Comparative analysis of the ribosomal DNA repeat unit (rDNA) of *Perna viridis* and *Perna canaliculus*

Zhansheng Guo [1], **Leng Han** [1], **Zhenlin Liang** [1], **Xuguang Hou** Corresp. [1]

[1] Shandong University at weihai, Weihai, China

Corresponding Author: Xuguang Hou
Email address: houxuguang@sdu.edu.cn

Perna viridis and *P.canaliculus* are economically and ecologically important species of shellfish. In this study, the complete ribosomal DNA unit (rDNA) sequences of these species were determined for the first time. The gene order, 18S rRNA–internal transcribed spacer (ITS) 1–5.8S rRNA–ITS2–28S rRNA–intergenic spacer (IGS), was similar to that observed in other eukaryotes. The lengths of the *P. viridis* and *P. canaliculus* rDNA sequences ranged from 8432 to 8616 bp and from 7597 to 7610 bp, respectively, thisvariability was mainly attributable to the IGS region. The putative transcription termination site and initiation site were confirmed. *P. viridis* and *P. canaliculus* rDNA contained two and one repeat motifs, respectively. Individual intra-species differences mainly involved the copy numbers of repeat units. In *P. viridis*, three CpG sites were found to cover nearly the entire IGS sequence, whereas in *P. canaliculus*, two CpG islands with sizes of 361 and 484 bp were identified. The phylogenetic trees constructed with maximum likelihood and neighbour-joining methods and based on ITS sequences were identical and included three major clusters. Species of the same genus were easily clustered in a single clade.

Comparative analysis of the ribosomal DNA repeat unit (rDNA) of *Perna viridis* and *Perna canaliculus*

Zhansheng Guo, Leng Han, Zhenlin Liang, Xuguang Hou *

Shandong University at Weihai, Weihai, China

* Corresponding author: Xuguang Hou;

Address: Shandong University at Weihai, Weihai, China

E-mail: houxuguang@sdu.edu.cn

Abstract: *Perna viridis* and *P.canaliculus* are economically and ecologically important species of shellfish. In this study, the complete ribosomal DNA unit (rDNA) sequences of these species were determined for the first time. The gene order, 18S rRNA–internal transcribed spacer (ITS) 1–5.8S rRNA–ITS2–28S rRNA–intergenic spacer (IGS), was similar to that observed in other eukaryotes. The lengths of the *P. viridis* and *P. canaliculus* rDNA sequences ranged from 8432 to 8616 bp and from 7597 to 7610 bp, respectively, thisvariability was mainly attributable to the IGS region. The putative transcription termination site and initiation site were confirmed. *P. viridis* and *P. canaliculus* rDNA contained two and one repeat motifs, respectively. Individual intra-species differences mainly involved the copy numbers of repeat units. In *P. viridis*, three CpG sites were found to cover nearly the entire IGS sequence, whereas in *P. canaliculus*, two CpG islands with sizes of 361 and 484 bp were identified. The phylogenetic trees constructed with maximum likelihood and neighbour-joining methods and based on ITS sequences were identical and included three major clusters. Species of the same genus were easily clustered in a single clade.

1. Introduction

The family Mytilidae comprises a diverse group of bivalves that are broadly distributed in marine environments (Distel, 2000). This family includes mussels of the genus *Perna*, which encompasses three currently recognised species of intertidal mussels: *Perna viridis*, *P. canaliculus* and *P. perna*. All three species are considered economically, ecologically and environmentally important (Wood et al., 2007). The green mussel *P. viridis* is a warm water bivalve species distributed along the coasts and estuaries of the Asia-Pacific region. This mollusc species is highly prized both as a food source and an important aquaculture component in southeast Asia (Tan and Ransangan, 2017; Wang et al., 2018). The Greenshell™ mussel *P. canaliculus* is endemic to New Zealand, and its range extends from the warm temperate northern waters (35.5°S) to the cold temperate waters south of Stewart Island (47°S) (Wood et al., 2007). The aquaculture industry surrounding the Greenshell™ mussel has expanded rapidly and now represents three quarters of all aquaculture exports (by value) from New Zealand (Ibarrola et al., 2017).

To date, molecular phylogenetic studies of *Perna* species have been based mainly on mitochondrial cytochrome oxidase I (COI) and ribosomal internal transcribed spacer (ITS) sequence data (Wood et al., 2007; Cunha et al., 2014; Gardner et al., 2016). Although COI (623 bp in length) and ITS (711 bp) sequences can be used as molecular markers to distinguish *Perna* species, these data do not supply sufficient information for classification based on geographical identification or among strains (or even individuals). Within a given locus, shorter markers may include less genetic variation than longer markers, which would reduce the ability to distinguish closely related species (Heeger et al., 2018). Therefore, this study aimed to identify a sufficiently informative molecular marker useful for inter-species and intra-species phylogenetic analyses of *Perna* species.

In eukaryotes, nuclear ribosomal DNA (rDNA) is usually organised in long tandem repeats. These repeats are arranged head-to-tail in large rDNA clusters to form nucleolar organising regions (NOR) in chromosomes (Dyomin et al. 2016). Each repeated transcribed unit of rDNA comprises a coding region (18S, 5.8S and 28S rRNA genes) separated by two internal transcribed spacers (ITS1 and ITS2). Moreover, each transcribed unit is separated by an intergenic spacer (IGS) which can be subdivided into external transcribed spacers (ETSs) and the non-transcribed spacer (NTS). The coding regions are highly conserved between organisms, and have been selected as genetic markers for higher-level relationships within the molluscan (Distel, 2000; Passamaneck et al., 2004; Combosch and Giribet, 2016). Whereas non-transcribed regions (ITS1, ITS2 and IGS) exhibit high structural variability (Li et al., 2016; Huang et al., 2017; Guo et al., 2018 a). The IGS region often contains repeat fragments, promoter and enhancer sites. Both the IGS and its embedded promoters can evolve more rapidly than other parts of the ribosomal repeat unit, which can lead to considerable differences in the sequences between related species and considerable intra-individual variations in length. The IGS region was shown to be more variable and phylogenetically informative than the ITS (Zhao et al., 2011). The availability of a complete rDNA sequence would provide researchers with various options, depending on the molecular variability. For *Perna* species, however, the NCBI database only contains partial rDNA sequences (18S and ITS) and a complete rDNA unit has never been reported.

