Mitochondrial DNA, October 2010; 21(S1): 1–7

informa
healthcare

RESEARCH PAPER

Five years of FISH-BOL: Brief status report

SVEN BECKER, ROBERT HANNER, & DIRK STEINKE

Department of Integrative Biology, Biodiversity Institute of Ontario, University of Guelph, Guelph, Ont., Canada

(*Received 3 September 2010; revised 24 September 2010; accepted 18 October 2010*)

Abstract
The Fish Barcode of Life Initiative (FISH-BOL) is a concerted global research project launched in 2005, with the goal to collect and assemble standardized DNA barcode sequences and associated voucher provenance data in a curated reference sequence library to aid the molecular identification of all fish species. This article is a detailed progress report (July 2010) on the number of fish species that have been assigned a DNA barcode. Of the approximately 31,000 currently known fish species, 25% have been processed successfully, with at least one species from 89% of all families barcoded; in this report we give a progress overview by taxonomy and geographic region. Using standard analytical protocols, differences in the barcoding completion rate between orders and families are observed, suggesting a potential PCR amplification bias. Overall, between 3 and 9% of the species analyzed failed to yield a "BARCODE compliant" sequence, depending upon how the data are filtered. When species with only a single representative specimen are included, the failure rate was 9%. This might derive from several sources such as mismatched primers and degraded DNA templates. In an attempt to account for the latter, when the analysis is restricted to species with at least two specimens examined, the observed failure rate is significantly lower (3%), suggesting that template quality is a source of concern for FISH-BOL. We, therefore, conclude that using a standard protocol with several specimens per species and PCR primer cocktails is an efficient and successful approach because failures were evenly distributed among orders and families. Only six orders with low species numbers (Pristiformes, Torpediniformes, Albuliformes, Batrachoidiformes, Gobiesociformes, and Petromyzontiformes) showed failure rates between 10 and 33%. Besides outlining an overarching approach for FISH-BOL data curation, the goal of the present article is to give guidance in directing sampling campaigns toward neglected or underrepresented families in order to complete the FISH-BOL campaign most efficiently.

Keywords: *DNA barcode, cytochrome* c *oxidase subunit I, Fish Barcode of Life Initiative, Barcode of Life Data Systems*

Introduction

DNA barcoding is a molecular method for species-level identification of eukaryotic organisms based on the analysis of short, standardized gene sequences (Hebert et al. 2003). In most animals, the 5′ region of the cytochrome *c* oxidase subunit I (COI) in the mitochondrial genome has been used as the target sequence, which encodes part of the terminal enzyme of the respiratory chain in mitochondria (Folmer et al. 1994). Barcoding provides a rapid and cost-effective method for the identification of eukaryotes and is revolutionizing the application of taxonomy for taxa with validated data sets (e.g. fishes). DNA barcodes have provided new perspectives in ecology, diversity, and the taxonomy of fishes from many geographic regions; for example, Canada (Hubert et al. 2008; Steinke et al. 2009a), Central America (Valdez-Moreno et al. 2009), and Australia (Ward et al. 2005, 2008). Fishes constitute a morphologically diverse group of vertebrates that exhibit deep phenotypic changes during development. In this context, the identification of all fish species is challenging and practically impossible, when based on morphology alone.

DNA barcoding can be used to identify fish species from whole fish, fillets, fins, fragments, juveniles, larvae, eggs, or any properly preserved tissue available. The ability of barcoding to provide species assignments also has important implications for the discovery of cryptic species (Hebert et al. 2004; Sriwattanarothai et al. 2010). Applications of DNA barcoding are

Correspondence: S. Becker, Biodiversity Institute of Ontario, University of Guelph, 50 Stone Road East, Guelph, Ont., Canada N1G 2W1. Tel: + 1 519 824 4120 ext. 53479. Fax + 1 519 767 1656. E-mail: svenbecker@gmx.ch

