DNA barcoding for species identification in deep-sea clams (Mollusca: Bivalvia: Vesicomyidae)

Jun Liu & Haibin Zhang

To cite this article: Jun Liu & Haibin Zhang (2018): DNA barcoding for species identification in deep-sea clams (Mollusca: Bivalvia: Vesicomyidae), Mitochondrial DNA Part A, DOI: 10.1080/24701394.2018.1424843

To link to this article: https://doi.org/10.1080/24701394.2018.1424843

View supplementary material

Published online: 15 Jan 2018.

Submit your article to this journal

View related articles

View Crossmark data
DNA barcoding for species identification in deep-sea clams (Mollusca: Bivalvia: Vesicomyidae)

Jun Liu and Haibin Zhang

Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, China

ABSTRACT

Deep-sea clams (Bivalvia: Vesicomyidae) have been found in reduced environments over the world oceans, but taxonomy of this group remains confusing at species and supraspecific levels due to their high-morphological similarity and plasticity. In the present study, we collected mitochondrial COI sequences to evaluate the utility of DNA barcoding on identifying vesicomyid species. COI dataset identified 56 well-supported putative species/operational taxonomic units (OTUs), approximately covering half of the extant vesicomyid species. One species (OTU2) was first detected, and may represent a new species. Average distances between species ranged from 1.65 to 29.64%, generally higher than average intraspecific distances (0–1.41%) when excluding Pliocardia sp.10 cf. venusta (average intraspecific distance 1.91%). Local barcoding gap existed in 33 of the 35 species when comparing distances of maximum interspecific and minimum interspecific distances with two exceptions (Abyssogena southwardae and Calyptogena rectimargo-starobogatovi). The barcode index number (BIN) system determined 41 of the 56 species/OTUs, each with a unique BIN, indicating their validity. Three species were found to have two BINS, together with their high level of intraspecific variation, implying cryptic diversity within them. Although fewer 16S sequences were collected, similar results were obtained. Nineteen putative species were determined and no overlap observed between intra- and inter-specific variation. Implications of DNA barcoding for the Vesicomyidae taxonomy were then discussed. Findings of this study will provide important evidence for taxonomic revision in this problematic clam group, and accelerate the discovery of new vesicomyid species in the future.

Introduction

Vesicomyid clams (Bivalvia: Vesicomyidae) are one of the dominant organism groups in the reduced environments, including hydrothermal vents, cold seeps and organic sediments (Peek et al. 1997). The family are found worldwide from continental shelf to hadal depths, at depths of 100–9530 meters (Krylova and Sahling 2010). Most, if not all, species of the family are chemosymbiotic, with sulphine-oxidizing chemoaerotrophic bacteria in the gills (Peek et al. 1998; Krylova and Sahling 2010). Their high adaptation and restriction to the reduced environments make the clams a good subject to explore ecological and evolutionary issues concerning the reducing communities (Kojima et al. 1995; Krylova and Sahling 2010).

