Limitations of cytochrome oxidase I for the barcoding of Neritidae (Mollusca: Gastropoda) as revealed by Bayesian analysis

S.Y. Chee

Center for Marine and Coastal Studies, Universiti Sains Malaysia, Penang, Malaysia

Corresponding author: S.Y. Chee
E-mail: suyinchee@usm.my

ABSTRACT. The mitochondrial DNA (mtDNA) cytochrome oxidase I (COI) gene has been universally and successfully utilized as a barcoding gene, mainly because it can be amplified easily, applied across a wide range of taxa, and results can be obtained cheaply and quickly. However, in rare cases, the gene can fail to distinguish between species, particularly when exposed to highly sensitive methods of data analysis, such as the Bayesian method, or when taxa have undergone introgressive hybridization, over-splitting, or incomplete lineage sorting. Such cases require the use of alternative markers, and nuclear DNA markers are commonly used. In this study, a dendrogram produced by Bayesian analysis of an mtDNA COI dataset was compared with that of a nuclear DNA ATPS-α dataset, in order to evaluate the efficiency of COI in barcoding Malaysian nerites (Neritidae). In the COI dendrogram, most of the species were in individual clusters, except for two species: *Nerita chamaeleon* and *N. histrio*. These two species were placed in the same subcluster, whereas in the ATPS-α dendrogram they were in their own subclusters. Analysis of the ATPS-α gene also placed the two genera of nerites (*Nerita* and *Neritina*) in separate clusters, whereas COI gene

analysis placed both genera in the same cluster. Therefore, in the case of the Neritidae, the ATPS-α gene is a better barcoding gene than the COI gene.

Key words: Bayesian analysis; Cytochrome oxidase I; ATPS-α; Molecular taxonomy; Neritidae

INTRODUCTION

DNA barcoding is a technique used to characterize specimens of organisms using a short DNA sequence from a standard position in the genome. The mitochondrial DNA (mtD-NA) gene cytochrome oxidase I (COI) is becoming the standard barcode region for higher animals, as it is a relatively short sequence (648 nucleotide base pairs in most groups) and can be obtained reasonably quickly and cheaply. However, there are limitations of COI as a barcoding gene. The first reservation is its inability to distinguish between species if introgressive hybridization, over-splitting, or incomplete lineage sorting occurs (Steinke et al., 2009). Secondly, the COI gene is not sensitive enough to discriminate between species that have recently evolved, and thirdly, if the species in question exhibits extensive spatial differentiation, discovering new species is difficult (Hickerson et al., 2006; Rubinoff et al., 2006). In addition, COI gene amplification is not always successful for species-level identification. Some studies, including Meier et al. (2006) and Johnson et al. (2008), have not been completely successful in identifying species using DNA barcoding methods that involved COI. In studies such as Pleijel et al. (2008), which used the same method, specimens were misidentified.

In such cases alternative markers are required, such as nuclear DNA (nDNA) markers, as they are less sensitive to introgression or incomplete lineage sorting and are more appropriate for recently diverged species (Baker et al., 2009; Zhang, 2011). According to April et al. (2011), for morphologically distinct species that remain genetically indistinguishable, it is highly unlikely that they would be delineated, as they could be derived from the same gene pool. Indeed, nuclear genes have been used in identifying indistinguishable species of freshwater fish (April et al., 2011) when COI could not delineate them. Nuclear genes have also been used with mitochondrial genes to provide accurate genetic identification and overcome the problem of misinterpretation caused by hybridization and introgression in museum samples of sturgeons (Garrido-Ramos et al., 2009). In the barcoding and identification of 17 known and easily confused species of Muricidae species from China, nuclear genes, and COI, were used to offer an analytically powerful addition or even an alternative (Zou et al., 2012).

The specimens used in the present study were nerites (Neritidae) from the Prosobranchia family. Nerites are operculate sea snails with a short spiral shell. They are widely distributed in tropical and temperate coastal waters. Some species are found on rocky and/or sandy shores, whereas others can only be found in muddy mangroves, and are commonly spotted hiding in crevices to protect themselves from predators and physical stress. They are gregarious herbivores, and play important roles in the food web and as bioindicators. Some of the more abundant species are important to algal community structure in coastal areas. Nerites are abundant, easy to collect, hardy, and relatively long-lived. To date, no studies concerning the species identification of Malaysian nerites have been conducted; the paucity of gastropod taxonomists has also hampered the process of determining their biodiversity. This study attempted to identify nerites found in Malaysian intertidal zones, using the barcoding mtDNA

COI gene and the nDNA ATPS-α gene.