In the present study, the complete sequences of nrDNA in *P. viridis* and *P. canaliculus* are reported for the first time, and the molecular features of various rDNA components are characterized, especially IGS. The phylogenetic relationships among members of the family Mytilidae are revealed based on ITS sequences. Comparative analyses are performed with several known complete rDNA sequences to clarify the rDNA relationships among these molluscs and other eukaryotes.

2. Materials and Methods

2.1. Mussel material and DNA extraction

P. viridis samples were collected offshore of Leizhou Bay (20.968614° N, 110.325743° E) in China, and *P. canaliculus* samples were collected from Pegasus Bay (43.312595° S, 172.854629° E) in New Zealand. Three individuals from each species were used in this study. Total genomic DNA was extracted from muscle tissue using the TIANamp Marine Animals DNA Kit (TianGen Biotech Co., Ltd, Beijing, China) according to the manufacturer's instructions.

2.2. PCR amplification, cloning and sequencing

The rDNA genes were amplified using the primers listed in Table 1. The primers used to amplify the 18S and 28S rRNA regions were designed based on partial sequences corresponding to the 18S rRNA of *P. viridis* (EF613234) and 28S rRNA of *Geukensia demissa* (AY145405), respectively. The primers used to amplify the 18S–ITS–28S region were designed from the 3′ end of *P. viridis* (EF613234) and 5′ end of *G. demissa* (AY145405), while the IGS region were designed based on the 3′ end of *G. demissa* (AY145405) and 5′ end of *P. viridis* (EF613234). We used the primers ITS–ar1–br and ITS–ar2–br to amplify the 18S–ITS–28S regions of *P. canaliculus* and *P. viridis*, respectively, and the primers IGS–ar1–br1 and IGS–ar2–br2 to amplify the 28S–IGS–18S regions of these species, respectively. Each PCR had a total volume of 25 μl and contained the following: 1 μL of DNA template, 0.5 μL of each primer (10 μmol/L), 12.5 μL of 2 × TransTaq PCR SuperMix, and 10.5 μL of H_2O. The reactions were initially denatured at 95 ºC for 3 min and subjected to 35 cycles of denaturing at 95 ℃ for 45 s; annealing at 52 ℃ for 18S and 28S rRNA, 50 ℃ for 18S–ITS–28S and 56.5 ℃ for 28S–IGS–18S for 30 s, and extension at 72 ℃ for 1–3.5 min (1 kb/min). Finally, all reactions were subjected to a final extension at 72 ℃ for 7 min. The PCR products were electrophoresed on 0.8% agarose gels, extracted and purified. The purified PCR products were cloned using the pUCm-T Vector Cloning Kit and SanPrep Column Plasmid MiniPreps Kit (Sangon Biotech Co., Ltd.) according to the manufacturer's instructions. Three positive recombinant plasmid colonies corresponding to each amplified region were picked, cultivated and sequenced by Beijing Ruiboxingke Biotechnology Co., Ltd (Beijing, China).

2.3. Sequence analysis

The complete rDNA unit was assembled using DNAMAN software (Kaukonen et al., 2000). The boundaries of each region were confirmed using the NCBI database and BLAST software (Johnson et al., 2008). Multiple sequence alignment was performed using T-Coffee software (Notredame et al., 2000) with manual adjustments. The general molecular features of *P. viridis* and *P. canaliculus* rDNA were calculated using MEGA 7.0 (Kumar et al., 2016). Sub-repeat fragments and inverted repeats of the IGS region were analysed with the Tandem Repeats Finder (Benson, 1999) and Unipro UGENE (Okonechnikov et al., 2012), respectively. Predictions of putative transcription initiation sites (TIS) and transcription termination sites (TTSs) in the IGS regions were based on a comparative analysis of the sequences with data for various marine animal species from the literature (Ki et al., 2009; Luchetti et al., 2016; Chae et al., 2018; Guo et al., 2018 a). CpG islands were identified using CpGPlot (Polanco and Delavega, 1994). The sequence identity was confirmed using DNAMAN software. The genetic distance between *P. viridis* and *P. canaliculus* was calculated with MEGA 7.0 software according to a Kimura two-parameter model.

For phylogenetic reconstruction, the ITS sequences of *P. viridis* and *P. canaliculus* were first determined. Next, nine complete or near-complete sequences from the following Mytilidae species were obtained from the GenBank database: *Mytilus edulis* (AY695798), *M. unguiculatus* (former *M. coruscus*) (MK157201), *M. galloprovincialis* (JX081670), *M. trossulus* (JX081669), *Aulacomya maoriana* (former *Aulacomya atra maoriana*) (DQ924557), *P. perna* southern Indian population (former *P. indica*) (JQ622200), *P. perna* Omani population (KC692037), *P. perna* Mauritanian population (former *P. picta*) (DQ924548) and *Modiolus rectus* (EF035114). Maximum likelihood (ML) and neighbour-joining (NJ) phylogenetic trees based on ITS sequences were constructed with 1000 bootstrap replicates using MEGA 7.0 software and the GTR+G nucleotide substitution and Kimura two-parameter models, respectively. *Crassostrea ariakensis* (EU252081) was selected as the outgroup.

