DEVELOPING MONITORING TOOLS FOR TOMORROW'S INVASIVES:
SPECIES LISTS, DNA BARCODES, AND IMAGES FOR ORNAMENTAL FISH

YI YOUGUANG

NATIONAL UNIVERSITY OF SINGAPORE

2014

DEVELOPING MONITORING TOOLS FOR TOMORROW'S INVASIVES:
SPECIES LISTS, DNA BARCODES, AND IMAGES FOR ORNAMENTAL FISH

YI YOUGUANG

(B.Sc. (Hons 2nd Upper), NUS)

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2014

ACKNOWLEDGEMENTS

Table of Contents

Summary

Ornamental fish are a major source of invasive species in freshwater habitats. In order to control and monitor introductions, it is important to know which species are in the trade and to develop identification tools for these species. Here I first study the species diversity in the trade by comparing two published lists with trade data for Singapore (2009-2011). I establish that a very large number of species (4769) are being traded, the lists and trade data are inconsistent, many species in Singapore's trade are wild-caught, and that new species are continuously added. I then image and generate DNA barcodes for 1448 specimens belonging to 554 species of which 334 species had not previously been barcoded. The images are used to build an online image database for ornamental fish while the DNA barcodes are used for testing species-specificity at three levels; local, global, and systematic. First, I establish whether DNA barcodes can be used for identifying the 89 of the 105 freshwater fish species living in Singapore. An identification efficiency of 77% to 89% indicates that COI can be used to allocate specimen to species at an island scale. I then determine identification success rates of DNA barcodes at a global scale based on all my data and all available COI sequences in Genbank. An identification efficiency of 77% to 91% indicates that COI can be used to allocate specimen to species at a global scale. Lastly, I collaborate with colleagues in New Zealand to test whether DNA barcodes are diagnostic for cypriniform fishes. An identification

efficiency of 90% to 99% is established for the 172 ornamental cyprinid

fish species sampled. Results indicate that COI can be used effectively for

identifying fish at local, global and systematic level.

Lists of tables

Lists of figures

Publication in this thesis

1. Collins, R. A., K. F. Armstrong, R. Meier, **Y. Yi**, S. D. J. Brown, R. H. Cruickshank, S. Keeling and C. Johnston (2012). "Barcoding and Border Biosecurity: Identifying Cyprinid Fishes in the Aquarium Trade." <u>Plos One</u> 7(1).

CHAPTER I

_____

General Introduction

1.1. Introduction to DNA barcoding and its applications

One of the main problems faced by biologists today is the taxonomic impediment; there are simply too many species, but too few taxonomists to discover, describe and identify all the specimens that are collected by applied biologists such as those interested in bio-protection (Ball & Armstrong, 2008; Bleeker et al., 2008; Chown et al., 2008), conservation biologists (Blaxter, 2006; Holmes et al., 2009; Logan et al., 2008) and scientists identifying food items (Logan et al., 2008; Yancy et al., 2008). They all agree that there are too few taxonomic experts thus creating an imbalance between needs and availability of taxonomic expertise (Tautz et al., 2003). One solution that is promoted by Hebert (2003) under the name "DNA barcoding" is to use 650bp piece of cytochrome oxidase I (COI) to identify and delimit species; i.e., DNA barcoding has been proposed as a remedy for resolving the taxonomic impediment (Hebert et al., 2003a; Hebert et al., 2003b). Hebert (2003) assumes that inter- and intraspecific distances are non-overlapping constituting a barcoding gap; a senario which makes DNA barcoding a perfect application. Hebert's proposal sparked off a decade-long debate over the strength and weaknesses of DNA barcodes.

Meyer & Paulay (2005) pointed out flaws in the current methodology that most proponents of barcoding have used for delimiting species and discover cryptic species. Species concepts were

rarely specified and barcoding researchers seemed to delimit species based on undefined concepts. Meyer also demonstrated that DNA barcoding can only yield high identification success rates if sampling was complete. In a study involving a well sampled group of marine gastropod – delimited based on phylogenetic species concept – misidentifications were 4% and 17% in phylogenetically well and poorly sampled groups respectively (Meyer & Paulay, 2005). While morphologically over-split species will suffer from an artificially low interspecific distance, over lumped morpho- and cryptic species will exhibit high intraspecific variation, often aiding in their discovery. However, morphologically well defined but recently radiated species could suffer from incomplete lineage sorting; a natural phenomenon that could give rise to high intraspecific variation when both derived and ancestral alleles were sampled within species, and low interspecific variations when ancestral alleles were sampled across species. Despite rarer, convergence might cause distantly related species to share COI with similar sequence.

Many researchers have since pointed out that DNA barcoding can only be successful if it is based on a solid taxonomic foundation, which is elusive for many taxa given that most animal species are undescribed and few are well studied. This also applies to the numerous species in the ornamental species trade that has recently become a source of many invasive species.

Many criticisms of DNA barcoding have been methodological and numerous researchers have pointed out that the analysis techniques were poorly developed. In addition, as more sequences have become available, the initial proposal of a universal COI barcode for each species was revealed to be incorrect. Indeed, quite a few studies have provided evidence that COI has limitations for species identification and delimitation, and that there is no barcoding gap in most taxa. For example, Mallet and Willmott (2003) mentioned that closely related species often share COI sequences and that a tendency to hybridize can make the situation even more confusing. Meier et al. (2006) pointed out that the lack of a barcoding gap was even more apparent when the smallest interspecific pairwise distances was used instead of average pairwise distances.

Further studies by Wiemers and Fiedler (2007) demonstrated that not all butterflies can be readily identified by their COI DNA barcode. Their analysis showed that there was an 18% overlap between the intra- and interspecific COI sequence divergence due to low interspecific divergence between many closely related species in the Lycaenidae which includes the well-sampled clade of *Agrodiaetus*. The authors showed that the lack of a barcoding gap resulted in a misidentification rate of 16%. Wiemers and Fiedler (2007) concluded that the "barcoding gap" is an artefact of insufficient sampling across taxa (Martin, 2007). Another test of the applicability of DNA barcoding to a diverse community of butterflies from the upper Amazon only

yielded a 77% identification success rate, a figure that dropped to 68% for species represented in the analyses by more than one geographical race and at least one congener (Elias et al., 2007). These studies as well as many other studies on Lepidoptera (Kaila & Stahls, 2006; Roe & Sperling, 2007) indicated that the initial claim of 100% identification success for lepidopterans was due to insufficient sampling.

Many other barcoding studies have also subsequently revealed that not all groups of birds, mammals and insects can be successfully identified based on DNA barcodes. Some congeneric species of New Zealand grasshoppers (Orthoptera: Acrididae) (Trewick, 2008) within the genus *Sigaus* possess similar DNA barcodes while *Sigaus australis* has more than one mitochondrial haplotype. Studies also revealed that COI alone cannot be used for successfully identifying parapatric avian species and that more than one gene was needed (Aliabadian et al., 2009). DNA barcoding also has its limitation for identifying different groups of Diptera. Many species have high intraspecific pairwise distances that result in low identification success rates of 65% (Meier et al., 2006). Other studies showed that DNA barcodes cannot be reliably used to identify species of the blowfly genus *Protocalliphora* (Whitworth et al., 2007) while COI DNA barcodes were shown to be effective in identifying some of the species within the Pipunculidae (big headed flies).

An additional problem for DNA barcoding is the occasional presence of nuclear copies of mitochondrial DNA (NUMT) (Song et al., 2008; Buhay, 2009). This is particularly well documented for grasshoppers, crustaceans, and primates. In order to circumvent this problems, primate specific primers had to be designed and reverse transcription was used for amplification (Lorenz et al., 2005). Fortunately for fish, there seems to be no evidence of NUMTs and all previous reports of NUMTs in *Fugu* were shown to be erroneous and due to aligning mtDNA with nuclear DNA (Antunes & Ramos, 2005; Venkatesh et al., 2006).

Ten years after proposing DNA barcodes, it is becoming clear that the technique works for most but not all species of fish (Ward et al., 2005; Ivanova et al., 2007; Hubert et al., 2008;), birds (Yoo et al., 2006; Rudnick et al., 2007; Dove et al., 2008; Johnsen et al., 2010;) and butterflies (Janzen et al., 2009; Lohman & Samarita, 2009; Hausmann et al., 2011;), while it has lower success rates in other taxa such as cnidarians and crustaceans (Buhay, 2009) and sepsids (Meier et al., 2004; Meier et al., 2006). While it has been quite clear that the taxonomic impediment cannot be fully removed by the use of half a gene segment, DNA barcodes have proven useful for many purposes. Some of the uncontroversial applications are matching of life history stages (Victor et al., 2009; Valdez-Moreno et al., 2010; Victor et al., 2010), verifying the identity of food sources (Wong & Hanner, 2008; Chen et al., 2009;), and monitoring the movement of endangered and

6

invasive species in the wildlife trade (Bleeker et al., 2008; Chown et al., 2008).

Currently, the main challenge for DNA barcoding is the sparse species coverage in the available public databases (GenBank and BOLD). Species without barcodes cannot be identified and barcodes for only ca. 60,000 of the 1.5 million described animal species are publically available via Genbank (Kwong et al., 2012). In the following few paragraphs, I will discuss the challenges and opportunities for using COI for monitoring invasive species in the ornamental fish trade.

1.2. Establishing what are the species in the ornamental fish trade (Chapter II)

The aquarium trade is a major source of invasive species. Most of the species in the trade are tropical fish and much of the trade is conducted in the tropics which makes the accidental release and establishment of species in tropical water systems very likely (Paine, 1966; Moynihan, 1971). Hence, there is a need to monitor the movement of trade fish in order to prevent invasive species from destroying native habitats. A common strategy by governmental agencies is to monitor and regulate the trade via lists of approved or disallowed species (FISORNIC.ALL, 2011;

http://www.cefas.defra.gov.uk/;

http://www.dpi.nsw.gov.au/fisheries/pests-diseases/noxious-fish-and-marine-vegetation), but it is unclear how accurate and complete these lists are. In chapter two, I tested the completeness by first comparing the consistency between two published lists for ornamental fish and then comparing both to the list of species that were traded in Singapore between 2009 and 2012. The comparison with the Singapore trade is useful because Singapore is globally one of the most active trading hubs for ornamental fish.

1.3. DNA barcoding as a solution for monitoring invasive species (Chapters III, IV & V)

DNA barcoding is a possible solution for monitoring the ornamental fish trade and identifying species introductions. However, this requires barcode databases with good nominal species coverage (Genbank and BOLD). While the fish barcoding campaign "FISH-BOL" estimates that there are DNA barcodes for about 10,267 fish species in their database (www.Fish-BOL.org), the number of publically available COI sequences in Genbank is only 8,327 species. Prior to my thesis, it was unknown how many of these species are aquarium fish species. In chapter IV, I investigated whether the species coverage of aquarium fish COI in both databases are broad enough for monitoring invasive species that originate from the ornamental trade.

A recent survey of Singapores' water system reveals that exotic species constitute 70% of Singapores' local fish diversity (Baker & Lim, 2008; Ng & Tan, 2010; Yi et al., 2012). While many are hypothesized to be invasive, some have already established breeding populations in Singapores' water systems. Singapore is also the largest trading hub of ornamental fish in the world (Livengood & Chapman, 2009), which suggests that the trade may be the source of these non-native species; either through accidental release by wholesaler or through release by hobbyists. Since many scientists have proposed that DNA barcoding will be more effective at a regional scale, I tested in Chapter III whether all freshwater fish species found in Singapore can be identified based

on DNA barcodes. This test was applied to a curious assemblage of species because Singapore has more exotic than native species. Chapter III provides insights into what are the identification success rates for DNA barcodes and since many of the species are non-native, the chapter also provides information on whether COI could be used effectively to detect and monitor invasive fish in Singapore.

The fish diversity in the ornamental trade is known to be high. The freshwater fish diversity in the trade is recorded by Ornamental Fish International to include 4,769 species, which is approximately one sixth of all described fish diversity (28,000 to 32,700 species) and one third of freshwater fish diversity on earth (11,676 to 13,635 species) (Axelrod et al., 2007; Froese et al., 2013; Nelson, 2006). This amazing diversity of fish in the ornamental trade provides us with the opportunity to collect many barcodes quickly. In Chapter IV, I created a COI database for 522 species of freshwater fish from the Singapore ornamental trade and test the identification efficiency of COI for my dataset and all sequences in Genbank. In addition, I provided high quality images for voucher specimens that are provided online to supplement the DNA sequences. Both the aquarium fish COI database and the image database will serve as important tools to monitor and regulate the movements of invasive species in the highly mobile ornamental fish trade.

Chapters III and IV also address which analysis technique should be used for species identifications based on DNA barcodes. Some very popular methods require global alignments (e.g., Best close match, Neighbour-joining) before a query sequence can be matched to a species. However, generating these alignments and analyzing large datasets containing substantial numbers of sequences can be time consuming and requires large amounts of computational power. Hence, computational biologists have proposed alternatives that are heuristic and do not require a global alignment (Little, 2011). In the third and fourth chapters, I investigate and compare the efficiency of these different methods of analyses: 1) global alignment-based methods involving "best match" and "best close match", 2) BLAST; a heuristic method based on pairwise alignments, and 3) BRONX, a method based on small diagnostic markers.

Nations with aquaculture and agricultural resources often show high levels of concern with regard to biological invasion because it his high priority to protect their environment and the commercially important species (e.g., salmonids). The New Zealand authorities are particularly concerned about the possibility of cyprinids in the aquarium trade invading and destroying their natural freshwater habitats. In Chapter V, I collaborated with Rupert Collins and Karen Armstrong from Lincoln University, New Zealand to investigate the effectiveness of COI to identify the 172 species of cyprinids collected from the ornamental trade. I studied the effectiveness of DNA barcodes for a particular

taxon in chapter V as opposed to the effectiveness at a regional scale (chapter III) and global scale (chapter IV).

Overall, my thesis investigated the efficiency of COI for identifying fish species at local, global and systematic level. The identification tools (COI database and image database) created are designed to lay the foundation for monitoring and regulating the movement of invasive fish in the trade. Further increasing the species coverage of ornamental fish in these databases will be important because many ornamental fish species still lack DNA barcodes. Once a more complete database is available, it will become possible to monitor the fish fauna via environmental DNA extracted from water. Currently, these techniques are mostly used for monitoring the diversity of unicellular species in water and soil (Johnson, 1992; Vilchez-Vargas et al., 2013). Recent research has shown that these techniques can be extended to multicellular species (Blanchet, 2012; Bronnenhuber & Wilson, 2013; Jerde et al., 2013). While several technical problems remain to be resolved (e.g., increasing the detection sensitivity for animal DNA), many authors are convinced that environmental DNA (eDNA) will become an important source of biological knowledge (Casey et al., 2012; Jerde et al., 2012). Thus, we must continue to build reference databases because they are required for species identification via eDNA.

1.4. References for Chapter I

Aliabadian, M., Kaboli, M., Nijman, V., & Vences, M. (2009). Molecular Identification of Birds: Performance of Distance-Based DNA Barcoding in Three Genes to Delimit Parapatric Species. *Plos One, 4*(1). doi: e411910.1371/journal.pone.0004119

Antunes, A., & Ramos, M. J. (2005). Discovery of a large number of previously unrecognized mitochondrial pseudogenes in fish genomes. *Genomics, 86*(6), 708-717. doi: 10.1016/j.ygeno.2005.08.002

Axelrod, G. S., Burgess, W. E., Pronek, N., Axelrod, H. R., & Walls, J. G. (2007). Dr. Axelrod's Atlas of Freshwater Aquarium Fishes, 11th Edition. *T.F.H. Publications, Inc.*, 1160.

Baker, N., & Lim, K. K. P. (2008). Wild Animals of Singapore. *Draco Publishing and Distribution Pte. Ltd. and Nature Society (Singapore).*

Ball, S. L., & Armstrong, K. F. (2008). Rapid, one-step DNA extraction for insect pest identification by using DNA barcodes. *Journal of Economic Entomology, 101*(2), 523-532.

Blanchet, S. (2012). The use of molecular tools in invasion biology: an emphasis on freshwater ecosystems. *Fisheries Management and Ecology, 19*(2), 120-132. doi: 10.1111/j.1365-2400.2011.00832.x

Blaxter, M. (2006). Will DNA barcoding advance efforts to conserve biodiversity more efficiently than traditional taxonomic methods? Reply. *Frontiers in Ecology and the Environment, 4*(5), 272-272.

Bleeker, W., Klausmeyer, S., Peintinger, M., & Dienst, M. (2008). DNA sequences identify invasive alien Cardamine at Lake Constance. *Biological Conservation, 141*(3), 692-698. doi: 10.1016/j.biocon.2007.12.015

Bronnenhuber, J. E., & Wilson, C. C. (2013). Combining species-specific COI primers with environmental DNA analysis for targeted detection of rare freshwater species. *Conservation Genetics Resources, 5*(4), 971-975. doi: 10.1007/s12686-013-9946-0

Buhay, J. E. (2009). "COI-like" sequences are becoming problematic in molecular systematic and barcoding studies. *Journal of Crustacean Biology, 29*(1), 96-110. doi: 10.1651/08-3020.1

Casey, J. W., Bowser, P. R., Getchell, R. G., & Bain, M. B. (2012). Sensitivity of detecting environmental DNA. *Conservation Letters, 5*(3), 240-240. doi: 10.1111/j.1755-263X.2012.00232.x

Chen, Y. J., Huang, K. M., Jen, H. C., Su, M. W., Liu, S. M., & Hwang, D. F. (2009). Bighead Carp Meat Adulterated in Commercial Milkfish Ball Products. *Journal of Food and Drug Analysis, 17*(3), 178-182.

Chown, S. L., Sinclair, B. J., & van Vuuren, B. J. (2008). DNA barcoding and the documentation of alien species establishment on sub-Antarctic Marion Island. *Polar Biology, 31*(5), 651-655. doi: 10.1007/s00300-007-0402-z

Dove, C. J., Rotzel, N. C., Heacker, M. & Weigt, L. A. (2008). Using DNA barcodes to identify bird species involved in birdstrikes. *Journal of Wildlife Management, 72*(5), 1231-1236. doi: 10.2193/2007-272

Elias, M., Hill, R. I., Willmott, K. R., Dasmahapatra, K. K., Brower, A. V. Z., Malllet, J. & Jiggins, C. D. (2007). Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B-Biological Sciences, 274*(1627), 2881-2889. doi: 10.1098/rspb.2007.1035

Froese, R., Paul, D. & Editors. (2013). Fishbase. *World Wide Web electronic publication. www.fishbase.org, version (12/1013).*

Hausmann, A., Haszprunar, G., Segerer, A. H., Speidel, W., Behounek, G. & Hebert, P. D. N. (2011). Now DNA-barcoded: the butterflies and larger moths of Germany. *Spixiana, 34*(1), 47-58.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & de Waard, J. R. (2003a). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences, 270*(1512), 313-321. doi: 10.1098/rspb.2002.2218

Hebert, P. D. N., Ratnasingham, S. & de Waard, J. R. (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London Series B-Biological Sciences, 270*, S96-S99. doi: 10.1098/rsbl.2003.0025

Holmes, B. H., Steinke, D. & Ward, R. D. (2009). Identification of shark and ray fins using DNA barcoding. *Fisheries Research, 95*(2-3), 280-288. doi: 10.1016/j.fishres.2008.09.036

Hubert, N., Hanner, R., Holm, E., Mandrak, N. E., Taylor, E., Burridge, M. & Bernatchez, L. (2008). Identifying Canadian Freshwater

Fishes through DNA Barcodes. *Plos One, 3*(6). doi: 10.1371/journal.pone.0002490

Ivanova, N. V., Zemlak, T. S., Hanner, R. H. & Hebert, P. D. N. (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes, 7*(4), 544-548. doi: 10.1111/j.1471-8286.2007.01748.x

Janzen, D. H., Hallwachs, W., Blandin, P., Burns, J. M., Cadiou, J. M., Chacon, I. & Wilson, J. J. (2009). Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources, 9*, 1-26. doi: 10.1111/j.1755-0998.2009.02628.x

Jerde, C. L., Chadderton, W. L., Mahon, A. R., Renshaw, M. A., Corush, J., Budny, M. L. & Lodge, D. M. (2013). Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. *Canadian Journal of Fisheries and Aquatic Sciences, 70*(4), 522-526. doi: 10.1139/cjfas-2012-0478

Jerde, C. L., Miller, D. J., Mahon, A. R., Chadderton, W. L. & Lodge, D. M. (2012). Response to Casey et al.'s sensitivity of detecting environmental DNA comment. *Conservation Letters, 5*(3), 241-242. doi: 10.1111/j.1755-263X.2012.00231.x

Johnson, A., Rindal, E., Ericson, P. G. P., Zuccon, D., Kerr, K. C. R., Stoeckle, M. Y. & Lifjeld, J. T. (2010). DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. *Journal of Ornithology, 151*(3), 565-578. doi: 10.1007/s10336-009-0490-3

Johnson, B. T. (1992). Potential genotoxicity of sediments from the Great Lakes. *Environmental Toxicology and Water Quality, 7*(4), 373-390. doi: 10.1002/tox.2530070407

Kaila, L. & Stahls, G. (2006). DNA barcodes: Evaluating the potential of COI to diffentiate closely related species of Elachista (Lepidoptera : Gelechioidea : Elachistidae) from Australia. *Zootaxa*(1170), 1-26.

Kwong, S., Srivathsan, A. & Meier, R. (2012). An update on DNA barcoding: low species coverage and numerous unidentified sequences. *Cladistics, 28*(6), 639-644. doi: 10.1111/j.1096-0031.2012.00408.x

Little, D. P. (2011). DNA Barcode Sequence Identification Incorporating Taxonomic Hierarchy and within Taxon Variability. *Plos One, 6*(8). doi: 10.1371/journal.pone.0020552

Livengood, E. J. & Chapman, F. A. (2009). The ornamental fish trade: an introduction with perspectives for responsible aquarium fish ownership. *University of Florida IFAS Extension, FA124*, 1-8.

Logan, C. A., Alter, S. E., Haupt, A. J., Tomalty, K. & Palumbi, S. R. (2008). An impediment to consumer choice: Overfished species are sold as Pacific red snapper. *Biological Conservation, 141*(6), 1591-1599. doi: 10.1016/j.biocon.2008.04.007

Lohman, D. J. & Samarita, V. U. (2009). The biology of carnivorous butterfly larvae (Lepidoptera: Lycaenidae: Miletinae: Miletini) and their ant-tended hemipteran prey in Thailand and the Philippines. *Journal of Natural History, 43*(9-10), 569-581. doi: 10.1080/00222930802610485

Lorenz, J. G., Jackson, W. E., Beck, J. C. & Hanner, R. (2005). The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philosophical Transactions of the Royal Society B-Biological Sciences, 360*(1462), 1869-1877. doi: 10.1098/rstb.2005.1718

Martin, W., Fiedler, K. (2007). Does the DNA barcoding gap exist? A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology, 4*(8), 1 - 8.

Meier, R., Kwong, S., Vaidya, G. & Ng, P. K. L. (2004). An empirical test of DNA taxonomy in Diptera based on cox-1 sequences. *Cladistics, 20*(6), 600-600.

Meier, R., Shiyang, K., Vaidya, G. & Ng, P. K. L. (2006). DNA barcoding and taxonomy in diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology, 55*(5), 715-728. doi: 10.1080/10635150600969864

Meyer, C. P. & Paulay, G. (2005). DNA barcoding: Error rates based on comprehensive sampling. *Plos Biology, 3*(12), 2229-2238. doi: e42210.1371/journal.pbio.0030422

Moynihan, M. (1971). Successes and failures of tropical mammals and birds. *American Naturalist, 105*(944), 371-&. doi: 10.1086/282731

Nelson, J. S. (2006). Fishes of the world, 4th Edition. *John Wiley & Sons, Inc.*, 601.

Ng, H. H., & Tan, H. H. (2010). An annotated checklist of the non-native Freshwater fish species in the reservoirs of Singapore. *COSMOS, 6*(1), 95-116.

Paine, R. T. (1966). Food web complexity and species diversity. *Amer. Natur., 100*, 65 - 75.

Roe, A. D. & Sperling, F. A. H. (2007). Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution, 44*(1), 325-345. doi: 10.1016/j.ympev.2006.12.005

Rudnick, J. A., Katzner, T. E., Bragin, E. A. & DeWoody, J. A. (2007). Species identification of birds through genetic analysis of naturally shed feathers. *Molecular Ecology Notes, 7*(5), 757-762. doi: 10.1111/j.1471-8286.2007.01796.x

Song, H., Buhay, J. E., Whiting, M. F. & Crandall, K. A. (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America, 105*(36), 13486-13491. doi: 10.1073/pnas.0803076105

Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. & Vogler, A. P. (2003). A plea for DNA taxonomy. *Trends in Ecology & Evolution, 18*(2), 70-74.

Trewick, S. A. (2008). DNA Barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera : Acrididae). *Cladistics, 24*(2), 240-254. doi: 10.1111/j.1096-0031.2007.00174.x

Valdez-Moreno, M., Vasquez-Yeomans, L., Elias-Gutierrez, M., Ivanova, N. V. & Hebert, P. D. N. (2010). Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. *Marine and Freshwater Research, 61*(6), 665-671. doi: 10.1071/mf09222

Venkatesh, B., Dandona, N. & Brenner, S. (2006). Fugu genome does not contain mitochondrial pseudogenes. *Genomics, 87*(2), 307-310. doi: 10.1016/j.ygeno.2005.11.007

Victor, B. C., Hanner, R., Shivji, M., Hyde, J. & Caldow, C. (2009). Identification of the larval and juvenile stages of the Cubera Snapper, Lutjanus cyanopterus, using DNA barcoding. *Zootaxa*(2215), 24-36.

Victor, B. C., Vasquez-Yeomans, L., Valdez-Moreno, M., Wilk, L., Jones, D. L., Lara, M. R. & Shivji, M. (2010). The larval, juvenile, and adult stages of the Caribbean goby, Coryphopterus kuna (Teleostei: Gobiidae): a reef fish with a pelagic larval duration longer than the post-settlement lifespan. *Zootaxa*(2346), 53-61.

Vilchez-Vargas, R., Geffers, R., Suarez-Diez, M., Conte, I., Waliczek, A., Kaser, V. S. & Pieper, D. H. (2013). Analysis of the microbial gene landscape and transcriptome for aromatic pollutants and alkane degradation using a novel internally calibrated microarray system. *Environmental Microbiology, 15*(4), 1016-1039. doi: 10.1111/j.1462-2920.2012.02752.x

Ward, R. D., Hanner, R. & Hebert, P. D. N. (2009). The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology, 74*(2), 329-356. doi: 10.1111/j.1095-8649.2008.02080.x

Whitworth, T. L., Dawson, R. D., Magalon, H. & Baudry, E. (2007). DNA barcoding cannot reliably identify species of the blowfly genus Protocalliphora (Diptera : Calliphoridae). *Proceedings of the Royal Society B-Biological Sciences, 274*(1619), 1731-1739. doi: 10.1098/rspb.2007.0062

Wiemers, M. & Fiedler, K. (2007). Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae). Frontiers in Zoology, 4(8), 16.

Wong, E. H. K. & Hanner, R. H. (2008). DNA barcoding detects market substitution in North American seafood. *Food Research International, 41*(8), 828-837. doi: 10.1016/j.foodres.2008.07.005

Yancy, H. F., Zemlak, T. S., Mason, J. A., Washington, J. D., Tenge, B. J., Nguyen, N. L. T. & Hebert, P. D. N. (2008). Potential use of DNA barcodes in regulatory science: Applications of the Regulatory Fish Encyclopedia. *Journal of Food Protection, 71*(1), 210-217.

Yi, Y. G., Tan, H. H. & Meier, R. (2012). COI is effective for the identification of fishes in Singapore's freshwater system. *pending publication.*

Yoo, H. S., Eah, J. Y., Kim, J. S., Kim, Y. J., Min, M. S., Paek, W. K. & Kim, C. B. (2006). DNA barcoding Korean birds. *Molecules and Cells, 22*(3), 323-327.

CHAPTER II

_____

Tracking a moving target:

ornamental fish in the pet trade

Abstract

One significant source of invasive species is ornamental plant and animal species that are sold to amateurs through the pet trade. The same trade also constitutes a significant problem for conservation biology because it is not uncommon that it includes endangered species that are taken from the wild. Government agencies have responded by either maintaining lists of approved or disallowed species, but it is unclear how accurate and complete these lists are. I tested for completeness by first comparing the consistency between two existing, published lists for ornamental fish and then comparing both to the list of species that were traded in Singapore from 2009 to 2012. Both published lists combined comprise 4,769 species of freshwater fish, of which 2,705 are only found on one list. However, both lists are still incomplete, because the 895 species that were traded in Singapore include 97 species that are on neither list. These 895 species traded in Singapore between 2009 and 2013 belong to 377 genera in 90 families. The majority were tropical species (95%) while subtropical (4%) and temperate (1%) species were rare. At least 62 of the traded species are now also found in Singapore's freshwater bodies and 44 (70%) of them are introduced species. This proportion of introduced species is likely an underestimate because non-Singaporean populations of many species indigenous to Singapore are in the pet trade and have likely been released. I find that 71% of all species in Singapore's trade were wild-caught with 79% of them being

20

from Asia (predominately Southeast Asia). For the latter the proportion of wild-caught species was even larger (86%). Of the species in the trade, 72-77% were correctly identified while the remaining ones suffer from incorrect or imprecise identification.

2.1. Introduction

The international trade of ornamental plants (Darbyshire & Francis, 2008; Hussner, 2012) and animals has been widely acknowledged to be a major source of introductions of invasive species into aquatic environments (Avila et al., 2012; Ayala, et al., 2007; Chang et al., 2009; Copp et al., 2005; de Magalhaes & de Carvalho, 2007; Gerstner et al., 2006; Maceda-Veiga et al., 2013; Magalhaes & Vitule, 2013). Indeed, the ornamental trade was responsible for the largest proportion of intentionally (73%) and unintentionally (34%) introduced species in Great Britain (Keller et al., 2009). Biological invasion through the process of releasing non-native fish into new habitats can have many negative effects. Firstly, it often results in the direct loss of native freshwater biodiversity. This can be particularly devastating when a new predator arrives. The best known example was the release of the predacious Nile perch (*Lates niloticus*) into Lake Victoria (East Africa) in the 1950s (Kolar & Lodge, 2001) which has been blamed for the mass extinction of over 200 endemic species. Other examples include the Oriental weatherfish *Misgurnus anguillicaudatus* which has become invasive in many temperate areas (Franch et al., 2008).

Secondly, many ornamental fish can not only establish viable, non-native populations in new habitats, but they can even modify the water chemistry. For example, recent studies have shown that the presence of an introduced catfish *Clarias gariepinus* with phosphate-rich body stoichiometry affects the nutrient dynamic of an entire aquatic

ecosystem (Capps & Flecker, 2009). Thirdly, invasive species can do damage to aquaculture by introducing new pathogens. Examples include the spread of Goldfish ulcer disease to salmon and trout farms and the accidental introduction of *Gourami iridovirus* to Murray cod [DAFF website (2010): http://www.daff.gov.au/]. Lastly, direct harm to humans can come from the introduction of dangerous species. This includes piranhas and freshwater stingrays (*Potamotrygon motoro*) (Ng & Tan, 2010).

The releases of ornamental fish and accidental escapees from aquaculture are the main source of non-native fish in water systems including Germany and Austria (Wolter, 2010). Fortunately, in many temperate countries only a small proportion of released, ornamental fish are likely to survive in their new environment because most species in the trade are adapted to subtropical or tropical climates. Survival of these species is more likely in tropical climates (Moynihan, 1971, Paine, 1966) as is evident from Singapore's freshwater fish fauna. Singapore has breeding populations for 108 species of fish of which 75 are aliens (Baker, 2008, Chapter 3, present volume) and the number keeps rising. Recent additions are *Acarichthys heckelii* (Tan, 2008), *Potamotrygon motoro* (Ng, 2010), and *Scleropages formosus* (Meier pers comm. 2009).

Additionally, the ornamental fish trade is not only a significant problem for the receiving nation. The same trade often also damages

the biodiversity in the country of origin because it is not uncommon that the trade includes endangered species that were taken from the wild.

Given these numerous problems caused by ornamental fish, it is not surprising that governments use regulatory and legal mechanisms as counteraction measures. However, all measures ultimately rely on accurate species-level data that are critical for preventing invasions and mitigating their consequences (Simberloff et al., 2013). Species-level data are thus important because government agencies maintain either positive lists of approved species or a mixture of positive and negative lists. The latter usually list particularly invasive species and endangered species that are on red-lists

and/or CITES. For example, Australia

(http://www.dpi.nsw.gov.au/fisheries/pests-diseases/noxious-fish-and-marine-vegetation), New Zealand (FISORNIC.ALL, 2011), United Kingdom (http://www.cefas.defra.gov.uk/), the European Union and some states in the United States maintain lists of approved organisms as well as lists of invasive species that are illegal to import for ornamental purposes. Accuracy and completeness of these lists are an important precondition for the success of these control measures.

Currently, there are two available lists of ornamental fish that are recognized by the trade. One was drafted by Axelrod in 2006 (Axelrod, 2007) and the other by Ornamental Fish International in 2010 (OFI) (Hensen, 2010). In this study, I first investigate the consistency

between these two lists. I then compared the combined lists with the list of species in Singapore's ornamental trade (2009-2012). Ideally, one would find that the two lists are consistent and largely overlapping and that the list for Singapore's trade is a subset of the other two lists. This would indicate that governments could use published species lists for selecting permitted and/or prohibited species.

2.2. Materials and Methods

2.2.1. Obtaining the international list of ornamental fish

The species list of Axelrod (2006) and Hensen (2010) were scanned and converted to word format using OCR (Adobe Acrobat 2010) before copying the species names into a worksheet database. A total of 2,705 and 4,769 species were recorded for the 2006 Axelrod and 2010 OFI lists respectively. Names of varieties were removed because I were only interested in species-level information. The combined list initially included 5,968 names. However, some names were synonyms and other names constituted new combinations. In order to obtain a list of unique species, the genus and species were separated into different columns and the list was sorted by species epithet. Identical and/or near identical species epithets were checked for new combinations (many in *Nandopsis*, *Vieja*, *Cichlasoma*). I also removed duplicate

25

names that only differed by genus gender. These standardizations were applied to all lists in my study. In addition, synonymy transcending genus boundaries was identified manually with the help of taxonomists or by searching for genus names with known, recent changes. Whenever encountered, the most recent name accepted by the Catalog of fishes (2014)

hosted by California Academy of Sciences (CAS;

http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp) was used.

2.2.2. Obtaining the list for the Singapore trade

In order to establish a list of ornamental fish species in Singapore's trade, 35 ornamental fish retail stores were surveyed over a period of 2 years (2007 and 2008) by visiting them once every two weeks (Lee, 2007; Lee, 2008). I also visited two major exporters of freshwater ornamental fish at the same interval for the duration of three and a half years (Feb 2009 to Jun 2012) and recorded the species in the trade.

Accurate identification and allocation of correct and standardized names were assessed for the survey conducted between February 2009 and June 2012. The purchased specimens were carefully identified using taxonomic keys, species descriptions and Fishbase (Roberts, 1989; Kottelat, 1990; Talwar, 1991; Kottelat, 1993;

Rainboth, 1996; Kottelat, 2001; Inger, 2002; Norris, 2002; Nelson, 2006; Tan, 2006; Axelrod, 2007; Hensen, 2010). Taxonomists Dr Tan Heok Hui (for cyprinids and silurids identities) and Dr Ng Heok Hee (for silurids and channids identities) from the Raffles Museum of Biodiversity Research (RMBR) were consulted when in doubt. Nomenclatures follows Fishbase (Froese, 2013) and the Catalog of Fishes web database maintained by the California Academy of Sciences (William, 2013).

Cases of mislabelling and misidentification by fish farms were recorded in order to investigate the reliability of fish farm identifications. This part of the study was restricted to the Qian Hu Fish Farm, a major importer in Singapore, whose fish tanks were properly labelled with species names. The other importer that I studied did not label its tanks regularly enough for us to carry out this part of the study. Similarly, the retail trade could not be assessed because it rarely uses scientific names.

In addition to species names, additional information was recorded such as whether fish were captive-bred or wild-caught and the supplier's country of origin. Obtaining this information for species in the retail trade proved difficult and in some cases only regional information ('Asia', 'South America', etc.) or climatic data ('tropical', 'temperate', 'sub-tropical') was available via a secondary source (Fishbase; see Figure 2.3.2.1)(Froese, 2013). In order to distinguish between popular

and rarely traded groups, I ranked the families according to the number of species traded in each family (Table 2.3.1.I). The full list of species and families is included as supplementary information (Appendix I: Species List). I also established the relationship between the traded fish, region of origin, and source (wild-caught or captive-bred: Figure 2.3.2.3).

