Research Article　　　　　　　　　　　　　　　　　　**Open Access**

Benjamin C. Victor*, Martha Valdez-Moreno, Lourdes Vásquez-Yeomans

Status of DNA Barcoding Coverage for the Tropical Western Atlantic Shorefishes and Reef Fishes

Abstract: Background: Barcode coverage is difficult to assess for large regions due to incomplete species lists, inaccurate identifications, and cryptic diversity. However, as coverage approaches completion, it becomes possible to critically evaluate identifications and validate barcode lineages. We collate the results of the FISH-BOL barcode project and assess coverage for each family of bony shorefishes and reef fishes from the tropical western Atlantic Ocean. **Methodology**: We identify to species the public and private barcode lineages from the region on BOLD, confirming identifications by vouchers, phylogeographic deduction, and the process of elimination. The lineages and BINs are assigned to species from a comprehensive species list for the region. **Results**: We estimate 1029 of 1311 total bony shorefish species in the region are barcoded (78.5%). For reef-associated fishes, 902 of 1083 species are barcoded (83.3%). About 70 of the 181 species not yet barcoded are endemic species from Florida/ Gulf of Mexico or Venezuela, leaving about 90% of the central Caribbean reef fish species barcoded to date. Most species are represented by one barcode lineage, but among the gobioids and blennioids there are many more lineages (BINs) than species, indicating substantial cryptic diversity. **Conclusions**: As barcode coverage for a region approaches completion, a robust assessment of coverage can be made. The reef fish fauna of the tropical western Atlantic now has the highest coverage for a large marine area, from about 80 to 90% depending on definitions and geographic limits.

Keywords: Mitochondrial DNA, coral reefs, Caribbean, taxonomy, biogeography, species list, phylogenetics

───────

*Corresponding author: Benjamin C. Victor, Guy Harvey Research Institute Nova Southeastern University, Ft. Lauderdale, Florida, USA, Email: ben@coralreeffish.com

Martha Valdez-Moreno,　Lourdes Vásquez-Yeomans, Departamento de Ecologia y Sistematica Acuatica, El Colegio de la Frontera Sur, Chetumal, Quintana Roo, Mexico

1 Introduction

The Fish Barcode of Life campaign, FISH-BOL, has reached a general milestone, with about 10,000 species barcoded from more than 100,000 specimens sequenced in BOLD projects (www.fishbol.org) and the data compiled on the barcode of life database, BOLD (www.boldsystems.org) [1-4]. This total is derived from projects in the campaign, and does not include independent GenBank COI sequences, which could perhaps add a few thousand more species. Many, but not all, of these GenBank sequences have been added to the BOLD database (the process has delays), however they are especially difficult to assess since there is virtually no quality control on sequences or identifications in GenBank. If GenBank metadata are accepted (as in BOLD's taxonomy browser), the degree of confidence for any barcode coverage estimate quickly erodes to impractical, especially for speciose or lesser known taxa. Based on the more rigorous estimates by the FISH-BOL program for actual BOLD projects, about one-third of the known fish species and about 40% of perciform fishes have been barcoded in BOLD to date, as of early 2015 (www.fishbol.org).

As a broad generalization from the FISH-BOL compilations, most of the large-bodied and prominent families of perciform fishes have about 50% barcode coverage, with the small-bodied and speciose gobies and blennioids lower, with about 30% coverage. None of the seven most speciose perciform families (with more than 300 species each) have more than about 60% of their species barcoded (the most complete are labrids and pomacentrids). Only three out of the fifteen of the largest perciform families (with more than 100 species each) have over 80% coverage: i.e. the carangids, lutjanids and percids. The non-perciform orders of Actinopterygii are less well covered: all but one of the fifteen largest orders (with more than 300 species each) have only about a third or fewer of their species barcoded. Note, however, that the accuracy of species identifications is not critically evaluated for any of these broad estimates,

but, in all likelihood (and based on nothing rigorous), the numbers are unlikely to change to a large degree since overestimated coverage by incorrect species IDs would be counterbalanced by additional lineages with an incorrect duplicate species ID, or without a species ID at all.

Regional coverage is a particularly difficult measure to assess for a number of reasons. Many large regions do not have complete species lists compiled for native fishes, and proposed or published species lists are frequently discordant. The most common problem is defining the habitat limits of a fish fauna; for example, marine fishes can include shorefishes, deeper water fishes, euryhaline fishes, and pelagics and those are variously included or excluded from most regional marine species lists. In addition, defining the geographic limits of a fauna is not always simple, since most regions have smaller satellite locations that can be variously included.

