



I.A.M.C.-CNR U.O.S. di Capo Granitola

DNA Barcoding as a tool for Zoological Taxonomy: Identification of bony fish in the Mediterranean Sea

Biondo G.^A, Armeri G.M.^A, Bennici C. D.^A, Nicosia A.^A, Tagliavia M.^A,
Maneiro I.^A, Torri M.^A, Masullo T.^A, Monastero L.^A, Di Natale M.V.^A,
Mangiaracina F.^A, Giaramita N.^A, Maggio T.^D, Lo Brutto S.^B, Arculeo M.^B,
Mazzola S.^C and Cuttitta A.^A.

A- Laboratory of Molecular Ecology and Biotechnology, National Research Council of Italy (CNR), Institute for Coastal Marine Environment (IAMC), UOS of Capo Granitola, via del Mare 3 – 91021, Torretta Granitola (Campobello di Mazara, Tp), Italy.

B- Dipartimento di Scienze e Tecnologie Biologiche, Chimiche e Farmaceutiche (STEBICEF), University of Palermo, Palermo, Italy.

C- National Research Council of Italy (CNR), Institute for Coastal Marine Environment (IAMC), UOS of Capo Granitola, via del Mare 3 – 91021, Torretta Granitola (Campobello di Mazara, Tp), Italy.

D- Ispra (Istituto Superiore per la Protezione e la Ricerca Ambientale), Palermo, Italy

Index

Introduction..... 3

Materials and Methods 5

 Sampling 5

 Genetic analysis..... 6

 Extraction of DNA total from tissue 6

 Amplification and sequencing of COI region 7

References..... 12

Thanks..... 13

Introduction

The Mediterranean Sea only covers 0.82% of Earth's surface and contains a 0.32 % the total water volume of marine waters. However, in this region there is about 14% of the world's fish fauna (Bianchi & Morri, 2000), so it is essential to study this geographical area where in a same surface area and volume of water it is possible capture a larger number of species than other seas.

The description of all the species present in nature is a vast task to be fulfilled by using the classical approach of morphological description of the organisms. In recent years, the traditional taxonomy, based primarily on identification keys of species, has shown a number of limitations in the use of the distinctive features in many animal taxa and inconsistencies with the genetic data. Furthermore, the increasing need to get a true estimate of biodiversity has led Zoological Taxonomy to seek new approaches and methodologies to support the traditional methods. The classification procedure has added modern criteriasuch as the evolutionary relationships and the genetic, biochemical and morphological characteristics of the organisms. Until now the Linnean binomial was the only abbreviated code associated with the description of the morphology of a species. The new technologies aim to achieve a short nucleotide sequence of the DNA to be used as an unique and solely label for a particular species, a specific genetic barcode. For both morphological and genetic approaches, skills and experience are required.

Taxonomy is one of zoological disciplines that has been benefited from the achievements reached by modern molecular biotechnology. Using a molecular approach it is possible to identify cryptic species, to establish a family relationship between species and their membership of taxonomic categories or to reconstruct the evolutionary history of a taxon.

Genetic approaches for the diagnosis of taxa exploit the differences between the DNA sequences to identify organisms (Wilson, 2005). The system proposed by P. Hebert et al. (2003) locates inside the mitochondrial COI gene (cytochrome oxidase subunit I) the bio-identification system useful in taxonomic identification of species (Lo Brutto et al., 2007).

The COI region, used for genetic identification " - DNA barcode - is short enough to allow, with the current technology, to decode sequence (the pairs of nucleotide bases) in a single step. Despite, this region only represents a tiny fraction of the mitochondrial DNA content in

each cell, the COI region has sufficient variability to distinguish the majority of species among them

Several studies have been conducted to test the effectiveness of the "COI code" in various terrestrial and marine animal groups, spread from the tropics to the poles. It was found that using only the "codes" it is possible to recognize 98% of species previously identified by classic taxonomy. In the other 2%, the "COI" narrows the identification code to a pair of closely related species, a small group that typically represents lineages whose divergence is very recent, or species that hybridize with each other regularly (Stoeckle & Hebert, 2008). Thanks to the DNA barcode, this concept can be extended to the whole animal kingdom (Hebert et al., 2003). Once it has been found the suitable barcode, it is possible to build a reference library of these segments, starting from specimens whose identity is already established. The comparison of the DNA barcode of an unknown sample with those of "type specimens" allows to determine whether that organism is a species already known. The mechanism to create the library is simple: DNA is extracted from a tissue sample, the sequence of base pairs on the COI region is determined and the information is entered in a database of genetic codes.

