Advantages and Limitations of DNA Barcoding in Identifying Commercially-Exploited Fish Species

Andreea Dudu, Teodora Barbălată, Gina-Oana Popa*, Sergiu Emil Georgescu, Marieta Costache

University of Bucharest, Department of Biochemistry and Molecular Biology, Splaiul Independentei, 91-95, zip code 050095, Bucharest, Romania

Abstract
DNA barcoding aims to be an effective tool for species identification based on partial sequence of the mitochondrial cytochrome c oxidase 1 gene (*CO I)*. The method offers several advantages like small amount of biological samples needed, applicability for all life stages and differentiation among phenotypically alike species. Our study aimed at barcoding economically and ecologically important salmonids. The sequences were determined by using specific primers when the so-call universal primers failed to give amplifications. The sequences were BLAST with GenBank and BOLD databases for species identity confirmation. With one exception, the results of barcoding were congruent with the ones based on morphology in the cases when this last type of diagnostic was done. The method can be regular carried out for species identification, but is totally inefficient when is dealing with hybrids and should be avoided for species that can naturally hybridize.

Keywords: DNA barcoding, *CO I* gene, salmonids, molecular diagnostic.

1. Introduction

DNA barcoding is a molecular tool for species detection based on the analysis of standardized gene sequences [1]. In higher animals, the 5' region of the mitochondrial cytochrome oxidase gene I (*CO I*) is considered to be a very efficient genetic „barcode". The method has several advantages like small amount of biological samples needed (including raw or processed food), applicability for all life stages and differentiation among phenotypically alike species. So, in this light, DNA barcoding appear to be a reliable method to identify species in the absence of morphological diagnostic, to discover species mislabeling and even intentional species substitutions.

* Corresponding author: Gina-Oana Popa, +403181575, +403181575/ 102, popa.gina.oana@gmail.com

Due to their high phenotypic variation during different life stages, fishes among other vertebrates are very challenging in terms of correctly identifying at the species level, when based on morphology alone. In this context, DNA barcoding is applicable for species detection from any tissue that was properly preserved (fins, muscle, scales, eggs, etc.) and in every life stage (larvae, juvenile, and adult) [2].
The efficacy of the barcoding has been demonstrated by the identification fish species with over 90% success rate [3].
DNA barcoding has a broad field of applications, including biodiversity monitoring and fish conservation [4], in monitoring illegal trade of wild species, to identify species origin of raw and processed food, this last topic being highly important for food safety and fraud detection [5].
Salmonids are a group of fish well represented in the Romanian fauna, having a great ecological and economical value. Among salmonid species, brown trout (*Salmo trutta fario*) is widely spread in a large number of water streams from the

mountain area, while the non-native rainbow trout (*Oncorhynchus mykiss*) is predominantly raised in fish farms for food consumption [6].

Until present several molecular methods have been proposed for salmonids' identification including HPLC (High Pressure Liquid Chromatography) [7], PCR-RFLP (Polymerase Chain Reaction Restriction Fragment Length Polymorphisms) [8, 9] and multiplex PCR [10].

The current study aims to barcode two salmonid species from Romania (brown trout and rainbow trout) in order to identify any misclassification based on morphology or species commercial substitution.

2. Materials and methods

Samples collection and DNA extraction

Salmonid samples were collected from water streams from Fagaras Mountains area, fish farms and fish market (Table 1).

The genomic DNA was isolated from fin fragments or from muscle tissue by a classic phenol/ chloroform/ isoamylic alcohol method. The DNA extracts were spectrophotometrically assessed by using NanoDrop 8000 (Thermo Scientific).

Table 1. Origin and identity of the salmonid samples barcoded in the study.