A more profound problem for assessing coverage is the accuracy of species identifications: without some high degree of confidence in identifications, the numbers of species barcoded can be an artifact of various contributors' imaginations. An additional problem is unresolved or difficult taxonomy, either in traditional practice or unexpected cryptic diversity, which can account for up to 10% of the species in a list, even in the better known fauna such as the US/Canadian North American freshwater fauna [5], or much more, as in the exceptionally speciose tropical freshwater fish fauna of South America, Africa, and SE Asia. With these impediments, many large (and small) regions cannot be assessed to any degree of certainty; nevertheless, since these regions are also undersampled in the barcode database, we can assume they do not have higher coverage than the examples we discuss here.

At this time, the highest barcode coverage for a large fauna is, unsurprisingly, the northern temperate freshwater fish community. For example, the coverage of European FW fish is about 86% of the approximately 600 species total (Jörg Freyhof, pers. comm.). The highest recorded coverage for a relatively large-scale region is the 98% coverage of the 500 freshwater fishes of the Mediterranean Basin [6]. The US/Canadian North American freshwater fauna coverage is also high, with about 83% of the 900 species barcoded at the last review [5].

Marine fish coverage is typically lower; for example, in contrast to the completion of the Mediterranean FW fishes, only a small fraction of the 650 species comprising the marine fish fauna of the Mediterranean Sea have been barcoded [7-9]. For combined marine and FW fishes, most large regions have less than 50% coverage: one of the more complete examples being Argentina, where almost half of the 1000 species have been barcoded [10,11]. If the surveys are limited to shorefishes (deep-water fishes are seriously undersampled), marine fish coverage is typically less than 50%, and well lower in undersampled regions such as the eastern Atlantic, the Red Sea, and the eastern Pacific. After Europe, Canada may have the highest coverage for combined marine and FW fish species, with more than 200 of the 350 shorefishes of Pacific Canada barcoded [12], 95% of the 200 FW species barcoded [13], and more than half of the 500 species of Atlantic fishes of Canada barcoded (Dirk Steinke, pers. comm.).

There have been extensive efforts at barcoding tropical reef fishes of the broad Indo-Pacific, especially French Polynesia [14], Queensland and Bali [15], Southern Africa [16], and the South China Sea [17]. However, the extreme number of coral-reef species, peaking with as many as 1700 species co-occurring on reefs in the West Papua region of Indonesia [18], still results in barcode coverage of only about half of the total, at best.

The shorefish fauna of the tropical western Atlantic (TWA) has been the most completely barcoded large marine region to date. Fortunately, it is also well studied and inventoried, with a comprehensive guide now available online for all of the shorefishes [19]. This barcoding achievement is mainly the result of three independent large FISH-BOL projects focusing on the region: the ECOSUR group with about 5000 records on BOLD [20], the Smithsonian with about 4000 (21), and the OSF/Victor project with about 3000 records. An important, and novel, factor promoting the reliability of our coverage estimates is the emergent property of positive feedback in identification of a limited set: as we approach completion of coverage for any particular taxonomic group, the identification of the remaining unassigned lineages becomes easier. This is facilitated by two important aids: the process of elimination combined with phylogeographic deduction, i.e. the improved resolution of phylogenetic relationships when most, or almost all, of the potential relatives have been identified and the range of each species is well documented.

The TWA is defined here as the northern tropical and warm subtropical W. Atlantic, excluding Brazil and including S. Florida and the Gulf of Mexico, or what could be called the Greater Caribbean region [22]. The species list for shorefishes of the region varies depending on how many peripheral species are included and the definition of a shorefish, especially considering depth on the continental shelf. In general, however, the number of species ranges up to 1500 [22], and we consider here about 1300 bony shorefish species for the region (excluding elasmobranchs, which number well less than 100 spp.).

The number of "reef species" is subject to a more fluid definition: for the Greater Caribbean, various large-scale surveys list 605 reef species [22], 774 reef species [23], and 885 reef species [24].

The goal of this survey is to introduce a more rigorous evaluation process for assessing coverage and applying it to the TWA shorefishes. With the complete inventory of species and their ranges well established [19,22], and the number of sequences of shorefishes from the region approaching 15,000 (with many well-vouchered), it becomes possible to critically assess species identifications independently- by phylogeographic deduction in combination with the process of elimination, backed up by expert evaluation of voucher metadata, particularly the location and photographs. In all, we estimate the barcode coverage for general shorefishes of the TWA to range up to 80% and the coverage for smaller subsets, such as the strictly coral-reef fishes of the Caribbean Sea, to approach 90%.

