



OPEN

Molecular phylogeny of selected dorid nudibranchs based on complete mitochondrial genome

Thinh Dinh Do^{1,2}, Dae-Wui Jung¹ & Chang-Bae Kim¹✉

Dorid nudibranchs are a large group of mollusks with approximately 2,000 recorded species. Although agreement exists on the monophyletic nature of the dorid nudibranch group, the interfamily relationships of the suborder are subject to debate. Despite efforts to elucidate this issue using short molecular markers, the conclusiveness of the findings has been hindered by branching polytomy. Mitogenomes are known to be effective markers for use in phylogenetic investigations. In this study, eight mitogenomes of dorid nudibranchs were decoded and analyzed. Gene content and structure showed little change among species, reflecting the conserved mitogenomes of dorid nudibranchs. For most genes, the direction was typical for nudibranchs; nevertheless, tRNA^{Cys} had an inverse direction in Cadlinidae species. Phylogenetic trees based on nucleotide and amino acid datasets revealed a relatively consistent pattern of interfamily relationships with little difference for positions of Phyllidiidae and Cadlinidae. Species of Cadlinidae were clustered together and did not form a clade with Chromodoridae. Additionally, Goniodorididae was sister to Aegiridae, whereas Discodoridae was sister to Dorididae. This finding was supported by tree topology test based on mitogenome data. The results of the present study indicate that complete mitogenomes are promising markers for investigating interfamily relationships among dorid nudibranchs.

Doridina (~2,000 species) and its sister group Cladobranchia (~1,000 species) are two suborders of mollusk nudibranchs^{1,2}. The dorid nudibranchs are a diverse group of marine mollusks found worldwide that play an important role in the marine ecosystem. Dorid species are carnivorous; they feed mainly on sedentary invertebrates such as sponges, cnidarians, tunicates, and bryozoans³. To deter their own predators, many dorid species synthesize unpleasant or toxic compounds from their foods³, this ability makes dorid nudibranchs potentially interesting subjects in the search for chemical compounds with pharmaceutical relevance⁴. Thus, several species have been used in pharmaceutical science and developmental studies. *Cadlina luteomarginata* has been of particular interest in biochemical investigations, whereas other nudibranch species, such as *Aldisa andersoni*, *Aldisa cooperi*, *Cadlina pellucida*, *Cadlina laevis*, *Doriprismatica atromarginata*, and *Jorunna funebris*, have become important subjects in bioactive substance studies³⁻⁷. However, the first step in the practical application of dorid nudibranch compounds is the elucidation of the group's taxonomy and phylogenetic relationships.

Despite their importance in marine ecology and pharmaceutical studies, the interfamily relationships of dorid nudibranchs have long been disputed⁸. Previously, only the morphological characteristics of dorid families, such as the rhinophores, mantle, gill, gill cavity, and radula, were used for classification; however, molecular markers are now used to study dorid nudibranch phylogeny^{9,10}. In such studies, a single marker or combination of several short markers is usually used. Although some studies have been conducted to determine relationships within a genus or family, only a few studies have dealt with the higher-level groups in Doridina and Cladobranchia^{2,11}. Moreover, the application of short markers has been difficult to elucidate the phylogenies containing these higher-level groups. To date, few attempts have been made to study the interfamily relationships of dorid nudibranchs using cladistic methods. Notable research has been published by Hallas et al.¹¹ and Korshunova et al.⁸ related to the families within the suborder Doridina. Nevertheless, interfamily relationships remain poorly understood because of conflicting phylogenies, tree polytomy, and inadequate sampling¹¹. For example, controversy surrounds the relationships between Discodoridae + Dorididae and Goniodorididae + Aegiridae. Conventionally,

¹Department of Biotechnology, Sangmyung University, Seoul 03016, South Korea. ²Institute of Marine Environment and Resources, Vietnam Academy of Science and Technology, Haiphong 04000, Vietnam. ✉email: evodevo@smu.ac.kr

Discodoridae was considered to have a close relationship with Dorididae, whereas Goniodorididae was believed to have a close relationship with Aegiridae¹². Nevertheless, recent molecular analyses have shown that Discodoridae is a sister group to Goniodorididae, whereas Aegiridae is a sister group to Dorididae⁸ or has an unstable position (depending on the analysis method used)¹¹. To improve the systematics of dorid nudibranchs, phylogenetic relationships must be explicitly determined, and the effective use of DNA sequences to elucidate phylogenies is one potential strategy. Another issue is the phylogenetic classification of Cadlinidae, which has long been controversial. Traditionally, Cadlinidae has been considered a member of Chromodorididae; however, recent taxonomic evaluation indicated that Cadlinidae is an independent family that is separate from Chromodorididae^{8,13}.

The mitochondrial genome is a powerful molecular marker used to explore phylogenetic relationships, and it has been applied to reveal the molecular evolution of mollusks^{14–16}. The typical mitogenome of mollusks contains 13 protein-coding genes, 22 transfer RNA (tRNA) genes, and two ribosomal RNA (rRNA) genes¹⁴. Given their importance in systematics and phylogenetic reconstruction, the mitogenomes of nudibranchs are now being characterized; however, too few of the mitogenomes from more than 2,000 dorid nudibranchs have been sequenced and analyzed. Molecular phylogenetic analyses and taxon-sampling schemes are known as effective tools in phylogenetic research¹³. Previously, the phylogenetic position of nudibranchs has been studied on the basis of partial *cox1*, 16S rRNA, 18S rRNA, and 28S rRNA sequences. Nevertheless, the complete mitogenome provides better phylogenetic resolution and accuracy than single genetic markers¹⁷.

