UNIVERSITA' DEGLI STUDI DI MILANO-BICOCCA
Facoltà di Scienze Matematiche, Fisiche e Naturali
Ph.D. in Biology – Cycle XXIII

DNA BARCODING: A LINK BETWEEN BASIC AND APPLIED SCIENCE

Project Supervisor: Dott. Maurizio CASIRAGHI

PhD thesis by:
Andrea GALIMBERTI
Matr. N° 040536

Academic years: 2007/2008 - 2009/2010

INDEX

ABSTRACT

DNA barcoding is a recent and widely used molecular-based identification system that aims to identify biological specimens, and to assign them to a given species. However, DNA barcoding is even more than this, and besides many practical uses, it can be considered the core of an integrated taxonomic system, where bioinformatics plays a key role. Quite soon since its development (in 2003) it became clear that DNA barcoding was suitable for two different purposes: (*i*) the molecular identification of already described species and (*ii*) the discovery of undescribed species (the so called 'DNA taxonomy'). However, such a method has generated a vast debate in the scientific community, which has been from the beginning, deeply divided into pros and cons.

The main objective of this research project was to investigate the strength of coherence reached in combining a standardized molecular methodology with classical biological information (e.g. morphology, ecology, host specificity), toward the synthesis of an integrated approach to taxonomy. In order to satisfy this requirement, nine case studies encompassing a wide panel of taxa (i.e. animal, plant and environmental samples) subjected to different taxonomic uncertainties or potentially dealing with economical, conservation or health implications (e.g. food traceability, parasites infectiveness, etc.) have been investigated.

More than 500 hundreds biological samples were collected directly in the field or retrieved from museum, herbariums or other institutional collections, allowing to create a synergic network among different disciplines and research fields. Standardization in the collection and processing of biological samples, as well as in the bioinformatic approaches used to manage and analyse molecular data has been a fundamental point in the experimental workflow we adopted.

The results obtained with our analyses clearly showed that DNA barcoding represents a powerful tool for taxonomy and it can act as an effective supporting tool for the traceability of food products, for the diagnosis of endoparasites and for the characterization of environmental biodiversity. Although some limitations arise from the incomplete coverage of the existing diversity, the inherent characteristics of the molecular markers adopted as barcodes and other factors, the method showed to be more flexible than expected.

1. INTRODUCTION

1.1 UNKNOWN BIODIVERSITY AND THE RENAISSANCE OF TAXONOMY

Since the advent of Linnaean nomenclature in 1758, taxonomists have been describing and naming an astounding number of species. A clear example concerns animals: currently around 15.000-20.000 new species are described every year (Polaszek et al., 2005; Zhang, 2008). This trend is rapidly increasing for many groups of organisms, mainly due to the availability of new tools and technologies and the recent exploration of poorly known areas of the planet (Köhler et al., 2005; Padial et al., 2006; Sangster G, 2009; Vieites et al., 2009). Despite these recent advances, it should be acknowledged that the great part of species currently extant on our planet are far to be completely characterized; at least concerning eukaryotes (Padial et al., 2010). The total number of known eukaryotic species is currently estimated as ~1.8 to 1.9 million, (may be 1.6 million or fewer if synonymies are removed; May & Harvey, 2009). Despite some classes are presumed to be known very well in terms of number of species (e.g. birds and mammals) especially in widely investigated regions such as the Western Palearctic, our knowledge of invertebrate biodiversity (insects, helminths, and others) is largely inadequate. In this context, credible estimates of the true number of only eukaryotic species raise a total of 10 million, but suggestions as low as 3 million or as high as 100 million can be defended with several implications on surveillance and conservation of many endangered species (Blaxter, 2003; May & Harvey, 2009; see also Figure 1.1). It is clear that to be adequately monitored and protected, a taxon requires first to be described, classified and named; this is the main task of taxonomy. There is little doubt that the central unit for taxonomy is the species, and that associating scientific names unequivocally to species is essential for a reliable reference system of biological information (Wheeler, 2004; Padial et al., 2010).

Traditionally, species identification and classification have been the specialist domain of taxonomists, providing a nomenclature and a several key prerequisites for numerous biological studies. The identification of species depends on the knowledge held by

taxonomists, whose work cannot cover all of the possible taxa for which identification could be requested by non-specialists.

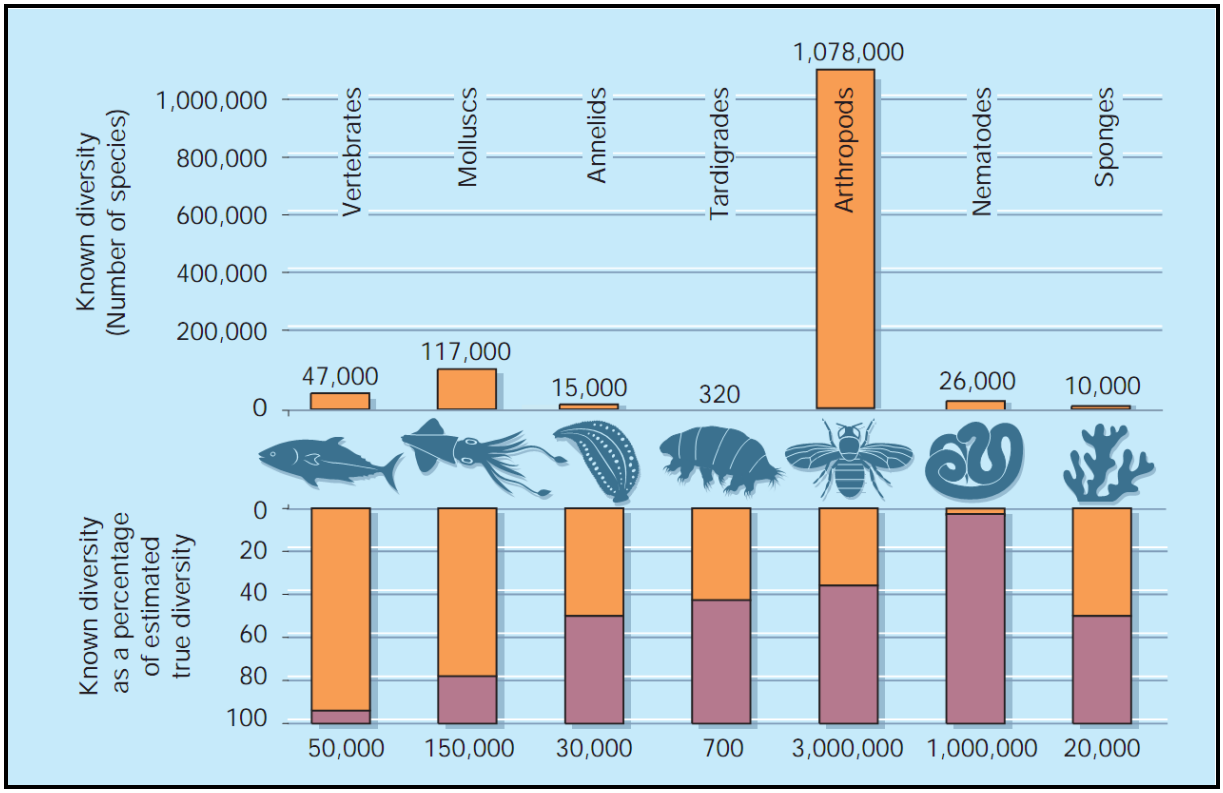


Figure 1.1: Known and estimated diversity of a selection of animal phyla. The histogram above each representative image indicates known diversity (in terms of number of species). That below represents known diversity as a percentage of estimated true diversity, shown by the number at the bottom. [Image modified from Blaxter, 2003]

It has been estimated that few taxonomists can reliably diagnose even 1,000 species and if we hypothesize to deal with a number of species on Earth comprised from 10 to 100 million, it means that up to 100,000 taxonomists should be required simply to sustain the ability to recognize them, once they have all been described. It is trivial to observe that this number of taxonomists is nowadays far to be available. Moreover, the recognition of several taxa is an activity that often requires big amounts of time and resources for a taxonomist. A clear example could be found in the context of investigations on meiofaunal diversity and in particular nematodes for which morphological assignment of individuals to named species is often not technically possible due to sheer abundance, small size, and lack of expert knowledge of the groups encountered. As terrestrial nematodes can easily exceed one million

individuals per square metre of soil, any attempt to exhaustively describe a local nematode fauna will become a difficult and time-consuming work. In addition, many taxa can be diagnosed only from adult male or female-specific structures, or from a comprehensive population survey of some morphological characters. In such cases, larvae, individuals of the 'wrong' sex or individual specimens may not be identifiable therefore causing a partial or even insufficient characterization of soil diversity (Floyd et al., 2002). A similar situation could be observed in many other animal taxa (e.g. rotifers, crustaceans) that are largely unknown with several species (and even higher taxonomic categories) that have yet to be formally described morphologically. Thus, to overcome the clear limits of classical taxonomy, a robust and transferable system of identification, applicable to all individuals and taxa, is sorely needed. This is especially true if requirements of today's society related to the need of resolving crucial biological issues are considered such as to preserve biodiversity, to ensure biosecurity, to protect species and to avoid pandemics (Frézal et al., 2008).

In the last decades, advances in molecular biology and bioinformatics helped resolving some of these difficulties and contributed to increase the process of discovering of new taxa (even clarifying taxonomic uncertainties discussed/debated so far) through the development of standardized, rapid and inexpensive methods even accessible to non-specialists. In particular, the use of DNA sequences in taxonomy dates back 30 years to when ribosomal RNA probes were developed for the identification and phylogenetics of eubacteria and archaebacteria (Fox et al., 1980). The idea of optimizing a molecular identification system for species recognition that was as standardized as possible emerged progressively during the 1990s with the development of PCR-based approaches. At the beginning, molecular identification was largely applied in the context of bacterial studies, in order to avoid the classical time-consuming culture methods usually performed for microbial biodiversity surveys (e.g. Woese, 1996; Zhou et al., 1997) and routine pathogenic strains diagnoses (e.g. Maiden et al., 1996; Sugita et al., 1998; Wirth et al., 2006). The use of PCR-based methods moved then to the application in multidisciplinary fields related to eukaryotes such as taxonomy, food traceability, forensic identification, medical and veterinary diagnostic of pathogens and vectors (e.g. Baker & Palumbi, 1994; Sperling et al., 1994; DeSalle & Birstein, 1996). Some attempts of developing universal systems for molecular-based identification

have been proposed for specific groups (e.g. nematodes, Floyd et al., 2002) although they were not successfully implemented to cover a wide range of taxa (Frézal & Leblois, 2002).

This large-scale application of molecular data has progressively led to a revolution of taxonomy (Savolainen et al., 2005) providing evidences for the existence of thousands of new candidate species. Thanks to these laboratory progresses, modern taxonomy is now resurging as a solid scientific discipline incorporating several technological and methodological tools (e.g. virtual access to museum collections or high-throughput DNA sequencing) toward a progressive digitalization of taxonomic data which direct outputs have been diffusing through several global initiatives, such as Species2000 - http://www.sp2000.org/ (Padial et al., 2010). Some authors envisioned the next future as a sort of "cybertaxonomy" with online description and publication of new species, and where updated taxonomic information would be publicly accessible (Schram, 2004). However, despite these highly positive claims, modern taxonomy still works in attempt to balance two major requirements: to provide empirical rigor to species hypotheses and nomenclature, which lead to a careful and often time-consuming labor of species delimitation and, on the other hand, an acceleration in the 'rate' of species description aware of the peril to encounter erroneous species hypotheses and thus to provide unstable nomenclature. Validity and practicalities of these modern molecular approaches have in fact been subject to a variety of criticisms concerning for example the way to analyze sequence data and their integration in a taxonomic context (Vogler & Monaghan, 2006; Padial et al., 2010).

In order to deal with this lack of a general consensus on the best way to manage and standardize molecular data in a "modern taxonomy" context, in 2003, researchers at the University of Guelph in Canada published a paper proposing a new system of species identification and discovery using a very short genetic sequence from a standard part of the genome (Hebert et al., 2003). They called the new molecular identification system "DNA barcoding".

1.2 THE DNA BARCODING REVOLUTION

DNA barcoding is a molecular and bioinformatic tool that aims to identify biological specimens and to assign them to a given species (Hebert et al., 2003). The basic idea is quite simple: through the analysis of the variability in the nucleotide sequence of a single (or few) standard molecular marker(s), usually named as 'barcode(s)' it is possible to discriminate biological entities (hopefully belonging to the species taxonomic rank). The name of this approach figuratively refers to the way a supermarket scanner univocally distinguishes products using the black stripes of the Universal Product Code (UPC). Similarly, DNA barcoding is based on the assumption that every species will most likely have a unique DNA barcode; indeed there are 4^{650} possible ATGC-combinations compared to an estimated 10 million species remaining to be discovered (Wilson, 2004; see also Figure 1.2).

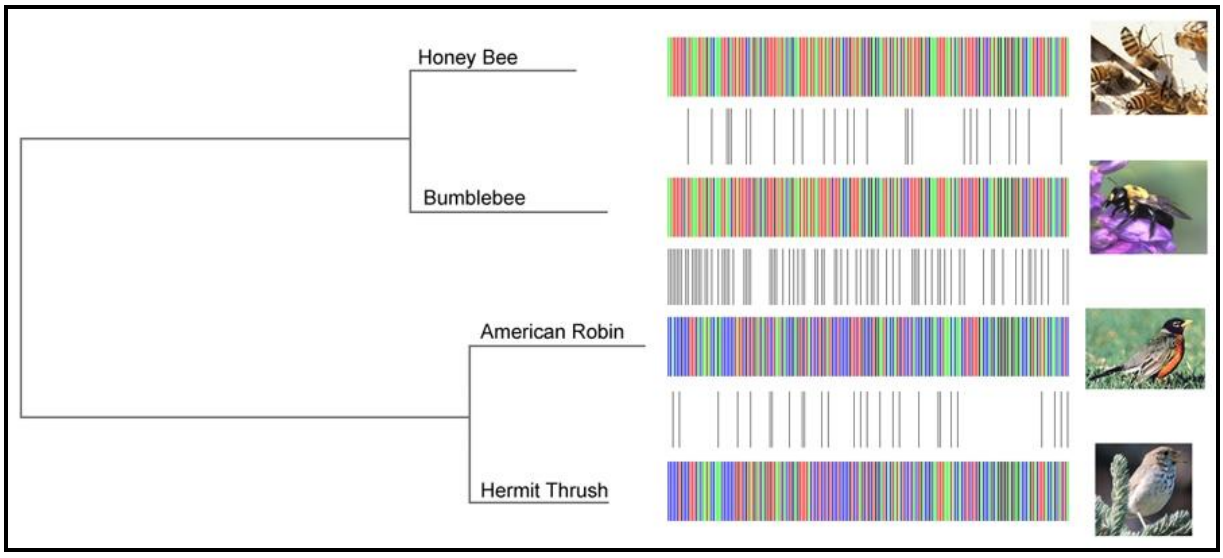


Figure 1.2: Schematic representation of differences in barcode sequences among different animal groups (e.g. insects and birds). Each coloured bar on the right corresponds to the nucleotide (i.e. A,T, C, G) occurring in a particular position within the amplified barcode region (e.g. the mitochondrial *coxI*). Black bars represent the variable sites between each pair of compared barcodes. On the left a NJ reconstruction of these differences is reported. [Image modified from http://www.eurekalert.org/pub_releases/2004-09/plos-atc092104.php]

An ideal DNA barcode require two fundamental characteristics in order to work properly: high taxonomic coverage and high resolution (Ficetola et al., 2010). A high taxonomic coverage (also called universality) would allow the applicability of the gene chosen as DNA

barcode to a number of taxa as large as possible (including undescribed or rare species), while a high resolution capacity refers to the ability of a certain barcode to differentiate species and relies on the amount of interspecific differences among DNA sequences. As an example, the typical DNA barcode region for metazoans is the mitochondrial cytochrome c oxidase 1 (*coxI*) and in particular it consists of a 648 bp region 58–705 from the 5' end of the gene (using the mouse mitochondrial genome as a reference) (Frézal et al., 2008).

Another assumption is that the molecular marker chosen as barcode should show a genetic variation between species exceeding that within species. Consequently, the ideal DNA barcoding analysis mirrors the distributions of intra- and inter-specific variability separated by a distance called 'DNA barcoding gap' (Meyer & Paulay, 2005; Wiemers et al., 2005; see also Figure 1.3).