ISSN 1940-1736 print/ISSN 1940-1744 online
DOI: 10.3109/19401736.2010.535528

emerging in the fields of fish conservation (e.g. Holmes et al. 2009; Steinke et al. 2009b), and management aspects such as quota, by-catch monitoring and sustainable fisheries (Rasmussen et al. 2009). In the fields of food safety and consumer fraud, DNA barcoding has demonstrated that 25% of fish samples from markets and restaurants in New York (USA) and Toronto (Canada) were mislabeled or substituted (Wong and Hanner 2008). DNA barcoding can also be applied successfully to cooked or processed fish (Smith et al. 2008), grilled or deep-fried fillets (Wong and Hanner 2008), and boiled samples (Cohen et al. 2009). Samples with degraded DNA, due to a combination of high pressure and temperature as used in canning, require the use of shorter fragments, the so-called mini-barcodes (Hajibabaei et al. 2006; Meusnier et al. 2008; Rasmussen et al. 2009; Ward et al. 2009). Given its utility, barcoding is being used by the US Food and Drug Administration as a replacement for the time-consuming technique of protein isoelectric focusing for fish and fish product identification (Yancy et al. 2008).

The Fish Barcode of Life Initiative (FISH-BOL; http://www.fishbol.org; Ward et al. 2009) is a concerted global effort to aid assembly of a standardized reference sequence library for all fish species; one that is derived from voucher specimens with authoritative taxonomic identifications. The benefits of barcoding fishes include facilitating species identification for all potential users, including taxonomists; highlighting specimens that represent a range expansion of known species; flagging previously unrecognized species; and, perhaps most importantly, enabling identifications where traditional methods are not applicable. FISH-BOL has the primary goal of gathering DNA barcode records for all the world's fishes, about 31,000 species (Ward et al. 2009; Eschmeyer 2010).

The present brief status report provides a taxonomic and geographic overview of the FISH-BOL progress since its inception in 2005, and it aims to direct ongoing or future sampling campaigns toward neglected or underrepresented taxa or geographic regions. The overview also discusses current laboratory protocols and aims to provide an outline for the concerted curation of barcode data pertaining to fishes. The data presented in this article were taken from the FISH-BOL web site (http://www.fishbol.org), version July 2010. Additional data from the taxonomy browser of the Barcode of Life Data Systems (BOLD; http://www.boldsystems.org) (Ratnasingham and Hebert 2007) were used to calculate success rates.

Taxonomic coverage and failure rates

The Catalog of Fishes currently recognizes about 31,000 species, distributed among six classes, 62 orders, and 540 families (Eschmeyer 2010). Overall, as of July 2010, about 7800 (25%) species have been characterized with at least one DNA barcode (Table I), and associated provenance data for these species have been deposited in BOLD. For an overview at the ordinal level (Table I), we counted species that failed to yield a "BARCODE compliant" sequence (see BARCODE data standard at: http://www.barcoding.si.edu/PDF/DWG_data_standards-final.pdf) after attempted analysis of at least two specimens, which resulted in a low failure rate of 3% (241 species) of all fish species so far attempted for DNA barcoding. A more detailed overview at the taxonomic family level is given in Table S1 in the supplementary material. At least one species from most families (89%) has been barcoded, and by extending the analysis to all specimens with at least one (failed) specimen per species the failure rate increases to 8.6% (734 species) of all species processed to date (see Table S1). Since single specimen failures cannot be readily attributed to any specific error such as problematic tissue quality, low DNA yield, or poor PCR primer hybridization, a more detailed analysis is needed to see whether the observed maximum failure rate is significantly higher than 3%. There are no orders or families that fail systematically (Table I and Table S1), and failures are relatively evenly distributed among these taxonomic levels. However, there are a few notable exceptions with failure rates above 10% in some orders (Table I): Pristiformes and Torpediniformes (Elasmobranchii); Albuliformes, Batrachoidiformes, and Gobiesociformes (Actinopterygii); Petromyzontiformes (Cephalaspidomorphi).

