CrossMark

ORIGINAL ARTICLE

Pay Attention to the Overlooked Cryptic Diversity in Existing Barcoding Data: the Case of Mollusca with Character-Based DNA Barcoding

Shanmei Zou[1,2] · Qi Li[1]

Abstract With the global biodiversity crisis, DNA barcoding aims for fast species identification and cryptic species diversity revelation. For more than 10 years, large amounts of DNA barcode data have been accumulating in publicly available databases, most of which were conducted by distance or tree-building methods that have often been argued, especially for cryptic species revelation. In this context, overlooked cryptic diversity may exist in the available barcoding data. The character-based DNA barcoding, however, has a good chance for detecting the overlooked cryptic diversity. In this study, marine mollusk was as the ideal case for detecting the overlooked potential cryptic species from existing cytochrome *c* oxidase I (COI) sequences with character-based DNA barcode. A total of 1081 COI sequences of mollusks, belonging to 176 species of 25 families of Gastropoda, Cephalopoda, and Lamellibranchia, were conducted by character analysis. As a whole, the character-based barcoding results were consistent with previous distance and tree-building analysis for species discrimination. More importantly, quite a number of species analyzed were divided into distinct clades with unique diagnostical characters. Based on the concept of cryptic species revelation of character-based barcoding, these species divided into

✉ Qi Li
  qili66@ouc.edu.cn

1 Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, China

2 Jiangsu Provincial Key Laboratory of Marine Biology, College of Resources and Environmental Science, Nanjing Agricultural University, Nanjing 210095, China

separate taxonomic groups might be potential cryptic species. The detection of the overlooked potential cryptic diversity proves that the character-based barcoding mode possesses more advantages of revealing cryptic biodiversity. With the development of DNA barcoding, making the best use of barcoding data is worthy of our attention for species conservation.

Introduction

Due to the current biodiversity crisis at local, regional, to global scales, it is pressing to catalogue the species on earth (Novacek and Cleland 2001; Bellwood et al. 2004; Zou et al. 2012a, b; Ristau et al. 2013; Sosa et al. 2013; Hyde et al. 2014). In this sense, it is important that comprehensive species identification is firmly established. However, reliable identification and biodiversity monitoring of organism in the field is still a major challenge for traditional taxonomy (Hopkins and Freckleton 2002). It is particularly difficult for closely related species that lack obvious structural features. In this context, it is urgent to detect the cryptic biodiversity for species conservation.

The ambitious idea of using DNA barcoding for large-scale species identification and cryptic species revelation has been already a powerful tool for scientists, and the application of this standard technique seems promising in a range of fields (Hebert et al. 2003a, b; Waugh 2007; Ratnasingham and Hebert 2007; Bertolazzi et al. 2009; Wong et al. 2011; Dong et al. 2014). DNA barcoding is the application of short sequences of DNA to species discrimination across all forms of life (Hebert et al. 2003a, b). The use of a partial fragment

(approximately 648 bp from the 5/end) of cytochrome *c* oxidase I (COI) gene has been proposed to be a promising species-level marker (Hebert et al. 2003a, b) due to its high interspecies variation, low intraspecies variation, and universality of priming site for large taxonomic group (Folmer et al. 1994) and has gained enormous attention worldwide (Waugh 2007; Shneer 2009). However, one of the major issues concerning the inclusion of molecular information into taxonomic aspects of biology that has yet to be discussed in detail in the commentaries is concerning the best way to interpret the barcodes (Desalle et al. 2005; Desalle 2006). That is, the barcoding analytical methods may play an important role in species identification and new species discovery.

The conventional DNA barcoding approach relies on genetic distance for the species identification, which can fall into two broadly defined classes, distance, or phylogeny based. The "barcoding gap" in distance-based method and the monophyletic clades in phylogenetic tree-based method have been popularly used as the criterions for species identification of DNA barcoding. The barcoding gap relies on the separation of intra- and interspecific genetic variation, and the monophyletic clades in phylogenetic trees allow us to estimate ancestral states using information from extant taxa (Hebert et al. 2003a, b). However, the effectiveness of both of these criterions in cryptic species discovery is questionable (Raupach et al. 2014; Rougerie et al. 2014). The shortcoming of distance and tree-building barcoding is mainly related to the need for diagnostic characters that classical studies use to validate the existence of a species, the lack of an objective set of criteria to delineate taxa when using distances, and the fact that similarity scores often do not give the nearest neighbor as the closest relative (Desalle et al. 2005). One of the most essential arguments focuses on the so-called barcoding gap. Advocates of barcoding claim that interspecific genetic divergence exceeds intraspecific variation to such an extent that a clear gap exists which enables the assignment of unidentified individuals to their species with a negligible error rate. However, the barcoding gap between intra- and interspecific genetic variations in most organism groups often did not exist (Will and Rubinoff 2004; Rubinoff 2006; Rubinoff et al. 2006; DeSalle et al. 2005; Hickerson et al. 2006). Proponents of barcoding argue that the main reason for the absence of barcoding gap is the poor taxonomy of these groups, e.g., cryptic species may have been overlooked which are differentiated genetically but very similar or even identical in morphology. Another argument focusing on the tree-building method is that the gene trees, which are constructed based on the genetic distance, are not necessarily congruent with the "true" species trees, the reason for which may be due to the inability to identify morphological similar cryptic species (Hudson and Coyne 2002; Knowles and Carstens 2007; Kizirian and Donnelly 2004; Coissac et al. 2012).

