

Monophyly, Distance and Character-Based Multigene Barcoding Reveal Extraordinary Cryptic Diversity in *Nassarius*: A Complex and Dangerous Community

Shanmei Zou¹, Qi Li^{2*}, Lingfeng Kong³

1 Key Laboratory of Mariculture Ministry of Education, Ocean University of China, Qingdao, China, **2** Key Laboratory of Mariculture Ministry of Education, Ocean University of China, Qingdao, China, **3** Key Laboratory of Mariculture Ministry of Education, Ocean University of China, Qingdao, China

Abstract

Background: Correct identification and cryptic biodiversity revelation for marine organisms are pressing since the marine life is important in maintaining the balance of ecological system and is facing the problem of biodiversity crisis or food safety. DNA barcoding has been proved successful to provide resolution beyond the boundaries of morphological information. *Nassarius*, the common mudsnail, plays an important role in marine environment and has problem in food safety, but the classification of it is quite confused because of the complex morphological diversity.

Methodology/Principal Findings: Here we report a comprehensive barcoding analysis of 22 *Nassarius* species. We integrated the mitochondrial and nuclear sequences and the morphological characters to determine 13 *Nassarius* species studied and reveal four cryptic species and one pair synonyms. Distance, monophyly, and character-based barcoding methods were employed.

Conclusions/Significance: Such successful identification and unexpected cryptic discovery is significant for *Nassarius* in food safety and species conversation and remind us to pay more attention to the hidden cryptic biodiversity ignored in marine life. Distance, monophyly, and character-based barcoding methods are all very helpful in identification but the character-based method shows some advantages.

Citation: Zou S, Li Q, Kong L (2012) Monophyly, Distance and Character-Based Multigene Barcoding Reveal Extraordinary Cryptic Diversity in *Nassarius*: A Complex and Dangerous Community. PLoS ONE 7(10): e47276. doi:10.1371/journal.pone.0047276

Editor: Zhanjiang Liu, Auburn University, United States of America

Received: July 3, 2012; **Accepted:** September 10, 2012; **Published:** October 11, 2012

Copyright: © 2012 Zou et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by research grants from 973 Program (2010CB126406) and National Marine Public Welfare Research Program (201205023). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: qjli66@ouc.edu.cn

Introduction

It is pressing to catalogue the earth's species since the world is facing a global biodiversity crisis [1,2]. The rapid loss of marine biodiversity has prompted efforts to catalogue the biodiversity, such as the Census of Marine Life (www.coml.org). Large numbers of marine organisms are important in maintaining the balance of ecological system, and many of them are consumed as seafood. Thus, correct species identification and revelation of cryptic species diversity for marine life is important to nature conversation, food safety and better understanding the patterns of ecosystem functioning. Nevertheless, due to the declining number of taxonomists [3], the insufficient funding for taxonomy and the confused morphological diversity, it is hard for traditional taxonomy to undertake the huge taxonomic task for marine organisms.

While the traditional taxonomy has been declining, DNA-based techniques, such as DNA barcoding [4], often provide resolution beyond the boundaries of morphological information [5]. DNA barcoding, which involves taxon identification using standardized DNA regions, has recently received much attention [6,7,8,9]. It is an aid to the discrimination and identification of species and can recover new or cryptic species [10,11]. Until now, DNA barcoding

has been successfully applied to many animals (e.g. [12,13,14,15,16,17]). Two broad methods of DNA barcoding (distance and monophyly-based methods) have been originally used. Distance-based method is based on the "barcoding gap", the degree of DNA sequence variation within and between species. Monophyly-based method requires the recovery of species as discrete clades (monophyly) on a phylogenetic tree [6]. Nevertheless, some issues complicate the use of both methods [18,19,20,21,22,23]. A recently applied new technique, the character-based DNA barcode approach, characterizes species through a unique combination of diagnostic characters [18,23,24,25] and has been proved useful for species identification and discovery of cryptic species [15,23,24].

Nassariidae is a large gastropod group, comprising about 300 extant and almost 600 extinct nassariid species that are organized into 12 genera and 31 subgenera [26]. Three nassariid subfamilies are commonly recognized [26,27]: the Dorsaninae, the Cylleninae and the cosmopolitan Nassariinae. *Nassarius*, the common mudsnail, is a species-rich genus of Nassariinae and is distributed throughout worldwide oceans [28]. The nassariids of *Nassarius* are usually less than 50 mm in adult shell height [29]. Ecologically, most nassariids of *Nassarius* are thought to be facultative scavengers