2.3. Results

2.3.1. Comparison between the existing species lists

After comparing the two lists, it became clear that Axelrod (2006) is a subset of OFI (2010). Comparison between these lists reveals that 2,705 new species names have been added between 2006 and 2010. The combined lists contain 4,769 species records while the trade list for Singapore contain only 895 species. However, 97 of these are new additions to the list of ornamental fish in the trade; i.e., only 798 species are already found in Axelrod's list (Axelrod, 2007) and OFI list (Hensen, 2010).
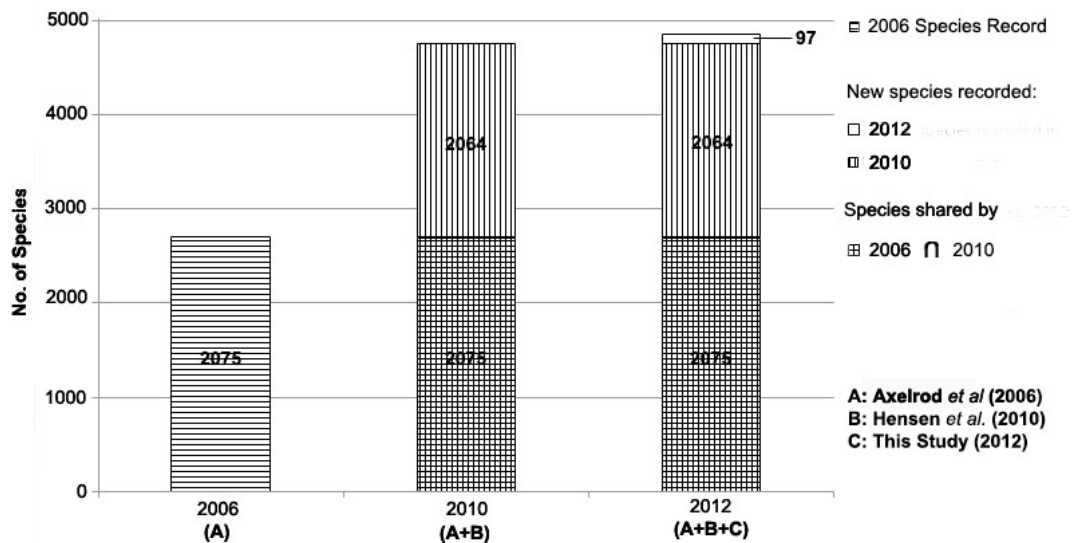
Figure 2.3.1.1: Freshwater fish recorded in the global ornmental trade
from 2006-2012
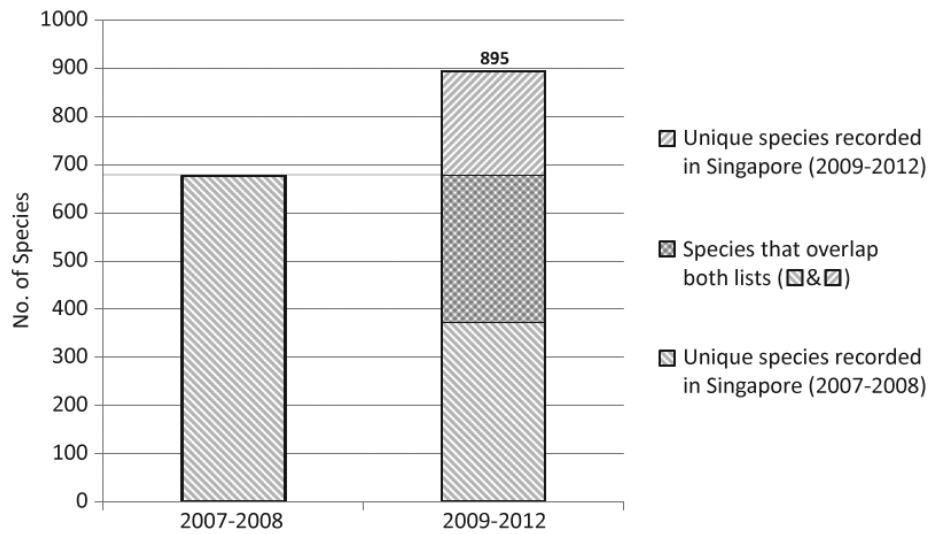
Figure 2.3.1.2: Freshwater fish recorded in the Singapore ornamental trade from 2007 to 2012

2.3.2. Statistics for the Singapore aquarium trade

2.3.2.1. Species distribution according to region of origin

The Singapore trade list in this study comes from two sources. Some records were collected from 2007 to 2008 by Lester (2007-2008) in his UROPs project during his undergraduate course. This study yielded 678 species that were recorded as part of a trade surveilence. The second survey was conducted from 2009 to 2013 for my PhD course and involved specimen collection and DNA barcoding. It contributed 217 new species while 310 species were already on the previous lists.

A total of 895 species of freshwater fishes from 377 genera in 90 families were recorded. The majority of these are tropical freshwater species (95%) originating from Asia (mainly Southeast Asia and tropical region of China and India), South America and Africa. Subtropical (4%) and temperate (1%) species contribute less than 5% of the species in the trade (Figure 2.3.2.1).
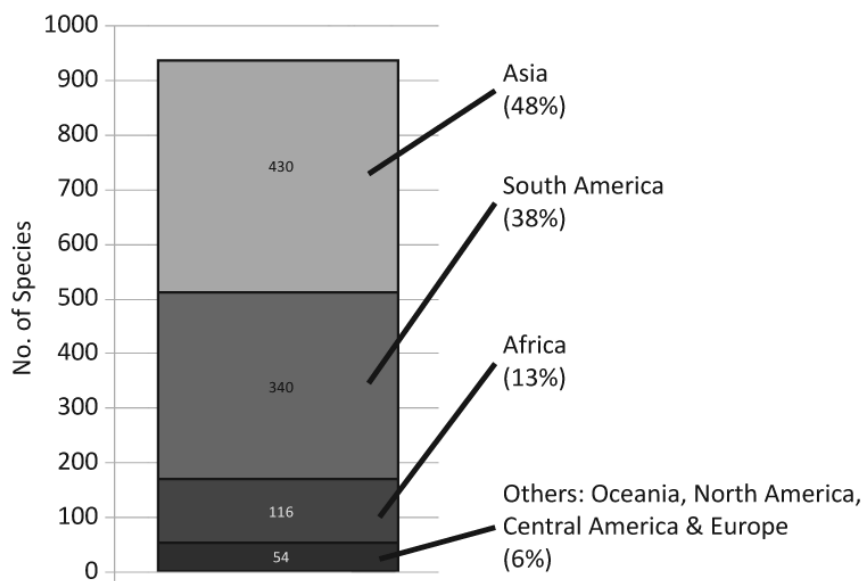
Figure 2.3.2.1: Regional distribution of ornamental fish in the Singapore trade

2.3.2.2. Species distribution according to family

A total of 90 families are on the Singapore trade list, 80% (731) of the species in the trade are found in the first 21 families (rank 1 to 20) as listed in fig. 2.: Cyprinidae (Carps), Cichidae, Loricariidae (Sucker Catfish), Osphronemidae (Gouramis & Betta), Characidae

(Tetra) and Callichthyidae (Armoured Catfish) contained 55% (492) of

the recorded species in the trade.

Table 2.3.1.I. Species distribution within families for ornamental fish
recorded from the Singapore trade

Rank	Family/Families	Species per family
1	Cyprinidae	143
2	Cichlidae	111
3	Characidae	63
4	Callichthyidae	62
5	Osphronemidae	62
6	Loricariidae	51
7	Balitoridae	35
8	Cobitidae	32
9	Channidae	20
10	Tetraodontidae	20
11	Bagridae	16
12	Melanotaeniidae	16
13	Gobiidae	15
14	Pimelodidae	14
15	Poeciliidae	12
16	Polypteridae	11
17	Mastacembelidae	10
18	Mochokidae	10
19	Siluridae	10
20	Potamotrygonidae, Nothobranchiidae (2)	9
21	Lebiasinidae	8
22	Hemirhamphidae	7
23	Osteoglossidae, Mormyridae (2)	6
24	Gasteropelecidae, Doradidae, Datnioididae, Anostomidae, Ambassidae, Alestiidae (6)	5
Continue on next page		

32

Rank	Family/Families	Species per family
25	Gymnotidae, Erethistidae, Eleotridae, Auchenipteridae	4
26	Toxotidae, Sissoridae, Pseudomugilidae, Pangasiidae, Notopteridae, Lepisosteidae, Clariidae, Catostomidae, Badidae, Aplocheilidae, Anabantidae, Akysidae (12)	3
27	Schilbeidae, Rivulidae, Nandidae, Malapteruridae, Gyrinocheilidae, Erythrinidae, Cynodontidae, Citharinidae, Centrarchidae, Arapaimidae, Apteronotidae, Adrianichthyidae (12)	2
28	Telmaterihnidae, Syngnathidae, Sternopygidae, Soleidae, Pseudopimelodidae, Protopteridae, Procheilodontidae, Phallostethidae, Pantodontidae, Lepidosirenidae, Latidae, Indostomidae, Heteropneustidae, Hepsitidae, Hemiodontidae, Helostomatidae, Gymnarchidae, Esocidae, Electrophoridae, Dasyatidae, Cyprinodontidae, Ctenoluciidae, Crenuchidae, Claroteidae, Cheilodontidae, Chacidae, Cetopsidae, Bedotiidae, Asphredinidae, Amblycipitidae, Acestrorhynchiidae (31)	1

2.3.2.3. Relationship between wild-caught and captive-bred species and supplier countries

Information on source and supplier (i.e., wild-caught vs. captive-bred and supplier information at the country level) are available for 464 of the 895 species. The majority of the species (357; 76%) are wild-caught while 182 (39%) species are captive-bred. Twenty-five of these are included in both categories. The majority of the freshwater fish originating from Asia are wild-caught and only 24% are captive-bred (Figure 2.3.2.3).
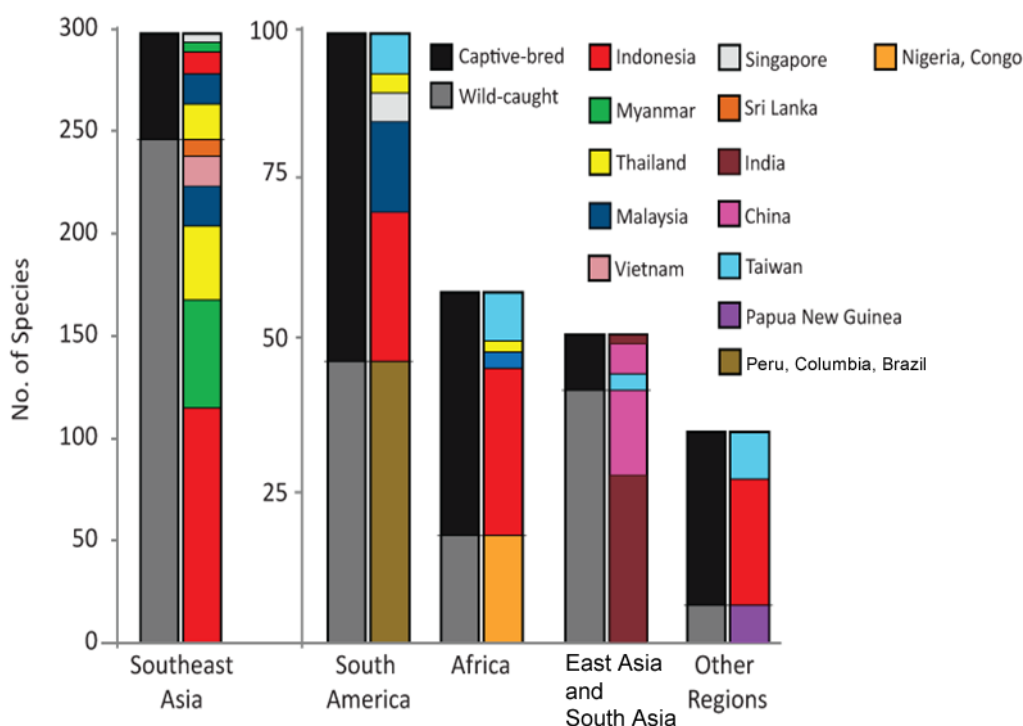
Figure 2.3.2.3: Distribution of wild-caught and captive-captive-bred species according to supplier countries

The majority of species (245 spp.) that are native to Southeast Asia are wild-caught and bought from suppliers in Indonesia (115 spp.),

Myanmar (53 species) and Thailand (36 spp.). In fact, the proportion of Southeast Asian wild-caught species is 53% and constitutes two-thirds (68%) of the total wild-caught species. Indonesia alone supplies 47% of the wild-caught species in Southeast Asia and about one-third (32%) of the wild-caught species in the world. The remaining 28% of Southeast Asian species (53) are captive-bred in Thailand (18), Malaysia (15) and Indonesia (6). All Southeast Asian species are supplied by suppliers within the region (Figure 2.3.2.3).

The ratio of wild-caught to captive-bred South American species in the Singapore trade is about 1:1; 46 species to 54 species respectively. While all wild-caught South American species are from Peru, Brazil and Colombia, the majority of captive-bred species are from Southeast Asian suppliers with Indonesia and Malaysia supplying 46% and 28% of the total South American captive-bred species respectively. As recorded, Indonesia also serves as the largest supplier of captive-bred African and Oceanian species in the Singapore trade. Overall, Indonesia supplies 50% (91 spp.) of all recorded captive-bred species (Figure 2.3.2.3).

Sixty-eight percent of the African species in the Singapore trade are captive-bred and supplied by Southeast Asian fish farms, 69% (27 spp.) of these are from Indonesian breeders, the remaining 31% (12 spp.) of African species are from Taiwan, Malaysia and Thailand in descending order based on the number of species recorded for each

country. Wild-caught African species are mainly from Nigeria and Congo, but the source information is vague for most species the country of origin is "Africa" (Figure 2.3.2.3).

Eighty percent of South Asian species are wild-caught from India (28) and China (12), while the remaining 9 species are captive-bred in Taiwan, China and India. Wild-caught Asian species constitute 86% of the Asian fish in the Singapore trade (Figure 2.3.2.3).

I do not indicate how many fish are traded for each species because this information is not available to us. It is considered confidential information by the fish farms. Instead, my results pertain to species numbers and only apply to the survey conducted from 2009 to 2013 because information on wild-caught, captive-bred, and supplier country were not recorded in the previous survey. Only 80%, 29%, 49% and 64% of all recorded Asian, South American, African and "other regions" species have their 'wild-caught or captive-bred' information being recorded.

2.3.3. Misidentification and mislabeling in the Singapore trade

I assess identification accuracy based on 358 species that were collected from Qian Hu fish farm for 722 shipments. At the species level, 259 species were always correctly identifed (72%), 47 species were misidentified (13%), and the remaining 15% of species were ambiguous (tank not labeled, identification only to genus, species occasionally misidentified). Identification efficiency increases to 78% (562 shipments) when considering the number of shipments, while misidentification increases by 1% (10 cases) to 14% (101 cases). The misidentification rate is ca. 28% (species level) and 23% (tank level; for every 100 shipments, 23 are misidentified) if counting ambiguity as misidentification.
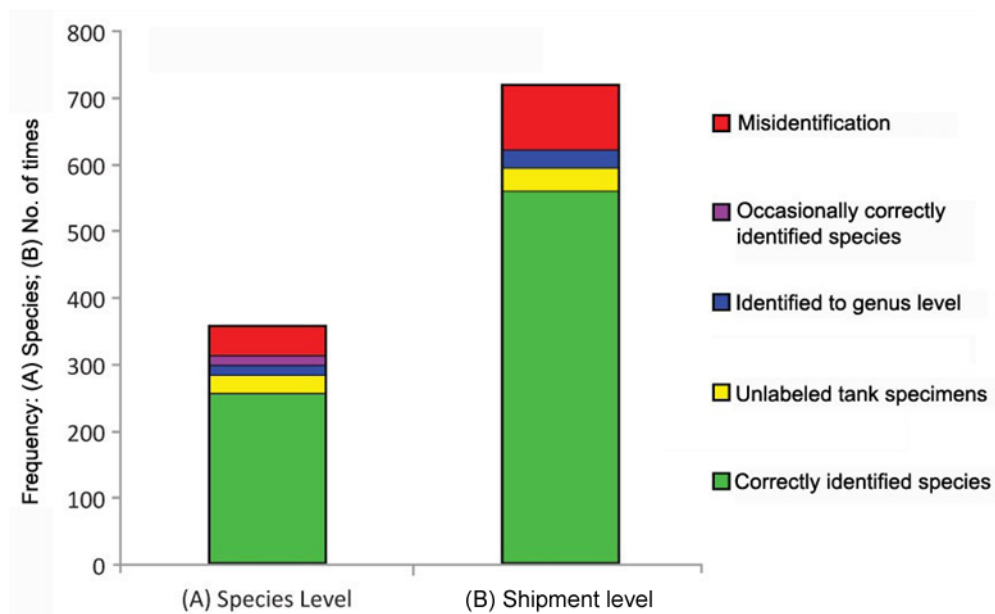
Figure 2.3.3.1: The status of species identification in Qian Hu fish farm

2.4. Discussion

The ornamental trade of plants and animals is well known to be a major route for the introduction of invasive species, and subsequent loss of biodiversity (Avila and Troca, 2012; Ayala, 2007; Chang, 2009; Copp, 2005; Gerstner, 2006; Magalhaes, 2013; de Magalhaes, 2007; Maceda-Veiga, 2013). I here test whether governments can trust existing species lists for creating positive lists of allowed and negative lists of disallowed species. In addition, I compare the lists to what is traded in Singapore.

I find that that the ornamental fish trade is apparently evolving very quickly. The first comprehensive list was published in 2006 (Axelrod, 2007). Yet, four years later, 2,705 "new" species were added (Hensen, 2010). One might expect that the list of species in Singapore's trade would be a subset of both the combined international lists, but this is not the case. In fact, 97 "new" species were identified, suggesting that the international lists are incomplete or the trade has a high turnover of species; i.e., new species are regularly added or replacing "old" species. I favor the latter explanation, because a comparison between the two lists for Singapore (2007-2008 and 2009-2013) indicates a high turnover with 217 "new" species turning up in Singapore's trade.

I therefore believe that it will be very difficult for governments to maintain up-to-date lists for species in the ornamental fish trade. This

has repercussions for which bioprotection strategies can be applied. Negative lists of species that are now allowed to be traded appear unrealistic because they do not allow for the fast turnover that I observe in the trade. The numerous new species that appear in the trade every year would not be captured by negative lists and the latter would have to be updated on a monthly basis in order to avoid undesirable species. In theory, positive lists appear a better strategy, but they will meet with stiff resistance fromt trade and aquarists. The fast turnover in species is clearly indicating that there is strong demand for novelty and such demand is incompatible with positive lists.

A particuarly worrisome aspect of the trade is that many species currently sold in Singapore are wild-caught and from Southeast Asia (approximately 53% of wild-caught species recorded), a region that is rich in biodiversity and where new species continue to be discovered (Giam & Ng et al. 2010). New wild-caught species introduced into the trade contributes disprotionally to the species turnover. I find that 90% (88 of 97) of the "new" species recorded in the Singapore trade are wild-caught Southeast Asian species. The high frequency and the poor regulation of the trade in these undocumented species across international borders will increase the chances of species invasion, and horizontal bio-invasion across the region.

In fact, a recent survey of Singapore freshwater habitats revealed that as many as 70% of the freshwater fish species are non-native.

Most are tropical with most coming from tropical Southeast Asia and South America (Baker, 2008; Chapter 3, present volume). Arguably the risk for species introductions is higher for tropical environments because a large proportion of the species in the trade and fish farms are tropical. The introduction of temperate species is somewhat more difficult because few temperate species are in the trade and introdcutions often require a transfer from the Northern to the Southern hemisphere (Moynihan, 1971; Paine, 1966).

Besides causing bioprotection problems, the ornamental fish trade will also have negative effects on native populations given that a large proportion of the trade is in wild-caught fish. This will affect the conservation of freshwater fish in Southeast Asia which has a larger species diversity than most other parts of the world. Note that Southeast Asia contributes about 68% of the wild-caught species in the Singapore ornamental fish trade. This is only partially explained by geography because Singapore is one of the largest clearing hubs for the trade and part of an extensive, global network. Most of the fish traded through Singapore come from Indonesia, Thailand and Myanmar with Indonesia being by far the most important source. Overall, the data suggest that developing nations are the source of the majority of the wild-caught fish species in the trade which is similar to what has been found for other wildlife trades within the region (Nijman, 2010) and globally (Van der Knaap, 2013). Southeast Asia is a sink of biodiversity resources including freshwater fish which is caused by

global demand for wildlife and low labour cost in developing nation such as Indonesia, Thailand and Myanmar.

These factors make it more likely that fish farms established within these countries, while poverty and unemployment within the region makes it more likely that fish are taken from the wild. Although I do not have data on the number of specimens traded for each species, I noticed that many wild-caught species are repeatedly found in the quarantine facilities of fish farms; i.e., there is a sustained trade of wild-caught specimens for many species. The presence of wild-caught and captive-bred specimens of the same species side-by-side in the trade also suggests that taking fish from the wild is economically viable even for those species that can be captive-bred by fish farms.

Besides being the leading contributor of wild-caught species, Southeast Asia now also serves as an important contributor of captive-bred species that originated from around the world; i.e., many of the popular South American, African and Oceania species are captive-bred in Southeast Asia and supplied by countries such as Indonesia (50% of recorded captive-bred species). While captive breeding operations within the region will reduce the pressure on wild populations, they create new problems because accidental introductions are more likely to occur. In addition, it does not solve the problem of wild-caught species within the region. Only 50 out of the 300 trade species native to Southeast Asia are captive-bred. It appears likely that captive

breeding only starts if there is sustained demand, low-cost captive breeding is feasible, and the wild populations are shrinking. This means that many species will have to decline dramatically before the pressure on wild populations will ease. In order to avoid extinction, better trade data for these fish are needed.

Bioprotection and conservation can only be carried out efficiently when consensus is reached between the regulators, operators and buyers. Education and promotion of conservation is key for affecting buyers' choices (Chang, 2009), but it is equally important that species identification and proper labeling of ornamental fish becomes mandatory so that regulators can monitor the trade. Currently, even fish farms with comparatively sound policies can only identify 72-78% of all species correctly. This means that it would be very difficult to monitor and regulate the trade given that 28% of species and/or holding tanks have incorrect labels.

Continuous documentation of the species in the ornamental trade is important for maintaining an updated species list. Stricter regulation could yield regular updates as long as it involves taxonomic experts and includes photographic documentation of traded species online. Regulators might not be able to identify all species, but they could use such tools for obtaining expert advise. This is why I documented all species in the Singapore trade from 2009-2013 with high-quality photographs and generated DNA barcodes. By using images and/or

DNA sequences, species can be identified and the trade can be monitored.

Indeed, the next step toward regulation could be based on standardised molecular markers. The use of short sequences of DNA to identify species has been advocated by the scientific community and is now known as DNA barcoding (Hebert, 2005; Lorenz, 2005). It appears that 80-90% of all species have unique DNA barcodes so that this tool may become important for automated identifiction of species that do not require the involvement of taxonomic experts. Evidence from studies that investigate identification success in fish COI often reveal above 90% species identification success rates so that the technique may be useful for ornamental fish (Ward, 2009; Ward et al., 2005; Collins, 2012). DNA barcodes may also be particularly useful because many species have recently been split based on microscopic or behavioural evidence. These differences are inaccessible to regulators while the amplification of DNA sequences remains feasible. However, before this technique can be promoted, it is important to know more about the intraspecific and interspecific variability of DNA barcodes for ornamental fish species. These issues are addressed in the remaining chapters of my PhD thesis.

2.5. References for Chapter 2

Avila T. D. F. & Vieira, J. P. (2012). Potential invasive non-native fish farmed in the coastal region of Rio Grande do Sul, Brazil. *Boletim Do Instituto De Pesca, 38*(2), 109-120.

Axelrod, G. S., Burgess, W. E., Pronek, N., Axelrod, H. R. & Walls, J. G. (2007). Dr. Axelrod's Atlas of Freshwater Aquarium Fishes, 11th Edition. *T.F.H. Publications, Inc.*, 1160.

Ayala, J. R., Rader, R. B., Belk, M. C. & Schaalje, G. B. (2007). Ground-truthing the impact of invasive species: spatio-temporal overlap between native least chub and introduced western mosquitofish. *Biological Invasions, 9*(7), 857-869. doi: 10.1007/s10530-006-9087-4

Baker, N. & Lim, K. K. P. (2008). Wild Animals of Singapore. *Draco Publishing and Distribution Pte. Ltd. and Nature Society (Singapore).*

Capps, K. A. & Flecker, A. (2013) Invasive aquarium fish transform ecosystem nutrient dynamics. *Proc. R. Soc. B., 1471-2954.* doi : 10.1098/rspb.2013.1520

Chang, A. L., Grossman, J. D., Spezio, T. S., Weiskel, H. W., Blum, J. C., Burt, J. W. & Grosholz, E. D. (2009). Tackling aquatic invasions: risks and opportunities for the aquarium fish industry. *Biological Invasions, 11*(4), 773-785. doi: 10.1007/s10530-008-9292-4

Collins, R. A., Armstrong, K. F., Meier, R., Yi, Y. G., Brown, S. D. J., Cruickshank, R. H. & Johnston, C. (2012). Barcoding and Border Biosecurity: Identifying Cyprinid Fishes in the Aquarium Trade. *Plos One, 7*(1). doi: 10.1371/journal.pone.0028381

Copp, G. H., Wesley, K. J. & Vilizzi, L. (2005). Pathways of ornamental and aquarium fish introductions into urban ponds of Epping Forest (London, England): the human vector. *Journal of Applied Ichthyology, 21*(4), 263-274. doi: 10.1111/j.1439-0426.2005.00673.x

DAFF webiste: Biosecurity materials for ornamental fishes. Retrieved 10 March 2013, from http://www.daff.gov.au/animal-plant-health/pests-diseases-weeds/disease_watch_aquatic_ animal_health_awareness/biosecurity_awareness_materials/ other_aquatic_biosecurity_material

Darbyshire, S. J. & Francis, A. (2008). The biology of invasive alien plants in Canada. 10. Nymphoides peltata (S. G. Gmel.) Kuntze. *Canadian Journal of Plant Science, 88*(4), 811-829.

Franch, N., Clavero, M., Garrido, M., Gaya, N., Lopez, V., Pou-Rovira, Q. & Queral, J. M. (2008). On the establishment and range expansion of oriental weatherfish (Misgurnus anguillicaudatus) in NE Iberian Peninsula. *Biological Invasions, 10*(8), 1327-1331. doi: 10.1007/s10530-007-9207-9

Froese, R., Pauly, D. & Editors. (2013). Fishbase. *World Wide Web electronic publication., www.fishbase.org, version (12/1013).*

Gerstner, C. L., Ortega, H., Sanchez, H. & Graham, D. L. (2006). Effects of the freshwater aquarium trade on wild fish populations in differentially-fished areas of the Peruvian Amazon. *Journal of Fish Biology, 68*(3), 862-875. doi: 10.1111/j.1095.8649.2006.00978.x

Giam, X. L., Ng, T. H., Yap, V., & Tan, H. T. W. (2010). The extent of undiscovered species in Southeast Asia. *Biodiversity and Conservation, 19*(4), 943-954. doi: 10.1007/s10531-010-9792-2.

Hensen, R. R., Ploeg, A. & Fossa, S. A. (2010). Standard names for freshwater fishes in the Ornamental Aquatic Industry. *OFI educational publication, 5*, 145.

Hussner, A. (2012). Alien aquatic plant species in European countries. *Weed Research, 52*(4), 297-306. doi: 10.1111/j.1365-3180.2012.00926.x

Inger, R. F. & Chin, P. K. (2002). The Freshwater Fishes of North Borneo. *Natural History Publications (Borneo) Sdn. Bhd.*, 262.

Keller, R. P., Ermgassen, P. & Aldridge, D. C. (2009). Vectors and Timing of Freshwater Invasions in Great Britain. *Conservation Biology, 23*(6), 1526-1534. doi: 10.1111/j.1523-1739.2009.01249.x

Kolar, C. S. & Lodge, D. M. (2001). Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution, 16*(4), 199-204. doi: 10.1016/s0169-5347(01)02101-2

Kottelat, M. (1990). Indochinese nemacheilines, A revision of nemacheiline loaches (Pisces: Cypriniformes) of Thailand, Burma, Laos, Cambodia and southern Viet Nam. *Verlag Dr . Friedrich Pfeil, Munchen, FRG*, 262.

Kottelat, M. (2001). Fishes of Laos. *WHT Publications (Pte) Ltd.*, 198.

Kottelat, M., Whitten, A. J., Kartikasari, S. N. & Wirjoatmodjo, S. (1993). Freshwater Fishes of Western Indonesia and Sulawesi. *Periplus Editions (HK) Ltd*, 293.

Lee, L. (2007). The pet fish trade and invasive freshwater fish species in Singapore. *Undergraduate Research Opportunities in Science Project Report, Department of Biological Sciences, The National University of Singapore.*, 37.

Lee, L. (2008). The ornamental fish trade and invasive freshwater fish species in Singapore. *A thesis submitted to the Department of Biological Sciences, The National University of Singapore in partial fulfilment for the Degree of Bachelor of Science with Honours in Life Sciences*, 141.

Lorenz, J. G., Jackson, W. E., Beck, J. C. & Hanner, R. (2005). The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philosophical Transactions of the Royal Society B-Biological Sciences, 360*(1462), 1869-1877. doi: 10.1098/rstb.2005.1718

de Maceda-Veiga, A., Escribano-Alacid, J., de Sostoa, A. & Garcia-Berthou, E. (2013). The aquarium trade as a potential source of fish introductions in southwestern Europe. *Biological Invasions, 15*(12), 2707-2716. doi: 10.1007/s10530-013-0485-0

de Magalhaes, A. L. B. & de Carvalho, P. A. (2007). Occurrence of exotic ornamental fish in streams in the states of Minas Gerias and Rio de Janeiro, Brazil. *Natureza & Conservacao, 5*(2), 124-129.

de Magalhaes, A. L. B. & Vitule, J. R. S. (2013). Aquarium Industry Threatens Biodiversity. *Science, 341*(6145), 457-457.

FISORNIC.ALL 2011, Import Health Standard for Ornamental Fish and Marine Invertebrates from all Countries. Ministry of Agriculture and Forestry, New Zealand

Moynihan, M. (1971). Successes and failures of tropical mammals and birds. *American Naturalist, 105*(944), 371-&. doi: 10.1086/282731

Nelson, J. S. (2006). Fishes of the world, 4th Edition. *John Wiley & Sons, Inc.*, 601.

Ng, H. H., Tan, H. H., Yeo, D. C. J. & Ng, P. K. L. (2010). Stingers in a strange land: South American freshwater stingrays (Potamotrygonidae) in Singapore. *Biological Invasions, 12*(8), 2385-2388. doi: 10.1007/s10530-009-9663-5

Nijman, V. (2010). An overview of international wildlife trade from Southeast Asia. *Biodiversity and Conservation, 19*(4), 1101-1114. doi: 10.1007/s10531-009-9758-4

Norris, S. M. (2002). A Revision of the African electric catfishes, Family Malapteruridae (Teleostei, Siluriformes), with erection of a new genus and descriptions of fourteen new species, and an annotated bibliography. *Annales Sciences Zoologiques, 289*, 155.

Paine, R. T. (1966). Food web complexity and species diversity. *Amer. Natur., 100*, 65 - 75.

Rainboth, W. J. (1996). Fishes of the Cambodian Mekong. 263.

Roberts, T. R. (1989). The Freshwater Fishes of Western Borneo (Kalimantan Barat, Indonesia). *Memoirs of the California Academy of Sciences, 14*, 210.

Simberloff, D., Martin, J. L., Genovesi, P., Maris, V., Wardle, D. A., Aronson, J., Courchamp F. & Vila, M. (2013). Impacts of biological invasions: what's what and the way forward. Trends in Ecology & Evolution, 28(1), 58-66.

Talwar, P. K. & Jhingran, A. G. (1991). Inland fishes of India and Adjacent Country. *Oxford & IBH Pub. Co.*, 1158.

Tan, H. H. (2006). The Borneo Suckers. *Natural History Publications (Borneo) Sdn. Bhd.*, 242.

Tan, H. H. & Lim, K. P. (2008). *Acarichthys heckelii* (Muller & Troschell), an introduced cichlid fish in Singapore. *Nature in Singapore, 1*, 129-133.

Van der Knaap, M. (2013). May we eat biodiversity? How to solve the impasse of conservation and exploitation of biodiversity and fishery resources. *Aquatic Ecosystem Health & Management, 16*(2), 164-171. doi: 10.1080/14634988.2013.799986

Ward, R. D., Zemlak T. S., Bronwyn H. I., Last, P.R. & Hebert, P.D.N. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B-Biological Sciences, 1716*, 11. doi: 10.1098

Ward, R. D., Hanner, R. & Hebert, P. D. N. (2009). The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology, 74*(2), 329-356. doi: 10.1111/j.1095-8649.2008.02080.x

William, N. E. & Jon, D. F. (2013). Catalog of Fishes. *California Academy of Sciences (CAS)*.