In a small percentage of fishes, the usual gene marker COI used for DNA barcoding may not enable rigorous species discrimination (Hubert et al. 2008), including some sharks (Wong et al. 2009). Thus, longer sequences or alternative markers might be required for complete resolution of all species. This demonstrates the need to develop new and more specific PCR oligonucleotide primers, and also to conduct research into alternative target gene markers. This will be essential to complete and validate the existing database; indeed, the major issue that hinders DNA barcoding as a taxonomic tool in broad applications is the completion (Ekrem et al. 2007) and validation (Dawnay et al. 2007) of the database. Although there is the possibility of taxon-specific PCR primer mismatches with failures more frequent in certain taxa, our results demonstrate that the protocols currently used to retrieve COI DNA barcodes (see below) are sufficient for most fishes and that the primary constraint for FISH-BOL lies in gaining access to well-identified specimens and associated tissues preserved with good quality DNA.

Geographic coverage

The primary work of the FISH-BOL campaign is led by 10 collaborative research groups who have responsibility for overseeing sampling, identification,

Table I. Number of described fish species, species successfully DNA barcoded, and species failed for barcoding; overview by class and order.

Class/order	Species	Species barcoded	Species barcoded (%)	Species failed*	Species failed* (%)
Elasmobranchii					
Carcharhiniformes	270	154	57.0	2	1.3
Heterodontiformes	9	5	55.6	0	0.0
Hexanchiformes	5	5	100.0	0	0.0
Lamniformes	16	15	93.8	0	0.0
Orectolobiformes	41	23	56.1	1	4.2
Pristiformes	7	6	85.7	1	14.3
Pristiophoriformes	6	3	50.0	0	0.0
Rajiformes	550	239	43.5	14	5.5
Squaliformes	124	59	47.6	0	0.0
Squatiniformes	22	13	59.1	0	0.0
Torpediniformes	64	17	26.6	5	22.7
Actinopterygii					
Acipenseriformes	30	21	70.0	0	0.0
Albuliformes	10	4	40.0	2	33.3
Amiiformes	1	1	100.0	0	0.0
Anguilliformes	892	181	20.3	5	2.7
Ateleopodiformes	13	2	15.4	0	0.0
Atheriniformes	326	55	16.9	0	0.0
Aulopiformes	254	102	40.2	3	2.9
Batrachoidiformes	81	14	17.3	4	22.2
Beloniformes	267	42	15.7	4	8.7
Beryciformes	159	70	44.0	1	1.4
Cetomimiformes	34	5	14.7	0	0.0
Characiformes	1919	281	14.6	12	4.1
Clupeiformes	386	96	24.9	7	6.8
Cypriniformes	3819	467	12.2	16	3.3
Cyprinodontiformes	1189	116	9.8	5	4.1
Elopiformes	8	6	75.0	0	0.0
Esociformes	13	10	76.9	0	0.0
Gadiformes	608	198	32.6	9	4.3
Gasterosteiformes	27	10	37.0	0	0.0
Gobiesociformes	158	16	10.1	4	20.0
Gonorynchiformes	37	7	18.9	0	0.0
Gymnotiformes	158	16	10.1	0	0.0
Lampriformes	24	18	75.0	0	0.0
Lepisosteiformes	7	6	85.7	0	0.0
Lophiiformes	327	69	21.1	5	6.8
Mugiliformes	81	29	35.8	0	0.0
Myctophiformes	254	120	47.2	1	0.8
Notacanthiformes	28	10	35.7	1	9.1
Ophidiiformes	491	71	14.5	3	4.1
Osmeriformes	318	95	29.9	2	2.1
Osteoglossiformes	230	40	17.4	0	0.0
Perciformes	10,461	3445	32.9	91	2.6
Percopsiformes	9	6	66.7	0	0.0
Pleuronectiformes	766	231	30.2	8	3.3
Polymixiiformes	10	3	30.0	0	0.0
Polypteriformes	18	3	16.7	0	0.0
Saccopharyngiformes	29	5	17.2	0	0.0
Salmoniformes	206	29	14.1	0	0.0
Scorpaeniformes	1561	500	32.0	10	2.0
Siluriformes	3377	360	10.7	13	3.5
Stephanoberyciformes	44	15	34.1	0	0.0
Stomiiformes	412	98	23.8	0	0.0
Synbranchiformes	109	7	6.4	0	0.0
Syngnathiformes	318	95	29.9	0	0.0
Tetraodontiformes	431	193	44.8	3	1.5
Zeiformes	33	23	69.7	0	0.0
Myxini					
Myxiniformes	74	13	17.6	1	7.1
Cephalaspidomorphi					
Petromyzontiformes	42	24	57.1	6	20.0
Holocephali					
Chimaeriformes	46	27	58.7	2	6.9

