Application of DNA barcoding for controlling of the species from *Octopus* genus

Francesco Debenedetti,[1]
Alessandra Dalmasso,[1]
Maria Teresa Bottero,[1]
Maurizio Gilli,[2] Stefano Gili,[2]
Valentina Tepedino,[3] Tiziana Civera[1]

[1]Dipartimento di Scienze Veterinarie, Università degli Studi di Torino, Grugliasco (TO); [2]ASLTO1, Torino; [3]Eurofishmarket, Bologna, Italy

Abstract

The DNA barcoding proposes the use of a particular sequence from a single genomic region as the base for an identifying system capable to determine all animal species. This methodology comprises the analysis of a 655 base-pair region from the mithocondrial cytochrome *C oxidase* gene (COI). Its application in the species identification of fishery products has been very promising. However, in the last years some doubts about its usage have emerged. In this work, we make use of the DNA barcoding for the identification of some of the octopus species with higher commercial interest (*Octopus membranaceus*, *Octopus vulgaris*, *Octopus aegina*, *Octopus cyanea*) focusing the attention on the reliability and completeness of the available information on the databases. The study looked over 51 individuals apparently belonging to the *Octopus* genus. For the identification of *O.aegina*, *O.cyanea*, *O.vulgaris* species no particular problems were found. On the other hand, most of the samples of *O.membranaceus*, though they clearly presented the morphological characteristics of the species, were not identified with the biomolecular analyses.

Introduction

The DNA barcoding puts forward the use of a sequence from a single genomic region (defined as *barcode*), as the base of a recognition system capable to identify all animal species (Hebert *et al.*, 2003). This biomolecular methodology consists in the identification of the belonging species through the sequencing of a fragment of mitochondrial genes, its alignment and comparison with the information available on the databases. In particular, the barcoding approach comprises the analysis of a 655 bp region located at the 5' end of the subunit I of the mithocondrial cytochrome *C*

oxidase (COI) gene. This gene codes for one part of the terminal enzyme of the mithocondrial respiratory chain. The COI gene has been chosen as a universal molecular target since it allows the design of universal, robust and functional primers for almost all the members of the animal Phyla (Folmer *et al.*, 1994). The effectiveness of this fragment for the identification of species has been demonstrated for several animal species, from vertebrates to invertebrates (Waugh, 2007; Ward and Holmes, 2007). Recently, this technique was recognised as the best method in forensics for species identification, and was proposed by the United States Food and Drug Administration (USFDA) as the methodology for the authentication of commercial fish products (Dawnay *et al.*, 2007). The US agency has also the intention of introduce the DNA barcoding data in the Regulatory Fish Encyclopedia (RFE) to support the investigations of mislabeling and the substitution of the fish species (Yancy *et al.*, 2008). In May 2004, the Consortium for the Barcoding of Life (CBOL) was formed, comprising several international patterns. Since the beginning, its mission has been the exploration and the development of the potentiality of the DNA barcoding as a practical research tool for species identification. The use of this methodology for species identification of fishery products has been very promising from the outset, given the large number of species already identified. It has been shown that 98 and 93% of the marine and fresh water species, respectively, can be differentiated using barcodes (Savolainen *et al.*, 2005; Ward *et al.*, 2009).

This favourable outcome as well as the need of a complete and reliable instrument for identification of the species, have led to the formation of the initiative Fish Barcode of Life (FISH-BOL) (www.fishbol.org). This campaign, launched in 2005, has as main objective the collection of DNA barcodes of all the fish in the world, equivalent to 31,000 species approximately. It endorses the FishBase (www.fishbase.org) as the taxonomic authority, and the BOL database (BOLD) as the working bioinformatic platform. FISH-BOL represents one of the most complete resources for the species identification of the fishery products (Ward *et al*., 2009).

However, in the past years, some doubts regarding the use of DNA barcoding have emerged because of the difficulties to discriminate recently spread species, or species with a wide spatial differentiation; as well as the inability to differentiate new and hybrid species (Moritz and Cicero, 2004; Hickerson *et al.*, 2006; Rubinoff, 2006). Furthermore, this technique has been extensively evaluated for the identification of fish species, still there are few data supporting the applicability in the identification of the most common species of

cephalopods and octopus.