This study aimed to analyze the mitogenome structure of dorid nudibranchs and use mitogenomes as molecular markers to investigate the interfamily relationships of this group. To achieve this aim, the mitogenomes of different dorid nudibranchs were decoded and analyzed. The structure of dorid nudibranch mitogenomes was examined and compared with the mitogenome sequences already available in public databases. Additionally, phylogenetic trees showing interfamily relationships were determined on the basis of the examined dorid nudibranch mitogenomes.

Results

General mitogenome features. Eight complete mitogenomes of dorid nudibranchs were sequenced in the present study, including those of *Aldisa cooperi*, *Cadlina japonica*, *Cadlina koreana*, *Cadlina umiushi*, *Carminodoris armata*, *Doris odhneri*, *Triopha modesta*, and *Verconia nivalis*. Mitogenome lengths ranged from 14,397 bp (*T. modesta*) to 14,982 bp (*C. japonica*) (Tables S1–8; Figs. S1–8). All eight mitogenomes had negative AT skew values (from -0.167 in *V. nivalis* to -0.089 in *D. odhneri*) and positive GC skew values (from 0.008 in *D. odhneri* to 0.152 in *Cadlina umiushi*), suggesting a bias for T and G nucleotides (Table S9). Generally, mitogenomes contained 13 protein-coding genes, two rRNA genes, and 22 tRNA genes. The mitogenomes of most species comprised 37 genes; however, that of *C. japonica* contained 38 genes due to a duplication of tRNA^{Leu}. The mitogenome of each species contained 13 protein-coding genes (PCGs), including nine genes (*cox1*, *cox2*, *cytb*, *nd1*, *nd2*, *nd4*, *nd4l*, *nd5*, and *nd6*) encoded by the H-strand and four genes (*atp6*, *atp8*, *cox3*, and *nd3*) encoded by the L-strand (Tables S1–8). In terms of start and stop codons, ATN was the most frequent initiation codon, ATG was most commonly used, and ATA, GTG, and TTG were also used for initiation; TAA was the most common termination codon, with TAG and incomplete T– were also used for the termination of several genes. Codon usage and relative synonymous codon usage (RSCU) are indicated in Table S10. For all mitogenomes, the amino acids most frequently found in PCGs were leucine followed by serine; by contrast, glutamine, arginine and cysteine were the least common amino acids. The RSCU values of 13 PCGs in the eight examined mitogenomes showed a bias toward amino acids encoded by codons rich in A and T, such as UUA-Leu, AUU-Ile, UUU-Phe, and AUA-Met (Table S10).

There were 22 tRNA genes in most dorid species, except the *C. japonica* mitogenome carried 23 genes due to a tRNA^{Leu} duplication. Similar to other nudibranchs, different anticodons were observed for tRNA^{Leu} and tRNA^{Ser}. Two rRNA genes were detected in the mitogenomes of dorid nudibranchs. The large (16S rRNA) and small (12S rRNA) rRNAs were encoded by the H-strand and L-strand, respectively. Overall, the intergenic regions in the eight dorid nudibranch mitogenomes were short in length. In the present study, the sizes of the intergenic regions varied according to species. The longest noncoding region was located between tRNA^{His} and tRNA^{Cys} in Cadlinidae species (324 bp in *C. japonica*). The overlapping regions were also short and variable among species; nevertheless, the longest overlapping region was always located between *nd5* and *nd1* genes.

The gene contents and order were similar to those in typically structured Nudibranchia mitogenomes. Gene direction was similar across the examined mitogenomes, although the direction of tRNA^{Cys} was inverse in species from Cadlinidae, i.e., *A. cooperi*, *C. japonica*, *C. koreana*, and *C. umiushi*, compared with its direction in other nudibranchs (Fig. 1). In the four Cadlinidae species, tRNA^{Cys} was encoded by the L-strand; in the other nudibranchs, it was encoded by the H-strand. Overall, gene order within Doridina mitogenomes is pretty conservative and identical to that of arrangement pattern in Nudibranchia. An exception was observed in the mitogenomes of *Hypselodoris* which present translocation of the second tRNA^{Ser} (GCU) and *nd4*^{18,19}. In most nudibranch, the second tRNA^{Ser} and *nd4* are located between the first tRNA^{Ser} and tRNA^{Thr}. However, in three recorded *Hypselodoris* species, this block is translocated to the position between tRNA^{Cys} and tRNA^{Gln} (Fig. 1).

Phylogeny of dorid nudibranchs based on complete mitogenome sequences. Phylogenetic trees of dorid nudibranchs were constructed with and without Gblocks based on nucleotide and amino acid datasets of mitogenome sequences. For each dataset, two tree construction methods, Bayesian inference (BI) and Maximum likelihood (ML), were used. Because the BI and ML analyses showed similar phylogenetic topologies for all datasets, the trees were combined to show the interfamily relationships of the dorid nudibranchs (Fig. 2; Figs. S9–15). The use of three outgroup species revealed little change in tree reliability compared to two outgroup species (Fig. S16–S19). As shown in phylogenetic trees, the interfamily relationships were well resolved. In gen-