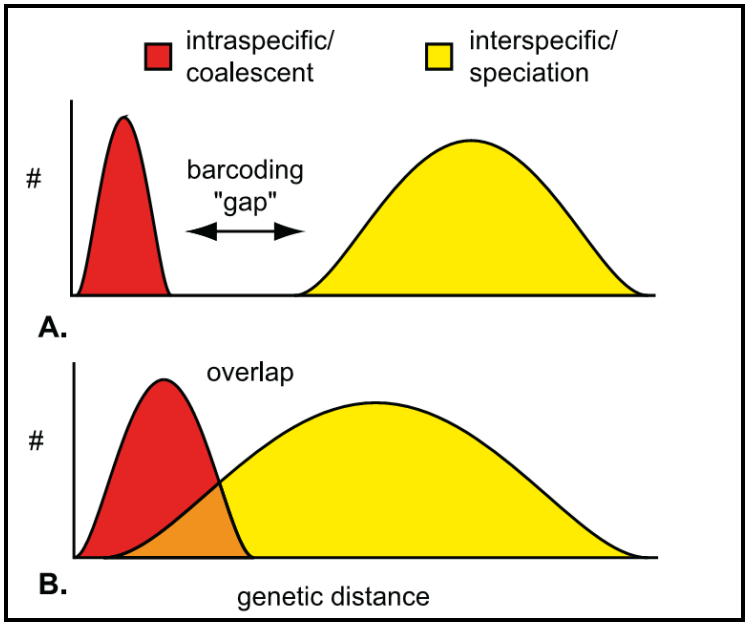


Figure 1.3: Schematic of the inferred 'barcoding gap' according to Meyer & Paulay (2005). The distribution of intraspecific variation is shown in red, and interspecific divergence in yellow. (A) Ideal world for barcoding, with discrete distributions and no overlap. (B) An alternative version of the world with significant overlap and no gap.

The original idea was to apply DNA barcoding systematically to all metazoans, by the use of one or few (mitochondrial) markers (e.g. *coxI* – Hebert et al., 2003). Rapidly, but with less coherent results, the idea was extended to flowering plants (Kress et al., 2005; Hollingsworth et al., 2009) and fungi (Min et al., 2007), and now the DNA barcoding initiative can be considered a tool suitable for all the tree of life branches (even for Bacteria and Archaea,

using a multi-locus sequences typing, see Unwin et al., 2003; Casiraghi et al., 2010), despite some problems with the discriminatory power of barcode(s) region(s) in plants and other taxa occur (see paragraph 1.8). In support of this emerging trend in the use of such a kind of molecular system, it should be considered that while the firsts studies involving DNA barcoding focused only on local faunas (Hogg & Hebert, 2004; Ball et al., 2005) these are now dealing with a broader set of fields of applications. As an example, DNA barcoding has been already used in biodiversity assessment, forensics, food traceability, diet analysis and paleoecological studies (Valentini et al., 2009a; Dawnay et al., 2007; Barbuto et al., 2010; Valentini et al., 2009b; Willerslev et al., 2003).

Over the past seven years, a number of campaigns have started to collect and register DNA barcodes from specific families and regions of life. All of these global or regional campaigns work closely with the 'The Consortium for the Barcode of Life' (CBOL - http://www.barcodeoflife.org/), the recently launched 'International Barcode of Life Project' (iBOL - http://ibol.org/) and the 'Barcode of Life Database Systems' (BOLD - http://www.barcodeoflife.org/), an online workbench that aids collection, management, analysis, and use of DNA barcodes towards the ultimate goal of a barcode reference library of all life on Earth (Ratnasingham & Hebert, 2007). A list of the most important campaigns could be found in Table 1.1.

Table 1.1: List of global DNA barcoding campaigns

PROJECT NAME	LOGO	TARGET ORGANISMS	SPECIES BARCODED	EXTANT SPECIES	PROJECT WEBSITE
ABBI		Birds	3,445	9,933	http://www.barcodingbirds.org/
Formicidae barcode of Life		Ants	798	12,205	http://www.formicidaebol.org/index.php
Trichoptera Barcode of Life		Caddisflies	2,474	13,165	http://trichopterabol.org/

PROJECT NAME	LOGO	TARGET ORGANISMS	SPECIES BARCODED	EXTANT SPECIES	PROJECT WEBSITE
Coral Reef Barcode of Life		Coral Reef fauna	5,431	16,807	http://www.reefbarcoding.org/
Fish-BOL		Fishes	7,989	31,220	http://www.fishbol.org/
Lepidoptera Barcode of Life		Butterflies and Moths	48,676	165,000	http://www.lepbarcoding.org/index.php
Mammalia Barcode of Life		Mammals	858	5,426	http://www.mammaliabol.org/index.php
Marine Barcode of Life		Marine fauna	6199	55451	http://www.marinebarcoding.org/

Table 1.1 *continued*

1.3 UTILITY OF A DNA BARCODING APPROACH

DNA barcoding has generated a vast debate in the scientific community, which has been from the beginning, deeply divided into pros and cons. Indeed, since its development in 2003, it became clear that DNA barcoding was suitable for two different aims: (i) the molecular identification of already described species (Hebert et al., 2003) and (ii) the discovery of undescribed species (Hebert et al., 2004). It is important to underline that the validity of this last purpose (also called 'DNA Taxonomy') would be carefully considered because DNA barcoding in itself cannot act as an absolute criterion to describe new taxa; however, it can be used to support and integrate other diagnostic characteristics (e.g. geography, morphology) (DeSalle et al., 2005; Vogler & Monaghan, 2006).

The most important advantages of DNA barcoding comes from the rapid acquisition of molecular data with relatively low costs of the analyses. As a contrast, classical identification systems based on morphological data require large amount of time and sometimes it could lead to results totally confusing. In other cases it could be almost impossible to reach a certain identification [for example in the case of dinoflagellates (e.g. Litaker et al., 2007) or diatoms (Evans et al., 2007)]. Furthermore, the use of a DNA barcoding approach could help resolving some of the most frequent identification difficulties that occurs in a biological context. As an example, it could be useful to identify damaged organisms from parts/pieces (e.g. endoparasites recovered by physicians or veterinarians, stomach extracts, slices of food) with several implications in the fields of forensic sciences, food traceability, diet analyses and in the monitoring of illegal trade of endangered species. A molecular-based identification also allows to identify a species independently from its life history stage or gender when they substantially differ in morphology, living behaviors and habitat (e.g. fish larvae, Pegg et al., 2006; amphibians, Randrianiaina et al., 2007; coleoptera, Caterino and Tishechkin, 2006; fungal sexual stage, Shenoy et al., 2007). A third category of problems that requires the use of molecular tools (e.g. DNA barcoding) is when morphological traits cannot successfully discriminate species (e.g. red algal species, Saunders, 2005; fungal species, Jaklitsch et al., 2006), also in the case of small-size organisms (for which an easy visual identification is often precluded; see Floyd et al., 2002; Blaxter et al., 2005).

The cost and time-effectiveness of DNA barcoding and the recent development of new-generation sequencing technologies (i.e. pyrosequencing) allow a certain degree of automation in species identification, which is particularly useful in large sampling campaigns (e.g. Rusch et al., 2007) or in the emerging field of metagenomics to characterize biodiversity of environmental matrixes (e.g. soil, water; see for example Valentini et al., 2008; Porazinska et al., 2009).

In conclusion it can be assumed that the big leap of DNA barcoding resides not only in the discrimination power itself, but also in the conjugation of three innovations of modern taxonomy: (i) molecularization (i.e. the use of the variability in a molecular marker as a discriminator); (ii) computerization (i.e. the not redundant transposition of the data using

informatics supports) and (iii) standardization (i.e. the extension of the approach to vast groups of organisms not deeply related) of the taxonomic approach. While molecularization (Tautz et al., 2003; Pennisi, 2003; Lee, 2004) and computerization (Bisby et al., 2002; Godfray, 2007; Kress & Erickson, 2008) have been independently present in the taxonomic world for a long time, standardization was randomly present in the taxonomic world, and can be considered the main contribution given by the DNA barcoding approach (Frézal et al., 2008; Casiraghi et al., 2010).

1.4 THE DNA BARCODING LANDSCAPE

As a global research initiative, DNA barcoding project encompasses some of the characteristics of large, coordinated, projects (like the Human Genome), and some characteristics of more traditional (and often individualistic) taxonomic research projects. The goal of DNA barcoding is the construction of an enormous, online, freely available sequence database of (at least) all eukaryotes although barcoding research-lines are often done by researchers who are focusing on one or a few taxonomic group(s) (see also Figure 1.4). In order to properly integrate the numerous outputs produced by those individual projects into the global initiative a sort of 'standardization code' has been developed in collaboration between the Database Working Group (i.e. a branch of the DNA barcoding community) and the International Nucleotide Sequence Database Collaboration (INSDC). This document provides a list of important directives concerning how to produce and archive DNA barcoding data in dedicated databases (http://www.barcodeoflife.org/sites/default/files/legacy-/pdf/DWG_data_standards-Final.pdf).

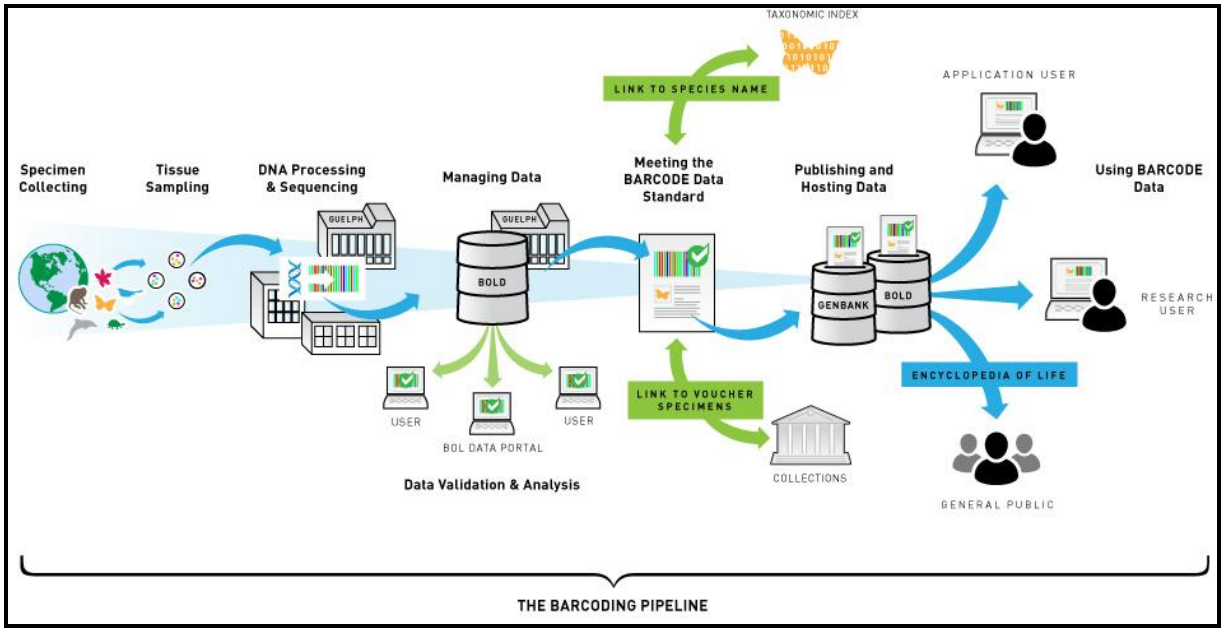


Figura 1.4: The DNA barcoding pipeline. [image retrieved from http://www.barcodeoflife.org/]

Participants in the DNA barcoding initiative can hold different configurations, including consortia, databases, networks, labs, and projects ranging in size from local entities ranging in size from local to global scale. Consortia and databases play a key-role in the DNA barcoding system. The most important and largest consortia are can be listed as follow:

- iBOL (http://ibol.org/), the International Barcode of Life Project. Founded in 2010; it is organized by the Biodiversity Institute of Ontario at the University of Guelph. This consortium encompasses 25 nations and 10 working groups specialized on different taxonomic groups or habitat types. The aim of iBOL is to produce 5 million barcode records from 500,000 species by 2015. iBOL's partners consist of national, regional and central nodes, each of which is a network of projects, institutions and labs.

- CBOL (http://www.barcodeoflife.org/), the Consortium for the Barcode of Life. Established in 2004, it is based at the Smithsonian Institution's National Museum of Natural History (Washington, DC). Differently from iBOL, CBOL is not involved in the generation of barcode data but it represents the designated lead organization for iBOL's working group for outreach and collaborations. The CBOL's network currently counts on more than 200 Member Organizations from 50 countries among which the University of Milano-Bicocca is

one of the most active partners in Italy.

Concerning databases, two central platforms are currently used to deposit DNA barcode sequences:

- GenBank, EMBL, and DDBJ which have comprised the International Nucleotide Sequence Database Collaboration (INSDC, see above), are the permanent public repositories for barcode data. As previously stated, DNA barcoding data are meant to be easily and widely accessed. In order to gain this purpose a dedicated sequence submission procedure is available for GenBank (http://www.ncbi.nlm.nih.gov/WebSub/?tool=barcode). This procedure has slightly modified the standard procedures, introducing a 'DNA barcoding' label to the sequences in order to simplify database querying and searching.

- BOLD (http://www.boldsystems.org/views/login.php). The Barcode of Life Data System, located at the University of Guelph is a public interactive workbench that aids collection, management, analysis and use of DNA barcodes. BOLD (as like as the members of the INSDC) interact with other databases of voucher specimens and taxonomic names (e.g. museums, herbaria). The system is still in constant evolution and update. By January 2011, the total number of available DNA Barcode records were at 1,067,533 *coxI* sequences belonging to 91,840 species, of which 229,067 *coxI* sequences (belonging to 30,154 species) satisfied DNA barcoding criteria (i.e. minimum sequence length of 500bp and a minimum of three individuals barcoded per species; see Ratnasingham & Hebert, 2007). Moreover, the platform consists of some components among which two are the most useful for the public use of the database (see Ratnasingham & Hebert, 2007):

* MANAGEMENT & ANALYSIS TOOL (BOLD-MAS) provides a standardized archive for barcode data coupled with analytical tools (e.g. distance summary, accumulation curve) which allow researchers to assemble, test and analyze their data before uploading them to GenBank (or to one of the others international databases cited above).

* IDENTIFICATION ENGINE TOOL (BOLD-IDS) provides a species identification tool that accepts DNA barcode sequences and returns a taxonomic assignment to the species level whenever possible.

With respect to a classical sequence submission procedure (http://www.ncbi.nlm.nih.gov/genbank/-submit.html), both the databases described above (i.e. GenBank and BOLD) require additional data in order to link barcode data to their voucher specimens. This standardization is mirrored by the establishment of the Registry of Biological Repositories initiative (http://www.biorepositories.org/), an online registry of organisms linked to DNA sequences.

1.5 WEAK POINTS OF DNA BARCODING

DNA barcoding has deeply impacted the scientific community becoming a widely used approach. Moreover, it has generated a vast debate which has been from the beginning, deeply divided into pros and cons (Mallet et al., 2003; Moritz & Cicero, 2004; Will & Rubinoff, 2004; Will et al., 2005; Waugh, 2007; Frézal et al., 2008, Casiraghi et al., 2010). As a diagnostic method, DNA barcoding approach can be more or less fallacious, and it should be taken into account that failures are mainly in the essence of biological species, in the patterns of molecular evolution, in the completeness of sampling, in the hybridization events and in the heteroplasmy of sequences from different tissues rather than in the method (Bensasson et al., 2001; Funk and Omland, 2003; Hurst & Higgins, 2005; Rubinoff et al., 2006; Ekrem et al., 2007; Song et al., 2008; Meier et al., 2008; Naciri et al., 2009; Zhang et al., 2009; Magnacca et al., 2010). Some of these issues deserve to be further remarked. As an example, in a DNA barcode reference database (e.g. BOLD), an unbalanced taxonomic sampling from within and among populations of a certain species could lead to uncertain identifications (Lefébure et al., 2006) due to the fact that the individuals chosen to represent each taxon do not properly cover the major part of the existing molecular variability (Rubinoff et al., 2006; Sungmin et al., 2010). Another problem related to sampling may arise in those taxa for which total biodiversity has been not yet completely characterized (e.g. insects); exuberant species richness may severely constrain the ability of the DNA barcode reference databases to adequately represent the overwhelming taxonomic diversity (Virgilio et al., 2010). Some risks in the discriminatory power of DNA barcoding could also be related to the maternal inheritance of mitochondrial markers. The amount of mtDNA diversity is strongly linked to the influence of ecological or biological constraints on females that could

lead to an overestimation (or underestimation) of sample divergence therefore rendering conclusions on taxonomic status unclear (Petit et al., 2009). Recently, it has been demonstrated that mitochondrial inheritance within species can also be confounded by symbiont infection (e.g. *Wolbachia*; see Hurst & Higgins, 2005; Frézal et al., 2008).

Finally, one of the main issues affecting the aimed standardization of DNA barcoding is the absence of a shared analytical procedure to manage barcode data (Casiraghi et al., 2010). In the recent years, the scientific community has indeed produced a large amount of alternative approaches but no consensus has yet emerged on the best method to integrate DNA-based information with the existing taxonomic system (Vogler & Monaghan, 2006; Padial et al., 2010).

1.6 DNA BARCODING *SENSU STRICTO* AND *SENSU LATO*

Most of the questions raised by the use of DNA barcoding are directly linked to the essence of an identification method. As stated by Helbig et al. (2002), species could be defined only in relation to other species and thus, in a strict sense, to identify means simply to differentiate (Casiraghi et al., 2010). The choice of the discriminator is essential, because it is (almost) always possible to differentiate: the difficulty is in giving a biological meaning to what it has been discriminated (Vogler & Monaghan, 2006; Padial et al., 2010; Casiraghi et al., 2010). Even if not always fully acknowledged, DNA barcoding implies two different approaches to discrimination. DNA barcoding *sensu stricto* is a simple sorting method that could differentiate (not define) biological entities as a sort of dichotomic key in the traditional taxonomy framework. On the other hand, DNA barcoding *sensu lato* represents a system implementing all the aspects of taxonomy towards the representation of the living world as a whole (e.g. in the field of integrated taxonomy; see chapter 1.10) (Casiraghi et al., 2010).