As one of the most diverse groups in the chemosynthetic ecosystems, more than 100 extant vesicomyid species have been recognized and described around the world oceans (Taylor and Glover 2010) with morphological and/or anatomical characters (Krylova and Sahling 2006; Krylova and Sahling 2010). However, taxonomy of this group is poorly understood at both species and supraspecific levels as the morphospecies are often incongruent with molecular assignments (Vrijenhoek et al. 1994; Kojima et al. 1995; Peek et al. 1997; von Cosel and Salas 2001; Goffredi et al. 2003; Kojima et al. 2006; von Cosel and Olu 2009; Krylova and Sahling 2010; Audzijonyte et al. 2012; Decker et al. 2012; Johnson et al. 2017). The taxonomic ambiguity might be attributed to both the existence of morphological convergence and plasticity among these clams. Different morphospecies from remote locations could be conspecific based on molecular data as found in some transoceanic species (Peek et al. 2000; Audzijonyte et al. 2012). On the other hand, cryptic species (i.e. indistinguishable morphology) were revealed in some cases as well. For example, the Vesicomya gigas/Calyptogena kilmeri complex was detected to comprise multiple evolutionary lineages (Peek et al. 1997). In addition, due to their remoteness, the vesicomyid specimens are often limited by availability, which could lead to the misidentification of widely distributed species. Despite several revision studies (von Cosel and Salas 2001; Krylova and Sahling 2006; Krylova and Sahling 2010; Johnson et al. 2017), species identification of the vesicomyid clams remains difficult. Along with more exploration of new sites and collection of new specimens, a fast and efficient molecular tool to identify species is required and necessary for this problematic bivalve group.
DNA barcoding uses a short standardized mitochondrial sequence (cytochrome c oxidase subunit I, COI for animals) to identify species (Hebert et al. 2003). The molecular tool has been widely used and proved useful in species identification, delimitation and discovery (Taylor and Harris 2012). Despite some crucial criticisms (Meyer and Paulay 2005; Knowles and Carstens 2007), DNA barcoding is much helpful in discrimination of morphologically indistinguishable taxon (Liu et al. 2011) or organisms at different life stages (Footit et al. 2009; Versteirt et al. 2015). Further, the standard system allows comparison of results between different studies, which could facilitate delimitation and discovery for species occupying wide ranges or difficult to obtain.

For vesicomyid species, mitochondrial COI sequences have been extensively used to evaluate phylogenetic relationships within and among species (Kojima et al. 1995; Peek et al. 1997; Goffredi et al. 2003; Kojima et al. 2006; Audzijonyte et al. 2012; Decker et al. 2012). However, most studies focused on regional fauna and limited sample sequences (Audzijonyte et al. 2012; Decker et al. 2012), and rare studies comprehensively investigated the utility of DNA barcoding on the identification of vesicomyid species. In this study, we evaluated the utility of DNA barcoding techniques in identifying the poorly resolved vesicomyid species. A comprehensive mitochondrial COI dataset of vesicomyids occurring worldwide were collected and analyzed. Findings of this study will supplement and accelerate revisions of the Vesicomyidae taxonomy, and thus aid species discovery for this deep-sea bivalve group.

Materials and methods

Data collection

For the DNA barcoding analyses, two datasets (mitochondrial COI and 16S rDNA) were achieved from this study and the GenBank database (Tables S1 and S2, Supporting information): the COI dataset involving 56 putative species (OTUs), and 16S dataset involving 19 species. Sequences from closely related veneroid species (Dreissena polymorpha, Gemma gemma and Mercenaria mercenaria) were used as outgroup based on previous phylogenetic studies (Mikkelsen et al. 2006; Bieler et al. 2014). In addition, sequences of mitochondrial (COI, 16S rRNA) and nuclear (18S rRNA, 28S rRNA, and histone H3) genes were also collected for a subsample of 14 vesicomyid and 18 veneroid species (Table S3, Supporting information).

Because barcode gap (i.e. interspecific divergence larger than intraspecific variation) could be affected by increasing sampling scale (Bergsten et al. 2012), or by biological processes such as interspecific hybridization or incomplete lineage sorting (Funk and Omland 2003). In this study, we collected COI sequences as many as possible to minimize sampling bias (Table S1 and Figure S1, Supporting information). Almost two-thirds of the species (n = 35) included two or more COI sequences with three on average (Table 1). Nearly one third of the species (n = 17) had specimens located in three or more different locations (Table 1). It is noticed that a few sequences submitted to the GenBank did not include location information, making them less useful for further studies.

DNA sequencing

Genomic DNA was extracted from the adductor muscles or foot using the TIANamp Marine Animals DNA kit (TIANGEN, Beijing, China) according to the manufacturer’s instructions. Partial sequences of COI, 16S, 18S, 28S and H3 genes were amplified by polymerase chain reaction (PCR). Because most of the samples failed using the universal COI primers (Folmer et al. 1994) in an initial PCR, a shorter COI sequence was amplified with vesicomyid specific primers (VesLCOI and VeshCO; Peek et al. 1997). The 50 μL PCR reaction was performed with 0.5 μM of each primer, 50–100 ng genomic DNA template and 1.25 U Takara Premix Taq (Takara, Dalian, China). Thermal conditions included 94 °C for 10 min, followed by 30 cycles of 94 °C for 45 s, annealing temperature for 45–60 s, 72 °C for 45–90 s and final extension at 72 °C for 10 min. For detailed primer sequences and annealing temperatures, see Table S4 (Supporting information). PCR products were sequenced in both directions on an ABI 3730x1 DNA analyzer using the BigDye Terminator v3.1 Cycle sequencing kit (Applied Biosystems, Inc., Foster City, CA, USA). Sequences were assembled and edited using the program Lasergene SeqMan v7.2 (DNASTAR, Inc., Madison, WI, USA). All new sequences were submitted to GenBank.