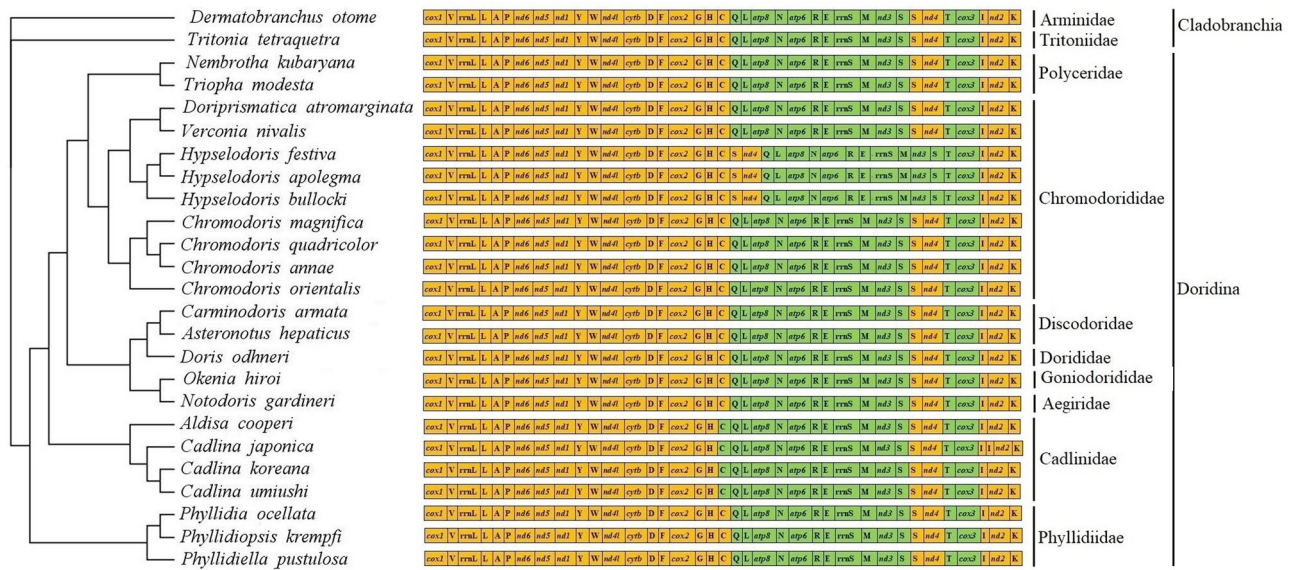


Figure 1. Linearized mitochondrial gene arrangement patterns of the suborders Doridina and Cladobranchia. Species relationships were based on phylogenetic analyses presented in Fig. 2. Yellow and green represent genes encoded on the H-strand and L-strand, respectively. Single-letter abbreviation of the amino acid code represents tRNAs.

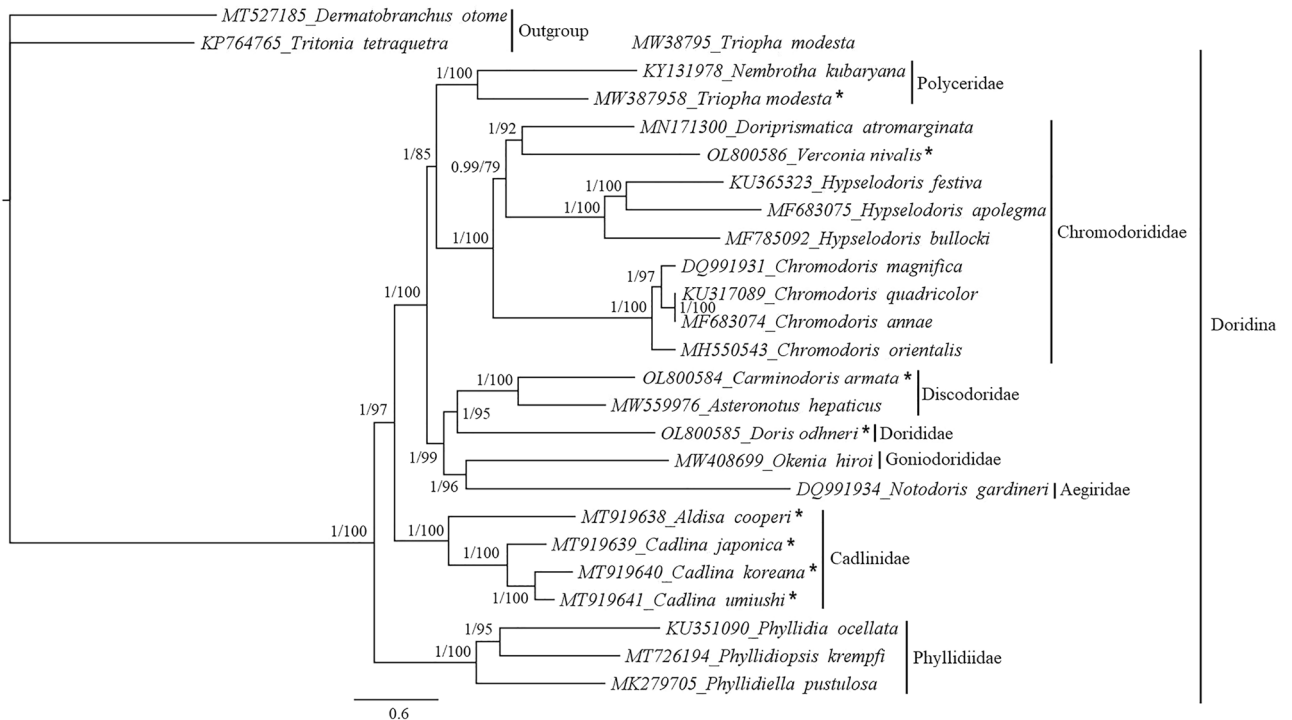


Figure 2. Phylogenetic tree showing the interfamily relationships of dorid nudibranchs based on the nucleotide sequences of 12 PCGs + 2 rRNAs + 22 tRNAs from mitogenomes (*nd4l* excluded). Sequences generated in this study are marked with stars. GenBank accession numbers are indicated next to species names. Gblocks was used after sequence alignment. Posterior possibility values (left) and ultrafast bootstrap values (right) are shown at the nodes. Species of the suborder Cladobranchia were used as outgroup.