An alternative to existing phenetic approaches, the character-based DNA barcoding (DeSalle et al. 2005), however, is consistent with and can serve as a complement to the approaches of traditional morphological identification systems since previously established taxonomic groups could be identified through the presence of diagnostic characters or combination of characters within short stretches of DNA sequences (Sarkar et al. 2002, 2008; Desalle et al. 2005; Rach et al. 2008). The character analysis is based on the fundamental concept that members of a given taxonomic group share attributes (e.g., polymorphisms) that are absent from comparable groups (Sarkar et al. 2002, 2008; Rach et al. 2008). In this sense, the diagnostic characters (termed CAs) among taxonomic groupings could be used for distinguishing species. For example, a threshold of more than three CAs is originally proposed for character-based DNA barcoding to separate the taxonomic groupings (Rach et al. 2008; Yassin et al. 2010; Zou et al. 2011, 2012b; Yu et al. 2015). That is, the taxonomic groupings are more easily distinguished when more CAs among them are detected. In this context, we could reveal some taxonomic groupings as potential cryptic species by comparing their unique combination of diagnostic characters with other taxonomic groups. The character-based DNA barcode has already been proved useful in species identification and discovery of cryptic species in a few previous studies (Zou et al. 2011, 2012b; Rach et al. 2008; Reid et al. 2011; Damm et al. 2010; Yassin et al. 2010; Goldstein and DeSalle 2010). However, it is not yet a commonplace in barcoding practice for both animals and plants.

DNA barcoding has been provided as a rapid and effective means to assess primary biodiversity for more than 10 years (Hebert et al. 2003a). Since then, large amounts of DNA barcode data have been accumulating in publicly available databases. As of October 2013, the Barcode of Life Data System (BOLD: www.boldsystems.org) (Ratnasingham and Hebert 2007) had a total of 192,350 species and 2,509,708 eukaryote specimens with DNA barcode data (Joly et al. 2014). Among the large amounts of barcoding data, most of them were conducted by distance and tree-building analysis. In this context, due to the shortcoming of both distance and tree-building barcoding methods, some potential cryptic species may be still undetected in the exiting barcoding data. Thus, the overlooked cryptic diversity needs to be further revealed by different barcoding modes, and the character-based DNA barcode would provide a good chance to detect them.

Mollusca is the second largest animal phylum in the earth, only next to the arthropod. There is extraordinary species diversity in Mollusca, especially for the marine mollusks. However, the identification of marine mollusks is often difficult due to the phenotypic plasticity and environment effects. In the past several years, marine Mollusca diversity was studied by DNA barcoding, particularly for the quantities of mollusks along the China coast, including the Gastropoda, Cephalopoda, and Lamellibranchia (Zou et al. 2011, 2012a,

b; Feng et al. 2011; Dai et al. 2012; Sun et al. 2012; Chen et al. 2011; Liu et al. 2011). These previous studies generally showed that COI gene was effective in species identification of marine Mollusca. Nevertheless, all the barcoding analysis was just based on tree-building or distance-based methods, in which sometimes the barcoding gap did not exist and the neighbor-joining (NJ) and Bayesian trees failed in identifying the closely related species. For example, Dai et al. (2012) found that high levels of genetic differentiation within *Loliolus beka* led to an overlap between intra- and interspecific distances, and some unknown species could not be determined to the species level in Chen et al. (2011). Thus, it is possible that a good deal of cryptic species is still not revealed in the existing barcoding data of mollusks. In this sense, Mollusca will provide an ideal case for detecting the overlooked cryptic diversity with a different barcoding mode.

In this paper, the character-based DNA barcoding was employed to recover the overlooked cryptic diversity in quantities of existing COI barcoding data of marine mollusks, in comparison with previous tree-building and distance-based barcoding results. Revelation of the overlooked cryptic diversity would caution us to correctly read the generated organismal barcode and make the best of the wealth of barcoding information to completely detect the cryptic biodiversity, which is significant to the species conservation.

Materials and Methods

Data Sets

A total of six COI data sets of mollusk samples from Feng et al. (2011), Dai et al. (2012), Sun et al. (2012), Chen et al. (2011), Liu et al. (2011), and Zou et al. (2012a) were analyzed based on character-based barcoding respectively, including Coleoidea (Mollusca: Cephalopoda), venerid species, Arcoida species (Bivalvia: Pteriomorphia), the pen shell *Atrina pectinata* (Bivalvia: Pinnidae), muricids (Gastropoda: Neogastropoda), and Caenogastropoda. They were from 1081 samples, belonging to 176 species of 86 genera of 25 families of mollusks. The COI sequences of each study of Feng et al. (2011), Dai et al. (2012), Sun et al. (2012), Chen et al. (2011), Liu et al. (2011), and Zou et al. (2012a) were obtained from GenBank as one data set, respectively. Then, the data sets were edited using the software program Sequencher 4.5 (Gene Codes Corporation, Ann Arbor, MI) and were aligned using MAFFT 6.717 (Katoh et al. 2009).

Character-Based Barcode Analysis

The character-based barcode analysis was conducted in characteristic attribute organization system (CAOS) and CAOS-Analyzer (http://bol.uvm.edu/caos-workbench/) (Sarkar et al.

2008; Bergmann et al. 2009). As described in detail previously (Zou et al. 2011), the neighbor-joining trees of COI gene loci were incorporated into NEXUS files with their DNA data matrix in MacClade v4.06 (Maddison and Maddison 2005). Then, the incorporated NEXUS data sets were carried out in CAOS system. The most variable sites that distinguish all the taxa were chosen, and the character states at these nucleotide positions were listed.

Sequences analyzed were deposited in GenBank (http://www.ncbi.nlm.nih.gov), corresponding to the following accession numbers: (1) HQ529502–HQ529542; HQ846076–HQ846163; JN315869–JN315874 (Dai et al. 2012); (2) HQ703031–HQ703342 (Chen et al. 2011); (3) HQ258822–HQ258880 (Feng et al. 2011); (4) HQ449254–Q449388 (Liu et al. 2011); (5) JF693339–JF693448 (Sun et al. 2012); and (6) GU188166–GU188271 (Zou et al. 2012a).

Results

In total, six data sets from 1081 COI sequences (about 658 bp) of 176 species of 86 genera of 25 families of mollusks were included in the character analysis, and their sequencing was conducted respectively. As described in detail previously (Rach et al. 2008; Zou et al. 2011), due to the high number of diagnostic characters (CAs), the particular nucleotide positions were chosen to distinguish the taxonomic groups. As expected, a good deal of mollusk species that could be potential cryptic species but were overlooked in previous distance- and monophyly-based barcoding analysis was further recovered by character-based barcoding in this study.