Wolter, C. & Rohr, F. (2010). Distribution history of non-native freshwater fish species in Germany: how invasive are they? *Journal of Applied Ichthyology, 26*, 19-27. doi: 10.1111/j.1439-0426.2010.01505.x

Appendix I: Species list of fish in Singapore Ornamental Trade

Acestrorhynchiidae
Acestrorhynchus isalineae

Adrianichthyidae
Oryzias dancena
Oryzias pectoralis

Akysidae
Akysis longifilis

Akysis prashadi
Akysis vespa

Alestiidae
Alestopetersius caudalis
Brycinus longipinnis
Hemigrammopetersius caudalis
Hydrocynus goliath
Phenacogrammus interruptus

Ambassidae
Parambassis ranga
Gymnochanda filamentosa
Parambassis apogonoides
Parambassis pulcinella
Parambassis siamensis

Amblycipitidae
Amblyceps murraystuarti

Anabantidae
Ctenopoma acutirostre
Ctenopoma muriei
Anabus testudineus

Anostomidae
Abramites hypselonotus
Anostomus anostomus
Laemolyta taeniata
Leporinus affinis
Leporinus fasciatus

Aplocheilidae
Pseudapiplatys annulatus
Aplocheilus lineatus
Aplocheilus panchax

Apteronotidae
Apteronotus albifrons
Apteronotus leptorhynchus

Arapaimidae
Heterotis niloticus
Arapaima gigas

Asphredinidae
Bunocephalus amaurus

Auchenipteridae
Asterophysus batrachus
Liosomadoras oncinus
Tatia intermedia
Tatia perugiae

Badidae
Badis badis
Dario dario

Dario hysginon

Bagridae
Auchenoglanis occidentalis
Bagrichthys hypselopterus
Bagrichthys macracanthus
Bagroides melapterus
Hemibagrus wyckii
Horabagrus brachysoma
Horabagrus nigricollaris
Hyalobagrus flavus
Hyalobagrus ornatus
Leiocassis micropogon
Mystus bimaculatus
Mystus bocourti
Mystus tengara
Pelteobagrus ornatus

Pseudomystus heokhuii
Pseudomystus siamensis
Pelteobagrus fulvidraco

Balitoridae
Sinogastromyzon nantaiensis
Sinogastromyzon puliensis
Beaufortia kweichowensis
Micronoemacheilus pulcher
Pseudogastromyzon fasciatus
Aborichthys elongatus
Acanthocobitis botia
Beaufortia leveretti
Gastromyzon ctenocephalus
Gastromyzon ocellatus
Gastromyzon punctulatus
Gastromyzon scitulus
Gastromyzon zebrinus
Homaloptera bilineata
Homaloptera nebulosus
Homaloptera orthogoniata
Homaloptera parclitella
Homaloptera tweediei
Homaloptera zollingeri
Liniparhomaloptera disparis
Mesonoemacheilus triangularis
Physoschistura rivulicola
Schistura balteata
Schistura magnifluvis
Schistura marhneti
Schistura notostigma
Schistura pridii
Schistura vinciguerrae
Sewellia cf albisuera
Sewellia lineolata
Sewellia speciosa
Vaillantella maassi
Yunnanilus brevis
Yunnanilus cruciatus
Triplophysa siluroides

Bedotiidae
Bedotia geayi

Callichthyidae
Aspidoras pauciradiatus
Brochis britskii
Brochis splendens
Corydoras acutus
Corydoras adolfoi
Corydoras aeneus
Corydoras amandajanea
Corydoras arcuatus
Corydoras caudimaculatus
Corydoras cervinus
Corydoras concolor
Corydoras copei
Corydoras crypticus
Corydoras davidsandsi
Corydoras diphyes
Corydoras duplicareus
Corydoras ehrhardti
Corydoras elegans
Corydoras eques
Corydoras fowleri
Corydoras gossei
Corydoras habrosus
Corydoras imitator
Corydoras julii
Corydoras kanei
Corydoras leopardus
Corydoras loxozonus
Corydoras melanistius
Corydoras melanotaenia
Corydoras melini
Corydoras metae
Corydoras napoensis
Corydoras narcissus
Corydoras nattereri
Corydoras ornatus
Corydoras orphnopterus
Corydoras osteocarus
Corydoras ourastigma

Corydoras paleatus
Corydoras panda

Corydoras pantanalensis
Corydoras pulcher
Corydoras reynoldsi
Corydoras robineae
Corydoras schwartzi
Corydoras semiaquilus
Corydoras septentrionalis
Corydoras serratus
Corydoras seussi
Corydoras sodalis
Corydoras steindachneri
Corydoras sterbai
Corydoras surinamensis
Corydoras sychri
Corydoras trilineatus
Corydoras virginiae
Dianema longibarbis
Dianema urostriata
Megalechis thoracata
Scleromystax barbatus

Catostomidae
Myxocyprinus asiaticus
Minytrema melanops
Carpiodes velifer

Centrarchidae
Lepomis gibbosus
Lepomis megalotis

Cetopsidae
Cetopsis coecutiens

Chacidae
Chaca chaca

Channidae
Parachanna africana
Channa asiatica
Channa aurantimaculata
Channa bankanensis
Channa barca

Channa bleheri
Channa gachua

Channa harcourtbutleri
Channa lucius
Channa marulioides
Channa marulius
Channa melasoma
Channa micropeltes
Channa orientalis
Channa ornatipinnis
Channa panaw
Channa pleurophthalma
Channa pulchra
Channa sp. (Fire Blood)
Channa stewartii

Characidae
Aphyocharax anisitsi
Astyanax jordani
Astyanax mexicanus
Aphyocharax erythrurus
Aphyocharax paraguayensis
Aphyocharax rathbuni
Axelrodia riesei
Boehlkea fredcochui
Chalceus erythrurus
Chalceus macrolepidotus
Colossoma macropomum
Corynopoma riisei
Exodon paradoxus
Gymnocorymbus ternetzi
Hasemania nana
Hemigrammus bleheri
Hemigrammus erythrozonus
Hemigrammus marginatus
Hemigrammus ocellifer
Hemigrammus rhodostomus
Hemigrammus rodwayi
Hemigrammus stictus
Hemigrammus ulreyi
Hemigrammus unilineatus
Hyphessobrycon amandae
Hyphessobrycon anisitsi
Hyphessobrycon bentosi
Hyphessobrycon columbianus
Hyphessobrycon eques

Characidae

Hyphessobrycon erythrostigma
Hyphessobrycon flammeus
Hyphessobrycon griemi
Hyphessobrycon haraldschultzi
Hyphessobrycon herbertaxlrodi

Hyphessobrycon megalopterus
Hyphessobrycon pulchripinnis
Hyphessobrycon rosaceus
Hyphessobrycon roseus
Hyphessobrycon serpae
Hyphessobrycon socolofi
Hyphessobrycon sweglesi
Hyphessobrycon takasei
Inpaichthys kerri
Megalamphodus sweglesi
Metynnis argenteus
Metynnis hypsauchen
Metynnis maculatus
Moenkhausia pittieri
Moenkhausia sanctaefilomenae
Myleus rubripinnis
Myleus schomburgkii
Nematobrycon palmeri
Paracheirodon axelrodi
Paracheirodon innesi
Paracheirodon simulans
Petitella georgiae
Piaractus brachypomus
Prionobrama filigera
Pristella maxillaris
Saccoderma melanostigma
Salminus brasiliensis
Thayeria boehlkei
Thayeria obliqua

Cheilodontidae

Chilodus punctatus

Cichlidae

Gymnogeophagus balzanii
Etroplus suratensis
Cichla kelberi
Cichla ocellaris

Cichla orinocensis
Cichla temensis
Cichlasoma bimaculatum
Cichlasoma octofasciatum
Cichlasoma salvini
Altolamprologus calvus

Aulonocara baenschi

Aulonocara hueseri
Aulonocara nyassae
Aulonocara rubescens
Champsochromis caeruleus
Copadichromis mbenjii
Copadichromis verduyni
Cyphotilapia frontosa
Cyprichromis leptosoma
Cyrtocara moorii
Dimidiochromis compressiceps
Dimidiochromis kiwinge
Gephyrochromis moorii
Hemichromis bimaculatus

Hemichromis stellifer
Julidochromis marlieri
Julidochromis regani
Julidochromis transcriptus
Maylandia callainos
Maylandia zebra
Mchenga flavimanus
Melanochromis auratus
Neolamprologus cf brichardi
Neolamprologus leleupi
Neolamprologus tetracanthus
Neolamprologus tretocephalus
Nimbochromis venustus
Ophthalmotilapia ventralis
Oreochromis mossambicus
Paratilapia polleni
Paretroplus damii
Paretroplus maculatus
Paretroplus nourissati
Pelvicachromis pulcher
Placidochromis milomo
Pseudotropheus auratus
Pseudotropheus demasonii
Pseudotropheus elongatus

Pseudotropheus estherae
Pseudotropheus lanisticola
Pseudotropheus tropheops
Sciaenochromis ahli
Steatocranus casuarius
Steatocranus tinanti
Stigmatochromis pleurospilus

Tilapia buttikoferi
Tilapia cessiana
Tropheus duboisi
Tropheus moorii
Tropheus sp. 'Ikola'
Etroplus maculatus
Archocentrus nigrofasciatus
Hypsophrys nicaraguensis
Parachromis dovii
Parachromis loisellei
Parachromis managuensis
Thorichthys helleri
Thorichthys meeki

Vieja argentea
Vieja synspila
Herichthys carpintis
Herichthys cyanoguttatus
Rocio octofasciata
Acarichthys heckelii
Aequidens pulcher
Aequidens rivulatus
Apistogramma agassizii
Apistogramma bitaeniata
Apistogramma cacatuoides
Apistogramma elizabethae
Apistogramma iniridae
Apistogramma macmasteri
Apistogramma nijsseni
Apistogramma trifasciata
Apistogramma viejita
Astronotus ocellatus
Crenicichla compressiceps
Crenicichla johanna
Crenicichla saxatilis
Crenicichla strigata
Dicrossus filamentosus
Dicrossus maculatus

Cichlidae
Geophagus altifrons
Geophagus brasiliensis
Geophagus surinamensis
Heros severus
Hypselecara coryphaenoides
Hypselecara temporalis
Laetacara thayeri
Mesonauta festivus
Microgeophagus altispinosus
Microgeophagus ramirezi
Pterophyllum altum
Pterophyllum leopoldi
Pterophyllum scalare
Retroculus lapidifer
Satanoperca acuticeps
Symphysodon aequifasciatus
Symphysodon discus
Uaru amphiacanthoides

Citharinidae
Distichodus affinis
Distichodus sexfasciatus

Clariidae
Clarius gariepinus
Gymnallabes typus
Clarias batrachus

Cobitidae
Serpenticobitis octozona
Cobitis sinensis
Leptobotia elongata
Misgurnus anguillicaudatus
Acantopsis choirorhychos
Botia dario
Botia histrionica
Botia kubotai
Botia lohachata
Botia rostrata
Botia striata
Chromobotia macracanthus
Cobitis laoensis
Kottelatlimia pristes

Pangio anguillaris
Pangio doriae
Pangio kuhlii
Pangio malayana
Pangio myersi
Pangio pangia
Pangio semicincta
Syncrossus beauforti
Syncrossus berdmorei
Syncrossus helodes
Syncrossus hymenophysa
Tuberoschistura sp. arakanensis
Yasuhikotakia eos
Yasuhikotakia modesta
Yasuhikotakia morleti
Yasuhikotakia nigrolineata
Yasuhikotakia sidthimunki
Yasuhikotakia splendida
Leptobotia taeniops

Crenuchidae
Poecilocharax weitzmani

Ctenoluciidae
Ctenolucius hujeta

Cynodontidae
Hydrolycus scomberoides
Rhaphiodon vulpinus

Cyprinidae
Rhodeus ocellatus
Tanakia himantegus
Zacco platypus
Candidia barbatus
Carassius auratus
Paracheilognathus himantegus
Cyprinus carpio
Acheilognathus macropterus
Acrossocheilus fasciatus
Balantiocheilus melanopterus
Barbonymus schwanenfeldii
Barilius dogarsinghi
Barilius hukaungensis

Boraras brigittae
Boraras maculatus
Boraras merah
Boraras micros
Boraras urophthalmoides
Celestichthys margaritatus
Chela dadiburjori
Chela laubuca
Crossocheilus atrilimes
Crossocheilus langei
Crossocheilus reticulatus
Crossocheilus siamensis
Cyclocheilichthys janthochir
Danio albolineatus
Danio choprai
Danio dangila
Danio erythromicron
Danio freegradei
Danio kerri
Danio kyathit
Danio nigrofasciatus
Danio pathirana
Danio rerio
Danio roseus
Danio sp. (blue red stripe)
Danio sp. Hikari
Danio tinwini
Danionella dracula
Devario auropurpureus
Devario browni
Devario malabaricus
Devario shanensis
Devario sondhii
Devario sp. Giraffe
Eirmotus furvus
Eirmotus insignis
Eirmotus isthmus
Eirmotus octozona
Epalzeorhynchos bicolor
Epalzeorhynchos frenatum
Epalzeorhynchos kalopterus
Esomus metallicus
Garra annandalei
Garra ceylonensis

Cyprinidae

Garra flavatra
Garra gotyla
Garra gravelyi
Garra rufa
Hampala ampalong
Hampala bimaculata
Hampala macrolepidota
Hypselobarbus curmuca
Hypsibarbus vernayi
Hypsibarbus wetmorei
Labiobarbus ocellatus
Leptobarbus hoevenii
Macrochirichthys macrochirus
Microdevario cf kubotai
Microdevario cf rubescens
Microdevario nana
Mystacoleucus atridorsalis
Mystacoleucus marginatus
Neolissochilus hexagonolepis
Neolissochilus sumatranus
Opsarius pulchellus
Oreichthys parvus
Osteochilus pentalineatus
Osteochilus vittatus
Paedocypris micromegethes
Paedocypris progenetica
Pectenocypris korthausae
Probarbus jullieni
Puntius conchonius
Puntius denisonii
Puntius everetti
Puntius fasciatus
Puntius filamentosus
Puntius foerschi
Puntius gelius
Puntius lateristriga
Puntius lineatus
Puntius oligolepis
Puntius orphoides
Puntius padamya
Puntius pentazona
Puntius phutunio
Puntius rhomboocellatus

Puntius semifasciolatus
Puntius stoliczkanus
Puntius tetrazona
Puntius ticto
Puntius titteya
Rasbora agilis
Rasbora bankanensis
Rasbora borapetensis
Rasbora brittani
Rasbora caudimaculata
Rasbora cephalotaenia
Rasbora dorsiocellata
Rasbora elegans
Rasbora ennealepis
Rasbora gerlachi
Rasbora gracilis
Rasbora kalochroma
Rasbora patrickyapi
Rasbora pauciperforata
Rasbora paucisqualis
Rasbora rubrodorsalis
Rasbora rutteni
Rasbora sarawakensis
Rasbora trilineata
Rasbora vulcanus
Rasbora wilpita
Rasboroides vaterifloris
Sawbwa resplendens
Sundadanio axelrodi
Tanichthys albonubes
Tanichthys micagemmae
Tor soro
Tor tambra
Tor tor
Trigonostigma espei
Trigonostigma hengeli
Trigonostigma heteromorpha
Cyprinella lutrensis
Catlocarpio siamensis
Elopichthys bambusa

Cyprinodontidae

Jordanella floridae

Dasyatidae

Himantura oxyrhynchus

Datnioididae

Datnioides campbelli
Datnioides microlepis
Datnioides pulcher
Datnioides quadrifasciatus
Datnioides undecimradiatus

Doradidae

Acanthodoras cataphractus
Agamyxis pectinifrons
Megalodoras uranoscopus
Oxydoras niger
Platydoras costatus

Electrophoridae

Electrophorus electricus

Eleotridae

Hypseleotris compressa
Mogurnda pulchra
Tateurndina ocellicauda

Erethistidae

Erethistes filamentosa
Erethistes hara
Erethistes jerdoni
Hara minuscula

Erythrinidae

Erythrinus erythrinus
Hoplias malabaricus

Esocidae

Esox lucius

Gasteropelecidae

Carnegiella strigata
Gasteropelecus levis
Gasteropelecus maculatus
Gasteropelecus sternicla
Thoracocharax stellatus

Gobiidae
Rhinogobius brunneus
Sicyopterus japonicus
Pseudogobiopsis oligactis
Sicyopus leprurus
Brachygobius doriae
Brachygobius nunus
Gobiopterus brachypterus
Gobiopterus chuno
Oxyeleotris marmorata
Rhinogobius leavelli
Sicyopterus cynocephalus
Sicyopterus fasciatus
Stigmatogobius sadanundio
Stiphodon carisa
Stiphodon ornatus
Stiphodon carinata
Stiphodon elegans

Gymnarchidae
Gymnarchus niloticus

Gymnotidae
Gymnotus carapo
Gymnotus pedanopterus
Gymnotus tigre
Gymnotus varzea

Gyrinocheilidae
Gyrinocheilus aymonieri
Gyrinocheilus pennocki

Helostomatidae
Helostoma temminckii

Hemiodontidae
Hemiodopsis gracilis

Hemirhamphidae
Dermogenys puscilla
Hemirhamphodon chrysopunctatus
Hemirhamphodon kuekenthali
Hemirhamphodon pogonognathus

Hemirhamphodon tengah

Nomorhamphus ebrardtii
Nomorhamphus liemi

Hepsetidae
Hepsetus odoe

Heteropneustidae
Heteropneustes fossilis

Indostomidae
Indostomus crocodilus

Latidae
Lates mariae

Lebiasinidae
Copella eigenmanni
Copella nattereri
Nannostomus beckfordi
Nannostomus eques
Nannostomus marginatus
Nannostomus mortenthaleri
Nannostomus trifasciatus
Nannostomus unifasciatus

Lepidosirenidae
Lepidosiren paradoxa

Lepisosteidae
Atractosteus spatula
Atractosteus tristoechus
Lepisosteus oculatus

Loricariidae
Rineloricaria parva
Rineloricaria sp. (L010a)
Acanthicus adonis
Acanthicus hystrix
Ancistrini sp. (L239)
Ancistrus dolichopterus

Ancistrus ranunculus
Ancistrus sp. (L071)

Ancistrus sp. (L107)

Ancistrus sp. (L144)

Ancistrus sp. (L213)
Ancistrus sp. (L255)
Baryancistrus demantoides
Baryancistrus niveatus
Baryancistrus sp. (L018)
Baryancistrus sp. (L047)
Baryancistrus sp. (L081)
Baryancistrus sp. (L142)
Baryancistrus sp. (L177)
Dekeyseria brachyura
Dekeyseria pulchra
Dekeyseria vittata
Farlowella acus
Farlowella gracilis
Glyptoperichthys gibbiceps
Glyptoperichthys joselimaianus
Hemiancistrus sp. (L128)
Hemiancistrus subviridis
Hopliancistrus tricornis
Hypancistrus inspector
Hypancistrus sp. (L066)
Hypancistrus sp. (L174)
Hypancistrus sp. (L260)
Hypancistrus sp. (L262)
Hypancistrus sp. (L270)
Hypancistrus sp. (L333)
Hypancistrus sp. (L340)
Hypancistrus zebra
Hypostomus punctatus
Hypostomus sp. (L360)
Lamontichthys llanero
Leporacanthicus cf galaxias (L007)
Leporacanthicus joselimai
Leporacanthicus sp. (L314)
Leporacanthicus triactis
Megalancistrus parananus
Oligancistrus sp. (L020)
Oligancistrus sp. (L030)
Oligancistrus sp. (L354)
Otocinclus cocama

Otocinclus flexilis

Otocinclus vestitus

Otocinclus vittatus
Panaque cf nigrolineatus (L027 Tapajos)

Loricariidae
Panaque cf nigrolineatus (L027 Xingu)
Panaque cf nigrolineatus (L027a)
Panaque cf nigrolineatus (L027b)
Panaque maccus
Panaque sp. (L090)
Panaque sp. (L191)
Panaque sp. (L204)
Panaque sp. (L271)
Parancistrus aurantiacus
Parancistrus magnum
Peckoltia sabaji
Peckoltia sp. (L049)
Peckoltia sp. (L134)
Peckoltia sp. (L209)
Peckoltia sp. (L243)
Peckoltia vittata
Pseudacanthicus leopardus
Pseudacanthicus serratus
Pseudacanthicus sp. (L024)
Pseudacanthicus sp. (L025)
Pseudacanthicus sp. (L064)
Pseudacanthicus sp. (L097)
Pseudacanthicus sp. (L273)
Pseudacanthicus sp. (L600)
Pseudancistrus sp. (L056)
Pseudohemiodon lamina
Pseudorinelepis genibarbis
Pterygoplichthys gibbiceps
Pterygoplichthys pardalis
Scobinancistrus aureatus
Scobinancistrus cf pariolispos
Sturisoma aureum
Sturisoma foerschi
Sturisoma panamense
Sturisoma pursochi

Malapteruridae
Malapterurus electricus
Malapterurus microstoma

Mastacembelidae
Macrognathus aral
Macrognathus circumcinctus

Macrognathus panacalus

Macrognathus siamensis
Macrognathus tapirus
Macrognathus zebrinus
Mastacembelus armatus
Mastacembelus erythrotaenia
Mastacembelus flavidus

Melanotaeniidae
Chilatherina sentaniensis
Glossolepis incisus
Glossolepis multisquamatus
Iriatherina werneri
Melanotaenia australis
Melanotaenia boesmani
Melanotaenia duboulayi
Melanotaenia herbertaxelrodi
Melanotaenia lacustris
Melanotaenia maccullochi
Melanotaenia mubiensis
Melanotaenia multisquamatus
Melanotaenia nigrans
Melanotaenia parkinsoni
Melanotaenia praecox
Melanotaenia trifasciata

Mochokidae
Synodontis angelicus
Synodontis brichardi
Synodontis caudalis
Synodontis decorus
Synodontis eupterus
Synodontis greshoffi
Synodontis multipunctatus
Synodontis nigriventris
Synodontis njassae
Synodontis schoutedeni

Mormyridae
Brienomyrus brachyistius
Campylomormyrus cassaicus
Gnathonemus petersil
Gnathonemus tamandua
Mormyrus hasselquistii

Mormyrus longirostris

Nandidae
Nandus nandus
Pristolepis grooti

Nothobranchiidae
Aphyosemion australe
Aphyosemion gardneri
Aphyosemion striatum
Fundulopanchax gardneri
Nothobranchius eggersi
Nothobranchius foerschi
Nothobranchius guentheri
Nothobranchius korthausae
Nothobranchius orthonotus
Nothobranchius rachovii

Notopteridae
Notopterus notopterus
Chitala blanci
Chitala ornata

Osphronemidae
Betta albimarginata
Betta anabatoides
Betta antoni
Betta brownorum
Betta channoides
Betta coccina
Betta edithae
Betta enisae
Betta falx
Betta hipposideros
Betta ideii
Betta krataios
Betta macrostoma
Betta mandor
Betta ocellata
Betta pallifina
Betta patoti
Betta persephone
Betta pi
Betta prima

Osphronemidae

Betta raja
Betta renata
Betta rutilans
Betta simorum
Betta smaragdina
Betta splendens
Betta tussyae
Betta uberis
Colisa chuna
Colisa labiosa
Colisa lalia
Ctenops nobilis
Luciocephalus aura
Luciocephalus pulcher
Luciosoma setigerum
Macropodus concolor
Macropodus opercularis
Macropodus spechti
Malpulutta kretseri
Osphronemus goramy
Osphronemus laticlavius
Parasphaerichthys lineatus
Parasphaerichthys ocellatus
Parosphromenus anjunganensis
Parosphromenus bintan
Parosphromenus deissneri
Parosphromenus linkei
Parosphromenus nagyi
Parosphromenus opallios
Parosphromenus ornaticauda
Parosphromenus sumatranus
Sphaerichthys acrostoma
Sphaerichthys osphromenoides
Sphaerichthys selatanensis
Sphaerichthys vaillanti
Trichogaster chuna
Trichogaster leerii
Trichogaster microlepis
Trichogaster pectoralis
Trichopodus trichopterus
Trichopsis pumila
Trichopsis schalleri
Trichopsis vittata

Osteoglossidae

Scleropages formosus
Scleropages jardinii
Scleropages legendrei
Scleropages leichardti
Osteoglossum bicirrhosum
Osteoglossum ferreirai

Pangasianodon

Pangasianodon gigas
Pangasianodon hypophthalmus
Pangasianodon sanitwongsei

Pantodontidae

Pantodon buchholzi

Pimelodidae

Aguarunichthys torosus
Brachyplatystoma juruense
Brachyplatystoma tigrinum
Brachyplatystoma vaillantii
Perrunichthys perruno
Phractocephalus hemioliopterus
Pimelodus pictus

Pinirampus pirinampu
Platystomatichthys sturio
Pseudoplatystoma fasciatum
Pseudoplatystoma tigrinum
Sorubim lima
Sorubimichthys planiceps

Poeciliidae

Poecilia latipinna
Gambusia affinis
Aplocheilichthys normani
Poecilia sphenops
Poecilia velifera
Poecilia wingei
Xiphophorus hellerii
Xiphophorus maculatus
Xiphophorus nigrensis
Xiphophorus variatus
Poecilia reticulata
Micropoecilia minima

Polypteridae

Erpetoichthys calabaricus
Polypterus bichir
Polypterus delhezi
Polypterus endlicheri
Polypterus mokelembembe
Polypterus ornatipinnis
Polypterus palmas
Polypterus retropinnis
Polypterus senegalus
Polypterus teugelsi
Polypterus weeksii

Potamotrygonidae

Paratrygon aiereba
Potamotrygon castexi
Potamotrygon henlei
Potamotrygon hystrix
Potamotrygon leopoldi
Potamotrygon menchacai
Potamotrygon motoro
Potamotrygon orbignyi
Potamotrygon schroederi

Procheilodontidae

Semaprochilodus taeniurus

Protopteridae

Protopterus dolloi

Pseudomugilidae

Pseudomugil gertrudae
Pseudomugil furcatus
Pseudomugil signifer

Pseudopimelodidae

Pseudopimelodus bufonius
Pseudopimelodus zungaro

Rivulidae

Austrolebias bellottii
Austrolebias nigripinnis

Schilbeidae

Parailia pellucida
Neotropius acutirostris

Siluridae

Belodontichthys dinema
Kryptopterus apogon
Kryptopterus bicirrhis
Kryptopterus macrocephalus
Kryptopterus minor
Ompok bimaculatus
Ompok fumidus
Pterocryptis berdmorei
Silurichthys phaiosoma
Wallago leerii

Sissoridae

Glyptothorax trilineatus
Bagarius bagarius
Gagata dolichonema

Soleidae

Brachirus panoides

Sternopygidae

Eigenmannia virescens

Syngnathidae

Microphis brachyurus
Dorichthys doekhatoides
Dorichthys martensii

Telmatherinidae

Marosatherina ladigesi
Telmatherina ladigesi

Tetraodontidae

Takifugu ocellatus
Tetraodon lineatus
Tetraodon mbu
Tetraodon miurus
Auriglobus modestus
Carinotetraodon borneensis
Carinotetraodon irrubesco

Carinotetraodon lorteti
Carinotetraodon salivator
Carinotetraodon travancoricus
Tetraodon baileyi
Tetraodon biocellatus
Tetraodon cochinchinensis
Tetraodon erythrotaenia
Tetraodon fluviatilis
Tetraodon leiurus
Tetraodon nigroviridis
Tetraodon palembangensis
Tetraodon suvattii
Colomesus asellus

Toxotidae

Toxotes blythii
Toxotes chatareus
Toxotes jaculatrix

CHAPTER III

_____

Testing the effectiveness of COI barcodes

for the identification of the native and

invasive freshwater fish of Singapore

Abstract

One of the main challenges for identifying species using DNA sequences ('DNA barcoding') is obtaining complete or near-complete species coverage within a taxon group. It is only with such coverage that one can test whether (1) all species within this taxon group have distinct barcodes and (2) intra- and interspecific pairwise distances are overlapping, which could interfere with identification success rates. Indeed, some authors have suggested that DNA barcoding can only work well at a local scale because the interspecific distances are more likely to be discrete for small species samples. However, this proposition is rarely tested because there are few studies with near-complete species coverage. I tested the feasibility of obtaining a complete species barcode database for the 108 of Singapore's native and invasive freshwater fish species as well as the effectiveness of COI barcodes for the identification of these Singapore's freshwater fish in a local and global setting. I obtained species coverage of 83% (89/108) for the freshwater fish species of Singapore (383 individuals: ca. 4 specimens/species) and demonstrated an identification efficiency of 79% to 97% depending on the method and stringency of analytical technique. 95% of the species in this study possess unique consensus barcodes and I found only two cases of species sharing barcodes. Obtaining complete COI coverage for the freshwater fish diversity proves to be challenging despite the study being restricted to a small area such as Singapore. Nevertheless, my high identification success

rates demonstrate that COI can be effectively used to allocate specimens to species at a country-scale (or even regional scale) as long as the species-genus ratio is low (our sample: 1.3:1).

3.1. Introduction

The use of DNA sequences (mainly COI) for identifying species, in short DNA barcoding, has been of great interest for many biologists because of its potential application value in biodiversity, conservation (Gompert *et al.* 2006; Holmes *et al.* 2009), and bio-protection against invasive flora and fauna (Bleeker *et al.* 2008; Chown *et al.* 2008). However, many studies in DNA barcoding have been criticized because they involved very incomplete DNA barcode databases. Such incomplete databases make it more likely that the sampled species have discrete DNA barcodes because not all sister species pairs have been sampled. Indeed, a number of studies including Meyer and Paulay's (2005) landmark work have demonstrated that DNA barcoding efficiency is lower in comprehensively and completely sampled clades. It is therefore generally acknowledged that the completeness of species coverage of a DNA barcode database will be an important factor when evaluating the usefulness of DNA barcodes.

The process of obtaining complete species coverage is by no means easy or convenient (Kwong *et al.* 2012). A survey of the approximately 1740 DNA barcoding publications in the Web of Science reveals that the majority of studies have relatively poor species coverage; this is even found in studies that only focus on one genus. Poor species coverage is similarly observed for most DNA barcoding studies with a regional focus (often looking at family-level taxa in a

particular country). The species coverage typically ranged from 9% to 78% (Valdez-Moreno *et al.* 2009; Page & Hughes 2010; Valdez-Moreno et al. 2010; Lakra et al. 2011; Park et al. 2011; Sonet et al. 2011; Tavares et al. 2011; Bergsten et al. 2012; Costa et al. 2012; Dai et al. 2012; Gattolliat et al. 2012; Kumar et al. 2012; Webb et al. 2012) and I only found seven studies with near complete species coverage (>90%). They are two studies on bats in the Neotropical and Palaearctic Regions (Clare et al. 2011; Kruskop et al. 2012), one study on flowering plants and conifers in Wales (de Vere et al. 2012), two studies on butterflies in Germany and Romania (Dinca *et al.* 2011; Hausmann et al. 2011), one study on the birds of North America, one study on the arthropods of Australia (Hendrich et al. 2010) and one study on Canadian freshwater fish (Hubert et al. 2008). Here, I add another data point by generating a DNA barcode dataset for Singapore's native and invasive fish species.

Species coverage is easier to achieve regionally because it reduces the political and geographical challenges that come with obtaining tissues at a global scale. For example, it avoids research permit applications for many countries that are time consuming and all the other costs that come with planning expeditions to many field sites for collecting specimens (Funk et al, 2005). An alternative is obtaining specimens from museums, but unfortunately they do not have molecular-grade tissues for most species given that a large proportion of species have only been collected once (Lim, Balke, & Meier, 2012)

and many fish specimens were preserved in formalin which interferes with the extraction of DNA (Zimmermann et al., 2008).

Focusing on regional barcode databases is thus attractive and has the additional advantage that it realistic because many identification problems are regional problems. For example, government agencies are mostly interested in documenting organisms within specific countries, for example, for the purpose of assessing biodiversity or controlling and identifying pests. Examples for regional or national-scale databases include DNA barcoding projects for fish of Nayband National Park in the Persian Gulf (Asgharian et al. 2011), species identification of Tanzanian antelopes using DNA barcoding (Bitanyi et al. 2011) and DNA barcoding of Canadian freshwater fish for the purpose of bio-protection (Hubert et al. 2008). Of course, there are exceptions. Any use of DNA barcoding for monitoring the international trade in food and ornamental species will require a more global perspective in that all or at least most traded species need to be covered.

Regional DNA barcode databases are also attractive because they may circumvent the biggest problem with DNA barcoding; i.e., the lacks of a barcoding gap between inter- and intraspecific variability (Meyer & Paulay, 2005). Regional coverage makes it more likely that the proportion of closely related species is lower and fewer populations of widespread species are sampled. Compared to global sampling, it

will increase interspecific variability and reduce intraspecific distances (Bergsten et al. 2012). Indeed, evidence suggests that COI will exhibit greater failure rates with increased species coverage as has been documented in, for example, a study on Neotropical butterflies (Martin 2007). This result is not surprising given that one could predict that COI barcodes should be more useful when coverage of closely related species is low as is often the case, for example, in the international food and aquarium trade, and my case study of freshwater fishes Singapore. For this species assembly, I find that most genera have only one or a few species (108 species from 82 genus; an average of 1.3 species per genus).

Currently, 108 species of extant freshwater fishes are known to occur in Singapore. Thirty-two of these are indigenous to the island. They are mainly found in the remaining forest streams that are usually found in nature reserves. The vast majority of species are exotics that inhabit Singapore's reservoirs. Many of the species have established viable populations in the reservoirs but some have also invaded the forest streams (e.g., the marble goby, *Oxyeleotris marmorata*). Faced with the irony that two thirds of the country's freshwater fishes comprises of exotic species, conservationists and environmental biologists in Singapore are concerned about the well being of the remaining native species. As such, DNA barcodes could serve as a valuable tool for detecting and monitoring exotic species in Singapore's natural freshwater systems. This is particularly important for the forest

streams habitat where the majority of the native freshwater fishes are residing. While many of the studies concerning environmental genomics are about the detection of micro-organisms via DNA (e.g., viruses, bacteria, nematodes, protists), there are already a few studies that have demonstrated the possibility to amplify fish DNA from freshwater (Jerde et al. 2011; Minamoto et al. 2012; Takahara et al. 2012; Thomsen et al. 2012). This area of research is quickly developing and demonstrating the presence of a species via eDNA will become routine before soon.

Here, I will test the feasibility of obtaining near complete species coverage for the fish fauna of Singapore. The efficiency of allocating specimens to species using COI for both native and exotic freshwater fish species will be tested and documented. This study also determines the identification success rates of DNA barcodes for a fauna with a low species to genus ratio using the different analysis methods ranging from traditional methods based on global alignments (e.g., Best Match: BM; Best Close Match: BCM) to methods such as BRONX that look for short DNA tags that are diagnostic at the species level.

3.2. Materials and Methods

3.2.1. List of species and specimen collection

The complete list of freshwater fish of Singapore was assembled using the literature (Baker & Lim 2008; Ng & Tan 2010; Yeo & Chia 2010; Lim & Ng 2012), the National Parks Board's (Singapore) website (http://goo.gl/CRd4bP), data from a reservoir survey conducted by Public Utility Board (PUB), anglers' records, and additional personal sightings of undocumented species such as *Arapaima gigas*. A list of 108 species of freshwater fishes was assembled; 32 are indigenous species while 76 are introduced species.

Specimens and tissues were collected from several sources: ornamental fish trade, Raffles Museum of Biodiversity Research (RMBR), from the PUB reservoir survey, and freshwater streams in Singapore. The majority of the indigenous species (12 species) were obtained from RMBR and the freshwater streams in Singapore, five species were collected from the aquarium trade, data for 18 species were obtained from Genbank, and seven species were sampled based on specimens/data from a combination of sources. A total of 145 COI sequences were obtained for the 28 indigenous species inclusive of 22 sequences from Genbank. Most indigenous species, except for four, are represented by multiple specimens and an average of 4.75 individuals per species was sampled. The majority of the introduced

species were collected from the trade (19 species) and tissues in RMBR and from the reservoir survey (19 species). Sequences from Genbank were utilized only for 14 introduced species. Nine species were represented by specimens from different sources. A total of 238 COI sequences inclusive of 53 sequences from Genbank were downloaded for the 61 introduced freshwater fish species. All but 11 exotic species are represented by multiple specimens and the average number of specimens per species is 3.9.

Overall, the study includes COI sequences from 383 individuals for 89 species (28 native + 61 introduced species) inclusive of 75 sequences from Genbank for 22 species. 68 individuals from 18 species (14 exotic and 4 native species) were represented by Genbank sequences. 16 species were singletons in the dataset, while the majority of the species were represented by multiple specimens (4.3 specimens/species). Because singletons were known to reduce the identification success rates in barcode datasets (Lim, Balke and Meier 2011), I also used a dataset of 367 sequences where each species was represented by multiple sequences. This dataset contains only 73 species. The details and sources for the specimens were recorded in Appendix II.

3.2.2. DNA Extraction, amplification and sequencing

Genomic DNA was mainly obtained by Phenol/Chloroform extraction. Polymerase chain reactions (PCR) were carried out in 25µl mixture of DNA template using a fish-specific COI primer cocktail:

(Table 3.3.2.I) Primers utilized for amplification of fish COI

.	Forward/ reverse	Primers	Primer sequences	Citation
1	forward	Fish COI FI	TCAACCAACCACA AAGACATTGGCAC	(Ward et al., 2005)
2	forward	Fish COI F2	TCGACTAATCATA AAGATATCGGCAC	(Ward et al., 2005)
3	reverse	Fish COI R1	TAGACTTCTGGGT GGCCAAAGAATCA	(Ward et al., 2005)
4	reverse	Fish COI R2	ACTTCAGGGTGAC CGAAGAATCAGAA	(Ward et al., 2005)

PCR amplifications were carried out with Takara Ex Taq ^{TM} DNA Polymerase, Ex Taq buffer and water. The PCR cycle conditions involves melting temperature of 95°C for 1.5min, annealing temperature of 50°C to 54°C for 1.5min and extension temperature of 72°C for 1.5min for 30 cycles. PCR amplicons were cleaned using BIOLINE SureClean. The purified amplicons served as the template for cycle sequencing reaction using big dye terminator (condition: 30 cycle of 95°C for 30s, 52.5°C for 30s and 60°C for 4min) with the respective primers. Sequences were generated using ABI3730 96-capillary sequencer. Sequencher 4.6 from Gene Code Corporation was used for sequence editing and contig joining.

3.2.3. Species Identification

I employed three methods for species identification: one requires a global alignment for all sequences and was implemented in SpeciesIdentifier (Meier, Shiyang et al. 2006), one uses pairwise local alignments as implemented by BLAST (Altschul, Gish et al. 1990), and one is character-based and implemented in BRONX (Little, 2011).

3.2.3.1. SpeciesIdentifier

Since COI is a protein encoding gene, the global alignment was based on amino acid translations as implemented in Alignmenthelper or Mega 4.1 which integrates Clustal W (Thompson, Higgins et al. 1994). The aligned data were analyzed in SpeciesIdentifier (Meier, Shiyang et al. 2006). SpeciesIdentifier was then used for identifications based on Best Match (BM) and Best Close Match (BCM) analyses. Both analyses were carried out under the following parameters: 1) uncorrected p-distance was used for measuring the distance between two sequences (Srivathsan and Meier, 2011) and 2) 300 bp minimum overlap was required between two sequences prior to distance measurement.

3.2.3.1.1. Best match (BM) analysis

In a BM analysis, each sequence was removed from the dataset and treated as a query for the remaining sequences in the dataset (query = DNA sequence belonging to an unknown species). The query was then matched to species in the dataset based on the smallest pairwise distance. Query identification was considered successful when the query and best matching sequences belong to the same species and the query was thus correctly identified; the identification was unsuccessful when the query and corresponding sequences were incorrectly matched and ambiguous when two or more sequences from different species have equally good matches to the query sequence.

3.2.3.1.2. Best close match (BCM) analysis

A best close match analysis (BCM) is essentially a best match analysis with a set cutoff point or threshold because best matches between a query and an identified sequence are unlikely to be correct when the distances are too large. In BCM a distance threshold was pre-set or determined by an initial assessment of proportion of correct and incorrect identifications when using different thresholds (here ranging from 0 to 12%). An optimal criterion was then selected as the threshold for BCM analyses. The optimal criterion was one where inaccurate

identifications were minimized without major losses in terms of proportion of sequences identified (see Results).

3.2.3.2. BLAST based identification

Multiple sequence alignment based analyses are time-consuming because even a single query sequence needs to be integrated into the alignment. Tools such as BLAST are based on pairwise alignments and thus avoid this complication. In my study I used analyses similar to BM and BCM for identifying query sequences with BLAST. A BLAST database was created using unaligned fasta sequences using NCBI BLAST+ v2.2.28. Each sequence was queried to the database, under settings of MEGABLAST and e-value cut-off of 1e-5 (Altschul et al., 1990). Given that every query sequence will give a hit to itself, all matches of the query sequence to itself were excluded from the result file. After the removal of this hit, I carried out analyses that corresponded to BM and BCM. For BM I used the best hit and for analyses corresponding to BCM I only used the best hit if it was within the distance threshold. The threshold used for BLAST was same as the one determined for SpeciesIdentifier in order to keep the results comparable. All analyses were conducted after excluding any hit <300bp in length.

3.2.3.3. BRONX

In recent years, alignment free approaches to DNA barcoding have gained traction; they are particularly useful because creating large multiple sequence alignments for examining query sequences can be time consuming and BLAST based approaches, although fast, yield approximate results (Little & Stevenson, 2007). I used BRONX (v2) (Little, 2011) which is a "Sequence Identification Engine" that can use short variable motifs of DNA within a sequence and scores the query sequence based on presence or absence of these motifs. Given that motifs are associated with different species, the best score can then be used to make identifications. The BRONX databases were built using tools provided in BRONX2 package. A modified version of the script that gives the top two best hits instead of only a single best hit was used to query sequences against the databases. Along with a hit, BRONX outputs a score corresponding to each hit. Analogous to BM and BCM analyses, I carried out BRONX analyses with or without thresholds; if the latter, the threshold was based on a BRONX score. A high score implies a close hit. Similar to SpeciesIdentifier based analyses, the score threshold was determined by an initial assessment of proportion of correct and incorrect identifications when using different thresholds ranging from 0-500 at intervals of 100. I did not use a threshold >500 as this led to exclusion of too many sequences. An optimum criterion was then selected as the threshold for the "with threshold" analyses.