3. Results

3.1. Complete rDNA sequences of *P. viridis* and *P. canaliculus*

The complete rDNA units of *P. viridis* and *P. canaliculus* were determined after sequencing, comparison and assembly. Fig. 1 presents a scheme of the structural organisation of the rDNA repeat unit, which includes the following elements in order: 18S–ITS1–5.8S–ITS2–28S–IGS. The complete rDNA units of *P. viridis* and *P. canaliculus* varied from 8432 to 8616 bp and from 7597 to 7609 bp, respectively. Both sequences have been deposited in the GenBank database (accession numbers: MK419104–MK419109).

The length, GC content, pairwise identity and variable sites of each region are shown in Table 2. An analysis of the *P. viridis* and *P. canaliculus* sequences revealed low levels of genetic variation in the 18S, 5.8S and 28S rRNA regions, which had pairwise identities of 99.56%, 100% and 98.94%, respectively. The variable (V) sites in 18S and 28S rRNA made respective contributions of only 0.83% and 1.22%, indicating that the coding regions (18S, 5.8S and 28S rRNA) were even more highly conserved than the complete rDNA unit (6.25%). Differences between *P. viridis* and *P. canaliculus* were observed mainly in non-coding regions (ITS1, ITS2 and IGS). In intra-species

comparisons of three individuals per species, the sequence similarities of ITS1 and ITS2 in *P. viridis* were 99.38% and 93.87%, respectively, and these regions differed by only five and one base, respectively. Among *P. canaliculus* individuals, both regions had sequence similarities of 100%. The respective sizes of ITS1 and ITS2 were 269 and 263 bp in *P. viridis* and 304 and 258 bp in *P. canaliculus*. Differences in the lengths and variable sites of the ITS1 and ITS2 regions resulted in relatively low sequence identities (66.99% in ITS1, 67.03% in ITS2).

Comparisons of the IGS sequences at various levels revealed that the highest (91.84% for *P. viridis*, 98.31% for *P. canaliculus*) and lowest similarities (33.32%) were found at the individual and interspecies levels, respectively, and the variable sites between the IGS sequences of the two species accounted for 15.65% of all sites in this region (Table 2, 3). The sizes of the IGS sequences ranged from 1399–1411 bp and 2267–2451 bp in *P. canaliculus* and *P. viridis*, respectively. This result suggested that the IGS sequences of them had large differences in length and a relatively high level of sequence heterogeneity. Due to heterogeneity in the length of IGSs at the individual and interspecies levels of the genus *Perna*, the longest IGS sequence of each species was chosen as a representative to outline the canonical structural organization of the IGSs from *P. canaliculus* and *P. viridis*. The detailed analysis of the molecular structure of IGS sequences of them contains six distinct regions: TTS, TIS, NTS, ETS, SR and CpG islands.

3.2. Beginning and ending sequences of transcripts, ETS and NTS

A poly (T) tract (5′-TTTTCGTTTGCCTTTTTCGTTCCTTTTTTTTT-3′) was identified in the *P. viridis* IGS region from 171 to 202 bp. The sequence 5′ -TTACTTGT-3′ was detected in the *P. canaliculus* IGS region from 120 to 127 bp. Both sequences were considered putative TTSs. Accordingly, the positions of the 3′ ETS sequences in the *P. viridis* and *P. canaliculus* 5′ IGS regions were determined to correspond to bp 1–202 and bp 1–127, respectively. The sequence ‘TTATTATGTGGAGTGGG’ was considered a putative TIS for RNA polymerase I. The RNA transcript initiation site at the +1 position corresponds to the C/T residue in the middle of the IGS region in *P. viridis* and *P. canaliculus*. The 5′ ETSs were localised between the putative TIS and the beginning of the 18S rRNA gene and had lengths of 581 bp and 500–503 bp in *P. viridis* and *P. canaliculus*, respectively. NTSs comprised the remainder of IGS. The high level of individual, intraspecies and interspecies divergence within the NTS sequences was attributed to lower functional constraints. In contrast, the nucleotide sequences were more conserved in the ETS region than in the NTSs, despite differences in the length of the former between *P. viridis* and *P. canaliculus*.

3.3. Sub-repeat (SR) regions and CpG islands in the IGS sequence

Next, the DNA repeats in the IGS sequences were analysed to explore the IGS genetic structure in greater detail. The identification of the numbers and lengths of the repetitive sequence motifs occurring in *P. viridis* and *P. canaliculus* enabled the definition of three different SRs with lengths of 45–131bp (Fig. 1). A comparison of the IGS regions indicated an extremely high level of sequence diversity within the SR region. The *P. canaliculus* sequence included only one type of SR (length: 131 bp) which was repeated 2.9 times, and three poly (A) tracts in the NTS area. The *P. viridis* sequence included two repeat motifs: SR1 (length: 93 bp), which was repeated 9.3–11.5 times

within the NTS, and SR2 (length: 45 bp), which was repeated 2.5 times within the 5′ETS. Differences between individuals of the same species were due to SR copy number variation. No inverted repeats were detected in either *Perna* species.

The GC content of the IGS region was slightly higher in *P. viridis* than in *P. canaliculus* (Table 2). Similarly, the number and lengths of the CpG islands were greater in *P. viridis* than in *P. canaliculus*. In *P. viridis*, three CpG sites were found to cover nearly the entire IGS sequence. By contrast, the *P. canaliculus* IGS region included two CpG islands located at the NTS and 5′ETS (lengths: 361 and 484 bp).