To support the library of codes, different researchers have found and put together big banks of tissue, under conditions that preserve the DNA (Stoeckle & Hebert, 2008).

Currently, the reference library of DNA Barcode for all fish species is provided by the Fish Barcode of Life (FISH-BOL) (Ward et al, 2009). (<http://www.fishbol.org/>)

Materials and Methods

Sampling

From May 2014 to December 2015, 144 specimens of bone fishes were collected . Individuals were captured in the Sicilian Channel (Fig. 1), at a depth between 80 and 300 meters, through scientific surveys.

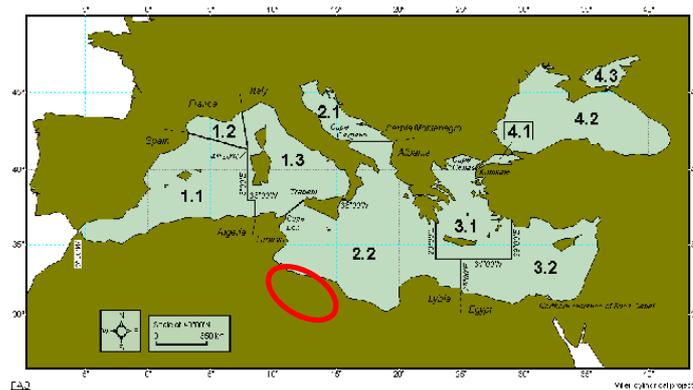


Figure. 1: Map of FAO area 37.2 where is indicated in red the area of sampling.

All samples were identified through morphological identification keys (Tortona, 1956; 1970; 1975) and then it was given to them a reference code (voucher-code) that matches the code to be inserted in the FISH -BOL (Fig.:2).

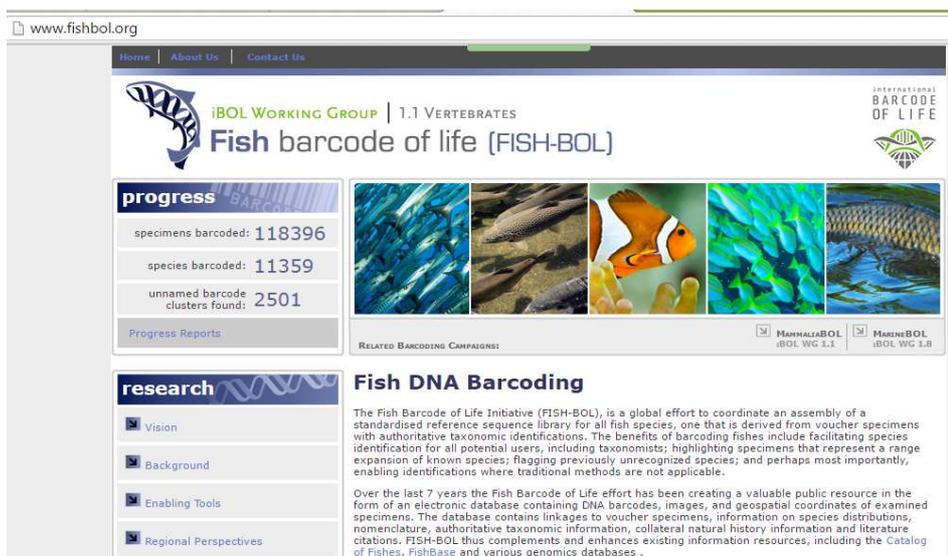


Figure 2: Homepage of FISHBOL (it is a coordination between the research groups on a global scale, aimed at the identification of taxonomic fish, through the application of traditional and molecular methods).

Following the protocol of the Barcode of Life Data System database (useful for the identification of all the animal species, <http://www.barcodinglife.org>), for each code it was also stored different characteristics related to each sample as photo, whether a piece of tissue was preserved in absolute ethanol, location and/or date of collection

One to five specimens of each species (according Barcode protocol), were processed following these steps:

- ✓ Taxonomical identification through identification keys species (Tortonese,1970, 1975);
- ✓ Photographed and archived of digital photos;
- ✓ Preservation of a piece of muscle tissue in ethyl alcohol 98% for total DNA extraction to performe genetic analysis and construct a tissue library.

Genetic analysis

For the DNA amplification process there are three steps: 1) the pre-PCR where it is carried out sample preparation, DNA extraction and the start of the amplification reaction; 2) PCR (Polymerase Chain Reaction) with the amplification of DNA template; and 3) the post-PCR, in which the amplified fragments are highlighting.