A total of 36 taxonomic groups of 31 Cephalopoda species (Dai et al. 2012) were analyzed by character analysis. It was showed that five species, *L. beka*, *Loliolus uyii*, *Euprymna morsei*, *Octopus minor*, and *Octopus* sp., clustered into two distinguishable monophyletic clades respectively in comparison with other well-supported monophyletic species (Fig. 1), which could either be potential cryptic species. Among the five species, *L. beka* has been already recovered as cryptic species in Dai et al. (2012) by monophyly- and distance-based barcoding analysis, which corresponded to the character results here that two distinct clades within *L. beka* were revealed with more CAs (Table 1). However, *L. uyii*, *E. morsei*, *O. minor*, and *O.* sp., which also could either be potential cryptic species, were not detected in their study. Here, it was clearly indicated that *L. uyii*, *E. morsei*, *O. minor*, and *O.* sp. were also separated as two distinct clades with at least three CAs for each clade in the character analysis (Table 1).

Sequencing of a total of 120 COI sequences of 17 known and easily confused muricid species from Zou et al. (2012a) was carried out by character analysis. The barcoding results based on monophyly- and distance-based barcoding methods in Zou et al. (2012a) demonstrated that COI gene was a

suitable marker for barcoding muricids, which can distinguish
all muricid species. However, no candidate cryptic species
were detected in this previous study. Here, it was showed that
three species, *Chicoreus torrefactus*, *Morula musiva*, and
Ergalatax margariticola, clustered into two distinguishable
clades respectively in COI NJ tree (Fig. 2), and the distin-
guishable clades also possessed a unique combination of char-
acter states with more than four CAs, in comparison with other
monophyletic species (Table 2).

The tree-building and distance barcoding analysis by Sun
et al. (2012) showed clear discrimination of 45 species of
Caenogastropoda. Similarly, the cryptic diversity was also
not involved in their study. Here, it was demonstrated that five
species which were recovered as two separate clades respec-
tively in COI NJ tree (Figure S1) were also divided into two
distinct clades with at least three CAs for each at 40 nucleotide
positions in character analysis (Table S1). These five species
might also be putative cryptic species.

For character-based barcoding of venerid species from
Chen et al. (2011), a total of 69 taxonomic groups were tested,
nine out of which clustered into two distinguishable mono-
phyletic clades respectively in COI phylogenetic trees
(Figure S2). These nine species were all divided into two clear
clades with more than three CAs in character analysis, includ-
ing five species (*Gafrarium dispar*, *Circe scripta*, *Meretrix
petechialis*, *Paphia gallus*, and *Periglypta puerpera*) which
were already recovered as putative hidden species by Chen
et al. (2011) (Figure S2, Table S2).

A total of 48 taxonomic groups of Arcoida species
(Bivalvia: Pteriomorphia) from Feng et al. (2011) were ana-
lyzed by character analysis (Figure S3). It was showed that
both *Scapharca broughtonii* and *Tegillarca granosa* clustered
into two distinguishable monophyletic clades in COI phylo-
genetic trees (Figure S3), as shown in Feng et al. (2011). This
was consistent with the character results here that two distin-
guishable monophyletic clades of *S. broughtonii* and
T. granosa in the COI tree both formed unique combinations
of character states with six and 16 CAs, respectively
(Table S3). Thus, these two species could either be potential
cryptic spices, which, however, were not involved as
crypticism in Feng et al. (2011).

Six taxonomic groups of *A. pectinata* were recovered as
potential cryptic species in Liu et al. (2011) with distance and
phylogenetic barcoding analysis of COI and nuclear ribosom-
al internal transcribed spacer (nrITS) genes, which were fur-
ther tested by character analysis in this study (Figure S4 and
Table S4). The character results indicated that four taxonomic
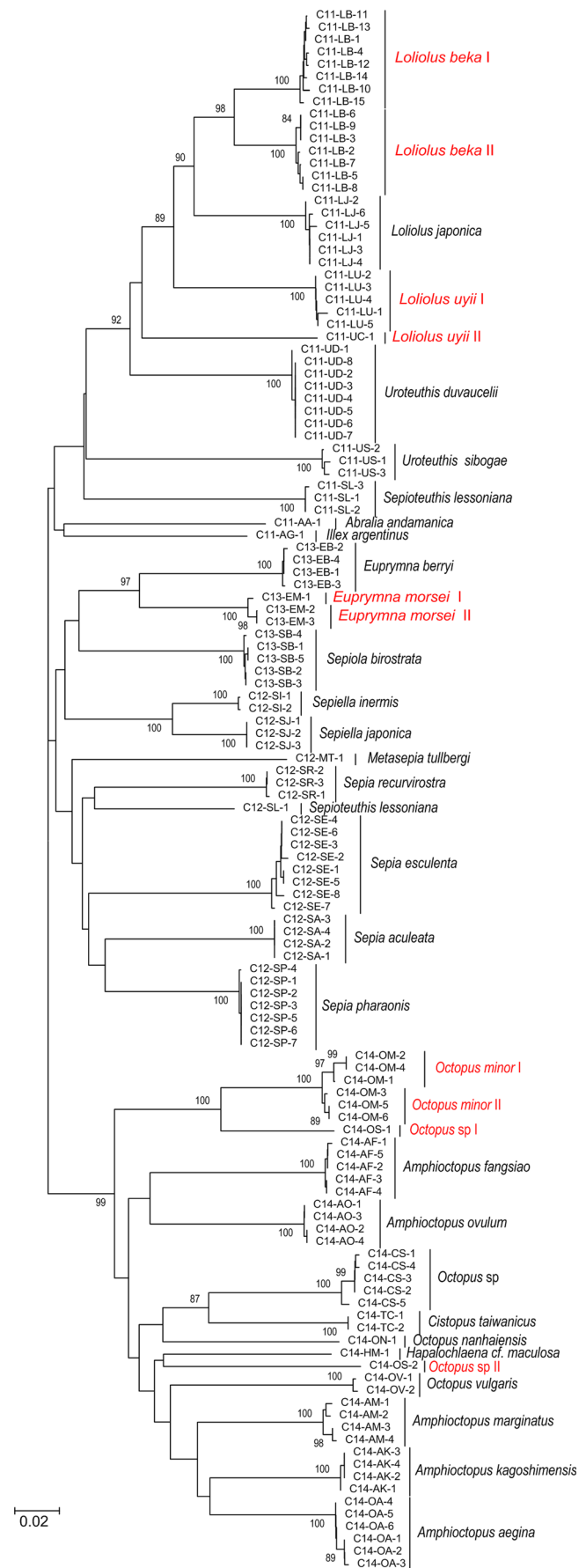groups of *A. pectinata* formed a unique combination of