Table I – *continued*

Class/order	Species	Species barcoded	Species barcoded (%)	Species failed*	Species failed* (%)
Sarcopterygii					
Coelacanthiformes	2	0	0.0	0	0.0
Ceratodontiformes	1	1	100.0	0	0.0
Lepidosireniformes	8	1	12.5	0	0.0
Total	31,220	7796	25.0	241	3.0

*Of all species tested with at least two specimens.

and DNA barcoding of the fish faunas in their geographic region. Species lists associated with 19 marine and seven inland Food and Agriculture Organization (FAO) statistical areas provided an organizational framework for these regional teams, with an initial goal of sampling five specimens from each species across each area. For certain species with broad geographic distribution, as many as 25 specimens were assumed to be sequenced under this scenario.

The current progress detailed by FISH-BOL region is presented in Table II. In addition to the 10 collaborative research group regions, we have added the two polar regions since there has been an increased sampling effort driven by the third International Polar year 2007–2009, and two Census of Marine Life projects (ArcOD). These efforts are reflected in a high coverage for the Arctic (74%) and Antarctic (50%). Other regions (e.g. Australia, Meso-America, North-America, and Oceania) show good progress with coverage near or above 20%. However, extremely species-rich regions such as Asia, South America, and Africa show lower progress. This might be explained with an observed bias toward the processing of marine species, because of the 7800 species recorded as barcoded in FISH-BOL about 5700 (73.1%) are marine. In turn, a large proportion of species in the regions with lower completion rates are indeed freshwater fishes. Most new described fish taxa from the past 10 years are freshwater species from South America and southeastern Asia (Eschmeyer 2010).

Table II. DNA barcoding progress overview by FISH-BOL region.

Region	Species	Species barcoded	Progress (%)
Arctic	240*	178	74.2
Antarctica	325†	162	49.8
Africa	8980	1216	13.5
Australia	8623	2449	28.4
Europe	2028	391	19.3
India	11,023	1918	17.4
Mesoamerica	7677	1713	22.3
North America	8112	2239	27.6
Northeast Asia	10,414	917	8.8
Oceania	5702	1394	24.4
South America	8981	1003	11.2
Southeast Asia	12,140	2019	16.6

*See Mecklenburg et al. (2010); †SCAR-Marine Biodiversity Information Network.

Therefore, sampling in the coming years should focus on the collection of freshwater species in Africa, Asia, and South America. The overview in this article will also make it possible to direct ongoing or future sampling campaigns toward neglected or under-represented orders and families.

DNA barcoding protocols

Most (88%) of the DNA barcode sequences for FISH-BOL have been generated at the Biodiversity Institute of Ontario (BIO) in Canada, with standardized protocols for DNA extraction, PCR, and sequencing. DNA is usually extracted from muscle tissue using an automated glass fiber protocol (Ivanova et al. 2006). PCR conditions are optimized with low oligonucleotide primer and dNTP concentrations to make additional PCR product purification unnecessary before sequencing.

While there are a number of conventional PCR primer pairs that can be used for successful DNA barcoding of fishes (see Table III), the routine strategy at BIO has been to use a fish primer cocktail (C_FishF1t1–C_FishR1t1; Ivanova et al. 2007) in a first PCR round, and a mammal cocktail (C_VF1LFt1–C_VR1LRt1; Ivanova et al. 2007) in an optional second (or third) attempt. Almost all (99.5%) of the fish amplicons generated for sequencing at BIO were generated with one (or both) of these two primer cocktails, in which the single primers have M13 tails (Messing 1983) to enable standardized (forward and reverse) bidirectional sequencing reactions with the M13 primer pair (Table III).