eral, phylogenetic trees indicated high posterior priority (PP) values, while ultrafast bootstrap (UFBoot) values were variable among datasets. High credibility was observed in tree generated based on the nucleotide sequences of 12 PCGs + 2 rRNAs + 22 tRNAs with Gblocks. BI tree of this dataset showed PP ≥ 0.99, meaning that polytomy is not observed if threshold < 0.99. For interfamily relationship in ML tree, the lowest support value was found

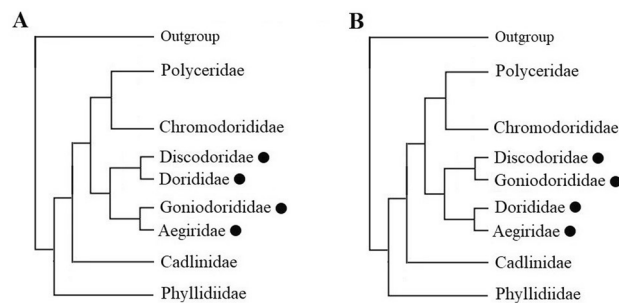


Figure 3. Interfamily relationship from this study (A) and constrained branching pattern (B) used for topology test. In the constrained tree, Discodoridae is sister to Goniodorididae and Aegiridae is sister to Dorididae. The families targeted for topology test were marked with black circles.

between Polyceridae and Chromodorididae (UFBoot = 85). Polytoamy of this node occurs if threshold is set as 85, while UFboot values for other family relationships show strong supports (Fig. 2).

All examined families were recovered as monophyly in the phylogenetic trees generated in this study. Moreover, the branching patterns for interfamily relationships in the trees were similar except trees generated from 1st and 2nd codon dataset. In most trees, three species of Phyllidiidae were branched early, while in 1st and 2nd codon tree, the members of Cadlinidae were branched early. Expectedly, three *Cadlina* species (*C. japonica*, *C. koreana*, and *C. umiushi*) were clustered together with absolute support values for both PP and UFBoot in all analyzed trees (PP = 1; UFBoot = 100). The clade containing these three *Cadlina* species was a sister to *A. cooperi* (PP = 1; UFBoot = 100). The phylogenetic trees also showed that Goniodorididae was a sister group to Aegiridae, whereas Discodoridae was a sister group to Dorididae. Support values for these relationships were variable among analyses. In a tree based on the nucleotide sequences of 12 PCGs + 2 rRNAs + 22 tRNAs with Gblocks, support values were as follows: Discodoridae + Dorididae: PP = 1 and UFBoot = 95; Aegiridae + Goniodorididae: PP = 1 and UFBoot = 96 (Fig. 2). Additionally, these two clades were clustered together with high support values (PP = 1; UFBoot = 99). The amino acid tree showed a similar pattern of interfamily relationships but lower support values relative to those of the nucleotide tree. In a tree based on 12 amino acid sequences with Gblocks, support values were as follows: Discodoridae + Dorididae: PP = 0.99 and UFBoot = 73; Aegiridae + Goniodorididae: PP = 0.98 and UFBoot = 65 (Fig. S14). The clade containing these four families was a sister to a clade that included Chromodorididae and Polyceridae. Although the relationships among the four Chromodorididae genera were slightly variable among datasets, this family was always a sister to Polyceridae in all generated trees.

A tree topology test was performed to investigate the relationships among Discodoridae, Dorididae, Goniodorididae, and Aegiridae. The tree topology and constrained tree pattern are shown in Fig. 3. The topology of the tree from the present study was confirmed as the most likely dorid nudibranch phylogeny ($p = 0.974$, i.e., $p > 0.05$), whereas the topology of the constrained tree was rejected ($p = 0.026$, i.e., $p < 0.05$). Therefore, the relationships of Discodoridae + Dorididae and Aegiridae + Goniodorididae were statistically supported, whereas those of Aegiridae + Dorididae and Goniodorididae + Discodoridae were rejected.

Discussion

Despite the rich diversity of dorid nudibranchs, a limited number of their mitogenomes are recorded in public databases; this hinders the use of mitogenomes in investigations of dorid nudibranch evolution and phylogeny. In the present study, mitogenomes from different families of dorid nudibranchs were sequenced and characterized. Generally, these mitogenomes were small (14,397–14,982 bp); thus, they were similar in size to those of other gastropods. The small mitogenome size is attributable to the low number of noncoding regions, the overlap of genes, and the reduced size of genes in all sequences¹⁴. Base skewness was also consistent among the studied species (negative A-T skew and positive G-C skew), and gene arrangement was similar to that recorded in other nudibranch mitogenomes. Among the PCGs in the mitogenomes, codons were typically rich in A and T. Generally, the gene arrangement of nudibranchs is conserved with little variation. Indeed, a change in gene arrangement has only been observed in *Hyselodoris*^{18,19}. Accordingly, tRNA^{Ser} and *nd4* were located between tRNA^{Cys} and tRNA^{Gln}. The prominent difference observed was the inverse direction of tRNA^{Cys} in all four Cadlinidae species. This characteristic was not found in the other nudibranchs studied here, but it has been observed in *Berthella* sp. from Pleurobranchida, which is the sister order of Nudibranchia²⁰. This change might have specifically occurred during evolution. Except for the inverse direction of tRNA^{Cys} in Cadlinidae, the gene direction in the studied dorid species was typical of other nudibranchs. We also identified the duplication of tRNA^{Leu} in *C. japonica*, which is the first time a gene duplication has been observed in a dorid nudibranch.