Table 1 Character-based COI barcodes for 36 defined clades of Coleoidea (Mollusca: Cephalopoda) in Fig. 1

Taxa	Position																																					
	79	85	91	112	169	172	184	208	217	265	280	286	292	295	307	313	316	334	337	343	373	382	388	403	412	469	478	493	499	508	514	517	535	538	556	559	580	619
Abralia andamanica	T	A	T	C	T	T	A	A	A	A	A	A	A	T	A	A	A	T	A	T	A	C	A	A	A	A	T	T	T	A	A	A	A	A	A	A	T	A
Illex argentinus	G		A	T	T	T	G	A	A	A	A	A	A	T	A	A	G	G	A	T	A	T	A	A	A	A	T	T	A	T	T	A	T	A	A	T	A	T
Loliolus beka I	A	G	A	A	T	C	A	A	A	A	A	G	C	T	A	G	A	G	G	C	A	T	A	A	T	T	A	T	T	A/G	C	A	A	A	A	T	G
Loliolus beka II	A	A	A	A	T	G	A	A	A	A	A	A	C	A	A	G	A	A	T	A	T	A	A	T	C/T	A	T	T	T	A	C	A	A	A	T	G
Loliolus uyii I	A	A	G	A	T	T	T	A	A	A	A	A	C	T	A	A	A	A	T	G	A	T	C	A	A	A	T	T	A	G	A	A	C	G
Loliolus uyii II	T	A	A	A	C	T	G	A	A	A	A	A	A	T	A	A	A	T	C	A	A	C	A	T	T	A	A	T	T	A	G	A	A	T	C
Loliolus japonica	G	A	A	A	T	T	G	A	A	A	A	A	A	C	C	A	G	A	A	T	A	T	A	A	T	T	T	A	T	C	T	C	A	A	C	A	T	A
Uroteuthis duvaucelii	A	T	A	A	T	T	A	A	A	A	A	A	A	T	A	A	C	A	A	C	T	A	T	A	T	A	T	A	A	A
Uroteuthis sibogae	G	T	C	A	T	T	A	A	C	T	A	A	C	A	A	C	A	A	T	A	C	T	A	A	A	C	A	A	A	A	T	T
Sepioteuthis lessoniana	T	C	A	A	T	T	C	C	A	G	A	C	A	G	T	A	A	C	C	A	A	T	G	T	A	T	C	T	C	T	C	A	T	A	A	A
Sepia aculeata	C	A	T	C	A	T	G	A	T	T	A	A	C	A	A	A	G	T	T	T	A	A	A	A	A	T	A	A	A	T	G	A	A	A	A
Sepia esculenta	T	C	C	A	T	A	A	T	A	A	A	A	T	G	G	A	T	C	T	A	C	G	A	T	T	C	A	T/C	T	T	A	A	G	G	A	A
Sepioteuthis lessoniana	C	A	T	A	T	A	A	A	A	A	A	A	A	T	A	T	A	T	A	A	A	A	A	T	A	A	A	A	T	A	G	A	A
Sepia pharaonis	T	T	C	G	T	A	A	A	A	A	A	A	A	A	T	A	A	T	A	T	A	A	A	A	A	T	T	A	T	A	A	A	A
Sepia recurvirostra	C	A	T	A	T	T	A	A	A	A	A	A	A	T	A	T	A	A	T	A	A	A	C	T	A	T	A	A	A	T	G	A	C	A
Sepiella japonica	A	A	T	A	T	T	T	A	G	A	A	T	A	A	A	T	A	C	A	A	A	T	G	A	T	A	T	C	A	A	A	A
Sepiella inermis	A	A	T	A	T	T	A	A	A	A	T	A	G	A	A	T	A	A	A	T	A	A	A	T	T	A	A	C	A	A	A	T
Metasepia tullbergi	C	A	A	A	T	T	G	A	C	A	A	C	T	A	A	G	A	C	T	A	G	A	A	A	T	T	G	T	T	A	A	A	A	C
Sepiola birostrata	T	T	T	A	T	A	A	A	A	A	A	A	T	A	A	A	C	C	T	A	C	T	A	T	A	A	T	T	A	A	A	A	G	C
Euprymna berryi	T	T	A	A	C	A	A	A	G	G	A	C	G	T	T	A	A	T	A	C	G	A	G	A	T	C	A	A	T	A	A	A
Euprymna morsei I	T	T	A	A	G	A	A	A	A	A	C	C	A	A	T	G	C	T	A	C	G	A	T	G	T	A	A	A	G
Euprymna morsei II	T	T	A	A	C	A	A	A	A	A	C	C	A	A	T	G	C	T	A	C	G	A	T	A	T	A	A	A	G
Amphioctopus aegina	T	A	A	A	A	A	T	A	A	T	C	A	A	C	A	C	A	C	A	A	T	A	A	A	T	T	A	T
Amphioctopus fangsiao	T	A	T	A	T	A	T	A	T	C	A	A	A	C	A	C	A	C	T	A	A	A	T	C	A	A	T	T	T
Amphioctopus kagoshimensis	A	C	A	A	T	A	A	A	A	T	A	A	C	A	A	C	A	A	C	A	A	A	T	A	A	C	T	A
Amphioctopus marginatus	A	T	A	C	A	A	A	A	A	T	A	A	A	T	A	A	A	G	A	A	A	A	A	C	T	A
Amphioctopus ovulum	T	A	A	A	T	A	A	A	A	A	A	A	G	A	A	A	A	A	A	T	C	A	A	A	T
Cistopus taiwanicus	A	A	T	A	C	A	A	T	A	A	A	A	A	A	A	T	T	T	A	A	C	T	T
Cistopus sp.	A	A	T	A	A	A	A	T	A	A	A	A	A	G	A	C	A	A	T	T	A	A	C
Hapalochlaena cf. maculosa	A	A	C	A	T	A	A	A	C	A	T	A	A	C	T	T	C	T	T	A	A	G	T	A	T	A	A	T	A
Octopus minor I	T	A	A	T	T	A	A	A	C	T	A	A	C	A	C	A	C	A	T	T	T	A	T	T	A
Octopus minor II	T	A	A	T	T	A	A	A	C	T	G	T	A	A	C	A	C	A	T	T	A	C	A
Octopus nanhaiensis	T	G	T	A	C	T	C	A	A	C	A	A	G	A	A	C	C	A	A	T	A	C	A	G	A
Octopus vulgaris	A	C	A	A	A	A	C	T	T	T	A	C	C	T	G	A	A	A	A	A	C	T	T
Octopus sp 1.	T	A	A	T	A	A	A	T	A	A	G	A	T	T	A	C	A	A	C	T	A	T	C	A
Octopus sp 2.	A	A	T	A	A	A	A	A	A	T	A	T	A	T	T	A	T	A	A	A	A	A	T