In this study we test the DNA barcoding for the identification of some octopus species of greatest commercial interest (*Octopus membranaceus*, *Octopus vulgaris*, *Octopus aegina*, *Octopus cyanea*) focusing the attention on the reliability and completeness of the available databases.

Materials and Methods

The sampling involved the collection of 51 specimens belonging to the genus *Octopus*: 1 *O.dollfusi*, 2 *O.aegina*, 2 *O.cyanea*, 24 *O.vulgaris* and 22 *O.membranaceus*. All samples, collected in 2012 and 2013 from different suppliers, were morphologically identified and kept at -20°C. The DNA extraction was done using the DNeasy Tissue and Blood Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions for animal tissue. The extracted DNA was quantified with NanoDrop 1000 (Thermo Scientific, Waltham, MA, USA). For all samples, segments of the COI genes were amplified following the protocol proposed by Folmer *et al*. (1994). Since there were no available reference sequences for *O.membranaceus* in FISH-BOL, a second sequencing protocol was introduced. It contemplated the usage of *cytochrome b* (*cytb*), a historical target used for species identification (Espineira *et al*., 2010).

Amplicons were visualised by electrophoresis on a 2% agarose gel and coloured with Eurosafe Nucleic Acid Stain. The amplified products were purified with the enzyme ExoSap-IT (USB) and sequence with the

BigDye terminator kit (Applied Biosystems, Carlsbad, CA, USA) according to the manufacturer's instructions. The extension products were purified using the DyeEx 2.0 spin kit (Qiagen), denatured in formamide and analysed with the ABI Prism 310 Genetic Analyzer (Applied Biosystems). The obtained sequences were examined, corrected and analysed with the MEGA5 software and subjected to identification with BOLD-IDS for the COI gene, and to BLASTN (GenBank) for both genes.

Finally phylogenetic relationships among the studied samples were investigated with differences method (Nei and Kumar, 2000). For the distance matrix, phylogenetic trees were constructed using the Neighbour-Joining method. The bootstrap method (500 replicates) was used to obtain the support of different groups included in the phylogenetic (Felsenstein, 1985).

Results

All samples originated an amplicon of 655 bp for the COI gene and 651 bp for the *cytb* gene. The sequencing results are presented in Table 1. For the species *O.aegina*, *O.cyanea*, *O.vulgaris* and *O.dollfusi* there were no identification problems. We were able to identify 27 samples, out of 29, with a similarity of 99-100%. Out of these, 22 samples were in accordance with the labels of the products, while 5 did not correspond to what was declared in the product description. Finally, for 2 samples labeled as *O.vulgaris* it was not possible to obtain the biomolecular identification, since they presented a low identity (91%) with *Amphioctopus rex* and *Amphioctopus marginatus,* species not reported in the FAO catalogue.

Further difficulties were encountered in the identification of samples labelled as *O.membranaceus.* One sample only was identified from both databases in a non-ambiguous way (100% similarity) as *O.aegina*, while for the other 21 samples it was not possible to obtain a certain identification. In particular, regarding the *cytb,* the similarity obtained values ranged from 93-94% with *O.membranaceus,* 93% with *O.aegina* and 90% with *Cistopus taiwanicus* (Figure 1). The results were even more complex for the COI gene because of the

lack of reference sequences of *O.membranaceus* in both GenBank and BOLD databases (Figure 2).

Discussion

The expansion in food preservation, processing technologies and market liberalisation has contributed significantly to the globalisation of fish trade, both in terms of species and products. In a globalised market, the eagerness to offer leading products appreciated and recognised to derive maximum revenue, can introduce the habit of replacing valuable species with others very similar but economically or qualitatively inferior, thus incurring into commercial fraud (Jacquet and Pauly, 2008).

