunchhorst D J 1993 The systematic and phylogeny of Phyllidiid nudibranchs (Doridoidea) *Rec Aust Mus, Suppl* 16 1-107
[4] Wu Q, Chen W T, Li S W, Ye J Y, Huan X J, Gavagnin M, Yao L G, Wang H, Miao Z H, II X W and Guo Y W 2019 Cytotoxic nitrogenous terpenoids from two south China Sea nudibranchs *Phyllidiella pustulosa, Phyllidia coelestis,* and their sponge-prey *Acanthella cavernosa* Mar. *Drugs* 17 1-12
[5] Alqudah A, Saad S B, Hadry F N and Susanti D 2015 Identification and phylogenetic inference in different molluscs nudibranch species via mitochondrial 16S rDNA *Braz. J. Biol. Sci.* 2 295-302