Bayesian data analysis was employed in this study. Bayesian refers to methods in probability and statistics, and is considered to be fundamentally sound, very flexible, produces clear and direct inferences, and makes use of all available information (Pernestal, 2009). This method has been used to provide accurate estimations of the evolutionary rates, morphological parallelism, and biogeography of the Littorininae (Williams et al., 2003); phylogenetic relationships and speciation in calyptraeid gastropods (Collin, 2003); and the phylogeny and taxonomy of topshells (Trochidae) (Donald et al., 2005). We tested the sensitivity of COI for barcoding using this established mode of data analysis.

MATERIAL AND METHODS

Sampling

Samples were collected from selected beach areas along the coasts of the Malaysian Peninsular, Sarawak, and Sabah, providing a complete geographical coverage throughout the distribution of nerites. In total, 14 locations were sampled. Samples of each morphospecies were collected from every location they could be found to avoid overlooking cryptic species. When possible, at least five individuals per morphospecies were collected per site. Photographs of dorsal and ventral views were taken of each species to aid morphological descriptions and to catalogue the diversity of the Neritidae. Dissection was conducted in the field whenever possible, or in the laboratory, after the muscles were relaxed using 10% $MgCl_2$ diluted in double-distilled water (ddH_2O). The foot tissues of each specimen were then cut and placed into 1.5-mL tubes filled with absolute ethanol (99%), and subsequently stored at 27°C.

Molecular analysis

DNA was extracted from 2 mm^3 of foot tissue using an AquaGenomic™ kit, following the protocol provided by the manufacturer (MultiTarget Pharmaceuticals LLC, Colorado Springs, CO, USA). The quality and quantity of the DNA extracted was checked at two wavelengths (260 and 280 nm) using a spectrophotometer (U-1900, Hitachi). Each DNA stock sample was diluted to a concentration of 50 ng. All DNA samples were then stored in a refrigerator at -2°C.

The COI gene was amplified with the following primers: HCO2198 5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3' and LCO1490 5'-GGT CAA ATC ATA AAG ATA TTG G-3' (Folmer et al., 1994). Each polymerase chain reaction (PCR) contained 2.5 µL 10X PCR buffer, 2.5 µL 25 mM $MgCl_2$, 0.5 µL 2.5 mM dNTPs, 0.25 µL 10 µM forward and reverse primers, 0.1 µL 5 U *Taq* polymerase, and 1.0 µL 50 ng DNA template. ddH_2O was added to make a total volume of 25 µL. Amplification consisted of 40 cycles of denaturation at 94°C for 30 s, annealing at 45°C for 1 min, and extension at 72°C for 1 min.

The nuclear ATPS-α gene was amplified using the following primers: ATPSαf1 5'-GAG CCM ATG CAG ACT GGT ATT AAG GCY GT-3' and ATPSαr1 5'-TTG AAN CKC TTC TGG TTG ATG GTG TC-3' (Jarman et al., 2002). Each PCR contained 2.5 µL 10X PCR buffer, 2.5 µL 25 mM $MgCl_2$, 1.0 µL 2.5 mM dNTPs, 1.0 µL 10 µM forward and reverse primers, 0.2 µL 5 U *Taq* polymerase, and 1.0 µL 50 ng DNA template. ddH_2O was added to make a total volume of 25 µL. The PCR amplifications consisted of an initial denaturation at 94°C for 5 min, 35 cycles of denaturation at 94°C for 45 s, annealing at 60°C for 45 s, extension at

72°C for 1 min, and a final extension at 72°C for 7 min. All of the PCRs were conducted in a G-Storm™ Thermal Cycler GS4822 (Somerset, United Kingdom).

Prior to sequencing, the PCR products were purified using a Wizard® Genomic DNA Purification Kit (Promega) to remove residual primers and dNTPs. The purified products were then sent for sequencing at the Universiti Sains Malaysia Center for Chemical Biology facility and run on a 3130 Capillary Electrophoresis Genetic Analyzer (Applied Biosystems Inc.), following the manufacturer protocol. Unique sequences were deposited in GenBank and BOLD systems (version 2.5).

