Mitochondrial DNA

http://informahealthcare.com/mdn
ISSN: 1940-1736 (print), 1940-1744 (electronic)

Mitochondrial DNA, Early Online: 1–4
 DOI: 10.3109/19401736.2014.945579

informa
healthcare

SHORT COMMUNICATION

Evaluation of the efficacy of twelve mitochondrial protein-coding genes as barcodes for mollusk DNA barcoding

Hong Yu, Lingfeng Kong, and Qi Li

The Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao, China

Abstract

In this study, we evaluated the efficacy of 12 mitochondrial protein-coding genes from 238 mitochondrial genomes of 140 molluscan species as potential DNA barcodes for mollusks. Three barcoding methods (distance, monophyly and character-based methods) were used in species identification. The species recovery rates based on genetic distances for the 12 genes ranged from 70.83 to 83.33%. There were no significant differences in intra- or interspecific variability among the 12 genes. The monophyly and character-based methods provided higher resolution than the distance-based method in species delimitation. Especially in closely related taxa, the character-based method showed some advantages. The results suggested that besides the standard COI barcode, other 11 mitochondrial protein-coding genes could also be potentially used as a molecular diagnostic for molluscan species discrimination. Our results also showed that the combination of mitochondrial genes did not enhance the efficacy for species identification and a single mitochondrial gene would be fully competent.

Keywords

DNA barcoding, mitochondrial, mollusks

History

Received 22 April 2014
Revised 16 June 2014
Accepted 27 June 2014
Published online 11 August 2014

Introduction

DNA barcoding is the derivation of short DNA sequence (s) that enables species identification, closely related species discrimination, and discovery of cryptic or new species (Hebert et al., 2003). Since Hebert et al. (2003) proposed that the mitochondrial (mt) gene, COI could serve as a genetic barcode for all animal life, the implementation of the idea has seen a rapid rise. At present, DNA barcoding is a global enterprise, attracting large amounts of funding (Taylor & Harris, 2012), and applying to animals, plants, fungi and protists. As a species identifier, DNA barcoding has been used to handle a wide variety of problems, from conservation of biodiversity to food safety (Bucklin et al., 2011).

The efficacy of DNA barcoding primarily depends on the choice of suitable genetic markers. For metazoan barcode genes, mitochondrial protein coding genes (PCGs) were good candidates with numerous advantages: lack of introns and indels, rapid evolution, limited exposure to recombination, and haploid character. At present, a 658-bp region at the 5′ end of the COI gene proposed by Hebert et al. (2003) is regarded as the default DNA barcode region for most animal groups. This universal COI barcode has been highly successful in species identification through many studies of diverse metazoans (Bucklin et al., 2011; Feng et al., 2011; Hebert et al., 2004). COI was initially chosen as the barcode by Hebert et al. (2003) due to two important advantages: the universal primers and the great range of phylogenetic signal. However, along with the increase in the number of investigated species, the COI barcode has failed to deliver the reliable DNA barcode in some animal groups. For example, in insects, gastropods and amphibians, intraspecific variation in COI is high and usually overlaps with interspecific variation (Davison et al., 2009; Meier et al., 2006). In addition, the universal primers (Folmer et al., 1994) often failed to amplify the fragment of COI barcode and other primers were needed (Chen et al., 2011; Hoareau and Boissin, 2010; Lohman et al., 2009; Zou et al., 2012). Therefore, it is necessary to search for alternative DNA barcodes to avoid an exclusive reliance on COI. Actually, other gene regions have proved to have potential too. In insects, the mt ND1 gene region has proved to be another suitable marker especially for the identification of lower level taxonomic entities (Bergmann et al., 2013). In mammals, the estimated variability in ND1 is slightly higher than that in COI (Saccone et al., 1999). The species-recovery rate of COI in eutherian mammals is similar to those of other 11 mt genes (Luo et al., 2011).

Mollusca as the second largest animal phylum are a highly diverse group, with up to 200,000 species occurring worldwide (Collen et al., 2012). Morphological variability is a common characteristic of mollusks and species identification of mollusks is often difficult. DNA barcoding provides an ideal opportunity to offer fresh insights into the taxonomy and biodiversity of mollusks. To date, DNA barcoding has been successfully used in species discrimination of many molluscan groups (Feng et al., 2011; Mikkelsen et al., 2007). Most DNA barcoding of mollusks used the COI barcode. However, the overlap between intra- and interspecific variability, and primer universality were still problematic for some species (Chen et al., 2011; Davison et al., 2009; Zou et al., 2012). Other mt PCGs have never been used or evaluated for their barcoding utility in mollusks, and there is no evidence to prove that COI gene is a better identifier than other mt PCGs in mollusks.