Key to the assessment is the categorization of mtDNA lineages, which need to be enumerated and defined by an algorithm, an "operational taxonomic unit"- in BOLD these units are BINs, or Barcode Index Numbers [25]. BINs are not set groups of lineages separated by a certain percentage distance from each other, but a cluster calculated by an algorithm taking into account similarity and connectivity and assessing cluster boundaries. Of course, a cluster of sequences does not a species make, and the taxonomic decision of the relationship of a BIN, or any DNA lineage, to a species is a much more complex analysis, i.e. what is a species? [4]. Nevertheless, the BIN provides the framework for categorizing mtDNA lineages, and, in the large majority of TWA shorefishes at least, proves to correspond one-to-one with known species or suspected sub-species.

2 Methods

A complete species list of the bony shorefishes of the tropical western Atlantic/Greater Caribbean (including the Gulf of Mexico and S. Florida and excluding the Brazilian fauna) was assembled by reviewing taxonomic literature, guidebooks [26-29] and assessing published species lists (22-24). Shorefishes were defined as those associated with the substrate in waters up to 200m depth, excluding mid-water species, but including pelagic fish families. This definition has wide usage among tropical fish taxonomy books [26-29]. Taxonomic validity mostly follows Eschmeyer (2015) [30], with a few practical exceptions, and thus undescribed cryptic lineages were not considered as species.

The DNA lineages present on the barcode database (specifically collected in the TWA) were assigned to the shorefish species list. Almost all TWA lineages (including unique sequences) in the database were assessed, public and private, as well as lineages with no identification data at all. Some private records were made available to us by the owners allowing us to share projects. Such requests were facilitated by the BOLD ID engine showing related private DNA lineages (stripped of metadata and the sequences themselves hidden) on a neighbor-joining tree in the ID-engine procedure initiated by one of our sequences. Only a rare lineage with only private and unshared sequences and not a single nearby relative (from any ocean) would be invisible to us. The BIN application is also very helpful: the BIN summaries on BOLD list private sequences within the BIN (also without any private associated data), as well as the nearest-neighbor BIN code (even if made up of only private sequences), meaning that virtually all barcode lineages, including GenBank downloads and private projects, could be assessed to some degree by our combined research groups.

We did not accept species identifications from ID metadata on BOLD, which are determined by various submitters to the barcode database or GenBank. The general lack of quality control has led to a proliferation of misidentifications on databases, exacerbated by the desire, or even perceived requirement, by contributors to identify specimens to species, often without the expertise to make species-level determinations. This flaw is one of the greatest limitations of specimen-record databases, both for DNA sequences or general occurrence records (such as FishBase or GBIF). BOLD fortunately connects sequences to voucher specimen records, often with photographs. In many cases, the photographs alone contain diagnostic information for species-level identification. Voucher specimens were retained and examined for almost all specimens sequenced in the projects by ECOSUR and the OSF/Victor collection (the majority of BOLD TWA records). DNA lineages without a diagnostic voucher or photograph were assigned to species with varying degrees of certainty based on a combination of three cumulative methods: 1) phylogenetic deduction from the nearest-neighbor species (either from the region, or sibling species from the eastern Pacific), 2) phylogeographic deduction, adding geographic range matching (i.e. an unassigned lineage from a set of locations known to represent the range of a particular candidate species), and 3) the process of elimination; for example, when only one candidate species in a genus is left unassigned and there is one remaining unidentified DNA lineage. With this procedure,

almost all DNA lineages could be assigned a species identification (although for gobioids and blennioids, it often was assignment to a local subspecies/population of a nominal species), with only a small fraction considered by us to be questionably assigned (but certainty was not quantifiable). As complete coverage for any particular taxon is approached, the positive feedback property for identification moves many tentative species assignments to confident species assignments.

3 Results

3.1 Overall Coverage

For the broadest category of bony shorefishes of the Greater Caribbean, we include fishes from the Gulf of Mexico, Florida, and the Caribbean Sea- excluding fishes only found in Guyana to Brazil. All bottom-associated species down to 200 m are included, along with pelagic families that can be found nearshore, such as the carangids, scombrids, ariommatids, nomeids, echeniids, molids, belonids, and hemiramphids. Euryhaline fishes such as the eleotrids, atherinopsids, and atherinids are also included. That list totals 1311 bony shorefish species, and 1029 of them are barcoded to date (78.5%) (Table 1).

Our subset of bony reef-fish species of the region (found above 100 m and associated with reefs) numbers 1083 species, higher than other published reef-fish compilations, since we generally follow taxonomic guidebook format and thus include pelagic fish families with members that can be observed over reefs (including all of the carangids, scombrids, echeneids, belonids, and the elopiformes and albulids); as well as the soft-substrate species that are found in sandbeds and grassbeds around reefs (such as bothids, paralichthyids, cynoglossids, triglids, ogcocephalids, bythitids, chlopsids, and ophichthids); as well as a small subset of the clupeids, atherinids, congrids, and sciaenids that are typically seen near reefs (Table 1). Of our bony reef-fish species total of 1083 species, 902 are barcoded to date (83.3%).