Huge efforts have been made to investigate interfamily relationships of dorid nudibranchs^{8–11}. Recently, molecular markers¹¹ or a combination of molecular markers and morphological characteristics⁸ were applied for this purpose. Although short markers, such as partial *cox1* and 16S rRNA genes, have previously been used to build phylogenetic trees for dorid nudibranchs, satisfactory conclusions could not be drawn regarding interfamily relationships due to polytoamy and instability of family branching^{8,11}. In the present study, we attempted to address this issue using mitogenomes as markers. Despite the small number of mitogenome sequences used in this study relative to the total number of dorid nudibranch families, our findings are valuable, and promising.

First, the branching of trees based on different datasets was pretty clear and consistent. Comparing tree topology among datasets, little difference in branching pattern was observed when first and second codons of PCGs were used. The trees based on this dataset showed Cadlinidae early branched instead of Phyllidiidae (Fig. S12–13). However, PP and UFBoot values between Cadlinidae and Phyllidiidae with inside branch were not high. This may be caused by the lack of samples for the whole suborder Doridina. Therefore, increase in sample coverage for different families is necessary to gain an insight into phylogeny of dorid nudibranch.

Through phylogenetic analyses, significant interfamily relationships were detected for the suborder Doridina. Congruent with previous studies, our results revealed the existence of a cluster containing Polyceridae and Chromodorididae^{8,11}. Importantly, our phylogenetic trees showed that Discodoridae was a sister group to Dorididae and that these two families were a sister group to a clade containing Goniodorididae and Aegiridae. The identified relationships among these families are not consistent with those reported in previous studies^{8,11}. However, our findings are congruent with morphological evidence because Discodoridae and Dorididae possess characteristics of cryptobranchs, whereas Goniodorididae and Aegiridae possess characteristics of phanerobranchs^{8,10}. Additionally, the (Aegiridae + Goniodorididae) + (Discodoridae + Dorididae) was in agreement with conventional classification. It is widely accepted that dorid nudibranchs can be divided into two groups: cryptobranchia (with a gill cavity, e.g., Dorididae and Discodoridae) and phanerobranchia (without a gill cavity, e.g., Aegiridae and Goniodorididae)⁸. Moreover, our topology test based on mitogenome sequences significantly supported this pattern and rejected the sister relationship between Discodoridae and Goniodorididae as well as that between Aegiridae and Dorididae.

Additionally, our study supported the separation of Cadlinidae from Chromodorididae. Historically, the classification of Chromodorididae was based on morphological similarities, primarily radular and reproductive morphology, and *Cadlina* was thought to be closely related to members of Chromodorididae such as *Chromodoris* and *Hypselodoris*¹³. However, besides shared characteristics, some of the characteristics possessed by *Cadlina* are not present in Chromodorididae. For example, contrary to chromodorids, *Cadlina* possesses penial spines, spicules in the mantle tissues, and tubercles on the mantle surface^{8,13}. Similar to *Cadlina*, the taxonomic position of *Aldisa* has been disputed. Nevertheless, the denticulate teeth of the radular suggest that both genera are associated with chromodorids^{8,13}. By contrast, some characteristics are shared by *Cadlina* and *Aldisa*, e.g., the tuberculate mantle and differentiated stomach, whereas these characteristics are not found in chromodorids¹³. Besides morphology, the geographical distribution differs between Cadlinidae and Chromodorididae: Chromodorididae is commonly found in tropical and subtropical waters, whereas Cadlinidae is distributed in temperate and cold waters⁸. Unlike the disputed morphological characteristics, molecular evidence suggests that Cadlinidae is distant from Chromodorididae^{13,21}. Besides the morphological evidence, a phylogeny based on *cox1* and 16S rRNA markers indicated that Cadlinidae was a distinct family¹³. Following a study by Johnson¹³, additional molecular evidence has been accumulated and analyzed. Koroshunova et al.⁸ recently investigated the relationships among dorid nudibranchs, including Cadlinidae and Chromodorididae, by concatenating *cox1* + 16S rRNA + 18S rRNA + 28S rRNA sequences; they also showed the distinct separation of Cadlinidae. Hence, consistent with previous reports, our phylogenetic trees based on mitogenome data confirmed that Cadlinidae is a distinct family.

Conclusion

In this study, eight new mitogenomes of dorid nudibranchs were characterized and compared with previously recorded mitogenomes from this suborder. Little change in gene content and structure revealed the conserved mitogenome of dorid nudibranchs. Variation in gene direction was only observed in Cadlinidae with the inversion of tRNA^{Cys}. We have provided the most comprehensive phylogeny of dorid nudibranchs to date based on mitogenomes. From analyses of nucleotide and amino acid datasets, we revealed a pretty consistent pattern of branching among interfamily relationships. Well branched phylogeny revealed that complete mitogenomes are promising markers for investigating the phylogenies of dorid nudibranchs. Despite our promising results, we were unable to cover all families of dorid nudibranchs. To better understand the overall phylogeny of this group, additional mitogenome sequences from different families should be sequenced and analyzed. Moreover, nuclear genes and transcriptomic data should be used to provide better phylogenetic resolution.

Methods

Sample collection and mitogenome sequencing. Specimens were collected from different localities in South Korea during scuba diving sessions (Table S11). Upon collection, samples were preserved in 95% ethanol in preparation for DNA extraction. Before starting mitogenome sequencing, all species were identified using DNA barcoding (data not shown). Following species confirmation, total DNA was extracted from the feet of the specimens using E.Z.N.A.® Mollusk DNA Kit (Omega Bio-tek, Norcross, USA). Library preparation was conducted using Illumina TruSeq Nano DNA Sample Prep Kit (Illumina, San Diego, USA) following the manufacturer's instructions. Paired-end reads of mitogenomes were generated from prepared libraries using an Illumina MiSeq system (Illumina, San Diego, USA). First, data quality was checked, and clean reads were generated by trimming adapters and low-quality bases. Subsequently, MITObim was used to assemble the mitogenome sequences from clean reads²². Fragments of *cox1* from the same species were used as bait for assembly. Mitogenome sequences were annotated on the MITOS web server using the invertebrate genetic code²³.