Correspondence: Q. Li, Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, China. Tel: +8653282031622. Fax: +8653282031765. E-mail: qili66@ouc.edu.cn

In this study, we evaluated the efficacy of each 12 mt PCGs as the potential barcodes in mollusks to explore alternative mitochondrial barcoding regions. Three barcoding methods were used in species identification. We also investigated whether multiple mt PCGs performed better than a single mt PCG in DNA barcoding analyses.

Methods

Raw mitogenomes from 273 molluscan specimens available in GenBank were downloaded. Some molluscan species possess doubly uniparental inheritance of mtDNA, of which only F mitogenomes were included for subsequent analyses. The mitogenome data of *Platynereis dumerilii* was used as the outgroup in subsequent monophyly-based evaluation.

Genome sequences that contain large ambiguous regions in PCGs were removed. A total of 238 mt genomes of 140 molluscan species were obtained (see details on http://pan.baidu.com/s/1mg3EzVi). The ATP8 gene was excluded because many bivalve species lack ATP8. The nucleotide sequences of 12 PCGs were aligned using Clustal Omega (Sievers et al., 2011). The aligned sequences of 12 PCGs were concatenated in the same order for each species, generating the combination of 12 PCGs termed as the ''genome profile'' in this study.

For distance analyses, pairwise sequence divergences of the 13 datasets were calculated using a Kimura 2-parameter (K2P) distance model and analyzed at species and genus level in MEGA 5.05 (Kumar et al., 2008). Statistical significance of differences among different datasets respectively was estimated by nonparametric tests using IBM SPSS Statistics 19. Neighbour-joining (NJ) analyses were conducted independently for the 13 datasets using the K2P model with a bootstrap support analysis (1000 replicates) in MEGA 5.05, and discrete clades on a phylogenetic tree are required for the recovery of species. The characteristic attribute organization system (CAOS) (Sarkar et al., 2008) was used for the character-based method at species level. The CAOS algorithm identified diagnostic characters, termed ''characteristic attributes'' (CAs), for all clades at each branching node within the given guide tree. The program MacClade (Maddison & Maddison, 2000) was used to produce the nexus files for P-Gnome.

Results

Distance-based analyses

Most species showed low intraspecific K2P distances in the 13 datasets (<3%, data available on http://pan.baidu.com/s/1mg3EzVi), resulting in similar mean distances for each of the 13 datasets (1%–2.05%, Figure 1). However, a high level of intraspecific variation was observed in two species (*Oncomelania hupensis* and *Sthenoteuthis oualaniensis*) in all the 13 datasets, with mean intraspecific K2P distances greater than 5%. The largest intraspecific K2P distances within the 13 datasets were observed in *O. hupensis* (13.78–18.61%). Although the mean intraspecific distances for the 13 datasets differed slightly from each other (Figure 1), no significant difference in the mean intraspecific distances was detected among the 13 datasets or between any pairwise comparisons ($p > 0.05$). Mean interspecific distances were more than 3% within most genera, with some exceptions including three genera *Crassostrea* (Bivalvia, Ostreoida, Ostreidae), *Meretrix* (Bivalvia, Veneroida, Veneridae) and *Mytilus* (Bivalvia, Mytiloida, Mytilidae). The interspecific distance between *Meretrix meretrix* and *Meretrix petechialis* were