3.4. Phylogenetic tree analysis based on ITS sequences

ML and NJ phylogenetic trees were constructed based on the ITS sequences of Mytilidae species (Fig. 2). clustering pattern analysis revealed that the NJ and ML trees were identical. Both included three major clusters, and species of the same genus clearly clustered in single clades with relatively high supporting values. For both *P. viridis* and *P. canaliculus*, the individuals were classified into single clusters with high sequence identity. All Mytilidae species included in the phylogenetic analysis belong to the subfamily Mytilinae, except *M. rectus* which belongs to the subfamily Modiolinae. During tree formation, *M. rectus* and *A. maoriana* were first clustered together, and then clustered with *Perna* and *Mytilus* species.

4. Discussion

In this study, the complete rDNA units of *P. viridis* and *P. canaliculus* were sequenced for the first time. The rDNA gene orders in these organisms were identical to those in other eukaryotes. Reported rDNA unit lengths vary among taxa, as shown in Table 4, with values of 43 kb (human) (Gonzalez and Sylvester, 1995) and 45 kb (mouse) for mammals (Grozdanov et al., 2003); 13.67 kb for fish (*Cyprinus carpio*) (Vera et al., 2003); 12.26 kb (*Brachiola algerae*) (Belkorchia et al., 2008), 9.5 kb (*Plasmodiophora brassicae*) (Niwa et al., 2011) and 8.3 kb (*Eurytrema pancreaticum*) (Su et al., 2018) for parasites; 7.9 kb (*Paracyclopina nana*) (Ki et al., 2011), 7.7 kb (*Aurelia* sp.1) (Ki et al., 2009) and 9.6–10.7 kb (*Haliotis* species) (Guo et al., 2017, 2018 a b) for marine invertebrates; 7.9–8.9 kb (*Oryza sativa*) (Fujisawa et al., 2006) and 8.0–8.9 kb (*Stipa* spp.) (Krawczyk et al., 2017) for land plants and 11.76–12.57 kb (*Bangia*) (Xu et al., 2016) and 13.65 kb (*Pyropia yezoensis*) (Li et al., 2016) for sea algae. Meanwhile, the rDNA lengths of *P. viridis* and *P. canaliculus* were 8.6 and 7.6 kb, respectively.

Table 4 compares each region of the rDNA unit between the two *Perna* species and other eukaryotes. The length of the coding region (18S, 5.8S and 28S rRNA genes) exhibits conservative features. The length of 18S rRNA varies from 1799 bp to 1996 bp except for *Plasmodiophora brassicae* (3105 bp) and *Brachiola algerae* (1391 bp). The sizes of 5.8S rRNA in most eukaryotes are around 160 bp except for *Entamoeba invadens* (116 bp) and *Brachiola algerae*, In *B. algerae*, the ITS size was estimated at 22 bp and the 5.8S rRNA gene was not detected (Belkorchia et al., 2008), while the length of 28S rRNA ranged from 3 to 5 kb. Differences in rDNA lengths among species are mainly attributable to the IGS. According to the database of rDNA unit sequences submitted to GenBank, IGS lengths vary from 0.58 kb in *Paramphistomum cervi* to over 30 kb in mammals.

This study observed variations in the lengths of the IGS region in comparisons at both intra- and inter-species levels. Sequence differences in this region were largely responsible for variations in the rDNA lengths of *P. viridis* and *P. canaliculus* at both levels and are mainly attributable to differences in indels between individuals. Compared with the entire IGS region and the NTS, the beginning of the 3′ ETS and latter half of the 5′ ETS sequences were more highly conserved and exhibited greater sequence identity (>65%). All rDNA IGS sequences shared characteristics such as the presence of SR elements. Therefore, heterogeneity in the IGS length can be attributed to duplications or deletions of the SR region, which may be present in different copy numbers in almost all species (Ambrose and Crease, 2011; Huang et al., 2017). In an intra-species comparison of individuals, length variants were mainly attributable to the SR zone; for example, 9.3–11.5 copies of the repeat motif SR1 were present in *P. viridis* individuals. Huang et al. (2017) speculated that indels are restricted to SR-rich regions. Therefore, length polymorphisms in IGS result from concerted evolution, and unequal crossover between SR elements might be a major driving force underlying the evolution of rDNA units. The repeat fragments were found to be species-specific and rarely adhered to any predictable behaviour or similarities even within a genus. For example, *P. viridis* and *P. canaliculus* exhibited totally different repeat patterns, and this phenomenon has also been observed in other eukaryotes (Krawczyk et al., 2017; Guo et al., 2018 a). Possibly, the repeat pattern could be used as a genetic marker for species identification.

The IGS contains several functional elements, including TTS and TIS regions. According to Ki et al. (2009, 2011), Guo et al. (2018 a, b) and Chae et al. (2018), a poly(T) tract on the 5′ side of the IGS can be considered a the putative TTS in marine invertebrate. The typical structure of the region near the beginning of the 5′ IGS was also observed in *P. viridis* and *P. canaliculus*, even though the sequences were not totally identical. The core promoter contains TATA and GGGG boxes, which appear to be a general feature of rRNA transcription, and this site is immediately preceded by an AT-rich region, which is commonly found in both plants and animals (Vera et al., 2003; Wang et al., 2003; Maggini et al., 2008; Krawczyk et al., 2017) . However, the beginnings of the *P. viridis* and *P. canaliculus* IGS genes contained only two and one TATA-box sequences, respectively, and lacked AT-rich regions. Only one previous report described a similar putative TIS sequence (TTATTATGTGGAGTG GG). Specifically, Krawczyk et al. (2017), after an analysis of Poaceae species, also reported that an AT-rich region upstream of the TIS was not ubiquitous. The rDNA genes are widely used to resolve phylogenetic relationships between species at various taxonomic levels. However, few studies have used IGS to identify species or reveal phylogenetic relationships among mussels. A comparison of the genetic distances of rDNA genes between *P. viridis* and *P. canaliculus* yielded the largest value in the IGS region (0.304–0.313), which suggests that this region could be used to identify species even at a sub-genus level. Given its relatively rapid evolution (compared with other rDNA sequences) and differences in the SR sequences (e.g., copy number, sequence, and length) among species, the IGS region may be more suitable than ITS and 18/28S rRNA for reconstructing both inter- and intra-species phylogenetic relationships.