The identification of species in fisherystock products, essential at the moment of the commercialisation, has been based for a long time on morphological characteristics solely, and it is still the official method for taxonomic classification. However, in the last years this morphological classification has been supported by
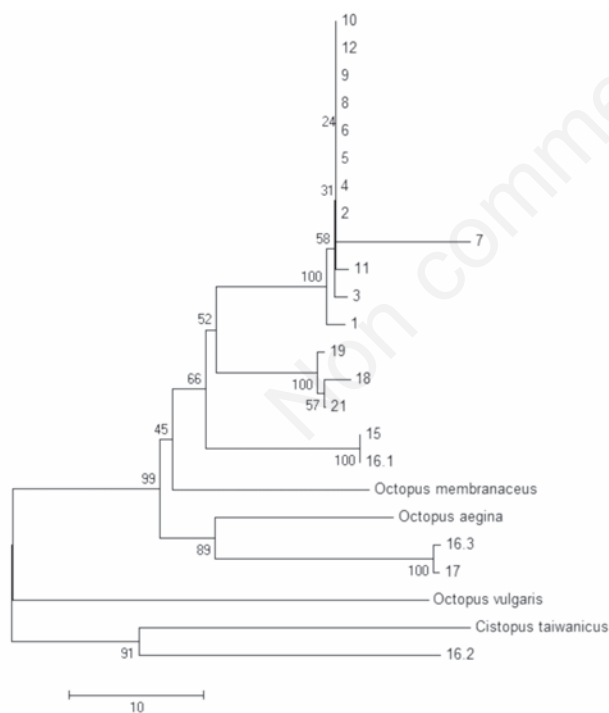
Figure 1. Phylogenetic tree showing the relationship among the studied samples carried from the alignment of the *cytochrome b* gene. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The distances were computed using the number of differences method and are in the units of the number of base differences per sequence.
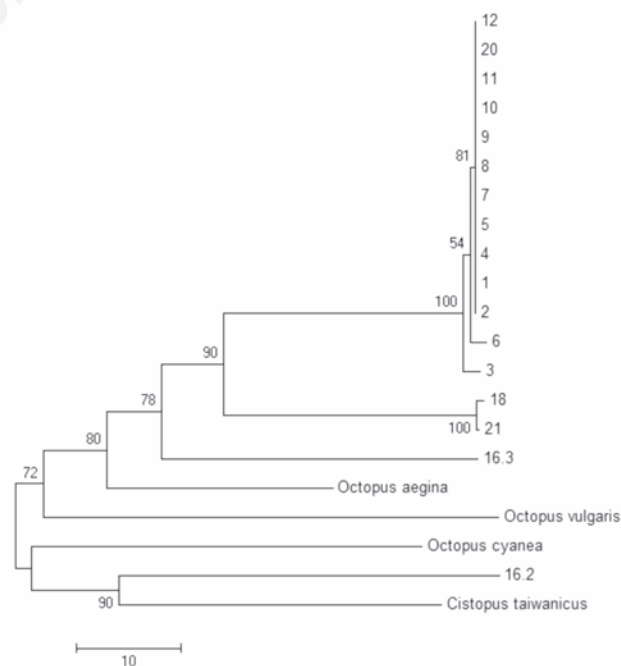
Figure 2. Phylogenetic tree showing the relationship among the studied samples carried from the alignment of the *C oxidase* gene. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The distances were computed using the number of differences method and are in the units of the number of base differences per sequence.

biomolecular laboratory techniques based on the sequencing of several mitochondrial genes (*cytb*, 16S rRNA, 12SrRNA, NADH). Precisely, the DNA barcoding, based on the sequencing of the COI gene, might be a reliable tool used in the daily practice for the control of commercial fraud. The ability of barcoding to provide unambiguous designation of species from a complete specimen or part of it, has important implications in several fields: in the retail (accidental or intentional replacement, consumer protection and trade regulation); in fisheries management (monitoring of fishing stocks, sustainable fishing); in the conservation of fish stock (identification of endangered species, protected, damaged or parts of them) (Costa and Carvalho, 2007).