Extraction of DNA total from tissue

The nucleic acid extraction was conducted by QUIAquick DNeasy tissue and blood extraction kit (Qiagen). This kit uses a membrane to the high affinity of silica gel for DNA which allows a rapid and efficient extraction of total DNA from the tissue, without organic extractions or ethanol precipitates. The buffer system included in the kit allows cell lysis and subsequent selective binding of DNA to the membrane. The protocol followed for the extraction of total DNA from tissue, is that suggested by the kit. A small portion of tissue (35 mg) is placed in a column with 20 µl of Proteinase K. The mixture was positioned on a mechanical shaker and incubated in an oven at 37°C until complete digestion of the tissue. The lysate is subjected to several steps, with different buffers as indicated in kit protocol. Finally, the extracted DNA was suspended in a solution of sterile distilled water.

Amplification and sequencing of COI region

The technique of polymerase chain or PCR (Polymerase Chain Reaction) was introduced by Kary Mullis in the mid-80s and has revolutionized molecular genetics since it allows the amplification of a specific region of DNA. The PCR reconstructs in vitro a specific passage of cellular reproduction: the synthesis of a "complete" DNA region (double-stranded segment) starting with one single-stranded DNA. The missing filament is reconstructed from a series of nucleotides (the "building blocks" which constitute the nucleic acids) that are arranged in the correct sequence, complementary to that of DNA concerned.

The PCR procedure presents 3 operational stages: denaturation, annealing, extension.

In the denaturation step, the double-stranded DNA is heated to a temperature between 95 °C and 98 °C so it is separated into the two single filaments that compose it. In the annealing step the, primers (a strand of short nucleic acid sequence, generally about 10 base pairs, that serves as a starting point for DNA synthesis) bind to the single strands in the points in which their sequence is complementary. In the extension phase, (about 70 °C) the primers are extended by DNA polymerase. The set of the three phases constitutes one cycle. In order to guarantee the complete denaturation of DNA at the beginning of the PCR procedure an initial denaturation is performed. In the same sense, at the end of the cycles a final extension was carried out to let DNA fragments were completed. The number of target DNA copies is doubled each cycle, so, even if the procedure was started with a single chain of nucleic acid, a number of copies equal to 2^n , where n corresponds to the number of cycles performed, is obtained.

The reaction mixture for the PCR contains: a quantity, also a minimum of template DNA; the mixture of the four nucleotide precursors (dNTPs: 2 'deoxynucleosides 5'triphosphate); appropriate primers complementary to the segment to be amplified; other support elements (eg. Mg^{++} ions) necessary to constitute the environment suitable for the reaction; DNA polymerase. In the specific of bone fish samples, the COI region was amplified using the primer pair and FishF1 FishR1 (Ward et al., 2005) and the PCR reactions are set up using 1 μ l of total DNA extract as template in a total volume of 25 μ l according to the following reaction scheme reported (respecting the final concentrations):

Component	Reaction of 25 μ l	Final concentration
dH ₂ O	Fino a 25 μ l	
5x Reaction buffer	5 μ l	1x

10 mM dNTs	X μ l	200 μ M
Primer mix (FishF1+FishR1)	X μ l	0,5 μ M
DNA	1 μ l	
Phyre Hot Start DNA Polymerase	0,5 μ l	

The amplification was carried out using Perkin-Elmer® PCR thermocycler.

The PCR reaction profile used for the samples analyzed are described in the following table.

Reaction profile
<i>Initial denaturation</i>
15 seconds at 98° C
<i>40 cycles of the following profile</i>
10 seconds at 98° C
10 seconds at 55 ° C
15 seconds at 72 ° C
<i>Final extension</i>
1 minute at 72C°

The presence of amplified DNA fragments (PCR products) was verified by electrophoresis. A 2% agarose gel was prepared by placing 1 g of agarose in 50 ml of a solution TAE 1x. This solution was heated in a microwave oven until it assumed a transparent appearance. Finally, 1 ml of DNA dye was added (GelRed™ : a sensitive, stable and safe for the environment dye which has replaced the historical carcinogenic ethidium bromide) to develop DNA fluorescence when illuminated with UV light.

The prepared solution was poured into an electrophoresis bowl in which a suitable comb allows the generation of wells for the loading of samples.

The gel solidified at room temperature, was transferred into the electrophoresis tank containing running buffer, a 1x TAE solution.