Data analysis

Sequences were edited and trimmed using MEGA (version 5.10) (Tamura et al., 2007), aligned with ClustalW, and checked for accuracy by eye in Chromas (version 2.33) (Technelysium, 2009). All sequence identities were confirmed using BLAST in GenBank. For the COI dataset, the outgroups included in the analysis were the Guamanian nerite *Clithon oualaniensis* (EU732364), the bathyal gastropod *Bathynerita naticoidea* (FJ977768), and the Danube nerite *Theodoxus danubialis* (GQ365726). For the ATPS-α dataset, the horned nerite *Clithon corona* (EU732513), *Vittina variegata* (EU732517), and the blacklip abalone *Haliotis rubra* (AY043205) were included as outgroups.

The COI and ATPS-α datasets were then analyzed using the MrBayes software (Huelsenbeck and Ronquist, 2001) for the inference and estimation of phylogeny, using the Markov chain Monte Carlo (MCMC) simulation technique to approximate the posterior probabilities of trees. The number of generations included was based on the standard deviation of the split frequencies (below 0.01). A plot of generation against the log probability of the data (the log likelihood value) was generated using the sump command. The distribution of the plot was checked to ensure that the runs were adequate for tree-building. A tree was then constructed using the sumt command that generated output files. The FigTree (version 1.3.1) software (Rambaut, 2009) was used to run the file, which had the suffix .con. The consensus tree that was built was then added, with its confidence values generated by MrBayes. A representative sequence for each species was included from GenBank to confirm the phylogenetic placement of species in the tree.

RESULTS

Nine species of nerite were used to build dendrograms by the Bayesian method. Of these nine species, eight were from the genus *Nerita* (*N. costata, N. undata, N. albicilla, N. balteata, N. litterata, N. chamaeleon, N. histrio,* and *N. planospira*) and one was from the genus *Neritina* (*N. violacea*).

Four clusters were produced in the dendrogram built using the COI dataset (Figure 1). The first cluster consisted of *N. costata*, the second of *N. undata* and *N. albicilla*, and the third of *N. balteata, N. litterata, N. chamaeleon*, and *N. histrio*. The fourth cluster consisted of *N. planospira* and *N. violacea*. In the third cluster, *N. balteata* and *N. litterata* formed their own subcluster. Most of the specimens of *N. chamaeleon* formed their own cluster, but one specimen (a sequence obtained from GenBank) was found in the same subcluster as *N. histrio*. When more *N. chamaeleon* sequences were added to the dataset, some specimens from this study also grouped together with *N. histrio* (tree not shown). The bootstrap values for the tree

were high in most cases, with the lowest value being 52% and the highest being 99%.