Character states (nucleotides) at 38 selected positions of the COI gene region (ranging from position 79 to 619); taxa name according to Fig. 1; species showing cryptic diversity are marked in *red*

character states with more than ten CAs respectively (Table S4). Other two clades, *A. pectinata* III and IV, corresponding to lineages 5 and 6 in Liu et al. (2011), were recovered as two separate clades with only one distinct CA.

Discussion

DNA barcoding has been developed as an effective tool for fast species identification and new species revelation for more than 10 years (Hebert et al. 2003a; Damm et al. 2010; Stern et al. 2010; Schoch et al. 2012; Nevill et al. 2013; Kulsantiwong et al. 2014), one important goal of which is revealing cryptic diversity for species conservation. The major issue about successful application of DNA barcoding is concerning the best way to read the barcodes. However, the distance and tree-building methods, originally used for DNA barcoding, have often been argued, especially for the use in cryptic species revelation. On the other hand, presently, the large amounts of existing barcoding data for identifying animals, plants, and microorganism are generally conducted by distance or tree-building analysis. In this context, some

potential cryptic species may be still undetected in the scores of existing barcoding data. The character-based DNA barcoding, complementary to the traditional approaches, however, has a good chance for recovering the overlooked species diversity.

In this paper, Mollusca was employed as a case to test the performance of character-based DNA barcoding for revealing overlooked cryptic diversity in existing COI barcoding sequences, in comparison with traditional barcoding results. First of all, as a whole, the character-based barcoding results in this study were consistent with previous distance and tree-building analysis, in which the species studied were mostly clearly distinguished with unique attribute diagnostical characters. More importantly, the overlooked cryptic diversity was recovered by character-based barcoding. For example, among the species of Cephalopoda, Bivalvia, and Gastropoda from Feng et al. (2011), Dai et al. (2012), Sun et al. (2012), Chen et al. (2011), Liu et al. (2011), and Zou et al. (2012a), quite a number of them were divided into distinct clades with unique diagnostical characters in character-based analysis. Based on the concept of cryptic species revelation of character-based barcoding, these distinct clades could be putative cryptic
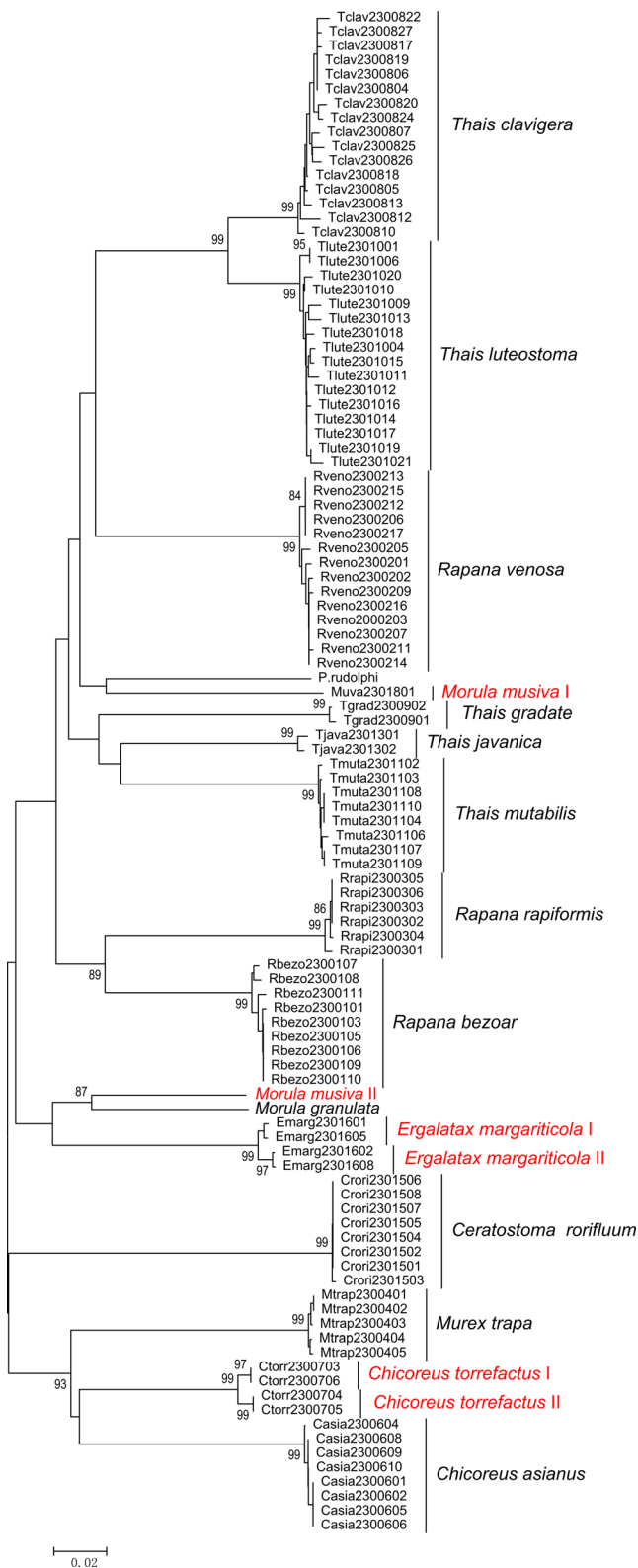
Fig. 2 COI neighbor-joining tree of muricid species. Taxonomic groups in *red* might be potential cryptic species and would be analyzed by character-based barcoding