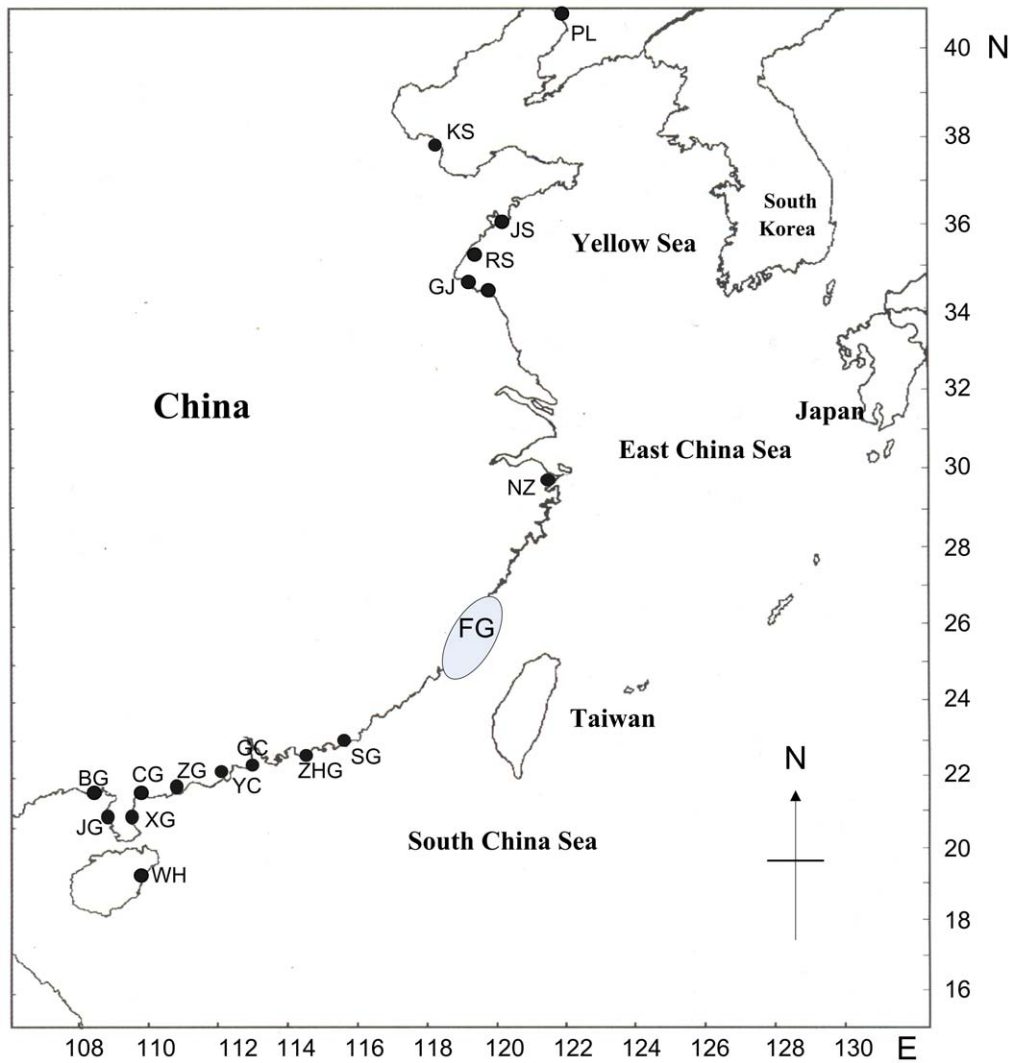


Figure 1. Sampling sites in this study. The letter codes correspond to geographic locations listed in table S1.
doi:10.1371/journal.pone.0047276.g001

inhabiting inter- to subtidal shallow marine environments [26]. As scavengers, nassariids of *Nassarius* are important in maintaining the balance of ecological system, especially for the balance of benthic community. They are also useful in the biomonitoring of Tributyltin (TBT) pollution in marine environment. Due to the high specificity and sensitivity to TBT, imposex phenomenon is found in some *Nassarius* species. In fact, imposex is considered the best biological indicator of TBT pollution in marine waters [30]. More importantly, food safety problem exists in *Nassarius*. Most species of *Nassarius* are consumed as food in China where they are widely distributed. Nevertheless, maybe due to the food nassariids of *Nassarius* get from marine waters, different toxins are concentrated in *Nassarius* sp's body. Recent studies find that the toxicity of *Nassarius* is relative to species [31,32]. For example, *N. hepaticus* are toxic gastropods, *N. festiva* are non-toxic gastropods, while the toxicity of *N. succinctus* probably change with the season [32].

Despite the importance of maintaining the balance of ecological system, the usefulness of monitoring TBT pollution and the danger of eating, the taxonomy of *Nassarius* species is still confusing. Discrimination of the *Nassarius* species is mainly based on the shell morphology, especially the sculpture [33]. However, due to the

intraspecific shell variation affected by biotic and abiotic factors [34,35] and the various shell forms in different species, the identification of *Nassarius* species is often difficult. Environment adaptive intraspecific morphological variation can lead to ambiguous identification of closely related species [36], and interspecific uniformity may also present difficulties in species identification [37]. Thus, it is arbitrary to identify *Nassarius* species only using morphological characters and there is probably some misidentification and a significant amount of cryptic diversity within *Nassarius*. Some *Nassarius* species that are considered as single may be erroneously classified under one species name. Unfortunately, until now there are few large-scale reliable genetic studies to identify *Nassarius* species and estimate the level of cryptic diversity within *Nassarius*. Li *et al.* [38] employed mitochondrial sequences to study the identification and phylogeny of *Nassarius*. Nevertheless, due to the very limited samples, the status of *Nassarius* species is still unclear.

In this study we reported a barcoding analysis of 22 *Nassarius* species. Many of the species have diverse morphological characters and are easily confused. Two mitochondrial genes COI and 16S rRNA and one nuclear gene ITS-1 were employed. Distance, monophyly, and character-based barcoding methods were con-

ducted. We integrated the molecular and morphological data: (1) to identify the species and reveal the cryptic diversity within *Nassarius* (2) to test the performance of DNA barcoding and three barcode methods for morphologically complex species.

Materials and Methods

Ethics Statement

No specific permits were required for the described field studies. The field studies did not involve endangered or protected species. No specific permissions were required for the locations. The locations are not privately-owned or protected in any way.

Sample Collections

A total of 220 samples representing 22 *Nassarius* species were used in this study (Table S1). Thereinto, 208 specimens were collected across the whole China coast from 2005 to 2011 (Figure 1). One or more specimens were chosen from each locality in order to include as many morphologically distinguishable individuals per site as possible. Specimens were collected and stored in 90–100% ethanol.

DNA Extraction, PCR Amplification and Sequencing

DNA was extracted from small pieces of foot tissue by the CTAB method as modified by Winnepenninckx *et al.* [39]. PCR reactions were carried out in a total volume of 50 μ L, using 1.5 mM MgCl₂, 0.2 mM of each dNTPs, 1 μ M of both forward and reverse PCR primers, 10 \times buffer and 2.5 U *Taq* DNA polymerase. Thermal cyclings were performed with an initial denaturation for 3 min at 95°C, 45 s at primer-specific annealing temperatures, and 1 min at 72°C, followed by 35 cycles of 30 s at 95°C, 45 s at primer-specific annealing temperatures, 1 min at 72°C, with a final extension of 10 min at 72°C. PCR and sequencing primers for COI, 16S rRNA and ITS-1 genes were listed in Table 1. The PCR products were confirmed by 1.5% agarose gel electrophoresis and stained with ethidium bromide. The fragment of interest was purified using EZ Spin Column PCR Product Purification Kit, Sangon. Purified products were sequenced in both directions using the BigDye Terminator Cycle Sequencing Kit (ver. 3.1, Applied Biosystems) and an AB PRISM 3730 (Applied Biosystems) automatic sequencer.

Distance and Phylogenetic Analysis

Forward and reverse sequences of each gene were edited, assembled and merged into consensus sequences using the software program Sequencher 4.5 (Genecodes Corporation, Ann Arbor, MI). Sequences were aligned using the program, *ffnisi*, which is implemented in MAFFT 6.717 [40]. Alignment of COI nucleotide sequences was unproblematic since indels were absent. For 16S rDNA and ITS-1 sequences, areas of uncertain alignment were omitted by the software Gblocks 0.91b [41], with minimum number of sequences for a conserved position set to 50% of the total, minimum number of sequences for a flanking position set to 90% of the total, maximum number of contiguous non-conserved positions set to 3, minimum length of a block set to 5, and half gap positions allowed.