3.2.3.4. Species specific consensus sequences and diagnostic characters

In order to determine whether each species has a unique consensus barcode (and conceivably diagnostic markers); I generated consensus sequences for each of the 89 species in my dataset, and subjected the dataset to a BCM analysis with a threshold of 0%.

3.3.   Results

3.3.1. Determining the threshold for Best Close Match analysis

I determine the optimal threshold for BCM as being between 2.3 to 2.8% for Singapore freshwater fish. This is the point where identification success, misidentification and ambiguous identification stabilize. The BCM analysis with cutoff point of 2.8% is here used for comparisons with the other methods.
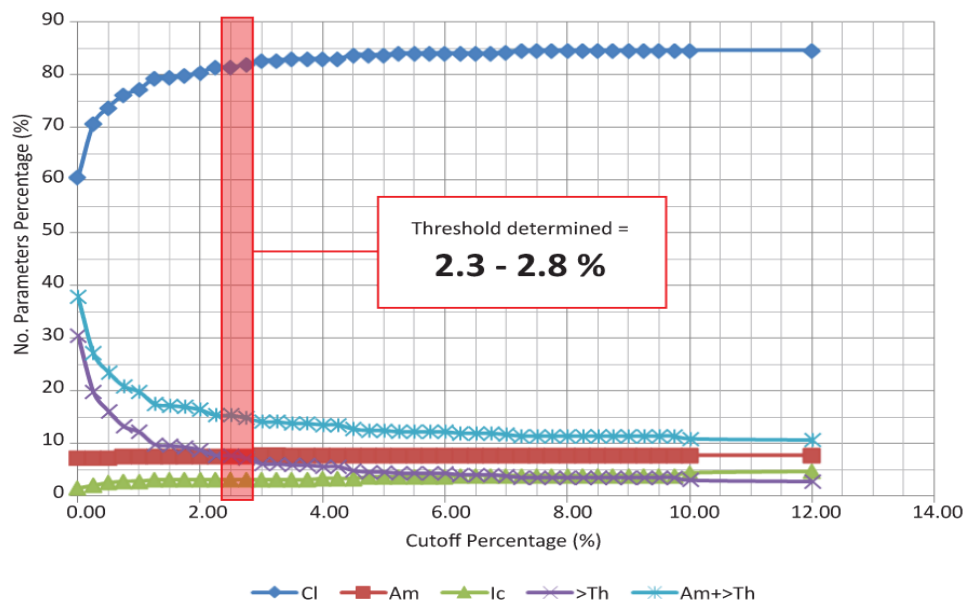
Fig. 3.4.1. The optimal cutoff point for the BCM is 2.3 to 2.8%. CI = correct identification; Am = ambiguous identification; Ic = incorrect identification; >Th = best match is above threshold

3.3.2. Determining the thresholds for the BRONX analysis

The thresholds of BRONX are inversely proportionate to stringency. After conducting BRONX analysis using score thresholds ranging from 0 to 500, a score threshold of 100/200 is observed to maximize the number of sequences identified and while reducing misidentifications (Figure 3.4.2). Since the Singapore dataset is expected to give high identification success in analysis because of the low species to genus ratio, a more stringent score threshold of 200 was here chosen for analysis. This also corresponded to threshold determined in Chapter IV.
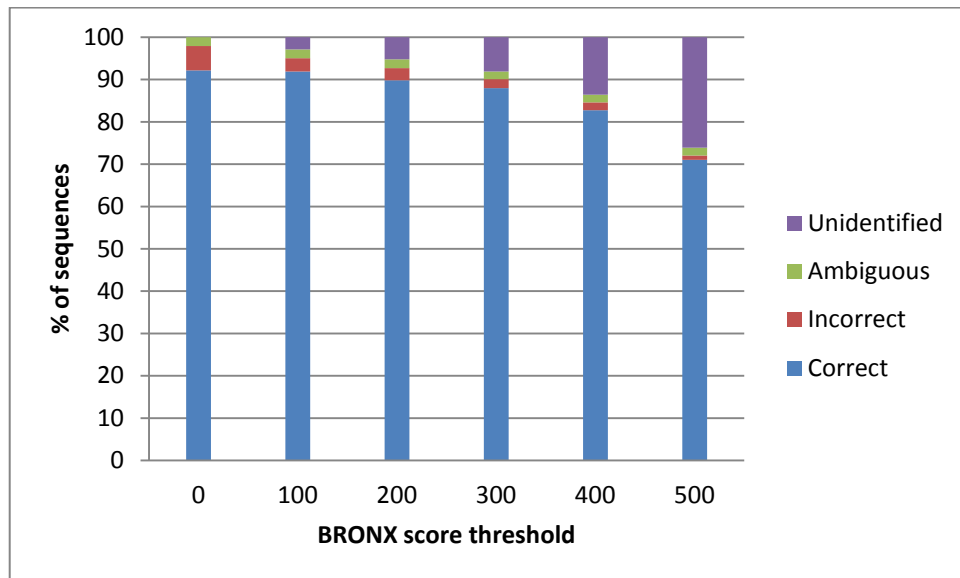
Figure 3.4.2. Percentage of sequences identified using BRONX at different score thresholds

3.3.3. Identification of Singapore's freshwater fish

The identification success of ornamental fish varied from 79-97% depending on dataset and method of identification. Because singletons will contribute to misidentification once they are treated as queries, the dataset without singletons (85-97%) performs better than the corresponding dataset with singletons (79-92%) for all methods of analysis. The singleton species in the dataset are *Pterygoplichthys joselimaianus*, *Puntius binotatus*, *Glossogobius giuris*, *Nomorhamphus liemi*, *Vieja synspila*, *Poecilia sphenops*, *Osphronemus goramy*, *Bostrychus sinensis*, *Hemibagrus nemurus*, *Xiphophorus helleri*, *Monopterus albus*, *Cichla temensis*, *Barbonymus altus*, *Rasbora einthovenii* and *Trichopsis schalleri*. These 15 species yield incorrect identifications in analysis without threshold. However, there are also 34 sequences from four species in the dataset that have an allospecific identical match (between *Oreochromis mossambicus* and *O. niloticus*; between *Pterygoplichthys joselimaianus* and *P. pardalis*) i.e., 4.4% (full dataset) or 5.4% (dataset without singletons) yield misidentifications even in analyses with thresholds.

All analyses with thresholds are able to reduce misidentifications by leaving queries with poor matches unidentified. Therefore, the number of incorrect identifications declines once the dataset without singletons is analyzed (from 5.74-6.78% to 2.87-3.13%). Correspondingly, successful identifications range from 82-94%.

Table 3.4.3.I: Identification efficiency for the different methods of assigning species to specimen

Analysis type	Correct S / W	Incorrect S / W	Ambiguous S / W	Unidentified S / W
BM	85.37/90.19	6.78/2.45	7.83/7.35	- / -
BLAST	84.56/89.37	6.53/2.45	8.09/7.36	0.78/0.54
BRONX	92.16/96.46	5.74/1.63	2.09/1.91	- / -
BCM (2.8%)	82.5 / 88.01	3.13 / 2.17	7.57 / 7.35	6.78 / 2.45
BLAST (3%)	79.63 / 84.20	2.35 / 1.65	8.09 / 7.35	9.92 / 6.82
BRONX (200)	89.81 / 94.01	2.87 / 1.63	2.09 / 1.91	5.22 / 2.45

S = dataset with singleton; W = dataset without singletons; (n%) = threshold
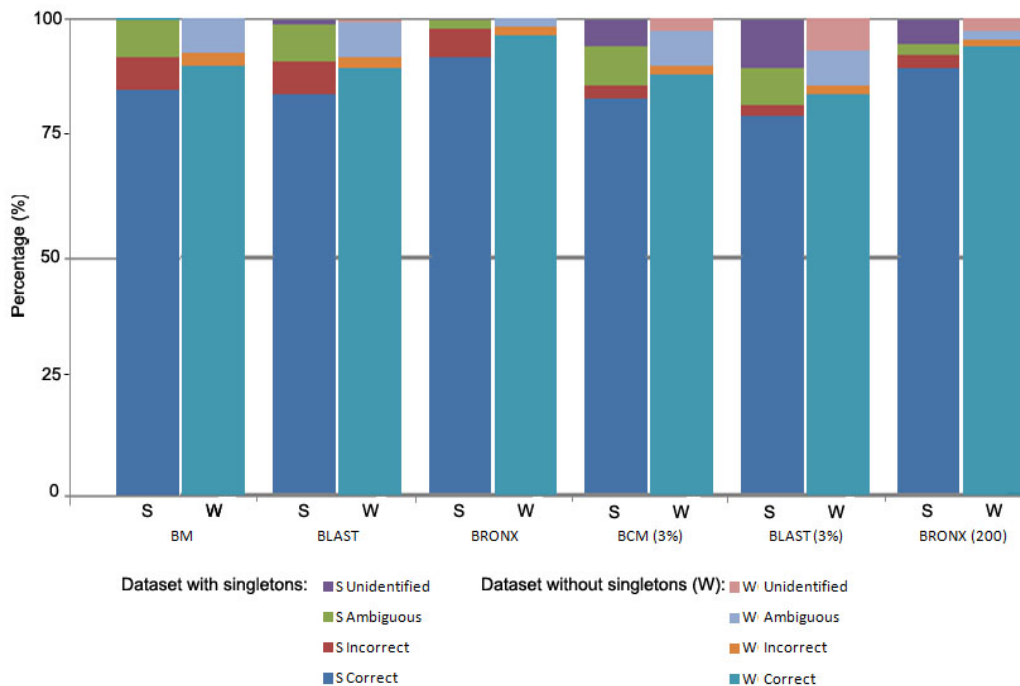
Figure 3.4.3.1: Identification efficiency for the different methods of assigning species to specimen

3.3.4. Comparison of the different methods of analysis

BRONX analyses outperform BLAST and BM/BCM analysis, exhibiting identification success rates of 89.81-96.46% and an identification success of 92% is attainable even in a dataset with singletons. Notably, BRONX identifies greater number of sequences than the distance based approaches by minimizing ambiguous identifications. Misidentification rates are similar for all analyses with or without threshold regardless of whether a dataset with or without singletons is analysed. BLAST without threshold and BCM exhibit similar identification efficiency in terms of identification success, misidentification and ambiguous identification rates. Upon including a threshold, the identification success rate of BCM is higher than for BLAST for both datasets (with or without singletons). However, BLAST was able to prevent more misidentifications than BCM.

3.3.5. Consensus barcodes

Species barcodes *sensu stricto* constructed as an union-based consensus of all conspecific barcodes revealed that four species are involved in two cases of two species sharing identical barcodes (between *Oreochromis mossambicus* and *O. niloticus* and between *Pterygoplichthys joselimaianus* and *P. pardalis*). This implies that 4.5% of the species lack unique barcodes. The result indicates that the

remaining 70 species contain diagnostic characters in their COI barcode.

3.3.6. Species coverage

383 COI sequences were obtained for 89 species of freshwater fish in Singapore, of these 75 were from Genbank for 22 species. 18 of the 22 species are only represented by COI sequences from Genbank (68 sequences). I sequenced COI for 318 individuals from 71 species. Only tissues from two species could not be sequenced (*Rhinogobius giurinus*, *Synodontis eupterus*). In both cases, I suspect that primer specificity could be the main issue but it is also noted that specimens for *R. giurinus* were kept in 70% ethanol for several years before DNA extraction was attempted. Such storage conditions are not conducive for DNA extraction.

DNA extraction and amplification was unsuccessful for the following formalin-preserved specimens; *Eugnathogobius siamensis, Mystus castaneus, Satanoperca jurupari*. I could not obtain collecting permits for obtaining specimen for *Arapaima gigas* from the trade and unfortunately there is no Genbank sequence for COI for this species. In general, I could not obtain any specimens for the following species due to their rarity in the wild and/or the ornamental fish trade: *Glyptothorax major, Nemacheilus selangoricus, Neostethus lankesteri, Parakysis verrucosus, Puntius johorensis, Puntius partipentazona, Silurichthys hasseltii, Clarias teijismanni,* and *Nandus*

nebulosa. The overall species coverage of my study is thus 83% inclusive of Genbank sequences. If I only count species that were collected from museum, ornamental fish trade and habitat, the actual species coverage would be much lower (65%) thus illustrating the problems with getting good species coverage even for a small country.

3.4. Discussion

The native species coverage obtained is 87.5% (28/32) while the exotic species coverage is 80.2% (61/76) thus yielding an overall species coverage of 83% inclusive of Genbank sequences for 18 species. My species coverage is thus higher than in a study conducted at a similar scale by Page and Hughes (2010) on freshwater water fish of Queensland (Australia), whereby 22 of the 28 native fish species were obtained. This together with my own example suggests that obtaining full species coverage challenging even at a regional scale. One would expect that obtaining complete COI coverage for all freshwater fish on a small island would be straightforward. However, without Genbank data, I would have only achieved 65% coverage and even with the aid of Genbank data, the overall species coverage increased only to 83%. I lack barcodes for 27 species but only three species are native (*Eugnathogobius siamensis*, *Clarias teijismanni, Nandus nebulosa*). All three are rare and *Eugnathogobius siamensis* was thought to be locally extinct until a few specimens were collected during a recent reservoir survey. Unfortunately, they were stored in formalin so that the DNA could not be successfully amplified.

Overall, it is also difficult to obtain specimens for some exotic species because many were not seen in the trade over the three years of collection for this project. Obviously, the ornamental trade includes numerous species and the species change over time. This means that

a species that may have been introduced into a Singapore reservoir 10 years ago may no longer available in the trade. This does not apply to species with a steady trade such as many cichlids and *Acaricthys heckelii*. It is noteworthy that the main reason why it was easier to get samples for indigenous species is due to collecting activity of the RMBR. my species coverage would have been much worse if I had to rely on freshly collected specimens for this project.

Appropriate specimen storage is an important issue. In my case, many specimens were preserved in formalin without first taking tissue samples for preservation in ethanol. It is well known that formalin affects DNA preservation (Diaz-Viloria et al., 2005) and this generates particularly serious problems for fish, because until recently it was standard practice to preserve specimens in formalin for a week before changing to ethanol. This fixation procedure, while important for preserving morphological traits, generates cross-links between DNA and histone proteins so that only very short DNA fragments can be obtained during a DNA extraction (Skage & Schander, 2007; Zhang, 2010).

Overall, I find that identification success rates are high ranging from 79-97% depending on the stringency of the identification criteria. 79% percent is the success rate of a BLAST analysis with a cutoff point of 3% with the full dataset, while 97% is obtained when BRONX is used for a dataset from which all singleton sequences have been removed.

The highest success rate observed in my study (97%) is within range of the ones published in the fish barcoding literature. For example, Ward et al. (2005) concluded that 98% of the 270 mostly Australian marine species of fish, which include commercially important groups such as *Thunnus* (tunas), *Platycephalus* (flatheads) and *Squalus* (dogfish or spurdogs), could be identified based on COI. Subsequently, similar studies have been carried out on fish species in other geographical regions such as Canada (Hubert et al. 2008), deep water sharks from the north-eastern Atlantic (Moura et al. 2008), widespread species with populations between the north Atlantic and Australasia (Ward et al. 2008), North Pacific skates from Alaska (Spies et al. 2006), *Sardinella tawilis* (a Philippine endemic freshwater Sardine) could be differentiated from its marine relatives (Quilang et al. 2011), and Mexico and Guatamala (Valdez-Moreno et al. 2010). Based on these studies, success rates ranged from 91- 97% and it appeared that most species of fish have small intraspecific and high interspecific diversity; i.e. they are good candidates for identification through DNA barcodes.

Comparison between the different methods of analysis shows that BRONX (Little, 2011) is capable of improving identification efficiency when assigning species to specimen for both datasets with and without singletons. These results are similar those found by Little (2011) where BRONX outperformed distance based approaches to DNA barcoding. I observed that BRONX minimized ambiguous identifications. For example, several sequences of *Oreochromis nilloticus* and *O.*

mossambicus were ambiguously identified or misidentified using distance based approaches, while BRONX was able to resolve 3/4 of these identifications. Thus it is able to discriminate closely related sequences. BRONX analyses are additionally able to analyse large datasets which require large amounts of computational power and time if a global alignment has to be prepared before running a BM and BMC analysis. In my study, BRONX yields promising result when analysing the small Singapore fish dataset and identification success rates from 89-97%. In the next chapter of the thesis (Chapter IV), BRONX is used to analyze a much larger dataset comprising of fish from the ornamental trade and Genbank to determine identification efficiency for determining identification success at a global scale.

I find a number of misidentification cases. Closely related species with similar barcodes have been discussed in the literature; they could not be readily identified by their COI sequences. For example, 7% of the barcoded Canadian fish species belong to this category. They include the lampreys *Ichthyomyzon fossor* and *I. unicuspis*, shiners *Notropis volucellus* and *N. buchanani*, the shad *Alosa aestivalis* and *A. pseudoharengus*, putative species in the cisco species flock (*Coregonus artedi, C. hoyi, C. kiyi, C. nigripinnis* and *C. zenithicus*), and darters (*Etheostoma nigrum* and *E. olmstedi*) (Hubert et al. 2008). In the Scotia Sea sample, both COI and *cytb* dataset lacked sufficient sensitivity for resolving species within the *Bathydraco* and *Artedidraco* (11 of the 35 species tested) and the identification success rate was

only 68% (Rock et al. 2008). In the study conducted by Quilang et al. (2011), Genbank COI sequences of *Sardinella atricauda* and *S. melanura* were identical. Similarly, I can also find species that cannot be identified by their COI in my dataset. In my study, misidentifications and ambiguities were caused by the sharing of highly similar or identical barcodes between some individuals of 8 species: *Cichlasoma uropthalmus* and *Amphilopus citrinellus*, between some individuals of *Oreochromis niloticus* and *Oreochromis mossambicus*, between *Puntius banksi* and *Puntius binotatus*, and between *Pterygoplichthys pardalis* and *Pterygoplichthys joselimaianus*. Four of these species shared identical barcodes and contributed the largest number of misidentifications (between *Oreochromis niloticus* and *Oreochromis mossambicus,* and between *Pterygoplichthys pardalis* and *Pterygoplichthys joselimaianus*). For most of the specimens, I re-identified the vouchers so that the barcode sharing was not due to misidentification (except for species with Genbank sequences: *Pterygoplichthys pardalis*, *P. joselimaianus*, and some *Oreochromis niloticus*). The sharing of COI barcode by cichlid species was not surprising, because it had been documented by many studies that many species of cichlids exhibit low interspecific variation in their DNA sequences (Shirak et al., 2009; Valdez-Moreno & Ivanova, 2009). As for *Puntius banksi* and *P. binotatus*, some taxonomists had considered them distinct species based on slight morphological differences (*P. banksi* has a dark wedge-shaped marking while *P. binotatus* has a round spot, both below the dorsal fin) (Kottelat & Lim 1995) while

others have posited that that *P. banksi* and *P. binotatus* represent two extreme colour forms of a single species (Ng & Tan 1999). One could argue that this position is consistent with barcode sharing, but more evidence is needed before this question can be decided. Both *P. banksi* and *P. binotatus* can be found schooling together in the local stream at Venus Drive in Singapore, and this creates opportunities for studying whether interbreeding occurs under natural conditions.

Here, I determine that the best threshold for a BCM analysis is 2.3% to 2.8% which is similar to what has been suggested by Hebert's 2003 study on genetic distances from a wide variety of taxonomic groups (e.g., Mammalia, Cnideria, Arthropoda), albeit with poor species coverage. He predicted a 3% distance threshold for separating species. However, this threshold was subsequently lowered to 1% without discussing the reasons. The lower threshold increases the stringency of query identification, but it is inappropriate for species with large sequence variability and will not prevent misidentification for species with very similar barcodes (e.g., Cichlidae). In this study, the optimal thresholds and cut off points for BRONX and BCM are determined empirically, but this cannot mask that thresholds are problematic because there is no biological reason to expect the same threshold to apply to many species. If identification is particularly critical in a real-time application, I recommend using the more stringent threshold (300 to 400 for BRONX, 1% for BCM and BLAST).

78

Most of the species in my data set have species-specific consensus barcodes which are constructed as union-based consensus sequences of all conspecific barcodes. The only exceptions are four species that are involved in two cases of two species sharing identical barcodes (between *Oreochromis mossambicus* and *O. niloticus*, and between *Pterygoplichthys joselimaianus* and *P. pardalis*). This result is in contrast to the popular belief by opponents of DNA barcoding that COI lacks diagnostic characters. However, my test here is not very rigorous because few genera are represented by more than one species and many species have only one sequence. I would predict that species barcodes will become less diagnostic as more individuals and species are sampled.

3.5. Conclusion

I do not obtain a 100% success rate in allocating specimen to species even for the Singapore dataset with its low species to genus ratio (1.3). However, identification success rates of 85-97% in most of my empirical analyses indicate that the COI fragment is fairly effective for identifying Singapore's freshwater fishes. The alignment free analysis BRONX is shown to improve the identification efficiency of assigning species to specimen. The best results are obtained with a 2.3 to 2.8% threshold for BCM and 200 for BRONX which should be used for future query sequences. Only identifications in some genera should be

interpreted with care because they have high intra- or low interspecific sequence diversity. These genera and species are identified in my analysis. However, not all species could be included because it very difficult to comprehensively sample even a small fish fauna of a small island such as Singapore. This means that future studies will generate queries that cannot be matched based on my dataset.

3.6   References for Chapter 3

Altschul, S.F., Gish, S.F. & Miller, W. (1990). Basic local alignment search tool. *Journal of Molecular Biology 215(3): 403-410.*

Asgharian, H., H. H. Sahafi, et al. (2011). Cytochrome c oxidase subunit 1 barcode data of fish of the Nayband National Park in the Persian Gulf and analysis using meta-data flag several cryptic species. *Molecular Ecology Resources 11(3): 461-472.*

Baker, N. & Lim, K. K. P. (2008). Wild Animals of Singapore. *Draco Publishing and Distribution Pte. Ltd. and Nature Society (Singapore).*

Bergsten, J., D. T. Bilton, et al. (2012). The Effect of Geographical Scale of Sampling on DNA Barcoding. *Systematic Biology 61(5): 851-869.*

Bitanyi, S., G. Bjornstad, et al. (2011). Species identification of Tanzanian antelopes using DNA barcoding. *Molecular Ecology Resources 11(3): 442-449.*

Bleeker, W., Klausmeyer, S., Peintinger, M., & Dienst, M. (2008). DNA sequences identify invasive alien Cardamine at Lake Constance. *Biological Conservation, 141*(3), 692-698. doi: 10.1016/j.biocon.2007.12.015

Chown, S. L., Sinclair, B. J., & van Vuuren, B. J. (2008). DNA barcoding and the documentation of alien species establishment on sub-Antarctic Marion Island. *Polar Biology, 31*(5), 651-655. doi: 10.1007/s00300-007-0402-z

Clare, E. L., B. K. Lim, et al. (2011). Neotropical Bats: Estimating Species Diversity with DNA Barcodes. *Plos One 6(7).*

Costa, F. O., M. Landi, et al. (2012). A Ranking System for Reference Libraries of DNA Barcodes: Application to Marine Fish Species from Portugal. *Plos One 7(4).*

Dai, L. N., X. D. Zheng, et al. (2012). DNA barcoding analysis of Coleoidea (Mollusca: Cephalopoda) from Chinese waters. *Molecular Ecology Resources 12(3): 437-447.*

De Vere, N., T. C. G. Rich, et al. (2012). DNA Barcoding the Native Flowering Plants and Conifers of Wales. *Plos One 7(6).*

Diaz-Viloria, N., Sanchez-Velasco, L., & Perez-Enriquez, R. (2005). Inhibition of DNA amplification in marine fish larvae preserved in formalin. *Journal of Plankton Research, 27*(8), 787-792. doi: 10.1093/plankt/fbi052

Dinca, V., E. V. Zakharov, et al. (2011). Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. *Proceedings of the Royal Society B-Biological Sciences 278(1704): 347-355.*

Funk, V. A., Richardson, K. S., & Ferrier, S. (2005). Survey-gap analysis in expeditionary research: where do we go from here? *Biological Journal of the Linnean Society, 85*(4), 549-567. doi: 10.1111/j.1095-8312.2005.00520.x

Gattolliat, J. L., L. Vuataz, et al. (2012). First contribution to the mayflies of Jordan. *Zoology in the Middle East 56: 91-110.*

Gompert, Z., C. C. Nice, et al. (2006). Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Molecular Ecology 15(7): 1759-1768.*

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences, 270(1512), 313-321. doi: 10.1098/rspb.2002.2218*

Hausmann, A., G. Haszprunar, et al. (2011). Now DNA-barcoded: the butterflies and larger moths of Germany. *Spixiana 34(1): 47-58.*

Hendrich, L., Pons, J., Ribera, I., & Balke, M. (2010). Mitochondrial Cox1 Sequence Data Reliably Uncover Patterns of Insect Diversity But Suffer from High Lineage-Idiosyncratic Error Rates. *PLoS ONE, 5*(12). doi: 10.1371/journal.pone.0014448

Holmes, B. H., Steinke, D. & Ward, R. D. (2009). Identification of shark and ray fins using DNA barcoding. *Fisheries Research, 95*(2-3), 280-288. doi: 10.1016/j.fishres.2008.09.036

Hubert, N., Hanner, R., Holm, E., Mandrak, N. E., Taylor, E., Burridge, M. & Bernatchez, L. (2008). Identifying Canadian Freshwater Fishes through DNA Barcodes. *Plos One, 3*(6). doi: 10.1371/journal.pone.0002490

Jerde, C. L., A. R. Mahon, et al. (2011). Sight-unseen detection of rare aquatic species using environmental DNA. *Conservation Letters 4(2): 150-157.*

Kottelat, M. (1995). 4 New species of fishes from the middle Kapaus basin, Indonesia Borneo (Osteichthyes, Cyprinidae and Belontiidae). *Raffles Bulletin of Zoology, 43*(1), 51-64.

Kruskop, S. V., A. V. Borisenko, et al. (2012). Genetic diversity of northeastern Palaearctic bats as revealed by DNA barcodes. *Acta Chiropterologica 14(1): 1-14.*

Kumar, N. P., R. Srinivasan, et al. (2012). DNA barcoding for identification of sand flies (Diptera: Psychodidae) in India. *Molecular Ecology Resources 12(3): 414-420.*

Kwong, S., Srivathsan, A. & Meier, R. (2012). An update on DNA barcoding: low species coverage and numerous unidentified sequences. *Cladistics, 28*(6), 639-644. doi: 10.1111/j.1096-0031.2012.00408.x

Lim, G. S., Balke, M., & Meier, R. (2012). Determining Species Boundaries in a World Full of Rarity: Singletons, Species Delimitation Methods. *Systematic Biology, 61*(1), 165-169. doi: 10.1093/sysbio/syr030

Lakra, W. S., M. S. Verma, et al. (2011). DNA barcoding Indian marine fishes. *Molecular Ecology Resources 11(1): 60-71.*

Lim, K. K. P. & Ng P. K. L. (2012). A guide to the freshwater fishes of Singapore.

Lim, G.S., Balke, M. & Meier R. (2011). Determining species boundaries in a world full of rarity: singletons, species delimitation methods. *Systematic Biology 61(1): 165-169.*

Little, D.P. & Stevenson D.W. (2007). A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. *Cladistics 23(1): 1-21.*

Little, D. P. (2011). DNA Barcode Sequence Identification Incorporating Taxonomic Hierarchy and within Taxon Variability. *Plos One, 6*(8). doi: 10.1371/journal.pone.0020552

Martin, W. & Fiedler, K. (2007). Does the DNA barcoding gap exist? A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology 4(8): 1 - 8.*

Meier, R., Kwong, S., Vaidya, G. & Ng, P. K. L. (2004). An empirical test of DNA taxonomy in Diptera based on cox-1 sequences. *Cladistics, 20*(6), 600-600.

Meier, R., Shiyang, K., Vaidya, G. & Ng, P. K. L. (2006). DNA barcoding and taxonomy in diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology, 55*(5), 715-728. doi: 10.1080/10635150600969864

Meyer, C.P. & Paulay G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLOS Biology 3(12): e422.*

Minamoto, T., H. Yamanaka, et al. (2012). Surveillance of fish species composition using environmental DNA. *Limnology 13(2): 193-197.*

Moura, T., Silva, M. C., Figueiredo, I., Neves, A., Muñoz, P. D., Coelho, M. M., & Gordo, L. S. (2008) Molecular barcoding of north-east Atlantic deep-water sharks: species identification and application to fisheries management and conservation. *Marine and Freshwater Research 59, 214–223.*

Ng, H. H. & Tan, H. H. (2010). An annotated checklist of the non-native Freshwater fish species in the reservoirs of Singapore. *COSMOS 6(1): 95-116.*

Page, T. J. & Hughes, J. M. (2010). Comparing the performance of multiple mitochondrial genes in the analysis of Australian freshwater fishes. *Journal of Fish Biology 77(9): 2093-2122.*

Park, H. Y., H. S. Yoo, et al. (2011). New DNA barcodes for identification of Korean birds. *Genes & Genomics 33(2): 91-95.*

Quilang, J. P., Santos, B. S., Ong, P. S., Basiao, Z. U., Fontanilla, I. K. C., Cao & E. P. (2011) DNA Barcoding of the Philippine Endemic Freshwater Sardine *Sardinella tawilis* (Clupeiformes: Clupeidae) and Its Marine Relatives *The Philippine Agricultural Scientist 94(3): 10-15.*

Shirak, A., Cohen-Zinder, M., Barroso, R. M., Seroussi, E., Ron, M., & Hulata, G. (2009). DNA Barcoding of Israeli Indigenous and Introduced Cichlids. *Israeli Journal of Aquaculture-Bamidgeh, 61*(2), 83-88.

Skage, M., & Schander, C. (2007). DNA from formalin-fixed tissue: extraction or repair? That is the question. *Marine Biology Research, 3*(5), 289-295. doi: 10.1080/17451000701473942

Sonet, G., F. C. Breman, et al. (2011). Applicability of DNA barcoding to museum specimens of birds from the Deomcratic Republic of the Congo. *Bonner Zoologische Monographien (57): 117-131.*

Spies, I. B., Gaichas, S., Stevenson, D. E., Orr, J. W. & Canino, M. F. (2006) DNA-based identification of Alaska skates (Amblyraja, Bathyraja and Raja: Rajidae) using cytochrome c oxidase subunit I (COI) variation *Journal of Fish Biology 69: 283-292.*

Srivathsan, A. & Meier R. (2012). On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics 28(2): 190-194.*

Takahara, T., T. Minamoto, et al. (2012). Estimation of Fish Biomass Using Environmental DNA. *Plos One 7(4).*

Tavares, E. S., P. Goncalves, et al. (2011). DNA Barcode Detects High Genetic Structure within Neotropical Bird Species. *Plos One 6(12).*

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matric choice *Nucleic Acids Research 22(22): 4673-4680.*

Thomson, P. F., J. Kielgast, et al. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology 21(11): 2565-2573.*

Valdez-Moreno, M., N. V. Ivanova, et al. (2009). Probing diversity in freshwater fishes from Mexico and Guatemala with DNA barcodes. *Journal of Fish Biology 74(2): 377-402.*

Valdez-Moreno, M., Vasquez-Yeomans, L., Elias-Gutierrez, M., Ivanova, N. V. & Hebert, P. D. N. (2010). Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. *Marine and Freshwater Research, 61*(6), 665-671. doi: 10.1071/mf09222

Ward, R. D., Zemlak T. S., Bronwyn H. I., Last, P.R. & Hebert, P.D.N. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B-Biological Sciences, 1716,* 11. doi: 10.1098

Webb, J. M., L. M. Jacobus, et al. (2012). A DNA Barcode Library for North American Ephemeroptera: Progress and Prospects. *Plos One 7(5).*

Yeo, D. C. J. & Chia C. S. W. (2010). Introduced species in Singapore: An Overview. *COSMOS 6(1): 23-37.*

Zhang, J. (2010). Exploiting formalin-preserved fish specimens for resources of DNA barcoding. *Molecular Ecology Resources, 10*(6), 935-941. doi: 10.1111/j.1755-0998.2010.02838.x

Zimmermann, J., Hajibabaei, M., Blackburn, D. C., Hanken, J., Cantin, E., Posfai, J., & Evans, T. C. (2008). DNA damage in preserved specimens and tissue samples: a molecular assessment. *Frontiers in Zoology, 5*. doi: 10.1186/1742-9994-5-18

Appendix II: Species list of freshwater fish in Singapore water systems

No.	Species Names	Native or non-native	No. of specimens	Source/s
1	*Acarichthys heckelii*	non-native	10	*Reservoir survey*
2	*Amphilopus citrinellus*	non-native	8	*Reservoir survey*
3	*Anabus testudineus*	non-native	3	*Reservoir survey*
4	*Aplocheilus panchax*	Native	3	*Reservoir survey& Trade*
5	*Astronotus ocellatus*	non-native	3	Genbank
6	*Atractosteus spatula*	non-native	2	*RMBR*
7	*Barbonymus altus*	non-native	1	Genbank
8	*Barbonymus schwanenfeldii*	non-native	3	*RMBR*
9	*Betta pugnax*	Native	5	*RMBR*
10	*Boraras maculatus*	Native	4	*Trade*
11	*Bostrychus sinensis*	Native	1	Genbank
12	*Carassius auratus*	non-native	6	*Trade & RMBR*
13	*Channa gachua*	Native	4	*Trade*
14	*Channa lucius*	Native	2	*Reservoir survey*
15	*Channa melasoma*	Native	1	*Trade & RMBR*
16	*Channa micropeltes*	non-native	5	*Trade*
17	*Channa striata*	Native	3	*RMBR*
18	*Chitala ornata*	non-native	3	*RMBR*
19	*Chromobotia macracantha*	non-native	6	*Trade*
20	*Cichla orinocensis*	non-native	3	*Reservoir survey*
21	*Cichla temensis*	non-native	1	*Reservoir survey*
22	*Cichlasoma urophthalmus*	non-native	3	*Reservoir survey*
23	*Clarias batrachus*	Native	14	*Trade*
24	*Clarias gariepinus*	non-native	5	Genbank
25	*Ctenopharyngodon idella*	non-native	2	Genbank
26	*Cyclocheilichthys apogon*	Native	2	*Trade*
27	*Cyprinus carpio*	non-native	6	*Trade*
28	*Datnioides microlepis*	non-native	7	*Trade*
29	*Dermogeny collettei*	Native	11	*Trade*
30	*Esomus metallicus*	non-native	2	*Trade*
31	*Gambusia affinis*	non-native	6	Genbank
32	*Geophagus altifrons*	non-native	4	*Reservoir survey*
33	*Glossogobius aureus*	Native	13	Genbank
34	*Glossogobius giuris*	Native	1	Genbank
35	*Gobiopterus brachypterus*	Native	2	Genbank
36	*Hampala macrolepidota*	non-native	4	*RMBR*

No.	Species Names	Native or non-native	No. of specimens	Source
37	*Hemibagrus nemurus*	Native	1	Genbank
38	*Hemirhamphodon pogonognathus*	Native	3	Genbank
39	*Heros severus*	non-native	3	*Trade*
40	*Hypophthalmichthys nobilis*	non-native	6	Genbank
41	*Labeo rohita*	non-native	5	Genbank
42	*Leptobarbus hoevenii*	non-native	3	*Trade*
43	*Luciocephalus pulcher*	Native	4	*Trade & RMBR*
44	*Macrognathus zebrinus*	non-native	2	*Trade*
45	*Megalops cyprinoides*	Native	2	*RMBR*
46	*Monopterus albus*	Native	1	*RMBR*
47	*Nomorhamphus liemi*	non-native	1	*RMBR*
48	*Notopterus notopterus*	non-native	4	*RMBR*
49	*Oreochromis mossambicus*	non-native	5	*Genbank*
50	*Oreochromis niloticus*	non-native	28	Genbank
51	*Osphronemus goramy*	non-native	1	*Trade*
52	*Osteochilus hasseltii*	non-native	6	*Trade*
53	*Osteoglossum bicirrhis*	non-native	3	*Trade & RMBR*
54	*Oxyeleotris marmorata*	non-native	3	*Rservoir survey*
55	*Pangasianodon hypophthalmus*	non-native	6	*Trade*
56	*Pangio muraeniformis*	Native	12	*Reservoir survey*
57	*Pangio semicincta*	Native	4	*Reservoir survey*
58	*Parachromis managuensis*	non-native	7	*Reservoir survey*
59	*Parambassis siamensis*	non-native	5	*Reservoir survey*
60	*Phractocephalus hemioliopterus*	non-native	3	*Trade*
61	*Piaractus brachypomus*	non-native	4	*RMBR*
62	*Poecilia reticulata*	non-native	3	*RMBR*
63	*Poecilia sphenops*	non-native	1	*RMBR*
64	*Potamotrygon motoro*	non-native	2	*RMBR*
65	*Pterygoplichthys disjunctivus*	non-native	2	Genbank
66	*Pterygoplichthys joselimaianus*	non-native	1	Genbank
67	*Pterygoplichthys pardalis*	non-native	4	Genbank
68	*Puntius banksi*	Native	4	*RMBR*
69	*Puntius binotatus*	non-native	1	*RMBR*
70	*Puntius hexazona*	Native	4	*RMBR*
71	*Puntius lateristriga*	non-native	2	*RMBR*
72	*Puntius semifasciatus*	non-native	2	*RMBR*

No.	Species Names	Native or non-native	No. of specimens	Source
73	Puntius tetrazona	non-native	4	RMBR
74	Rasbora bankanensis	non-native	4	Trade
75	Rasbora einthovenii	Native	1	Genbank
76	Rasbora elegans	Native	5	RMBR
77	Scleropages formosus	non-native	3	RMBR
78	Stigmatogobius sadanundio	non-native	2	Trade
79	Tilapia buttikoferi	non-native	2	Reservoir survey
80	Tor tambra	non-native	3	RMBR
81	Toxotes chatareus	Native	7	Trade
82	Toxotes jaculatrix	Native	3	Trade
83	Trichogaster leeri	non-native	5	Reservoir survey
84	Trichopodus trichopterus	Native	20	Reservoir survey
85	Trichopsis schalleri	non-native	1	RMBR
86	Trichopsis vittata	Native	2	Reservoir survey& Trade
87	Trigonostigma heteromorpha	Native	6	Reservoir survey & Trade
88	Vieja synspila	non-native	1	Trade
89	Xiphophorus helleri	non-native	1	RMBR
90	Arapaima gigas	Non-native	0	n/a
91	Brachygobius kabiliensis	native	0	n/a
92	Clarias leicanthus	non-native	0	n/a
93	Esomus metallicus	non-native	0	n/a
94	Eugnathogobius siamensis	native	0	n/a
95	Glyptothorax major	non-native	0	n/a
96	Leptobarbus rubripinna	non-native	0	n/a
97	Mystus castaneus	non-native	0	n/a
98	Mystus wolffii	non-native	0	n/a
99	Nandus nebulosa	non-native	0	n/a
100	Nemacheilus selangoricus	non-native	0	n/a
101	Neostethus lankesteri	non-native	0	n/a
102	Parakysis verrucosus	non-native	0	n/a
103	Puntius johorensis	non-native	0	n/a
104	Puntius partipentazona	non-native	0	n/a
105	Rhinogobius giurinus	non-native	0	n/a
106	Satanoperca jurupari	non-native	0	n/a
107	Silurichthys hasseltii	non-native	0	n/a
108	Etroplus suratensis	non-native	0	n/a

CHAPTER IV

_____

Towards a Comprehensive DNA Barcode

Database for Freshwater Aquarium Fish: A

Pragmatic Approach to Increasing Species

Coverage

Abstract

The main problems with contemporary DNA barcode databases are poor species coverage, slow growth, insufficient intraspecific sampling, the inclusion of sequences from partially identified specimens, and the lack of readily accessible vouchers. In particular, species coverage is critical for successful species identification with DNA barcodes. This also applies to the identification of fish species that are in the ornamental trade. Here, I have provided 1,450 COI DNA barcodes for 522 freshwater fish species that were obtained in Singapore's ornamental fish trade. Of these, 334 species previously lacked DNA barcodes in Genbank. All specimens that were sequenced were not only vouchered, but also documented with high-resolution photographs that are made available online. When testing the ability of DNA barcodes to identify species, I find success rates that range from 77% to 91% depending on choice of database and stringency of identification criteria. Despite generating barcodes for many additional species, I find that only 1,225 species of the 4,769 freshwater ornamental fish recorded by Ornamental Fish International OFI (2010) have COI barcodes in Genbank, while the remaining ca. 3500 species have yet to be barcoded. Barcoding fish species from the ornamental trade has the downside of imprecise locality information, but I argue that it should nevertheless be pursued further because obtaining species from the trade is faster than field colleting and have a higher chance of becoming invasive. Given the quick rise in the use of eDNA

from water for species identification, a more complete database of potentially invasive species is urgently needed.