Order	Species	n	Sample code	Sample origin
Salmoniformes	*S. trutta fario*	14	St_3.2 St_3.4 St_3.5 St_3.18	Porumbacu River, Fagaras Mountains
			St_4.1 St_4.4 St_4.5 St_4.7 St_4.18	Avrig River, Fagaras Mountains
			St_5.1 St_5.12 St_5.16 St_5.24 St_5.28	Topolog River, Fagaras Mountains
	O. mykiss	12	Omy_1a Omy_3a Omy_4a Omy_5a Omy_12a	Fish farm, Tulcea County
			Omy_I11 (*S. trutta fario* mislabeled as *O. mykiss*) Omy_II2 Omy_02-Damb Omy_05- Cand	Fish Farm, Valcea County
			Omy_09-1m Omy_09-3m Omy_09-5m	Fish market

Amplification and sequencing

The standard fish DNA barcoding primers, FishF1 (5′tcaaccaaccacaaagacattggcac3′), Fish F2 (5' tcgactaatcataaagatatcggcac3'), Fish R1 (5′ tagacttctgggtggccgaagaatca 3′) and Fish R2 (5' acttcagggtgaccgaagaatcagaa3') [11] (Fish F1: Fish F2: Fish R1: Fish R2 ratio = 1:1:1:1) were used to amplify a 5' fragment of the *CO I* gene. Some samples failed to amplify when the previous mentioned primer cocktail was used, so specific primers were designed with Primer 3 software ([12], [13]). The primers specifically designed for the two salmonid species have the following sequences (StSfo F: 5'tccaccgcttaaactctcag3'/ StSfo R: 5'ccgggtcaaagaaagtggta3') and allow the amplification of a fragment of 678 bp.

Amplifications were performed in 25 µl reaction volume containing 50 ng of DNA template, 1X PCR buffer, 1.5 mM $MgCl_2$, 0.4 mM of each dNTP, 0.4 µM of StSfo F primer, 0.4 µM of StSfo R primer, 1 U of AmpliTaq Gold Polymerase (*Applied Biosystems*). The PCR reaction conditions were: 95 °C for 10 min, 35 cycles of 95 °C for 30 s, 59 °C for 30 s, and 72 ° C for 60 s, followed by a final extension of 72 °C for 10 min. Amplifications were performed in a GenAmp PCR System 9700 thermocycler (*Applied Biosystems*).

After purification with Wizard SV Gel and PCR Clean-Up System (*Promega*), the PCR products were sequenced using BigDye Terminator v3.1 kit *Applied Biosystems*) and analyzed on ABI3130 DNA Genetic Analyzer (*Applied Biosystems*).

Genetic diversity and phylogenetic analysis

Sequences were edited for rectifying ambiguities with BioEdit Sequence Alignment Editor [14] and ulterior screened by using the BLAST algorithm in order to identify the nearest matching sequences GenBank database.

For the phylogenetic and genetic diversity analyses our sequences were truncated to 661 bp in order to accommodate to similar GenBank sequences. The final data set comprised 38 sequences (26 obtained for our salmonids and 12 retrieved from GenBank) (Table 2). The best-fitting model of molecular evolution was inferred with ModelTest [15] and Kimura2-Parameter (K2P) was selected.

The phylogenetic tree was generated using the Neighbor-Joining algorithm implemented in MEGA5 software [16]. The bootstrap consensus tree was inferred from 1000 replicates.

The genetic diversity in terms of number of haplotypes, haplotype diversity (H_d), nucleotide diversity (Pi) and K (average number of nucleotide differences) was inferred using DnaSP v5.1 [17].

Table 2. Partial *CO I* sequences from GenBank used in the phylogenetic analysis.

Species	GenBank accesion number
Oncorhynchus mykiss	L29771
	KP013084
	LC050735
	NC_001717
	HQ167682
	DQ288270
	DQ288268
	KP218514
Salmo trutta fario	KT633607
	JQ390057
	HQ167684
	AM910409

3. Results and discussion

A total number of 26 *CO I* partial sequences that act as DNA barcodes were obtained in this study. The sequences were longer than 600 bp and no INDELs or STOP codons were present in the analyzed data set suggesting that these represent valid DNA barcodes.

With one exception, the BLAST analysis revealed a 99% of identity at an E-value equal with 0 for the individuals identified (the case of wild and in the fish farms samples) or labeled (the case of fish market samples) by morphological features.