Data analysis

The protein-coding sequences (COI and H3) were aligned using TranslatorX program (Abascal et al. 2010) based on the amino acid translations. The rest gene sequences were aligned using the software Muscle (Edgar 2004). The best-fitting nucleotide substitution models were selected for each codon position of COI and H3, and the rest genes under the Akaike Information Criterion (AIC) within jModelTest v0.1 (Posada 2008).

DNA barcoding analyses were performed with COI (n = 170) and 16S (n = 76) datasets, respectively. Phylogenetic trees with Neighbor-joining (NJ) and Bayesian inference (BI) methods were generated. The GTR+G model was selected as the best-fitting nucleotide substitution model for the three codon positions of COI and 16S sequences. The NJ tree was constructed with TrN+G model and 1000 bootstrap replicates, as implemented in Mega v6.06 (Tamura et al. 2013). The BI method was carried out in MrBayes v.3.2 (Ronquist et al. 2012). Two Markov Chain Monte Carlo (MCMC) runs, each with six chains were performed for 15,000,000 generations with sampling trees every 1000 generations and a burn-in of 25%. Convergence diagnostics were checked with the ESS values (>200) in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). A majority-rule consensus tree was obtained with the posterior probability for each node. To compare our results with those in the previous studies, the putative species names were determined mainly according to Johnson et al. (2017) and Krylova et al. (2015). Those which could not be able to assign to a certain species were
### Table 1. K2P distances of mitochondrial COI within species.