4.1. Introduction

The use of DNA sequences for identifying specimens has a long history, but it was only formalized as "DNA barcoding" by Hebert et al. (2003) who proposed the use of a 650bp fragment of Cytochrome Oxidase I (COI) for identifying all animal species on earth (Hebert et al. 2003). This led to a decade of building COI databases and methodological discussions and debates on utility of DNA barcoding for identification as well as species delimitation (Will and Rubinoff 2004, Meier et al. 2006, Ratnasingham and Hebert 2007). It is now well understood that many species can be identified by their COI sequences (Ward et al. 2005, Rudnick, Katzner et al. 2007, Lohman and Samarita 2009, Hausmann et al. 2011), while other species – often concentrated in particular taxa – are prone to low inter- and high intraspecific sequence variability (Meier et al. 2006, Huang et al. 2008). Overall, the initial claim of "one species, one barcode" is now known to be unrealistic. Nevertheless, a substantial number of species can be identified using COI and therefore numerous applications of DNA barcoding have been proposed and applied on taxa such as fish, birds, and butterflies, where identification success rates based on COI can be high as long as some clades are avoided (e.g., cichlid radiations). Indeed, DNA barcodes have become invaluable for many purposes. Good examples are the matching of juveniles with adults (Robertson et al. 2007, Victor et al. 2009, Valdez-Moreno et al. 2010), identifying the origin of food ingredients such as the fish species used in fish fillets

(Wong and Hanner 2008, Yancy et al. 2008, Barbuto et al. 2010), and using DNA sequences for monitoring the movement of invasive and endangered animals in the international wild life trade (Bleeker et al. 2008, Chown et al. 2008, Saunders 2009).

Fundamental to any of these applications is the availability of DNA barcode databases with sequences from reliably identified specimens. Examples of DNA sequence repositories include Genbank and BOLD with the latter being the official repository of COI sequences from many barcoding projects. However, while sequence quality, quantity and accessibility of these databases are important for the accurate identification of specimen, many databases suffer currently from poor species coverage, sequences that are only identified to genus or family, and lack of ready access to voucher specimens (Kwong et al. 2012). In particular, the species coverage is poor considering that only 60,000 Metazoa species have been barcoded.

The main challenge of the DNA barcoding campaign is providing DNA barcodes for all species given that 1.8 million species have been described, 5-10 million species are estimated to exist (Camilo et al., 2011; Costello, May, & Stork, 2013), and many of these species have only been collected once (Lim, Balke & Meier, 2011). These problems were recognized early-on and therefore the International Barcode of Life initiative decided to target certain groups of Metazoa for their first barcoding campaigns; these included "Fish Bol" (Ward, Hanner, &

Hebert, 2009), the "All Birds Barcoding Initiative (ABBI)" (Tavares & Baker, 2008), and "Lepidoptera Barcode of Life (Lepbarcoding)" (Hebert et al., 2004). Despite 10 years of barcoding and choosing these less challenging targets, most species in these groups still lack barcodes. Overall, fewer than 14% of the species in the campaign taxa have been barcoded and made available on public databases. This greatly limits the utility of these databases for precise and accurate identification of unknown specimens. For example, the number of species of COI barcode available for one of the target organism ("fish") is estimated to be 10,620 and 8,035, in Fish Bol and Genbank respectively. This constitutes less than a third of the ca. 32,700 described fish species. One of the main challenges or Fish Bol is to find faster ways to grow the barcode database.

The aquarium trade is an important source of invasive fish and several authors have proposed the use of DNA barcoding to monitor and regulate the trade including the movement of invasive species for the purpose of protecting the native biodiversity and commercially important fisheries (Collins et al. 2012, Cote et al. 2013). However, the extent to which the currently available fish COI databases are useful for this application is currently poorly understood. Here, I investigate the COI coverage for the 4,769 species of ornamental fish in the Ornamental Fish International database (OFI, 2010) in Genbank, in order to determine the usefulness of the public database for bioprotection and bio-monitoring application.

A high quality barcode database should have good taxonomic coverage and every barcode should be associated with a species name and a voucher specimen. A voucher allows for the revision of the taxonomic information associated with a sequence in the future. This is frequently needed in case there are re-descriptions of closely related species or the DNA sequence implies misidentification. Good voucher information is nevertheless still rare. In addition, it is not uncommon to encounter COI barcodes that are only associated with partial taxonomic information. In an extreme case, some barcodes have been submitted to Genbank that were only identified to order ("Diptera sp.")(Kwong et al. 2012). While such extreme cases are not found for fish, it is still common to find barcodes for specimens identified only to genus. This is particularly problematic if high-quality vouchers are not available because it makes it unlikely that the identification level will ever be improved. While this problem has long been recognized, voucher documentation is still underappreciated. For example, in a recent publication "Barcoding Nemo", voucher images are of poor quality because they were obtained using desktop scanner (Steinke et al. 2009). Indeed, many of the voucher images in BOLD (Ratnasingham & Hebert, 2007) suffer from such problems and they are usually insufficient for determining species. In this study, I firstly contribute barcode sequences for 522 species in order to work toward the goal of obtaining DNA barcodes for all ornamental fish. Many of the here contributed species (334 spp) are new to Genbank. Secondly, I provide online access to high resolution voucher photographs of my

specimens. Along with building the database, I assess the efficiency with which DNA barcodes can be used to identify fish species in the ornamental trade. I test whether the species in this database can be distinguished from each other but also whether they can be distinguished from all other species of fish in Genbank. As identification criteria, I employ multiple techniques including two that are distance-based, one that is based on BLAST, and one that is based on diagnostic markers.

The ability to detect fish COI in water system is as essential as a DNA barcode library for monitoring introduced species. The field of eDNA had always been dominated by studies with emphasis to detect microorganisms from the environment as a form of biosecurity measure to ensure water safety. Recently, there had been growing interest in environmental DNA (eDNA) as surveillance tools for identifying the presence of targeted animals (macroorganism) via DNA sequences (Collins et al., 2013; Goldberg et al., 2011; Jerde et al., 2013; Minamoto et al., 2012; Takahara et al., 2013; Thomsen et al., 2012). However, the extent of this growth is undocumented. Hence, I will also be documenting the publication trend of eDNA related studies for the pass decades to reveal insights into the future of eDNA for detecting invasive species.

4.2. Materials and Methods

4.2.1. Determining the species coverage of aquarium fish COI in Genbank

In order to obtain a species list for all fish with COI barcodes in GenBank, I followed Kwong et al. (2012) by carrying out a "taxonomy" search in NCBI for "Chondrichthyes, Actinopterygii, and Hyperoartia" and adding the search term [COI(Gene Name) OR "cytochrome oxidase subunit 1"(Gene Name) OR "cytochrome c oxidase subunit 1"(Gene Name) OR "cytochrome c oxidase subunit I"(Gene Name) OR "cytochrome oxidase subunit I"(Gene Name) OR COX1(Gene Name)] in NCBI (http://www.ncbi.nlm.nih.gov/). All sequences found with this strategy were downloaded in FASTA format. I then used the approach developed by Hunt et al. (2007) for extracting sequences with only the barcode region of the COI. This pipeline used a curated subset of DNA barcodes that were matched against the downloaded sequences and only extracts the barcode region of the downloaded sequences. For this subset the sequences were translated in all six possible reading frames using the "vertebrate mitochondrial" genetic code. The "best frame" – the one minimizing stop codons – was identified using transAlign (Bininda-Emmonds, 2005). If the best translation contained a stop codon, the sequence was removed. Thus, I obtained a FASTA file containing all translatable sequences corresponding to the 650 bp barcode region of COI.

In order to determine the species names of these, the "Species Summary" feature of SpeciesIdentifier (Meier et al. 2006) was used. Thus, a raw set of species names was obtained. This raw list of names was likely to contain synonyms, sequences identified to genus only, and other types of undesirable variations. In order to obtain a list of unique species, the genus and species names were separated into different columns and the list was sorted by species epithet. Identical and/or near identical species epithets were checked for new combinations. Names including "aff.", "cf." and "sp" were deleted as were unidentified sequences from environmental genomic studies.

In order to determine whether the species coverage of aquarium fish COI in Genbank was adequate for identifying invasive freshwater fish originating in the trade, Icompared the freshwater aquarium fish species list created by Ornamental Fish International (Hensen et al. 2010) (refer to chapter 2, section 2.2.1 for obtaining the list) to the Genbank fish COI species list. The results were presented in the barchart showing the number of overlapping species between the two lists (Figure. 4.3.1).

4.2.2. Specimen collection and identification

Specimens were collected from major ornamental fish farms (Qian Hu Fish Farm and Aquatech), and aquarium retailers (mainly Clementi

Florists) in Singapore for duration of 3.5 years (Febuary 2009 to June 2012) (refer to electronic Appendix I: species list). Alcohol preserved specimens from the Raffles Museum of Biodiversity Research (RMBR) were also used to increase species coverage. A total of 1450 specimens from 522 species were processed as described above.

Fish farms and retailers vary in their standards of identifying and labelling fish species. Fish farms frequently provide scientific names for their specimens, whereas retailers generally use common names. In order to generate datasets with accurate species names for my analyses based on scientific names, all specimens obtained from the trade were identified using published fish identification keys and monographs (Roberts 1989, Kottelat 1990, Talwar and Jhingran 1991, Kottelat et al. 1993, Rainboth 1996, Kottelat 2001, Inger and Chin 2002, Norris 2002, Kottelat and Widjanarti 2005, Nelson 2006, Tan 2006, Axelrod et al. 2007, Hensen et al. 2010). In some cases, fish systematists from the RMBR (e.g., Drs. Tan Heok Hui, Ng Heok Hee) were consulted in order to ensure the accuracy of identifications.

4.2.3 Specimen vouchering, tissue sampling and image vouchering

High-resolution photographs were taken for vouchering purposes while also preserving a physical voucher. All images were made available online for public access as important supplementary information for the sequences generated in this study (http://evolution.science.nus.edu.sg/Ornamental_fish.html). A fish comparator website is also available that allows users to compare two fish specimens side by side (http://evolution.science.nus.edu.sg/Ornamental_Fish_Comparator.html). Fish specimens were documented on the lateral habitus (e.g. Fig 4.4.1, A), and where required, dorsal (Fig. 4.4.1, B) and ventral (Fig. 4.4.1, C) orientations as well. In addition, some specimens were also imaged after preservation in formalin (E.g., Fig. 4.4.2, D3) in order to document specimen discoloration. Specimens were imaged using a Nikon D300 (60mm Macro Lens) and a slaved flash system. The image resolution generated with this system is sufficient enough to show chromatophores on the fish (Fig. 4.4.2E), and is comparable in quality to observations obtained with a good-quality stereomicroscope. Images were edited with Adobe® Photoshop® CS4 and exported in Zoomify™ format (as a Zoomifyer) and embedded into a fish specimen image database website. The Zoomifyer is a specialized Flash object that allows users to stream high-magnification images of morphological features that are critical for species identification: it divides an image

into a series of smaller-sized picture tiles at different resolutions and sizes that are presented onto a fixed frame. Because the viewer frame requires only few picture tiles to be loaded at any time, viewing is fast and smooth. As Zoomifyers comprises a simple package of HTML code, small image files and a simple Flash movie code; it can be played readily on any browser with Flash support.

All voucher specimens, tissue samples and images were labelled with unique serial numbers and species names to allow voucher tracing. One set of tissues was used in my study, while a second set of tissues was excised and stored in RMBR's cryo facility. Physical vouchers were prepared via preservation in formalin with subsequent storage in 75% ethanol in RMBR.

4.2.4 DNA Extraction, amplification and sequencing

In order to create a DNA barcode sequence library of aquarium fish in the Singapore trade, tissues extracted from the collected specimens were subjected to the following treatment. Genomic DNA was obtained by Phenol/Chloroform extraction. Polymerase chain reactions (PCR) were carried out in 25µl mixture using Takara Ex Taq ™ DNA Polymerase following manufacturer's recommendation. Fish specific COI primer cocktails were used. Forward primers included: FishF2: TCGACTAATCATAAAGATATCGGCAC and FishF1:

TCAACCAACCACAAAGACATTGGCAC. Reverse primers included: FishR1: TAGACTTCTGGGTGGCCAAAGAATCA and FishR2: ACTTCAGGGTGACCGAAGAATCAGAA (Ward et al. 2005). The PCR cycling conditions were as follows: initial denaturation at 95°C for 1.5min, annealing at 50°C to 54°C for 1.5min and extension at 72°C for 1.5min. PCR amplicons were cleaned using Sure Clean (Bioline, Luckenwald, Germany). The purified amplicon served as the template for cycle sequencing reaction using BigDye terminator (condition: 30 cycle of 95°C for 30s, 52.5°C for 30s and 60°C for 4min) with the respective primers. Sequences were obtained using ABI3730 96-capillary sequencer. Sequencher 4.6 from Gene Code Corporation was used for sequence editing and forming contigs.

4.2.5 Datasets used for analyses for DNA barcodes

Three datasets were used in the current study to analyse the utility of COI as a DNA barcode to identify ornamental trade fish. The first dataset corresponded to all ornamental fish sequenced in this project. This database consists of 1,450 COI from 522 species and was named the "local COI dataset". The second dataset is for all fish in the ornamental trade, which comprised of sequences generated in the current study as well as COI sequences for ornamental fish in GenBank. This database was named "Ornamental Fish COI dataset". It contained 14,981 sequences from 1,578 species. The third dataset

contained all fish COI sequences from GenBank is named "Global fish COI dataset". It contained 62,624 COI sequences for 8,335 fish species and was obtained by combining data generated in this study with COI sequences obtained from GenBank.

Singletons, i.e., species represented by one sequence only, present particular challenges in testing species identification via DNA barcodes. Once treated as a query sequence, they lack representation in the barcode database against which the query can be identified. Identifying the query will thus always lead to an inaccurate identification. However, such misidentifications are an artefact of sampling instead of indicating of the failure of a DNA barcode to be diagnostic for the species. I therefore conducted analyses using both the full datasets and datasets from which the singletons had been removed. One such dataset was produced for each of the three databases described previously. The local COI dataset without singletons included 1,265 COI sequences for 359 species, the ornamental fish COI dataset without singletons included 14,612 sequences for 1,207 species and the global fish COI dataset without singletons included 60,336 sequences for 6,039 species.

4.2.6. Sequence analysis

4.2.6.1. Global alignment-based analyses

Best Match (BM) and Best Close Match (BCM) analyses were carried out using SpeciesIdentifier (Meier et al. 2006). In order to identify sequences using BM and BCM, I had to create datasets with aligned sequences for the three datasets. Since COI is a protein encoding gene, alignment was based on amino acid translations as implemented in Alignmenthelper or Mega 4.1 (The algorithm employed was Clustal W (Thompson et al. 1994). The aligned data were analysed in SpeciesIdentifier (Meier et al. 2006). The Global fish dataset was aligned using MAFFT (Katoh et al. 2002).

4.2.6.2. SpeciesIdentifier

For BM and BCM analyses the query was matched to species in the dataset with the nearest pair-wise distance (uncorrected p-distance) as long as at least 300bp of overlap existed between the sequences. A BCM analyses requires a set cut-off distance threshold (Meier et al. 2006). I used 3% (see Chapter 2), as this threshold was recommended and commonly used by Hebert et al. (2003). Query identification was considered successful when the query and corresponding sequences of the same species name were correctly

matched, unsuccessful when the query and corresponding sequences were incorrectly matched and ambiguous when two or more sequences from different species had equally good matches to the query sequence.

4.2.6.3. BLAST based identification

I also tested species identification with DNA barcodes using a local pairwise alignment tool; i.e., BLAST. For conducting BLAST-based analyses, I created BLAST databases using the makeblastdb command. Then each sequence was queried to the database using blastn, under settings of MEGABLAST and e-value cut-off of 1e-5 (Altschul et al., 1990). In order to summarize and compare results with SpeciesIdentifier based analyses, I used the same criteria of classifying sequences as correct, incorrect and ambiguous identifications. A match of the query sequence to itself was excluded and the results were parsed under the criterion of minimum hit length of 300 bp as in the BM and BCM analyses. For analyses corresponding to BM I used the best hit, and for analyses corresponding to BCM, I applied a 97% identity threshold (corresponding 3% difference) as in case of BCM prior to determining the sequence with closest identity to the query sequence.

4.2.6.3. Alignment-free Analyses using BRONX

BRONX identities species based on short diagnostic barcodes. A BRONX database was built using each of the three datasets using fasta2bdb.pl script provided in BRONX(v2) (Little 2011). Sequences were then matched to the database using a modified script for BRONX that gave the best hits as well as a second best hit (bronx.pl), and any hit of the query sequence to itself was removed. After removal, the best hit was determined. I also obtained a corresponding score and determined identifications at various score thresholds for the Global fish dataset (Fig 4.2.6.2.2.). I found that a score threshold of 200/300 maximized the number of sequences identified and while reducing misidentifications. Given that I found similar results for the dataset in Chapter 3, I used a score of 200 consistently across all analyses.
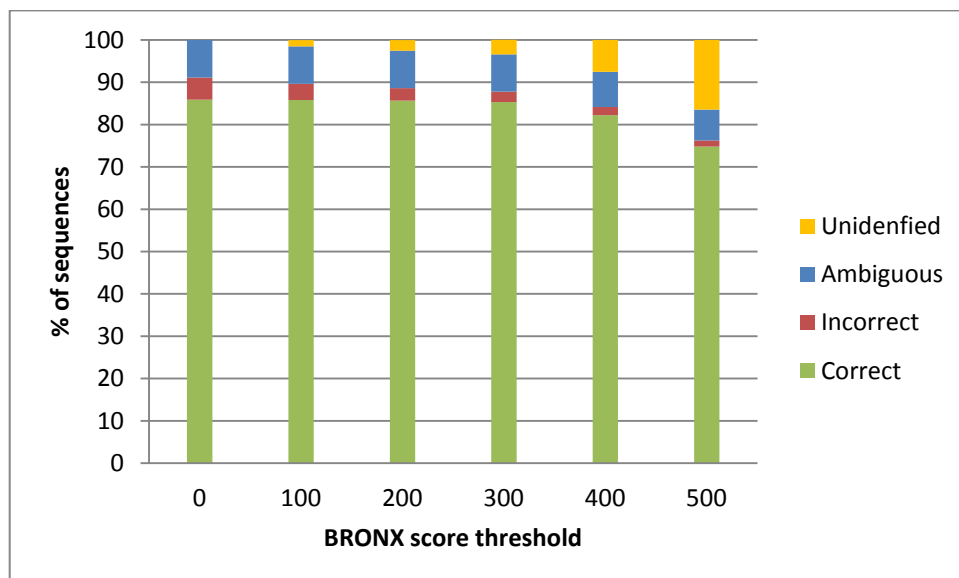
Fig 4.2.6.2.2. Percentage of sequences 1) correctly identified 2) incorrectly identified and 3) unidentified for the Global COI dataset at various score thresholds for BRONX.

4.2.7. Determining number of publications in fish barcoding and Environmental DNA over time

In order to assess whether fish barcoding for freshwater fish and eDNA studies were increasing in the scientific literature, I determined queried Web of Science™ database hosted by Thomson Reuters using the following search criteria: *((fish or teleos* or chondric*)(coi or dna barcod*))*. This search found 612 publications with most being related to fish disease and parasites. A search within the 612 publications using *(freshwater or stream or bay or river)(fish*)* identified 205 publications related to DNA barcodes of freshwater fish. Both numbers were combined to create the overall publication trend in fish barcoding because this provided a more specific trend than using the 612 publications in the initial search. I analyzed the results further to obtain annual publication numbers (Figure 4.3.3.1).

Note that the above mentioned criteria were used after initially examining several other strategies such as: *"fish DNA barcod*" or "Teleos* DNA barcod" or "Chondric* DNA barcod*" or "fish COI" or "Teleos* COI" or "Chondric* COI" or "fish coxI" or "Teleos* coxI" or "Chondric* coxI"*. These search criteria were too stringent and produced only 72 publications and excluded many fish barcoding papers. Another option would be to take all the citations from the five most cited fish barcoding papers. However, this yielded too many unspecific results (1,600 citations of which most were not related to fish).

Similarly, the publication trends related to environmental genomics were assessed. I used the search criteria: "Environmental genomic*" or "eDNA" or "environmental DNA" to obtain publication numbers. The publications were exported into Endnote and 727 publications were manually assessed to determine the relevance to identification of eukaryotes via eDNA and organized by publication year. Furthermore, these were categorized as eukaryotes (animals and fish). This was necessary because most publications dealt with microbial eDNA.

4.3. Results

4.3.1. Species coverage of Freshwater Aquarium Fish in Genbank

While Genbank contains COI for 8,335 fish species, only 1,225 species are freshwater ornamental fish according to OFI (Hensen et al. 2010). The remaining 3,453 aquarium fish species do not have COI sequences (Figure 4.3.1).

	Number of species
Ornamental fish without COI in Genbank	3453
OFI species record	4679
Ornamental fish COI in Genbank	1225
Genbank species record (COI)	8335

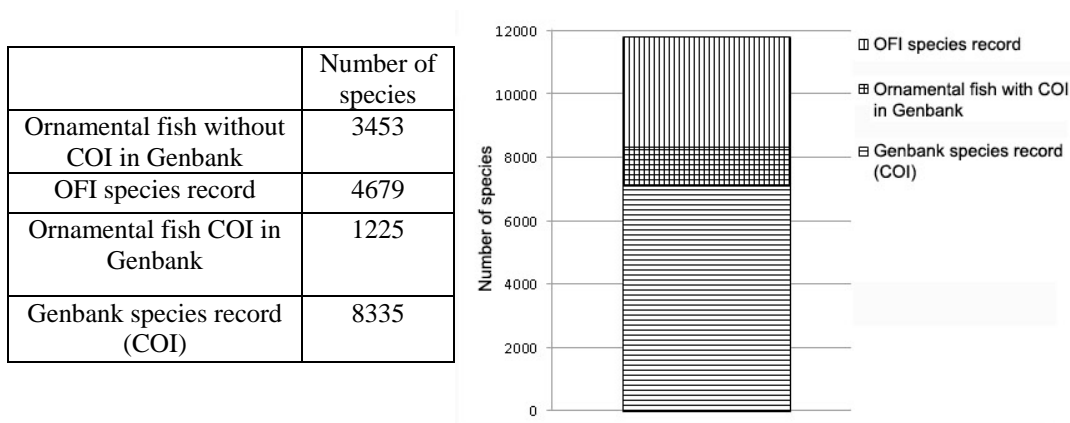Figure 4.3.1. Species coverage of freshwater aquarium fish in Genbank

4.3.2. Identification of Ornamental Fish

The identification success of ornamental fish varied from 77-92% depending on dataset used and method of identification. An initial test assessed whether the species in the local COI dataset could be distinguished from each other. Here I observed ~77-81% identification success across the different methods

of analyses (Table 4(I)). However upon excluding the singletons in this dataset, the identification success increased dramatically to 87-92%. Overall, Best Match analyses or the corresponding analyses for BLAST and BRONX without any identity/score threshold yielded a large number of incorrect identifications. This effect was larger when datasets containing singletons were analyzed; upon excluding singletons, the number of incorrect identifications dramatically declined. Nonetheless the most accurate identifications were obtained either by using a percentage threshold as in case of SpeciesIdentifier's BCM, BLAST or a score threshold in case of BRONX. Here, the local COI dataset yielded 3.5-4.8% incorrect identifications while the local COI dataset without singletons yielded 1.89-2.46% incorrect identifications. Overall 3-5% of the identifications were ambiguous at species level across the different methods.

In the next step, I analyzed whether ornamental trade fish can be identified in the context of all available fish COI. This implied analyses of the ornamental fish datasets against the Global COI dataset (Table 4 (II). Here, in comparison to the local dataset against local dataset analysis, the results were better for the dataset including singletons, with percentage of correct identifications increasing to ~83-87% depending on the method of analyses. Incorrect identifications varied from 2-4% while ambiguities were much higher at 10-11%. Removing singletons further improved identification success to 89-90%. Overall, I observed that, singletons created fewer problems in this dataset, which is likely due to singletons accounting for ~20% of the taxa in this dataset in comparison to ~30% in case of the local datasets.

Next I analysed the efficiency of COI in identifying any fish species (Table 4 (III). This was done by querying every available fish COI in the Global fish COI database. Here, identification success varied from 83-87 % when singletons were included and from 87% to 89% when singletons were excluded. Upon comparing this with the previous analyses of identification of ornamental trade fish in context of all fish, I found that the results are very similar.

Table 4.I. Identification efficiency determined for COI dataset of ornamental fish in the Singapore trade using SpeciesIdentifier (Best Match and Best Close Match analysis), BLAST and BRONX. Values represent results for Full dataset (1a) / Dataset without singletons (1b).

Methodology	Correct identification (%) Full / Data without singletons	Incorrect identification (%)	Ambiguous identification (%)	Sequences without any match closer than cut off point in Best close match analysis (%)
SpeciesIdentifier (Best match)	79.07 / 91.06	15.88 / 4.03	5.04 / 4.9	na
SpeciesIdentifier (Best close match) (3%)	78.10 / 89.01	3.66 / 1.89	4.55 / 4.82	13.67 / 4.26
BLAST (without cut-off point)	81.22/90.83	15.19/5.69	3.45/3.48	0.14/0.32
BLAST (3%)	77.69/87.51	3.52/2.45	3.45/3.48	15.33 /5.93
BRONX (without cut-off point)	81.59/92.22	14.97/4.37	3.45/3.42	na
BRONX (200)	79.93/90.79	4.82/2.46	3.45/3.42	11.79/3.33

Table 4.II. Identification efficiency of ornamental fish COI upon querying against global fish COI database using SpeciesIdentifier (Best Match and Best Close Match analysis), BLAST and BRONX. Values represent results for Full dataset / Dataset without singletons

Methodology	Correct identification (%)	Incorrect identification (%)	Ambiguous identification (%)	Sequences without any match closer than cut off point in Best close match analysis (%)
Best match analysis	83.96/na	4.5/na	11.54/na	-
Best close match analysis (3%)	83.21/na	2.43/na	11.32/na	2.51/na
BLAST (without distance cut-off)	85.09/87.24	4.89/2.48	9.99/10.25	0.03/0.03
BLAST (3%)	84.34/86.5	2.68/1.95	9.98/10.23	2.98/1.32
BRONX (without score cut-off)	86.7/89.63	4.06/1.64	9.25/8.73	-
BRONX (200)	86.47/89.86	2.32/1.23	9.22/8.69	2.00/0.75

Table 4.III.Identification efficiency determined for the global fish COI dataset in SpeciesIdentifier (Best Match and Best Close Match analysis), BLAST and BRONX. Values represent results for Full dataset (3a) / Data without singletons dataset (3b).

Methodology	Correct identification (%) Full / Data without singletons	Incorrect identification (%)	Ambiguous identification (%)	Sequences without any match closer than cut off point in Best close match analysis (%)
Best match analysis	83.87 / na	5.91 / na	10.2 / na	-
Best close match analysis (3%)	83.3 / na	3.08 / na	9.99 / na	3.61 / na
BLAST (without cut-off point)	84.02 /88.05	6.43 /2.61	9.54 /9.31	0.032/0.031
BLAST (3%)	83.11 / 87.43	3.46 / 2.07	9.53 / 9.29	3.57 / 1.21
BRONX (without cut-off point)	85.88/89.77	5.25/1.77	8.87/8.45	-
BRONX (200)	85.67/89.52	2.96/1.36	8.84/8.42	2.53/0.69

4.3.2. Comparison of SpeciesIdentifier, BLAST and BRONX for identifying fish COI

I used two distance based approaches and one character based method for identifying species. SpeciesIdentifier uses a global alignment for all species, BLAST uses a local pairwise alignment while BRONX uses an alignment free approach. Across all analyses I find that in terms of correct identifications BRONX outperforms the other approaches, while the distance based approaches yield similar results. In terms of ambiguous identifications, a similar pattern was observed; BRONX yielded fewer ambiguities followed by BLAST and SpeciesIdentifier. In terms of incorrect identifications, results were more variable. While distance based methods outperformed BRONX in the databases with sparser taxon samples (local database), I find the reverse for the databases with denser taxon sampling.

4.3.3. Publication trends in the field of fish barcoding, and environmental genomics

The number of fish barcoding publications has been increasing rapidly since the start of fish barcoding campaign in the year 2005. The number of marine fish and freshwater fish barcoding publications were initially similar, but after 2006 there are more publications on marine fish in every year (refer to Figure 4.4, series Fish total, Marine and Freshwater). The number of publications

related to environmental genomics is increasing linearly since 1999. The majority of environmental genomics publications were related to microbial diversity. Publications related to detecting environmental DNA of fish, frogs and other animals started to only appear after 2011 (refer to Figure 4.4, series eDNA total and eDNA eukaryotes).
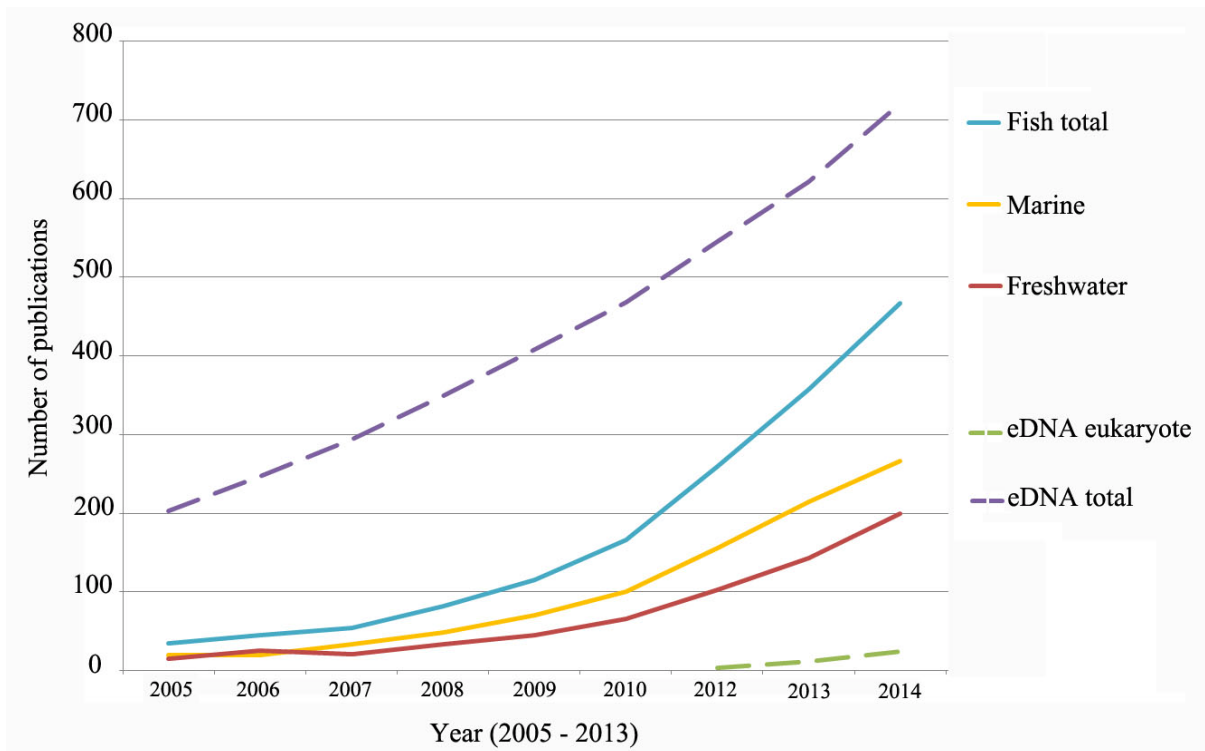
Figure 4.3.3. Publication trends of publications associated with environmental genomics (eDNA) & fish DNA barcoding. Legend: Fish total = all fish barcoding publications; Marine = fish barcoding publications associated with Marine fish; Freshwater = fish barcoding publications associated with freshwater fish; eDNA total = all environmental genomics publications; eDNA eukaryote: environmental genomics publications associated with animal DNA detection in water and soil.

4.4. Discussion

The trade in ornamental aquatic animals is one of the largest contributors to the transport of live organisms worldwide and estimated to be worth up to $25 billion US dollars (Hensen *et al.* 2009). Yet, despite its economic significance, the industry is poorly regulated which poses a significant risk to biodiversity and economic activity through the spread of invasive species and exotic pathogens (Collins *et al.* 2012). One of the factors impeding effective regulation is the difficulty of getting specimens identified to species. DNA barcoding has been proposed as a solution to this problem because it allows for the work to be carried out by laboratory technicians instead of fish systematists with doctoral degrees (Steinke et al. 2009). However, it has been shown that species coverage is crucial for the precise and accurate identification of unknown species via DNA barcodes (Ekrem et al. 2007) and that the overall COI species coverage in Genbank and BOLD is far from sufficient even for those taxa that have been subject to DNA barcoding campaigns (e.g., fish, birds, butterflies) (Kwong *et al.* 2012).

To date, I estimate that Genbank contains barcodes for ca. 8,335 species of fish that are distributed across 62,624 specimens. Of these, only 1,225 species are fish species in the freshwater ornamental trade, which represents only about a quarter of all ornamental species recorded in the OFI list of freshwater aquarium fish (Hensen et al. 2010). This means that I are far from having sufficient species

116

coverage for using DNA barcodes for regulating the trade and/or using eDNA from water to diagnose the arrival of new invasive fish species. It has been estimated by Nelson (2006) that freshwater fish contribute ca. 42% of the global described fish diversity (ca. 14,000 species). Many of these species are in the trade and obtaining specimens from the trade may be the fastest way to increase the number of barcoded species. The alternative fresh collecting will be much slower and much more expensive although it has the advantage of generating specimens with precise locality information. Furthermore, concentrating on the fish in the trade has the advantage of generating barcodes for those species that are most likely to become invasive in the future.