One sample (Omy_I11) was incorrectly classified as rainbow trout based on morphology, while the BLAST analysis revealed 99% identity with *S. trutta* sequences available in GenBank and BOLD (Barcode of Life Data Systems) databases.

The genetic diversity analysis results for the complete data set of 38 sequences were summarized in table 3.

The results regarding the number of variable sites, number of haplotypes and haplotype diversity were highly similar for the two species. Instead the Pi and K indices were higher in *S. trutta fario* indicating a higher variation within this species. The overall analysis results showed higher values for Pi and K when compared with the same indices for individual species. This represents a consequence of between species variation.

Table 3. Diversity indices for the analyzed salmonid species.

Taxon	No. of analyzed sequences	No. of analyzed sites	No. of variable sites	No. of haplotypes	Haplotype diversity (Hd) ± SD	Nucleotide diversity (Pi)	Average no. of nucleotide differences (K)
S. trutta fario	18 (plus one mislabeled as *O.mykiss*)	661	7	5	0.771 ± 0.051	0.00389 ± 0.00044	2.569
O. mykiss	19	661	7	6	0.726 ± 0.090	0.00283 ± 0.00048	1.868
Salmoniformes (*S. trutta fario* + *O.mykiss*)	38	661	86	11	0.876 ± 0.029	0.06171 ± 0.00242	40.791

The phylogenetic analysis of the full dataset showed a clear and well supported separation at the species level (Figure 1).

With two exceptions, all the individuals were assigned to the correspondent clade, according to their classification. The individual coded as Omy_I11, misclassified as rainbow trout, was distributed along with brown trout sequences in a monophyletic group. Also, the KP218514 sequence from GenBank, corresponding to an interspecific hybrid between *O. mykiss* and *S. salar*, was placed together with the sequences of rainbow trout showing the limitation of DNA barcoding in discriminating pure species from interspecific hybrids. The *O. mykiss* fish market samples were correctly classified, no species substitution being detected.

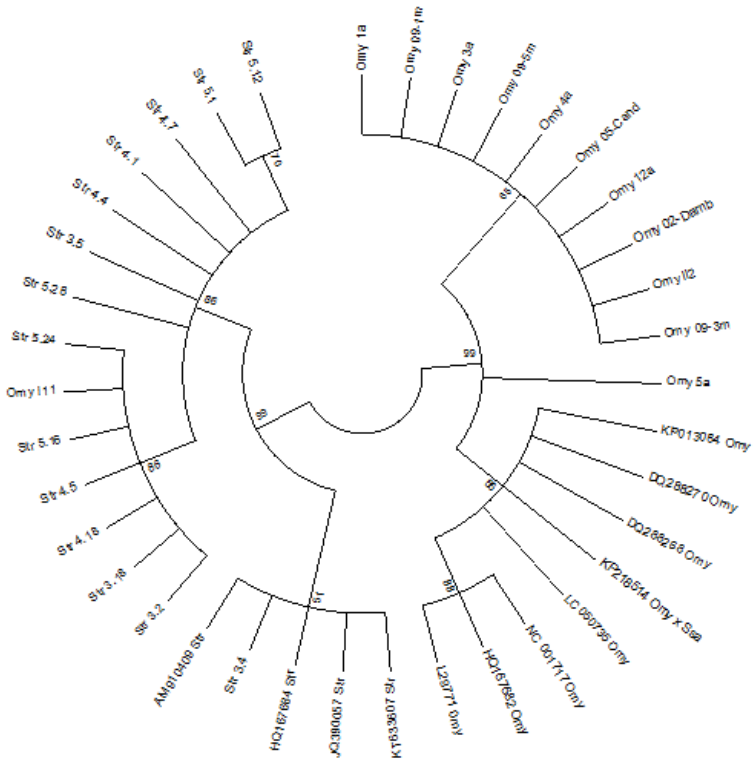

Figure 1. Phylogenetic tree based on NJ algorithm illustrating the relationships between the DNA barcodes sequences in the investigated salmonid species. Bootstrap values (>50) out of 1000 iterations are showed for each node.