Given the limited applications of DNA barcoding in the field of cephalopods, the present work evaluated the possibility of using it to control the *Octopus* genus and the most commonly found species in the Italian fish markets. In fact, octopus phylogeny has been subjected to confusion and controversy throughout the past years (Carlini *et al.*, 2001). The *Octopus* genus is the largest within the family *Octopidae*, as it comprises more than 200 species. Of these, 90% were inserted just in the above-mentioned genus, for this reason it is called *catch-all* (Lü *et al.*, 2013). Several taxonomic studies based on morphology, have recognised more complexity within the genus *Octopus* spp. (*Species group*). At the beginning there were 9 groups of species identified, many of which are under review. For example, the *O.aegina* complex was classified as *Amphioctopus* spp. Currently many species, previously classified as *Octopus* spp., were included in the *Amphioctopus* genus (Huffard and Hochberg, 2005). Furthermore, the analysis of morphological parameters does not seem sufficient to describe such a complex variability, as also confirmed by the data obtained in

this work. If there were no identification problems for the species *O.aegina, O.vulgaris* and *O.cyanea*, the same cannot be said for *O.membranaceus*, important species of commercial value, fished in the Indo-Pacific area and widely spread in Italian fisheries. From the analysis of the COI gene, the majority of the *O.membranaceus* samples, despite they clearly presented the morphological characteristics of the species (a conspicuous, dark, ringed ocellus on web at base of arms II, enteroventral to the eyes) (FAO, 1984), were not identified with molecular analyses. In particular regarding the COI gene, there were no reference sequences, while in the case of *cytb* the samples showed values of similarity with *O.membranaceus*, unexpectedly low (93-94%). From the alignments of the sequences of the *O.membranaceus* samples, both the COI and *cytb* presented a great variability not only with respect to the sequences deposited in GenBank, but also between the same samples. The only exception, represented by a group of homologous samples (identity of 99-100%) is justified by the fact that the 12 specimens were from the same batch (Figures 1 and 2).

Moreover, this considerable variability was also found in *O. vulgaris* species: in fact from the alignment of the sequences of the COI gene a range of similarity between the 96 and 100% was observed. As example, in Table 2, ten representative sequences of this gene were aligned, highliting the variable sites between them. There were found show 35 single nucleotide polymorphism (SNPs) out of 655 bp, which demonstrate a higher variability. On one hand, these results showed that the extreme genotypic variability is not correlated with the phenotypic variability, and on the other hand they are confirmed by the literature related to the *Octopus* genus. The polyphyletic of the *Octopus* genus has been demonstrated by a number of molecular studies (Lü *et al.*, 2013). The reconsideration of generic names and the major revision of these taxa have been proposed by some authors (Guzik *et al.*, 2005).

Conclusions

The DNA barcoding applied in the control of

Table 1. Results of sequencing identification.

Number of samples	Labelling	Species	Similarity (%)
2	*O. aegina*	*O. aegina*	99
2	*O. cyaneus*	*O. cyanea*	99-100
1	*O. dollfusi*	*O. aegina*	100
4	*O. vulgaris*	*O. cyanea*	99-100
18	*O. vulgaris*	*O. vulgaris*	99-100
2	*O. vulgaris*	n.a.	n.a.
1	*O. membranaceus*	*O. aegina*	99-100
21	*O. membranaceus*	n.a.	n.a.

n.a., not available.

Table 2. Example of variable sites in *C oxidase* gene from the alignment of ten *Octopus vulgaris* representative sequences.

Octopus vulgaris	107	145	175	196	212	247	262	277	301	310	313	340	342	367	373	383	385	388	406	415	424	439	481	493	505	514	539	562	580	601	616	617	618	631	650
1	T	T	T	A	A	C	C	C	T	A	T	A	C	A	C	C	A	A	A	T	T	C	A	T	G	T	T	A	A	C	A	A	G	C	T
2	.	C	C	.	G	T	.	T	A	.	.	G	T	T	.	T	C	T	.	C	.	T	.	A	A	C	.	T	T	T	G	A	T	.	.
3	.	C	C	G	G	T	.	T	A	.	C	G	.	.	T	.	.	C	.	.	C	.	.	.	A	A	C	.	.	T	T	.	.	T	.
4	.	C	C	.	G	T	.	T	A	.	C	G	.	.	T	.	.	C	.	C	C	.	.	.	A	A	C	.	.	T	T	.	.	T	.
5	.	C	C	.	G	T	.	T	A	.	C	G	.	.	T	.	.	C	.	.	C	.	.	.	A	A	C	.	.	T	T	.	.	T	.
6	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	C	A	.	C	G	.	.
7	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	.	C	G	.	.
8	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	.	C	G	.	.
9	.	.	.	.	G	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.
10	C	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C