Here, I generate COI data for 522 species of aquarium fish collected from the Singapore trade. Of these, 334 species are new to Genbank. For all species that I sequenced, I also prepared high-quality digital images that were processed and made available online (http://evolution.science.nus.edu.sg/Ornamental_fish.html). To facilitate future species identification, all images in the database are associated with species names and serial numbers that refer to vouchers and tissues in the main collection of the RMBR and its cryo-collection, respectively. Several authors have earlier suggested that COI barcodes be supplemented with high-quality images of voucher specimens (Steinke *et al.* 2009, Collins *et al.* 2012) and some authors have provided voucher photographs (Collins *et al.* 2012). However, the quality of many of these in the BOLDSYSTEMS

(http://www.boldsystems.org) is so poor (e.g., small size) that they are insufficient for taxonomic purposes. In addition, many of the sequences in BOLDSYSTEMS lack all voucher images. For example, only 14 out of the 30 COI for *Barbonymus gonionotus* contain voucher photographs. The corresponding number for *Cyprinus carpio* is 15 out of the 164 barcodes. Many of the BOLDSYSTEMS images are also species-specific while specimen images would be needed for checking vouchers. The species-specific images are apparently taken from Fishbase (Froese et al., 2013) which is now affiliated with BOLD. To my knowledge, my database is thus the first for fish where DNA barcodes are consistently associated with high-quality images.

My databases of aquarium fish barcodes and voucher photographs are important steps towards creating identification tools for detecting invasive freshwater species. Traditionally, marine fishes have received more attention than freshwater fish although the latter are arguably more important in invasion biology. Hence, it is time to refocus the barcoding efforts to freshwater fish with the goal to rapidly increase species coverage until all ornamental fish are covered. Barcoding all aquarium fish will then also help with covering all freshwater fish at a global scale because 30% of the species are available in the trade. A full dataset would be a boon for the budding field of fish detection via eDNA from water. eDNA has been used for a long time, but in the past it was mostly utilized for detecting and documenting microbial diversity in soil and water because of applied concerns about water and food

118

safety. However, eDNA studies designed to identify animal DNA (e.g., frogs, fish) are starting to appear in the literature (figure 4.3.3). Overall, I find that COI barcodes could potentially be an effective tool for identifying ornamental trade fish if the species coverage was better.

Based on current species coverage, I find that >83% of the ornamental fish can be identified when an unknown sequence is matched against all available fish barcodes. The identification success rates can be improved by optimizing the choice of analysis method. In the current study I find that a new character based-method, BRONX, can identify ~87% of the taxa accurately and minimize the incorrect and ambiguous identifications. I also find that ornamental fish have a similar identification success rate as all fishes; i.e., identification success did not decrease when I compared the sparser local dataset of ~1,000 sequences and ornamental datasets to the global datasets of ~60,000. This is somewhat surprising because it is generally assumed that sparser taxon sampling leads to higher identification success rates because fewer closely related species are included in the sample. Further improvements of species identification success rates can be obtained if all species have multiple sequences and if all sequences could be removed that are likely to be misidentified. It is well known that Genbank contains misidentified sequences and they will make a contribution to the "misidentified" sequences in my analysis. Note that the sequences that are classified in my analyses as "misidentified" and "ambiguous" have this status either because a specimen was

misidentified or because the DNA barcode for a species is not diagnostic.

Overall, the method requiring a global alignment has similar success rates to those that are based on pairwise or local alignments. I therefore recommend the latter because they are more efficient. Firstly, it is time-consuming to obtain a global alignment and each new query sequence will have to be added to the existing alignment before it can be identified. Even after the alignment has been obtained, a new BM and BCM would have to be run and these analyses are slow. For example, for my largest data set, the analysis required 7 days to complete although 60 gigabytes of computer memory had been assigned. In contrast, BLAST and BRONX each took much less time to complete the analyses of the same dataset. Therefore, these methods are overall more conducive for real-time application without jeopardizing identification accuracy.

My results are overall promising given that more than 85% of all species can be identified based on DNA barcodes. This identification success rate is somewhat lower than those reported in some other fish DNA barcoding projects that claim identification success rates of 99% to 100%. However, these studies often use problematic identification techniques (e.g., NJ trees) and focus on geographically very limited samples. Examples include the results of a preliminary fish barcoding initiative started by Ward *et al* (2005) for barcoding Australia's native

fishes. The lower identification success rates in my study are expected given that I sample fish at a global scale which is more likely to yield high intra-specific and low interspecific differences. I would also argue that identification success rates above 80% are sufficiently high for most applied purposes, because many other Taxa including Polifera, Cnidarians (Huang et al. 2008) and Sepsids (Meier et al. 2006) have exhibited success rates of below 65%. Moreover, I can identify which species do not have discrete barcodes so that the users of my database will be aware which identifications are trustworthy. Note also that misidentification rates based on DNA barcodes may overall not be very different from what can be achieved by experts in the field because morphology is sometimes not sufficient for identification. For example, many fish taxonomists use geographical information to delimit morphologically similar species. An example is *Sinogastromyzon puliensis* and *S. nantaiensis* that are morphologically indistinguishable Taiwanese species collected from two different streams in Southern Taiwan (Chen *et al.* 2002; Shao & Lim, 1991; Shen, 1993). Morphometric data for these species is overlapping and identification based on morphology requires geographic information. Putting aside the philosophical problem of describing species based on geographical isolation, precise geographic information is usually not available for ornamental fish. Often, the lack of precise geographical information for trade specimens impedes with distinguishing morphologically similar species from geographically isolated populations. This is another reason why the image database becomes

essential, as these images are specimen specific; hence they serve as essential information to complement the corresponding COI sequences.

There are two functions in the fish image specimen website: A specimen browser and a specimen comparator. In the specimen browser: (http://evolution.science.nus.edu.sg/Ornamental_fish.html; Fig. 4.4.2), the left frame is used for taxon navigation (A-C), while the right frame displays the specimen images (D-E). The user can select the desired Order tab (A), which then links to the list of Families (B). Clicking the desired Family tab further opens up a list of species specimens included in the database (C). Here, *Trigonostigma heteromorpha* specimen YGN187 has been selected, which is then displayed on the right frame (D1-D3). The main habitus image (freshly killed specimen) is provided as a Zoomifyer flash object (D1), which can be zoomed in to show high-resolution details such as the chromatic cells on the fish (E). The specimen information is also provided (D2), as well as any additional images (here, the same but discoloured formalin-preserved specimen is shown in D3).

In the specimen comparator: (http://evolution.science.nus.edu.sg/Ornamental_Fish_Comparator.html; Fig. 4.4.3), two taxonavigation frames are provided side by side (A), similar to the left frame in the specimen browser website (See Fig 4.4.1). Upon selection of desired specimen in each frame (B), they will

link to the specimen images side by side for comparison by the user (C). The image database is currently hosted by the Evolutionary Biology Laboratory (National University of Singapore), but will eventually be hosted by the Raffles Museum of Biodiversity to ensure permanency. Because this database functions on simple HTML and flash scripts, it can be sustained easily by anyone familiar with basic HTML. Furthermore, as most of the essential data (images, locality,
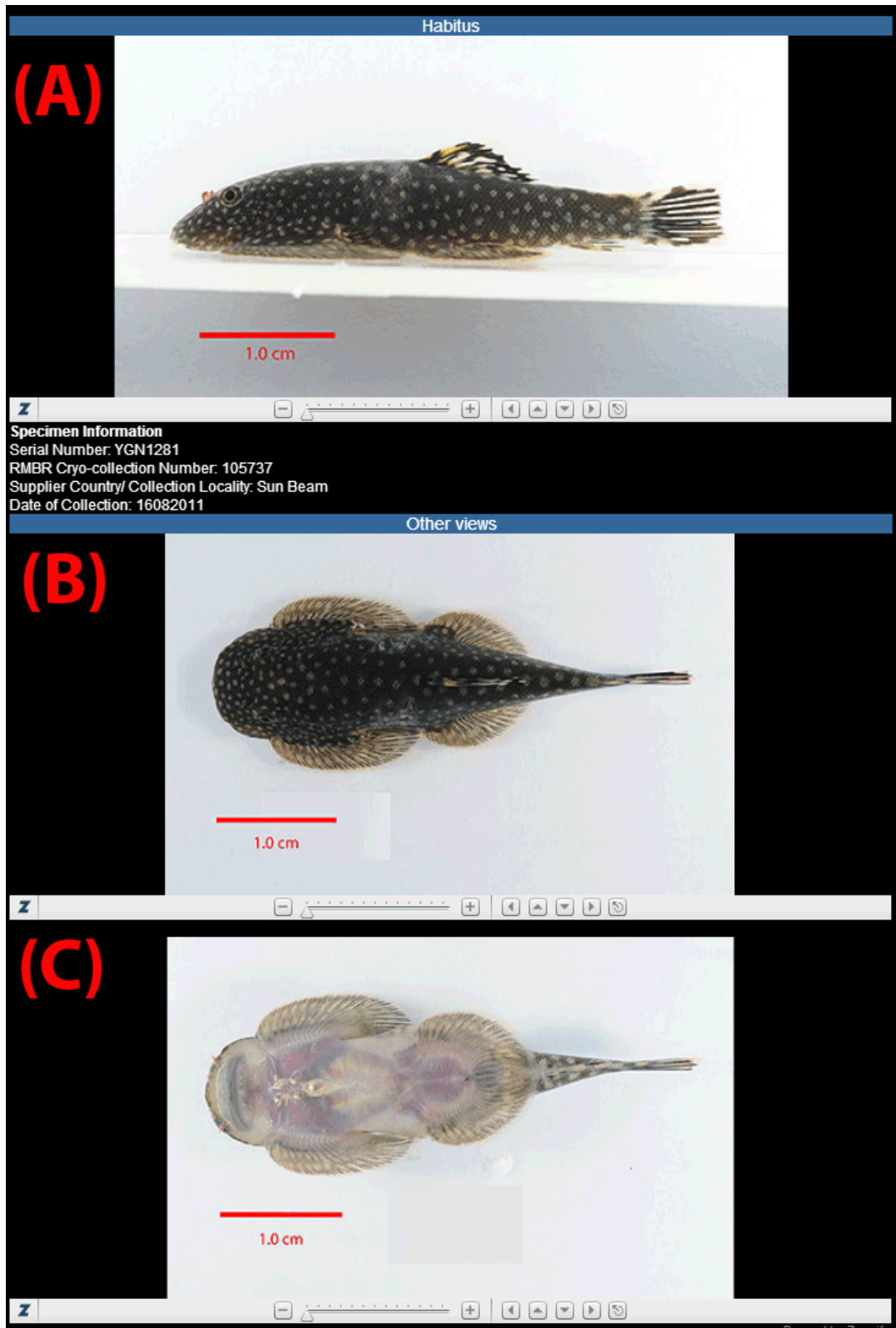
Figure. 4.4.1: Examplar presentation of habitus images for a freshly-killed *Gastromyzon ctenocephalus* specimen (YGN1281). Lateral (A), Dorsal (B) and Ventral (C) views of the habitus are displayed separately in three Zoomifyer Flash objects, which allows the viewer to zoom in to see more minute details on the fish.

Figure 4.4.2: A visual guide to using the visual specimen database (specimen browser) website. At the home page, the left frame is used for taxonavigation (A-C), while the right frame displays the specimen images (D-E). The user can select the desired Order tab (A), which then links to the list of Families (B). Clicking the desired Family tab further opens up a list of species specimens included in the database (C). Here, Trigonostigma heteromorpha specimen YGN187 has been selected, which is then displayed on the right frame (D1-D3). The main habitus image (freshly killed specimen) is provided as a Zoomifyer flash object (D1), which can be zoomed in to show high-resolution details such as the chromatic cells on the fish (E). The specimen information is also provided (D2), as well as any additional images (here, the same specimen is imaged again to show the effect of discoloration due to preservative formalin; D3).

125
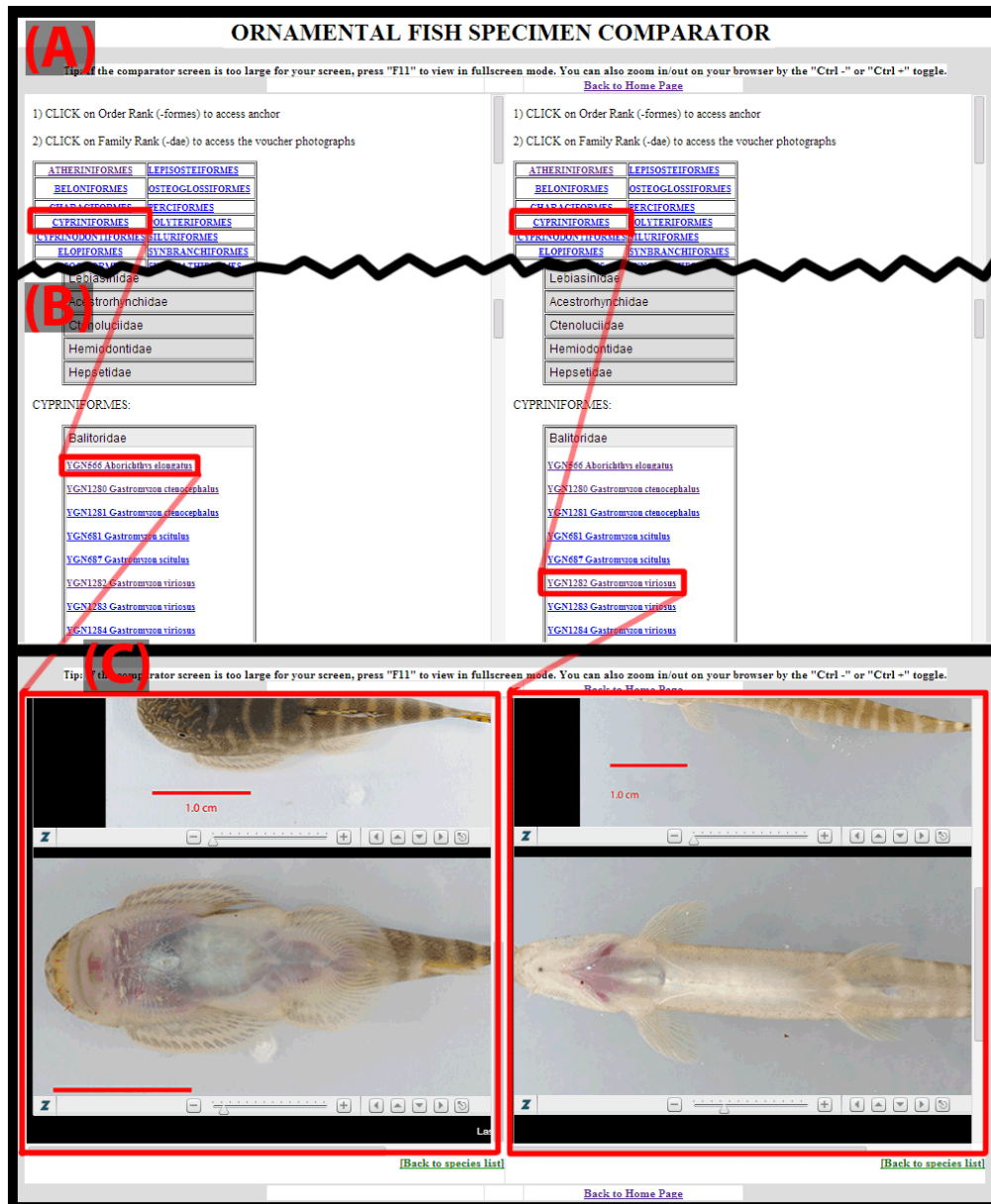
Figure 4.4.3: A visual guide to using the visual specimen database (specimen comparator) website. At the home page, two taxonavigation frames are provided side by side (A), similar to the left frame in the specimen browser website (See Fig 4.1). Upon selection of desired specimen in each frame (B), they will link to the specimen images side by side for comparison by the user (C).

126

DNA barcode and specimen information remain stored in universal formats (text and .tiff images), they can easily be imported to another platform should any current format fall out of favour online. For example, the Zoomifyer FLASH engine can easily be imported into a HTML5 version if needed in the future.

The database of aquarium fish COI and voucher images are important tools for monitoring the trade and regulating invasive species. Recently, there has been growing interest in environmental DNA (eDNA) as surveillance tools for identifying the presence of invasive species via DNA sequences (Figure 4.3.3). eDNA has increased sensitivity compared to traditional methods and can be efficient, which makes it effective for early detection of invasive species (Goldberg, Pilliod et al. 2011). But this method can only be used when a sufficiently large number of species has been barcoded. Unfortunately, this is not the case for freshwater ornamental fish although my sampling of aquarium fish from the Singapore trade has effectively increased the number of freshwater ornamental fish barcodes in Genbank by 27%. But there is much room for improvement and even with the addition of these 334 new species, coverage in Genbank is still unsatisfactory. Greater species coverage is clearly needed for DNA barcoding to be an effective biosecurity tool for rapidly and accurately identifying ornamental fish in the global trade

4.5. Conclusion

The different analysis methods have shown that COI is efficient enough for identifying fish for datasets of different scale, with identification success ranging from 77-91%. BRONX did exceptionally well, attaining identification success of 85% even for the global dataset containing singletons and large numbers of sequences. The species coverage in Genbank for ornamental fish is found to be sparse, with more than 3,000 recorded freshwater aquarium species being left out. Hence, it is crucial to include COI of these species before currently available databases can be used to monitor invasive species in the trade. Sequences in this study are supplemented with high resolution voucher photographs that are made easily accessible online for verification purpose. These images are superior in quality when compared to the inadequately available voucher images provided in BOLDSYSTEM database. The sequence and image database created for this chapter shall serve as important identification tools for detecting invasive species in water system in the future.

4.6. References for Chapter 4

Altschul, S.F., Gish, S.F. & Miller, W. (1990). Basic local alignment search tool. *Journal of Molecular Biology 215(3): 403-410.*

Axelrod, G. S., Burgess, W. E., Pronek, N., Axelrod, H. R., & Walls, J. G. (2007). Dr. Axelrod's Atlas of Freshwater Aquarium Fishes, 11th Edition. *T.F.H. Publications, Inc.*, 1160.

Barbuto, M., Galimberti, A., Ferri, E., Labra, M., Malandra, R., Galli, P., & Casiraghi, M. (2010). DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of "palombo" (Mustelus spp.). *Food Research International, 43*(1), 376-381. doi: 10.1016/j.foodres.2009.10.009

Bininda-Emmonds, O. R. P. (2005) transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioimformatics 6: 156*

Bleeker, W., Klausmeyer, S., Peintinger, M., & Dienst, M. (2008). DNA sequences identify invasive alien Cardamine at Lake Constance. *Biological Conservation, 141*(3), 692-698. doi: 10.1016/j.biocon.2007.12.015

Costello, M. J., May, R. M., & Stork, N. E. (2013). Can I Name Earth's Species Before They Go Extinct? *Science, 339*(6118), 413-416. doi: 10.1126/science.1230318

Collins, R. A., Armstrong, K. F., Holyoake, A. J., & Keeling, S. (2013). Something in the water: biosecurity monitoring of ornamental fish imports using environmental DNA. *Biological Invasions, 15*(6), 1209-1215. doi: 10.1007/s10530-012-0376-9

Chen, I., Han, C., & Fang, L. (2002). Sinogastromyzon nantaiensis, a new balitorid fish from southern Taiwan (Teleostei: Balitoridae). *Ichthyol. Explor. Freshwat., 13*(3), 4.

Chown, S. L., Sinclair, B. J., & van Vuuren, B. J. (2008). DNA barcoding and the documentation of alien species establishment on sub-Antarctic Marion Island. *Polar Biology, 31*(5), 651-655. doi: 10.1007/s00300-007-0402-z

Collins, R. A., Armstrong, K. F., Meier, R., Yi, Y. G., Brown, S. D. J., Cruickshank, R. H., Johnston, C. (2012). Barcoding and Border Biosecurity: Identifying Cyprinid Fishes in the Aquarium Trade. *Plos One, 7*(1). doi: 10.1371/journal.pone.0028381

Cote, I. M., Green, S. J., Morris, J. A., Jr., Akins, J. L., & Steinke, D. (2013). Diet richness of invasive Indo-Pacific lionfish revealed by DNA barcoding. *Marine Ecology Progress Series, 472*, 249-256. doi: 10.3354/meps09992

Ekrem, T., Willassen, E., & Stur, E. (2007). A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution, 43*(2), 530-542. doi: 10.1016/j.ympev.2006.11.021

Froese, R., Pauly, D., & Editors. (2013). Fishbase. *World Wide Web electronic publication., www.fishbase.org, version (12/1013).*

Goldberg, C. S., Pilliod, D. S., Arkle, R. S., & Waits, L. P. (2011). Molecular Detection of Vertebrates in Stream Water: A Demonstration Using Rocky Mountain Tailed Frogs and Idaho Giant Salamanders. *PLoS ONE, 6*(7). doi: 10.1371/journal.pone.0022746

Hausmann, A., Haszprunar, G., Segerer, A. H., Speidel, W., Behounek, G., & Hebert, P. D. N. (2011). Now DNA-barcoded: the butterflies and larger moths of Germany. *Spixiana, 34*(1), 47-58.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences, 270*(1512), 313-321. doi: 10.1098/rspb.2002.2218

Hebert, P. D. N., Ratnasingham, S., & deWaard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London Series B-Biological Sciences, 270*, S96-S99. doi: 10.1098/rsbl.2003.0025

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences of the United States of America, 101*(41), 14812-14817. doi: 10.1073/pnas.0406166101

Hensen, R. R., Ploeg, A., & Fossa, S. A. (2010). Standard names for freshwater fishes in the Ornamental Aquatic Industry. *OFI educational publication, 5*, 145.

Hensen, R. R., Ploeg, A., & Fossa, S. A. (2010). Standard names for freshwater fishes in the ornamental aquatic industry. *OFI educational publication, 5*, 145.

Huang, D. W., Meier, R., Todd, P. A., & Chou, L. M. (2008). Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. *Journal of Molecular Evolution, 66*(2), 167-174. doi: 10.1007/s00239-008-9069-5

Inger, R. F., & Chin, P. K. (2002). The Freshwater Fishes of North Borneo. *Natural History Publications (Borneo) Sdn. Bhd.*, 262.

Jerde, C. L., Chadderton, W. L., Mahon, A. R., Renshaw, M. A., Corush, J., Budny, M. L., . . . Lodge, D. M. (2013). Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. *Canadian Journal of Fisheries and Aquatic Sciences, 70*(4), 522-526. doi: 10.1139/cjfas-2012-0478

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research, 30*(14), 3059-3066. doi: 10.1093/nar/gkf436

Kottelat, M. (1990). Indochinese nemacheilines, A revision of nemacheiline loaches (Pisces: Cypriniformes) of Thailand, Burma, Laos, Cambodia and southern Viet Nam. *Verlag Dr . Friedrich Pfeil, Munchen, FRG*, 262.

Kottelat, M. (2001). Fishes of Laos. *WHT Publications (Pte) Ltd.*, 198.

Kottelat, M., Whitten, A. J., Kartikasari, S. N., & Wirjoatmodjo, S. (1993). Freshwater Fishes of Western Indonesia and Sulawesi. *Periplus Editions (HK) Ltd*, 293.

Kottelat, M., & Widjanarti, E. (2005). The fishes of Danau Sentarum National Park and the Kapuas Lakes Area, Kalimantan Barat, Indonesia. *Raffles Bulletin of Zoology*, 139-173.

Kwong, S., Srivathsan, A., & Meier, R. (2012). An update on DNA barcoding: low species coverage and numerous unidentified sequences. *Cladistics, 28*(6), 639-644. doi: 10.1111/j.1096-0031.2012.00408.x

Lim, G.S., Balke, M. & Meier R. (2011). Determining species boundaries in a world full of rarity: singletons, species delimitation methods. *Systematic Biology 61(1): 165-169.*

Lohman, D. J., & Samarita, V. U. (2009). The biology of carnivorous butterfly larvae (Lepidoptera: Lycaenidae: Miletinae: Miletini) and their ant-tended hemipteran prey in Thailand and the Philippines. *Journal of Natural History, 43*(9-10), 569-581. doi: 10.1080/00222930802610485

Meier, R., Shiyang, K., Vaidya, G., & Ng, P. K. L. (2006). DNA barcoding and taxonomy in diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology, 55*(5), 715-728. doi: 10.1080/10635150600969864

Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M. N., & Kawabata, Z. i. (2012). Surveillance of fish species composition using environmental DNA. *Limnology, 13*(2), 193-197. doi: 10.1007/s10201-011-0362-4

Nelson, J. S. (2006). Fishes of the world, 4th Edition. *John Wiley & Sons, Inc.*, 601.

Norris, S. M. (2002). A Revision of the African electric catfishes, Family Malapteruridae (Teleostei, Siluriformes), with erection of a new genus and descriptions of fourteen new species, and an annotated bibliography. *Annales Sciences Zoologiques, 289*, 155.

Rainboth, W. J. (1996). Fishes of the Cambodian Mekong. 263.

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes, 7*(3), 355-364. doi: 10.1111/j.1471-8286.2006.01678.x

Roberts, T. R. (1989). The Freshwater Fishes of Western Borneo (Kalimantan Barat, Indonesia). *Memoirs of the California Academy of Sciences, 14*, 210.

Robertson, M. D., Ovenden, J. R., & Barker, S. C. (2007). Identification of small juvenile scombrids from northwest tropical Australia using mitochondrial DNA cytochrome b sequences. *Ichthyological Research, 54*(3), 246-252. doi: 10.1007/s10228-007-0397-z

Rudnick, J. A., Katzner, T. E., Bragin, E. A., & DeWoody, J. A. (2007). Species identification of birds through genetic analysis of naturally shed feathers. *Molecular Ecology Notes, 7*(5), 757-762. doi: 10.1111/j.1471-8286.2007.01796.

Saunders, G. W. (2009). Routine DNA barcoding of Canadian Gracilariales (Rhodophyta) reveals the invasive species Gracilaria vermiculophylla in British Columbia. *Molecular Ecology Resources, 9*, 140-150. doi: 10.1111/j.1755-0998.2009.02639.

Shao, K., & Lim, P. (1991). Fishes of freshwater and estuary. Encyclopedia of field guidein Taiwan. *Recreation Press, Co., Ltd., Taipei., 31*, 240.

Shen, S. (1993). Fishes of Taiwan. *Department of Zoology, National Taiwan University, Taipei*, 1028.

Steinke, D., Zemlak, T. S., & Hebert, P. D. N. (2009). Barcoding Nemo: DNA-Based Identifications for the Ornamental Fish Trade. *Plos One, 4*(7). doi: 10.1371/journal.pone.0006300

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System ([www.barcodinglife.org)](www.barcodinglife.org). *Molecular Ecology Notes, 7*(3), 355-364. doi: 10.1111/j.1471-8286.2006.01678.x

Takahara, T., Minamoto, T., & Doi, H. (2013). Using Environmental DNA to Estimate the Distribution of an Invasive Fish Species in Ponds. *PLoS ONE, 8*(2). doi: 10.1371/journal.pone.0056584

Tavares, E. S., & Baker, A. J. (2008). Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. Bmc Evolutionary Biology, 8. doi: 8110.1186/1471-2148-8-81

Talwar, P. K., & Jhingran, A. G. (1991). Inland fishes of India and Adjacent Country. *Oxford & IBH Pub. Co.*, 1158.

Tan, H. H. (2006). The Borneo Suckers. *Natural History Publications (Borneo) Sdn. Bhd.*, 242.

Takahara, T., Minamoto, T., & Doi, H. (2013). Using Environmental DNA to Estimate the Distribution of an Invasive Fish Species in Ponds. *PLoS ONE, 8*(2). doi: 10.1371/journal.pone.0056584

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matric choice *Nucleic Acids Research 22(22): 4673-4680.*

Thomsen, P. F., Kielgast, J., Iversen, L. L., Moller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples. *PLoS ONE, 7*(8). doi: 10.1371/journal.pone.0041732

Valdez-Moreno, M., Vasquez-Yeomans, L., Elias-Gutierrez, M., Ivanova, N. V., & Hebert, P. D. N. (2010). Using DNA barcodes to connect adults and early life stages of marin

Victor, B. C., Hanner, R., Shivji, M., Hyde, J., & Caldow, C. (2009). Identification of the larval and juvenile stages of the Cubera Snapper, Lutjanus cyanopterus, using DNA barcoding. *Zootaxa*(2215), 24-36.

Ward et al., R. D., Tyler S. Zemlak, Bronwyn H. Innes, Peter R. Last, Paul D. N. Hebert. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B-Biological Sciences, 1716*, 11. doi: 10.1098

Will, K. W., & Rubinoff, D. (2004). Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics-the International Journal of the Willi Hennig Society, 20*(1), 47-55.

Wong, E. H. K., & Hanner, R. H. (2008). DNA barcoding detects market substitution in North American seafood. *Food Research International, 41*(8), 828-837. doi: 10.1016/j.foodres.2008.07.005

Yancy, H. F., Zemlak, T. S., Mason, J. A., Washington, J. D., Tenge, B. J., Nguyen, N. L. T., . . . Hebert, P. D. N. (2008). Potential use of DNA barcodes in regulatory science: Applications of the Regulatory Fish Encyclopedia. *Journal of Food Protection, 71*(1), 210-217.

CHAPTER V

_____

Barcoding and Border Biosecurity:

Identifying Cyprinid Fishes in the Aquarium

Trade

Note: This chapter is presented in its original format in the published journal, Plos One:

Collins, R. A., Armstrong, K. F., Meier, R., **Yi, Y.**, Brown, S. D. J., Cruickshank, R. H., Keeling, S. & Johnston, C. (2012). "Barcoding and Border Biosecurity: Identifying Cyprinid Fishes in the Aquarium Trade."

Plos One 7(1)

My Contributions to this publication

In Chapter III & IV, I have shown that COI of freshwater fish have high diagnostic value, exhibiting more than 80% efficiency for datasets of different scale (regional scale, Singapore ornamental trade scale, international trade scale and global scale). In chapter V, I have collaborated with Rupert A. Collins and Karen F. Armstrong from Lincoln University (New Zealand) to investigate the diagnostic value of Cyprinid's COI; an example of diagnostic analysis at taxonomic level. I have contributed 30 species of cyprinids for 120 specimens in this study, increasing the species and sample coverage of cyprinids in Rupert's original dataset. I have also taken part in debates regarding the identities of many specimens that have shown low interspecific variations and high intraspecific variations, hence influencing the outcome of the written discussion. I have also taken part in experimental design and performed the necessary experiments to obtain many of the COI and Rho sequences that have been used in this study.

Barcoding and Border Biosecurity: Identifying Cyprinid Fishes in the Aquarium Trade

Rupert A. Collins[1]*, Karen F. Armstrong[1], Rudolf Meier[2], Youguang Yi[2], Samuel D. J. Brown[3], Robert H. Cruickshank[3], Suzanne Keeling[4], Colin Johnston[4]

1 Bio-Protection Research Centre, Lincoln University, Canterbury, New Zealand, 2 Department of Biological Sciences and University Scholars Programme, National University of Singapore, Singapore, Singapore, 3 Department of Ecology, Faculty of Agriculture and Life Sciences, Lincoln University, Canterbury, New Zealand, 4 Animal Health Laboratory, Investigation and Diagnostic Centre, Ministry of Agriculture and Forestry, Upper Hutt, New Zealand

Abstract

Background: Poorly regulated international trade in ornamental fishes poses risks to both biodiversity and economic activity via invasive alien species and exotic pathogens. Border security officials need robust tools to confirm identifications, often requiring hard-to-obtain taxonomic literature and expertise. DNA barcoding offers a potentially attractive tool for quarantine inspection, but has yet to be scrutinised for aquarium fishes. Here, we present a barcoding approach for ornamental cyprinid fishes by: (1) expanding current barcode reference libraries; (2) assessing barcode congruence with morphological identifications under numerous scenarios (e.g. inclusion of GenBank data, presence of singleton species, choice of analytical method); and (3) providing supplementary information to identify difficult species.

Methodology/Principal Findings: We sampled 172 ornamental cyprinid fish species from the international trade, and provide data for 91 species currently unrepresented in reference libraries (GenBank/Bold). DNA barcodes were found to be highly congruent with our morphological assignments, achieving success rates of 90–99%, depending on the method used (neighbour-joining monophyly, bootstrap, nearest neighbour, GMYC, percent threshold). Inclusion of data from GenBank (additional 157 spp.) resulted in a more comprehensive library, but at a cost to success rate due to the increased number of singleton species. In addition to DNA barcodes, our study also provides supporting data in the form of specimen images, morphological characters, taxonomic bibliography, preserved vouchers, and nuclear rhodopsin sequences. Using this nuclear rhodopsin data we also uncovered evidence of interspecific hybridisation, and highlighted unrecognised diversity within popular aquarium species, including the endangered Indian barb *Puntius denisonii*.

Conclusions/Significance: We demonstrate that DNA barcoding provides a highly effective biosecurity tool for rapidly identifying ornamental fishes. In cases where DNA barcodes are unable to offer an identification, we improve on previous studies by consolidating supplementary information from multiple data sources, and empower biosecurity agencies to confidently identify high-risk fishes in the aquarium trade.

Introduction

Globalisation in the form of international trade breaches biogeographical as well as administrative boundaries, enabling organisms to colonise regions beyond their contemporaneous natural ranges [1]. The impacts of invasive alien species are well documented as a leading cause of global biodiversity decline and economic loss [2,3], and particularly as a driving force in the biotic homogenisation and degradation of freshwater ecosystems [4–6]. Biosecurity challenges exist in effectively monitoring and managing the complex pathways involved [1,7,8], with a key issue for risk assessment being the identification of traded biological materials to species [9–11]. Effective cataloguing of both potential propagules (all traded species) and known invasive alien species, can inform risk analyses and facilitate pre- or post-border control measures

(i.e., import restrictions and quarantine). In circumstances where species cannot be diagnosed easily by morphology and/or only certain life history stages can be identified, standardised molecular protocols for species identification are important for biosecurity [9–11]. However, these techniques still require further testing and reference libraries need to be expanded to encompass more species.

The ornamental aquatic industry is among the world's largest transporters of live animals and plants, with an annual trade volume estimated at US$15–25 billion [12,13]. Data from the United States implicates the industry as the primary transport vector in 37 of 59 fish introductions [6]. In Singapore–a global aquarium fish trading hub–at least 14 invasive ornamental fish species were reported to be resident in reservoirs in 1993 [14]. The risks presented by this industry are not, however, limited to traded

invasive fishes. Associated pathogenic organisms such as protozoa, bacteria and viruses are equally undesirable, with exotic pathogens known to cause harm to native species [15], industrial food aquaculture [16–18], and also the ornamental fish trade [13]. Compounding this, some pathogens can be vectored by carrier hosts with no clinical signs of disease [13,15,18], and host-taxon specific pathogens may also require special quarantine measures [13,18].

Aquarium fishes are both wild caught, and captive bred at aquaculture facilities, with over one billion fishes traded through more than 100 countries in 2000 [18]. In the case of freshwater fishes, $\geq 90\%$ of the trade volume is in a relatively small number of popular species sourced from commercial farms [19], while more diverse wild caught exports contribute the remainder. A complex supply chain exists for these ornamental fishes, and before they arrive at a retailer they may have passed though a series of regional and international distribution centres where consignments can be consolidated, reconsolidated and subdivided [13]. This potentially increases the number of access points for undesirable organisms to enter each shipment [13], as well as opportunities for mislabelling. While statistics are available on total volumes sold, little quantitative data exist on the number and composition of species involved in the aquarium trade, but it has been estimated that up to 5,300 species have been available at some point [20]. The industry in aquatic ornamentals for the aquarium hobby is a dynamic business, with new and undescribed species frequently appearing from new areas. Some, such as *Puntius denisonii* have quickly moved from obscurity to becoming a major Indian export and a conservation concern within a few years [21,22].

Approaches to addressing biosecurity threats from ornamental fishes are varied; the United States and United Kingdom adopt a "blacklist", whereby a small group of known high-risk species are subject to controls [23,24], while countries such as Australia and New Zealand who view this industry as a greater biosecurity threat, permit only fishes included on a "whitelist" of manageable species [17,18,24,25]. A total of 82 cyprinid (Teleostei: Cypriniformes: Cyprinidae) fish species are permitted for import as ornamentals in New Zealand [25]. Of these 82 species, 27 are further classified "high-risk" in terms of disease susceptibility, and require specific mitigation measures [25]. For the enforcement of these restrictions, an effective biosecurity procedure requires fast and accurate early detection of potentially harmful fishes at the pre-retail quarantine stage. For a variety of reasons, however, it may be difficult for inspectors to definitively identify all species likely to be encountered [17,26].

Use of the standardised mitochondrial cytochrome *c* oxidase subunit I (COI) DNA barcoding protocol, *sensu* Hebert *et al.* [27,28], has been demonstrated as an effective fish identification tool in situations including consumer protection [29–31] and fisheries management/conservation [32,33]. Steinke *et al.* [34] also effectively demonstrated application of this technique for the trade in marine ornamental fish species, with their study reporting a high rate of identification success.