For distance analyses, pairwise sequence divergences were calculated using a Kimura 2-parameter (K2P) distance model in MEGA 4.0 [42]. Phylogenetic analysis of COI, 16S rDNA and ITS-1 sequences were carried out using neighbour joining (NJ) and Bayesian methods. The species *Fusinus longicaudus*, *Euphica scripta*, *Mitrella burchardi* and *Pseudamycla formosa* were selected as the outgroups. NJ analyses were conducted using K2P distance model as recommended by Hebert *et al.* [4] in MEGA 4.0 [42]. Bayesian analyses were carried out using the Monte Carlo Markov Chainmethod (MCMC) implemented on MrBayes v.3.1.2 [43]. Nucleotide substitution models for Bayesian analyses were selected separately for each gene using the Akaike Information Criterion (AIC) as implemented in the jModeltest v.0.1.1 [44]. The most appropriate models for Bayesian analyses were HKY+I+G for COI, HKY+I+G for 16S and GTR+G for ITS-1. Four chains were run twice in parallel for 10⁷ generations, and trees were sampled every 100 generations. Stationarity was considered to be reached when the average standard deviation of split frequencies shown in MrBayes were less than 0.01 [43]. Chain convergence was further verified by ensuring potential scale reduction factors neared 1 and using Tracer v.1.5 to confirm sufficiently large ESS values. Burn-ins were determined by visually inspecting the $-\ln L$ trace plot in Tracer.

Character-Based Barcode Analysis

The characteristic attribute organization system (CAOS) [45,46] was used for the character-based identification method. The CAOS algorithm identifies character-based diagnostics, here termed “characteristic attributes” (CAs), for every clade at

Table 1. Sequences of the primers used in the PCRs.

Name	Sequence 5'-3'	Annealing Temperature (°C)	Source
COI			
LCO1490 (F)	GGTCAACAAATCATAAAGATATTGG	45–50	[56]
HCO2198 (R)	TTAACTTCAGGGTGACCAAAAATCA	45–50	[56]
16S			
16Sar	CGCCTGTTTATCAAAAACAT	51	[57]
16Sbr	CCGGTCTGAACCTCAGATCACGT	51	[57]
16SarM	GCGGTACTCTGACCGTGCAA	48–50	[16]
16SbrM	TCACGTAGAATTTAATGGTCG	48–50	[16]
ITS-1			
ITS-1 (F)	TAACAAGGTTCCGTAGGTGAA	52	[58]
ITS-1 (R)	GCTGCGTTCTTCATCGATGC	52	[59]

doi:10.1371/journal.pone.0047276.t001

branching node within a guide tree that is first produced from a given dataset. The system comprises two programs: P-Gnome and P-Elf [45]. The program Macclade [47] was used to produce the nexus files for P-Gnome in accordance with the CAOS manual. The most variable sites that distinguish all the taxa were chosen and the character states at these nucleotide positions were listed.

Results

In total, we analyzed 187 COI (652 bp), 171 16S rDNA (440–530 bp) and 82 ITS-1 (470–560 bp) sequences from 220 *Nassarius* individuals. Sequences from this study were submitted to the GenBank Barcode database with accession numbers JQ975421–JQ975808 listed in Table S1. 40 COI sequences and 12 16S rDNA sequences were obtained from previous studies.

Phylogenetic, Distance and Character Assignments in COI Barcoding

The NJ and Bayesian trees of COI locus supported the monophyly of *Nassarius* (Figure 2). For the 20 *Nassarius* species analyzed, the species *N. hepaticus*, *N. acuminatus*, *N. algidus*, *N. conoidalis*, *N. succinctus*, *N. pullus*, *N. siquijorensis* and *N. semiplicata* formed distinct barcode clusters allowing their unambiguous identification. Two separate clades within *N. festiva* and *N. livescens* were clearly recovered respectively (Figure 2). *N. sp1* fell within the *N. variciferus* clade. *N. sp* and one individual of *N. hepaticus* (FJ660644) and *N. sufflatus* and *N. dorsatus* were lumped into one lineage respectively.

The COI pairwise genetic divergences among conspecific individuals ranged from 0% to 16.2% with a mean of 1.19%. Between specimens of different species, the variation was from 0% to 24.80%. The mean interspecific distance was from 0.30% to 22.90% (Table S2). No “distance-gap” was found between intraspecific and interspecific divergences of COI sequences within *Nassarius* (Figure 3). The mean distances between two clades within *N. festiva* and *N. livescens* were 6.00% and 4.5% respectively. The mean genetic divergences between *N. sp1* and *N. variciferus* and *N. sufflatus* and *N. dorsatus* were only 0.60% and 0.30% respectively.

The COI NJ tree was the guide tree for COI CAOS analysis. 22 defined clades in Figure 2 were analyzed: *N. festiva I*, *N. festiva II*, *N. hepaticus*, *N. succinctus*, *N. siquijorensis*, *N. dorsatus*, *N. pullus*, *N. semiplicata*, *N. conoidalis*, *N. livescens I*, *N. livescens II*, *N. sp* (including FJ660644), *N. variciferus*, *N. acuminatus*, *N. sp1*, *N. algidus*, *N. sufflatus*, *N. burchardi*, *N. kraussianus*, *N. nitidus*, *N. reticulatus* and *N. sp2*. In the COI gene region of 22 clades character states at 41 nucleotide positions were detected (Table S3). The particular nucleotide positions were chosen due to the high number of CAs at the important nodes or because of the presence of CAs for groups with highly similar sequences. All the clades except *N. variciferus*, *N. sp1*, *N. dorsatus* and *N. sufflatus* revealed a unique combination of character states at 41 nucleotide positions with at least 3 CAs for each. *N. festiva I* and *N. festiva II* and *N. livescens I* and *N. livescens II* were clearly separated respectively with more than 8 CAs. Two separate clades within *N. acuminatus* and *N. algidus* in COI phylogenetic tree (see Figure 2) were also detected with 3 and 5 CAs respectively.

Phylogenetic and Character Assignments in 16S rDNA Barcoding

Generally, the 16S rDNA NJ and Bayesian trees revealed same resolution to COI trees (Figure 4). For the 16 *Nassarius* species analyzed, the monophyly of *N. hepaticus*, *N. succinctus*, *N. conoidalis*, *N. pullus* were strongly supported. Although with weak support, *N.*

festiva was separated into two clades (Figure 4). *N. livescens* was also clearly separated into two clusters. *N. variciferus* and *N. sp1* and *N. sufflatus* and *N. dorsatus* grouped together with each other respectively. Unexpectedly, one individual of *N. semiplicata* (EU076706) failed to group together with other individuals.