The unbarcoded species are predominantly species endemic to the Gulf of Mexico (GOM) or Venezuela and/or the South American continental shore. Of the 181 unbarcoded species in the reef-fish species list, 30 are Florida/GOM endemics (2.8% of total) and 40 are Venezuelan/S. Caribbean endemics (3.7% of total), leaving only 111 remaining species unbarcoded (10.2%), indicating overall coverage of about 90% for the central reef-fish fauna in the region. The remaining unbarcoded species are mostly rare species (either deeper water, local endemics, or unusual or secretive habitats), those with uncertain taxonomy (or unique holotypes), and a very few regular reef-fish species that have just been overlooked in collections, coincidentally by all research groups concentrating on the region.

3.2 Coverage by Family

Three of the 30 large bony shorefish families (more than 15 species each in the TWA) have been completely barcoded for the region (Table 1). They are prominent commercially important families, comprising the snappers (Lutjanidae), grunts (Haemulidae), and tunas (Scombridae). Several more of the large families are almost complete, with only one or two species missing: i.e. the cardinalfishes (Apogonidae), damselfishes (Pomacentridae), wrasses (Labridae, excluding parrotfishes), porgies (Sparidae), and jacks (Carangidae). The second largest family of shorefishes in the region, the basses (Serranidae, including the groupers of Epinephelidae), is 86% barcoded, mainly missing a few deepwater and/or rare species. The largest fish family is the gobies (Gobiidae), with only 76% of the 134 regional species barcoded, mainly missing a set of deeper-water species and some rare and/or microendemic species that have not been sampled. The reef-fish subset within families is somewhat better covered, with almost all families having higher coverage of their reef-associated members (Table 1).

3.3 Cryptic Diversity

Numerous species of tropical marine fishes, especially reef fishes, show evidence of undescribed cryptic diversity after genetic analyses [4]. The pattern of which species show extensive cryptic diversity is not clear in the vast scale of the Indo-Pacific, where cryptic diversity is frequent among many quite different fish families. However, in the smaller Greater Caribbean region, the pattern is very clear- only families with lower dispersal ability, i.e. benthic brooded eggs and relatively short larval lives, less than about 30 days, break up into cryptic species complexes [4]. A number of cryptic Caribbean species-complexes have been described in recent years (e.g. [31-37]), and many more remain to be explored. This pattern is clearly apparent in the number of BINs associated with a single nominal species in our review: the number of BINs approximates the number of species barcoded in most families of shorefishes (rarely there are fewer BINs than species, see below), with the main exception being a markedly greater number of BINs in the families with benthic eggs and short larval lives, i.e. the Gobiidae and

Table 1. Barcode coverage of tropical western Atlantic bony fishes, by family; in descending order of number of species known for the region. Numbers in bold highlight families that include non-reef species (i.e. #reef species less than total #species).

Family	ALL SPECIES				REEF SPECIES			
	#sp	#barcoded	%barcoded	#BINs	#reef sp.	#barcoded	%barcoded	#BINs
Gobiidae	134	102	76%	157	**130**	**102**	**78%**	**157**
Serranidae	96	82	85%	84	**92**	**81**	**88%**	**84**
Labrisomidae	58	50	86%	88	**57**	**50**	**88%**	**88**
Sciaenidae	54	40	74%	44	**8**	**8**	**100%**	**9**
Chaenopsidae	50	39	78%	85	50	39	78%	85
Ophichthidae	45	29	64%	38	**37**	**27**	**73%**	**26**
Paralichthyidae	35	25	71%	31	**30**	**22**	**73%**	**28**
Carangidae	32	30	94%	30	32	30	94%	30
Congridae	32	22	69%	27	**10**	**8**	**80%**	**11**
Syngnathidae	27	22	81%	22	**25**	**22**	**88%**	**22**
Gobiesocidae	26	16	62%	24	**21**	**16**	**76%**	**24**
Haemulidae	25	25	100%	26	25	25	100%	26
Bythitidae	25	16	64%	20	25	16	64%	20
Scorpaenidae	25	22	88%	25	**24**	**22**	**92%**	**25**
Apogonidae	24	23	96%	28	24	23	96%	28
Batrachoididae	24	12	50%	15	**23**	**12**	**52%**	**15**
Ophidiidae	22	14	64%	17	22	14	64%	17
Labridae (wrasses)	21	20	95%	23	21	20	95%	23
Muraenidae	21	17	81%	17	**20**	**16**	**80%**	**16**
Blenniidae	20	14	70%	10	20	14	70%	10
Lutjanidae	18	18	100%	17	18	18	100%	17
Sparidae	18	16	89%	15	18	16	89%	15
Engraulidae	18	12	67%	12				
Triglidae	18	15	83%	16	18	15	83%	16
Cynoglossidae	17	11	65%	11	**13**	**11**	**85%**	**11**
Dactyloscopidae	17	7	41%	7	17	7	41%	7
Scombridae	16	16	100%	12	16	16	100%	12
Pomacentridae	16	15	94%	17	16	15	94%	17
Scarinae	16	13	81%	13	16	13	81%	13
Tetraodontidae	15	11	73%	12	**14**	**11**	**79%**	**12**
Clupeidae	14	11	79%	12	**7**	**6**	**86%**	**6**
Ogcocephalidae	14	6	43%	5	14	6	43%	5
Atherinopsidae	14	7	50%	8				
Bothidae	14	10	71%	11	**9**	**7**	**78%**	**8**
Opistognathidae	14	7	50%	9	**11**	**7**	**64%**	**9**
Nettastomatidae	13	7	54%	8				
Synodontidae	12	12	100%	15	12	12	100%	15
Holocentridae	12	9	75%	11	12	9	75%	11
Grammatidae	12	8	67%	8	**10**	**8**	**80%**	**8**