The dendrogram built with the ATPS-α dataset also produced four clusters (Figure 2). The first cluster consisted of *N. planospira*, the second of *N. costata, N. chamaeleon, N. histrio, N. undata, N. litterata*, and *N. balteata*, the third of *N. albicilla*, and the fourth of *N. violacea*. In the second cluster, every species formed its own subcluster, as did *N. histrio* and *N. chamaeleon*. The bootstrap values were high, and most of them reached 100% confidence.
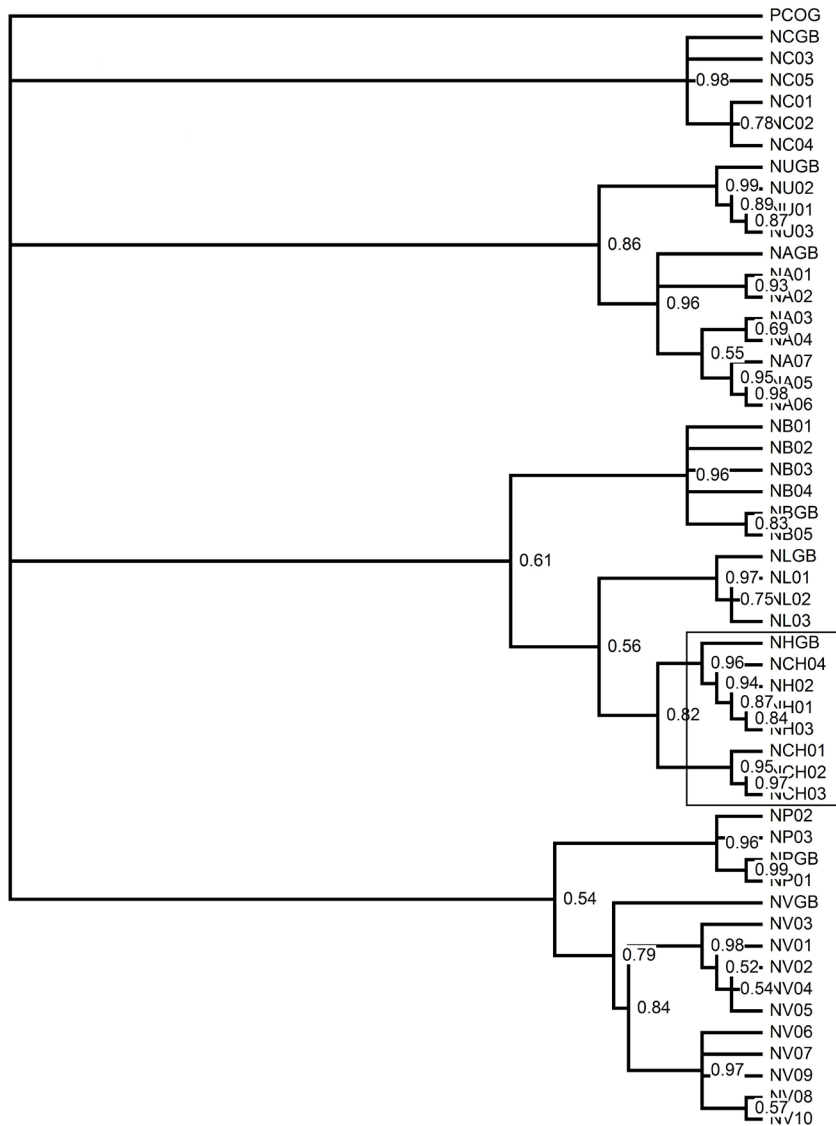