To facilitate the loading of samples into the wells and to monitor the electrophoresis to each sample was added a dye (made from glycerol and TrackIt™ Cyan / Orange Loading Buffer). Moreover, an appropriate reference ladder (ie a sample of DNA fragments of known molecular weight) was also loaded to check the quality of the run and the size of the amplified fragments.

The electrophoretic run was carried out with the normal equipment available for this purpose (Figure 3). at 50 volts for 1 hour and 30 minutes.

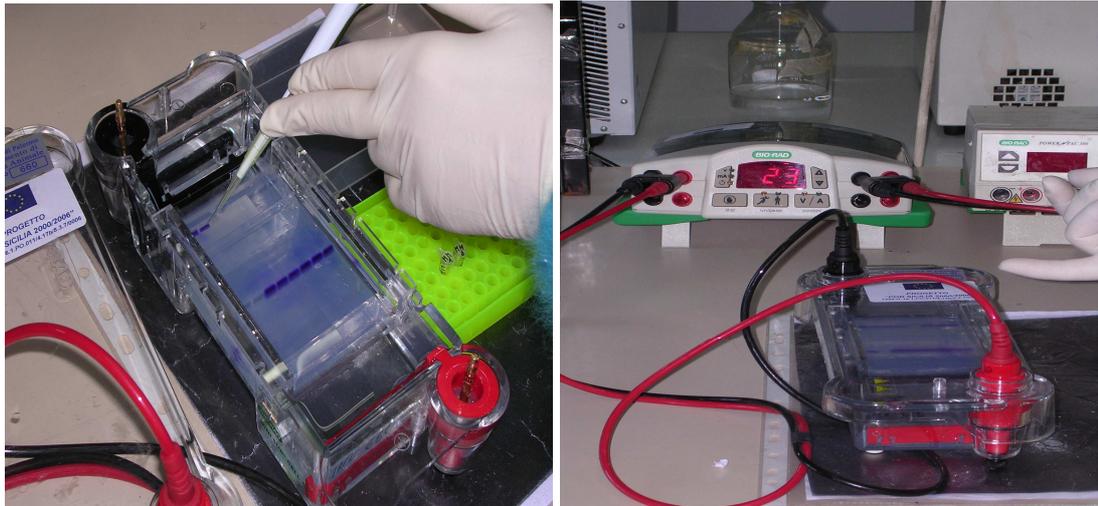


Figure 3: On the left: loading of PCR product samples into the wells of the agarose gel; on the right: electrophoretic scroll.

After the electrophoretic run the gel was observed at the UV transilluminator. The bands, related to the COI region amplified by PCR after electrophoretic flow, revealed a dimension of approximately 650 bp (Figure 4).