Here, we test this DNA barcoding approach for identification of ornamental cyprinid fishes obtained from aquarium retailers and wholesalers. Of the global diversity of $>2,400$ cyprinid fish species [35], some such as the barbs, danios and rasboras are popular aquarium or pond fishes, and are commonly available in petshops. Many are difficult to identify based on morphological features, and some represent risks in terms of their potential as invasive species and pathogen vectors [6,18,25]. We test the DNA barcoding method by comparing congruence of taxonomic identifications based on morphological features, with the patterns in DNA barcodes. In order to expand taxon coverage we also evaluate the utility of extra data from GenBank and the Barcode of Life Data System, Bold [28]. These databases will include sequences for additional species, but may also include sequences from misidentified specimens or specimens collected from otherwise unsampled, divergent populations [26,36,37]. Therefore, we conduct separate analyses for our own data, GenBank/Bold data, and all data combined. In addition, we use a range of different identification techniques in order to address criticisms of some commonly employed methods [37–41], and also incorporate a measure of how rare species affect identification success [42].

As well as testing barcodes against morphological data, nuclear loci are increasingly used to validate mitochondrial results and also provide an independent, additional source of data for both identification, systematics or taxonomy [38]. In the case of aquarium fishes, a nuclear marker may also offer advantages in detecting natural introgression patterns, or interspecific hybridisation events that may have occurred during indiscriminate or deliberate breeding at ornamental fish farms. We will assess the utility of nuclear rhodopsin (RHO), a marker having been observed to show variation at the species level for molecular systematic questions [43], and also demonstrated to serve as an effective component of a multi-locus fish identification tool [44].

With the tendency of DNA barcoding studies to discover putatively cryptic taxa [45], it is likely that our study also uncovers previously unrecognised lineages that may represent species [46]. Some researchers have even questioned the validity of cryptic taxa as reported by divergences in mtDNA analyses [47–49], insisting species status be additionally supported with independent datasets, *sensu* the "integration by congruence" of Padial *et al.* [50]. Nuclear markers can assist in the critical assessment of these lineage divergences, so to this effect, RHO will also be used here to test support for these hypotheses.

Materials and Methods

Ethics Statement

Where applicable, this study was carried out in accordance with the recommendations of the National University of Singapore Institutional Animal Care and Use Committee (IACUC) under approved IACUC protocol number B10/06 (proposal entitled "Raffles Museum of Biodiversity Research Day to Day Operations"); living fishes were kept, photographed, and handled according to these rules in the cryo-collection of the Raffles Museum of Biodiversity Research.

Data Collection and Sampling

Specimens of ornamental cyprinid fishes were acquired from aquarium retailers, wholesalers and exporters in the United Kingdom, Singapore and New Zealand from 2008 to 2010. The non-cyprinid taxa *Gyrinocheilus* and *Myxocyprinus* were also included due to their ubiquity and superficial morphological similarity to some cyprinid fishes. Specimens were euthanised with MS-222 (tricaine methane sulfonate), before a tissue sample was excised from the right-hand caudal peduncle and stored at $-20°C$ in 100% ethanol. Specimens were subsequently formalin fixed and preserved in 70% ethanol as vouchers, following procedures outlined by Kottelat and Freyhof [51]. At least one specimen from each sample was photographed alive (left-hand side) prior to tissue sampling, with the remainder photographed after preservation. Voucher specimens for each COI barcode were deposited at the Raffles Museum of Biodiversity Research (ZRC), National University of Singapore.

Specimens were identified morphologically using scientific literature relevant to the group, and original descriptions were

consulted where possible. The use of "sp.", "cf." and "aff." notation in reference specimen identification follows Kottelat and Freyhof [51]. For analytical purposes, individuals designated "cf." are treated as conspecific with taxa of the same specific name, while those designated "aff." are treated as non-conspecific. Nomenclature follows Eschmeyer [52], unless otherwise stated. To assess the coverage of the project, a list of species believed to be in the aquarium trade was consulted as the most up-to-date and accurate guide available at this time [20]; we also used the MAF Biosecurity New Zealand Import Health Standard list of species [25].

Whenever possible, multiple individuals of each species were sampled. In order to better assess intraspecific genetic diversity, we tried to purchase multiple specimens at different times and from different vendors. Sampling efficiency was tested by correlating the number of haplotypes observed in each species with the number of individuals collected and the number of samples taken. For this purpose, a sample was considered as all conspecific specimens acquired from the same holding tank at the same premises on the same visit. These analyses were carried out in R version 2.12.1 [53], using a generalised, linear regression model with poisson distributions for count data; singleton species (species represented by one individual) were omitted.

DNA Protocols

Approximately 2–3 mm^2 of white muscle tissue was prepared for genomic DNA extraction using the Quick-gDNA spin-column kit (Zymo Research Corporation) following the manufacturer's protocol, but scaled to use a 50% volume of pre-elution reagents. Optimised PCR reactions were carried out using a GeneAmp 9700 thermocycler (Applied Biosystems) in 10 μl reactions. Amplification of the COI barcode marker comprised reactions of the following reagents: 2.385 μl ultrapure water; 1.0 μl Expand High Fidelity 10× PCR buffer (Roche Diagnostics); 0.54 μl MgCl$_2$ (25.0 mM); 2.0 μl dNTPs (1.0 mM); 1.5 μl forward and reverse primer (2.0 μM); 1.0 μl DNA template; 0.075 μl Expand High Fidelity polymerase (Roche Diagnostics). The COI fragment was amplified using one of the following primer pairs: FishF1 and FishR1 [54], LCO1490 and HCO2198 [55], or LCO1490A and HCO2198A [56]. Thermocycler settings for COI amplification were as follows: 2 min at 94°C; 40 cycles of 15 s at 94.0°C, 30 s at 48.0–52.0°C and 45 s at 72.0°C; 7 min at 72.0°C; ∞ at 4.0°C.

The nuclear RHO data were generated as per the COI protocol, but using the primers RH28F [57] and RH1039R [58], and the following reagents: 1.7 μl ultrapure water; 1.0 μl Expand High Fidelity 10× PCR buffer (Roche Diagnostics); 2.0 μl Q-Solution (Qiagen); 0.2 μl MgCl$_2$ (25.0 mM); 2.0 μl dNTPs (1.0 mM); 1.0 μl forward and reverse primer (2.0 μM); 1.0 μl DNA template; 0.1 μl Expand High Fidelity polymerase (Roche Diagnostics). Thermocycler settings for RHO amplification were as follows: 4 min at 94.0°C; 40 cycles of 20 s at 94.0°C, 30 s at 54.0–56.0°C and 60 s at 72.0°C; 7 min at 72.0°C; ∞ at 4.0°C.

Prior to sequencing, PCR products were checked visually for quality and length conformity on a 1% agarose gel. Bidirectional sequencing was carried out following the manufacturer's protocol on a Prism 3130xl Genetic Analyser (Applied Biosystems) using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). The same primer combinations as for PCR amplification were used for sequencing. Sequencing products were purified using the Agencourt CleanSEQ system (Beckman Coulter Genomics). Steps undertaken here to avoid or identify cross-amplification of nuclear mitochondrial pseudogenes (NUMTs) are outlined by Buhay [59] and Song et al. [60]. Sequence chromatograms were inspected visually for quality and exported

using FinchTV 1.4 (Geospiza). Trimmed nucleotide sequences were aligned according to the translated vertebrate mitochondrial amino acid code in the program Mega 4.1 [61]. The resulting COI fragment comprised a sequence read length of 651 base pairs (bp), positionally homologous to nucleotides 6,476 through 7,126 of the Danio rerio mitochondrial genome presented by Broughton et al. [62]. The RHO fragment corresponded to an 858 bp length (sites 58–915) of the Astyanax mexicanus rhodopsin gene, GenBank accession U12328 [44,63]. For COI and RHO, sequence data, chromatogram trace files, images and supplementary information were uploaded to Bold, and are available in the "Ornamental Cyprinidae" [RCYY] project. In addition to sequence data generated here, public databases including GenBank and Bold were searched under the following terms: "Cyprinidae", "COI", "CO1" and "COX1". Records were retained if the taxon in question was believed to occur in the aquarium trade [20], or if congeneric to a species we had already collected in our sampling. To facilitate analysis, nomenclature and spellings of GenBank/Bold records were updated or corrected following Eschmeyer [52].

Analysis

The suitability of COI barcodes as a species identification tool was tested using five primary metrics, thereby quantifying different properties of the data. Rather than simply providing a species-based descriptive summary, we simulated a real identification problem for a biosecurity official by treating each individual as an identification query. In effect, this means that each sequence is considered an unknown while the remaining sequences in the dataset constitute the DNA barcoding database that is used for identification. Identification rates for these queries were divided into four categories: "correct" or "incorrect", and "no identification" or "ambiguous" if applicable to the method. The extent to which rare, singleton specimens (one specimen per species) affect identification success rates is rarely explored, and is a problem for DNA barcode identification systems [42]. As few taxon-specific barcoding projects (i.e., databases) are complete [42], we aim to examine how the data perform for these singletons. It is therefore important for our analyses to distinguish between two identification scenarios. First, a query specimen belongs to a species that has already been barcoded and whose DNA barcode is maintained in a DNA barcoding database. Once sequenced, the best identification result for such a specimen is a "correct identification". Second, the query specimen belongs to a species that remains to be barcoded (it is a singleton). The best result here is "no identification", since the specimen has no conspecific barcode match in the database. The best overall identification technique is one that maximises identification success for scenario one, and yields a "no identification" result under scenario two. In light of this, we report results with both singleton species included (scenario two) and excluded (scenario one). When the analyses were carried out, however, the singletons remained in all datasets as possible matches for non-singletons. We term the success rates for scenario one (singletons excluded) as the "re-identification rate".

Unless otherwise stated, all descriptive statistics and analyses were conducted using Spider, Brown et al.'s DNA barcode analysis package for R [64,65]. Distance matrices and neighbour-joining (NJ) phylograms were generated under Kimura's two-parameter model (K2P/K80), with missing data treated under the "pairwise deletion" option. The K2P model was only used here to ensure consistency and comparability with other barcoding studies, but see Collins et al. [66] and Srivathsan and Meier [67] for more general discussion on the applicability of the K2P model. Negative branch lengths were set to zero [68,69]. Terminology of

topological relationships follows phylogenetic nomenclature consistent with literature but applies only to the gene tree relationships (e.g. monophyly, paraphyly, polyphyly). NJ phylograms were rendered in Web-based jsPhyloSVG format [70], following conversion from Nexus format into phyloXML using Archaeopteryx [71]. This creates an interactive vector-graphic phylogram with links to specimen database records and supplementary data (e.g. images) via embedded URLs.

The five primary metrics measuring identification success rates in this study are described as follows: (1) We employed a tree-based test of species monophyly, with this measurement reporting the exclusivity of the genetic clusters in an NJ phylogram. The procedure returns each species as either monophyletic (correct identification), non-monophyletic (incorrect identification) or singleton (incorrect identification). This per-species measure was then scaled to include the number of individuals in each species. We also incorporated a bootstrap test of node support, with correct identifications scored if values were greater than 70% [72]; 1,000 replications and codon resample constraints (block $=3$ option) were used for the bootstrap analysis. (2) A test using the k-nearest neighbour (k-NN) or "best match" classification approach [37,73] was employed on the K2P distance matrix. A nearest neighbour ($k=1$) conspecific with the query returned a correct identification, otherwise an incorrect identification; singletons were reported as an incorrect identification, and ties were broken by majority, followed by random assignment. (3) We used the "best close match" (BCM) method presented by Meier et al. [37]. In BCM, ties are reported as ambiguous and matches must be within a pre-specified threshold value (i.e., 1%) otherwise no identification is returned [37]. (4) Fourthly, the data were tested with a technique approximating the threshold method used by the Bold-IDS identification engine [28]. Bold-IDS will return a positive identification if a query shares a $>99\%$ similar unambiguous match with a reference specimen [28]. Here, data were tested on a per-individual basis, using the K2P distance matrix. A correct identification was returned if all distances within 1% of the query were conspecific, an incorrect identification resulted when all distances within the threshold were different species, while an ambiguous identification result was given when multiple species, including the correct species, were present within the threshold. This method is similar to BCM, but operates upon all matches within the threshold, rather than just nearest neighbour matches.

Lastly, we used a method incorporating an estimation of group membership; the general mixed Yule-coalescent (GMYC) models the probability of transition between speciation-level (Yule model) and population-level (coalescent model) processes of lineage branching [74,75]. This offers a likelihood based test of biological pattern in the data, i.e., approximating the "barcoding gap" of intraspecific versus interspecific variation. Following Monaghan et al. [75], data were reduced to haplotypes using Alter [76], with gaps treated as missing data (ambiguous bases were first transformed to gap characters). Next, ultrametric chronograms were generated in Beast v1.6.1 [77,78] under the following settings: site models as suggested by the BIC in jModelTest [79,80]; strict molecular clock; $1/x$ Yule tree prior; two independent MCMC chains with random starting topologies; chain length 20 million; total 20,000 trees; burn-in 10%; all other settings and priors default. The GMYC model was fitted in the Splits package for R [75], using the single threshold method under default settings. An individual was scored as a correct identification if it formed a GMYC cluster with at least one other conspecific individual. An incorrect identification was made when an individual clustered with members of other species, and a "no

identification" was made when an individual formed a single entity (did not cluster with anything else). Exploratory results (data not shown) suggested that more sophisticated Beast and GMYC analyses using relaxed clocks, codon partitioned site models, outgroups, and multiple threshold GMYC resulted in a poorer fit to the morphologically identified species names, as did a full dataset (sequences not collapsed into haplotypes).

The use of a universal (e.g. 1%) threshold has been questioned repeatedly [37,41,81,82], and although no single threshold is likely to suit all species, error can be minimised across a dataset for different threshold values. We tested a range of threshold percent values for their effect on both the false positive (α) and false negative (β) error rates. Categorisation of these error rates follows Meyer and Paulay [82]: "False positives are the identification of spurious novel taxa (splitting) within a species whose intraspecific variation extends deeper than the threshold value; false negatives are inaccurate identification (lumping) within a cluster of taxa whose interspecific divergences are shallower than the proposed value" (p. 2230). The optimum threshold is found where cumulative errors are minimised. Positive identifications were recorded when only conspecific matches were delivered within the threshold percent of the query. False negative identifications occurred when more than one species was recorded within the threshold, and a false positive was returned when there were no matches within the threshold value although conspecific species were available in the dataset. We incorporated a modification of the Bold and BCM analyses, using the revised threshold values generated during this procedure.

To evaluate the performance of the COI barcodes in terms of their agreement with nuclear RHO, a subset ($n=200$) of individuals were amplified for this marker. This yielded reduced datasets of 82 species (1–10 individuals per species) for which both the COI and RHO sequences were available. Barbs (Puntius) and danios (Danionini) were targeted, along with other taxa showing COI divergences. Patterns in the matched RHO and COI subsets were investigated using the NJ monophyly and k-NN methods. When a sufficient number of specimens were available (≥ 5) for aquarium species showing multiple COI clusters, we were able to explore this possibly unrecognised diversity with RHO, and assess an approach complementary to COI barcoding. We used four methods in assessing support for unrecognised or cryptic species: mean intergroup K2P distances; a character based approach using diagnostic, fixed character states between lineages, i.e., pure, simple "characteristic attributes" (CAs) [29,83]; bootstrap estimates of NJ clade support (settings as described above); and Rosenberg's P, a statistical measure testing the probability of reciprocal monophyly over random branching processes [84].

Results

A total of 678 cyprinid fish specimens were collected during the study, and these were identified to 172 species in 45 genera using morphological characters (refer to Table S1 for identifications, characters, taxonomic comments and bibliography). The survey of GenBank and Bold databases contributed a further 562 COI sequences from 157 species, with 81 of the species represented in both GenBank/Bold data and our data. With regard to the aquarium trade, the taxon coverage of this study represents 131 (39%) of the 333 aquarium cyprinid fishes listed in Hensen et al. [20], a proportion which increased to 56% coverage when GenBank/Bold data were also included. An additional 41 species not present in this inventory [20] were reported from our survey of the trade. In terms of biosecurity risk, our taxon sample covered 78% (85% including GenBank/Bold) of the 27 cyprinid fish

species listed as high-risk allowable imports to New Zealand [25]; of the total 82 permitted cyprinid fishes, our data represented 79% of these (90% including GenBank/Bold).

DNA barcodes were successfully amplified from all samples in the study with the primers reported. All nucleotides translated into functional protein sequences in the correct reading frame, with no stop codons or indels observed in the data. In our COI barcode dataset, each species was represented by an average of 3.9 individuals (2.32 sampling events), with twenty species by one individual (11.6%), and 102 (59%) by ≥ 3 individuals. The average number of haplotypes per species was 1.97, with sampling effort (sampling events and number of individuals per sp.) and haplotype diversity correlated ($P < 0.001$). Table 1 provides a further summary of barcode statistics, and links to Bold and GenBank database records for all sequences used in this study are presented as URLs in Figure S1 and Figure S2. All sequence data used in this study are also provided as supplementary text files (Fasta format): Dataset S1 (COI) and Dataset S2 (RHO).

Genetic diversity was generally lower within species than between, with 95% of total intraspecific variation less than 5.48% K2P distance. Of the interspecific distances to a closest non-conspecific neighbour (i.e., the "smallest interspecific distance" of Meier et al. [85]), 95% were above 1.72% K2P distance. Mean distance to closest non-conspecific was $10\times$ mean intraspecific distance. Of the intraspecific values, 13.5% were over 2% K2P distance, while 19.0% were above 1%. Graphical structure of the distance data is shown in the NJ phylogram presented as Figure S1, and indicates cohesive clusters for the majority of species. Many morphologically similar species were well differentiated with DNA barcodes, and Figure 1 illustrates an example.

Identification Success Rates using DNA Barcodes

When appraising the identification power of the barcode data, success rates were generally high ($>93\%$) when singletons were excluded (i.e., re-identification). The only exception was the NJ bootstrap analysis (89.7%). When GenBank/Bold data were added, correct re-identification rates dropped between 4% and 15% depending on identification technique. If singleton species were included in the results, the reduction in success rate was between 2.7% and 2.9% for the data generated in this study, and 5.2% and 7.4% when GenBank/Bold data were combined. When just the GenBank/Bold data were considered, success rates decreased between 13.6% and 20.8% depending on the method. Optimised distance thresholds were 1.4% for the barcodes in this study and 0.8% when combined with GenBank/Bold (Figure 2). A breakdown of identification success rate for each method and for each dataset is presented in Table 2.

Incongruence between Morphology, DNA Barcodes, and GenBank/Bold Data

Cases of incongruence and inconsistency for some common aquarium species are presented in a reduced NJ phylogram (Figure 3). Of the data generated in this study, barcode sharing was observed in two groups: between two Eirmotus species (E. cf. insignis and E. cf. octozona), and between two Rasbora species (R. brigittae and R. merah). Additionally, a polyphyletic species was observed: an individual of Danio cf. dangila (RC0343) clustered closer to D. meghalayensis than to other D. dangila. When GenBank data were added, several additional species were also non-monophyletic on the COI phylogram, with these added data conflicting with some barcodes generated in this study. For example, D. albolineatus became polyphyletic with the inclusion of D. albolineatus HM224143, as did D. roseus when D. roseus HM224151 was added. The topology of the NJ phylogram (Figure 3) is misleading for identification purposes, however, as all D. roseus remain diagnosable from D. albolineatus by a single transversion at position 564, while the remaining differences in D. roseus HM224151 are autapomorphies. Other aquarium species that were affected by GenBank data inclusion include (refer to Figure S1): haplotype sharing between a possibly undescribed Devario ("TW04") and D. annandalei HM224155; haplotype sharing and polyphyly of R. daniconius and R. cf. dandia; paraphyly of Barbonymus schwanenfeldii by Balantiocheilos melanopterus HM536894;

Table 1. Summary of descriptive barcode statistics for the three data partitions analysed in the study.

Statistic	This study	GenBank/Bold	Combined
Individuals	678	562	1240
Species (no. unique sp.)	172 (91)	238 (157)	329
Mean individuals per sp. (range)	3.9 (1–12)	2.4 (1–42)	3.8
Singletons	20	125	97
Genera	45	63	65
Mean sampling events per sp. (range)	2.32 (1–8)	-	-
Mean seq. length bp (range)	645 (378–651)	639 (441–651)	643 (378–651)
No. barcodes <500 bp	5	1	6
Mean haplotypes per species	1.97 (1–7)	1.61 (1–8)	2.07 (1–10)
Mean intraspecific dist. (range)	0.90% (0–14.7%)	0.86% (0–24.1%)	1.13% (0–24.1%)
Mean smallest interspecific dist. (range)	9.11% (0–23.2%)	8.40% (0–26.0%)	8.06% (0–26.0%)
95% intraspecific var. $\leq$	5.48%	2.13%	6.85%
95% smallest interspecific dist. $\geq$	1.72%	0.00%	0.15%
Prop. intraspecific dist. >1%	19.0%	32.2%	28.3%
Prop. intraspecific dist. >2%	13.5%	5.90%	12.7%

Ranges or subsets are presented in parentheses. Abbreviations: dist. = distance(s); no. = number; prop. = proportion; seq. = sequence; sp. = species; tot. = total; var. = variation. "Combined" refers to data generated in this study combined with collected GenBank/Bold data.
doi:10.1371/journal.pone.0028381.t001

Figure 1. Illustrating the utility of DNA barcodes in biosecurity.
Puntius filamentosus (A) and *P. assimilis* (B) are two species strikingly similar in appearance; morphological differences are especially difficult to discern when these are exported as juveniles. Here, we demonstrate they can be readily separated by DNA barcodes, with the two specimens pictured here differing by a 17.6% divergence in K2P distance for COI.
doi:10.1371/journal.pone.0028381.g001

paraphyly of *Devario* cf. *devario* by *D. devario* EF452866; polyphyly of *Paedocypris carbunculus*; paraphyly of *Puntius stoliczkanus* with polyphyletic *P. ticto*; polyphyly of *R. paviana* with regard to *R. hobelmani* HM224229 and *R. vulgaris* HM224243; polyphyly of *Esomus metallicus*.

Nuclear Data and Unrecognised Diversity

When comparing suitability of COI and RHO as a species level marker in our reduced, matched datasets, the NJ monophyly analysis yielded 98.6% success rate for COI, and 87.8% for RHO.

The rates for the nearest neighbour analyses (*k*-NN) were 99.0% for COI, and 92.2% for RHO. The two genes representing two different genomes produced consistent results, but with the nuclear data performing slightly poorer at discriminating some closely related species. A NJ phylogram of RHO data is presented in Figure S2. Taxa unable to be resolved by RHO include some members of the *Puntius conchonius* group including *P. padamya*, *P. tiantian* and *P. manipurensis*. *Danio albolineatus*/*D. roseus* were also unresolved, as were *Microdevario kubotai*/*M. nana*, plus *Devario* cf. *browni* and other associated undescribed/unidentified *Devario* species. The hybrid *Puntius* clustered close to *P. arulius* in the COI NJ phylogram (Figure S1), while it clustered with *P. denisonii* in the RHO phylogram (Figure S2). This result indeed supports its identification as a hybrid, and potentially identifies the parental species.

In the COI data, divergent lineages (e.g. >3%) were found to be present within several common aquarium species, including: *Danio choprae, D. dangila, D. kyathit, Devario devario, Epalzeorhynchos kalopterus, Microdevario kubotai, Microrasbora rubescens, Puntius assimilis, P. denisonii, P. fasciatus, P. gelius, P. lateristriga, P. stoliczkanus, Rasbora dorsiocellata, R. einthovenii, R. heteromorpha, R. maculata, R. pauciperforata* and *Sundadanio axelrodi*. Some were expected, based on the morphological examination process, to be unrecognised diversity (noted by "sp.", "cf." or "aff."), and some were divergent in the absence of apparent morphological differences (i.e., so-called "cryptic" species). Divergent COI lineages of species sequenced in this study are represented as an NJ phylogram in Figure 4. A numerical summary of some of these is presented in Table 3, where nuclear RHO data were used to explore whether the COI relationships were supported [48]. We find here that when COI splits were large, the RHO distances were also large, albeit on average 9.9× smaller (range 3.8–22.7×). Discrete character states were observed for all species in both genes, but were again fewer at the nuclear locus and also corresponded to lower bootstrap support. Rosenberg's *P* statistic of reciprocal monophyly showed adequate sample sizes for most comparisons, but highlighted where further sampling would be beneficial.
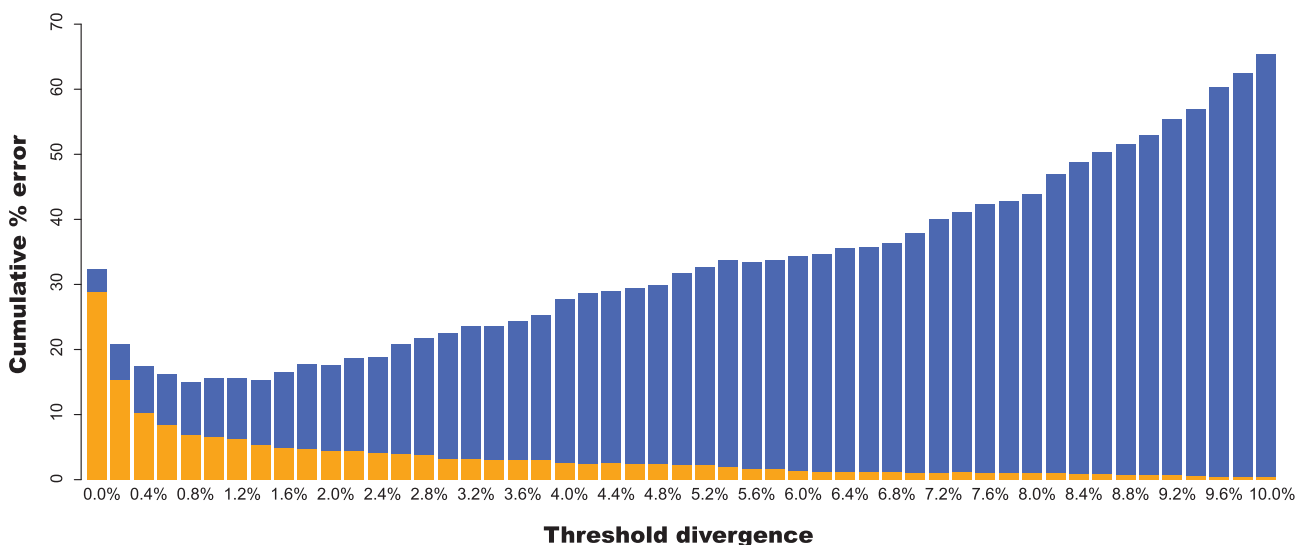

Figure 2. Cumulative error and threshold optimisation. False positive (orange) and false negative (blue) identification error rates summed across a range of distance thresholds from 0–10% in 0.2% increments (combined data). Definition of errors follows Meyer and Paulay [82]. Optimum threshold is 0.8%.
doi:10.1371/journal.pone.0028381.g002

Table 2. Identification percent success rates for each of the five primary analytical methods across three data partitions (with singletons both included and excluded from results), plus optimum threshold values from cumulative error estimation.

Measure	Singletons	This study (%)	GenBank/Bold (%)	Combined (%)
NJ mono.	excl.	96.7 (3.3)	83.5 (16.5)	84.7 (15.3)
	incl.	93.8 (6.2)	64.9 (35.1)	78.1 (21.9)
NJ mono. boot.	excl.	89.7 (10.3)	78.7 (21.3)	74.7 (25.3)
	incl.	87.0 (13.0)	61.2 (38.8)	68.9 (31.1)
k-NN ($k=1$)	excl.	98.9 (1.1)	93.6 (6.4)	94.8 (5.2)
	incl.	96.0 (3.9)	72.8 (27.2)	87.4 (12.6)
GMYC	excl.	94.2 (3.6, 2.1)	72.1 (17.3, 10.5)	82.2 (12.5, 5.3)
	incl.	91.4 (3.5, 5.0)	58.5 (14.1, 27.4)	77.0 (11.7, 11.3)
Bold: 1% thresh.	excl.	93.2 (0.0, 3.2, 3.6)	75.3 (2.5, 12.8, 9.4)	82.9 (1.5, 6.6, 8.9)
	incl.	90.4 (0.0, 6.0, 3.6)	58.5 (5.3, 28.8, 7.3)	76.5 (2.8, 12.5, 8.2)
Bold: opt. thresh.	excl.	93.9 (0.0, 2.4, 3.6)	75.3 (2.5, 12.8, 9.4)	83.4 (1.7, 6.9, 8.0)
	incl.	91.2 (0.0, 5.3, 3.5)	58.5 (5.3, 28.8, 7.3)	76.9 (2.9, 12.0, 7.3)
BCM: 1% thresh.	excl.	94.8 (0.2, 3.2, 1.8)	77.6 (3.4, 12.8, 6.2)	86.7 (2.4, 6.6, 4.2)
	incl.	92.0 (0.1, 6.0, 1.8)	60.3 (6.0, 28.8, 4.8)	79.9 (3.7, 12.5, 3.9)
BCM: opt. thresh.	excl.	95.6 (0.2, 2.4, 1.8)	77.6 (3.4, 12.8, 6.2)	86.5 (2.4, 6.9, 4.2)
	incl.	92.8 (0.1, 5.3, 1.8)	60.3 (6.0, 28.8, 4.8)	79.8 (3.5, 12.9, 3.9)
Opt. thresh. value		1.4	1.0	0.8

Values in parentheses show failure rate broken down into "misidentification", "no identification" and "ambiguous" (BCM and Bold only) respectively. "Combined" refers to data generated in this study combined with collected GenBank/Bold data. Abbreviations: BCM = "best close match"; boot. = bootstrap ($>70\%$); excl. = excluded; incl. = included; mono. = monophyly; opt. = optimum; thresh. = threshold.
doi:10.1371/journal.pone.0028381.t002

Discussion

Sampling

Accurately assigning correct taxonomic names to voucher specimens and barcodes is a critical first step in assembling a useful reference library for non-expert users. Unlike previous studies of regional faunas [86,87], scientific publications covering all taxa likely to be encountered in the aquarium trade were not available. In some cases, reliable guides to local faunas and up-to-date revisions existed, but in other cases such as Indian fishes, little taxonomic research has been conducted since the original descriptions from the early 19th century. Liberal use of the "cf." notation where specimens examined differed from diagnoses in the literature (29 examples), is testament to the uncertainty in identification based on these data.

Our survey of the trade revealed that 24% of species available were not listed in the most recent and thorough reference list for the trade [20], indicating a mismatch between actual availability and published literature. Conversely, many species listed in this reference did not appear to be available at the wholesalers and retailers visited. Some of these discrepancies surely arise from identification and nomenclatural issues, but is otherwise likely due to changing export patterns through different regions and time.

A strong relationship between haplotype diversity and sample frequency was observed, indicating that expanding the reference library will result in the discovery of further genetic variability. In terms of the patterns of trade, we predict that farmed species will have a lower genetic diversity and fewer observed haplotypes than those of wild caught species, which may make them easier to identify with DNA barcodes. Preliminary investigations have suggested that this may well be the case, but due to difficulties obtaining reliable information through the supply chain and problems with establishing independence of samples (i.e., "inde-

pendent" samples may have derived from a single source), these observations should be investigated further.

Identification Success Rates using DNA barcodes

For biosecurity applications, relying upon the names provided by aquarium fish suppliers is likely to be highly inaccurate, and DNA barcoding represents a defensible approach. When we compared our morphological identifications to trade names or names in popular references used by the trade [88], we estimate that up to 25% of cyprinid species could be mislabelled. The DNA barcode library generated in this study provides an ideal tool to test this preliminary observation in more detail and provide a future quantified study of supplier mislabelling in the ornamental industry.

A particular challenge to biosecurity is the steady change in the number and identity of species that are traded. Any useful identification method must be robust to these changes; i.e., sequences from new species in the trade should not be erroneously matched to species with barcodes in the database, while a good identification technique should allow for the re-identification of species that are already represented. We do not present a full assessment of all identification methodologies, but we can here discuss the advantages and disadvantages of the methods covered in our study.

Many barcoding studies employ terminology describing, for example, species forming "cohesive clusters" differentiated from one another by greater interspecific than intraspecific divergence, i.e., the barcode gap of Meyer and Paulay [82]. In our study, we measured clustering in terms of monophyly in NJ phylograms, a tree-based method which performed well on data generated here, but suffered when combined with GenBank/Bold information. This method requires strict monophyly of each species, resulting in a situation where the inclusion of a single misidentified specimen
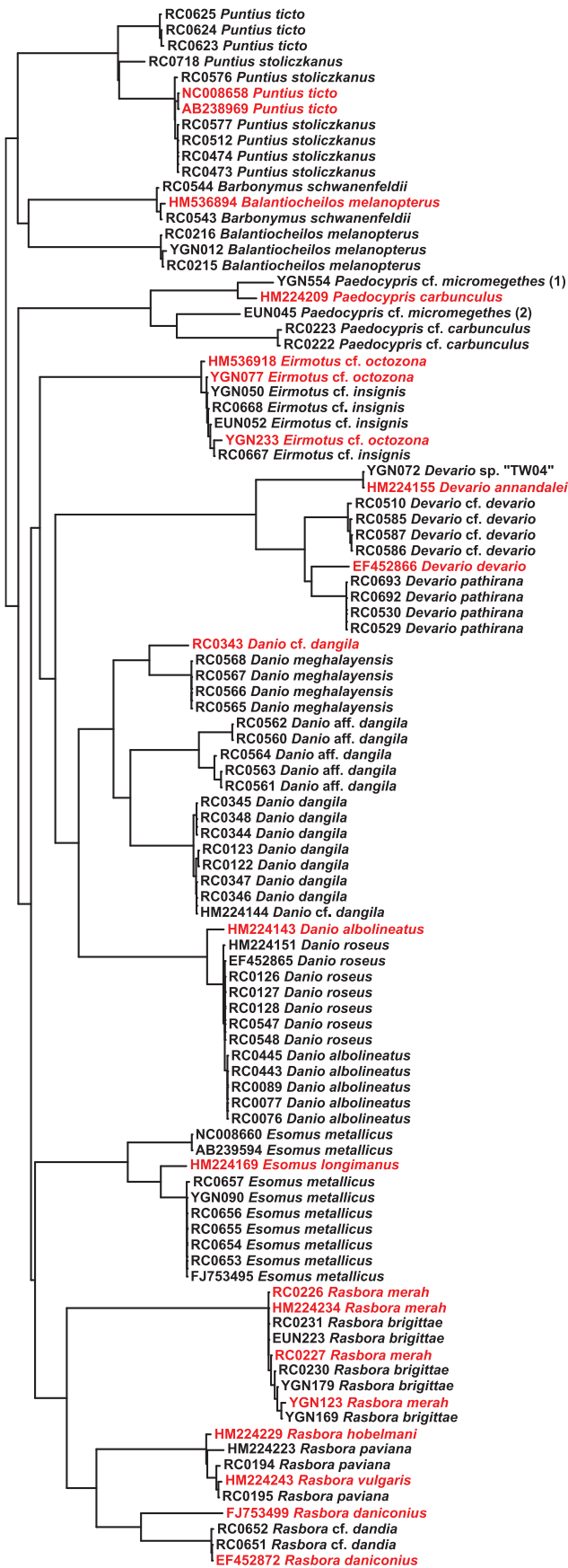
RC0625 *Puntius ticto*
RC0624 *Puntius ticto*
RC0623 *Puntius ticto*
RC0718 *Puntius stoliczkanus*
RC0576 *Puntius stoliczkanus*
NC008658 *Puntius ticto*
AB238969 *Puntius ticto*
RC0577 *Puntius stoliczkanus*
RC0512 *Puntius stoliczkanus*
RC0474 *Puntius stoliczkanus*
RC0473 *Puntius stoliczkanus*
RC0544 *Barbonymus schwanenfeldii*
HM536894 *Balantiocheilos melanopterus*
RC0543 *Barbonymus schwanenfeldii*
RC0216 *Balantiocheilos melanopterus*
YGN012 *Balantiocheilos melanopterus*
RC0215 *Balantiocheilos melanopterus*
YGN554 *Paedocypris* cf. *micromegethes* (1)
HM224209 *Paedocypris carbunculus*
EUN045 *Paedocypris* cf. *micromegethes* (2)
RC0223 *Paedocypris* cf. *carbunculus*
RC0222 *Paedocypris* cf. *carbunculus*
HM536918 *Eirmotus* cf. *octozona*
YGN077 *Eirmotus* cf. *octozona*
YGN050 *Eirmotus* cf. *insignis*
RC0668 *Eirmotus* cf. *insignis*
EUN052 *Eirmotus* cf. *insignis*
YGN233 *Eirmotus* cf. *octozona*
RC0667 *Eirmotus* cf. *insignis*
YGN072 *Devario* sp. "TW04"
HM224155 *Devario annandalei*
RC0510 *Devario* cf. *devario*
RC0585 *Devario* cf. *devario*
RC0587 *Devario* cf. *devario*
RC0586 *Devario* cf. *devario*
EF452866 *Devario devario*
RC0693 *Devario pathirana*
RC0692 *Devario pathirana*
RC0530 *Devario pathirana*
RC0529 *Devario pathirana*
RC0343 *Danio* cf. *dangila*
RC0568 *Danio meghalayensis*
RC0567 *Danio meghalayensis*
RC0566 *Danio meghalayensis*
RC0565 *Danio meghalayensis*
RC0562 *Danio* aff. *dangila*
RC0560 *Danio* aff. *dangila*
RC0564 *Danio* aff. *dangila*
RC0563 *Danio* aff. *dangila*
RC0561 *Danio* aff. *dangila*
RC0345 *Danio dangila*
RC0348 *Danio dangila*
RC0344 *Danio dangila*
RC0123 *Danio dangila*
RC0122 *Danio dangila*
RC0347 *Danio dangila*
RC0346 *Danio dangila*
HM224144 *Danio* cf. *dangila*
HM224143 *Danio albolineatus*
HM224151 *Danio roseus*
EF452865 *Danio roseus*
RC0126 *Danio roseus*
RC0127 *Danio roseus*
RC0128 *Danio roseus*
RC0547 *Danio roseus*
RC0548 *Danio roseus*
RC0445 *Danio albolineatus*
RC0443 *Danio albolineatus*
RC0089 *Danio albolineatus*
RC0077 *Danio albolineatus*
RC0076 *Danio albolineatus*
NC008660 *Esomus metallicus*
AB239594 *Esomus metallicus*
HM224169 *Esomus longimanus*
RC0657 *Esomus metallicus*
YGN090 *Esomus metallicus*
RC0656 *Esomus metallicus*
RC0655 *Esomus metallicus*
RC0654 *Esomus metallicus*
RC0653 *Esomus metallicus*
FJ753495 *Esomus metallicus*
RC0226 *Rasbora merah*
HM224234 *Rasbora merah*
RC0231 *Rasbora brigittae*
EUN223 *Rasbora brigittae*
RC0227 *Rasbora merah*
RC0230 *Rasbora brigittae*
YGN179 *Rasbora brigittae*
YGN123 *Rasbora merah*
YGN169 *Rasbora brigittae*
HM224229 *Rasbora hobelmani*
HM224223 *Rasbora paviana*
RC0194 *Rasbora paviana*
HM224243 *Rasbora vulgaris*
RC0195 *Rasbora paviana*
FJ753499 *Rasbora daniconius*
RC0652 *Rasbora* cf. *dandia*
RC0651 *Rasbora* cf. *dandia*
EF452872 *Rasbora daniconius*

— 1%

Figure 3. Incongruences and inconsistencies in barcode data. This reduced-taxon NJ phylogram highlights cases of haplotype sharing and paraphyly/polyphyly between nominal species. Data generated in this study are prefixed "RC0", "YGN" and "EUN" (otherwise GenBank), with anomalous individuals represented in red.
doi:10.1371/journal.pone.0028381.g003

renders all queries in that species as misidentifications. Although alternative tree-based measures are available (e.g. Ross *et al.* [39]), the use of NJ trees in general is questionable due their method of construction [29,37] and topological uncertainty [37,89]. Furthermore, for a variety of reasons, "good species" may not always be monophyletic at mtDNA loci, so this method may fail to recognise species with either a history of introgression, or young species with large effective population sizes retaining ancestral polymorphisms [49,73,90]. These problems are not resolved through the use of bootstrap values, as we observed a significant reduction in identification success rate when node support was considered (up to 10%); recently divergent sister species on short branches were often not supported, even if they were monophyletic and diagnosable. DNA barcoding aims to maximise congruence between morphological identifications and sequence information while minimising misdiagnosis, but this is seriously undermined when bootstrap support values are included. For the reasons stated above, NJ trees are best avoided as a sole identification method [91], but can be a useful way to visualise and summarise patterns within barcode data.