Elevated failure rates in some orders (e.g. Albuliformes; Table I) might call for the design of new and more specific PCR oligonucleotide primers. However, the established PCR strategy with fish and mammal primer cocktails in different PCR rounds has been successful, and failure rates are high in a few orders with low species numbers only (Table I). Poor tissue preservation, low or high DNA yield, or the presence of so-called PCR inhibitors might explain many of the 241 failed species (with at least two specimens processed). Since we have no data to disambiguate these assumptions, this paves the way for a deeper investigation of failures that hinder faster success of the FISH-BOL enterprise.

Table III. PCR and sequencing primers for fish DNA barcoding.

Name	Cocktail name/5′ − 3′ sequence	Reference
	Primers without M13 tails	
FishF1	TCAACCAACCACAAAGACATTGGCAC	Ward et al. (2005)
FishF2	TCGACTAATCATAAAGATATCGGCAC	Ward et al. (2005)
FishR1	TAGACTTCTGGGTGGCCAAAGAATCA	Ward et al. (2005)
FishR2	ACTTCAGGGTGACCGAAGAATCAGAA	Ward et al. (2005)
Fish-BCH	ACTTCYGGGTGRCCRAARAATCA	Baldwin et al. (2009)
Fish-BCL	TCAACYAATCAYAAAGATATYGGCAC	Baldwin et al. (2009)
VF1	TTCTCAACCAACCACAAAGACATTGG	Ward et al. (2005)
VF2	TCAACCAACCACAAAGACATTGGCAC	Ivanova et al. (2007)
VR1	TAGACTTCTGGGTGGCCAAAGAATCA	Ward et al. (2005)
	M13-tailed primers	
	C_FishF1t1–C_FishR1t1 (ratio 1:1:1:1)	Ivanova et al. (2007)
VF2_t1	TGTAAAACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC	
FishF2_t1	TGTAAAACGACGGCCAGTCGACTAATCATAAAGATATCGGCAC	
FishR2_t1	CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA	
FR1d_t1	CAGGAAACAGCTATGACACCTCAGGGTGTCCGAARAAYCARAA	
	C_VF1LFt1–C_VR1LRt1 (ratio 1:1:1:3:1:1:1:3)	Ivanova et al. (2007)
LepF1_t1	TGTAAAACGACGGCCAGTATTCAACCAATCATAAAGATATTGG	
VF1_t1	TGTAAAACGACGGCCAGTTCTCAACCAACCACAAAGACATTGG	
VF1d_t1	TGTAAAACGACGGCCAGTTCTCAACCAACCACAARGAYATYGG	
VF1i_t1	TGTAAAACGACGGCCAGTTCTCAACCAACCAIAAIGAIATIGG	
LepR1_t1	CAGGAAACAGCTATGACTAAACTTCTGGATGTCCAAAAAATCA	
VR1d_t1	CAGGAAACAGCTATGACTAGACTTCTGGGTGGCCRAARAAYCA	
VR1_t1	CAGGAAACAGCTATGACTAGACTTCTGGGTGGCCAAAGAATCA	
VR1i_t1	CAGGAAACAGCTATGACTAGACTTCTGGGTGICCIAAIAAICA	
	Sequencing primers for M13-tailed PCR products	
M13F	TGTAAAACGACGGCCAGT	Messing (1983)
M13R	CAGGAAACAGCTATGAC	Messing (1983)

Another possible pitfall in the DNA barcoding process is the amplification of COI pseudogenes. It has been suggested that inserts of mtDNA sequences into the nuclear genome (the so-called NUMTs) are released from selection and prevail as pseudogenes, which might hinder identifications (Thalmann et al. 2004; Sword et al. 2007; Song et al. 2008; Buhay 2009). Rare pseudogenes in fishes (Venkatesh et al. 2006) can usually be identified because they contain stop codons and BOLD has a function to identify and flag such sequences when they are uploaded accidentally.