PCGs and rRNA genes were aligned with homologous genes from other nudibranchs and confirmed using BLAST searches in GenBank. Additionally, tRNA structures were predicted and identified using the MITOS web server²³ and ARWEN²⁴. Circular maps of complete mitogenomes were generated and annotated using Geneious v9.1²⁵. Skewness was assessed using the following formulas: AT skew = $[A - T] / [A + T]$; GC skew = $[G - C] / [G + C]$ ²⁶. RSCU values were calculated using MEGA X to evaluate the level of nucleotide bias in each codon²⁷.

Family	Species	Length (bp)	Genbank no	References
Cadlinidae	<i>Aldisa cooperi</i>	14,517	MT919638	This study
Cadlinidae	<i>Cadlina japonica</i>	14,982	MT919639	This study
Cadlinidae	<i>Cadlina koreana</i>	14,707	MT919640	This study
Cadlinidae	<i>Cadlina umiushi</i>	14,731	MT919641	This study
Dorididae	<i>Doris odhneri</i>	14,445	OL800585	This study
Discodorididae	<i>Carminodoris armata</i>	14,424	OL800584	This study
Discodorididae	<i>Asteronotus hepaticus</i>	14,464	MW559976	Unpublished
Polyceridae	<i>Triopha modesta</i>	14,397	MW387958	This study
Polyceridae	<i>Nembrotha kubaryana</i>	14,395	KY131978	Xiang et al. ²⁸
Chromodorididae	<i>Verconia nivalis</i>	14,595	OL800584	This study
Chromodorididae	<i>Chromodoris magnifica</i>	14,446	DQ991931	Medina et al. ¹⁷
Chromodorididae	<i>Chromodoris quadricolor</i>	14,259	KU317089	Xiang et al. ²⁹
Chromodorididae	<i>Chromodoris annae</i>	14,260	MF683074	Lin et al. ³⁰
Chromodorididae	<i>Chromodoris orientalis</i>	14,260	MH550543	Yu et al. ³¹
Chromodorididae	<i>Hypselodoris festiva</i>	14,880	KU365323	Karagozlu et al. ¹⁸
Chromodorididae	<i>Hypselodoris apolegma</i>	14,749	MF683075	Lin et al. ¹⁹
Chromodorididae	<i>Hypselodoris bullocki</i>	14,666	MF785092	Lin et al. ¹⁹
Chromodorididae	<i>Doriprismatica atromarginata</i>	14,421	MN171300	Do et al. 2019 ³²
Phyllidiidae	<i>Phyllidia ocellata</i>	14,598	KU351090	Xiang et al. ³³
Phyllidiidae	<i>Phyllidiella pustulosa</i>	14,717	MK279705	Do et al. ³⁴
Phyllidiidae	<i>Phyllidiopsis krempfi</i>	14,970	MT726194	Kim et al. ³⁵
Aegiridae	<i>Notodoris gardineri</i>	14,424	DQ991934	Medina et al. ¹⁷
Goniodorididae	<i>Okenia hiroi</i>	14,606	MW408699	Do et al. ³⁶
Arminidae	<i>Dermatobranchus otome</i>	14,559	MT527185	Do et al. ³⁷
Tritoniidae	<i>Tritonia tetraquetta</i>	14,540	KP764765	Sevigny et al. ¹⁴
Tethydidae	<i>Melibe leonina</i>	14,513	KP764764	Sevigny et al. ¹⁴
Pleurolidiidae	<i>Protaeolidiella atra</i>	14,445	MN911169	Do et al. ³⁸
Facelinidae	<i>Sakuraeolis japonica</i>	15,059	KX610997	Karagozlu et al. ³⁹

Table 1. Complete mitogenomes used in this study. Species names and systematics are used following the World Register of Marine Species at <http://www.marinespecies.org>.

Phylogenetic analysis. The mitogenomes generated in this study and those obtained from GenBank for other nudibranchs were used for phylogenetic analyses (Table 1). Two species of the suborder Cladobranchia, *Dermatobranchus otome* and *Tritonia tetraquetta* were used as outgroup. Both amino acid and nucleotide sequences were applied to construct phylogenetic trees. Because *nd4l* gene of *Notodoris gardineri* was half as short as that of other species, this gene was excluded from the analyses. For both amino acid and nucleotide sequences, each sequence was extracted and aligned using MAFFT v7⁴⁰ in Geneious v9.1²⁵. To check the impact of the variable regions of mitogenomes on phylogenetic trees, two alignment schemes were used¹¹. In the first scheme, following alignment, sequences were directly concatenated without the use of Gblocks. In the second scheme, Gblocks v0.91b was used to remove poorly aligned regions⁴¹. Four datasets were used to build the phylogenetic trees: nucleotide sequences of 12 PCGs + 2 rRNAs + 22 tRNAs, nucleotide sequences of 12 PCGs, 1st and 2nd codons of 12 PCGs and amino acid sequences of 12 PCGs. For the nucleotide and amino acid sequences of the 12 PCGs, each of the 12 sequences was set for a separate partition. The sequences of each mitogenome were concatenated using Geneious v9.1²⁵. The best partition scheme and the best fit model were determined using Partition Finder 2⁴². For a dataset with 36 nucleotide sequences, each of the 12 PCGs and two rRNA genes were set for separate partitions and the 22 tRNA genes were set for a partition. Also, for testing the impact of outgroup on phylogenetic reliability, different outgroup including three species of the suborder Cladobranchia, *Melibe leonine*, *Protaeolidiella atra* and *Sakuraeolis japonica* were used to compare BI and UFboot values among trees. Data for phylogenetic analyses were prepared as description above with the use of Gblocks⁴¹.