In the scientific community, the expression 'DNA barcoding *sensu lato*' is generally used in a less comprehensive meaning associated to the idea of 'DNA Taxonomy' (Valentini et al., 2009a; Ficetola et al., 2010). The aim in DNA taxonomy is species circumscription and delineation (Voegler & Monaghan, 2006) and therefore to identify groups of sequences (belonging to one or more regions of mtDNA or nuclear DNA) that correspond to entities of

reproductively coherent individuals (the species) through phylogenetic and clustering methods. DNA barcoding *sensu lato* can thus be useful to assess a species rank to unknown or unrepresented taxa but it should be clear to the users which kind of DNA barcoding philosophy (i.e. *sensu stricto* or *sensu lato*) they are going to use because they rely on assumptions that are substantially different.

A clear schematization of these differences has been provided by Voegler & Monaghan (2006) and is showed in Figure 1.5. As stated by these authors, a first important distinction is to be made in the treatment of the individual organisms as the basic items of the two identification approaches, and the taxonomic entities into which these individuals are grouped.

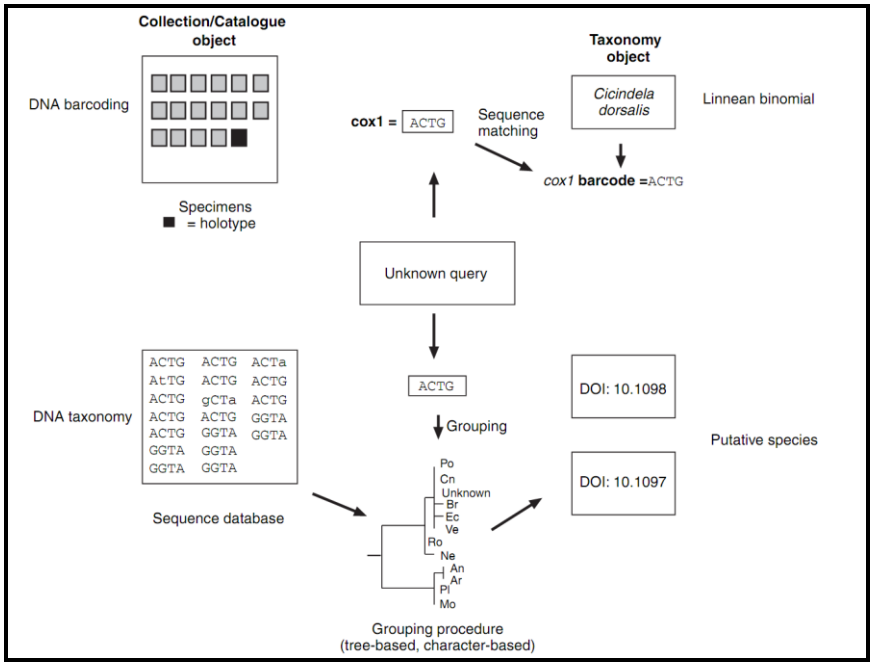


Figura 1.5: Schematic illustration of the differences between DNA barcoding and DNA taxonomy. (1) In traditional taxonomy, the collection/catalogue objects are grouped an named based on morphological traits; the taxonomy objects are the Linnean binomials. Specimen(s) identified as members of a given species are then sequenced to provide the 'barcode'. The identity of a query sequence obtained from an unknown individual (in italics) is determined by a match to the database of barcodes. (2) DNA taxonomy uses the sequences as the primary catalogue/collection objects, which are grouped to represent the taxonomy objects. As these sequences might differ slightly, a range of grouping procedures can be used to identify the species-level entities from the sequence information (e.g a NJ tree). Identification of unknowns is against this set of sequences. The group so defined can be assigned any type of name. Many sets of sequences will correspond to existing Linnean names. [modified from Vogler & Monaghan, 2006]

In the schematization showed above, individuals are treated as 'collection objects' (or 'catalogue objects'), as they represent the objects in a physical collection (e.g. a museum) or the specimen entries in a database, while the 'taxonomy objects' correspond to the Linnean binomials under which the individuals are subsumed. In a typical DNA barcoding approach (i.e. DNA barcoding *sensu stricto*), the collection objects are stored with a Linnean name (the taxonomy object) by means of traditional procedures (e.g. morphology), and the barcode sequence is fitted to the taxonomy object retrospectively. Inexact matches are either grouped with taxa present in the database or identified as new to the database based on whether they fall within a threshold of sequence similarity (Voegler & Monaghan, 2006). In contrast, a 'DNA Taxonomy' approach is conventionally based on the notion that the DNA sequences themselves serve as the taxonomic reference system. In the schematization framework of Figure 1.5, the DNA sequences constitute the catalogue objects, from which the taxonomy objects have to be derived. The latters are groups of sequences assumed to correspond to Linnean binomials in the traditional taxonomy (i.e. they serve as the term to which biological information is being associated; Thiele & Yeates, 2002).

1.7 THE UNIT OF IDENTIFICATION: 'MOLECULAR ENTITIES' VS 'SPECIES' DEBATE

It is well known that no identification method (morphological, biochemical, genetic or whatsoever based) can truly identify species, because species are entities in continuous evolution and it is theoretically impossible to define statically such dynamic matter (Casiraghi et al., 2010). It is important to underline that DNA barcoding does not claim for a new species concept because this is not necessary to the success of the method. However, the choice of a proper species concept is crucial to perform a reasoned analysis of DNA barcode data (Casiraghi et al., 2010). Since its original generalization, DNA barcoding has been following the typological species approach, a concept that theoretically fails because it freezes the evolutionary continuum of species. To cope with this limitations, some development of DNA

barcoding shifted towards other species concepts (e.g. the evolutionary species concept; see Casiraghi et al., 2010).

A critical step in a DNA barcoding approach also relies on the criteria and assumptions used to group barcode sequences in defined entities as their share some characteristics (e.g. sequence similarity, geographical provenance of the specimen). Moreover, another difficult is to provide a correct interpretation of these clusters in a taxonomic or evolutionary context. The entities identified by molecular approaches have been named in several ways: 'Genospecies'; 'Phylospecies', *sensu* Eldredge & Cracraft (1980); 'Recognizable Taxonomic Units', RTUs, *sensu* Oliver et al. (1993); 'Phylotypes' *sensu* Moreira & López-García (2002); 'Molecular Operational Taxonomic Units', 'MOTUs', *sensu* Floyd et al. (2002). To date, MOTUs have been unofficially but widely adopted in a DNA barcoding context. As stated by Floyd et al. (2002) a series of specimens can be considered to belong to the same MOTU when their DNA barcoding sequences are characterized by a certain degree of similarity or relies under a fixed threshold of molecular divergence (e.g. the 1% K2P threshold fixed by the BOLD System, Ratnhasingham & Hebert, 2007). However, in a more general context, the major issue is how close are those, whatever named, molecular entities to what researchers are used to call 'species'. Even if the point has been clearly treated by different authors (see, for instance Blaxter, 2004; Blaxter et al., 2005; Vogler & Monaghan, 2006; Frézal et al., 2008;), a general assumption considers 'molecular entities' and 'species' as synonyms (Casiraghi et al., 2010). Given the fact that DNA barcoding deals with the boundaries among races, varieties, demes, populations and species, some doubts can still arise on the meaning of molecular entities that does not match a described taxon and on the criteria needed to equalize a molecular entity to a given species. This is the (almost) insurmountable problem for DNA barcoding *sensu stricto*: the biological meaning of the identified ranks cannot be directly derived, unless we have clearly and unequivocally linked a species to the variability pattern of a single DNA barcoding marker (e.g. *coxI*). In all the other cases, a DNA barcoding *sensu lato* approach should be used (Dayrat, 2005; De Ley et al., 2005; Ferri et al., 2009).

Moreover, in the context of this research project, we introduced an alternative concept to describe some kinds of molecular entities that we have here tentatively named IOTUs

(Integrated Operational Taxonomic Units). In particular, MOTUs and IOTUs are the results of two quite different DNA barcoding approaches. MOTUs derive from molecular discrimination only, while IOTUs are the result of an integration of different taxonomic tools (elsewhere defined as 'the taxonomic circle'; DeSalle et al., 2005). From the systematic point of view IOTUs are stronger than MOTUs, being one or more steps closer to the species concept (but, it is important to keep it in mind, IOTUs are not yet species).

1.8 UNIVERSALITY AND RESOLUTION OF A DNA BARCODING SYSTEM

As previously stated, among the properties of an ideal DNA barcoding system, high taxonomic coverage and high resolution are the most outstanding (Ficetola et al., 2010). The former factor (also called 'universality') constraints the genomic region chosen as a barcode to present five main characteristics which has been clearly synthesized by Taberlet et al. (2007). First, it should be enough variable to discriminate between species and at the same time, conserved enough to be less variable within species. Second, it should be standardized, with the same DNA region as far as possible used for a wide range of taxa (at least belonging to the same class or phylum). Third, the target DNA region should contain enough phylogenetic information to easily assign species to its taxonomic group (genus, family, etc.). Fourth, it should be extremely robust, with sufficiently conserved flanking regions enabling the design of universal primers and highly reliable DNA amplification and sequencing. This is particularly important when using environmental samples or similar (e.g. soil, stomach contents) where each extract contains a mixture of many taxa to be identified. Fifth, the target DNA region should be short enough to allow amplification of degraded DNA.

Hebert et al. (2003, 2004a,b) argue in favor of a precise portion in 5' end region of the mitochondrial cytochrome c oxidase 1 gene '*coxI'* (i.e. 648 bp region 58–705 from the 5' end of the using the mouse mitochondrial genome as a reference), a choice justified by its great resolving power in several animal groups (e.g. birds, mammals, lepidopterans). Ideally, a single pair of universal primers (e.g. Folmer primers; Folmer et al., 1994) would amplify the DNA barcode locus in almost any animal species. Even if *coxI* has proven to be useful to discriminate species in most groups tested, its limits in some animal taxa are already evident (Meyer & Paulay, 2005; Vences et al., 2005; Shearer & Coffroth, 2008). For these reasons,

there is no consensus in the scientific community and in some animal groups the mitochondrial 16S rRNA or the nuclear ribosomal DNA have also been proposed as barcoding markers (Tautz et al., 2003; Vences et al., 2005). In spite of these problems, *coxI* is the main marker for DNA barcoding purposes in metazoans, as revealed by the high number of published projects (see Project section at http://www.boldsystems.org and Table 1.1). However, the extension of DNA barcoding to other kingdoms is also progressing as like as the idea of a multi-locus DNA barcoding approach (Ferri et al., 2009).

The choice of regions suitable for DNA barcoding has been little investigated in many other eukaryotes. For instance, a marker was already available in fungi: the nuclear ITS region, which has been currently confirmed as the main DNA barcode for this group (Begerow et al., 2010), even if *coxI* has been successfully tested on these organisms (Seifert et al., 2007).

In land plants, compared to metazoans, the situation is much more difficult because their mitochondrial genome exhibits a lower rate of evolution, thus limiting the discriminatory capability of plant *coxI* for species identification (Rubinoff et al., 2006; Mower et al., 2007; Vijayan & Tsou, 2010). In 2007, the CBOL plant working group encouraged the use of a multilocus approach, with a marker serving as an ideal 'anchor' (i.e. universal across the whole plant kingdom as the use of *coxI* for animals) and one or more genomic regions serving as 'identifiers' (i.e. efficient to distinguish closely related species) (Bakker, Second International Barcode of Life Conference TAIPEI, September 2007). The search for an analogous to *coxI* or ITS in land plants, has focused on the plastid genome. Several plastid genes have been proposed, such as the most conserved *rpoB*, *rpoC1* and *rbcL* or a section of *matK* showing a rapid rate of evolution, but in some plant families, these genes showed amplification problems (Hollinsgworth et al., 2009). At the same time, intergenic spacers such as *trnH-psbA*, *atpF-atpH* and *psbK-psbI* were tested for their rapid evolution and a greater resolution among similar taxa (Fazekas et al., 2008; 2009), although they showed standardization and sequence alignment problems in some cases. Despite several combinations of DNA regions have been proposed to date (Kress et al., 2005; Chase et al., 2005; Kress and Erickson, 2007; Pennisi, 2007), the best set of candidate markers for plants DNA barcoding, has not been yet formalized (Pennisi, 2007). In 2009, the Canadian Plant

Barcoding Group, in attempt to standardize as well as possible the plant DNA barcoding system as a global initiative, proposed the use of at least two fixed coding coding genomic regions corresponding to the plastidial *rbcL* and *matK* (CBol Plant working group, 2009), although several studies provided clear evidences for the efficacy of the non coding *trnH-psbA* to discriminate closely related species (Bruni et al., 2010; De Mattia et al., *in press*).

As stated before, DNA barcoding is a standardized method which aimed to be as little as possible, taxon influenced. In order to satisfy this basic requirement, as a general principle, design *ad hoc* primers working on few or even only one genera is not really in agreement with DNA barcoding philosophy (Casiraghi et al., 2010). However, the development of taxon-specific primers and their combinations (the so called 'primer cocktails') could be sometimes necessary to ensure greater intrageneric accuracy (e.g. Neigel et al., 2007), or the widest coverage for particular taxa as in the case of fishes (Ward et al., 2005; Ivanova et al., 2007) and bats (Clare et al., 2007). As a first approach to a DNA barcoding work, before using taxon-specific oligonuclotides, some of the widely used primer pairs and conditions (e.g. Folmer primer LCO1490 – HCO2198) should be tested in advance. This approach should be applied especially when the user is dealing with taxa never tested before, with parts or fragments of an unrecognizable specimen or in the case of environmental matrices (e.g. soil, water, faeces). In BOLD, the registration of primers for the amplification of different barcode regions in animals, plants and fungi is available and encouraged (http://www.boldsystems.org/views/primerlist.php). Quite different is the situation in plants, where users are presently facing most of the difficulties for primer design. As an example, a universal primer combination suitable for all plant species was defined for the plastidial *rbcL* region only (Fazekas et al., 2009). On the contrary, *matK* was analysed in different plants and several reports were published regarding the universality of the primers, ranging from routine success (Lahaye et al., 2008), to more patchy recovery (Dunning et al., 2010).

1.9 BIOINFORMATICS OF DNA BARCODING

The objective of a typical DNA barcoding analysis is relatively simple: to assign each query sequence to a set of referenced (tagged-specimen) sequences belonging to reference

databases (i.e. DNA barcoding *sensu stricto*). A profusion of different bioinformatics approaches is available to reach this aim but, as stated before, there is still no consensus on the best method to analyse DNA barcoding data. This is confirmed by the publication of works in which researchers tested different approaches on the same dataset (Elias et al., 2007; Austerlitz et al., 2009; Kerr et al., 2009). In most of the cases, the result is the same: there is no analytical method outperforming the others, but the 'best method' is case related. In such a dynamic situation it is necessary that users should learn how to properly manage DNA barcoding data in order to avoid errors and incorrect interpretation of the results.

The majority of the published works based on DNA barcoding combine similarity methods with distance tree reconstruction using a Neighbour Joining (NJ) algorithm, with a Kimura 2 parameters (K2P) correction (see for instance Hebert et al., 2004; Ward et al., 2005; Borisenko et al., 2008; Shearer & Coffroth, 2008; Wong & Hanner, 2008). The feeling is that this combination of methods and the adoption of K2P correction has being largely used more by routine than reasoned choice (Casiraghi et al., 2010). Indeed, this approach is at least disputable (Meier et al., 2008), even if the Kimura correction was claimed as the best DNA substitution model for low genetic distances (Nei & Kumar, 2000) and NJ seems to be sufficient to satisfy the requirements of DNA barcoding, because users have to deal with terminal branches and not with deeper branching patterns (Kerr et al., 2009). For instance, in BOLD, the Identification Engine tool (BOLD-IDS) combines similarity methods with distance tree reconstruction in the following way: (i) first, the query barcode sequence is aligned to the global alignment through a Hidden Markov Model (HMM) profile of the *coxI* protein (Eddy, 1998), followed by a linear search of the reference library. The 100 best hits are selected as a pre-set of 'closely related tagged-specimens'; (ii) second, a NJ tree is reconstructed on this preset plus the query sequence to assess the relationship between the query sequence and its neighboring referenced sequences (Kelly et al., 2007). A certain identification at the species level is provided only if K2P distance between them do not exceed a fixed 1% threshold (Ratnasingham & Hebert, 2007) otherwise, the system returns the species name belonging to the nearest-neighboring reference sequence as shown in Figure 1.6. Although this method is direct and rapid as it can also be applied with large molecular datasets, it seems to suffer from a number of potential limitations (See the paragraph below; DeSalle et al., 2005; Meyer & Paulay, 2005; Frézal et al., 2008).