Generation and curation of DNA barcoding data

Previously, comparative genetic surveys have suggested that freshwater fishes generally exhibit higher levels of inter-population genetic diversity than marine fishes (Ward et al. 1994). This observation seems to be reflected in fish DNA barcoding results, because with the current DNA barcoding methodology it is possible to separate about 98% of probed marine species, and 93% of freshwater species (Ward et al. 2009). Besides these overall very satisfying success rates, and the ones reported in this study, DNA barcoding has the potential to identify cryptic species (Hebert et al. 2004). In this respect, Bucklin et al. (2011) calculated an average retrieval of 2% new species in larger fish DNA barcoding studies, and they extrapolated this rate to about 600 overlooked or cryptic species to await discovery through similar studies. From the 31,000 species currently listed in the Catalog of Fishes (compare Table I), about 4000 have been described new during the past 10 years (2000–2009), with 500 added in 2008 and 300 in 2009 (Eschmeyer 2010).

There have been reports on the inability of DNA barcoding to discriminate very recently radiated species; for example, in Canada since the end of the last ice age about 10,000 years ago (Hubert et al. 2008). Other concerns in relation to the use of fish DNA barcoding for species discrimination include hybridization, regional differentiation in barcode sequences, and shared haplotypes (Hubert et al. 2008; Ward et al. 2009). However, our experience with the accumulation of FISH-BOL data during the past 5 years suggests that initial specimen misidentification appears to be of considerably more concern than complications caused by the genetic mechanisms described above, and the results so far indicate only a minor failure rate due to these issues. Hence, disambiguation and reconciliation of the application of names across collections/institutions is emerging as a primary area of endeavor after sampling. This is somewhat surprising given that the taxonomy of fishes is well advanced compared with most other taxa and

suggests that even specialists have difficulty in applying names under the *status quo*.

The considerations above suggest that ongoing curation of the data resulting from the FISH-BOL campaign will require increasing diligence for the data to be fit-for-use in molecular diagnostic applications. To this end, new tools on BOLD are being developed to assist campaign leaders in communicating with project managers about which specimens under their charge require careful re-examination. The current BOLD 2.5 version allows data managers to flag individual records as outliers (e.g. potentially misidentified), and to alert relevant project managers to address records thus flagged within their respective projects. This approach underscores the extreme importance of retaining voucher specimens (Wheeler 2003). While certain large-bodied taxa will prove problematic in this respect, a thorough documentation of specimen collecting events that includes an e-Voucher (Monk and Baker 2001) and precise Global Positioning System (GPS) coordinates should be considered a bare minimum. Moreover, the community must endeavor to use existing data fields on BOLD (such as who identified the specimen and where the voucher is held) to aid project and campaign managers in resolving apparent outliers and name/sequence cluster conflicts in the FISH-BOL data set. This is critical given the fact that reference specimen misidentification appears to be the single largest factor contributing to errors in the FISH-BOL data set, which calls into question the reliability of other public sequence databases that place rather less emphasis on careful taxonomic identification (Harris 2003).

In the light of the above, the participation of collection managers and taxonomists in the FISH-BOL campaign is crucial if we are to reconcile and disambiguate the application of Linnean names to barcode clusters in order to use barcoding as a simplified method for the taxonomic identification of unknowns. Barring this, a more pragmatic approach might simply involve designating molecular operational taxonomic units (e.g. MOTUs *sensu* Floyd et al. 2002) such that the molecular registration of biodiversity can proceed ahead of traditional taxonomy. To this end, we encourage all who would describe a new species to include a DNA barcode of the type specimen in their routine (Victor 2007; Diaz de Astarloa et al. 2008) and note that integrating new technologies into standard taxonomic practice stands as a major challenge for the discipline (Padial et al. 2010), not barcoding in and of itself. Molecular-assisted alpha taxonomy is aiding the recognition of cryptic diversity and going a long way toward making the products of classical taxonomy accessible to its user community. This in turn will surely benefit the discipline by exposing the critical importance of both collections and the taxonomists who make use of them.

Supplementary Material

Supplementary table available online.