The 16S rDNA NJ tree was the guide tree for 16S rDNA CAOS analysis. 19 defined clades in Figure 4 were analyzed: *N. festiva I*, *N. festiva II*, *N. hepaticus*, *N. succinctus*, *N. siquijorensis*, *N. dorsatus*, *N. pullus*, *N. semiplicata I*, *N. semiplicata II*, *N. conoidalis*, *N. livescens I*, *N. livescens II*, *N. variciferus*, *N. sp1*, *N. sufflatus*, *N. semiplicatoides*, *N. acuminatus*, *N. algidus* and *N. castus*. In the 16S rDNA gene region of 19 clades character states at 30 nucleotide positions were detected (Table S4). All the clades except *N. variciferus*, *N. sp1*, *N. dorsatus* and *N. sufflatus* revealed a unique combination of character states at 30 nucleotide positions with at least 3 CAs for each. *N. festiva I* and *N. festiva II* and *N. livescens I* and *N. livescens II* were clearly separated respectively with more than 5 CAs. For *N. semiplicata I* and *N. semiplicata II*, 20 CAs were found.

Phylogenetic and Character Assignments in ITS-1 Barcoding

For the 10 *Nassarius* species analyzed, 10 distinct *Nassarius* lineages can be identified in ITS-1 NJ and Bayesian trees (Figure 5). The species *N. hepaticus*, *N. siquijorensis*, *N. succinctus*, *N. conoidalis*, *N. pullus*, *N. livescens* and *N. variciferus* were recovered as monophyletic. However, the ITS-1 region failed to separate *N. festiva I* and *N. festiva II* recovered in COI and 16S rDNA trees.

The ITS-1 NJ tree was the guide tree for ITS-1 CAOS analysis. 11 clades in Figure 5 were analyzed: *N. festiva I*, *N. festiva II*, *N. hepaticus*, *N. succinctus*, *N. siquijorensis*, *N. pullus*, *N. sp*, *N. semiplicata*, *N. conoidalis*, *N. livescens* and *N. variciferus*. In the ITS-1 gene region of 11 clades character states at 26 nucleotide positions were found (Table S5). All the clades except *N. festiva I* and *N. festiva II* revealed a unique combination of character states at 26 nucleotide positions with at least 3 CAs for each.

Discussion

Species Delimitation and Cryptic Diversity

DNA sequence data now offers an effective tool for taxonomic studies by greatly expanding the number of characters that can be used to distinguish species. The inclusion of such data, along with the traditional morphological variables, promises to rectify the problem of subjectivity in current species descriptions [48,49,50]. Our analyses of comprehensive samples of *Nassarius* species, combining genetic data with morphological characters (discussed below), led to the successful identification of 12 *Nassarius* species and the discovery of four cryptic species, one pair synonyms and one intraspecific morphologically diverse species.

First, the genetic data and the morphological characters provided the most obvious evidence for the existence of one cryptic species in *N. festiva*. In this study, all the individuals of *N. festiva* were separated into two different lineages (*N. festiva I* and *N. festiva II*) in COI and 16S rDNA phylogenetic trees. The two lineages were also clearly recovered in COI and 16S rDNA character assignments with many CAs. Moreover, the COI divergence between *N. festiva I* and *N. festiva II* was larger than the mean intraspecific divergence. However, the ITS-1 region failed to separate the two clades. The reason may be that ITS-1 gene does not have sufficient variation to distinguish the two recent diverged lineages since it evolve much more slowly than COI and 16S rDNA genes [51,52]. Slightly ambiguous, but still significant different morphological trait between *N. festiva I* and *N. festiva II* is that the verrucous protuberances on the shell of *N. festiva*