results. Due to the large amount of sequence data and the lack of morphological characters, presently, we cannot assign these putative cryptic species as real new species. However, the overlooked cryptic species diversity detected in this study could cause us to protect biological diversity.

Dai et al. (2012) used two mitochondrial fragments, COI and 16S, to assess whether 34 coleoids accounting for about one third of the Chinese coleoid fauna could be identified by DNA barcoding based on distance and phylogeny analysis. Their barcoding results showed existence of overlap between COI intra- and interspecific divergences and revealed one cryptic species *L. beka*. In this study, except *L. beka*, several other species, *L. uyii*, *E. morsei*, *O. minor*, and *O.* sp., were also divided into two separate clades with more than three CAs in character analysis; particularly, two clades of *L. uyii* were recovered with nine CAs. These four species thus could be as potential cryptic species, but were undetected by Dai et al. (2012). Similarly, barcoding overlap between COI intra- and interspecific variation also existed in Arcoida species (Feng et al. 2011). The COI phylogenetic trees in Feng et al. (2011) demonstrated that all species analyzed fell into reciprocally monophyletic clades with high bootstrap values. However, no cryptic species were detected in their study. In this paper, the character-based barcoding analysis of COI data showed that two separate clades within *S. broughtonii* and *T. granosa* both possessed unique combinations of character states with more than six CAs, which could also be hidden cryptic species but were overlooked by Feng et al. (2011). Chen et al. (2011) employed DNA barcoding to reveal the species boundaries of venerid species based on distance and tree-building analysis, in which nearly all individuals identified to species level based on morphological traits possessed distinct barcode clusters except for some specimens, and five species, *M. petechialis*, *C. scripta*, *G. dispar*, *P. gallus*, and *P. puerpera*, were recovered as putative hidden species. Here, in the character analysis, a total of ten species, including the above five putative hidden species already recovered, were all divided into more than two separate clades with distinct diagnostical characters; particularly, *Ruditapes philippinarum* was recovered as four separate clades. Thus, these five additional species, *Paphia semirugata*, *Calyptraea chinensis*, *Gibbula tumida*, *Gafrarium divaricatum*, and *R. philippinarum*, that were not detected by distance and tree-building analysis in Chen et al. (2011) were recovered as potential cryptic species in this study. Based on tree-building trees and genetic distance of COI and nrITS molecular data, Liu et al. (2011) revealed five cryptic species within *A. pectinata*, which corresponded with the character analysis in this study where the five cryptic lineages were also showed as separate clades with more than

species. However, most of them were not detected in distance and tree-building analysis of previous barcoding

Table 2 Character-based COI barcodes for 19 defined clades of Muricids (Gastropoda: Neogastropoda) in Fig. 2

Taxa	Position																															
	79	91	97	104	127	130	160	205	212	238	241	244	310	334	343	364	370	379	401	403	409	451	454	469	493	514	517	535	550	556	568	586
Rapana bezoar	T	A	T	G	A	A	A	T	T/C	T		C	T	A	T	A	T	T	T	T	A	A	A	T	T/C	A	T	G	T	T	T	A
Rapana venosa	A	G	A	G	T	T	A	T	C	T	A/G	A	A	T	T	T	T	T	T	A	A	T	A	T	G	A	T	T	A	T	T	G
Rapana rapiformis	T	A	A	G	A	A	A	T	T	T	C	T	A	G	A	C	T	A	C	A	A	G	T	T	A	T	G	T	T	A	C	G
Murex trapa	G	T	G	G	A	T	G	C	T	T	A	A	A	A	T	A	T	T	T	A	A	A	C	G	T	T	A	G	A	T	A	T
Chicoreus asianius	G	T	A	G	C	T	G/A	T	T	A	A	A	A	T	T	C	T	T	A	G	A	T	A	T	C	T	A	A	T	T	T	T
Chicoreus torrefactus I	A	T	A	G	T	C	A	T	C	T		A	A	A	T	T	T	T	A	A	T	A	T	T	A	G	A	T	T	T	T	T
Chicoreus torrefactus II	A	T	A	G	T	C	A	T		C	A	G	A	T	C	T	T	A	A	T	A	T	T	A	G	A	T	T	T	T	T	T
Thais clavigera	A	T	G	G	T	A	A	T	T	T		A	A	T	T/C	A	T	A	A	A	C	T	A	T	G	T	T	G	T	A	G	A
Thais gradate	A	G	G	G	C	A	A	G	C	T	G	T	A	C	G	T	C	G	C	A	G	G	G	T	C	G	T	C	T	A	T	A
Thais javanica	G	G	A	G	T	G	A	A	T	T	A	A	A	T	T	G	C	T	G	A	A	T	C	G	T	G	T	G	T	C	T	G
Thais luteostoma	A	T	A/G	G	T	A	G	T	T	T	T	A	G	G	T/C	T	A	T	A	A	A	C	T	A	T	A	T	T	T	G	T	A
Thais mutabilis	A	C	A	G	T	A	A	T	T	T	A	A	C	T	T/C	A	A	C	A	A	A	T	T	T	T	A	T	T	T	T	T	T
Purpura rudolphi	A	T	A	G	A	A	G	T	T	A	A	A	A	T	G	A	T	A	A	A	T	A	T	A	T	A	T	A	A	T	A	A
Ceratostoma rorifluum	A	T	T	G	G	A	T	T	T	A	G	G	G	T	C	C	T	T	G	A	T	T	T	A	A	G	T	G	T	G	T	T
Morula musiva I	G	A	A	G	A	A	A	A	T	A	T	T	T	T	G	G	T	A	A	T	T	A	T	T	T	T	G	A	G			G
Morula musiva II	A	T	A	G	G	A	A	T	T	T	A	T	T	T	T	C	C	A	A	C	A	T	T	C	A	T	T	T	A			A
Morula granulata	G	T	A	G	A	A	A	T	T	A	T	T	C	T	C	T	A	A	A	G	T	T	T	A	A	T	T	G				G
Ergalatax margariticola I	C	C	G	G	A	A	G	T	T	T	G	T	A	T	T	T	A	A	A	T	G	T	C	T	A	A	A	T	C			C
Ergalatax margariticola II	C	T	A	G	A	A	G	T	C	T	G	T	A	T	T	T	T	T	A	G	A	T	G	T	C	T	A	A	A	T		C