SNP, single nucleotide polymorphism. Dots indicate homolog nucleotides.

Octopus spp. highlights the need to enlarge information in the database by increasing the number of species and the number of sequences available for each one. In addition, from this study a complexity in the interpretation of the data emerges, pointing out the importance of the training and experience of the operator to avoid mistakes.

References

Carlini DB, Young RE, Vecchione M, 2001. A molecular phylogeny of the Octopoda (Mollusca: Cephalopoda) evaluated in light of the morphological evidence. Mol Phylogenet Evol 21:388-97.

Costa FO, Carvalho GR, 2007. The barcode of life initiative: synopsis and prospective societal impacts of DNA barcoding of fish. Genomics Soc Policy 3:29-40.

Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe RS, 2007. Validation of the barcoding gene CO1 for use in forensic genetic species identification. Forensic Sci Int 173:1-6.

Espineira M, Vieites JM, Santaclara FJ, 2010. Species authentication of octopus, cuttle fish, bobtail and bottle squids (families Octopodidae, Sepiidae and Sepiolidae) by FINS methodology in seafoods. Food Chem 121:527-32.

FAO, 1984. Cephalopods of the world: an annotated and illustrated catalogue of species of interest to fisheries. Food and Agriculture Organization ed., Rome, Italy.

Felsenstein J, 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-91.

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 Biotech 3:294-9.

Guzik MT, Norman MD, Crozier RH, 2005. Molecular phylogeny of the benthic shallow-water octopuses (Cephalopoda: Octopodinae). Mol Phylogenet Evol 37:235-48.

Hebert PDN, Cywinska A, Ball SL, de Waard JR, 2003. Biological identifications through DNA barcodes. Philos T Roy Soc B 270:313-22.

Hickerson MJ, Meyer CP, Moritz C, 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. Syst Biol 55:729-39.

Huffard CL, Hochberg FG, 2005. Description of a new species of the genus Amphioctopus (Mollusca: Octopodidae) from the Hawaiian islands. Molluscan Res 23:113-28.

Jacquet J, Pauly D, 2008. Trade secrets: renaming and mislabeling of seafood. Mar Policy 32:309-18.

Lü ZM, Cui WT, Liu LQ, Li HM, Wu CW, 2013. Phylogenetic relationships among Octopodidae species in coastal waters of China inferred from two mitochondrial DNA gene sequences. Genet Mol Res 12: 3755-65.

Moritz C, Cicero C, 2004. DNA barcoding: promise and pitfalls. PLoS Biology 2:1529-31.

Nei M, Kumar S, 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, NY, USA.

Rubinoff D, 2006. Utility of mitochondrial DNA barcodes in species conservation. Conserv Biol 20:1026-33.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R, 2005. Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philos T Roy Soc B 360:1805-11.

Ward RD, Hanner R, Hebert PDN, 2009. The campaign to DNA barcode all fishes, FISHBOL. J Fish Biol 74: 329-56.

Ward RD, Holmes BH 2007. An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (cox1) in fishes. Mol Ecol 7:899-907.

Waugh J, 2007. DNA barcoding in animal species: progress, potential and pitfalls. Bioessays 29:188-97.

Yancy HF, Zemlak TS, Mason JA, Washington JD, Tenge BJ, Nguyen N-LNT, Barnett JD, Savary WE, Hill WE, Moore MM, Fry FS, Randolph SC, Rogers PL, Hebert PDN, 2008. Potential use of DNA barcodes in regulatory science: applications of the regulatory fish encyclopedia. J Food Protect 71:210-7.