continued **Table 1.** Barcode coverage of tropical western Atlantic bony fishes, by family; in descending order of number of species known for the region. Numbers in bold highlight families that include non-reef species (i.e. #reef species less than total #species).

Family	ALL SPECIES				REEF SPECIES			
	#sp	#barcoded	%barcoded	#BINs	#reef sp.	#barcoded	%barcoded	#BINs
Mugilidae	12	11	92%	11	**3**	**3**	**100%**	**3**
Gerreidae	12	12	100%	16	12	12	100%	16
Monacanthidae	10	10	100%	9	10	10	100%	9
Chlopsidae	8	7	88%	8	**7**	**7**	**100%**	**7**
Echeneidae	8	8	100%	8	8	8	100%	8
Nomeidae	8	8	100%	8				
Achiridae	8	4	50%	7	**1**	**1**	**100%**	**2**
Tripterygiidae	8	8	100%	15	8	8	100%	15
Antennariidae	7	6	86%	6	7	6	86%	6
Belonidae	7	7	100%	8	7	7	100%	8
Hemiramphidae	7	6	86%	6				
Chaetodontidae	7	7	100%	7	7	7	100%	7
Pomacanthidae	7	7	100%	7	7	7	100%	5
Eleotridae	7	7	100%	7	7	7	100%	7
Diodontidae	7	7	100%	6	7	7	100%	6
Centropomidae	6	3	50%	3	**1**	**1**	**100%**	**1**
Microdesmidae	6	4	67%	6	6	4	67%	6
Balistidae	6	6	100%	6	6	6	100%	6
Ostraciidae	5	5	100%	6	5	5	100%	6
Albulidae	4	3	75%	3	**3**	**3**	**100%**	**3**
Pristigasteridae	4	1	25%	1				
Priacanthidae	4	4	100%	4	4	4	100%	4
Ariommatidae	4	3	75%	3				
Mullidae	4	4	100%	4	4	4	100%	4
Kyphosidae	4	4	100%	4	4	4	100%	4
Uranoscopidae	4	3	75%	3	4	3	75%	3
Callionymidae	4	3	75%	6	4	3	75%	6
Carapidae	3	1	33%	1	**1**	**1**	**100%**	**1**
Atherinidae	3	3	100%	4	**2**	**2**	**100%**	**2**
Sphyraenidae	3	3	100%	3	3	3	100%	3
Polynemidae	3	3	100%	3	3	3	100%	3
Ptereleotridae	3	2	67%	2	3	2	67%	2
Acanthuridae	3	3	100%	3	3	3	100%	3
Molidae	3	3	100%	3				
Elopidae	2	2	100%	2	2	2	100%	2
Moringuidae	2	2	100%	2	2	2	100%	2
Fistulariidae	2	2	100%	2	2	2	100%	2
Coryphaenidae	2	2	100%	2	2	2	100%	2
Stromateidae	2	2	100%	2				

continued Table 1. Barcode coverage of tropical western Atlantic bony fishes, by family; in descending order of number of species known for the region. Numbers in bold highlight families that include non-reef species (i.e. #reef species less than total #species).