Conclusions

The two salmonid species analyzed in this study presented a similar pattern of genetic diversity at *CO I* region. The phylogenetic analysis revealed that each represents a single cluster of highly similar sequences.

The DNA barcoding method is fast, cost efficient, and allows the identification of species from different biological samples, thus being a useful tool for conservation purposes. The method is inefficient for hybrids detection, but this is a minor drawback in salmonids due to the fact the phenomenon is limited in the wild. Instead, several salmonid hybrids are raised in fish farms and for an accurate identification the DNA barcoding should be applied along with nuclear marker analysis. Due to the limited number of fish market samples used in the current study, further research is needed to explore the presence of any species substitution.

Acknowledgements

This work was supported by the PN-II-PT-PCCA Project 116/2012.

References

1. Herbert, P.D.N., Cywinska, A., Ball., S.L., DeWaard, J.R., Biological identifications through DNA barcodes, Proc R Soc Lond B Biol Sci, 2003, 270, 313-321.
2. Becker, S., Hanner, R., Steinke, D., Five years of FISH-BOL: Brief status report, MDN, 2011, 22 (S1), 3-9.
3. Pereira, L., Hanner, R., Foresti, F., Oliveira, C., Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna?, 2013, BMC Genet, 14: 20.
4. Holmes, B.H., Steinke, D., Ward, R.D., Identification of shark and ray fins using DNA barcoding, 2009, Fish Res, 95, 280-288.
5. Nicole, S., Negrisolo, E., Eccher, G., Mantovani, R., Patarnello, T., Erickson, D. L., Kress, W. J., Barcaccia, G., DNA barcoding as a reliable method for the authentication of commercial seafood products, 2012, Food Technol. Biotechnol., 50(4), 387–398.
6. Dudu, A., Georgescu, S.E., Popa, O., Dinischiotu, A., Costache, M., Mitochondrial 16S and 12S rRNA sequence analysis in four salmonid species from Romania – phylogenetic aspects, Acta Zool. Acad. Sci. Hung., 2011, 57(3), 233-246.
7. Turujman, S. A., Warner, W. G., Wei, R. R., Albert, R. H., Rapid liquid chromatographic method to distinguish wild salmon from aquacultured salmon fed synthetic astaxanthin, J. AOAC Int., 1997, 80, 622–632.
8. Dudu, A., Georgescu, S. E., Dinischiotu, A., Costache, M. PCR-RFLP method to identify fish species of economic importance, Arch. zootech., 2010, 13 (1), 53-59.
9. Rasmussen, R. S., Morrissey, M. T., Walsh, J., Application of a PCR-RFLP method to identify salmon species in U.S. commercial products, 2010, J Aquat Food Prod T, 19(1), 3–15.
10. Hellberg, R. S., Rasmussen, Morrissey, M. T., Hanner, R. H., A multiplex PCR method for the identification of commercially important salmon and trout species (Oncorhynchus and Salmo) in North America. J. Food Sci., 2010, 75(7), C595–606.
11. Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., Hebert, P. D. N., DNA barcoding Australia's fish species, Phil. Trans. R. Soc. B, 2005, 360, 1847–1857.
12. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G. Primer3 - new capabilities and interfaces, 2012) Primer3 - new capabilities and interfaces, Nucleic Acids Res., 2012, 40(15), e115.
13. Koressaar, T., Remm, M., Enhancements and modifications of primer design program Primer3, Bioinformatics, 2007, 23(10), 1289-1291.
14. Hall, T.A., BioEdit: a user-friendly biological sequence alignement editor and analysis program for Windwos 95/98/NT, Nucleic Acids Symp Ser, 1999, 41, 95-98.
15. Posada, D., Crandall, K.A., Modeltest; testing the model of DNA substitution, Bionformatics, 1998, 14, 817-818.
16.Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Molecular Biology and Evolution, 2011, 28, 2731–2739.
17. Librado, P., Rozas J., DnaSPv5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 2009, 25, 1451-1452.