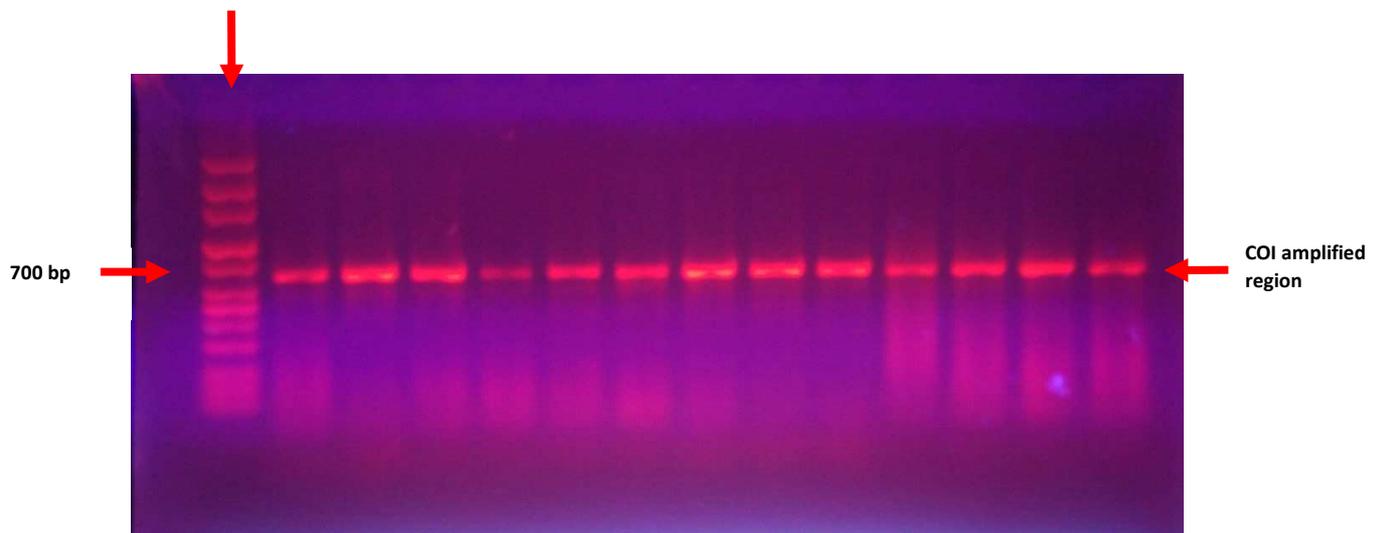


Figure 4: Photography of agarose gel (2%) observed at the UV transilluminator about the COI region. In all samples loaded, after electrophoretic run it has highlighted a distinct band of about 700 bp.

The PCR products appear as discrete bands inside the gel. This bands were cut from the gel with a scalpel and subsequently purified through "illustrata GFX PCR DNA and Gel Band Purification Kits". (GE healthcare Life Sciences).

The purified DNA extracted from the band was used to prepare samples for automated sequencing entrusted to MACROGEN[®] in Holland.

The display of the sequences produced by automated sequencers are shown as a graph, chromatogram (Fig.5), containing a succession of peaks infour-colors. Each peak represents a nucleotide, according to international conventions adenine corresponds to the green color, the guanine to the blue, the thymine to the red, the cytosine to the yellow, and all together correspond to the nucleotide base succession of the sequenced fragment. The chromatograms were analyzed by the Chromas program. Despite the program analyse the sequences in an automatic way sometimes there were some errors or poor quality of the peaks, so it is necessary to revise these inaccuracies made by the automated sequencer. The operator's experience in the chromatogram reading allows a higher accuracy in the interpretation of the sequences; so the chromatograms were always checked and manually corrected when necessary.

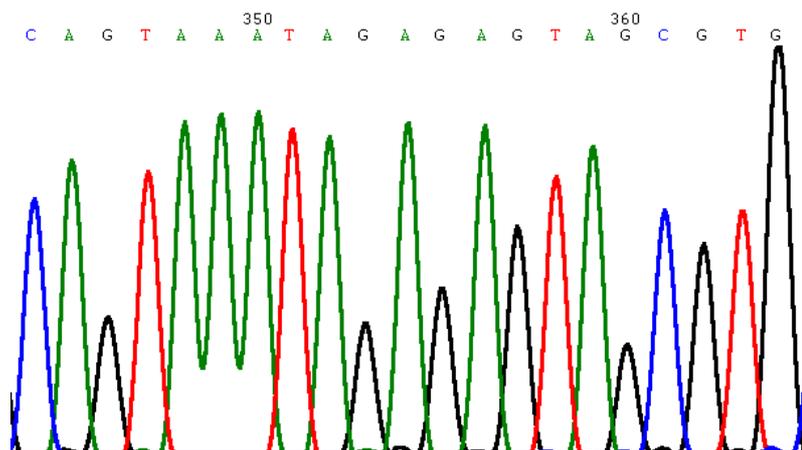


Figure 5 : Part of a chromatogram analyzed through Chromas software.

The sequences obtained represent the barcode of the analyzed species and together with the other information collected for each sample (sequence, photo, place and date of sample collection) were added to the information in the databases BOLDs (<http://www.boldsystems.org/>) and NCBI (<http://www.ncbi.nlm.nih.gov/>) available on line.

Concluding Remarks

The molecular approach in taxonomy, independently from identification based on morphology, has become crucial when there are no expert taxonomist skills, or in the presence of rare or invasive species (Bergsten et al., 2012). However, the ability to identify the species from tissue of unknown origin, presupposes the existence of a DNA barcode database when the entire specimen identification is not possible.

A reference library of barcode DNA is a robust support structure, for a wide variety of applications, as authentication of fish products (Barbuto et al., 2009; Hajibabaei M et al., 2006; Hanner et al., 2011), biosafety (Armstrong & Ball, 2005), manipulation of protected species or those whose catch is regulated (Rasmussen & Morrissey 2008), monitoring (Costa & Carvalho, 2007) or fisheries management (; Hanner et al., 2011). In fact, the main limitation is the relative insufficiency COI sequences of the species present in the reference database (Cerutti-Pereyra et al., 2012). The desirable expansion of DNA-barcode reference library for the marine fish fauna will probably reveal common misidentifications and taxonomic ambiguities, thus it will provide opportunities to review and clarification.