Family	ALL SPECIES				REEF SPECIES			
	#sp	#barcoded	%barcoded	#BINs	#reef sp.	#barcoded	%barcoded	#BINs
Emmelichthyidae	2	1	50%	1				
Pempheridae	2	1	50%	1	2	1	50%	1
Megalopidae	1	1	100%	1	1	1	100%	1
Dussumieriidae	1	1	100%	1				
Anguillidae	1	1	100%	1				
Heterenchelyidae	1	0	0%	0				
Colocongridae	1	0	0%	0				
Aulostomidae	1	1	100%	1	1	1	100%	1
Anomalopidae	1	0	0%	0	1	0	0%	0
Dactylopteridae	1	1	100%	1	1	1	100%	1
Malacanthidae	1	1	100%	1	1	1	100%	1
Cirrhitidae	1	1	100%	1	1	1	100%	1
Rachycentridae	1	1	100%	1	1	1	100%	1
Ephippidae	1	1	100%	1	1	1	100%	1
Muraenesocidae	1	0	0%	0				
Lobotidae	1	1	100%	1	1	1	100%	1
Totals	1311	1029		1247	1083	902		1099

the blennioid families Labrisomidae, Chaenopsidae, and Tripterygiidae (Table 1). (Interestingly, the pattern has not been found so far for the true blennies (Blenniidae), which are known to have relatively large wide-ranging larvae.) These gobioid and blennioid families have many species with multiple BINs, typically allopatric, but sometimes sympatric. Among the gobies, there are 102 species barcoded, but 157 barcode BINs within those 102 species. Similarly, among the Labrisomidae there are 88 BINs for 50 species, among the Chaenopsidae there are 85 BINs for 39 species, and among the Tripterygiidae there are 15 BINs for 8 species. The high number of cryptic lineages in the Tripterygiidae persists even though several new species have been recently described [37]. In that study, four cryptic species of *Enneanectes* (three new species) were found to be coexisting on reefs in the Lesser Antilles of the Caribbean Sea (sympatric cryptic species). Nevertheless, the typical finding is allopatric species complexes and, since not all subregions of the Caribbean have been well-sampled, especially Colombian and Venezuelan reefs, the number of cryptic lineages among this set of reef fish families is likely to continue to increase.

4 Discussion

This review of the barcode coverage for the tropical W. Atlantic bony fishes illustrates the changing priorities of a barcode program as it matures and approaches complete coverage. If the earliest phase of a barcode program is to promulgate the message, identify and develop collaborators, and field test the methodology, then the middle phase is intensive recruitment of more collaborators, accumulation of more specimens from unsequenced species and undersampled locations, and development of an optimal quality control approach. As completion is approached, quality control can be improved by more rigorous identification by using the methods of assessment exemplified by this review, i.e. confirmed voucher photographs and specimens, phylogeographic deduction, and process of elimination. This approach permits each BIN, i.e. algorithm-derived operational taxonomic unit (OTU) as defined by Ratnasingham and Hebert (2013) [25], to be assessed in comparison to other lineages in BOLD and a conclusion reached on the species identification for the BIN lineage or sub-lineage, thus creating a "validated BIN". A validated BIN allows an assessor to question species IDs that do not match the

validated identification- either to correct the specimen identification or to highlight exceptions to the "one BIN-one species" paradigm.

Exceptions to the "one BIN-one species" are particularly interesting, comprising two basic categories: A) different species that share a BIN, either distinct sub-BIN lineages that are under the threshold for full BIN assignment [25], or barcode "phenovariants", species that share COI haplotypes or co-occur within a lineage; or B) barcode "genovariants", or multiple distinct COI lineages that fit the description of a single nominal species. Barcode genovariants, which represent cryptic lineages, subspecies, or species, are lineages or OTUs that need to be assessed taxonomically. They are not automatically cryptic species, since the decision on what is a species is the prerogative of a trained taxonomist, not simply a genetic variant [4]. At present, there is a general consensus among fish taxonomists not to accept genovariants as species unless there is a morphological, meristic (countable features), or marking difference considered sufficiently consistent, reliable, and significant to merit species-level designation. This is frequently a difficult decision, especially when the intervening geographic range is unsampled, few (or unique) specimens are sampled or even retained, only preserved material is available, or the taxonomy of the group is unstable, unreviewed, or perhaps entirely unexamined.

Barcode phenovariants are apparently less frequent than genovariants in marine fishes, but are still not unusual, occurring perhaps among as many as 5-10% of all shorefish species (a difficult rough estimate, since "false phenovariants" because of misidentifications are ubiquitous in the databases). When two or more nominal species truly share COI sequences, it indicates that the species deserve a closer taxonomic examination– they could represent the same species that had erroneously been split, but frequently they are closely-related species that either hybridize occasionally or have not been separated for sufficient time for mutually exclusive sets of haplotypes in COI to develop (incomplete lineage sorting). Additional sequencing of faster evolving markers, such as the mitochondrial control region, could distinguish species that have not yet diverged sufficiently in the COI marker. An example of the latter is the important tuna genus *Thunnus*, which mostly share barcode BINs, but can be distinguished to some degree by additional mitochondrial and nuclear markers and/or character-based sequence analyses [38]. Almost all other cases of phenovariants among reef fishes have not been closely examined (e.g. *Hyporthodus* groupers or the *Helicolenus* and *Sebastes* rockfishes), but it is likely that many cases will prove to be similar to the tunas.