To date, various alternative approaches have been proposed to analyse DNA barcoding data (see Appendix Table II.1). Although they are based on a wide number of statistical and bioinformatic methods a provisional categorization, based on the characteristic of the molecular dataset to analyse, can be formalized as follow:

A) METHODS BASED ON THRESHOLD

This category encompasses those analytical methods that should be used when the user is dealing with large dataset with a reliable intra-specific sampling without having a clear idea of sequence variability levels. These methods are based on the analysis of similarity among barcode sequences compared to a reference dataset. In a strict sense these methods follow a typological species concept, and discriminate entities exceeding a certain level of variability called 'threshold value'. Threshold approaches rely on the assumption that intraspecific sequences variation does not exceed a certain distance value, otherwise they are considered as different species. In general, these methods perform DNA barcoding *sensu stricto* and are usually chosen because they are faster and require low knowledge on population structure or phylogenetic relationships. However, these methods imply the existence of a reference dataset, generated with the coordinated work of traditional and molecular taxonomists, to work. Hebert et al. (Hebert et al., 2004) firstly proposed the use of a divergence threshold following the '10-fold rule': the gap corresponds to a generic 10 times the value of intraspecific divergence. However, this rule has been deeply criticized (Moritz & Cicero, 2004; Meyer & Paulay, 2005; Ferri et al., 2009). Another example of fixed threshold is represented by the BOLD system approach (see above), which adopted 1% of K2P distance as universal threshold value for metazoans discriminations (Ratnasingham & Hebert, 2007; Figure 1.6).

Despite the initial success (e.g. Hajibabaei et al., 2006) pure distance-based methods could not be the most appropriate for species identification, because of several aspects clearly explained in (Ferguson, 2002). The main shortcomings rely on its strongly dependence from sampling accuracy, the lack of a strong biological support (that could cause high rates of false-positive assignments), the loss of character information and the absence of an objective criteria to delineate taxa when using distances (i.e. computing percentages of similarity or

genetic distances which are known to be sometimes irrelevant to infer taxonomic relationships; Ferguson, 2002).

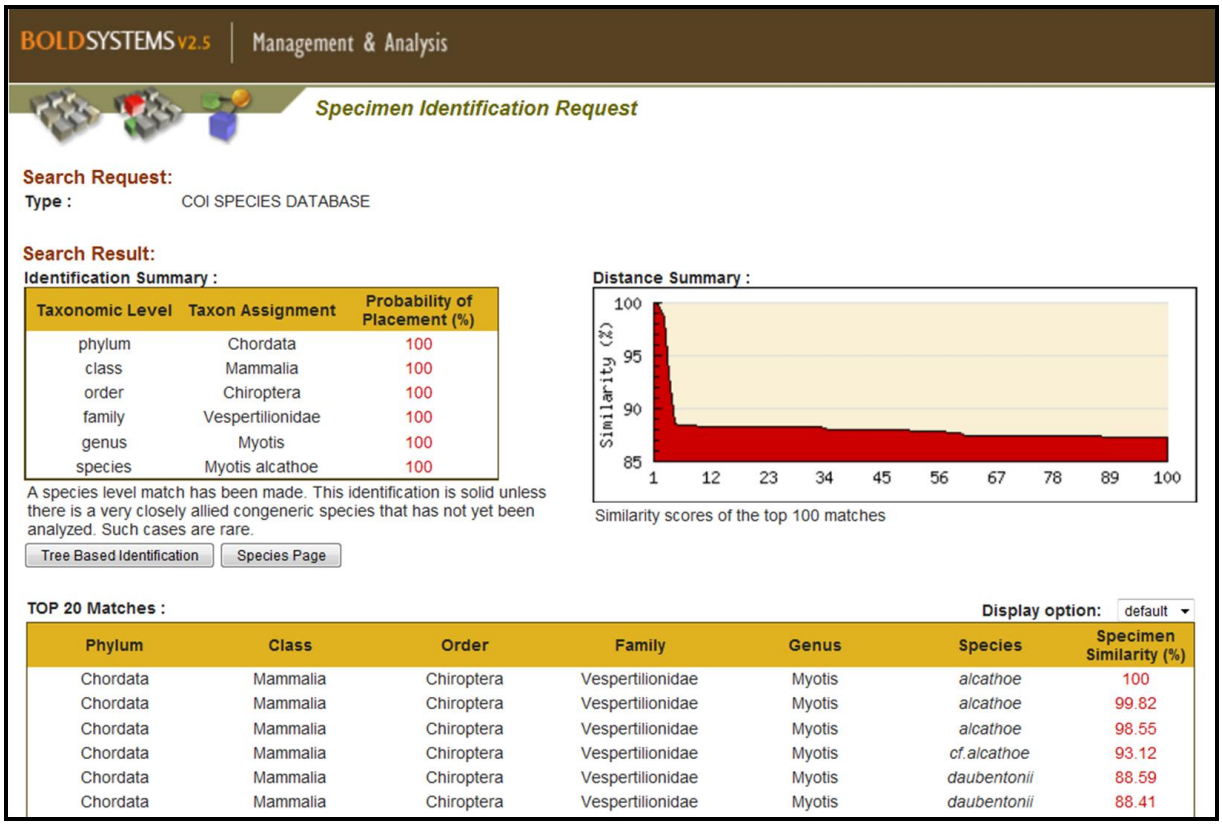


Figure 1.6: Typical output of a BOLD-IDS query. [image modified from http://www.boldsystems.org/views/-login.php]

Moreover, concerning tree reconstruction, Meier et al. (2008) already highlighted the limited performance of NJ, arguing that this method relies on the topology of one of all the possible NJ trees and does not consider the support of the nodes that separate and define species. Additionally, query sequences should be at least one node into monospecific clades, while simply clustering within a clade does not guarantee unambiguous identification. In order to avoid wrong or misleading interpretations of the results, some authors deliberately avoid to represent their data with a NJ reconstruction (e.g. Ferri et al., 2009; Virgilio et al., 2010).

Within the same group, there are similarity methods like BLAST (Altschul et al., 1990), BLAT (Kent, 2002) or FASTA (Pearson & Lipman, 1988) that were largely used to infer similarity between a query sequence and barcode reference sequences (DeLey et al., 2005; Vences et al., 2005) although they were also shown to cause incorrect or inconsistent identifications (Koski & Golding, 2001). In order to maximize the strength of coherence between DNA barcoding and other identification systems, Lefébure et al. (2006) and Ferri et al. (2009) formally tested the correlation between taxonomic ranks and genetic divergences toward the definition of a molecular threshold to help taxonomic decisions. In particular, Ferri et al. (2009) developed Perl scripts that allow to identify the optimal threshold value (OT) related to the minimum cumulative error (MCE, the minimum degree of discrepancies between two identification approaches, see Chapter 3.7) for rapid species diagnoses of filarial nematodes. This approach is the best choice in the context of studies which aim is to follow an integrated approach to taxonomy.

B) CLASSICAL PHYLOGENETIC APPROACHES

These methods follow a phylogenetic species concept and should be used with small datasets relative to groups that experienced different evolutive histories. Phylogenetic approaches have been developed and proposed for DNA barcoding data analysis in order to overcome the limits of threshold-based analyses. However, some confusion in the relationships between DNA barcoding and molecular phylogeny can emerge from their application (see also Moritz & Cicero, 2004; Vogler & Monaghan, 2006). In this context, it is important to underline that DNA barcoding is not, in a strict sense, a phylogenetic reconstruction because, as stated previously, to identify is different than to solve phylogenetic issues or to classify (Casiraghi et al., 2010). These methods could be tentatively organized in two subclasses: (i) pure phylogenetic methods and (ii) methods based on the coalescent theory. The formers are usually referred to NJ, maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) and are extensively used to identify query sequences by the reconstruction of a set of topologies. In order to include the query sequence in a specific group it is crucial to identify group membership variables in the reconstructed hierarchy. This procedure is highly subjective and consequently can cause disagreements in particular identification cases (Casiraghi et al., 2010). Moreover, other problems can occur

when a query sequence forms a new lineage or when the gene tree does not match previous classification caused by phenomena like incomplete lineage sorting (i.e. when the topology of the gene trees differs from that of the species tree; Nichols, 2001) or conservation of ancestral polymorphisms. It is known that in order to obtain robust phylogenetic reconstructions, the use of more taxa can perform better than the use of more genes (Huelsenbeck, 1995; Zwickl & Hillis, 2002; Pollock et al., 2002; Kolaczkowski & Thornton, 2004); however, in DNA barcoding the intrinsic need for large datasets causes high computational efforts that can be not easily supportable (Casiraghi et al., 2010; Virgilio et al., 2010). Given these considerations, it is clear that phylogenetic-based approaches should implement heuristics and simplified analytical methods to be used in this context.

Methods based on the coalescent theory rely on population genetics assumptions, and generally require a considerable collection of DNA sequences for each biological entity. Differently from threshold approaches, coalescent methods consider how the time since speciation influences patterns of genetic differentiation. The main principle is that between the speciation event and the observation of a reciprocal monophyly at a specific locus a time lag occurs and it can be used as a correction to perform ML infereces therefore providing a more realistic species modelling (Matz & Nielsen, 2005). In particular, Matz and Nielsen (2005) proposed one of the first efforts to introduce statistical formalisms in DNA barcoding data analysis. Their tree-based method takes into account phylogenetic uncertainty and uses population genetic theory to determine cut-offs for species assignment in ambiguous cases. A likelihood ratio test allows to evaluate possible boundaries of intra-specific variation (for each species) on the basis of reference datasets using population genetic inferences based on coalescent theory (the Markov Chain Monte Carlo method was implemented). Although the method has substantially improved since its first application in a DNA barcoding context (Nielsen & Matz, 2006; Abdo & Golding, 2007; Knowles & Carstens, 2007; Monaghan et al., 2009), it presents some main drawbacks such as the requirement of large intraspecific sampling (in order to extract population genetics data), intense computational times (as a direct consequence) and the inapplicability on a wide range of taxa (it requires a taxon-specific model of intra-specific variability).

C) CHARACTER-BASED APPROACHES

These approaches should be used with large dataset characterized by low intraspecific sampling. Differently from distance- and classical phylogenetic-based approaches, character-based methods rely only on diagnostic sites (i.e. presence/absence of discrete nucleotide substitutions), which being a small percentage of the total characters, make the application typically faster. These methods are considered consistent with the phylogenetic species concept (Goldstein & DeSalle, 2003) and can also handle other sources of data, such as morphological or ecological data therefore allowing to elude the distance 'nearest neighbour problem' through the reconstruction of hierarchical relationships (i.e. the common ancestor is inferred when two entities share derived characters). Some software tools have been developed to perform character-based analyses on DNA barcode datasets such as CAOS (Characteristic Attributes Organization System; Sarkar et al., 2002a; 2002b; 2008; DeSalle et al., 2005), Confind (Smagala et al., 2005), DNABAR (Dasgupta et al., 2005), DOME ID (Little & Stevenson, 2007).

D) METHODS INDEPENDENT FROM ALIGNMENT

Poor quality and the user influence on the alignment procedures could be responsible for the relatively poor performance of clustering hierarchical methods. In particular, NJ may be victim of low quality alignment. For these reasons, when large variable loci are present, it could be desirable to avoid methods that rely on alignments (e.g. Kuksa & Pavlovic, 2009). Since no universal alignment parameters are defined (hence gaps assignment into alignments is quite subjective, see Geiger, 2002) and there is no consensus on what defines a good or a best alignment, Chu et al. (2006; 2009) explored the feasibility of grouping taxa based on component vector (CV) analysis that does not require alignment.

1.10 TOWARD A MODERN APPROACH TO BIOLOGY: THE INTEGRATED TAXONOMY

As showed above, despite the growth of DNA barcode libraries, no consensus has yet emerged on the best method to analyze DNA barcode data (Casiraghi et al., 2010). The

situation is further complicated by the dual purposes of DNA barcoding: species identification and species discovery (DeSalle, 2006). The latter objective is the most controversial and concerns the definition of clusters of individuals that could be flagged as distinct taxa for further hypothesis testing (Rach et al., 2008). As a matter of fact, nowadays, species taxonomy is confronted with the challenge to incorporate new theories, methods and data from disciplines that study the origin, limits and evolution of species (e.g. morphology, biogeography, ecology) (Padial et al., 2010). These concepts have been fully acknowledged by the scientific community which has recently aimed for the adoption of an integrative approach to taxonomy (see for example: Dayrat, 2005; Will et al., 2005; Gibbs, 2009; Glaw et al., 2010; Schlick-Steiner et al., 2010). However, some disagreements emerged concerning mainly the degree of congruence that different characters must show to consider a group of individuals as a separate species. In this context, two putative kinds of integration could be formalized: i) 'integration by congruence' that occurs when the congruence between molecular variability and other characters (e.g. morphology) is considered as a necessary requisite (Dayrat, 2005; Meier, 2008; Cardoso et al., 2009) and ii) 'integration by cumulation' when *a priori* selection of character combinations is avoided in order to provide the strongest support to the integration (de Queiroz, 2007; Schlick-Steiner et al., 2010; Padial et al., 2010). A schematic representation of both integrative approaches is provided in Figures 1.7. and 1.8.

Congruence approaches have a long tradition in systematics (Padial et al., 2010); as an example, DeSalle et al. (2005), illustrated in a work diagram that congruence between two taxonomic characters is an important factor to reach a conclusion about species status (see Figure 1.9). Different combinations of taxonomic characters can be used to propose and support species (see Figure 1.8 *a*; Figure 1.9) such as the coherence between molecular and morphological characters (e.g. Ferri et al., 2009; Kerr et al., 2009), or even more restrictive combinations requiring evidence about reproductive isolation (Meiri & Mace, 2007; Alström et al., 2008). Integration by congruence promotes taxonomic stability because it relies on the assumption that a species is valid only if supported by several character sets.

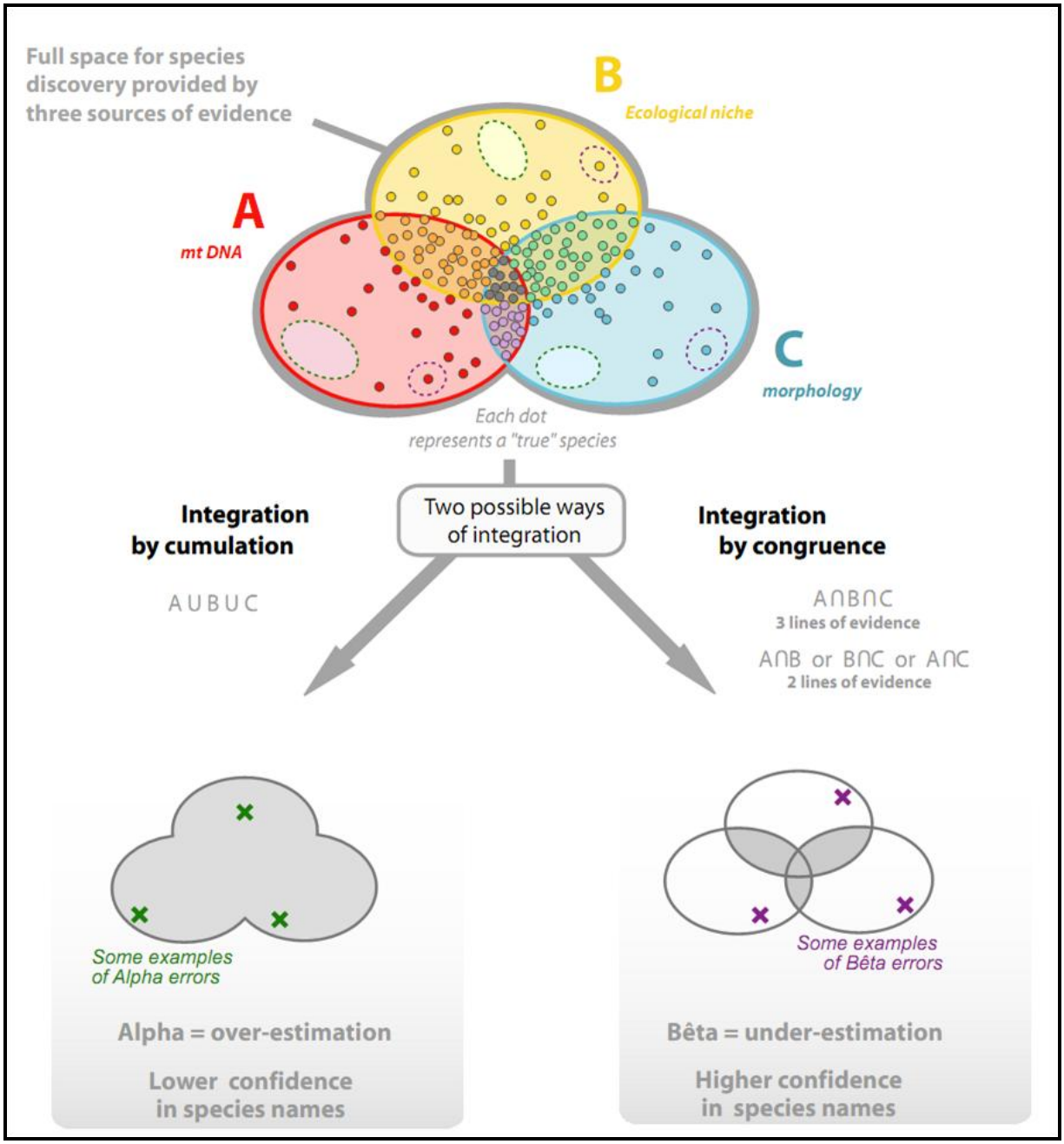


Figure 1.7: Schematic representation of two approaches of integrative taxonomy. Background yellow, red, and blue colors represent the spectrum of character variation, each dot being an independent evolutionary linage that requires identification and delimitation as separate species. Integration by cumulation (left) identifies species limits with divergence in one or more not necessarily overlapping taxonomic characters, whereas the integration by congruence (right) identifies species limits with the intersection of evidence from two or more independent taxonomic characters (e. g. mtDNA plus morphology). Both methods of integration have relevant limitations. The integration by cumulation approach may over-estimate the number of species by identifying distinct species where there is intraspecific character variation only. On the contrary, integration by congruence is a highly stringent approach that might under-estimate the number of species by being unable to detect cryptic or young species. [image retrieved from Padial et al., 2010]