In conclusion, we sequenced the complete rDNA unit sequences of *P. viridis* and *P. canaliculus* for the first time. In both species, the structural organisation of the rDNA unit was similar to those of many other eukaryotes.

The unit lengths in *P. viridis* and *P. canaliculus* were 8432–8616 bp and 7597–7610 bp, respectively, and variations in the length were mainly attributable to the IGS region. We further investigated the characteristics of the IGS and assessed sequence diversity at intra-species and inter-species levels. The *Perna* rDNA unit provides a structural model of nuclear rDNA for molecular comparisons, particularly among Mytilidae species. These research discoveries will hopefully pave the way for analyses of mollusc population genetics and evolution.

References

Ambrose, C.D., Crease, T.J., 2011. Evolution of the nuclear ribosomal DNA intergenic spacer in four species of the *Daphnia pulex* complex. BMC Genet. 12, 1–14.

Belkorchia, A., Biderre, C., Militon, C., Polonais, V., Wincker, P., Jubin, C., Delbac, F., Peyretaillade, E., Peyret, P., 2008. *In vitro* propagation of the microsporidian pathogen *Brachiola algerae* and studies of its chromosome and ribosomal DNA organization in the context of the complete genome sequencing project. Parasitol. Int. 57, 62–71.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27 (2), 573–580.

Chae, J., Seo, Y., Yu, W.B., Yoon, W.D., Lee, H.E., Chang, S.J., Ki, J.S., 2018. Comprehensive analysis of the jellyfish *Chrysaora pacifca* (Goette, 1886) (Semaeostomeae: Pelagiidae) with description of the complete rDNA sequence. Zool. Stud. 57, 51.

Combosch, D.J., Giribet, G., 2016. Clarifying phylogenetic relationships and the evolutionary history of the bivalve order Arcida (Mollusca: Bivalvia: Pteriomorphia). Mol. Phyl ogenet. Evol. 94,298–312.

Cunha, R.L., Nicastro, K.R., Costa, J., McQuaid, C.D., Serrao, E.A., Zardi, G.I., 2014. Wider sampling reveals a non-sister relationship for geographically contiguous lineages of a marine mussel. Ecol. Evol. 4 (11), 2070–2081.

Distel, D.L., 2000. Phylogenetic Relationships among Mytilidae (Bivalvia): 18S rRNA data suggest convergence in mytilid body plans. Mol. Phylogenet. Evol. 15, 25–33.

Dyomin, A.G., Koshel, E.I., Kiselev, A.M., Saifitdinova, A.F., Galkina, S.A., Fukagawa, T., Kostareva, A.A., Gaginskaya, E.R., 2016. Chicken rRNA gene cluster structure. PloS ONE 11 (6), e0157464.

Fujisawa, M., Yamagata, H., Kamiya, K., Nakamura, M., Saji, S., Kanamori, H., Wu, J., Matsumoto, T., Sasaki, T., 2006. Sequence comparison of distal and proximal ribosomal DNA arrays in rice (*Oryza sativa* L.) chromosome 9S and analysis of their flanking regions. Theor. Appl. Genet. 113, 419–428.

Gardner, J.P.A., Patterson, J., George, S., Edward, J.K.P., 2016. Combined evidence indicates that *Perna indica* Kuriakose and Nair 1976 is *Perna perna* (Linnaeus, 1758) from the Oman region introduced into southern India more than 100 years ago. Biol. Invasions 18, 1375–1390.

Gonzalez, I.L., Sylvester, J.E., 1995. Complete sequence of the 43 kb human ribosomal DNA repeat– analysis of the intergenic spacer. Genomics 27, 320–328.

Guo, Z.S., Ding, Y., Zhang, X.H., Hou, X.G., 2017. Complete nuclear ribosomal DNA sequence analysis of Pacific abalone *Haliotis discus hannai.* Fish. Sci. 83, 777–784.

Guo, Z.S., Han, L., Ding, Y., Hou, X.G., Liang, Z.L., 2018 a. Molecular characterisation of the complete nuclear ribosomal DNA sequence of the blacklip abalone *Haliotis rubra*. New Zeal. J. Mar. Fresh. 52 (3), 430–443.

Guo, Z.S., Hou, X.G., Han, L., 2018 b. Complete nuclear ribosomal DNA sequence analyses of the black-footed abalone *Haliot isiris*, New Zeal. J. Mar. Fresh. 52 (2), 232–246.

Grozdanov, P., Georgiev, O., Karagyozov, L., 2003. Complete sequence of the 45 kb mouse ribosomal DNA repeat: analysis of the intergenic spacer. Genomics 82, 637–643.

Heeger, F., Bourne, E.C., Baschien, C., Yurkov, A., Bunk, B., Sproer, C., Overmann, J., Mazzoni, C.J., Monaghan, M.T., 2018. Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. Mol. Ecol. Resour. 18, 1500–1514.

Huang, Y., Yu, F., Li, X., Luo, L., Wu, J., Yang, Y., Deng, Z., Chen, R., Zhang, M., 2017. Comparative genetic analysis of the 45S rDNA intergenic spacers from three *Saccharum* species. PLoS ONE 12 (8), e0183447.