Acknowledgements

The authors are grateful to Julie Stahlhut and Justin Schonfeld (both Biodiversity Institute of Ontario, University of Guelph) for data discussion, strategy development for PCR primer success evaluation, and improvement of an earlier version of the manuscript. Ariel Levitsky helped with data analysis. FISH-BOL has been supported by many individuals and institutions worldwide (see www.fishbol.org) with organizational support primarily derived from the Consortium for the Barcode of Life, and technical support primarily derived from the Canadian Barcode of Life Network and the International Barcode of Life (iBOL) Project, headquartered at the Biodiversity Institute of Ontario (BIO). Finally, the authors wish to acknowledge the many members of staff at BIO who have helped to make FISH-BOL a success.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper. D. S. was supported by funding from the Alfred P. Sloan Foundation to MarBOL. S. B. was supported by funding from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Baldwin CC, Mounts JH, Smith DG, Weigt LA. 2009. Genetic identification and color descriptions of early life-history stages of Belizean *Phaeoptyx* and *Astrapogon* (Teleostei: Apogonidae) with comments on identification of adult *Phaeoptyx*. Zootaxa 2008: 1–22.

Bucklin A, Steinke D, Blanco-Bercial L. 2011. DNA barcoding of marine metazoa. Ann Rev Mar Sci 3:18.1–18.38.

Buhay JE. 2009. "COI-Like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. J Crust Biol 29:96–110.

Cohen NJ, Deeds JR, Wong ES, Hanner R, Yancy HF, White KD, Thompson TM, Wahl M, Pham TD, Guichard FM, Huh I, Austin C, Dizikes G, Gerber SI. 2009. Public health response to puffer fish (tetrodotoxin) poisoning from mislabeled product. J Food Prot 72:810–817.

Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe RS. 2007. Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Sci Int 173:1–6.

Diaz de Astarloa JM, Mabragana E, Hanner R, Figueroa DE. 2008. Morphological and molecular evidence for a new species of longnose skate (Rajiformes: Rajidae: *Dipturus*) from Argentinean waters based on DNA barcoding. Zootaxa 1921:35–46.

Ekrem T, Willassen E, Stur E. 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. Mol Phylogenet Evol 43:530–542.

Eschmeyer WN. 2010. Catalog of fishes. Available at: http://research.calacademy.org/research/ichthyology/catalog/speciesbyfamily.asp

Floyd R, Abebe E, Papert A, Blaxter M. 2002. Molecular barcodes for soil nematode identification. Mol Ecol 11:839–850.

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

Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PDN. 2006. A minimalist barcode can identify a specimen whose DNA is degraded. Mol Ecol Notes 6:959–964.

Harris DJ. 2003. Can you bank on GenBank? Trends Ecol Evol 18: 317–319.

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

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Natl Acad Sci USA 101:14812–14817.

Holmes BH, Steinke D, Ward RD. 2009. Identification of shark and ray fins using DNA barcoding. Fish Res 95:280–288.

Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burridge M, Watkinson D, Dumont P, Curry A, Bentzen P, Zhang J, April J, Bernatchez L. 2008. Identifying Canadian freshwater fishes through DNA barcodes. PLoS ONE 3:e2490.

Ivanova NV, DeWaard JR, Hebert PDN. 2006. An inexpensive, automation-friendly protocol for recovering high-quality DNA. Mol Ecol Notes 6:998–1002.

Ivanova NV, Zemlak TS, Hanner R, Hebert PDN. 2007. Universal primer cocktails for fish DNA barcoding. Mol Ecol Notes 7:544–548.

Mecklenburg CW, Møller PR, Steinke D. 2010. Biodiversity of arctic marine fishes: Taxonomy and zoogeography Mar Biodiv in press.

Messing J. 1983. New M13 vectors for cloning. Meth Enzymol 101: 20–78.

Meusnier I, Singer G, Landry J, Hickey D, Hebert P, Hajibabaei M. 2008. A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9:214.

Monk RR, Baker RJ. 2001. e-Vouchers and the use of digital imagery in Natural History Collections. Museology of the Museum of Texas Tech University 10:1–8.