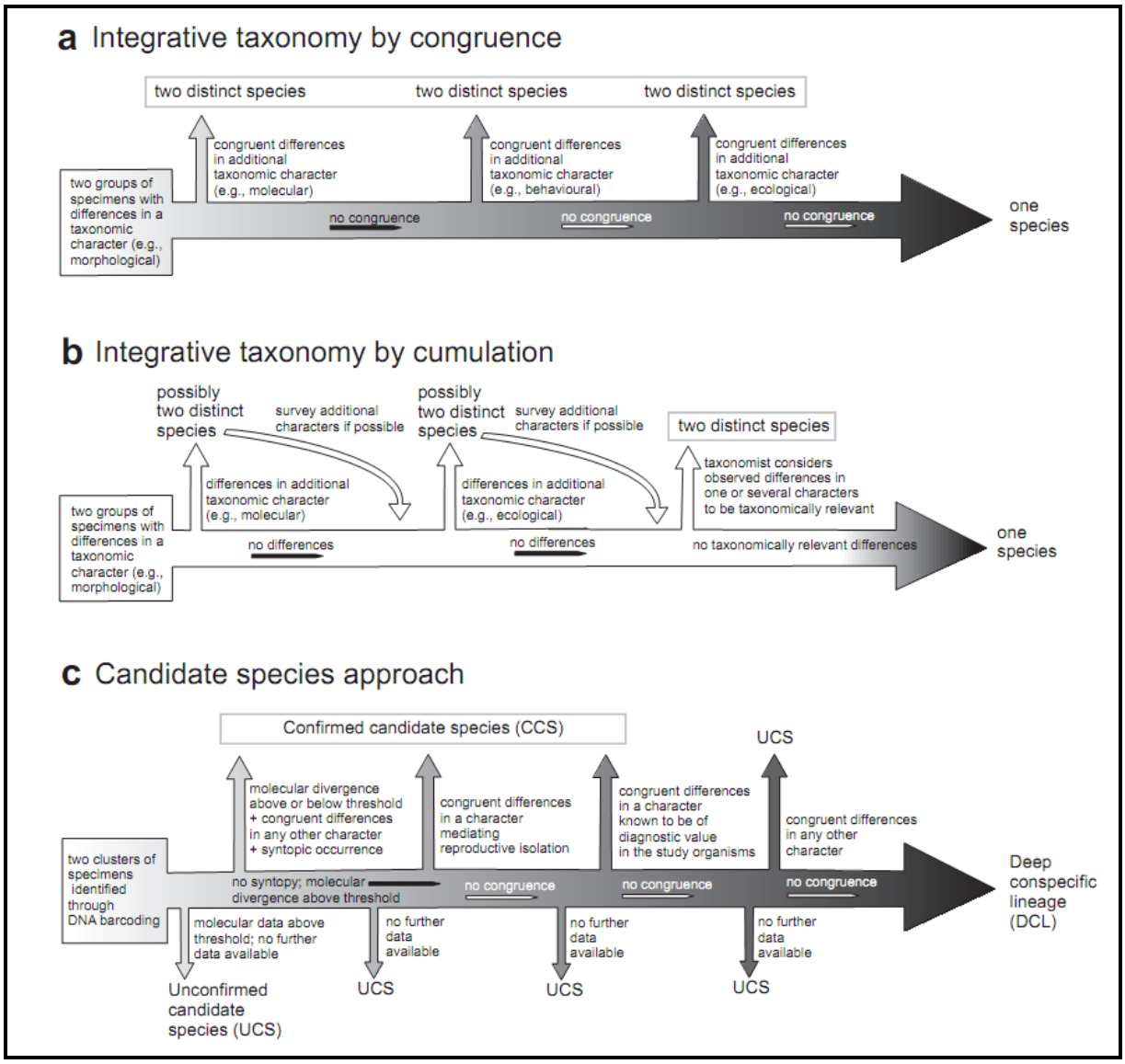


Figure 1.8: Schematic representation of work protocols in taxonomy. Workflow in (*a*) integrative taxonomy by congruence and (*b*) by cumulation; (*c*) work protocol to define Unconfirmed Candidate Species (UCS), Confirmed Candidate Species (CCS) and Deep Conspecific Lineages (DCL) in a DNA barcoding context. Increasing black color intensity in *a-c* represents increasing uncertainty about species status and the need of a more thorough evaluation of data. [modified from Padial et al., 2010]

However, some limitations of the method such as the risk of underestimating species numbers (e.g. Adams et al., 2009) and the heterogeneous rates of characters variability during lineage divergence should be taken into account. The lack of character congruence is indeed, more frequent than expected (e.g. in arthropods; Schlick-Steiner et al., 2010) resulting from the

different modes and circumstances of speciation (Shaffeer & Thomson, 2007; Degnan & Rosemberg, 2009).

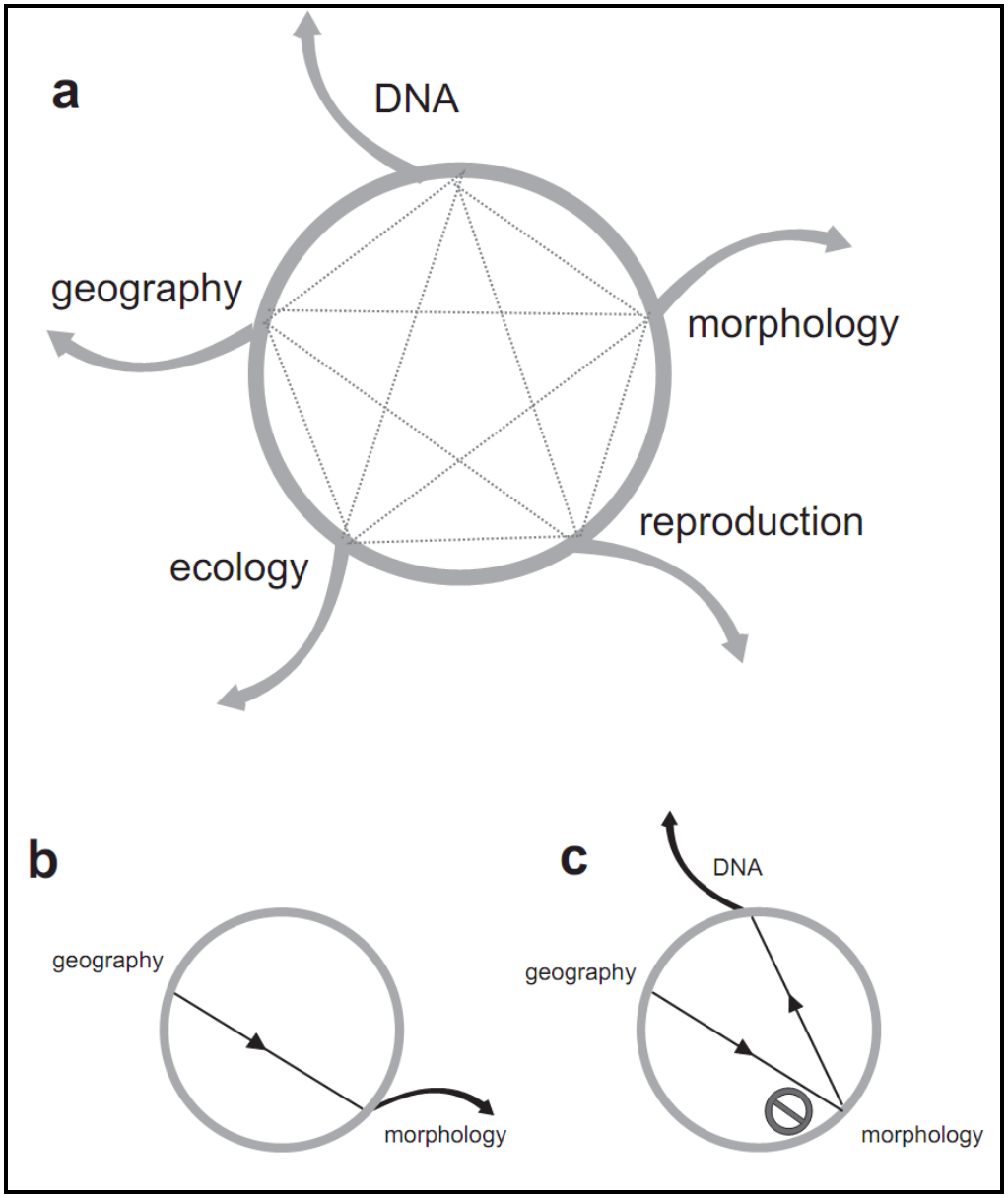


Figure 1.9: The "taxonomic circle" representation of work protocol for species recognition. This protocol (redrawn from DeSalle et al., 2005) is a schematic representation of the congruence approach to taxonomy. Dotted lines in (*a*) connect lines of evidence used to discover species or support previous hypotheses. The recognition of a species is considered when congruence between a taxonomic character and geography allows breaking out of the circle (black arrows). For example, in classical taxonomy (*b*) the occurrence of morphologically distinct specimens at different locations can be used to propose and support a species hypothesis. In the case of cryptic species (c), morphology fails to support the hypothesis but other characters (e.g. molecular) do provide support.

In "integration by cumulation" congruence of characters is desired but not necessary because the assumption is that whatever divergence in any of the organism attributes (i.e. the taxonomic characters) can provide evidence for the existence of a species (deQueiroz, 2007). Evidences from all character sets are therefore assembled cumulatively (Figure 1.8 *b*), and a taxonomic decision is made based on the available information, which can lead to recognition of a species on the basis of a single set of characters if these characters are considered good indicators of lineage divergence (Padial et al., 2010). The major advantages of this approach are that 'cumulation' does not force species delimitation to the identification of any particular biological property and it is most suitable choice when users are dealing with recently diverged species (Shaffer & Thomosn, 2007). On the contrary, the cumulative approach could be limited by the use of a single line of evidence (e.g. a single locus of mtDNA) that would lead to the overestimation of species numbers (e.g. the elevation of subspecies to the species rank; see for example Isaac et al., 2004).

Anyway, the recent adoption of integration processes has led to a synergy which has the advantage of minimizing the discrepancies between classical taxonomy and molecular approach (Casiraghi et al., 2010). In this context, in order to cope with the large number of molecular lineages and candidate species revealed by DNA barcoding studies of eukaryotes, Vieites et al. (2009) proposed a classification scheme for those units (see also Chapter 1.7) that will facilitate the subsequent formal description of new species under the Linnaean system. In particular, they recognize three subcategories of candidate species (Figure 1.8):

- UNCONFIRMED CANDIDATE SPECIES (UCS): Groups of individuals within nominal species showing large genetic distances, but further information are not available.

- DEEP CONSPECIFIC LINEAGES (DCL): Groups of individuals within nominal species showing large genetic distances, but additional data indicate that these units are not differentiated at the species level.

- CONFIRMED CANDIDATE SPECIES (CCS): Deep genealogical lineages that can be considered good species following standards of divergence for the group under study but that have not yet been formally described and named.

2. AIMS OF THE WORK AND CASE STUDIES

As extensively discussed in the previous sections it is clear how DNA barcoding has generated, in recent years, a wide impact in the biological studies becoming more than a simple molecular tool. Indeed, nowadays the widespread use of this method has led to an enormous acceleration in the identification of species (Frézal et al., 2008; Padial et al., 2010) therefore revolutionizing the classical vision of taxonomy and the processes to characterize biodiversity. On the other hand, despite its proved efficacy on a wide range of taxa and the existence of a number of free-available tools to manage barcode sequences (e.g. BOLD; see also Appendix Table II.1), a standard approach to analyze and to interpret DNA barcoding data is far to be formalized.

The aim of this research project was to test the efficacy of the DNA barcoding approach in identifying species belonging to different animal and plant taxa. In particular, the case studies here analysed allowed to cover some of the principal fields of application usually accomplished with this technique (i.e. classical taxonomy, medical and veterinary diagnostics, food traceability, assessment of environmental biodiversity). Each case has been analysed under the same experimental workflow which was developed to satisfy two general requirements: *i*) a general standardization in collection and processing of the samples following as better as possible the guidelines of CBOL (see also Material and Methods chapter) and *ii*) the adoption of technical procedures and analytical methods best fitting in the context of 'integrative taxonomy' (see chapter 1.10).

To cope with these requirements, samples have been collected by personnel experienced in the species or in the groups of taxa under examination (e.g. mammalogists, nematologists, ornithologists, botanists). Thanks to these professional figures it has been possible to obtain a preliminary assignment at the species rank for the most part of the samples collected basing mainly on disciplines belonging to classical taxonomy (e.g. morphology, ecology, geography, symptomatology). This approach also allowed to compare data inferred with classical methods and the molecular variability in the target barcode region(s) in order to produce a sort of 'reference dataset' for each studied group (see chapter 3.6). Bioinformatic analyses have been conducted according to the main assumptions of DNA barcoding and in particular

we decided to avoid those methods requiring high computational effort (e.g. methods based on intense phylogenetic or coalescence reconstructions) and accurate calibration of evolutionary models that are poorly suitable when dealing with a wide range of taxa. Otherwise, the procedures used to analyse most of the molecular dataset produced in this project were distance-based and/or based on threshold. Efficacy of these methods was evaluated for some datasets on the basis of the strength of coherence between the two identification approaches cited above and by calculating the amount of identification mismatches encountered during the analyses. Moreover, in attempt to propose a synergic synthesis between these two sources of identification characters we used for the first time an alternative approach to the identification of species in which this strength of coherence is used as a criterion to define the boundaries of entities here tentatively named IOTUs (Integrated Operational Taxonomic Units; see chapter 1.7). In some case studies, specific web tools (e.g. BOLD-IDS) have been used to identify species from their barcode sequence and identification results were compared to data obtained with our bioinformatic analyses (see Italian bats dataset as an example).

One or more case studies encompassing animal and/or plant taxa have been considered and developed for each one of the fields of application of DNA barcoding. Every case study relies on different problematic that have been tested with DNA barcoding in order to solve or at least clarify them. A synthetic list of the dataset treated in the context of this project is provided as follow.

2.1 CLASSICAL TAXONOMY

These case studies refer to animal groups which taxonomy has been subjected to several debates due to the occurrence of cryptic taxa (i.e. groups of species almost indistinguishable with classical approaches such as morphology), polymorphic populations, clinal variants and hybrid lineages. A DNA barcoding approach has been applied to shed more light on these situations.

A1) *Italian bats*

Mediterranean regions are renowned as areas of cryptic biodiversity (Hewitt, 1996, 2000; Taberlet et al., 1998; Jaarola & Searle, 2004; Pfenninger & Schwenk, 2007, García-Mudarra et al. 2009) and the occurrence of cryptic species have also been reported in the case of widely investigated taxa such as mammals and in particular bats. Indeed, several studies clearly confirmed that the order Chiroptera is characterized by a high incidence of overlooked or hardly distinguishable taxa (Jones, 1997; Clare et al. 2007). It is in fact almost impossible to recognize some species (especially in the case of females or juveniles) basing only on morphological characters or echolocation signals (Agnelli et al., 2006; Dietz & von Helversen, 2004). Moreover, diversity and biogeographical history of microchiropterans have been investigated mainly on those populations belonging to the Iberian Peninsula and other European regions, while little research has focused on the Italian peninsula (Ruedi et al., 2008) despite its biogeographical predisposition to harbour cryptic taxa.

Thirty-three microchiropteran species (belonging to four families) out of the almost 40 currently known to live in Europe are reported in Italy (Agnelli et al. 2006; Mayer et al. 2007). However, due to their cryptic morphology and habits, abundance, distribution and fragmentation of extant populations (Agnelli et al. 2006) are poorly known therefore negatively affecting the definition of appropriate conservation planning (e.g. in the case of the whiskered bats, genus *Myotis*; see Figure 2.1).

Figure 2.1: An example of cryptic species among Italian bats: the whiskered bats (Chiroptera: Vespertilionidae): *Myotis brandtii* (left), *M. mystacinus* (center) and *M. alcathoe* (right). [images modified from Dietz & Von Helversen, 2004]

In the last decade, thanks to the improvement of molecular techniques at least seven new cryptic species among vespertilionids (i.e. the most abundant microchiropteran family inhabiting Italy and Europe) have been molecularly-identified in Europe (Mayer et al., 2007, García-Mudarra et al. 2009). Four of these taxa have also been reported to occur in Italy (Agnelli et al., 2006). Moreover, other molecular studies demonstrated that despite their propensity to disperse, bats can display relatively high levels of intraspecific genetic differentiation and well structured phylogeographic patterns (Ruedi & Castella, 2003, Ibáñez et al., 2006, García-Mudarra et al., 2009).