Character states (nucleotides) at 32 selected positions of the COI gene region (ranging from position 79 to 586); taxa name according to Fig. 2; species showing cryptic diversity are marked in *red*

three CAs. For the taxonomic groups of Gastropoda, Sun et al. (2012) and Zou et al. (2012a) both found that DNA barcoding was effective in discriminating the species of Caenogastropoda and Neogastropoda with distance and tree-building barcoding methods. However, their studies did not involve the revelation of cryptic species. In the present study, character-based analysis of COI sequences of Caenogastropoda and Neogastropoda revealed some species that formed two distinct clades with more CAs and could be putative hidden species, which needs to be paid attention to in a further study.

With the existence of more and more barcoding data, it was once asked how to use this wealth of information (Rubinoff et al. 2006). Since DNA barcoding aims for species identification and new species discovery, it is urgent to make the best use of large numbers of existing barcoding sequences for recovering cryptic biodiversity. The criterion for species delimitation plays an important role in the use of DNA barcoding, including the new species discovery, e.g., the "10×rule" threshold, original criterion for species identification proposed by Hebert et al. (2003a), and any threshold of genetic distances (Blaxter et al. 2005; Jones et al. 2011), which, however, has often been argued. With the development of DNA barcoding, it is proposed that improved species-level phylogenetic trees in combination with new statistical methods will greatly improve our understanding of the biodiversity patterns (FitzJohn 2010; Joly et al. 2014). In this context, a combination of multiple DNA barcoding approaches may be more

effective to reveal cryptic biodiversity. It has been proposed that an optimal path to understand species boundaries is starting with a tree or distance-monophyletic population clusters (Hamilton et al. 2014). Then, the character-based approach is employed to confirm the initial identification. In this study, the character-based DNA barcoding recovered quite a few of species that could be potential cryptic species from the quantities of existing mollusk barcoding data, but were overlooked in previous distance and tree-building analysis. These results provide the evidence that the character-based DNA barcoding possesses the advantages of revealing overlooked cryptic biodiversity in traditional barcoding analysis. The revelation of overlooked cryptic diversity is significant to species conservation. On the other hand, with the development of DNA barcoding, making the best use of barcoding data is also worthy of our attention for better understanding of global biodiversity.

References

Bellwood DR, Hughes TP, Folke C, Nystrom M (2004) Confronting the coral reef crisis. Nature 429:827–833

Bergmann T, Hadrys H, Breves G, Schierwater B (2009) Character-based DNA barcoding: a superior tool for species classification. Berl Munch Tierarztl Wochenschr 122:446–450

Bertolazzi P, Felici G, Weitschek E (2009) Learning to classify species with barcodes. BMC Bioinformatics 10(Suppl14):S7

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C et al (2005) Defining operational taxonomy units using DNA barcode data. Philos Trans R Soc B 360:1935–1943

Chen J, Li Q, Kong L, Yu H (2011) How DNA barcodes complement taxonomy and explore species diversity: the case study of a poorly understood marine fauna. PLoS One 6(6):e21326

Coissac E, Riaz T, Puillandre N (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. Mol Ecol 21:1834–1847

Dai L, Zheng X, Kong L, Li Q (2012) DNA barcoding analysis of Coleoidea (Mollusca: Cephalopoda) from Chinese waters. Mol Ecol Resour 12:437–447

Damm S, Schierwater B, Hadrys H (2010) An integrative approach to species discovery in odonates: from character-based DNA barcoding to ecology. Mol Ecol 19:3881–3893

Desalle R (2006) Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. Conserv Biol 20:1545–1547

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philos Trans R Soc B 360:1905–1916

Dong W, Cheng T, Li C (2014) Discriminating plants using the DNA barcode *rbcL*b: an appraisal based on a large data set. Mol Ecol Resour 14(2):336–343

Feng Y, Li Q, Kong LF, Zheng X (2011) COI-based DNA barcoding of Arcoida species (Bivalvia: Pteriomorphia) along the coast of China. Mol Ecol Resour 11:435–441

FitzJohn RG (2010) Quantitative traits and diversification. Syst Biol 59:619–633

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3:294–299

Goldstein PZ, DeSalle R (2010) Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. BioEssays 33:135–147

Hamilton CA, Hendrixson BE, Brewer MS et al (2014) An evaluation of sampling effects on multiple DNA barcoding methods leads to an integrative approach for delimiting species: a case study of the North American tarantula genus *Aphonopelma* (Araneae, Mygalomorphae, Theraphosidae). Mol Phylogenet Evol 71:79–93

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003a) Biological identifications through DNA barcodes. Proc R Soc Lond Ser B 270:313–321