References

[1] Ratnasingham S., Hebert P., BOLD: The barcode of life data system, Mol. Ecol. Notes, 2007, 7, 355–364

[2] Ward R.D., Hanner R., Hebert P.D.N., The campaign to DNA barcode all fishes, FISH-BOL, J. Fish. Biol., 2009, 74, 329–356

[3] Steinke D., Hanner R., The FISH-BOL Collaborators Protocol, Mitochondrial DNA, 2011, 22, 10–14

[4] Victor B.C., How many coral reef fish species are there? Cryptic diversity and the new molecular taxonomy, In: Mora C. (Ed.), Ecology of Fishes on Coral Reefs, Cambridge University Press, Cambridge, 2015

[5] April J., Mayden R.L., Hanner R.H., Bernatchez L., Genetic calibration of species diversity among North America's freshwater fishes, Proc. Natl. Acad. Sci., 2011, 108, 10602–10607

[6] Geiger M.F., Herder F., Monaghan M.T., Almada V., Barbieri R., Bariche M., *et al.*, Spatial heterogeneity in the Mediterranean Biodiversity Hotspot affects barcoding accuracy of its

freshwater fishes, Mol. Ecol. Resour., 2014, 14 (6), 1210-21, doi: 10.1111/1755-0998.12257

[7] Coll M., Piroddi C., Steenbeek J., Kaschner K., Lasram F.B.R., Aguzzi J., *et al.*, The biodiversity of the Mediterranean Sea: Estimates, Patterns, and Threats, PLoS ONE, 2010, 5 (8), e11842. doi: 10.1371/journal.pone.0011842

[8] Landi M., Dimech M., Arculeo M., Biondo G., Martins R., Carneiro M., *et al.*, DNA Barcoding for Species Assignment: The Case of Mediterranean Marine Fishes, PLoS ONE, 2014, 9 (9), e106135. doi:10.1371/journal.pone.0106135

[9] Keskin E., Atar H.H., DNA barcoding commercially important fish species of Turkey, Mol. Ecol. Res., 2013, 13, 788–797

[10] Mabragaña E., Dıaz de Astarloa J.M., Hanner R., Zhang J., Gonzalez Castro M., DNA Barcoding Identifies Argentine Fishes from Marine and Brackish Waters, PLoS ONE, 2011, 6 (12), e28655. doi:10.1371/journal.pone.0028655

[11] Dıaz de Astarloa J.M., Ezequiel Mabragaña E., González-Castro M., Rosso J.J., Delpiani M., DNA Barcoding of Marine and Freshwater Fishes of Argentina: Progress after 8 Years of Research and Future Directions, in this volume

[12] Steinke D., Zemlak T.S., Boutillier J.A., Hebert P.D.N., DNA barcoding Pacific Canada's fishes, Mar. Biol., 2009, 156, 2641–2647, doi: 10.1007/s00227-009-1284-0

[13] Hubert N., Hanner R., Holm E., Mandrak N.E., Taylor E., Burridge M., *et al.*, Identifying Canadian freshwater fishes through DNA barcodes, PLoS One, 2008, 3, e2490

[14] Hubert N., Meyer C.P., Bruggemann H.J., Guérin F., Komeno R.J.L., Espiau B., *et al.*, Cryptic Diversity in Indo-Pacific Coral-Reef Fishes Revealed by DNA-Barcoding Provides New Support to the Centre-of-Overlap Hypothesis, PLoS ONE, 2012, 7 (3), doi:10.1371/journal.pone.0028987

[15] Ward R.D., Zemlak T.S., Innes B.H., Last P.R., Hebert P.D.N., DNA barcoding Australia's fish species, Phil. Trans. Roy. Soc. B-Bio. Sci., 2005, 360: 1847–1857

[16] Swartz E., Mwale M., Hanner R., A role for barcoding in the study of African fish diversity and conservation, S. Afr. J. Sci., 2008, 104, 293–298

[17] Zhang J., Hanner R., Molecular Approach to the Identification of Fish in the South China Sea, PLoS ONE, 2012, 7 (2), e30621. doi:10.1371/ journal.pone.0030621

[18] Allen G.R., Erdmann M.V., Reef fishes of the East Indies, Tropical Reef Research, Perth, Australia, 2012