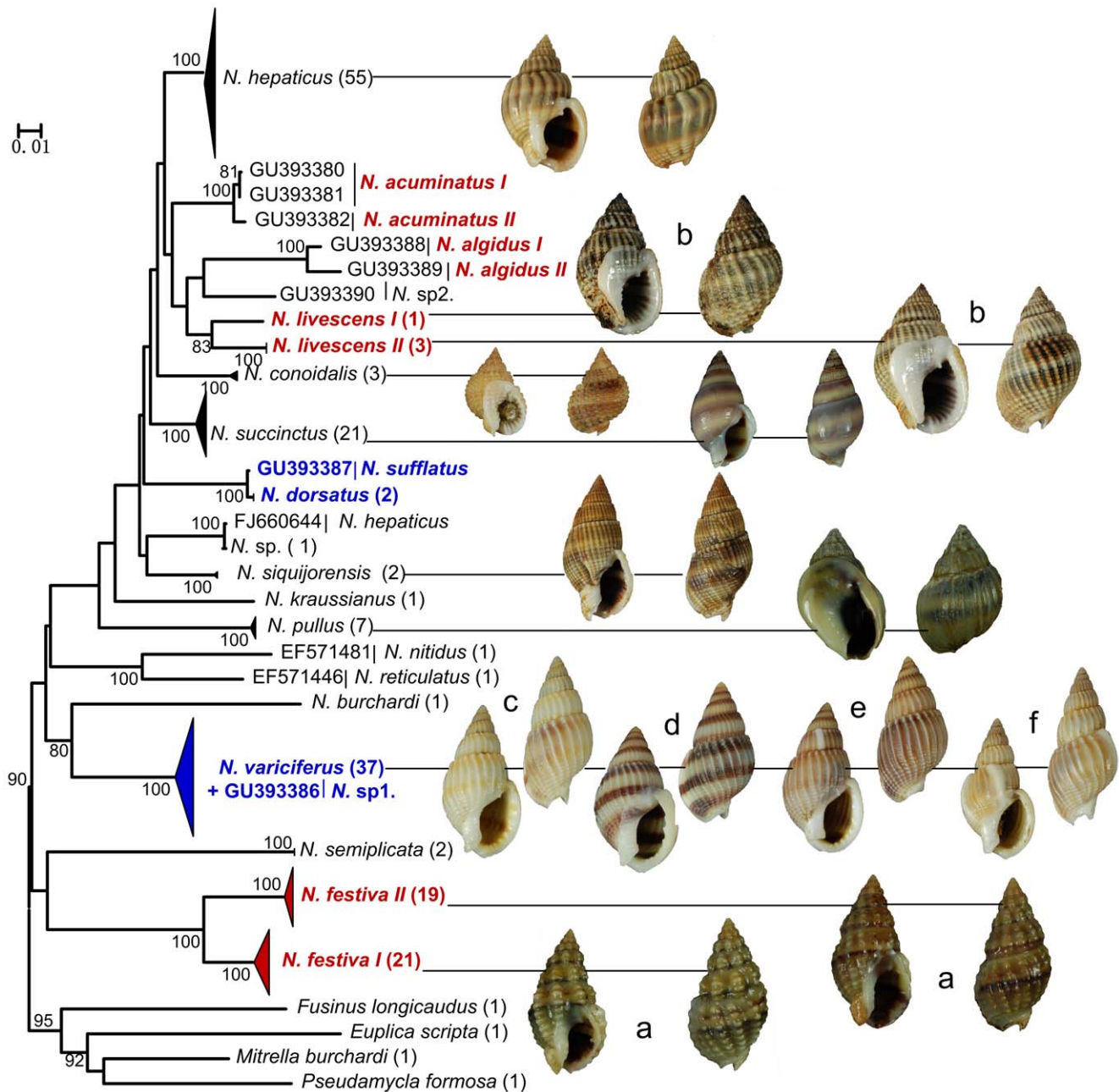


Figure 2. Bayesian tree of the COI locus. Posterior probabilities and bootstrap values were included. The number of individuals included in each species was shown in brackets by the species name. Species showing cryptic diversity were marked in red. Species that could be identified as synonyms were marked in blue. Representative shells of species available were illustrated. doi:10.1371/journal.pone.0047276.g002

I are bigger than that on the shell of *N. festiva II* (Figure 2 (a)). Geographically, the two clades are both represented in the same localities in this study. Wang *et al.* [53] suggested that *N. dealbatus* be the synonym of *N. festiva*. Here our study suggests that *N. festiva* be regarded as two separate species.

Second, both phylogenetic trees and character assignments of COI and 16S rDNA genes separated one individual of *N. livescens* (*N. livescens I*) from other individuals (*N. livescens II*). The mean COI distance between *N. livescens I* and *N. livescens II* was also larger than the mean intraspecific divergence. All the individuals of the two clades were collected from the same localities in this study. Specimens of the two clades are almost identical morphologically

but a putative difference may be that there are more axial ribs on the shell of *N. livescens II* than that on the shell of *N. livescens I* (Figure 2 (b)). Thus, a putative cryptic species within *N. livescens* is found and more individuals from more localities are needed to find more morphological and genetic differences between the two clades.

Third, the species *N. acuminatus* and *N. algidus* also showed cryptic genetic diversity. Although all the individuals of *N. acuminatus* and *N. algidus* fell into one cluster respectively in COI phylogenetic trees, two clades within each species in COI trees were clearly separated in COI character assignments. The cryptic diversity within the two species needs to be recognized.

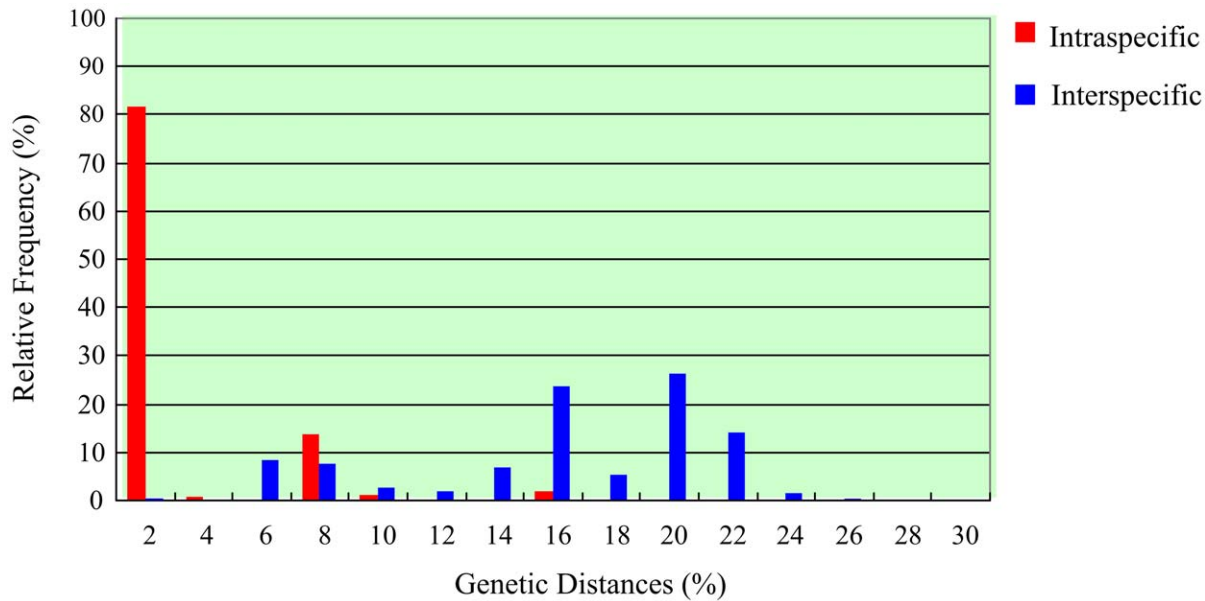


Figure 3. Frequency distribution of COI intraspecific and interspecific (congeneric) K2P distances in *Nassarius*.
doi:10.1371/journal.pone.0047276.g003

Fourth, although all analysis of COI, 16S rDNA and ITS-1 sequences supported the monophyly of *N. variciferus*, *N. variciferus* showed high intraspecific morphological diversity. First of all, it should be noted that some individuals of *N. variciferus* have no varices on the shell (Figure 2 (c) and (d)). Some individuals just have varices on body whorl or one spiral whorl (Figure 2 (e) and (f)). Thus, it is wrong that all individuals of *N. variciferus* have distinct varices on body whorl and all spiral whorls. In addition, the color of spiral bands of some individuals is much darker than that of other individuals (Figure 2 (d)) and the suture of some individuals is a little deeper than that of others (Figure 2 (f)). Therefore, we must be cautious to identify the specimens of *N. variciferus* since there is high morphological diversity within it. The unknown species *N. sp1* (GU393386 in Li *et al.* [38]) also fell into the *N. variciferus* clade in both COI and 16S rDNA phylogenetic trees, and same CAs were detected for them in both COI and 16S rDNA character assignments. Therefore, *N. sp1* can be identified as *N. variciferus*.