In order to cope with this lack of information a DNA barcoding approach has been performed on Italian microchiropteran populations. In particular, the main objectives of our investigation were: *i*) to investigate the genetic differentiation of the sampled taxa testing for the coherence of a molecular approach with the classically used morphological-based taxonomy toward the definition of IOTUs; *ii*) to test the efficacy of DNA barcoding in identifying undetermined samples belonging to the most important cryptic groups and *iii*) to investigate the intraspecific molecular variability of the barcode region in order to look for geographic divergent lineages within widespread species that should be considered as separate units in conservation planning.

A2) *Parrotbills (Aves: Paradoxornithidae)*

The family Paradoxornithidae (Order Passeriformes) includes, among the other species, the Vinous-throated Parrotbill - *Paradoxornis webbianus* (Gould 1852) and the Ashy-throated Parrotbill - *P. alphonsianus* (Verreaux 1870). Like many other members of the family, these two taxa show several problems in their taxonomic identification (Robson, 2007): some authors claim that differences in plumage pattern and biometric parameters are not enough for considering *P. alphonsianus* a separate species (Han, 1991; Dickinson, 2003), while others, following morphological and geographical criteria, prefer to consider them as two separate species (Sibley and Monroe, 1990, Clements, 2007; Robson, 2007; Penhallurick and Robson, 2009). These taxa are morphologically similar, but differ in the colour pattern of the plumage of the head, throat and breast, which are of a reddish-brown in *P. webbianus* and greyish in *P. alphonsianus*, although these differences are not always clear (see Figure 2.2).

Figure 2.2: These pictures represent the two naturalized parrotbills inhabiting the Palude Brabbia Natural Reserve (Varese, Northern Italy). *Paradoxornis alphonsianus* (left) and *P. webbianus* (right). The two taxa differ for slightly differences in the colour of head, throath, and breast. [images by Andrea Galimberti]

Following Clements (2007), *P. webbianus* includes six subspecies distributed from southeastern People's Republic of China to extreme southeastern Russia, including Korea, northern Vietnam and Taiwan; whereas *P. alphonsianus* includes two subspecies. Moreover, after an accidental introduction (dated 1995) of about 150 individuals of both taxa (Baratelli et al., 2008), they can now be found in northwestern Italy (Varese province), but the provenance of the founders is so far unknown. Throughout the years the population have naturalized (Boto et al., 1999; Fracasso et al., in press), and its size reached thousands of individuals (Baratelli et al., 2008), colonizing the entire Palude Brabbia Natural Reserve and neighboring areas. In this locality, the two taxa occur in mixed flocks with more ashy-throated individuals (Boto A., pers. comm.) and mixed pair bonds have been observed.

Despite these taxonomic uncertainties, no molecular studies are so far available on *P. webbianus* and *P. alphonsianus* genetic differentiation. A DNA barcoding approach based on four different mitochondrial markers was then applied in order to: *i*) investigate the genetic differentiation of these two taxa to test for the coherence of a molecular approach with the morphologically-based taxonomy; *ii*) study the molecular differentiation between Chinese and the naturalized Italian populations, *iii*) trace the provenance of the founders released in the Palude Brabbia Natural Reserve (Varese, Italy).

A3) *Terrestrial isopods*

The order Isopoda includes thousands of crustacean species that inhabit a wide variety of habitats, both aquatic and terrestrial. In particular the suborder Oniscidea includes about 4.000 species of terrestrial isopods, worldwide distributed and well adapted to almost all terrestrial environments, ranging from deserts to alpine fields (Schmalfuss, 2004; Klossa-Kilia et al., 2006). However, the extant biodiversity of this group is likely to be really underestimated. Indeed, an amazing number of new species are described every year, carrying on with this increasing trend.

Despite the astounding morphological variability showed within some species, such as *Porcellio lamellatus* (Montesanto et al., 2007), a large number of oniscidean species are extremely similar leading to difficulties for correct species identification (Figure 2.3). Moreover, in many species, a detailed taxonomical identification requires careful dissections of animals (Sutton, 1972; Oliver & Meechan, 1993) and only researchers experienced in this taxon can have major possibilities to reach a correct species assessment. In other cases, the combination of different approaches to achieve an efficient discrimination of species is required.

Figure 2.3: Three typical oniscidean isopods reported in Italy and analysed in the context of this study. *Armadillidium vulgare* (left), *Armadillidium nasatum* (center) and *Trichoniscus pusillus* (right).

A well-documented case regards the Italian populations of *Porcellio imbutus* that seem to be represented by a complex of species, morphologically indiscernible. Preliminary evidences of genetic diversity evaluated by electrophoretic analysis of 15 gene loci in 26 *P. imbutus* populations from Sicily, Egadi and Maltese islands, revealed four clusters probably belonging to previously undescribed species (Viglianisi et al., 1992). The problem is still open and new

populations collected out of that clusters seems to extend the level of diversity observed (Montesanto pers. comm.). In the context of this research project, an integrated approach based on traditional taxonomy and DNA barcoding was used to shed more light on the diversity of some terrestrial isopods (suborder Oniscidea).

2.2 MEDICAL AND VETERINARY DIAGNOSTICS

This category of DNA barcoding applications deals with the identification of endoparasites (e.g. nematodes and cestodes) that are among the most important agents of disease both for human and other animals of economical value. However, it is hard to identify parasite species by traditional procedures due to different limits involved in the direct analysis of their morphology. For instance, a lot of endoparasite taxa are transmitted by hematophagous intermediate hosts during the first phases of their life-cycle. The identification of juvenile stages is useful to detect any possible emergent zoonotic disease at its beginning is but it is seriously biased by the small size of the organisms (about 1 mm) and paucity of characters. Molecular techniques such as DNA barcoding could help to overcome these problems for many reasons: *i*) a fast identification engine, available not only for taxonomists, but validated by them, would be useful for quicker diagnoses of filariasis and other pathologies caused by endoparasites; *ii*) DNA barcoding can be useful for the identification in cases of co-infections with more than one parasite species (e.g. *Onchocerca volvulus* and *Loa loa*; Ferri et al., 2009); *iii*) a molecular survey can offer a reliable method for the identification of endoparasites in vectors, allowing widespread campaigns of epidemiological surveys.

B1) *Filarioid nematodes*

This case study focuses on an integrated approach at the identification of a group of nematodes, belonging to the order Spirurida, which includes the relevant superfamily Filarioidea. Several species of filarioid nematodes are agents of tropical diseases both for human and other animals of economical value. All the filarioids are transmitted through haematophagous vectors in which they span different juvenile stages (Besansky et al., 2003; Powers, 2004). In vertebrate definitive hosts, they are found in the digestive tract or in other

different tissues, from the lymphatic to blood vessels and heart chambers, from abdominal and thoracic cavities to skin and subcutaneous tissues. The identification of juvenile stages is a necessary condition for establishing the potential of transmission in endemic areas but it is difficult as discussed above. In addition, laboratories typically deals with fragments of parasitic nematodes recovered from host tissues, or with specimens representing a single developmental stage, and the diagnostic characters are often not present in these pieces of worms. Despite molecular data from representatives of filarioids deposited in public databases are quite abundant for those species of medical or veterinary relevance (Figure 2.4), very few DNA barcodes are available if compared with other taxa of similar dimensions; this is mostly caused by the difficulties of sampling many species of parasitic nematodes.

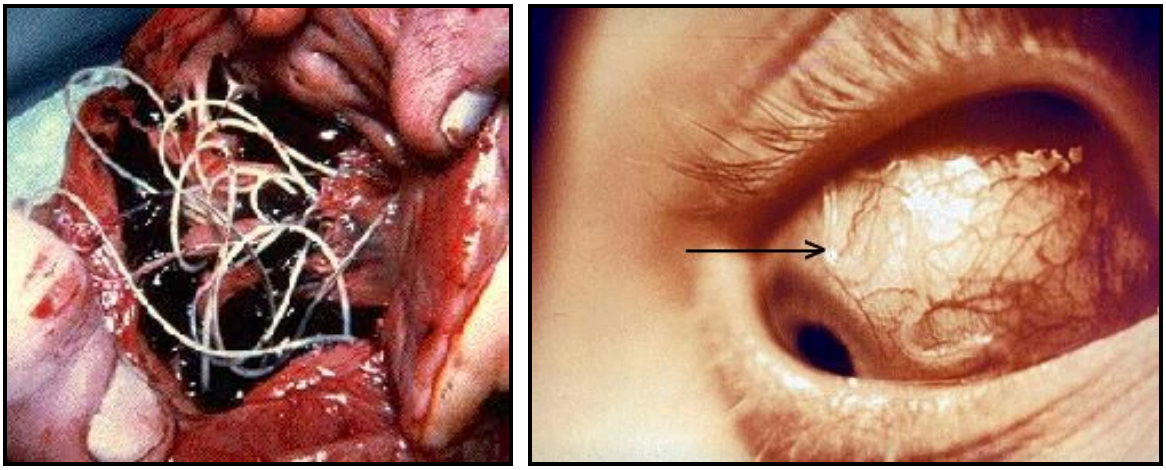


Figure 2.4: Examples of infection diseases caused by filarioid nematodes. On the left, a mass of *Dirofilaria immitis* is contained in the cardiac muscle of a dog; on the right, *Loa loa*: a nematode parasite causing filariasis in humans.

In this research a double approach (morphological and molecular) to the taxonomic identification of filarioids and related nematodes has been performed on the widest (in term of species number) molecular collection of these parasites ever achieved. Molecular distance estimation was performed with two different mitochondrial markers (*coxI* and 12S rDNA) and under different combinations of data handling (see chapter 3.7). In order to provide a useful tool for an easy identification of filarioid nematodes this work aims to answer the following questions: 1) which is the performance of DNA barcoding on filarioids and related nematodes? 2) which is the better marker (between *coxI* and 12S rDNA) for identification

these organisms at the species level? 3) can DNA barcoding be a useful tool for detection of putative new species?

B2) *Taeniids*

Traditionally, the taxonomical identification of cestodes has been based on both ecological (e.g. host specificity) and morphological criteria, but these approaches frequently led to unreliable results, especially when cryptic species complexes were considered (e.g. the case of *Taenia krabbei* and *T. ovis* reported by Priemer et al., 2002). In addition, the sampling of cestodes is often based on postmortem examination of the hosts, while less invasive approaches like the collection of eggs or proglottids in host feces cannot permit an easier identification due to the loss of many diagnostic tracts (Zhang et al., 2007). Besides these criticisms, one of the main aspects affecting the taxonomy of cestodes is the complex influence of their symbiotic relationship with hosts (Bouzid et al., 2008) and their facultative self-sufficient hermaphroditic reproduction.

In particular, a case of veterinary interest deserving to be further investigated concerns the taeniid panel typical of wildcat populations (*Felis silvestris* Schreber, 1777). The few studies available identified *Taenia taeniaeformis* as the most abundant cestode species harbored by European wildcat (Calvete, 1997; Krone et al., 2008; Figure 2.5) while it occurs less frequently in domestic cats (*F. s. catus*) populations belonging to the same study areas.

Figure 2.5: Adult stage of the taeniid *Taenia taeniaeformis*, one of the most common endoparasites harbored by wild and domestic cat populations (*Felis sylvestris* spp.)

Strong evidence for the presence of cryptic diversity within *T. taeniaeformis* were provided both morphologically and molecularly (Azuma et al., 1995; Okamoto et al., 1995; Lavikainen et al., 2008) but any attempt was performed to clearly synthesize these results toward a taxonomic reassessment of this taxon.

In this context, DNA barcoding could help clarifying this issue. In particular, a mitochondrial dataset of *coxI* sequences for several taeniid species of medical and veterinary interest was compiled in attempt to cover the following topics: *i*) to verify the efficacy of a DNA barcoding approach in the recognition of taeniid species, also testing for the coherence of a molecular approach with the morphology-based taxonomy; *ii*) to characterize taeniid spectrum in wildcat Italian populations and *iii*) to investigate the morphological and molecular variability of *T. taeniaeformis* .

2.3 FOOD TRACEABILITY

Some animal and plant taxa are of economic value as a food source (e.g. fishes, crustaceans, spices). Consumers are susceptible to any form of food alteration that may occur during the standard manufacturing processes and usually they pay attention to food ingredients in case of dietary nutritional requirements or medical conditions. Due to these reasons, the consumer is becoming more and more demanding on food quality, in particular for product traceability and for the use of detailed product labels. In the last years, media attention has repeatedly focused on the occurrence of species substitution that often implies a commercial fraud (e.g. in fishes: Marko et al., 2004; Smith et al., 2008; Wong & Hanner, 2008) because of the difficult in species recognition of products that are usually highly processed to meet consumer requirements (Blanco et al., 2008). It has also to be taken into account that substituted or mislabeled products offered in markets or restaurants may be potentially dangerous, due to the presence of unknown toxic or allergenic substances (e.g. poisonous plants) or hurtful in the case of commercial of endangered species (Ward et al., 2008; Wong & Hanner, 2008; Holmes et al., 2009). In this context, DNA barcoding can thus be a useful tool for food authentication and traceability.

C1) *Shark seafood products*

This survey has been focused on species substitution cases concerning the superorder Selachiomorpha (Chondrichthyes: Elasmobranchii), commonly named 'sharks', in the Italian fish trade. Due to the increasingly request of this group of fishes in food industry (Holmes et al., 2009), every month, in Italy, about fifteen tons of sharks are sold in one of the most important national fish market in Europe (i.e. Mercato Ittico di Milano-Milan, Italy - De Maddalena & Piscitelli, 2001). In particular, we studied sharks that are placed in the 'smooth-hound' complex (*Mustelus* spp., family Triakidae). Four shark species out of the 19 included in the Italian Regulation (G.U. No 45, February 22, 2008) belong to this genus but only *Mustelus mustelus* and *Mustelus asterias* can be labeled under the same vernacular name 'palombo'. As being the most requested by consumers, these two species are also the most subjected to commercial frauds (De Maddalena & Piscitelli, 2001). In fact, as showed in Figure 2.6, 'palombo' is typically sold as slices because of its length (usually between one and two meters) and after this kind of handling the organisms do not retain the diagnostic morphological details useful for the identification of the whole fish (Farrell et al., 2009).

Figure 2.6: Slices and fillets belonging to different shark species sold at the Milan fish market. Such a kind of processing treatment impedes an easy recognition of the species based on morphology. [Photo by Mario Lazzaro]

The objective of our approach was to verify the reliability of DNA barcoding in the recognition of commercialized shark species and to evaluate the amount of commercial frauds in the trading of shark slices labeled as 'palombo' in Italian markets.

C2) *Spices*

Some species belonging to the family Lamiaceae, due to the production of aromatic oils and secondary metabolites, are commonly used as spices for cooking (e.g. basil, marjoram, mint, etc.) therefore assuming a considerable commercial importance (Figure 2.7).

Figure 2.7: A typical herb belonging to the family Lamiaceae and used as a cooking spice: the Rosemary (*Rosmarinus officinalis* L.). Although the plant exhibits clear diagnostic characteristics that allows an easy recognition, the processed leaves (shredded or powdered) could be more difficult to identify from other spices.

Within the food market, internationally approved specifications provided by the American Spice Trade Association (ASTA - http://www.astaspice.org) and the European Spice Association (ESA - http://www.esa-spices.org/) define minimum quality thresholds for herbs and spices only by considering their phytochemical profile and the amount of their essential oil. However, herbs and spices tend to be grown by smallholders and it is difficult to monitor and control growers who might be inclined to use different herbs to increase the agricultural yield. Moreover, spices products, usually sold as enhancers for food flavor, could be accidentally or intentionally contaminated by other less-valuable plants species (Sasikumar et al., 2004; Dhanya & Sasikumar, 2010) with potential implications for human health.

The objective of this research was to evaluate the universal applicability of a DNA barcoding approach to reach a univocal identification of aromatic plant species starting from different plant portions and processed kitchen spices subjected to industrial modifications (e.g: drying, shredding, storage; Figure 2.7) that make them morphologically unrecognizable. Moreover,

amplification performances and identification success of four candidate barcoding loci (i.e. the plastidial *matK*, *rpoB*, *rbcL* and trnH-psbA) and different combinations of them were also investigated.

C3) *Poisonous plants*

Several spontaneous plants are potentially toxic for human beings (Vetter, 2000; Tinngi, 2003; Ngogang et al., 2008; Xia et al., 2008; Figure 2.8). Indeed, the accidental ingestion of toxic plant portions (for instance seed, fruit, root, etc.) can cause severe poisoning or even death (McIntire et al., 1990; Vetter, 2000).

Figure 2.8: Two examples of toxic plants causing a lot of accidental poisoning cases in humans: *Taxus baccata* L. (left) and *Colchicum speciosum* L. (right).

In recent years, the plant exposures are among the most frequent poisoning cases reported by poison control centres (Mrvos et al., 2001; Walker, 2004; Burrows & Tyrl, 2006; Gardner & Pfister, 2007). The clinical diagnosis of intoxicated patients is typically based on the morphological analysis of plant fragments in the stomach contents (Lawrence, 1997). This method is very tedious to perform, requires a considerable amount of training and usually a variable proportion of plant fragments remains unidentifiable. In addition, the plant species identifications can be difficult without residuals showing distinctive taxonomic elements.