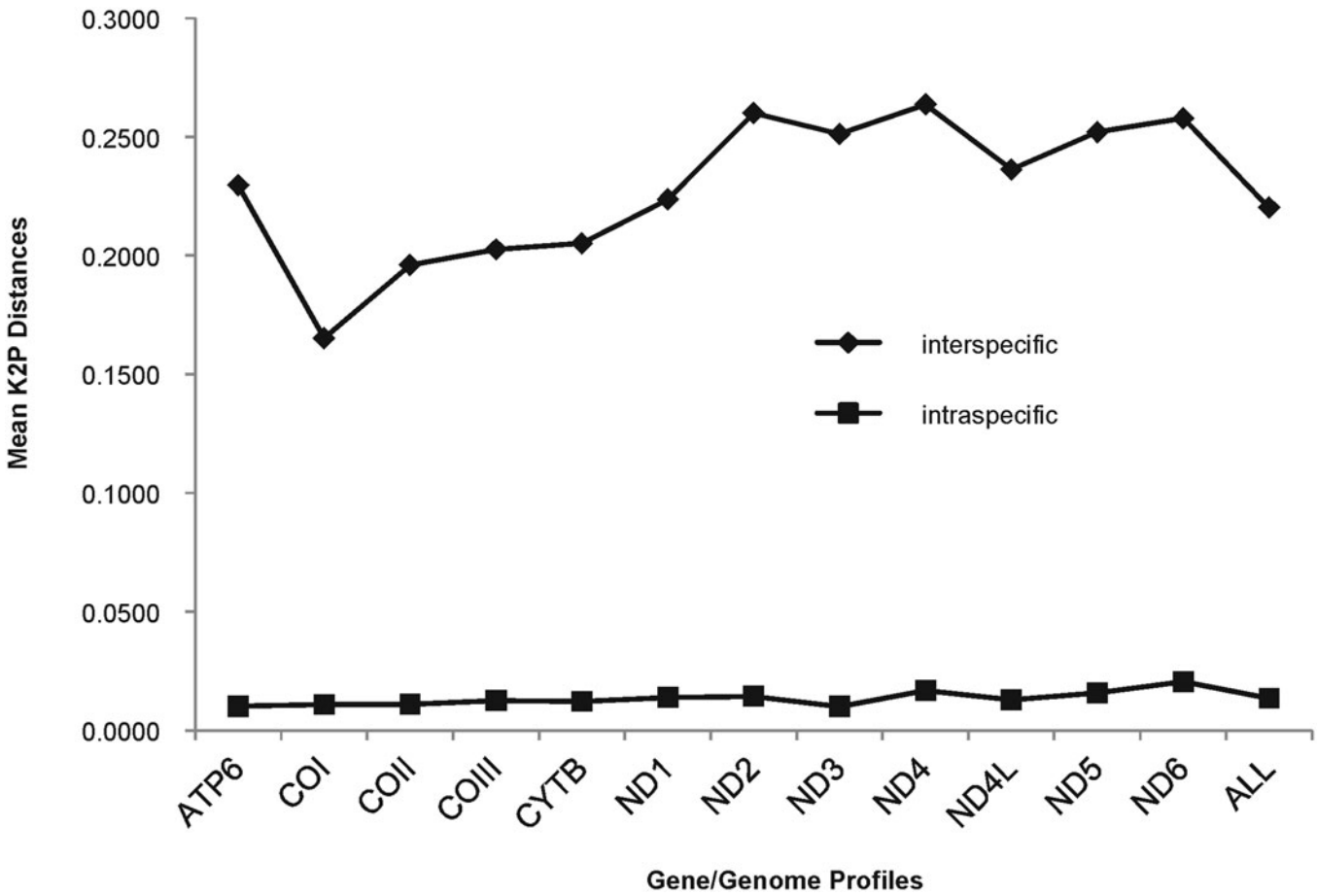

Figure 1. Mean intra- and interspecific distances from the 13 gene/genome dataset.

Table 1. Details of the 13 profiles in this study.

		Species						
Profile	Length (bp)	No.	Monophyly-based Barcode Sp. No.	Recovery rate (%)	Distance-based Barcode Sp. No.	Recovery rate (%)	Character-based Barcode Sp. No.	Recovery rate (%)
ATP6	633	24	22	91.67	18	75.00	22	91.67
COI	1530	24	22	91.67	18	75.00	22	91.67
COII	653	24	22	91.67	18	75.00	22	91.67
COIII	765	24	22	91.67	18	75.00	22	91.67
CYTB	1121	24	22	91.67	18	75.00	22	91.67
ND1	915	24	22	91.67	18	75.00	22	91.67
ND2	815	24	22	91.67	18	75.00	22	91.67
ND3	334	24	23	95.38	18	75.00	24	100
ND4	1262	24	23	95.38	20	83.33	24	100
ND4L	184	24	18	75.00	18	75.00	22	91.67
ND5	1547	24	23	95.38	20	83.33	24	100
ND6	424	24	22	91.67	17	70.83	22	91.67
All genes	10,183	24	23	95.38	18	75.00	24	100

''Species'' denotes the number of instances having at least two sequences representing each species.

less than 3% in all the 13 datasets. There was no significant difference in interspecific distance among the 13 datasets or between any pairwise comparisons ($p > 0.05$). The identification successes through the 12 genes and combination of the 12 genes were moderate (70.83–83.33%, Table 1).