Finally, *N. sufflatus* grouped together with *N. dorsatus* with 99% or 100% support in COI and 16S rDNA phylogenetic trees. Same CAs were also detected for *N. sufflatus* and *N. dorsatus* in both COI and 16S rDNA character assignments, and the mean COI distance between them was only 0.3%. Thus, *N. sufflatus* and *N. dorsatus* could be regarded as synonyms. Unexpectedly, one individual of *N. hepaticus* (FJ6606441 in Wang *et al.* [53]) was separated from all others and grouped together with *N. sp* in both COI phylogenetic and character analysis. The individual may be misidentified by Wang *et al.* [53]. Another unknown species *N. sp2* (GU393390 in Li *et al.* [38]) was not identified. It also failed to be identified to species level in the Barcode of Life Data Systems (BOLD). Thus, more sequences need to be produced in BOLD for the identification of unknown species. The other nominal species, e.g. *N. hepaticus*, *N. succinctus*, *N. pullus*, *N. semiplicata*, *N. conoidalis* and *N. siquijorensis* examined in this study, were successfully identified in phylogenetic trees and character assignments of COI, 16S rDNA and ITS-1 sequences. In these species clades, no geographical clusters could be detected.

DNA Barcoding and Three Methods

This study has shown that DNA barcoding is effective in identifying *Nassarius* species. It can reveal cryptic species that morphological characters can not distinguish alone due to the intraspecific variation and various intraspecific forms. The correct identification and revelation of cryptic diversity is important for *Nassarius* in species conversation, food safety and better understanding the patterns of ecosystem functioning. Actually, like *Nassarius*, the external morphology of most marine species is easily affected by the environmental factors, at least for the mollusk, which makes morphological characters sometimes unreliable to identify. Thus, DNA barcoding will be a powerful tool for revealing the marine biodiversity. At the fewest, DNA barcoding can flag species and reduce the candidate new species, after which the traditional characters can complement the identification. Whatever, the integration of distinct DNA characters and traditional information such as morphology and geography in a comprehensive character-based barcode database is needed for fast species identification and discovery.

In this case study, the traditional barcoding methods, monophyly and distance-based methods, were very helpful in revealing the diversity of *Nassarius* species. For example, all phylogenetic trees of COI, 16S rDNA and ITS-1 genes could recover most species (including the cryptic species) as monophyletic and the COI interspecific divergences were generally higher than the intraspecific divergences. Even so, compared with the character-based DNA barcoding, some limits of the traditional barcoding methods still appeared. First, identification does not hinge on monophyly and the use of reciprocal monophyly as a criterion for species recognition is arbitrary [54,55]. In this study, although some species were recovered as monophyletic in the phylogenetic trees, the cryptic diversity within the species could not be completely shown in the trees. For example, within the monophyletic species *N. festiva* in 16S rDNA trees, the clades *N. festiva I* and *N. festiva II* were weakly supported (see Figure 4), but they were clearly separated in 16S rDNA character assignments (Table S4). In addition, two closely related clades within *N. acuminatus* and *N. algidus* in COI trees (see Figure 2) were also detected respectively in COI character assignments (Table S3). Moreover, if one species is

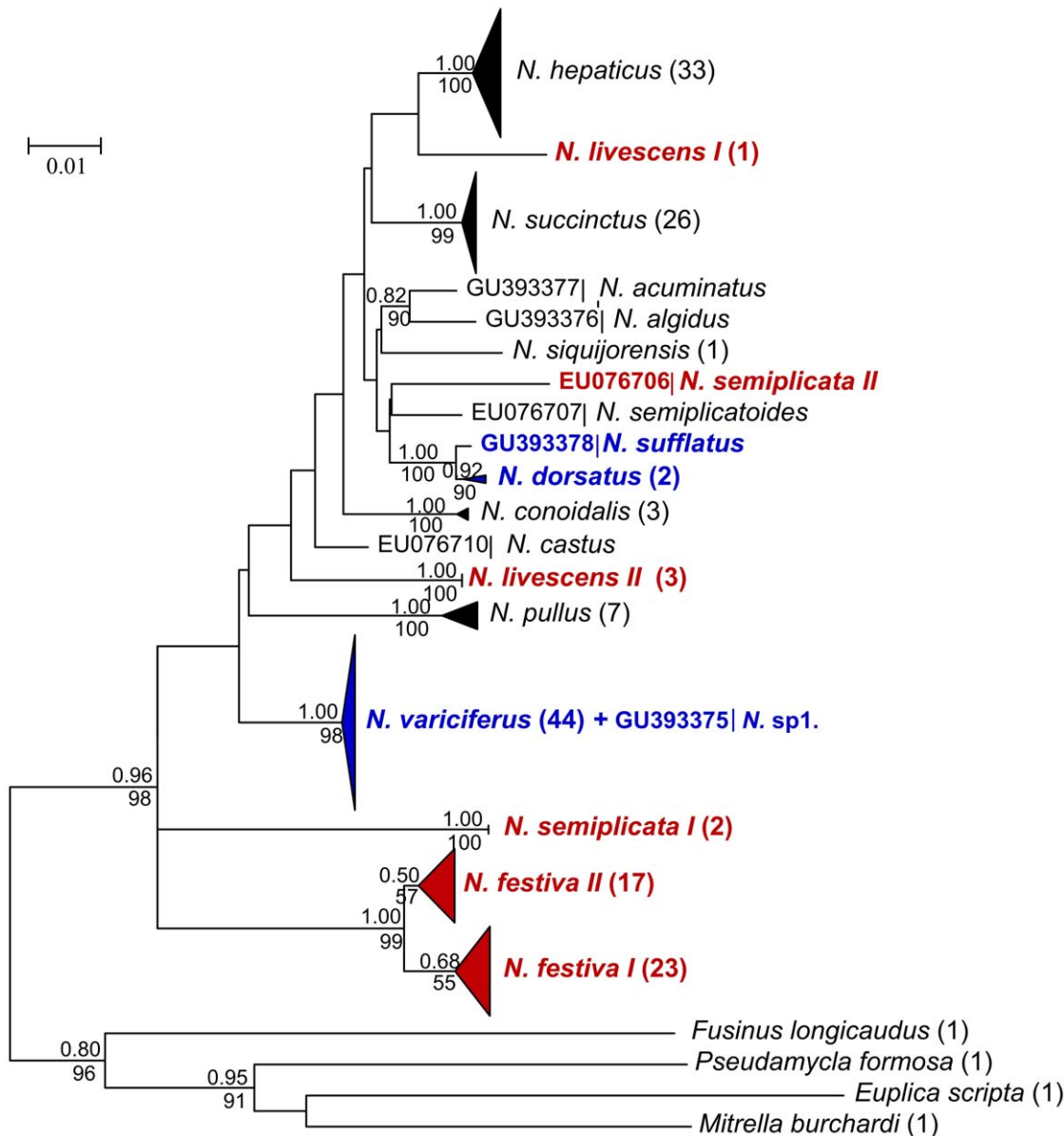


Figure 4. Bayesian tree of the 16S rDNA locus. Posterior probabilities and bootstrap values were included. The number of individuals included in each species was shown in brackets by the species name. Species showing cryptic diversity were marked in red. Species that could be identified as synonyms were marked in blue.