Figure 1. Bayesian tree built using COI dataset. NA = *Nerita albicilla*, NB = *Nerita balteata*, NC = *N. costata*, NCH = *Nerita chamaeleon*, NH = *Nerita histrio*, NL = *Nerita litterata*, NP = *Nerita planospira*, NU = *Nerita undata*, NV = *Neritina violacea*. Outgroups: TDOG = *Theodoxus danubialis*, BNOG = *Bathynerita naticoidea*, COOG = *Chlithon oualaniensis*.
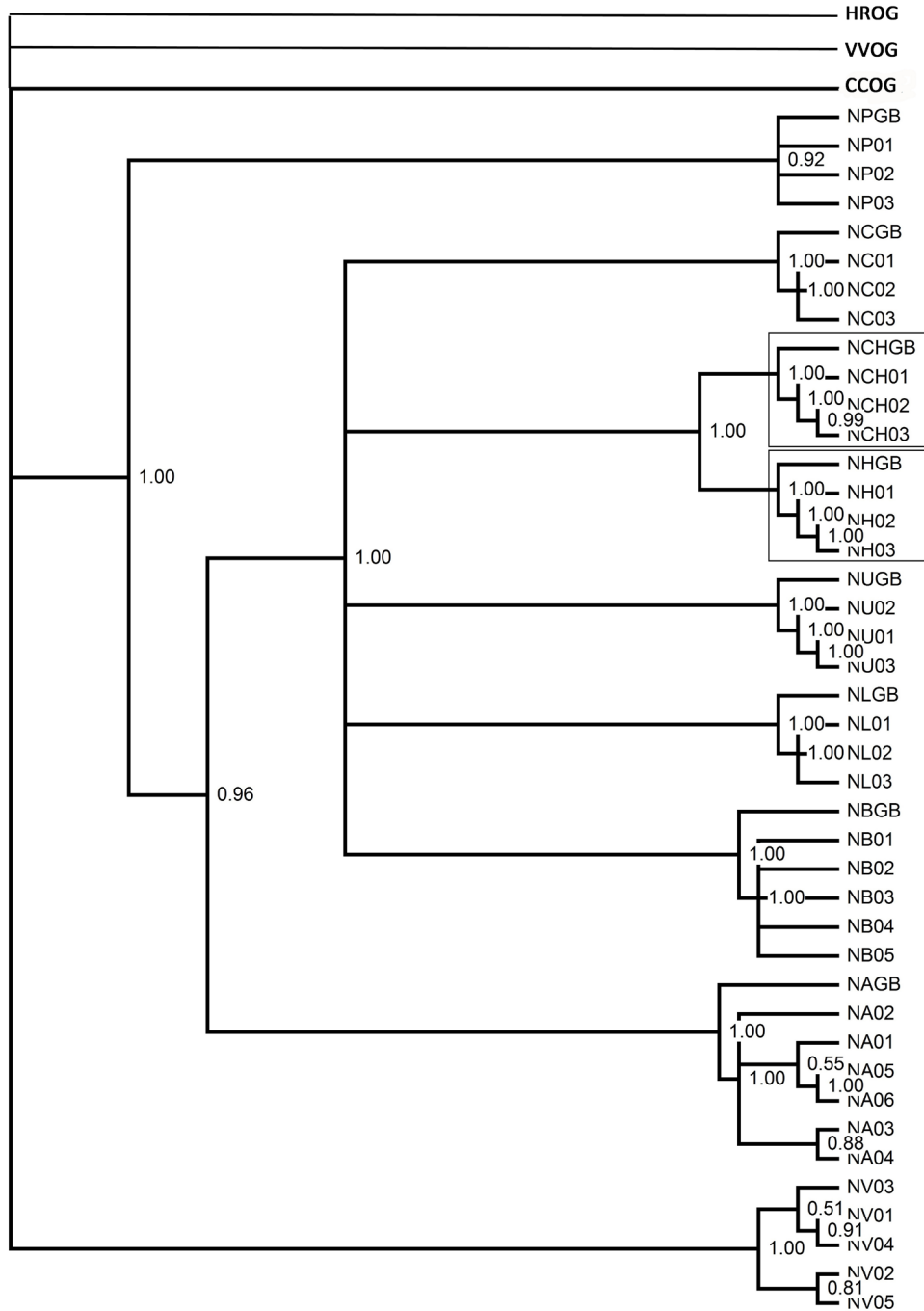
Figure 2. Bayesian tree built using ATPS-α dataset. NA = *Nerita albicilla*, NB = *Nerita balteata*, NC = *N. costata*, NCH = *Nerita chamaeleon*, NH = *Nerita histrio*, NL = *Nerita litterata*, NP = *Nerita planospira*, NU = *Nerita undata*, NV = *Neritina violacea*. Outgroups: CCOG = *Clithon corona*, VVOG = *Vittina variegata*, HROG = *Haliotis rubra*.

DISCUSSION

The nDNA ATPS-α gene proved a more powerful tool than the mtDNA COI gene for the identification of nerites. The bootstrap values produced in its dendrogram showed a high confidence of the taxonomic placement of the nerites included in this study. This gene was able to identify *N. chamaeleon* and *N. histrio* as individual species, whereas the COI gene was not. Moreover, the COI barcoding gene also placed *N. planospira* and *N. violacea* in the same cluster. An independent study of the genetic distances of all of the species included in the present study revealed that *N. planospira* was significantly different from the rest, even from those within the same genus (Frey and Vermeij, 2008). Such observations were also made by Frey and Vermeij (2008), leading to the perception that it is genetically quite different to other *Nerita* spp. Based on the interspecific values obtained, it is possible for *N. planospira* to be placed within the *Neritina*. Remarkably, the ATPS-α gene was able to distinguish this nerite by placing it in a separate cluster. The genetic differences between *N. planospira* and the other *Nerita* spp can also be seen in the ATPS-α dendrogram, where it is placed separately from the rest of the genus. *N. violacea* in this dendrogram, however, was isolated, and formed an entirely different cluster – confirming that it is within a separate genus.