<table>
<thead>
<tr>
<th>Species</th>
<th>No.</th>
<th>Mean (%)</th>
<th>Min (%)</th>
<th>Max (%)</th>
<th>Closest neighbor</th>
<th>D (%)</th>
<th>BIN</th>
<th>Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abyssogena mariana</td>
<td>2</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>Abyssogena southwardae</td>
<td>1.21</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Abyssogena phaseoliformis-1</td>
<td>4</td>
<td>1.04</td>
<td>0.40</td>
<td>1.42</td>
<td>Abyssogena phaseoliformis-2</td>
<td>2.24</td>
<td>WP, EP, WA, CA</td>
<td></td>
</tr>
<tr>
<td>Abyssogena phaseoliformis-2</td>
<td>2</td>
<td>1.21</td>
<td>1.21</td>
<td>1.21</td>
<td>Abyssogena phaseoliformis-1</td>
<td>2.24</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Abyssogena southwardae</td>
<td>7</td>
<td>0.73</td>
<td>0.00</td>
<td>1.41</td>
<td>Abyssogena mariana</td>
<td>1.21</td>
<td>WA, CA</td>
<td></td>
</tr>
<tr>
<td>Akebiconcha kawamurai</td>
<td>3</td>
<td>0.13</td>
<td>0.00</td>
<td>0.20</td>
<td>Phreagena okutanii</td>
<td>6.53</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Archisesia diagonalis-magnocellatus</td>
<td>4</td>
<td>0.50</td>
<td>0.20</td>
<td>0.80</td>
<td>AF008260</td>
<td>3.71</td>
<td>BOLD:ACF2739 WP</td>
<td></td>
</tr>
<tr>
<td>Archisesia fortunato</td>
<td>4</td>
<td>0.91</td>
<td>0.00</td>
<td>1.80</td>
<td>AB479083</td>
<td>5.17</td>
<td>WP, EA</td>
<td></td>
</tr>
<tr>
<td>Archisesia gigas</td>
<td>9</td>
<td>0.75</td>
<td>0.00</td>
<td>2.05</td>
<td>AF008260</td>
<td>2.44</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td><em>Archisesia</em> packardana</td>
<td>2</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>AB408260</td>
<td>7.52</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Archisesia sp.7</td>
<td>4</td>
<td>0.69</td>
<td>0.00</td>
<td>1.21</td>
<td>Archisesia diagonalis-magnocellatus</td>
<td>4.15</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Calyptogena costaricana</td>
<td>4</td>
<td>0.98</td>
<td>0.00</td>
<td>1.62</td>
<td>KC164258</td>
<td>2.24</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Calyptogena leptia</td>
<td>4</td>
<td>1.18</td>
<td>0.40</td>
<td>2.03</td>
<td>KC164258</td>
<td>3.28</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>“Calyptogena” magnifica</td>
<td>8</td>
<td>0.14</td>
<td>0.00</td>
<td>0.40</td>
<td>Vesicomya cordata</td>
<td>9.21</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Calyptogena pacifica</td>
<td>9</td>
<td>0.45</td>
<td>0.00</td>
<td>1.21</td>
<td>Calyptogena costaricana</td>
<td>7.90</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>“Cordata-group” ponderosa</td>
<td>4</td>
<td>0.10</td>
<td>0.00</td>
<td>0.20</td>
<td>Vesicomya stearnsi</td>
<td>7.21</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Calyptogena rectimargo-storobagotai</td>
<td>10</td>
<td>0.74</td>
<td>0.20</td>
<td>2.23</td>
<td>OTU4</td>
<td>2.23</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Calyptogena tuerkyai</td>
<td>2</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>Calyptogena tuerkyai</td>
<td>1.62</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Christineconcha regab</td>
<td>4</td>
<td>0.50</td>
<td>0.20</td>
<td>0.80</td>
<td>HE680074</td>
<td>6.57</td>
<td>EP</td>
<td></td>
</tr>
<tr>
<td>Ectenagena elongate</td>
<td>3</td>
<td>0.27</td>
<td>0.00</td>
<td>0.40</td>
<td>sp.2</td>
<td>9.44</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Isorropodon perplrcum</td>
<td>2</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>HE680074</td>
<td>7.43</td>
<td>EP</td>
<td></td>
</tr>
<tr>
<td>“Pheagena” extenta</td>
<td>7</td>
<td>0.44</td>
<td>0.00</td>
<td>1.01</td>
<td>Phreagena okutanii</td>
<td>5.21</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>“Pheagena” nankaiensis</td>
<td>4</td>
<td>1.05</td>
<td>0.00</td>
<td>1.83</td>
<td>AB110753</td>
<td>6.97</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Phreagena okutanii</td>
<td>5</td>
<td>0.72</td>
<td>0.00</td>
<td>1.21</td>
<td>Phreagena soyoae</td>
<td>3.71</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Phreagena soyoae</td>
<td>12</td>
<td>0.76</td>
<td>0.00</td>
<td>2.64</td>
<td>Phreagena okutanii</td>
<td>3.71</td>
<td>WP</td>
<td></td>
</tr>
<tr>
<td>Pliocardia sp.10 cf. venusta</td>
<td>3</td>
<td>1.91</td>
<td>0.60</td>
<td>2.66</td>
<td>AB110759</td>
<td>7.79</td>
<td>WP</td>
<td></td>
</tr>
</tbody>
</table>

**Note:**
- Closest neighbor species: Occurrence includes Western (WP) and Eastern (EP) Pacific, West (WA), Central (CA), and Eastern (EA) Atlantic and Indian Ocean (In).
- Species in bold are those with two BIN clusters.

**Number (No.) of samples; mean average (Mean); minimum (Min) and maximum (Max) distances in each species are shown.**
- D (%) represents minimum distance to the nearest neighbor species.
- Species with two or more specimens are recognized as operational taxonomic units (OTUs). Pairwise Kimura-2-parameter (K2P) distances were calculated within and between species using Mega. A local barcoding gap was tested by plotting the maximum intraspecific distance to the minimum interspecific distance (Collins and Cruickshank 2013) for species with two or more specimens.