Neighbour-joining clusters

More than 90% of the species which contained more than one individual formed monophyletic clusters in the NJ trees for 12 datasets (except for the ND4L gene), allowing their unambiguous identification. *Mytilus galloprovincialis* could not be recovered as a barcode species by any dataset, and *Mytilus edulis* could be recovered by the ND3, ND4, ND5 and genome profile. The NJ tree derived from ND4L gene recovered the least species, in which the taxa *Mytilus galloprovincialis* and *Mytilus edulis*, *S. oualaniensis*, *Crassostrea gigas*, *C. angulata* and *C. sikamea* were non-monophyletic.

Character-based DNA barcodes

Of 140 species, 136 species revealed unique base compositions, character-based DNA barcodes, with at least three CAs for each species in the 13 datasets. Two pairs of closely-related species (*Mytilus galloprovincialis*/*M. edulis* and *Meretrix. meretrix*/*M. petechialis*) had problems in some genes. No or less than three diagnostic characters were found in the four species in ATP6, COIII and ND4L gene regions. In the datasets of ND3, ND4, ND5 and genome profile, all the 140 species were clearly distinguished by diagnostic characters. The other six genes could not identify *Mytilus galloprovincialis* and *Mytilus edulis*. As expected, the genome profile showed the most CAs for each species among the 13 datasets.

Discussion

Utility of 12 mitochondrial protein-coding genes for species delimitation

Findings from the molluscan DNA barcoding analyses in this study demonstrated that all the 12 mt PCGs had similar performance and none was significantly better than others. This means that besides the universal COI gene region, other 11 mt PCGs can be alternative DNA barcodes for mollusks. Similar findings were also obtained in eutherians (Luo et al., 2011).

Although the distance-based method is most frequently used in DNA barcoding studies up to now (Taylor & Harris, 2012), the use of the distance-based method has been a major point of contention in the DNA barcoding. The use of genetic distances depends on the assumption that intraspecific distances are smaller than interspecific distances (Meyer & Paulay, 2005). However, an overlap between intra- and interspecific distances often occurred in analyses of COI barcode (Goldstein et al., 2000). In this study, overlaps between intra- and interspecific distances were not only detected in the COI gene, but also in the other 11 mt PCGs. The species which could not be discriminated were mostly closely related. Thus, the use of the distance-based method of mt gene barcodes for molluscan species identification should be done cautiously. A high level of intraspecific variation was observed in two controversial species (*O. hupensis* with 17 individuals and *S. oualaniensis* with three individuals) in all the 13 datasets, providing further evidence that there are two species or subspecies within these two species (Staaf et al., 2010; Wilke et al., 2000).

The NJ profile for identification depends on the coalescence of species but not an arbitrary level of divergence. High species recovery rates based on NJ trees were observed in 11 of 12 mt PCGs (except for ND4L). Some species (e.g. *C. gigas* and *C. angulata*) that failed recognition via the distance-based method could be recovered by NJ profiles. However, the recently divergent taxa are not reciprocally monophyletic because of lack of time for lineage sorting and the identification of the closely-related species *Mytilus galloprovincialis* and *Mytilus edulis* remained problematic via NJ profiles.

The character-based method of DNA barcoding is effective for species identification in this study. The recovery rates of the 12 mt PCGs were all larger than 91%. Especially, the closely-related species *Mytilus galloprovincialis* and *Mytilus edulis* that could not be clearly distinguished by distance and monophyly-based methods showed different diagnostic barcodes in the ND3, ND4, and ND5 genes. The effectiveness of the character-based barcoding method for discriminating closely-related species has also been reported (Bergmann et al., 2013; Zou et al., 2011).

From the above results, we can see that for some closely-related species, most mt PCGs failed in discrimination and barcode analysis should be done cautiously. The character-based method is more suitable for discrimination of closely-related species than distance and monophyly-based methods.

Efficacy of single mt genes and combination of mt genes for barcoding

Bergmann et al. (2013) found that the combination of the CO1 and ND1 fragment formed a better identifier than a single region

alone in the DNA barcoding analysis of odonate species. In this study, the results showed that there was no significant difference in efficacy for barcoding between the combination of 12 genes and the separate genes. The only difference was that the combination of 12 genes could identify more diagnostic characters than the separate genes. We also compared the combination of COI and ND1 genes with the single ND1 gene and no difference in the efficacy for barcoding was observed. Therefore, a single mt PCG can be representative of the efficacy of the whole mt genome. In another word, any one of the 12 PCGs can be potentially used as a molecular diagnostic for species identification in mollusks.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper. This study was supported by research grants from the National Natural Science Foundation of China (31201998, 41276138, 31372524).