The phylogenetic trees were constructed with Maximum Likelihood (ML) and Bayesian Inference (BI) methods. ML trees were searched using IQ-tree v2.1.2 with 1,000 bootstrap replicates⁴³. BI trees were searched using MrBayes v3.2.7 with four chains and 20,000,000 and 3,000,000 generations for the nucleotide dataset and amino acid dataset, respectively⁴⁴. Additionally, sampling was performed every 100 generations and 25% of the first tree was set as burn-in. Each run was checked for proper mixing and convergence on the basis of ESS values of > 200 in Tracer v1.7⁴⁵. The maximum clade credibility tree was visualized using FigTree v1.4.4⁴⁶.

A tree topology test was performed using the mitogenome sequences and IQ-tree v2.1.2, to compare the interfamilial relationships of dorid nudibranchs found in the present study with those found in previous reports⁴³. The tree topology from the present study was tested against a constrained tree in which Aegiridae was a sister

group to Dorididae and Goniodorididae was a sister group to Discodoridae. *P*-values for approximately unbiased tests were obtained from IQ-tree v2.1.2 using 20,000 bootstrap replicates⁴³.

Data availability

The mitogenome sequences generated during the current study are available in GenBank under accession numbers: *Aldisa cooperi* (MT919638), *Cadlina japonica* (MT919639), *Cadlina koreana* (MT919640), *Cadlina umiushi* (MT919641), *Carminodoris armata* (OL800584), *Doris odhneri* (OL800585), *Triopha modesta* (MW387958), and *Verconia nivalis* (OL800586).

Received: 24 February 2022; Accepted: 31 October 2022

Published online: 05 November 2022

References

- Valdés, Á. Phylogeography and phyloecology of dorid nudibranchs (Mollusca, Gastropoda). *Zool. J. Linn. Soc.* **83**, 551–559 (2004).
- Goodheart, J. A., Bazinet, A. L., Collins, A. G. & Cummings, M. P. Relationships within Cladobranchia (Gastropoda: Nudibranchia) based on RNA-Seq data: An initial investigation. *R. Soc. Open Sci.* **2**, 150196 (2015).
- Dean, L. J. & Prinsep, M. R. The chemistry and chemical ecology of nudibranchs. *Nat. Prod. Rep.* **34**, 1359–1390 (2017).
- Avila, C. & Angulo-Preckler, C. Bioactive compounds from marine heterobranchs. *Mar. Drugs*. **18**, 657 (2020).
- Hellou, J., Andersen, R. J. & Thompson, J. E. Terpenoids from the dorid nudibranch *Cadlina luteomarginata*. *Tetrahedron* **38**, 1875–1879 (1982).
- Dumdei, E. J. *et al.* New terpenoid metabolites from the skin extracts, an egg mass, and dietary sponges of the Northeastern Pacific dorid nudibranch *Cadlina luteomarginata*. *Can. J. Chem.* **75**, 773–789 (1997).
- Fontana, A. *et al.* New scalaranes from the nudibranch *Glossodoris atomarginata* and its sponge prey. *J. Nat. Prod.* **62**, 1367–1370 (1999).
- Korshunova, T. A. *et al.* The emperor *Cadlina*, hidden diversity and gill cavity evolution: New insights for the taxonomy and phylogeny of dorid nudibranchs (Mollusca: Gastropoda). *Zool. J. Linn. Soc.* **20**, 1–66 (2020).
- Wägele, H. & Willan, R. C. Phylogeny of the Nudibranchia. *Zool. J. Linn. Soc.* **130**, 83–181 (2000).
- Valdés, Á. A phylogenetic analysis and systematic revision of the cryptobranch dorids (Mollusca, Nudibranchia, Anthobranchia). *Zool. J. Linn. Soc.* **136**, 535–636 (2002).
- Hallas, J. M., Chichvarkhin, A. & Gosliner, T. M. Aligning evidence: Concerns regarding multiple sequence alignments in estimating the phylogeny of the Nudibranchia suborder Doridina. *R. Soc. Open Sci.* **4**, 171095 (2017).
- Bouchet, P. *et al.* Revised classification, nomenclator and typification of gastropod and monoplacophoran families. *Malacologia* **61**, 1–526 (2017).
- Johnson, R. F. Breaking family ties: Taxon sampling and molecular phylogeny of chromodorid nudibranchs (Mollusca, Gastropoda). *Zool. Scr.* **40**, 137–157 (2011).
- Sevigny, J. L. *et al.* The mitochondrial genomes of the nudibranch mollusks, *Melibe leonina* and *Tritonia diomedea*, and their impact on gastropod phylogeny. *PLoS ONE* **10**, e0127519 (2015).
- Lee, Y. *et al.* A mitochondrial genome phylogeny of Mytilidae (Bivalvia: Mytilida). *Mol. Phylogenet. Evol.* **139**, 106533 (2019).
- Varney, R. M. *et al.* Assessment of mitochondrial genomes for heterobranch gastropod phylogenetics. *BMC Ecol. Evol.* **21**(6), 1–14 (2021).
- Duchêne, S., Archer, F. I., Vilstrup, J., Caballero, S. & Morin, P. A. Mitogenome phylogenetics: The impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS ONE* **6**, e27138 (2011).
- Karagozlu, M. Z., Sung, J., Lee, J., Kwak, W. & Kim, C. B. Complete sequences of mitochondrial genome of *Hypselodoris festiva* (A. Adams, 1861) (Mollusca, Gastropoda, Nudibranchia). *Mitochondrial DNA Part B* **1**, 266–267 (2016).
- Lin, G. M., Xiang, P., Audira, G. & Hsiao, C. D. Low coverage whole genome sequencing yields the complete mitogenome of *Hypselodoris bullocki* and *Hypselodoris apolegma* (Mollusca: Chromodorididae). *J. Coast. Res.* **97**, 23–28 (2019).
- Medina, M. *et al.* Crawling through time: Transition of snails to slugs dating back to the Paleozoic, based on mitochondrial phylogenomics. *Mar. Genom.* **4**, 51–59 (2011).
- Turner, L. M. & Wilson, N. G. Polyphyly across oceans: A molecular phylogeny of the Chromodorididae (Mollusca, Nudibranchia). *Zool. Scr.* **37**, 23–42 (2008).
- Hahn, C., Bachmann, L. & Chevreaux, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucl. Acid. Res.* **41**, e129 (2013).
- Bernt, M. A. *et al.* MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69**, 313–319 (2013).
- Laslett, D. & Canbäck, B. ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**, 172–175 (2008).
- Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
- Perna, N. T. & Kocher, T. D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **41**, 353–358 (1995).
- Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
- Xiang, P. *et al.* The complete mitogenome of sea slug, *Nembrotha kubaryana* (Mollusca: Polyceridae). *Conserv. Genet. Resour.* **9**, 45–247 (2017).
- Xiang, P., Lin, M., Zhao, L., Shen, K. N. & Hsiao, C. D. Low-coverage genome sequencing yields the complete mitogenome of pyjama slug, *Chromodoris quadricolor* (Mollusca: Chromodorididae). *Mitochondrial DNA Part B* **1**, 94–95 (2016).
- Lin, G. M., Xiang, P., Sampurna, B. P. & Hsiao, C. D. Genome skimming yields the complete mitogenome of *Chromodoris annae* (Mollusca: Chromodorididae). *Mitochondrial DNA Part B* **2**, 609–610 (2017).
- Yu, C., Kim, H., Kim, H. J. & Jung, Y. H. The complete mitochondrial genome of the oriental sea slug: *Chromodoris orientalis* (Nudibranchia, Chromodorididae). *Mitochondrial DNA Part B* **3**, 1017–1018 (2018).
- Do, T. D., Karagozlu, M. Z., Nguyen, V. Q. & Kim, C. B. Sequencing and analysis of the complete mitogenome of *Doriprismatica atomarginata* (Cuvier, 1804). *Mitochondrial DNA Part B* **4**, 2894–2895 (2019).
- Xiang, P., Mao, L., Wang, Y., Shen, K. E. & Hsiao, C. D. The complete mitogenome of sea slug, *Phyllidia ocellata* (Mollusca: Phyllidiidae). *Mitochondrial DNA Part B* **1**, 96–97 (2016).
- Do, T. D. *et al.* The complete mitochondrial genome of *Phyllidiella pustulosa* (Cuvier, 1804) (Nudibranchia, Phyllidiidae). *Mitochondrial DNA Part B* **4**, 771–772 (2019).
- Kim, H., Yoon, M., Kim, K. Y. & Jung, Y. H. The complete mitochondrial genome of sea slug *Phyllidiopsis krempfi* Pruvot-Fol, 1957 (Nudibranchia, Phyllidiidae) from Pacific Ocean. *Mitochondrial DNA Part B* **6**, 1523–1524 (2021).