Ibarrola, I., Hilton, Z., Ragg, N.L.C., 2017. Physiological basis of inter-population, inter-familiar and intra-familiar differences in growth rate in the green-lipped mussel *Perna canaliculus*. Aquaculture 479, 544–555.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T.L., 2008. NCBI BLAST: a better web interface. Nucleic Acids Res. 36, W5–W9.

Kaukonen, J., Juselius, J.K., Tiranti, V., Kyttala, A., Zeviani, M., Comi, G.P., Keranen, S., Peltonen, L., Suomalainen, A., 2000. Role of adenine nucleotide translocator 1 in mtDNA maintenance. Science 289 (5480), 782–785.

Ki, J.S., Kim, I.C., Lee, J.S., 2009. Comparative analysis of nuclear ribosomal DNA from the moon jelly *Aurelia* sp.1 (Cnidaria: Scyphozoa) with characterizations of the 18S, 28S genes, and the intergenic spacer (IGS). Hydrobiologia 616, 229–239.

Ki, J.S., Park, H.G., Lee, J.S., 2011. Extensive analysis of nuclear cistron rDNA sequence of *Paracyclopina nana* (Cyclopoida: Cyclopettidae). Hydrobiologia 666, 3–9.

Krawczyk, K., Nobis, M., Nowak, A., Szczecińska, M., Sawicki, J., 2017. Phylogenetic implications of nuclear rRNA IGS variation in *Stipa* L. (Poaceae). Sci. Rep. 7, 11506.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Li, X.C., Xu, J.J., He, Y., Shen, S.S., Zhu, J.Y., Shen, Z.G., 2016. The complete nuclear ribosomal DNA (nrDNA) cistron sequence of *Pyropia yezoensis* (Bangiales, Rhodophyta). J. Appl. Phycol. 28, 663–669.

Luchetti, A., Scanabissi, F., Mantovani, B., 2006. Molecular characterization of ribosomal intergenic spacer in the tadpole shrimp *Triops cancriformis* (Crustacea, Branchiopoda, Notostraca). Genome 49, 888–893.

Maggini, F., Gelati, M.T., Spolverini, M., Frediani, M., 2008. The intergenic spacer region of the rDNA in *Olea europaea* L. Tree Genet. Genomes. 4:293–298.

Niwa, R., Kawahara, A., Murakami, H., Tanaka, S., Ezawa, T., 2011. Complete structure of nuclear rDNA of the obligate plant parasite *Plasmodiophora brassicae*: intraspecific polymorphisms in the exon and group I intron of the large subunit rDNA. Protist 162, 423–434.

298 Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: a novel method for multiple sequence alignments. J. Mol.
299 Biol. 302, 205–217.
300 Ojha, S., Singh, N., Bhattacharya, A., Bhattacharya, S., 2013. The ribosomal RNA transcription unit of *Entamoeba*
301 *invadens*: Accumulation of unprocessed pre-rRNA and a long non coding RNA during encystation. Mol.
302 Biochem. Parasit. 192, 30–38.
303 Okonechnikov, K., Golosova, O., Fursov, M., 2012. Unipro UGENE: a unified bioinformatics toolkit.
304 Bioinformatics 28 (8), 1166–1167.
305 Passamaneck, Y.J., Schander, C., Halanych, K.M., 2004. Investigation of molluscan phylogeny using large-subunit
306 and small-subunit nuclear rRNA sequences. Mol. Phylogenet. Evol. 32, 25–38.
307 Polanco, C., Delavega, M.P., 1994. The structure of the rDNA intergenic spacer of *Avena sativa* L.: a comparative
308 study. Plant Mol. Biol. 25 (4), 751–756.
309 Sone, T., Fujisawa, M., Takenaka, M., Nakagawa, S., Yamaoka, S., Sakaida, M., Nishiyama, R., Yamato,
310 K.T., Ohmido, N., Fukui, K., Fukuzawa, H., Ohyama, K., 1999. Bryophyte 5S rDNA was inserted into 45S
311 rDNA repeat units after the divergence from higher land plants. Plant Mol. Biol. 41, 679–685.
312 Su, X., Zhang, Y., Zheng, X., Wang, X.X., Li, Y., Li, Q., Wang, C.R., 2018. Characterization of the complete
313 nuclear ribosomal DNA sequences of *Eurytrema pancreaticum*. J. Helminthol. 92, 484–490.
314 Tan, K.S., Ransangan, J., 2017. Feeding behaviour of green mussels, *Perna viridis* farmed in Marudu Bay,
315 Malaysia. Aquaculture Res. 48, 1216–1231.
316 Vera, M.I., Molina, A., Pinto, R., Reyes, M., Alvarez, M., Krauskopf, E., Quezada, C., Torres, J., Krauskopf, M.,
317 2003. Genomic organization of the rDNA cistron of the teleost fish *Cyprinus carpio.* Biol. Res. 36, 241–251.
318 Wang, S.Q., Zhao, M.J., Li, T., 2003. Complete sequence of the 10.3 kb silkworm *Attacus ricini* rDNA repeat,
319 determination of the transcriptional initiation site and functional analysis of the intergenic spacer. DNA
320 Sequence 14 (2), 95–101.
321 Wang, Y.T., Zeng, Z.Y., Zhang, X.H., Shi, Q., Wang, C.G., Hu, Z.L., Li, H., 2018. Identification and
322 characterization of a novel defensin from Asian green mussel *Perna viridis*. Fish Shellfish Immun. 74, 242–
323 249.
324 Wood, A.R., Apte, S., MacAvoy, E.S., Gardner, J.P.A., 2007. A molecular phylogeny of the marine mussel genus
325 *Perna* (Bivalvia: Mytilidae) based on nuclear (ITS1&2) and mitochondrial (COI) DNA sequences. Mol.
326 Phylogenet. Evol. 44, 685–698.
327 Xu, J.J., Jiang, B., Chai, S.M., He, Y., Zhu, J.Y., Shen, Z.G., Shen, S.D., 2016. Complete nuclear ribosomal DNA
328 sequence amplification and molecular analyses of *Bangia* (Bangiales, Rhodophyta) from China. Chin. J.
329 Oceanol. Limnol. 34 (5), 1044–1053.
330 Zhao, G.H., Blair, D., Li, X.Y., Li, J., Lin, R.Q., Zou, F.C., Sugiyama, H., Mo, X.H., Yuan, Z.G., Song, H.Q., Zhu,
331 X.Q., 2011. The ribosomal intergenic spacer (IGS) region in *Schistosoma japonicum*: structure and
332 comparisons with related species. Infect. Genet. Evol. 11, 610–617.