CONCLUSIONS

The COI gene, which is the universal barcode for animals, and the nuclear gene ATPS-α have demonstrated their ability to identify nerite species. Although DNA barcoding is deemed a powerful tool in species identification, it does have its limitations. Nevertheless, it still represents an efficient tool in understanding the extent of biodiversity, and provides a simple way to identify species. Regarding the universality of priming sites, COI surpasses ATPS-α, but in terms of species resolution, ATPS-α is superior over COI. However, ATPS-α for species identification in the Neritidae should be undertaken with caution, as it is a rapidly evolving marker that may pose problems, particularly at the base of a tree. The results of this study provide useful data for future studies that involve DNA barcoding and gastropod identification.

ACKNOWLEDGMENTS

REFERENCES

April J, Mayden RL, Hanner RH and Bernatchez L (2011). Genetic calibration of species among north america's freshwater fishes. *Proc. Natl. Acad. Sci. U. S. A.* 108: 10602-10607.

Baker AJ, Tavares ES and Elbourne RF (2009). Countering criticisms of single mitochondrial DNA gene barcoding in birds. *Mol. Ecol. Resour.* 9: 257-268.

Collin R (2003). Phylogenetic relationships among calyptraeid gastropods and their implications for the biogeography of

marine speciation. *Syst. Biol*. 52: 618-640.

Donald KM, Kennedy M and Spencer HG (2005). The phylogeny and taxonomy of austral monodontine topshells (Mollusca: Gastropoda: Trochidae), inferred from DNA sequences. *Mol. Phylogenet. Evol*. 37: 474-483.

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

Frey MA and Vermeij GJ (2008). Molecular phylogenies and historical biogeography of a circumtropical group of gastropods (Genus: Nerita): Implications for regional diversity patterns in the marine tropics. *Mol. Phylogenet. Evol*. 48: 1067-1086.

Garrido-Ramos MA, Robles R, de la Herran R, Martinez-Espin E, et al. (2009). Analysis of mitochondrial and nuclear DNA markers in old museum sturgeons yield insights about the species existing in Western Europe: *A. sturio*, *A. naccarii* and *A. oxyrinchus*. In: Biology, conservation and sustainable development of sturgeons, Fish and Fisheries Series Vol. 29 (Carmona R, Domezain A, García-Gallego M, Hernando JA, et al., eds.). Springer, Berlin, 25-49.

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

Huelsenbeck JP and Rondquist F (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.

Jarman SN, Ward R and Elliott NG (2002). Oligonucleotide primers for PCR amplification of coelomate introns. *Mar. Biotechnol*. 4: 347-355.

Johnson SB, Waren A and Vrijenhoek RC (2008). DNA barcoding of *Lepetodrilus* limpets reveals cryptic species. *J. Shellfish Res*. 27: 43-51.

Meier R, Shiyang K, Vaidya G and Ng PKL (2006). DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Syst. Biol*. 55: 715-728.

Pernestal A (2009). Probabilistic fault diagniosis with automotive applications. LiU-Tryck, Linkoping, Sweden.

Pleijel F, Jondelius U, Norlinder E, Nygren A, et al. (2008). Phylogenies without roots? A plea for the use of vouchers in molecular phylogenetic studies. *Mol. Phylogenet. Evol*. 48: 369-371.

Rambaut A (2009). FigTree v1.3.1. Accessed October 12, 2011. Available at: http://tree.bio.ed.ac.uk/software/figtree/.

Rubinoff D, Cameron S and Will K (2006). A genomic perspective on the shortcomings of mitochondrial DNA for "Barcoding Identification." *J. Hered*. 97: 581-594.

Steinke D, Zemlak TS and Hebert PDN (2009). Barcoding Nemo: DNA-based identifications for the ornamental fish trade. *PLoS ONE* 4: e6300.

Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol*. 24:1596-1599.

Technelysium (2009). Chromas. Available at [http://technelysium.com.au/?page_id=13]. Accessed January 17, 2009.

Williams ST, Reid DG and Littlewood DTJ (2003). A molecular phylogeny of the Littorininae (Gastropoda: Littorininae): unequal evolutionary rates, morphological parallelism, and biogeography of the Southern Ocean. *Mol. Phylogenet. Evol*. 28: 60-86.

Zhang J (2011). Species identification of marine fishes in China with DNA barcoding. *Evidence-based Comp. Altern. Med*. doi: 10.1155/2011/978253.

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