doi:10.1371/journal.pone.0047276.g004

represented with a single individual in phylogenetic profile, it is not determinative the species is monophyletic or polyphyletic (e.g. *N. sufflatus* and *N. dorsatus* and *N. nitidus* and *N. reticulatus* in COI phylogenetic trees). Nevertheless, a character-based DNA barcode of a single individual is still useful and provides important information for this species within a group of interest. Second, the distance-based approach failed in some species in this study. No “barcoding gap” was found between COI intra- and interspecific variation. On the contrary, there was obvious overlap between them. In addition, since some cryptic species existed the “10× rule” threshold (11.9% in this study) proposed by Hebert *et al.* [52] was too liberal to recognize some distinct species. The character-based method of DNA barcoding, however, was effective for the identification of genetic entities. It could easily detect the cryptic species that could not be recovered with NJ profile and genetic distance and the species that were represented by a single

individual. Although there is no absolute certainty for a given CA to be fixed, the reliability of a barcode increases with each additional independent CA added [24]. Another advantage of character-based barcoding is the fact that it is compatible with classical approaches allowing the combination of classical morphological information.

Food Safety in *Nassarius*

Nassariids of *Nassarius* are popular with people in China since they are very delicious to eat. However, it is dangerous to consume them as food since different toxins are concentrated in *Nassarius* sp’s body. Food poisoning incidents caused by eating nassariids of *Nassarius* have been reported frequently in the last several years in China. Many people died of the poisoning incidents [31,32]. Thus, relevant departments of China have forbidden selling nassariids with toxins. The origin of the toxicity in *Nassarius* sp’s body is still unclear. It is inferred that the

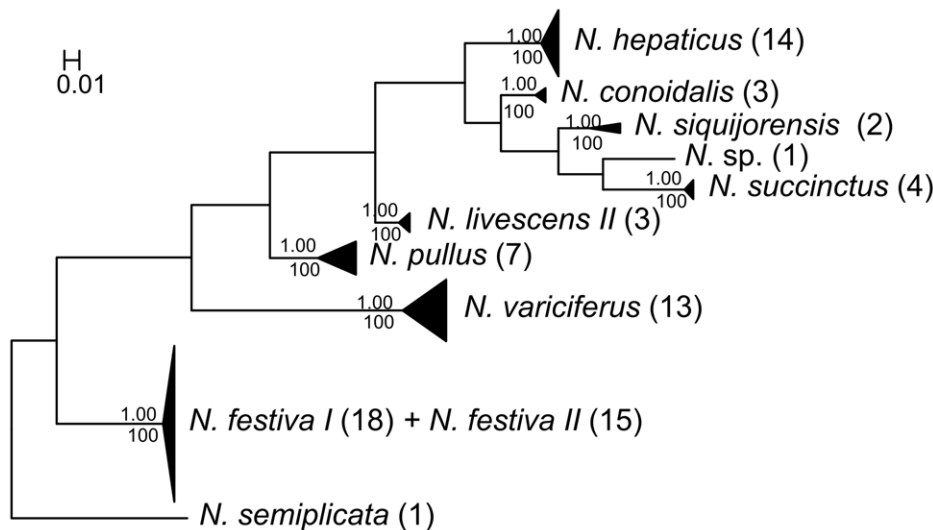


Figure 5. Bayesian tree of the ITS-1 locus. Posterior probabilities and bootstrap values were included. The number of individuals included in each species was shown in brackets by the species name. doi:10.1371/journal.pone.0047276.g005

toxicity probably originates from the food chains, actinomycetes in *Nassarius* sp's body or an enzyme produced by themselves. Some studies find that the toxicity of *Nassarius* is relative to species [31,32]. While some species are toxic and some species are non-toxic, the toxicity of some species changes with the season [31,32]. Therefore, correct species identification is the basis of studying *Nassarius* toxicity. However, the morphological confusion in *Nassarius* often results in error in virulence judgment. Here our comprehensive barcoding study for species delimitation and cryptic diversity revelation of *Nassarius* will greatly contribute to the virulence study of *Nassarius* since representatives of toxic, non-toxic and season-toxic species are all included in our study.

Supporting Information

Table S1 Sampling of *Nassarius* species and outgroups studied. (DOC)

References

- Novacek MJ, Cleland EE (2001) The current biodiversity extinction event: scenarios for mitigation and recovery. *Proc Natl Acad Sci USA* 98: 5466–5470.
- Bellwood DR, Hughes TP, Folke C, Nystrom M (2004) Confronting the coral reef crisis. *Nature* 429: 827–833.
- Hopkins GW, Freckleton RP (2002) Declines in the numbers of amateur and professional taxonomists: implications for conservation. *Anim Conserv* 5: 245–249.
- Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proc R Soc Lond B* 270: 313–321.
- Dincă V, Lukhtanov VA, Talavera G, Vila R (2011) unexpected layers of cryptic diversity in wood white *Leptidea* butterflies. *Nat. Commun*; doi:10.1038/ncomms1329.
- Hebert PDN, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond B* 270: S96–S99.
- Waugh J (2007) DNA barcoding in animal species: progress, potential and pitfalls. *Bioessays* 29: 188–197.
- Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol Ecol Notes* 7: 355–364.
- Bertolazzi P, Felici G, Weitschek E (2009) Learning to classify species with barcodes. *BMC Bioinformatics* 10: S7.
- Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, et al. (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *tnL* approach. *Mol Ecol Resour* 9: 51–60.
- Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Trends Ecol Evol* 24: 110–117.

Table S2 The mean interspecific divergences of COI sequences. (DOC)

Table S3 Character-based DNA barcodes for COI gene. (DOC)

Table S4 Character-based DNA barcodes for 16S rDNA gene. (DOC)

Table S5 Character-based DNA barcodes for ITS-1 gene. (DOC)

Author Contributions

Conceived and designed the experiments: SZ QL. Performed the experiments: SZ. Analyzed the data: SZ. Contributed reagents/materials/analysis tools: LK. Wrote the paper: SZ.

- Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia' fish species. *Phil Trans R Soc Lond B* 360: 1847–1857.
- Kerr KCR, Stoeckle MY, Dove C, Weigt LA, Francis CM, et al. (2007) Comprehensive DNA barcode coverage of North American birds. *Mol Ecol Notes* 7: 535–543.
- Wiemers M, Fiedler K (2007) Does the DNA barcoding gap exist?—a case study in blue butterflies (Lepidoptera: Lycaenidae). *Front Zool* 4: 8.
- 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.
- Zou S, Li Q, Kong L (2011) Multigene Barcoding and Phylogeny of Geographically Widespread Muricids (Gastropoda: Neogastropoda) Along the Coast of China. *Mar Biotechnol* 14: 21–34.
- 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(10): e26619.
- DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Phil Trans R Soc B* 360: 1905–1916.
- 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.
- Nielsen R, Matz M (2006) Statistical approaches for DNA barcoding. *Syst Biol* 55: 162–169.

22. Knowles LL, Carstens BC (2007) Delimiting species without monophyletic gene trees. *Syst Biol* 56: 887–895.
23. 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.
24. 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 B* 275: 237–247.
25. Reid BN, Le M, McCord WP, Iverson JB, Georges A, et al. (2011) Comparing and combining distance-based and character-based approaches for barcoding Turtles. *Mol Ecol Resour* 11: 956–967.
26. Cernohorsky WO (1984) Systematics of the family Nassariidae (Mollusca: Gastropoda). *Bull Auckl Inst Mus* 14: 1–356.
27. Cossmann M (1901) *Essais de Paléoconchologie Comparée*, volume 4. Self-published, Paris, 293.
28. Cernohorsky WO (1972) Indo-Pacific Nassariidae (Mollusca: Gastropoda). *Rec Auckland Inst Mus* 9: 125–194.
29. Boss KJ (1982) Phylum Mollusca. In: Parker SP, ed. *Synopsis and classification of living organisms*, McGraw-Hill, New York. 945–1166.
30. Svavarsson J, Granmo A, Ekelund R, Szpunar J (2001) Occurrence and effects of organotin on adult common whelk (*Buccinum undatum*) (Mollusca: Gastropoda) in harbours and in a simulated dredging situation. *Mar Pollut Bull* 42 (5): 370–376.
31. Xu J, Xu G, Chen Y, Qin P, Yu M, et al. (2007) Correlation between toxicity of poisonous *Nassarius* Sp and their habitats. *Chin J Health Lab Technol* 17(1): 63–67.
32. Zhang N, Su J, Liu H, Ye S, Li L, et al. (2009) The species and toxicities of Nassariidae collected from the coast of Southeast China Sea. *Asian J Ecotoxicol* 4(2): 289–294.
33. Haasl DM (2000) Phylogenetic relationships among Nassariid gastropoda. *J Vertebr Paleontol* 74(5): 839–852.
34. Wilke T, Falniowski A (2001) The genus *Adriolydrobia* (Hydrobiidae: Gastropoda): polytypic species or polymorphic populations? *J Zool Syst Evol Res* 39: 227–234.
35. Teske PR, Papadopoulos I, McQuaid CD, Newman BK, Barkeret NP (2007) Climate change, genetics or human choice: Why were the shells of mankind's earliest ornament larger in the Pleistocene than in the Holocene? *PLoS ONE* 2(7): e614.
36. Chiu YW, Chen HC, Lee SC, Chen CA (2002) Morphometric analysis of shell and operculum variations in the viviparid snail, *Cipangopaludina chinensis* (Mollusca: Gastropoda). *Zool Stud* 41(3): 321–331.
37. Barges MD, Mas-Coma S (1997) Phylogenetic analysis of Lymnaeid snails based on 18S rDNA sequences. *Mol Biol Evol* 14(5): 569–577.
38. Li H, Lin D, Fang H, Zhu A, Gao Y (2010) Species identification and phylogenetic analysis of genus *Nassarius* (Nassariidae) based on mitochondrial genes. *Chin J Oceanol Limnol* 28(3): 565–572.
39. Winnepeninckx B, Backeljau T, De Wachter R (1993) Extraction of high molecular weight DNA from molluscs. *Trends Genet* 9: 407.
40. Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol* 537: 39–64.
41. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17: 540–552.
42. Tamura K, Dudley J, Nei M, Kumar S (2007) Mega 4: molecular evolutionary genetics analyses (mega) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
43. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
44. Posada D (2008) jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25: 1253–1256.
45. Sarkar IN, Planet PJ, Desalle R (2008) CAOS software for use in character-based DNA barcoding. *Mol Ecol Resour* 8: 1256–1259.
46. Bergmann T, Hadrys H, Breves G, Schierwater B (2009) Character-based DNA barcoding: a superior tool for species classification. *Berl Münch tierärztl* 122: 446–450.
47. Maddison WP, Maddison DR (2005) *MACCLADE: Analysis of Phylogeny and Character Evolution*. Version 3.0. Sinauer Associates, Sunderland, Massachusetts.
48. Dayrat B (2005) Towards integrative taxonomy. *Biol J Linn Soc* 85: 407–415.
49. Rubinoff D (2006) DNA barcoding evolves into the familiar. *Conserv Biol* 20: 1548–1549.
50. Vogler AP (2006) Will DNA barcoding advance efforts to conserve biodiversity more efficiently than traditional taxonomic methods? *Front Ecol Environ* 5: 270–272.
51. Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proc R Soc B* 265: 2257–2263.
52. Hebert PDN, Stoeckle MY, Zemlack TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLoS Biol* 2: 1657–1663.
53. Wang W, Cai L, Liu W (2007) Morphological Classification of Nassariids in Fujian Coast. *J. Xiamen University (Nat Sci)* 46: 171–175.
54. Ross HA, Murugan S, Li WLS (2008) Testing the reliability of genetic methods of species identification via simulation. *Syst Biol* 57: 216–230.
55. Goldstein PZ, DeSalle R (2010) Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description. *Bioessays* 33: 135–147.
56. 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.
57. Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In: Hillis, D., Moritz C. (Eds.), *Molecular systematics*. Sinauer, Sunderland, pp. 205–247.
58. Armbruster GFJ, van Moorsel CHM, Gittenberger E (2000) Conserved sequence patterns in non-coding ribosomal ITS-1 of distantly related snail taxa. *J Moll Stud* 66: 570–573.
59. Van Moorsel CHM, van Nees WJ, Mengens HJ (2000) A quick, simple, and inexpensive Mollusc DNA extraction protocol for PCR-based techniques. *Malacologia* 42: 203–206.