For the subsamples, mitochondrial (COI + 16S) and nuclear (18S + 28S + H3) sequences were used to reconstruct phylogenetic relationships among the clams, respectively. BI trees were constructed with two MCMC runs each with four chains as implemented in MrBayes. Each run was conducted with 80,000,000 generations (sampling every 1000 generations) and a burnin of 25%. For the best-fitting model of each codon and genes, see Table S5 (Supporting information). Maximum likelihood (ML) method was conducted with RAXML (Stamatakis 2006) as run in the graphical program.
raxmlGUI v1.3.1 (Silvestro and Michalak 2012). ML with rapid bootstrap option was run with 1000 replicate bootstraps. GTR + G model was used for each partition.

**Results**

A total of 89 novel sequences (COI, 19; 16S, 19; 18S, 15; 28S, 19; H3, 17) were obtained from 19 vesicomyid specimens in this study (Table S3, Supporting information). The GenBank accession numbers are KT345510–KT345598. Due to the shortness of some sequences, the dataset of 170 COI sequences were finally trimmed to 501 bp. Both NJ and BI trees showed 56 terminal clades, that is putative species/OTUs (Figure 1). A unique OTU (OTU2) was first detected. Most of the terminal clades were supported by high NJ bootstrap/posterior probability (PP) values (99%/1). Despite that, the superspecies relationships were not well-resolved, seven major clades (Figure 1) were recognized in the genus level as proposed by Johnson et al. (2017) in both topologies. Among them, five were strongly supported (PP values ≥.96). Clade *Vesicomya*, assigned to the subfamily Vesicomyinae (Krylova et al. 2015), was placed in the basal position of the family, and divergent from all the rest clades assigned to the subfamily Pliocardinidae.

The average K2P distances within species were 0–1.91%. When excluding *Pliocardia* sp.10 cf. *venusta*, the values of average intraspecific divergence did not exceed 1.41%, lower than those of average interspecific divergence (1.65–29.64%). The minimum interspecific distances ranged from 1.21 to 23.24%, and the maximum intraspecific distances ranged from 0–2.66% (Table 1). Yet, local barcoding gap existed for 33 out of the 35 species with two or more specimens (Figure 2). The exceptions were *Abyssogena southwardae* and *Calyptogena rectimargo-starobogatovi*, whose minimum interspecific divergence was not greater than the maximum intraspecific divergence (Table 1). The distribution of pairwise intra- and inter-specific distances also revealed the existence of barcoding gap with minor overlaps (Figure 3). The overlaps occurred with distances ranging from 2 to 3%. The divergence value of 2.5% might be used as a threshold since 99.5% intra-specific pairwise distances were observed below this value. High-level intra-specific variation (>2%) most occurred in *Pheagena soyoae*, and several pairs within *Pliocardia* sp.10 cf. *venusta* and *C. rectimargo-starobogatovi*. Most low-level inter-specific variation (<2%) was mainly found between species among *Abyssogena*. The smallest inter-specific distance was 1.21% between *Abyssogena mariana* and *Abyssogena southwardae*.

It was noticed that average inter-specific distances between the two subfamilies were ≥20%, higher than the distances between species within each subfamily (<18%), which were also observed in the pairwise comparisons (Figure 3). The only exception was that *V. atlantica* (JX196991) was 23–25% divergent from *V. pacifica* and *Vesicomya* s.s. (KC429129) within the subfamily Vesicomyinae.

We also searched Vesicomyidae species in the barcode index number (BIN) system on BOLD v.3 (http://www.boldsystems.org), and found 44 of the 56 putative species (Table 1). Forty-one of them were found with unique BINs, and three (OTU1, *Abyssogena southwardae* and *Pliocardia* sp.10 cf. *venusta*) were with two BINs.