Zheng, X., Chang, Q.C., Zhang, Y., Tian, S.Q., Lou, Y., Duan, H., Guo, D.H., Wang, C.R., Zhu, X.Q. , 2014. Characterization of the complete nuclear ribosomal DNA sequences of *Paramphistomum cervi*. Sci. World J. 751907.

Table 1(on next page)

Table 1. Primers used in the present paper

Amplified region	Primers	Sequences (5′-3′)	Length (bp)
18S	18S-ar	CTTTCAAATGTCTGCCCTAT	1505
	18S-br	TTCACCTACGGAAACCTTGT	
18S-ITS-28S	ITS-ar1	AGGGACAAGTGGCGTTTAGC	1573
	ITS-ar2	TCGTAACAAGGTTTCCGTAG	1162
	ITS-br	TTACCTCTAAGCGGTTTAC	
28S1	28S1-ar	TTAGAGGTAAACGGGTGGAT	1093
	28S1-br	AGTTGATTCGGCAGGTGAG	
28S2	28S2-ar	GACGAAACGACCTCAACCTA	1957
	28S2-br	AATGATAGGATGAGCCGACA	
28S-IGS-18S	IGS-ar1	GGGATAACTGGCTTGTGGCA	2486
	IGS-br1	TGGATGTGGTAGCCGTTTCT	
	IGS-ar2	GGATAACTGGCTTGTGGCA	3497
	IGS-br2	CTGCCTTCCTTGGATGTGG	

1

2

Table 2(on next page)

Table 2. Characterization of nuclear ribosomal DNA (rDNA) from *Perna viridis* and *P. canaliculus*

bp: base pair; ITS: internal transcribed spacer; IGS: intergenic spacer; the pairwise distance and identity represent the mean value between *P. viridis* (F1-F3) and *P. canaliculus* (N1-N3), respectively.

Region	*P. viridis*		*P. canaliculus*		Alignment length (bp)	Pairwise identity (%)	Pairwise distance	Variable sites (V)
	Length (bp)	GC content (%)	Length (bp)	GC content (%)				
rDNA	8432–8616	53.75–53.90	7597–7610	53.02–53.08	8714	77.71	0.066–0.067	546 (6.26%)
18S rRNA	1799	50.03	1800	50.06–50.11	1800	99.56	0.003–0.007	15 (0.83%)
ITS1	269	52.04–53.53	304	57.89	314	66.99	0.183–0.188	45 (14.33%)
5.8S rRNA	157	56.05	157	56.05	157	100	0.000	0
ITS2	263	51.71–52.09	258	49.22	277	67.03	0.246–0.252	52 (18.77%)
28S rRNA	3678	54.59–54.65	3679	54.61–54.63	3679	98.94	0.009–0.010	45 (1.22%)
IGS	2267–2451	55.52–55.81	1399–1411	52.18–52.95	2486	33.32	0.304–0.313	389 (15.65%)

1

Table 3(on next page)

Table 3. pairwise identities (upper right matrix) and pairwise distances (lower left matrix) of IGS

F1–F3 and N1–N3 represent the three individuals of *P. viridis* and *P. canaliculus*, respectively

Sample	F1	F2	F3	N1	N2	N3
F1		91.48%	98.50%	33.32%	33.70%	35.75%
F2	0.013		92.21%	33.36%	32.97%	33.63%
F3	0.014	0.004		35.90%	35.39%	36.24%
N1	0.327	0.323	0.320		97.60%	98.58%
N2	0.314	0.310	0.307	0.009		98.30%
N3	0.318	0.314	0.311	0.007	0.007	

1

Table 4(on next page)