The BCM and *k*-NN methods do not require reciprocal monophyly of each species, but merely that the nearest neighbour (single closest match) is conspecific. Thus, even when conflicting GenBank/Bold data were included, identification success could still remain high. In cases of a tied closest match, the *k*-NN method ignores this uncertainty and will offer an identification based on majority, while the BCM method reports this as ambiguous. Similarly to NJ, practical difficulties can occur with *k*-NN when identifying a divergent query from an unsampled species or population, as there is no option for a "no identification". This is a serious problem for undersampled datasets, but the BCM and Bold are able to offer a "no identification" result by incorporating a heuristic measure of species membership (a threshold of 1% distance divergence). Despite fundamental criticisms of threshold methods (e.g. variable molecular clock rates between lineages [92]), it at least provides an approximate criterion for separating intraspecific from interspecific variation [91]. In assessing whether the threshold of 1% best-fitted data generated in this study, the analysis of cumulative error demonstrated that error was variable depending on the dataset. However, it did not grossly depart from Bold's 1% threshold, perhaps justifying the use of this metric at least in the cases presented here. When we modified the Bold and BCM methods to employ these revised thresholds, we found slight improvements in the identification success rates. Using the Bold method of identification, all matches within the threshold need to belong to conspecifics, rather than the single closest match (as in BCM and *k*-NN). So like NJ monophyly, the Bold technique is also confounded by even a single misidentified or haplotype sharing specimen in that cluster, and will return an ambiguous result in this situation. This is advantageous when all sources of uncertainty need to be considered, but can lower the number of successful identifications. As a biosecurity tool, it is worth noting that while the method used by Bold performed well, identification rates can be improved further by adopting a method such as BCM with a revised, data-derived threshold.

The GMYC is another method incorporating a measure of species membership (a "no identification"), but rather than an
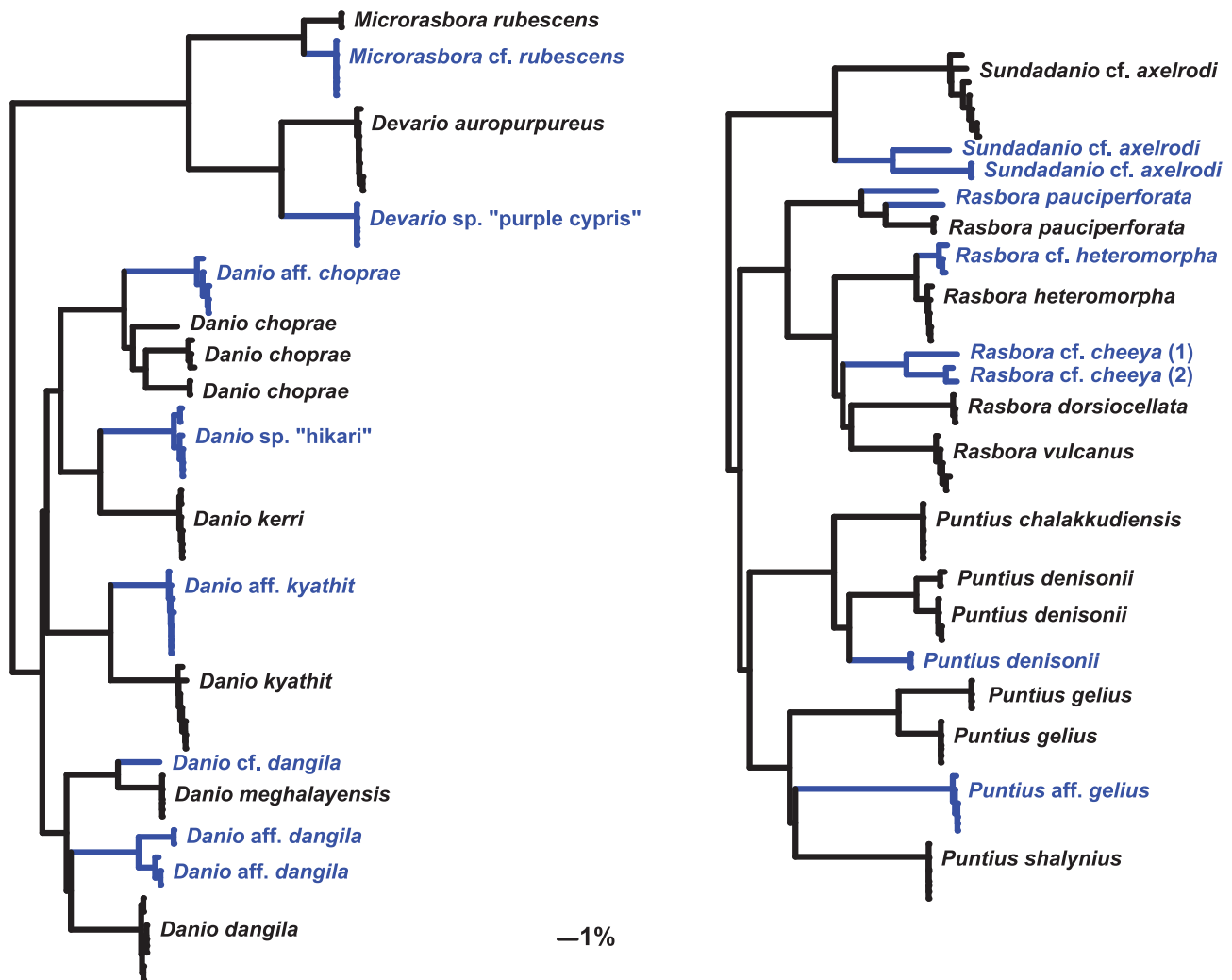
Figure 4. Cryptic and unrecognised species. An NJ phylogram showing deep COI barcode divergences in selected ornamental species. Taxa of interest are highlighted in blue.
doi:10.1371/journal.pone.0028381.g004

arbitrary or generalised cut-off, GMYC employs biological model specification, speciation patterns and coalescent theory in estimating species-like units. As a likelihood based approach, measures of probability and support can be incorporated. Results were highly congruent with the threshold analyses, suggesting the GMYC is picking up the same signal, but optimising the method for all situations may take prior experience or significant trial and error. Another drawback is that the GMYC is not a particularly user friendly technique, requiring many steps and intensive computation, perhaps precluding its use in some border biosecurity applications where fast identifications may be required [9]. Our analysis of 663 haplotypes took approximately five days on a dual processor desktop PC, and although unquantified here, the method also appears sensitive to initial tree-building methodologies.

We reported results with both singleton species included and excluded (Table 2). The exclusion of singletons represents a re-identification scenario where a barcode database is complete and no new species are to be encountered. However, this is an unrealistic assumption here, as the traded cyprinid fishes come from a much larger pool of these fishes not currently available in the trade, and the number of singletons in our trade survey shows

that it is likely that more singletons will be encountered in the future. These singleton species were usually rare/expensive species, contaminants, or bycatch. When singletons comprised a large proportion of the reference database (such as with the GenBank/Bold data), the correct identification rates were significantly reduced for all methods, but GMYC, Bold, and BCM were able to discriminate when a specimen could not be assigned to species. In this respect, the NJ and k-NN methods are poorly performing because they are not sensitive to the presence of singletons in a data set; they will always misidentify a query when a match is not available in the database, and this problem may preclude their use until reference databases are complete.

Incongruence between Morphology, DNA Barcodes, and GenBank/Bold Data

Although few in number, cases of incongruence between barcodes require careful interpretation, especially where the inclusion of GenBank or Bold data result in some common aquarium species becoming ambiguous to distinguish. However, with some background knowledge inferences can be made, and incongruence falls broadly into two categories: taxonomic uncertainty, and conflict due to misidentifications. In the example

Table 3. Exploring unrecognised diversity: undescribed and putative cryptic species were assessed with COI and nuclear RHO data in the context of their closest known congener or conspecifics.

Putative cryptic or unrecognised taxon	Taxon comparison	$n =$	Mean K2P % COI/RHO	No. CAs COI/RHO	Bootstrap % COI/RHO	Rosenberg's P COI/RHO
Danio aff. *choprae*	*D. choprae*	6	7.4/0.5	23/2	100/92.7*	Y/N*
Danio aff. *dangila*	*D. dangila*	7	9.0/1.3	21/10	100/89.9	Y/Y
Danio aff. *kyathit*	*D. kyathit*	6	7.0/1.1	40/7	100/100	Y/Y
Danio sp. "hikari"	*D.* cf. *kerri*	6	8.6/0.6	48/5	100/97.1	Y/Y
Devario sp. "purple cypris"	*D. auropurpureus*	6	8.1/0.6	47/5	100/99.8	Y/Y
Microrasbora cf. *rubescens*	*M. rubescens*	5	3.7/0.5	23/3	100/95.3	N/N
Puntius aff. *gelius*	*P. gelius*	7	17.2/4.1	76/27	100/100	Y/Y
Puntius denisonii	intraspecific	5	7.8/0.4	40/3	100/95.7	N†/N
Rasbora aff. *dorsiocellata* ‡	*R. dorsiocellata*	6	10.9/1.5	46/8	100/82.5	Y/Y
Rasbora cf. *heteromorpha*	*R. heteromorpha*	7	2.2/0.2	11/1	100/18.1	Y/N
Sundadanio cf. *axelrodi*	intraspecific	10	13.8/2.3	42/9	100/99.6	Y/Y

Notes: (*) renders *Danio choprae* paraphyletic; (†) P monophyly significant to the α 10^{-4} level with combined COI data (15 specimens); (‡) species likely described during manuscript preparation as *Brevibora cheeya* [99]. Abbreviations: CA = pure, simple characteristic attribute (i.e., discrete diagnostic character state); Y = Rosenberg's P, significant to $\alpha = 0.05$; N = not significant.
doi:10.1371/journal.pone.0028381.t003

of barcode sharing in *Eirmotus*, despite good quality specimens and the availability of a thorough, modern revision of the genus [93], our morphological identifications were uncertain (Table S1). DNA barcodes from this cluster could belong to either *E. octozona* or *E. insignis*, which is likely the result of these taxonomic/identification problems. Topotypic specimens would be required for a better understanding of the problem. Likewise in the case of *Rasbora brigittae* and *R. merah*, individuals of both species were observed to be inconsistent in diagnostic morphological character states (Table S1). Again, specimens clustering in this group could belong to either species, a finding which certainly warrants further taxonomic investigation. Haplotype sharing between the possibly undescribed *Devario* sp. "TW04" and GenBank *D. annandalei* is likely explained also by uncertainty in our identification of this individual, or the misidentification of the GenBank specimen. Due to the large number of undescribed *Devario* species in Asia, and few modern treatments, identification of many wild caught *Devario* is difficult. The aberrant specimen of *Danio dangila* (RC0343) displayed slight morphological differences to the other *D. dangila*, but with only one individual available, it was conservatively regarded as conspecific (Table S1). A similar observation was made with *Devario* cf. *devario* having divergent barcodes from GenBank *D. devario*, and an inconsistent morphology to that of the published *D. devario* literature. The example of *Danio albolineatus* and *D. roseus* shows a situation where all specimens from the trade are homogeneous and diagnosable, but rendered polyphyletic when data are included from other GenBank populations. This finding is perhaps expected given *D. albolineatus* (*sensu lato*) is a variable species with three synonyms, distributed across much of Southeast Asia [94].

Some examples certainly represent cases of misidentification, with specimens of GenBank "*Puntius ticto*" from the Mekong, grouping closer to *P. stoliczkanus*, a species with which it is often confused [95]. Other examples such as the paraphyly of *Barbonymus schwanenfeldii* by a GenBank *Balantiocheilos melanopterus* individual (HM536894), is probably a case of human error and poor quality control of data, given the marked morphological differences between the two species. Identifications made prior to recently published taxonomic works may also be subject to error,

which may explain GenBank's sequences of *Rasbora daniconius*, a species formerly considered to be widely distributed, but now likely restricted to the Ganges drainage of northern India [96].

So should GenBank data be included in "real life" biosecurity situations? GenBank certainly offers a formidable resource in terms of taxon coverage and extra information, providing sometimes expert-identified wild-caught specimens with published locality data. However, the absence in many cases of preserved vouchers and justified identifications in GenBank undermines its utility for identification purposes [26,36,37]. Bold data are certainly better curated, and with higher quality standards, but are also likely to suffer from misidentified specimens to some degree [37]. Our results do show a decrease in identification success when GenBank data were used, and this was generally due to the higher proportion of singleton species and misidentified specimens, rather than conflicting genetic data *per se*. Realistically though, as long as the practitioner is aware of alternative explanations for patterns, and is also aware of the relative disadvantages with each analytical technique, there is every reason for incorporating these additional data, especially when a smaller dataset is unable to provide a match. No database is immune to errors, but in this study identifications are transparent, and characters, photographs and preserved vouchers can be scrutinised and updated at any time via Bold.

Nuclear Data and Unrecognised Diversity

In terms of corroborating COI and assessing the suitability of a nuclear locus as a species identification tool, the RHO marker was found to be broadly consistent with mitochondrial COI and morphology. Although failing to distinguish a small number of closely related species, RHO served as a useful indicator of interspecific hybridisation in one case (*Puntius* spp. hybrid).

In terms of unrecognised diversity, significant within-species COI diversity was observed in several common ornamental species, and cases of otherwise unreported morphological variation was also recognised. For an exemplar group of aquarium species, and where sufficient numbers of individuals were available, additional support for these divergent COI lineages was assessed with the nuclear RHO marker using character-based analyses,

successfully demonstrating evidence in both genomes. Implications for conservation and sustainable management of fisheries are also apparent here; we find *Puntius denisonii*–a species at risk of over-exploitation [21]–may comprise at least two possibly morphologically cryptic lineages. Although sample sizes were relatively small, these findings certainly warrant further investigation into species limits of these particular taxa. Supporting methods using nuclear data attempt to build on the solely mitochondrial approach by providing congruence with an external dataset [47–49]. This process provides useful reference points, therefore generating further taxonomic questions for closer examination.

Conclusions

Despite the challenge of getting accurate identifications for many species, we have assembled a large database of demonstrably identified fishes and associated barcodes. We believe that DNA barcoding represents a significant move forward in providing identification tools for aquarium species in biosecurity situations. For the small number of cases where barcodes fail to offer unambiguous identifications, additional data such as Web-based images of live specimens, morphological characters, and nuclear loci can be called upon to resolve these problematic specimens. Benefits from barcoding extend beyond a simple quarantine tool, and provide a basis for the generation of accurate and consistent trade statistics, allowing auditing, record keeping and harmonisation between jurisdictions and agencies [97]. Benefits within the ornamental fish industry are also apparent, with accurately identified livestock providing a value added product suitable for export in compliance with international certification or legal standards [13]. Any country vulnerable to aquatic invasions of ornamental species can benefit, with barcode databases offering free and instant access to information. Additional benefits to conservation efforts arise in documenting the ornamental pet trade, with examples such as stock management, traceability, and effective regulation/enforcement of endangered and Cites controlled species [34]. Development of operational databases rely on solid taxonomic foundations [50,82,98], and studies such as these support taxonomy in generating new ideas as well as adding a suite of fine-scale characters and lab protocols, easily accessible via the Web.

Supporting Information

Figure S1 NJ phylogram (COI data) of all specimens (this study plus GenBank/Bold data), in phyloXML SVG (scalable vector graphic) format. Archived version of Figure S1 may require open-source archiving software such as "7-Zip" to unpack. The interactive Web version can be found at http://goo.gl/avNuz. Data including identifiers, sequences, trace files, museum voucher codes and specimen images are accessed via the Bold and GenBank Web sites using URLs embedded in the taxon names. This figure is best viewed with Mozilla Firefox to fully enjoy the benefits of SVG and URL linking. May take up to one minute to load. A scripting "error" may appear in some browsers–this is the browser taking time to render the complex diagram. Phylogram can be saved as a pdf by printing to file using a custom paper size (approximately 3,600 mm height). Links can be opened in a new tab using Ctrl+LeftClick.
(BZ2)

Figure S2 NJ phylogram (reduced RHO data) generated in phyloXML SVG (scalable vector graphic) format. Archived version of Figure S2 may require open-source archiving software such as "7-Zip" to unpack. The interactive Web version can be found at http://goo.gl/h9sY5. Data including identifiers, sequences, trace files, museum voucher codes and specimen images are accessed via the Bold and GenBank Web sites using URLs embedded in the taxon names. This figure is best viewed with Mozilla Firefox to fully enjoy the benefits of SVG and URL linking. May take up to one minute to load. A scripting "error" may appear in some browsers–this is the browser taking time to render the complex diagram. The phylogram can be saved as a pdf by printing to file using a custom paper size (approximately 750 mm height). Links can be opened in a new tab using Ctrl+LeftClick.
(BZ2)

Table S1 Full list of specimens, identifications, morphological characters, comments, and bibliography of samples generated in this study.
(PDF)

Dataset S1 Text file containing all COI sequences used in the study (Fasta format).
(TXT)

Dataset S2 Text file containing all RHO sequences used in the study (Fasta format).
(TXT)

Acknowledgments

Author Contributions

Conceived and designed the experiments: RAC CJ SK KFA RM YY RHC. Performed the experiments: RAC YY. Analyzed the data: RAC SDJB RM. Contributed reagents/materials/analysis tools: KFA RM CJ SK. Wrote the paper: RAC RM KFA SK CJ RHC.

References

1. Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. Journal of Applied Ecology 46: 10–18.

2. Chapin III FS, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM, et al. (2000) Consequences of changing biodiversity. Nature 405: 234–242.

3. Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics 52: 273–288.

4. Dudgeon D, Arthington AH, Gessner MO, Kawabata ZI, Knowler DJ, et al. (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews 81: 163–182.

5. Rahel FJ (2002) Homogenization of freshwater faunas. Annual Review of Ecology and Systematics 33: 291–315.

6. Rahel FJ (2007) Biogeographic barriers, connectivity and homogenization of freshwater faunas: it's a small world after all. Freshwater Biology 52: 696–710.

7. Rubinoff D, Holland BS, San Jose M, Powell JA (2011) Geographic proximity not a prerequisite for invasion: Hawaii not the source of California invasion by light brown apple moth (*Epiphyas postvittana*). PLoS ONE 6: e16361.

8. Wong YT, Meier R, Tan KS (2010) High haplotype variability in established Asian populations of the invasive Caribbean bivalve *Mytilopsis sallei* (Dreissenidae). Biological Invasions 13: 341–348.

9. Armstrong KF, Ball SL (2005) DNA barcodes for biosecurity: invasive species identification. Philosophical Transactions of the Royal Society B: Biological Sciences 360: 1813–1823.

10. Darling JA, Blum MJ (2007) DNA-based methods for monitoring invasive species: a review and prospectus. Biological Invasions 9: 751–765.

11. deWaard JR, Mitchell A, Keena MA, Gopurenko D, Boykin LM, et al. (2010) Towards a global barcode library for *Lymantria* (Lepidoptera: Lymantriinae) tussock moths of biosecurity concern. PLoS ONE 5: e14280.

12. Padilla DK, Williams SL (2004) Beyond ballast water: aquarium and ornamental trades as sources of invasive species in aquatic ecosystems. Frontiers in Ecology and the Environment 2: 131–138.

13. Ploeg A, Bassleer G, Hensen R (2009) Biosecurity in the Ornamental Aquatic Industry. Maarssen: Ornamental Fish International. 148 p.

14. Ng PKL, Chou LM, Lam TJ (1993) The status and impact of introduced freshwater animals in Singapore. Biological Conservation 64: 19–24.

15. Gozlan RE, St-Hilaire S, Feist SW, Martin P, Kent ML (2005) Disease threat to European fish. Nature 435: 1046.

16. Go J, Whittington R (2006) Experimental transmission and virulence of a megalocytivirus (family *Iridoviridae*) of dwarf gourami (*Colisa lalia*) from Asia in Murray cod (*Maccullochella peelii peelii*) in Australia. Aquaculture 258: 140–149.

17. McDowall RM (2004) Shoot first, and then ask questions: a look at aquarium fish imports and invasiveness in New Zealand. New Zealand Journal of Marine and Freshwater Research 38: 503–510.

18. Whittington RJ, Chong R (2007) Global trade in ornamental fish from an Australian perspective: the case for revised import risk analysis and management strategies. Preventive Veterinary Medicine 81: 92–116.

19. Gerstner CL, Ortega H, Sanchez H, Graham DL (2006) Effects of the freshwater aquarium trade on wild fish populations in differentially-fished areas of the Peruvian Amazon. Journal of Fish Biology 68: 862–875.

20. Hensen RR, Ploeg A, Fosså SA (2010) Standard names for freshwater fishes in the Ornamental Aquatic Industry. Maarssen: Ornamental Fish International. 145 p.

21. Raghavan R, Prasad G, Ali PHA, Sujarittanonta L (2007) \Boom and bust fishery" in a biodiversity hotspot - Is the Western Ghats losing its most celebrated native ornamental fish, *Puntius denisonii* Day? Current Science 92: 1671–1672.

22. Ali A, Raghavan R, Dahanukar N (2010) *Puntius denisonii*. In: IUCN 2011. IUCN Red List of Threatened Species. Version 2011.1. URL www.iucnredlist.org. (accessed 24 August 2011).

23. Copp GH, Vilizzi L, Gozlan RE (2010) The demography of introduction pathways, propagule pressure and occurrences of non-native freshwater fish in England. Aquatic Conservation: Marine and Freshwater Ecosystems 20: 595–601.

24. Ploeg A (2008) Invasive species in our industry? OFI Journal 58: 21–25.

25. MAF Biosecurity New Zealand (2011) Import health standard for ornamental fish and marine invertebrates from all countries URL http://www.biosecurity.govt.nz/files/ihs/fisornic.all.pdf. (accessed 24 August 2011).

26. Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. Journal of Fish Biology 74: 329–356.

27. Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270: 313–321.

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

29. Lowenstein JH, Amato G, Kolokotronis SO (2009) The real *maccoyii*: identifying tuna sushi with DNA barcodes - contrasting characteristic attributes and genetic distances. PLoS ONE 4: e7866.

30. Lowenstein JH, Burger J, Jeitner CW, Amato G, Kolokotronis SO, et al. (2010) DNA barcodes reveal species-specific mercury levels in tuna sushi that pose a health risk to consumers. Biology Letters 6: 692–695.

31. Cohen NJ, Deeds JR, Wong ES, Hanner RH, Yancy HF, et al. (2009) Public health response to puffer fish (tetrodotoxin) poisoning from mislabeled product. Journal of Food Protection 72: 810–817.

32. Holmes BH, Steinke D, Ward RD (2009) Identification of shark and ray fins using DNA barcoding. Fisheries Research 95: 280–288.

33. Wong EHK, Shivji MS, Hanner RH (2009) Identifying sharks with DNA barcodes: assessing the utility of a nucleotide diagnostic approach. Molecular Ecology Resources 9: 243–256.

34. Steinke D, Zemlak TS, Hebert PDN (2009) Barcoding Nemo: DNA-based identifications for the ornamental fish trade. PLoS ONE 4: e3600.

35. Nelson JS (2006) Fishes of the World. Hoboken, New Jersey: John Wiley & Sons, Inc. 601 p.

36. Harris J (2003) Can you bank on GenBank? Trends in Ecology and Evolution 18: 317–319.

37. Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Systematic Biology 55: 715–728.

38. Vogler AP, Monaghan MT (2007) Recent advances in DNA taxonomy. Journal of Zoological Systematics and Evolutionary Research 45: 1–10.

39. Ross HA, Murugan S, Li WLS (2008) Testing the reliability of genetic methods of species identification via simulation. Systematic Biology 57: 216–230.

40. Frézal L, Leblois R (2008) Four years of DNA barcoding: current advances and prospects. Infection, Genetics and Evolution 8: 727–736.

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

42. Lim GS, Balke M, Meier R (2011) Determining species boundaries in a world full of rarity: singletons, species delimitation methods. Systematic Biology;10.1093/sysbio/syr030.

43. Fang F, Norén M, Liao TY, Källersjö M, Kullander SO (2009) Molecular phylogenetic interrelationships of the south Asian cyprinid genera *Danio*, *Devario* and *Microrasbora* (Teleostei, Cyprinidae, Danioninae). Zoologica Scripta 38: 237–256.

44. Sevilla RG, Diez A, Norén M, Mouchel O, Jérôme M, et al. (2007) Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome *b* and nuclear rhodopsin genes. Molecular Ecology Notes 7: 730–734.

45. Zemlak TS, Ward RD, Connell AD, Holmes BH, Hebert PDN (2009) DNA barcoding reveals overlooked marine fishes. Molecular Ecology Resources 9: 237–242.

46. Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, et al. (2007) Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution 22: 148–155.

47. Dasmahapatra KK, Mallet J (2006) DNA barcodes: recent successes and future prospects. Heredity 97: 254–255.

48. Dasmahapatra KK, Elias M, Hill RI, Hoffman JI, Mallet J (2010) Mitochondrial DNA barcoding detects some species that are real, and some that are not. Molecular Ecology Resources 10: 264–273.

49. Elias M, Hill RI, Willmott KR, Dasmahapatra KK, Brower AVZ, et al. (2007) Limited performance of DNA barcoding in a diverse community of tropical butteries. Proceedings of the Royal Society B: Biological Sciences 274: 2881–2889.

50. Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. Frontiers in Zoology 7: 1–14.

51. Kottelat M, Freyhof J (2007) Handbook of European freshwater fishes. Cornol, Switzerland: Publications Kottelat. 646 p.

52. Eschmeyer WN (2010) Catalog of Fishes electronic version URL http://research.calacademy.org/ichthyology/catalog/fishcatmain.asp. (accessed 03 March 2011).

53. R Development Core Team (2010) R: A language and environment for statistical computing. Vienna, Austria. URL http://www.r-project.org.

54. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society B: Biological Sciences 360: 1847–1857.

55. 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. Molecular Marine Biology and Biotechnology 3: 294–299.

56. Tang KL, Agnew MK, Hirt MV, Sado T, Schneider LM, et al. (2010) Systematics of the subfamily Danioninae (Teleostei: Cypriniformes: Cyprinidae). Molecular Phylogenetics and Evolution 57: 198–214.

57. Chen WJ, Miya M, Saitoh K, Mayden RL (2008) Phylogenetic utility of two existing and four novel nuclear gene loci in reconstructing tree of life of ray-finned fishes: The order Cypriniformes (Ostariophysi) as a case study. Gene 423: 125–134.

58. Chen WJ, Bonillo C, Lecointre G (2003) Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. Molecular Phylogenetics and Evolution 26: 262–288.

59. Buhay JE (2009) "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. Journal of Crustacean Biology 29: 96–110.

60. Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proceedings of the National Academy of Sciences 105: 13486–13491.

61. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596–1599.

62. Broughton RE, Milam JE, Roe BA (2001) The complete sequence of the zebrafish (*Danio re-rio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. Genome Research 11: 1958–1967.

63. Yokoyama R, Knox BE, Yokoyama S (1995) Rhodopsin from the fish, *Astyanax*: role of tyrosine 261 in the red shift. Investigative Ophthalmology and Visual Science 36: 939–945.

64. Brown SDJ, Collins RA, Boyer S, Malumbres-Olarte J, Lefort MC, et al. (in press) Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Molecular Ecology Resources;10.1111/j.1755-0998.2011.03108.x.

65. Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289–290.

66. Collins RA, Boykin LM, Cruickshank RH, Armstrong KF (in press) Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. Methods in Ecology and Evolution;10.1111/j.2041-210X.2011.00176.x.

67. Srivathsan A, Meier R (2011) On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics;10.1111/j.1096-0031.2011.00370.x.

68. Ross HA, Lento GM, Dalebout ML, Goode M, Ewing G, et al. (2003) DNA surveillance: web-based molecular identification of whales, dolphins, and porpoises. Journal of Heredity 94: 111–114.

69. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406–425.

70. Smits SA, Ouverney CC (2010) jsPhyloSVG: a javascript library for visualizing interactive and vector-based phylogenetic trees on the Web. PLoS ONE 5: e12267.

71. Han MV, Zmasek CM (2009) phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics 10: 356.

72. Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42: 182–192.

73. Austerlitz F, David O, Schaeffer B, Bleakley K, Olteanu M, et al. (2009) DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. BMC Bioinformatics 10: S10.

74. Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, et al. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic Biology 55: 595–609.

75. Monaghan MT, Wild R, Elliot M, Fujisawa T, Balke M, et al. (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. Systematic Biology 58: 298–311.

76. Glez-Peña D, G_omez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D (2010) ALTER: program-oriented conversion of DNA and protein alignments. Nucleic Acids Research 38: W14–W18.

77. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7: 214.

78. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS Biology 4: 699–710.

79. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52: 696–704.

80. Posada D (2008) jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25: 1253–1256.

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

82. Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biology 3: 2229–2238.

83. Sarkar IN, Planet PJ, DeSalle R (2008) CAOS software for use in character-based DNA barcoding. Molecular Ecology Resources 8: 1256–1259.

84. Rosenberg NA (2007) Statistical tests for taxonomic distinctiveness from observations of monophyly. Evolution 61: 317–323.

85. Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the ''barcoding gap'' and leads to misidentification. Systematic Biology 57: 809–813.

86. Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, et al. (2008) Identifying Canadian freshwater fishes through DNA barcodes. PLoS ONE 3: e2490.

87. Valdez-Moreno M, Ivanova NV, Elías-Gutiérrez M, Contreras-Balderas S, Hebert PDN (2009) Probing diversity in freshwater fishes from Mexico and Guatemala with DNA barcodes. Journal of Fish Biology 74: 377–402.

88. Baensch HA, Fischer GW (2007) Aquarium atlas photo index 1–5. Melle, Germany: Mergus Verlag GmbH, third edition. 1211 p.

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

90. Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology, Evolution, and Systematics 34: 397–423.

91. Meier R (2008) DNA Sequences in Taxonomy: Opportunities and Challenges. In: Wheeler QD, ed. The New Taxonomy. New York: CRC Press. pp 95–127.

92. Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular Ecology 18: 4541–4550.

93. Tan HH, Kottelat M (2008) Revision of the cyprinid fish genus Eirmotus, with description of three new species from Sumatra and Borneo. Raffles Bulletin of Zoology 56: 423–433.

94. Fang F, Kottelat M (2000) Danio roseus, a new species from the Mekong basin in northeastern Thailand and northwestern Laos (Teleostei: Cyprinidae). Ichthyological Exploration of Freshwaters 11: 149–154.

95. Linthoingambi I, Vishwanath W (2007) Two new fish species of the genus Puntius Hamilton (Cyprinidae) from Manipur, India, with notes on P. ticto (Hamilton) and P. stoliczkanus (Day). Zootaxa 1450: 45–56.

96. Silva A, Maduwage K, Pethiyagoda R (2010) A review of the genus Rasbora in Sri Lanka, with description of two new species (Teleostei: Cyprinidae). Ichthyological Exploration of Freshwaters 21: 27–50.

97. Gerson H, Cudmore B, Mandrak NE, Coote LD, Farr K, et al. (2008) Monitoring international wildlife trade with coded species data. Conservation Biology 22: 4–7.

98. Dincă V, Zakharov EV, Hebert PDN, Vila R (2011) Complete DNA barcode reference library for a country's buttery fauna reveals high performance for temperate Europe. Proceedings of the Royal Society B: Biological Sciences 278: 347–355.

99. Liao TY, Tan HH (2011) Brevibora cheeya, a new species of cyprinid fish from Malay Peninsula and Sumatra. The Raffles Bulletin of Zoology 59: 77–82.

THESIS CONCLUSIONS

_____

When DNA barcoding was proposed ten years ago as a means for identifying species based on DNA sequences, few biologists imagined how and for what purpose DNA barcodes would be used 10 years later. Two of the unanticipated uses were monitoring invasive species and identifying such species via "eDNA" obtained from environmental samples. Ten years ago, the problems generated by invasive species were rarely discussed and the sequencing technologies for studying eDNA were either not available or too expensive. In my thesis, I explore whether an effective DNA barcode database for ornamental fish can be built. This database will then be available for use in eDNA studies.

In chapter II & IV, I explored how many species were in the ornamental fish trade and then determine whether they have DNA barcodes in Genbank or BOLD. I found the databases to be lacking because 3,453 of the 4,679 recorded species were not present in GenBank and most of them were probably also not in BOLD although the exact species coverage in this database was confidential. In a way, Chapter IV establishes the size of the target if one wanted to build a comprehensive DNA barcode database for ornamental fish. By sequencing 334 new species, I made a significant contribution toward this goal and argued that barcoding fish from the trade may be the fastest way to make progress in the FISH-BOLD project that aims to provide DNA barcodes for all fish species.

Chapter III of my thesis draws heavily on sequences from the ornamental fish trade, but the focus was on testing how difficult it was to obtain a near-complete barcode database for a relatively small region. I found that this seemingly easy task was difficult to complete despite the small size of Singapore and the good fish tissue holdings in the national museum. Based on those fishes for which I can obtain barcodes, I showed that COI can be effective for identifying the fish species in Singapore's freshwater water systems. This applies equally to the native and the non-native species. As in chapter IV, I compared different methods for species identifications and found BRONX to be the most effective. BRONX can discriminate closely related species and reduce cases of ambiguous identifications.

In Chapter III, I also explored whether DNA barcodes can be effective at a regional scale while in Chapter V, my New Zealand collaborators and I tested whether COI barcodes can be effectively used to identify a relatively dense sample of cyprinid species in the ornamental trade; i.e., I focused on testing DNA barcodes for a taxonomic group. Overall, the identification success rates were again similar to what I found for the global and the Singapore database.

In Chapter IV, I explored whether COI barcodes were effective as an identification tool. I compared the efficiency of COI in identifying aquarium fish and other fish taxa, and find no significant differences between ornamental trade fish and other fish species in GenBank.

Besides making a significant contribution to the species coverage in existing databases, I created the first easily accessible ornamental fish image database with high quality voucher images to supplement the COI sequences. In addition, I explored different analysis strategies and determine that BRONX may be the preferred analysis tool for query identification based on DNA barcodes.

Overall, I found consistently that the identification efficiency of DNA barcodes in fish ranges from ca. 85%-to 95% regardless of whether I used a regional, taxonomic, or global database. I also showed that a rapid alignment-free approach to DNA barcoding can yield highly accurate species identifications. This is particularly important given that with the increased use of Next Generation Sequencing technologies, datasets will become larger and different types of complex environmental DNA samples will have to be analyzed. The COI and image database created in my research will surely become valuable when investigating species introductions. However, nearly two thirds of all aquarium fish remain to be barcoded and I conclude that a concerted, international effort will be needed to achieve good species coverage for aquarium fish.

End of Thesis