In the 16S dataset, 75 sequences trimmed to 442 bp were analyzed, and 19 putative species were identified, most with strong support (bootstrap/PP values: 97–99%/99–1; Figure 1). Average distances within species were 0–0.53%, and the maximum intra-specific distances ranged from 0–0.68%. No overlap was observed between maximum intra-specific and minimum inter-specific distances (Figure S2, Supporting information), suggesting the presence of the local barcoding gap. Meanwhile, pairwise comparisons also showed no overlap between intra- (maximum 0.69%) and inter-specific (minimum 0.92%) distances (Figure S3, Supporting information). Pairwise inter-specific distances between subfamilies (species with sequence KC429291 vs. the other species) ranged from 15.43 to 20.52%, higher than the values between species within subfamily (0–11.67%). Several conflicts were observed between COI and 16S clusters, that is a specimen fell into different COI and 16S clusters. For example, a specimen had COI sequence as AF008258 and 16S as AF115072. It was grouped into *Archivesica diagonalis-magnocultellus* in the COI topology while into OTU5 in the 16S tree. It might be caused by information errors of the samples since there was no evidence of mitochondrial recombination in Vesicomyidae. Alternatively, it might be a result of phylogenetic inaccuracy due to fewer number of samples or limited sequence variation (Heath et al. 2008).

Phylogenetic trees generated by mitochondrial COI +16S and nuclear 18S + 28S + H3 were shown in Figure S4 (Supporting information). The tree topologies were not identical in the two molecular datasets, but four major clades with strong support were recognized (cordata-group, gigas-group, *Calyptogena*, and *Vesicomya*) in both trees. The analyses supported monophyly of the family Vesicomyidae (Peek et al. 1997).

**Discussion**

Classification of many marine bivalve groups based on shell characters often challenges taxonomists and phylogeneticists because of absence of diagnostic morphological traits. The problem is especially severe in the deep-sea Vesicomyidae species, which are relatively new to science and whose taxonomy are poorly resolved. The present study assessed the utility of DNA barcoding in the species identification of vesicomyid species. By collecting a comprehensive barcoding dataset, we found that combination of tree- and distance-based barcoding analysis was useful to identify vesicomyid specimens and thus to complement the current taxonomy as discussed below.

**Vesicomyid species identification based on COI sequences**

The COI dataset revealed 56 putative species/OTUs in total (Figure 1), approximately accounting for half of the extant vesicomyid species. Analyses of maximum intra-specific distances vs. minimum inter-specific distances, and the
Figure 1. Neighbor-joining (NJ) trees derived from mitochondrial (a) COI and (b) 16S sequences. NJ bootstrap and Bayesian posterior probability values >80% are labeled above and below branches, respectively. The detailed sequence information of each terminal clade is listed in Tables S1 and S2 (Supporting information).
distribution of pairwise intra- vs. inter-specific distances, both suggested existence of barcoding gap with several exceptions. Furthermore, the value of 2.5% could be a potential threshold to identify most vesicomyid species. The divergence was generally congruent to results in previous studies, where variation within well-defined species (lineages) was usually less than 2%, and variation between species often larger than 3% (Baco et al. 1999; Goffredi et al. 2003; Audzijonyte et al. 2012).

The level of inter-specific variation was compatible to that of venerid species in a comprehensive barcoding study (mean 3.17%; Chen et al. 2011). In addition, the results were also supported by the BIN system on BOLD, where 41 species were detected with unique BINs (Table 1).

One of the key principles of DNA barcoding is that inter-specific divergence is larger than intra-specific variation (Hebert et al. 2003). It was often violated in closely related taxa (Meyer and Paulay 2005; Wiemers and Fiedler 2007). In this study, several closely related species exhibited low level of divergence between species, especially among species of Abyssogena (1.21–2.85%). The extreme example was Abyssogena southwardae, in that its maximum intra-specific variation (1.41%) was larger than the variation (1.21%) to its nearest neighbor Abyssogena mariana. Unexpectedly, all the seven samples of Abyssogena southwardae were collected in the Atlantic Ocean, whereas the two samples of Abyssogena mariana were from the West Pacific. This phenomenon implied that cryptic species might exist within Abyssogena southwardae or that Abyssogena mariana might be invalid.