Table 4. the species were collected with rDNA repeat unit

Species	nrDNA	18S rRNA	ITS1	5.8S rRNA	ITS2	28S rRNA	IGS	GenBank accession NO.	Reference
Homo sapiens	42999	1871	1095	157	1155	5035	33686	U13369	Gonzalez & Sylvester, 1995
Mus musculus	45306	1870	1000	157	1088	4730	36462	BK000964	Grozdanov et al., 2003
Gallus gallus	11863	1823	2530	157	733	4441	2179	KT445934	Dyomin et al., 2016
Cyprinus carpio	13676	1861	367	158	390	4093	6807	AF133089 and AY260899	Vera et al., 2003
Entamoeba invadens	22481	1962	116	116	78	3188	16700	AY190083	Ojha et al. 2013
Brachiola algerae	12269	1391		22		2557	7915	AM422905	Belkorchia et al., 2008
Paramphistomum cervi	8493-10221	1994	1293	157	286	4186	577-2305	KJ459934–KJ459938	Zheng et al. 2014
Eurytrema pancreaticum	8306-8310	1996	1103	160	231	3669	1147-1151	KY490000– KY490004	Su et al., 2018
Plasmodiophora brassicae	9513	3105	143	154	165	3611	2332	/	Niwa et al., 2011
Aurelia sp.1	7731	1814	272	158	278	3606	1603	EU276014	Ki et al., 2009
Chrysaora pacifca	8167	1810	246	158	182	3609	2162	KY212123	Chae et al., 2018
Paracyclopina nana	7974	1808	299	157	216	3572	1922	FJ214952	Ki et al., 2011
Haliotis discus hannai	10668-10698	1871	329	160	301	3411	4624-4654	KY485141- KY485146, KY569413–KY569414	Guo et al., 2017
Haliotis iris	9579-9706	1858	321	160	296	3412	3560-3662	KY933301–KY933305, KY978225–KY978226	Guo et al., 2018
Haliotis rubra	9881	1858	327	160	297	3413	3854	MF099780, MF111108, MF106175, MF106174, MF099782, MF113042	Guo et al., 2018
Oryza sativa	7928-8934	/	/	/	/	/	/	OSJNOa063K24 and OSJNBb0013K10	Fujisawa et al., 2006
Stipa spp.	7791-8897	1811	221	164	205-207	2193-3098	3397	KY826229- KY826235	Krawczyk et al., 2017
Bangia	10976-11497	1839	217-699	158	398-437	3751	4613	KP279672–KP279682, KP311305–KP311315	Xu et al., 2016
Pyropia yezoensis	13650-13654	1834	371-372	159	532-535	4770	5984	KJ578745–KJ578748, KJ608639-KJ608640	Li et al., 2016
Marchantia polymorpha	16103	1820	1067	158	381	3468	9209	AB021684	Sone et al., 1999

Figure 1

Fig. 1 Structural organisation scheme of nuclear ribosomal DNA (nrDNA) repeat units in Perna species.

ITS: internal transcribed spacer; IGS: intergenic spacer; ETS: external transcribed spacer; NTS: non-transcribed spacer; TTS: transcription termination site; TIS: transcription initiation site; SR: sub-repeat; CpG island: cytosine-guanine island. The arrows represent primer positions.

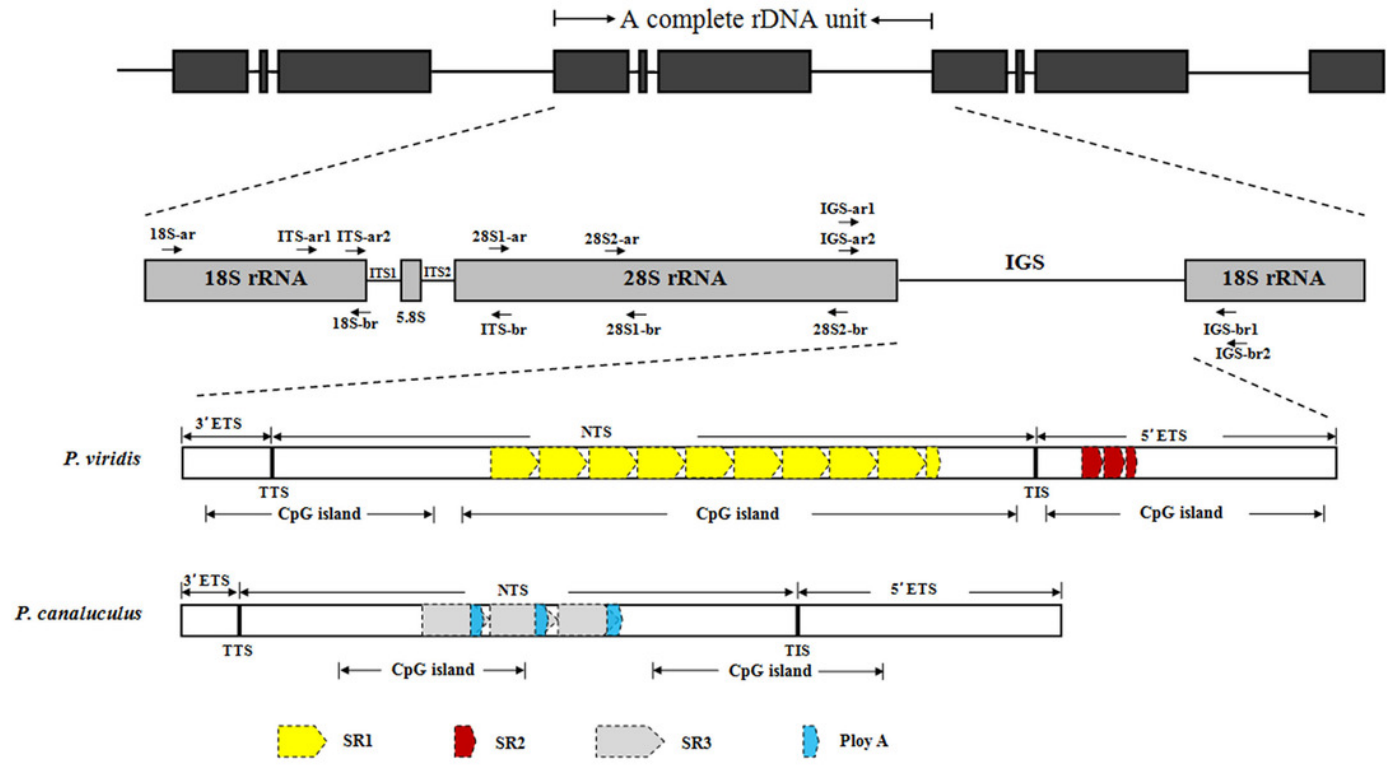

Figure 2

Fig. 2 Phylogenetic trees constructed using maximum likelihood (ML) and neighbour-joining (NJ) methods and based on ITS sequences of Mytilidae species.

Numbers around the branches indicate bootstrap support from 1000 tests. ML and NJ trees are topologically identical. F1–F3 and N1–N3 represent the three individuals of *P. viridis* and *P. canaliculus*, respectively.