[19] Robertson D.R., Van Tassell J., Shorefishes of the Greater Caribbean: online information system, Version 1.0 Smithsonian Tropical Research Institute, Balboa, Panamá, 2015

[20] Valdez-Moreno M., Vásquez-Yeomans L., Elías-Gutiérrez M., Ivanova N.V., Hebert P.D.N., Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management, Mar. Fresh. Res., 2010, 61, 655–671

[21] Weigt L.A., Baldwin C.C., Driskell A., Smith D.G., Ormos A., Reyier E.A., Using DNA Barcoding to Assess Caribbean Reef Fish Biodiversity: Expanding Taxonomic and Geographic Coverage, PLOS One, 2012, 7 (7), 1–7

[22] Robertson D.R., Cramer K.C., Defining and Dividing the Greater Caribbean: Insights from the Biogeography of Shorefishes, PLos ONE, 2014, 9, e102918

[23] Kulbicki M., Parravicini V., Bellwood D.R., Arias-González E., Chabanet P., Floeter S.R., *et al.* Global Biogeography of Reef Fishes: A Hierarchical Quantitative Delineation of Regions, PLoS ONE, 2013, 8 (12), e81847. doi:10.1371/journal.pone.0081847

[24] Floeter S.R., Rocha L.A., Robertson D.R., Joyeux J.C., Smith-Vaniz W.F., Wirtz P., *et al.*, Atlantic reef fish biogeography and evolution. J. Biogeog., 2008, 35, 22–47, doi: 10.1111/j.1365-2699.2007.01790.x

[25] Ratnasingham S., Hebert P.D.N., A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System, PLoS ONE, 2013, 8 (7), e66213. doi:10.1371/journal.pone.0066213

[26] Randall J.E., Caribbean Reef Fishes, T.F.H. Publications, Hong Kong, 1983

[27] Böhlke J.E., Chaplin C.C.G., Fishes of the Bahamas and adjacent tropical waters, 2nd ed., University of Texas Press, Austin, 1993

[28] McEachran J.D., Fechhelm J.D., Fishes of the Gulf of Mexico, University of Texas Press, Austin, 1998

[29] Carpenter K.E. (Ed.), The Living Marine Resources of the Western Central Atlantic, FAO Species Identification Guide for Fishery Purposes, FAO, Rome, 2002

[30] Eschmeyer W. N. (Ed.) Catalog of Fishes, 2015 http://research-archive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp

[31] Baldwin C.C., Weigt L.A., Smith D.G., Mounts J.H., Reconciling Genetic Lineages with Species in Western Atlantic *Coryphopterus* (Teleostei: Gobiidae), Smithson. Contrib. Mar. Sci., 2009, 38, 111–138

[32] Eytan R.I., Hellberg M.E., Nuclear and mitochondrial sequence data reveal and conceal different demographic histories and population genetic processes in Caribbean reef fishes, Evolution, 2010, 64, 3380–3397

[33] Tornabene L., Baldwin C.C., Weigt L.A., Pezold F., Exploring the diversity of western Atlantic *Bathygobius* (Teleostei: Gobiidae) using cytochrome-c oxidase subunit I, with descriptions of two new species, Aqua, Int. J. Ichthy., 2010, 16, 141–170

[34] Victor B.C. The Redcheek Paradox: the mismatch between genetic and phenotypic divergence among deeply divided mtDNA lineages in a coral-reef goby, with the description of two new cryptic species from the Caribbean Sea, J. Ocean Sci. Found., 2010, 3, 2–16

[35] Victor B.C. *Emblemariopsis carib* and *Emblemariopsis arawak*, two new chaenopsid blennies from the Caribbean Sea: DNA barcoding identifies males, females, and juveniles and distinguishes sympatric cryptic species, J. Ocean Sci. Found., 2010, 4, 1–29

[36] Baldwin C.C., Castillo C., Weigt L.A., Victor B.C., Seven new species within western Atlantic *Starksia atlantica*, *S. lepicoelia*, and *S. sluiteri* (Teleostei, Labrisomidae), with comments on congruence of DNA barcodes and species, ZooKeys, 2011, 79, 21–72

[37] Victor B.C., The Caribbean Roughhead Triplefin (*Enneanectes boehlkei*): DNA barcoding reveals a complex of four West Indian sympatric cryptic species (Teleostei: Blennioidei: Tripterygiidae), J. Ocean Sci. Found., 2013, 7, 44–73

[38] Abdullah A., Rehbein H., Authentication of raw and processed tuna from Indonesian markets using DNA barcoding, nuclear gene and character-based approach, Eur. Food Res. Tech., 2014, 239 (4), 695–706 DOI:10.1007/s00217-014-2266-0