Hebert PDN, Ratnasingham S, DeWaard JR (2003b) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proc R Soc Lond Ser B 270:S96–S99

Hickerson MJ, Meyer CP, Moritz C (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. Syst Biol 55:729–739

Hopkins GW, Freckleton RP (2002) Declines in the numbers of amateur and professional taxonomists: implications for conservation. Anim Conserv 5:245–249

Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. Evolution 56:1557–1565

Hyde JR, Underkoffler KE, Sundberg MA (2014) DNA barcoding provides support for a cryptic species complex within the globally distributed and fishery important opah (*Lampris guttatus*). Mol Ecol Resour 14:1239–1247

Joly S, Davies TJ, Archambault A, Bruneau A, Derry A et al (2014) Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. Mol Ecol Resour 14:221–232

Jones M, Ghoorah A, Blaxter M (2011) jMOTU and Taxonerator: turning DNA barcode sequences into annotated operational taxonomic units. PLoS One 6:e19259

Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol 537:39–64

Kizirian D, Donnelly MA (2004) The criterion of reciprocal monophyly and classification of nested diversity at the species level. Mol Phylogenet Evol 32:1072–1076

Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. Syst Biol 56:887–895

Kulsantiwong J, Prasopdee S, Ruangsittichai J, Ruangjirachuporn W, Boonmars T et al (2014) DNA barcode identification of freshwater snails in the family Bithyniidae from Thailand. PLoS One 8(11):e79144

Liu J, Kong L, Zheng X (2011) Cryptic diversity in the pen shell *Atrina pectinata* (Bivalvia: Pinnidae): high divergence and hybridization revealed by molecular and morphological data. Mol Ecol 20:4332–4345

Maddison WP, Maddison DR (2005) MACCLADE: analysis of phylogeny and character evolution. Version 3.0. Sinauer Associates, Sunderland

Nevill PG, Wallace MJ, Miller JT, Krauss SL (2013) DNA barcoding for conservation, seed banking and ecological restoration of Acacia in the Midwest of Western Australia. Mol Ecol Resour 13:1033–1042

Novacek MJ, Cleland EE (2001) The current biodiversity extinction event: scenarios for mitigation and recovery. Proc Natl Acad Sci U S A 98:5466–5470

Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H (2008) Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. Proc R Soc Lond Ser B 275:237–247

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (www.barcodinglife.org). Mol Ecol Notes 7:355–364

Raupach MJ, Hendrich L, Küchler SM, Deister F, Morinière J (2014) Building-up of a DNA barcode library for true bugs (Insecta: Hemiptera: Heteroptera) of Germany reveals taxonomic uncertainties and surprises. PLoS One 9(9):e106940

Reid BN, Le M, McCord WP et al (2011) Comparing and combining distance-based and character-based approaches for barcoding turtles. Mol Ecol Resour 11:956–967

Ristau K, Steinfartz S, Traunspurger W (2013) First evidence of cryptic species diversity and significant population structure in a widespread freshwater nematode morphospecies (*Tobrilus gracilis*). Mol Ecol 22:4562–4575

Rougerie R, Kitching IJ, Haxaire J, Miller SE, Hausmann A (2014) Australian Sphingidae—DNA barcodes challenge current species boundaries and distributions. PLoS One 9(7):e101108

Rubinoff D (2006) Utility of mitochondrial DNA barcodes in species conservation. Conserv Biol 20:1026–1033

Rubinoff D, Cameron S, Will K (2006) A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. J Hered 97:581–594

Sarkar IN, Thornton JW, Planet PJ et al (2002) An automated phylogenetic key for classifying homeoboxes. Mol Phylogenet Evol 24:388–399

Sarkar IN, Planet PJ, Desalle R (2008) CAOS software for use in character-based DNA barcoding. Mol Ecol Resour 8:1256–1259

Schoch CL, Seifertb KA, Huhndorfc S, Robertd V, Spougea JL et al (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc R Soc Lond Ser B 109(16):6241–6246

Shneer VS (2009) DNA barcoding of animal and plant species as an approach for their molecular identification and describing of diversity. Zh Obshch Biol 70:296–315

Sosa V, Mejía-Saules T, Cuéllar MA, Vovides AP (2013) DNA barcoding in endangered Mesoamerican groups of plants. Bot Rev 79:469–482

Stern RF, Horak A, Andrew RL (2010) Environmental barcoding reveals massive dinoflagellate diversity in marine environments. PLoS One 5(11):e13991

Sun Y, Li Q, Kong L, Zheng X (2012) DNA barcoding of Caenogastropoda along coast of China based on the COI gene. Mol Ecol Resour 12:209–218

Waugh J (2007) DNA barcoding in animal species: progress, potential and pitfalls. BioEssays 29(2):188–197

Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20:47–55

Wong LL, Peatman E, Lu J et al (2011) DNA barcoding of catfish: species authentication and phylogenetic assessment. PLoS One 6(3):e17812

Yassin A, Markow TA, Narechania A, OGrad PM, DeSalle R (2010) The genus Drosophila as a model for testing tree- and character-based methods of species identification using DNA barcoding. Mol Phylogenet Evol 57:509–517

Yu Z, Li Q, Kong L, Yu H (2015) Utility of DNA barcoding for Tellinoidea: a comparison of distance, coalescent and character-based methods on multiple genes. Mar Biotechnol 17(1):55–65

Zou S, Li Q, Kong L, Yu H, Zheng X (2011) Comparing the usefulness of distance, monophyly and character-based DNA barcoding methods in species identification: a case study of Neogastropoda. PLoS One 6:e26619

Zou S, Li Q, Kong L (2012a) Multigene barcoding and phylogeny of geographically widespread muricids (Gastropoda: Neogastropoda) along the coast of China. Mar Biotechnol 14(1):21–34

Zou S, Li Q, Kong L (2012b) Monophyly, distance and character-based multigene barcoding reveal extraordinary cryptic diversity in Nassarius: a complex and dangerous community. PLoS One 7(10):e47276