In this case study, the main objective was to develop a rapid molecular tool (based on DNA barcoding) for toxic plants identification starting from a small portion of plant tissue from

different biological samples such as stomach contents and/or faeces of patients. This tool could be useful for poison centres to identify the ingested plants univocally and to define the suitable medical treatments rapidly. An evaluation of the performance of five candidate barcoding loci on selected groups of poisonous plant species has been also provided.

2.4 ENVIRONMENTAL BIODIVERSITY

An environmental sample is a mix of organic and inorganic materials collected from the environment (e.g. soil, water, stomach contents). It can contain live individuals (i.e. microorganisms and small macroorganisms), parts or nucleic acids remains of those species that are/or have been present around the sampling site. To date, environmental samples have been used mainly for studying microbial communities with different molecular markers such as 16S rDNA or nuclear ITS (Gomez-Alvarez et al., 2007; Herrera et al., 2007; Zinger et al., 2008). In recent years, thanks to the combination between DNA barcoding and next-generation technologies (e.g. pyrosequencing), environmental samples could also be a good source of information for characterizing macro-organic species (e.g. plants or animals) in order to assess biodiversity, presence/absence of a certain taxon in different localities or environments and even diet composition (see for example Valentini et al., 2009a; 2009b ; 2009c; Porazinska et al., 2009; Deagle et al., 2010; Riemann et al., 2010).

D1) *Meiofaunal moss biodiversity*

Although meiofaunal taxa are the most abundant in every ecosystem (soil, freshwater and sea), knowledge about the real number of the species involved is low (Blaxter et al., 2004). The main difficulty relies on the fact that a detailed morphological analysis is based on few microscopic characters therefore requiring considerable taxonomic expertise and resources (Bhadury et al., 2006). In this context, as extensively discussed in the introduction part, a DNA barcoding approach could help in overcoming this issue.

In this case study, such a molecular tool has been adopted in attempt to characterize as better as possible the meiofauna of moss: a well-studied soil matrix which biodiversity has been previously investigated (Ramazzotti, 1958; Zullini, 1970; Barbuto & Zullini, 2006). Moss

environments harbor a rich and well diversified community of organisms which is characterized mainly by protozoans, rotifers, tardigrades and nematodes as showed in Figure 2.9 (see also Zullini, 1970). In addition to these groups, several less abundant taxa that inhabit moss surfaces are springtails, annelids and eggs or larval stages of insects and other invertebrates (Gatti & Parisi, 2004). The zoological component chosen as target in this research was the hydrobiont fauna: a group of organisms occupying water interstices among soil particles.

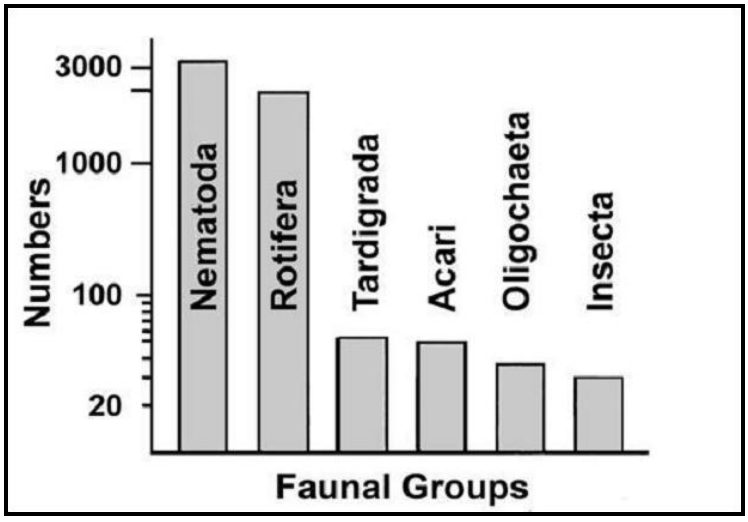


Figure 2.9: Comparison among the relative abundances (logarithmic scale) of the principal invertebrates inhabiting moss. [modified from Sayre & Brunson, 1971]

The study was performed in advance of a massive and fast DNA barcoding analysis on soil biodiversity based basing on 454 pyrosequencing in order to previoulsy: *i*) test the efficacy of universal primers used to amplify and sequence a fragment of the mitochondrial *coxI* for the principal groups of metazoans occurring in the environmental samples; *ii*) investigate the emergence of bias between the molecular characterization of biodiversity and the real number of taxa morphologically identified in the environmental sample.

3. MATERIALS AND METHODS

3.1 COLLECTION AND STORAGE OF BIOLOGICAL SAMPLES

The case studies here analysed encompass animal, plant and environmental biological samples belonging to organisms collected mainly in Italy, but also in the rest of the world. The most part of specimens have been recovered directly in the field (e.g. bats, birds, plants), while other samples belonged to institutional collections (e.g. museums, herbariums) or were found after necropsy or surgical intervention (e.g. filarioid nematodes and other of endoparasites). Sampled individuals were recognized at the species rank by researchers experienced in each single taxa of interest (e.g. mammalogists, nematologists, botanists). Due to the collector experience, sampling locality and/or host, we therefore assumed the correct identification of the samples. Several guides, dichotomous keys as well as primary literature reviewing or describing individual species were also used as a reference in order to reach the correct species assignment (e.g. Dietz & Von Helversen, 2004 for the identification of Italian bats; Robson, 2007 for the species and even subspecies identification of parrotbills, Pignatti, 1982 for plants identification). On the contrary, a correct assignment to the species rank for some samples was not possible due to the overlapping in morphological and ecological characters with congeneric cryptic species or in the case of processed material that lost every morphological diagnostic character (e.g. fish fillets, shredded leaves). In the context of those case studies concerning food traceability, these kinds of samples were named as ‘blind samples’. Capture and sampling of living animals (and plants) or parts of them were conducted by experienced personnel and authorized by appropriate permissions provided by local authorities. In some cases, animals were released soon after the collection of a small biological sample while in others, the entire organism was stored (e.g. in the case of insects or nematodes). Other samples come from different kinds of markets such as fisheries or supermarkets chains. In order to consider as well as possible the intraspecific geographical variation, samples from widely dispersed localities were preferentially collected for each species whenever possible.

Table 3.1 provides a synthetic view of the panel of samples collected in the context of this research project.

Table 3.1: Summarized list of biological samples collected for each category of application of DNA barcoding and related case studies (C.S. and C.S. Name) with reference to the total numbers of: samples (n samples), species identified by classical taxonomic approaches (n species), unknown/blind samples (n Unknown). The typologies of biological material collected, sampling source, storage conditions (E: ethanol; FM: frozen material; RT: room temperature) and reference to the corresponding detailed sampling table (Ref.) included in Appendix I are given for each case study.

CATEGORY	C.S.	C.S. NAME	n SAMPLES	n SPECIES	n UNKNOWN	MATERIAL	SOURCE	Storage	Ref.
CLASSICAL TAXONOMY	**A1**	italian bats	168	31	41	patagium	field, museums	E	I.1
	A2	parrotbills	35	2	3	feathers, muscolar tissue	field	E	I.2
	A3	terrestrial isopods	44	12	_	entire organism, internal organs	field	E	I.3
MEDICAL AND VETERINARY DIAGNOSTICS	**B1**	filarioid nematodes	76	28	13	entire organism, fragments	necropsy, museums	E	I.4
	B2	taeniids	62	1	_	proglottids	necropsy	E	I.5
FOOD TRACEABILITY	**C1**	shark seafood products	14	9	45	processed material	markets	E	I.6
	C2	spices	43	16	21	entire organism, processed material	herbariums, garden centers, markets	FM	I.7
	C3	poisonous plants	50	50	_	entire organism, fruits, berries, seeds	herbariums, garden centers, markets	FM	I.8
ENVIRONMENTAL BIODIVERSITY	**D1**	meiofaunal moss biodiversity	1	–	–	hydrobiont fauna (Baermann)	chestnut grove soil	RT	–

Concerning animals, typology and amount of tissue collected for successive genetic analyses varied depending on several factors such as dimension and conservation status of the individual (in the case of dead or museum stored specimens) and expected vulnerability to handling (in the case of taxa that had to be released soon after sampling; e.g. bats and birds). As a general condition, sampling was intended to be the less invasive as possible and biological samples from both museum and free-living individuals (e.g. feathers, *patagium* biopsies, hairs) were recovered following detailed guidelines provided in recent bibliography (e.g. Horvát et al., 2005; Trizio et al., 2005; Margam et al.; 2010; Martín-Gálvez et al., 2010). In order to have the possibility of performing alternative DNA extraction protocols (especially for oldest or degraded specimens), at least 25 mg of tissue were collected for each individual whenever possible. Obviously, in some situations it was not possible to satisfy this condition such as in the case of taxa characterized by reduced dimensions (e.g. nematodes, tardigrades). The collection of plant material was substantially easier due to the frequent high availability of fresh material. More than 100 mg of fresh tissues were collected for each living plant

while, in the case of specimens stored in herbarium collections, only 20 mg of tissue (usually leaves or roots) have been isolated. On the contrary, environmental or food samples were not subjected to particular risks deriving by an overexploitation of the original specimen; thus, a quantity of biological material largely exceeding one gram ore more (e.g. 200 g of soil were used for the characterization of environmental biodiversity, see chapter 2.4) was collected. In such a context it is important to underline that sampling coverage for some taxa was not exhaustive to describe completely the extant intra- and interspecific variability. Moreover, in some case studies (e.g. B1) sampling activities have been particularly difficult due to the emergence of complications associated with the collection step that required highly skilled personnel (not always available) and enduring logistic efforts all over the world.

The most part of animal samples was stored in plastic tubes containing 99% ethanol and then conserved at -20°C. Plant material was stored in plastic tubes without any solution and conserved at -20°C or at room temperature for fresh and dried samples respectively. However, some exceptions were encountered. As an example, in the case of organisms that required a morphological identification by microscopy (e.g. nematodes and other invertebrates), ethanol was substituted with another storage buffer (e.g. DMSO) to avoid morphological alteration of the sample. Environmental samples (e.g. moss) were conserved at 4°C because the isolation of hydrobiont fauna (the target of our DNA barcoding analysis) requires the presence of living individuals due to the limits of the extraction method (e.g. the 'Baermann' extraction protocol, see Figure 3.1; Hooper, 1990).

Once stored, samples were vouchered following the protocol specified by the Biorepositories initiative (http://www.biorepositories.org). The goal of this on-line registry is to create a system that permits records in nucleotide sequence databases (as well as other kinds of databases) to include links that point to the voucher specimens from which the DNA sequences were derived. In particular, each voucher name has to be composed by the combination of three parts: *i*) the universally-recognized acronym for the institution that holds the voucher specimen (in our case we used the acronym 'MIB' – University of Milano-Bicocca); *ii*) the institution's code to indicate the collection in which the voucher specimen is kept (in our case we used the code ZPL), and *iii*) the unique catalog number (or other identifier) in the catalog of specimens stored in that collection (e.g. MIB:ZPL:05248).

Plastic tubes containing vouchered samples have been also catalogued in a database called easytrack2d (http://www.easytrack2d.it/) on the base of a bidimensional code named '2D DataMatrix code' (a sort of barcode) stamped on the bottom of each tube. This software interface permits to track the 'life history' of each sample, helping the user in the management of biological collections and derived molecular data.

Additional sampling details (e.g. sampling localities, voucher numbers, haplotypes, etc.) for each dataset considered in this research are provided in Appendix I). Each table includes a list of samples and reference to the corresponding barcode sequence(s) produced in the context of this research project or retrieved from GenBank (these last were used as additional data in order to reach a better delineation of species boundaries; see the next chapters).

3.2 DNA EXTRACTIONS

In the most part of the cases, total genomic DNA was extracted starting from 5 - 25 mg of animal tissues or 20 - 100 mg of plant tissues. The quantity of biological material used for each extraction varied with age, conservation status and expected amount of extractable DNA of samples. For instance, in the case of animals, at least 25 mg of tissue (if available) were treated to extract DNA from museum specimens collected more than 20 years before or stored in low gradation ethanol or alkaline buffers. As a general assumption, in the case of dead conserved organisms, if any dimensional limit occurred (e.g. in the case of microscopic organisms such as free-living nematodes, rotifers, tardigrads), the biological sample to be used for DNA extraction was accurately chosen in order to avoid the loss of external diagnostic characters useful for morphological characterization (e.g. one leg in insects, a scale in fishes, some leaves or roots in plants; see also Table 3.1). Otherwise, in the case of small size taxa (e.g. juvenile stages of terrestrial isopods or filarioid nematodes), if more than one individual belonging to the same species and to the same geographical area were available, an entire organism was processed.

Preparation of samples consisted in one or more washes with cold (i.e. about 4°C) NaCl solution 0.9% in order to remove as well as possible ethanol or other buffers (e.g. DMSO) from tissues. Washing time, ranged from few minutes to overnight depending on the time elapsed since the sample was stored. Samples were then cut into small pieces or powderized after freezing with liquid nitrogen.

A different preparation procedure was performed in the case of environmental samples such as moss samples analysed in dataset D1 (see Table 3.1). Concerning this, a mixed sample of about 200 g of soil and its related moss surface was recovered in a chestnut grove in NW Italy (Lecco, 45° 51’ N, 09°24' E). Hydrobiont fauna was isolated according to a modified version of the Baermann method (Hooper, 1990): a bed of soil 3 cm high is laid on a piece of laboratory paper which is supported by a plastic net (mesh size: 0.5-1 cm); the structure is then transferred on a plate containing 300 ml of deionized water (see Figure 3.1).

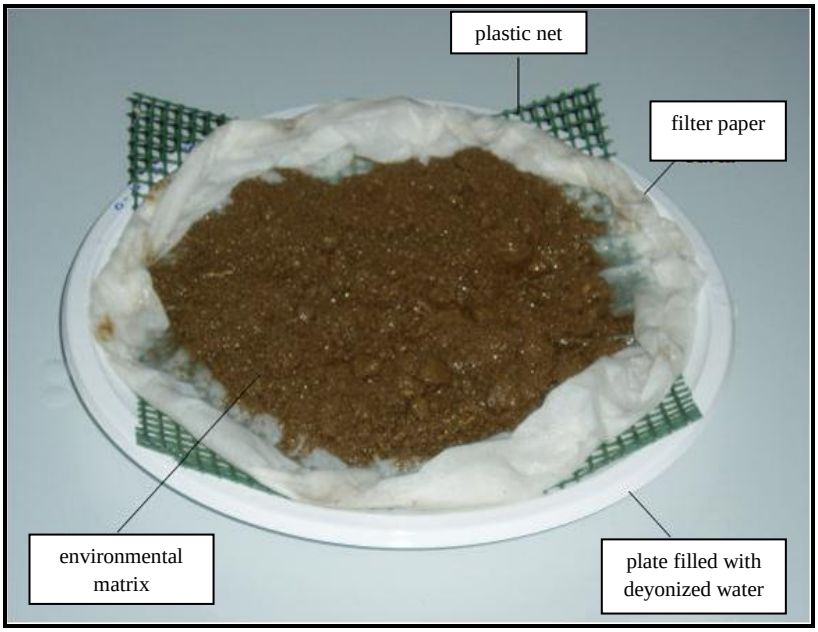


Figure 3.1: Baermann extraction (modified) system. Hooper et al., 1990

The system was incubated for 48 hours at room temperature allowing the hydrobiont organisms to move across the paper, the net and to precipitate in the water. The water was then transferred into a becker and progressively reduced to a final volume of 50 ml by using a vacuum pump. Following Seutin et al. (1991), the hydrobiont fauna was isolated from water with a 25 µm sieve and finally stored at room temperature in a NaCl saturated storage solution containing also EDTA 0.25M pH 7.5 and DMSO 20%. Once fixed, the sample was randomly divided in two parts. The first part was used to perform a preliminary morphological survey of the biodiversity of hydrobiont fauna in attempt to estimate the number of taxa (orders, families, genera) present in the extract..The second part was isolated to be processed with a DNA barcoding approach.

In particular, once the storage solution was replaced with deionized water, the total extract has been further reduced in volume, thoroughly resuspended and arbitrarily divided in 4 aliquots that were finally processed in order to extract DNA.

DNA extractions were performed with different commercial kits, following manufacturers' instructions. Animal and environmental samples were mainly processed using the 'ArchivePure DNA Purification Kit' (5 PRIME). Alternatively the 'DNeasy Blood & Tissue Kit' (QIAGEN) was used in case of difficult extractions (e.g. old samples, small organisms). Concerning plant samples, DNA was isolated using the 'DNeasy Isolation and Purification kit' (QIAGEN) in order to obtain high-quality DNA, free of polysaccharides or other metabolites that might interfere with successive amplification processes. Time of lysis and final eluition volumes were modified depending on quality and amount of the material to be extracted and varied from 30 minutes to overnight and from 20 µl to 200 µl respectively.

Purified DNA concentration of each sample was estimated both fluorometrically and by comparison of ethidium bromide-stained band intensities with λ DNA standard. Finally, DNA extracts were vouchered by adding a 'P' (which means 'Processed sample') to the original sample's voucher name (e.g. MIB:ZPL:05248P) and were stored in 2D Matrix tubes at -80°C.