On the other hand, high level of divergence within species was found in species Phreagena soyoae (maximum 2.64%), Calyptogena rectimargo-starobogatovi (maximum 2.23%) and Pliocardia sp.10 cf. venusta (maximum 2.66%). It is required to reexamine these species to clarify whether they are species complex or subdivided populations.

OTU1 included two species (“Phreagena similis, and an unidentified species) as reported previously (Kojima et al. 2004; Johnson et al. 2017). The two was confirmed by two BINs in this study (Table 1). However, they had a low value of intra-specific divergence as 1.41%, and the minimum inter-specific distance was five-fold higher than that value (7.18%).
As for OTU4, it had two species (Calyptogena makranensis and Calyptogena fausta) as in previous studies (Krylova et al. 2015; Johnson et al. 2017), but they were turned out as one BIN in this study. It also had a small intra-specific distance of 1.42%, lower than the distance to its nearest neighbor (1.62%). On the contrary, C. phaseoliformis considered as one species previously (Audzijonyte et al. 2012; Krylova et al. 2015) were divided into two putative species (C. phaseoliformis-1 and C. phaseoliformis-2) in our analysis, each represented by a unique BIN cluster. Controversial results had also emerged among previous studies (Peek et al. 2000; Audzijonyte et al. 2012; Decker et al. 2012). Classification discordance among these clades/subclades indicated potential cryptic diversity (Audzijonyte et al. 2012) and possible complex relationships between these sibling species. Evidence from more samples and nuclear markers and morphological characters is necessary to disentangle the genetic structures and evolutionary processes of these ambiguous lineages in order to determine their classification status and genetic diversity.

**COI vs 16 S DNA barcoding**

Mitochondrial COI sequences were initially advocated as barcode for animals (Hebert et al. 2003), while 16S was also used as a diagnostic barcode in some taxa due to difficulties in amplifying the COI fragment (slugs, Barr et al. 2009; sertulariids, Moura et al. 2011). In the deep-sea clams, fewer samples were sequenced with 16S. Despite fewer species recovered by the 16S compared to the COI sequences, it is not surprising to observe a similar tendency (Figures S2 and S3). Besides, no overlap was found in the 16S distances regarding maximum intra-specific vs. minimum inter-specific variation, or pairwise intra- and inter-specific variation. However, lower level of sequence variations in 16S suggested that COI was a better barcode for vesicomyid species. This phenomenon has been confirmed in some marine bivalves (Feng et al. 2011; Liu et al. 2011) and other animals (Xia et al. 2012).

**Implications of DNA barcoding in vesicomyidae taxonomy**

Although taxonomy and phylogeny of vesicomyid species have been investigated for decades (Boss 1970; Peek et al. 1997; Krylova and Sahling 2006; Audzijonyte et al. 2012; Krylova et al. 2015), their status and relationships remain confusing. In this study, specimens identified as Calyptogena elongata, C. magnifica, C. okutanii, “Cordata-group” ponderosa, Christineconcha regab, Isorropodon perlexum, Phreagena soyaee, V. cordata, Plicarcidia kuroshimana a priori, clustered into distinct COI lineages, suggesting that there were reliable diagnostic morphological characters to separate these species from the others. However, most putative species had sequences with different binomial names a priori or unidentified to species, suggesting that the morphological traits alone may not be enough to identify most vesicomyid species. Molecular data such as COI barcode would thus provide valuable information for the revision of the described vesicomyid species in that specimens from different locations and studies could be compared. It could also be used as an efficient tool to identify new samples of vesicomyids. In addition, unidentified clades (labeled with OTU or sp.) were determined in this and previous studies, indicating that these putative species might not yet be described, or they were not matched to the described species. In particular, a unique OTU (OTU2) was first sequenced and determined in this study, and may represent a new species. Morphology of these unrevised species should be scrutinized again to determine diagnostic characters. If not, the DNA variation might serve as diagnostic traits (Tautz et al. 2003). However, this should be interpreted with caution, because species delimitation using one locus (e.g. mitochondrial COI) has been proved problematic, especially for species under complex evolutionary histories (Funk and Omland 2003).