36. Do, T. D., Jung, D. W., Choi, T. J., An, H. E. & Kim, C. B. Characterization of the complete mitochondrial genome of *Okenia hiroi* (Baba, 1938) (Nudibranchia, Goniadorididae). *Mitochondrial DNA Part B* **6**, 1124–1125 (2021).
37. Do, T. D., Choi, Y., Jung, D. W. & Kim, C. B. Caution and curation for complete mitochondrial genome from next-generation sequencing: A case study from *Dermatobranchius otome* (Gastropoda, Nudibranchia). *Anim. Syst. Evol. Divers.* **36**, 336–346 (2020).
38. Do, T. D., Choi, T. J., Jung, D. W., An, H. E. & Kim, C. B. The mitochondrial genome analysis of *Protaeolidiella atra* Baba, 1955 from Korea. *Mitochondrial DNA Part B* **5**, 1277–1278 (2020).
39. Karagozlu, M. Z., Sung, J. M., Lee, J., Kim, S. G. & Kim, C. B. Complete mitochondrial genome analysis of *Sakuraeolis japonica* (Baba, 1937) (Mollusca, Gastropoda, Nudibranchia). *Mitochondrial DNA Part B* **1**, 720–721 (2016).
40. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
41. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Phylogenet. Evol.* **17**, 540–552 (2000).
42. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773 (2017).
43. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
44. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
45. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
46. Rambaut, A. (2018) FigTree, Version 1.4.4. Available at: <https://tree.bio.ed.ac.uk/software/figtree/>.

Acknowledgements

The authors would like to thank Jung-Il Kim, Tae-June Choi, Yiso Choi and Hyung-Eun for sample preparation and handling raw data. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIP) (No. NRF-2018R1D1A1B07042858) and the National Institute of Biological Resources grant funded by the Ministry of Environment (MOE) of the Republic of Korea (No. NIBR202102108).

Author contributions

Conceptualized the idea: C.B.K., T.D.D. Collected the samples: D.W.J., T.D.D. Data generation: C.B.K., T.D.D. Data analyses: C.B.K., T.D.D., D.W.J. Wrote the first draft of the manuscript: T.D.D., D.W.J. Substantially revised the manuscript: C.B.K., T.D.D., D.W.J. Reviewed and edited the manuscript: C.B.K., T.D.D.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-23400-9>.

Correspondence and requests for materials should be addressed to C.-B.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022