Padial J, Miralles A, De la Riva I, Vences M. 2010. The integrative future of taxonomy. Front Zool 7:16.

Rasmussen RS, Morrissey MT, Hebert PDN. 2009. DNA barcoding of commercially important salmon and trout species (*Oncorhynchus* and *Salmo*) from North America. J Agric Food Chem 57:8379–8385.

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

Smith PJ, McVeagh MS, Steinke D. 2008. Application of DNA barcoding for the identification of smoked fish products. J Fish Biol 72:464–471.

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 co-amplified. Proc Natl Acad Sci USA 105:13486–13491.

Sriwattanarothai N, Steinke D, Ruenwongsa P, Hanner R, Panijpan B. 2010. Molecular and morphological evidence supports the species status of the Mahachai fighter (*Betta* sp. *Mahachai*) and reveals new species of *Betta* from Thailand. J Fish Biol 77:414–424.

Steinke D, Zemlak TS, Boutillier JA, Hebert PDN. 2009a. DNA barcoding Pacific Canada's fishes. Mar Biol 156:2641–2647.

Steinke D, Zemlak TS, Hebert PDN. 2009b. Barcoding Nemo: DNA-based identifications for the ornamental fish trade. PLoS One 4:e6300.

Sword GA, Senior LB, Gaskin JF, Joern A. 2007. Double trouble for grasshopper molecular systematics: Intra-individual heterogeneity of both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer ribosomal DNA sequences in *Hesperotettix viridis* (Orthoptera; Acrididae). Syst Entomol 32:420–428.

Thalmann O, Hebler J, Poinar HN, Paabo S, Vigilant L. 2004. Unreliable mtDNA data due to nuclear insertions: A cautionary tale from analysis of humans and other great apes. Mol Ecol 13: 321–335.

Valdez-Moreno M, Ivanova NV, Elias-Gutierrez M, Contreras-Balderas S, Hebert PDN. 2009. Probing diversity in freshwater fishes from Mexico and Guatemala with DNA barcodes. J Fish Biol 74:377–402.

Venkatesh B, Dandona N, Brenner S. 2006. *Fugu* genome does not contain mitochondrial pseudogenes. Genomics 87:307–310.

Victor BC. 2007. *Coryphopterus kuna*, a new goby (Perciformes: Gobiidae: Gobiinae) from the western Caribbean, with the identification of the late larval stage and an estimate of the pelagic larval duration. Zootaxa 1526:51–61.

Ward RD, Woodwark M, Skininski DOF. 1994. A comparison of genetic diversity levels in marine, freshwater and anadromous fish. J Fish Biol 44:213–232.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. 2005. DNA barcoding Australia's fish species. Phil Trans R Soc B Biol Sci 360:1847–1857.

Ward RD, Holmes BH, White WT, Last PR. 2008. DNA barcoding Australasian chondrichthyans: Results and potential uses in conservation. Mar Freshw Res 59:57–71.

Ward RD, Hanner R, Hebert PDN. 2009. The campaign to DNA barcode all fishes, FISH-BOL. J Fish Biol 74:329–356.

Wheeler TA. 2003. The role of voucher specimens in validating faunistic and ecological research. Biological Survey of Canada (Terrestrial Arthropods) Document Series No. 9. 21 p. ISBN 0-9689321-2-6.

Wong EHK, Hanner R. 2008. DNA barcoding detects market substitution in North American seafood. Food Res Int 41: 828–837.

Wong E, Shivji M, Hanner R. 2009. Identifying sharks with DNA barcodes: Assessing the utility of a nucleotide diagnostic approach. Mol Ecol Res 9:243–256.

Yancy HF, Zemlak TS, Mason JA, Washington JD, Tenge BJ, Nguyen N-L, Barnett JD, Savary WE, Hill WE, Moore MM, Fry FS, Randolph SC, Rogers PL, Hebert PDN. 2008. The potential use of DNA barcodes in regulatory science: Applications of the regulatory fish encyclopedia. J Food Prot 71: 210–217.