The genera of the current taxonomy are quite ambiguous, likely due to the absence of efficient diagnosable characters, and generally not well supported and resolved by molecular data in this and previous studies (Baco et al. 1999; Kojima et al. 2004; Decker et al. 2012). Therefore, the genus assignment was not discussed here. However, three clusters (gigas-group, cordata-group and Calyptogena) were resolved by both mitochondrial and nuclear datasets (Figures 1 and S2), indicating the genetic cohesion within each cluster.

Two subfamilies, Pliocardinae and Vesicomyinae have been proposed based on morphological characters (Krylova and Sahling 2010) and molecular COI sequences (Krylova et al. 2015). In this study, the mitochondrial (COI, 16S, and COI +16S) data conformed this hypothesis, showing two subfamilies with strong support (Figures 1 and S2). Subfamily Vesicomyinae (exclusive including Vesicomya) was placed on the basal position and diverged from the other subfamily. However, the nuclear genes (18S +28S +H3) appeared not to support the separation of the two subfamilies (Figure S2) in that the clade Vesicomya was clustered with Calyptogena. The incongruence was probably a result of phylogenetic inaccuracy, which was caused by alignment ambiguity with a quite limited number of nuclear sequences. Nevertheless, since the subfamily Vesicomyinae has been less studied (Krylova et al. 2015), more samples and nuclear evidences are needed to further confirm its status.

**Geological distribution and subdivision**

Although a small number of specimens in each species were collected in the DNA barcoding analysis, a comprehensive dataset of one taxa group could provide insights into their distribution and intra-specific subdivision as well. In this study, a global data collection of vesicomyids was applied. Among the species occupying two or more locations, it was observed that many vesicomyids had wide distribution ranges over thousands of kilometers (Figure S1). At one extreme case, four specimens clustered into Archivesica fortunata were located in the East and West Pacific and the Atlantic Oceans, and the East and West Pacific samples had the identical COI sequences. The transoceanic occurrence of vesicomyid species have been detected in previous studies (Kojima et al.
2004; Audzijonyte et al. 2012; Decker et al. 2012). The widespread distribution of many vesicomyid species indicated that these species may have quite high-dispersal abilities, not as restricted as previously thought (Kojima et al. 2004).

On the other hand, wide distribution range may lead to population subdivision because of geographic separation. For example, two COI sequences of Abyssogena phaseoliformis-1 collected from the northwest of the Pacific were divergent from the two from the northeast of the Pacific with pairwise K2P distances of 1.21–1.42%, while within each region, the samples were much more closely related (0.40–0.60%). In contrast, relative high level of intra-specific variation (0.60–2.66%, mean 1.91%) was detected within Pliocardia sp. 10 cf. venusta where three samples were collected from the same location (Table S1). These subdivisions indicated that populations of vesicomyid species might undergo complex divergence and evolutionary histories. More studies should be conducted to be used to explain these subdivisions.

Conclusion
This study represents one of the most comprehensive analyses on the molecular taxonomy of vesicomyid species. Mitochondrial COI barcoding analysis revealed 56 putative species and potential barcode gap, suggesting that the DNA barcoding combining phylogenetic and distance methods was useful in identifying vesicomyid species. Minor overlap observed between inter- and intra-specific distances in several cases implied hidden cryptic diversity or close relationships among them. COI barcode performed better than 16S when identifying vesicomyid species. Although limited samples were collected within each species, a comprehensive data collection in the barcoding analyses also gave insights into the distribution ranges and population subdivision in the vesicomyid species around the world. This study will provide important evidence to revise the taxonomy of the problematic deep-sea clams, and accelerate the discovery of new clam species in the future.

Acknowledgements
The authors thank Dr. Vrijenhoek for his kindness in providing clam samples and editorial comments for the manuscript.

Disclosure statement
The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Funding
This work was supported by National Natural Science Foundation of China [41576127], Strategic Priority Research Program of the Chinese Academy of Sciences [CAS] [XDB060010104], Knowledge Innovation Program of CAS [SIDSSE-201401], Hundred Talents Program of CAS [SIDSSE-BR-201401], and Major scientific and technological projects of Hainan Province, China [ZDKJ2016009].

References