References

Bergmann T, Rach J, Damm S, Desalle R, Schierwater B, Hadrys H. (2013). The potential of distance-based thresholds and character-based DNA barcoding for defining problematic taxonomic entities by CO1 and ND1. Mol Ecol Res 13:1069–81.

Bucklin A, Steinke D, Blanco-Bercial L. (2011). DNA barcoding of marine metazoa. Annu Rev Mar Sci 3:471–508.

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:e21326.

Collen BF, Böhm M, Kemp R, Baillie JEM. (2012). Spineless: Status and trends of the world's invertebrates. London: Zoological Society of London.

Davison A, Blackie RL, Scothern GP. (2009). DNA barcoding of stylommatophoran land snails: A test of existing sequences. Mol Ecol Res 9:1092–101.

Feng Y, Li Q, Kong L, Zheng X. (2011). COI-based DNA barcoding of Arcoida species (Bivalvia: Pteriomorphia) along the coast of China. Mol Ecol Res 11:435–41.

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 Biotech 3:294–9.

Goldstein PZ, DeSalle R, Amato G, Vogler AP. (2000). Conservation genetics at the species boundary. Conserv Biol 14:120–31.

Hebert PD, Cywinska A, Ball SL, deWaard JR. (2003). Biological identifications through DNA barcodes. Proc R Soc Lond B 270: 313–21.

Hebert PD, Stoeckle MY, Zemlak TS, Francis CM. (2004). Identification of birds through DNA barcodes. PLoS Biol 2:e312.

Hoareau TB, Boissin E. (2010). Design of phylum-specific hybrid primers for DNA barcoding: Addressing the need for efficient COI amplification in the Echinodermata. Mol Ecol Res 10: 960–7.

Kumar S, Nei M, Dudley J, Tamura K. (2008). MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinformatics 9:299–306.

Lohman DJ, Prawiradilaga DM, Meier R. (2009). Improved COI barcoding primers for Southeast Asian perching birds (Aves: Passeriformes). Mol Ecol Res 9:37–40.

Luo A, Zhang A, Ho SY, Xu W, Zhang Y, Shi W. (2011). Potential efficacy of mitochondrial genes for animal DNA barcoding: A case study using eutherian mammals. BMC Genomics 12:84.

Maddison D, Maddison W. (2000). MacClade 4: Analysis of phylogeny and character evolution. Massachusetts: Sinauer Associates.

Meier R, Shiyang K, Vaidya G, Ng PK. (2006). DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. Syst Biol 55:715–28.

Meyer CP, Paulay G. (2005). DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol 3:e422.

Mikkelsen NT, Schander C, Willassen E. (2007). Local scale DNA barcoding of bivalves (Mollusca): A case study. Zool Scr 36: 455–63.

Saccone C, De Giorgi C, Gissi C, Pesole G, Reyes A. (1999). Evolutionary genomics in Metazoa: The mitochondrial DNA as a model system. Gene 238:195–209.

Sarkar IN, Planet PJ, DeSalle R. (2008). CAOS software for use in character-based DNA barcoding. Mol Ecol Res 8:1256–9.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.

Staaf DJ, Ruiz-Cooley RI, Elliger C, Lebaric Z, Campos B, Markaida U. (2010). Ommastrephid squids *Sthenoteuthis oualaniensis* and *Dosidicus gigas* in the eastern Pacific show convergent biogeographic breaks but contrasting population structures. Mar Ecol Prog Ser 418: 165–78.

Taylor HR, Harris WE. (2012). An emergent science on the brink of irrelevance: A review of the past 8 years of DNA barcoding. Mol Ecol Res 12:377–88.

Wilke T, Davis GM, Xiao-Nung Z, Xiao PZ, Yi Z, Spolsky CM. (2000). Oncomelania hupensis (Gastropoda: Rissooidea) in eastern China: Molecular phylogeny, population structure, and ecology. Acta Tropica 77:215–27.

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. (2012). Multigene barcoding and phylogeny of geographically widespread muricids (Gastropoda: Neogastropoda) along the coast of china. Mar Biotech 14:21–34.