The continual growth of the reference dataset will also facilitate the comparisons between regions and will contribute to the detection of unusual levels divergence among populations. It will also allow the discrimination between ecophenotypes and genotypes, the detection of cryptic species, the analysis of geographical patterns of distribution of species, the investigation about invasive species or the development of a mini-barcode system for the unique identification of the species of economic and commercial interests.

References

- Armstrong KF, Ball SL (2005) DNA barcodes for biosecurity: invasive species identification. *Philos T Roy Soc B* 360: 1813–1823.
- Barbuto M, Galimberti A, Ferri E, Labra M, Malandra R, et al. (2009) DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of ‘palombo’ (*Mustelus* spp.). *Food Res Int* 43: 376–381.
- Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monaghan MT, et al. (2012) The effect of Geographical Scale of Sampling on DNA Barcoding. *Syst Biol* 61: 1–19.
- Bianchi CN, Morri C (2000) Marine biodiversity of the Mediterranean Sea: situation, problems and prospects for future research. *Mar Pollut Bull* 40: 367–376.
- Cerutti-Pereyra F, Meekan MG, Wei N-WV, O’Shea O, Bradshaw CJA, et al. (2012) Identification of Rays through DNA Barcoding: An Application for Ecologists. *PLoS ONE* 7: e36479. doi:10.1371/journal.pone.0036479.
- Costa FO, Carvalho GR (2007) The Barcode of Life Initiative: synopsis and prospective societal impacts of DNA barcoding of fish. *Genomics, Society and Policy* 3: 29–40.
- Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, et al. (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Mol Ecol Notes* 6: 959–964.
- Hanner R, Becker S, Ivanova NV, Steinke D (2011) FISH-BOL and seafood identification: Geographically dispersed case studies reveal systemic market substitution across Canada. *Mitochondrial DNA* 22: 106–122.
- Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identification through DNA barcodes. *Proc R Soc Lond B* 270: 313–322.
- Lo Brutto S, Arculeo M, Grant WS (2011) Climate change and population genetic structure of marine species. *Chem Ecol* 27: 107–119.
- Rasmussen RS, Morrissey MT (2008) DNA-Based Methods for the Identification of Commercial Fish and Seafood Species. *Compr Rev Food Sci Food Saf* 7: 280–295.
- Stoeckle, M. Y.; Hebert, P.D.N. (2008): *Le scienze*. Dicembre, pp 70-74.
- Tortonese, E., 1956: *Leptocardia, Ciclostomata, Selachii*. Vol II. Ed. Calderini, Bologne, Italy, pp 69-74; 87-88.
- Tortonese, E., 1970: *Osteichthyes – Pesci ossei*, Vol X. Calderini Ed, Bologne, Italy, pp. 45-49;55-59.
- Tortonese, E., 1975: *Osteichthyes – Pesci ossei*, Vol XI. Ed. Calderini, Bologne, Italy, pp. 636.
- Ward RD, Hanner R, Hebert P.D.N. (2005) The campaign to DNA barcode all fishes, FISH-BOL. *J Fish Biol* 74: 329–356.

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

Ward, R. D.; Zemlak, T. S.; Innes, B. H.; Last, P. R.; Hebert, P. D. N. (2005) DNA barcoding Australia's fish species. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 1847–1857.

Wilson, E.O., 2005: Systematics and the future of biology. Proceeding of the National Academy of Science. 102, 6520–6521.

Thanks

- Project “Nuove Rotte: Blue Economy” Piano Sviluppo di Filiera PO FESR Sicilia 2007/2013 – Obiettivo Operativo 5.1.1, Linea d'intervento . TUTELA E VALORIZZAZIONE DEI PRODOTTI ITTICI FRESCHI, DI ALLEVAMENTO E TRASFORMATI 5.1.1.2

- Project Bandiera RITMARESP2_WP4_AZ2_UO04_D06. A. Cuttitta.

- Project: “Tecnologie e processi per il miglioramento della shelf-life dei prodotti del comparto agroalimentare attraverso l'uso di film edibili innovativi a base pectinica” (“PON FILM-EDIBILI”, Cod. PON01_02286 - CUP: B68F12000360007.

- Project: " Sistema di Comunicazione Informazione e Diffusione dell' Osservatorio della Biodiversità della Sicilia". PO FESR 2007/2013 linea di intervento 3.2.1.2 Regione Siciliana - Assessorato Regionale Territorio e Ambiente - Dipartimento Regionale dell' Ambiente Progetto