3.3 DNA AMPLIFICATION AND SEQUENCING

For each selected sample, at least 10 ng of the extracted DNA were used as a template in a PCR reaction to amplify the selected barcode region(s). Details on the genomic region(s) chosen as barcode, primer pairs, and expected length of amplicons have been provided in Table 3.2.

As showed in table, *coxI* was the target barcode region for all of the case studies involving metazoans although in dataset B1 the mitochondrial 12S rDNA was also amplified in order to evaluate its discrimination power in the identification of filarioid nematodes (and also because of the moderate abundance of reference 12S rDNA sequences in GenBank for this group of parasites). *cox*I has also been the amplified target region in the case of dataset D1 (i.e. meiofaunal moss biodiversity) for which, all of the four DNA extracts containing thousands of individuals belonging to different microinvertebrate taxa (see previous section) were amplified with the universal Folmer primers (i.e. LCO1490 – HCO2198; Folmer et al., 1994).

Table 3.2: List of molecular barcode regions analysed for each case study with reference to primer pairs (F: forward; R: reverse) used in PCR amplification and expected size (in bp) of amplicons. Further details (e.g. primer sequences and PCR conditions) can be retrieved from Appendix Table II.2.

CASE STUDY	DATASET NAME	Barcode region	Primer pair (F - R)	Expected amplicon size (bp)
A1	ITALIAN BATS	*coxI*	VF1d - VR1d	658
			LCO1490 - HCO2198	658
A2	PARROTBILLS	*coxI*	BIRDF1 - BIRDR1	700
A3	TERRESTRIAL ISOPODS	*coxI*	LCO1490 - HCO2198	658
B1	FILARIOID NEMATODES	*coxI*	COIintF - COIintR	700
		12S rDNA	12SF - 12SR	500
B2	TAENIIDS	*coxI*	JB3 - JB4.5	450
C1	SHARK SEAFOOD PRODUCTS	*coxI*	Shark_int - FishR2	616
C2	SPICES	*matK*	matK 390 - matK 1326	930
		trnH-psbA	psbA - trnH	700
		rpoB	rpoB 1F - rpoB 4R	470
		rbcL	rbcL 1F - rbcL 724R	720
C3	POISONOUS PLANTS	*matK*	matK 390 - matK 1326	930
		trnH-psbA	psbA - trnH	700
		rpoB	rpoB 1F - rpoB 4R	470
		Sqd1	sqd1F - sqd1R	270
		At103	At103F - At103R	300-430
D1	MEIOFAUNAL MOSS BIODIVERSITY	*coxI*	LCO1490 - HCO2198	658

On the contrary, several barcode regions have been tested in plant datasets (i.e. C2 and C3; see Table 3.2 and Appendix Table II.2) in order to assess their comparative performances. These included three coding (*rbcL*, *rpoB* and *matK*) and one non-coding (trnH-psbA intergenic spacer) plastidial DNA regions. Two additional nuclear DNA regions were analysed in dataset C3 (poisonous plants) in order to also evaluate the performance of nuclear genes. These unconventional regions are the *sqd1* (UDP sulfoquinovose synthase) and *At103* (Mg-protoporphyrin IX monomethyl ester cyclase) previously investigated by Li et al. (2008) as universally amplifiable markers for plants phylogenetic reconstructions. Amplicon lengths varied both in animal and plant taxa depending on the position of the specific region of annealing for each primer or due to the peculiar structure of some genes (e.g. 16S rDNA or the intergenic spacer trnH.-psbA).

Supplementary molecular analyses have been performed for almost all the samples included in datasets A2 (parrotbills) and for those individuals morphologically assigned to the

taxon *Myotis nattereri* included in the dataset A1 (Italian bats). In particular, three additional mitochondrial genes (i.e. c*yt b*, 12S rDNA and 16S rDNA; see Table 3.3 and Appendix Table II.2) were amplified and sequenced for parrotbills samples (see Appendix I.2) in order to study the molecular differentiation between Chinese and the naturalized Italian populations in a phylogeographic context. The genes listed above are the most used in this kind of analyses and *cyt b* and 16S rDNA have been largely used as barcode markers for the definition of species and/or subspecies boundaries in birds (Brambilla et al., 2008; Aliabadian et al., 2009; Song et al., 2009). Mitochondrial *cyt b* and *ND1* genes were also examined in *M. nattereri* samples (see Table 3.3.; Appendix Table II.2 and Appendix Table II.3) in order to have a comparison of samples collected in Northern, Central and Southern Italy with putative corresponding lineages (including the recently described taxon *M. escalerai*) sampled elsewhere in Western Palearctic.

Table 3.3: List of additional genes analysed for case studies A1 and A2 with reference to primer pairs (F: forward; R: reverse) used in PCR amplification and expected size (in bp) of amplicons. Further details (e.g. primer sequences and PCR conditions) can be retrieved from Appendix Table II.2.

CASE STUDY	DATASET NAME	Gene	Primer pair (F - R)	Expected Amplicon size (bp)
A1	ITALIAN BATS	*ND1*	ER65 - ER66	1000
		cyt b	Molcit_F - MVZ-16	800
A2	PARROTBILLS	*cyt b*	L14990 - H15916	930
		12S rDNA	L1549 - H1991	420
		16S rDNA	L3214 - H3783	550

For those datasets encompassing animal samples (i.e.: A1-A3; B1-B2; C1; D1) PCRs were performed in a volume of 20 µl under the following final conditions: 1X buffer including 1.5 – 2.5 mM MgCl2 (MasterTaq kit, Eppendorf^{TM}), 0.2 mM of each dNTP, 1 µM of each forward and reverse primers and 1U of DNA polymerase (MasterTaq kit, Eppendorf^{TM}). Barcode regions selected to test plant samples (datasets C2 and C3) were amplified in a volume of 25 µl using puReTaq Ready-To-Go PCR beads (Amersham Bioscience, Italy) according to the manufacturer's instructions. Primer sequences and thermal conditions used for each amplification are provided in Appendix Table II.2.

Concerning plant datasets, the criterion used to assess universality of the candidate barcode markers (see Table 3.2) involved the establishment of which regions could be

routinely amplified and sequenced in the maximum number of analysed plants. Only the most universal primer combinations were tested for each candidate DNA marker in order to facilitate the interpretation of successes and failures. For all taxa and loci, we conducted PCR amplification in a two stage trial. In the first stage, we used the standard PCR conditions described above, starting from 10 ng of DNA template. The second stage included only samples that did not amplify or that produced multiple PCR products (e.g. non-specific amplicons). These problematic samples were then reamplified using 1 and 25 ng of DNA template. The samples that failed to amplify were repeated at lower stringency (with a reduction of 5°C in the annealing temperature as described in Appendix Table II.2) and 40 PCR cycles. Only in case of negative amplification in all conditions the PCR reaction was considered a failure. Similarly, when DNA extracts belonging to animals were processed, slight modifications of annealing temperature, dosage of DNA and/or $MgCl_2$ in the reaction mix were performed in order to overcome amplification failures.

Amplicons size was assessed by electrophoresis in 1.5% agarose gels stained with ethidium bromide. PCRs products were gel purified (using the Perfectprep Gel Cleanup, Eppendorf^{TM}) and the heavy DNA strands were bidirectionally sequenced using an ABI 155 3730XL automated sequencer at Macrogen Inc., Korea. Manual editing of raw traces and subsequent alignments of forward and reverse sequences were conducted with Bioedit sequence alignment editor (version 7.0.5; Hall, 1999). Moreover, the 3′ and 5′ terminals were clipped to generate consensus sequences for each sample. After checked for the presence of pseudogenes and numts (i.e. nuclear mitochondrial pseudogenes; see next chapter), most of the sequences obtained have been deposited in the EMBL Data Library under the accession numbers listed in Appendix I tables.

3.4 SEARCH FOR NUMTs AND PSEUDOGENES INTERFERENCE

Nuclear mitochondrial pseudogenes, or ‘numts’, result from genetic material being translocated from the mitochondrial to the nuclear genome (Bensasson et al. 2001). Moreover, because they are nonfunctional, they may also accumulate insertions, deletions, and other nonsensical mutations. As reported in several recent studies, the inadvertent use of *coxI*–like sequences like numts in a DNA barcoding approach could affect the analyses leading to wrong identifications of undescribed molecular lineages or cryptic species (Song et al., 2008;

Buhay, 2009). The presence of numts among eukaryotes has been previously reported for different animal taxa (Zhang & Hewitt, 1996; Bensasson et al., 2001; Richly & Leister 2004; Kerr, 2010) and plants (see for example Naciri & Manen, 2009). Following the guidelines proposed by Song and colleagues (2008) and Buhay (2009) and according to the standard DNA barcoding procedures, quality controls of the sequences produced is requested (see Ratnasingham & Hebert, 2007). The barcode sequences produced in this study were therefore checked for the presence of numts. In particular each sequence was analyzed with different methods: NCBI Blast Search, quality of trace files, *in silico* translation (to check for indels and stop codons), comparison with published mitochondrial complete genomes and compositional biases. Those barcode sequences that did not pass the control procedures described above were removed from the dataset under analysis and a new PCR with different primers or conditions was performed in order to avoid the occurrence of pseudogene amplification.

3.5 CLONING ANALYSIS

Concerning the case study related to the characterization of environmental biodiversity (i.e. D1), a cloning analysis was performed in order to isolate as well as possible each *coxI* amplicon obtained from the total amplification of DNA extracted from different pools of hydrobiont fauna (i.e. four aliquots of the total Baermann's extraction). Quantification of DNA for the four respective PCR products was carried out using the QubitTM fluorimeter (Invitrogen^{TM}). Taking into account these concentration values, PCR products were ligated into the pGEM®-T Easy plasmid vector (Promega, Madison, Wisconsin) using T4 DNA Ligase and incubated overnight at 4°C. Recombinant plasmids were transformed into competent *E. coli* DH5a cells and, after overnight growth on LB medium plates at 37°C, colonies containing plasmids with inserts were screened by X-gal-mediated blue/white selection. A maximum of 96 clones for each one of the four extraction aliquots were picked up with a pipette tip and put into tubes containing 40 µl of sterile water. After denaturation at 95°C for 10 minutes, 1 µl of this solution was amplified for each clone by PCR using LCO1490 - HCO2198 primer pair (see Appendix Table II.2). It is important to underline that each one of the four Baermann's extractions aliquots was amplified and cloned twice, so we could potentially obtain 192 clones per aliquot (see also Figure 4.1). DNA purification and sequencing were then performed as described previously (see chapter 3.3).

3.6 ALIGNMENTS AND DATASETS DEFINITION

Barcode sequences for each molecular dataset were unambiguously aligned using Clustal X (Thompson et al., 1997). However, other alignment platforms were alternatively tested such as MAFFT (v 6.240 - http://mafft.cbrc.jp/alignment/software/) or MUSCLE (http://www.- ebi.ac.uk/Tools/msa /muscle/; Edgar, 2004). These online tools are particularly useful when the user has to deal with large datasets (e.g. dataset D1). In case study B1 (i.e. filarioid nematodes) homologous DNA barcode sequences of *coxI* and 12S rDNA belonging to the same individuals were aligned with MUSCLE – default options – and ClustalX – default options. This approach was performed in order to evaluate how the identification performances of DNA barcoding in this group of parasites varied with different molecular markers (i.e. *coxI* and 12S rDNA) and data handling such as alignment algorithm or gap treatment (see chapter 4.5.2). Due to differences in sequencing efficacy, the alignments were cut to yield the same length (in terms of bp) for all the barcode sequences included in each dataset (see Table 4.2).

As widely discussed in the introduction part, each case study was developed around a specific aim(s) (e.g. food traceability, identification of cryptic species, etc.). Thus, in order to meet the requirements of the bioinformatic analyses to be performed, barcode sequences were partitioned in one or more molecular sub-datasets. In particular, the DNA sequences used in these case studies were obtained by direct sequencing of PCR products or collected from GenBank; only sequences meeting a priory defined criteria of length, position, similarity and taxonomy were analysed. The high quality of the taxonomic details was mandatory, and allowed us to classify each barcode sequence into one of the four following groups: (1) sequences originated from organisms morphologically identified by international experts of our group; (2) sequences collected from GenBank and morphologically identified by international experts not affiliated to our group; (3) sequences originated from organisms collected by our group and morphologically undetermined; (4) sequences collected from GenBank and whose identification process is not certainly based on morphology.

Most of the case studies here presented were focused on the use of DNA barcoding to delimitate species boundaries. In particular, a sort of molecular divergence cutoff was calculated in order to identify species and also flag the presence of divergent lineages (i.e. case studies A1, A3, B1, B2, C1; see chapter 3.7).

These case studies were also characterized by a suitable sampling coverage at both intra- and interspecific level and almost all the samples collected have been identified with other approaches (e.g. morphology) allowing an efficient comparison between taxonomical identification and molecular variability in the barcode region(s). Barcode sequences belonging to datasets A1, A3, B1, B2 and C1 have been therefore subdivided in two molecular sub-datasets named as: 'reference dataset' and 'comprehensive dataset'. The former typology of molecular dataset encompasses only sequences deriving from specimens relying on a certain taxonomical identification (categories 1 and 2, see above). Reference datasets were used to test the coherence between classical taxonomy and molecular approaches following a 'typical' DNA barcoding analysis: generation of a K2P distance graph and cumulative error plots (see for example Figure 3.3). Comprehensive datasets encompass all the barcode sequences available (cases 1, 2, 3 and 4; see above) and were used for standard DNA barcoding analyses in attempt to assign those sequences deriving from morphologically undetermined organisms to a certain species (for example the blind samples of case studies A1 and C1) or alternatively to investigate the occurrence of previously undescribed taxa. Sampling tables provided in Appendix I show a list of samples used to generate barcode sequences and sequences retrieved from GenBank highlighting their 'role' as part of reference or comprehensive datasets.

Although both morphological and molecular data were available for the sampled taxa *Myotis nattereri* (case study A1: Italian bats), *T. taeniaeformis* and *T. polyacantha* (case study B2: taeniids), their *coxI* barcode sequences were removed from the respective reference datasets due to their uncertain taxonomic status as revealed by recent studies (see for example: García-Mudarra et al., 2009; Lavikainen et al., 2008). However, due to the astounding levels of intraspecific molecular variability observed for these taxa, they have been subsequently analysed in the next steps of our research (see chapter 3.9). Moreover, in the context of the case study B1 (i.e. filarioid nematodes), two further homologous sub-datasets were assembled (relative to *coxI* and 12S rDNA barcode sequences respectively). They encompass sequences derived from organisms belonging to cases 1 and 2 (see above) for which sequences of both genes were available (see Appendix Table I.4). These two types of datasets were assembled to compare the molecular identification performances of the two barcode markers under different kinds of data handling (see chapter 4.5.2).

Concerning plant datasets (case studies C2 and C3), barcode sequences were partitioned in smaller sub-datasets according to taxonomy (in the case of C2; see Appendix Table I.7) or toxicity (in the case of C3; see Appendix Table I.8) of the sampled taxa. In particular, in the case of poisonous plants (i.e. C3), barcode sequences belonging to different plant samples and showing different degrees of toxicity for humans have been grouped as follow:

- Group I: *Ornamental plants as well as spontaneous plants containing different toxic substances.* Several of these species show attractive poisonous portions such as fruits or flowers, and these are accidentally eaten, mainly by children. These samples were selected, because they represent a diverse set of species across the angiosperms (including monocots and eudicots) with various levels of phylogenetic distance. This provides a sound assessment of universality of the tested DNA barcode markers.

- Group II: *Phylogenetically related plants showing different poisonous portions and degrees of toxicity*. These taxa were selected in order to evaluate the discriminating power of different DNA barcode markers among closely related taxa. In this group, species belonging to the genera *Aconitum* (IIa) and *Sambucus* (IIb) were included (see Appendix Table I.8).

- Group III: *Congeneric edible and poisonous plant species*. In the clinical diagnosis of intoxicated patients, it is very important to distinguish fragments of edible plants from poisonous species. For this reason, some species of *Solanum* (group IIIa) and *Prunus* (group IIIb), two large angiosperm genera constituted by edible and poisonous plants, were analysed with different DNA barcode markers. The final goal was to test the power of DNA barcoding to distinguish common edible fruits from fragments of toxic plants accidentally ingested.

Finally, concerning dataset D1 (meiofaunal moss biodiversity), all the barcode sequences obtained after cloning were initially grouped in the same alignment because any morphological detail was available to associate each sequence to the corresponding organism. In attempt to provide a first taxonomical identification of the barcode sequences, each barcode was tentatively assigned to the belonging phylum by similarity match with GenBank sequences using BLAST algorithm (Altschul et al. 1990). In particular, sequences with a match of at least 90% identity were assigned to the recognized phylum (e.g. arthropods, nematodes, rotifers), while the remaining were tagged as “non-identified”. In the next step of

the analysis, these barcode sequences were partitioned in three smaller sub-datasets referred to nematodes, rotifers and tardigrades respectively.

3.7 ALIGNMENT CHARACTERISTICS AND IDENTIFICATION OF MOLECULAR ENTITIES

In order to evaluate the efficacy of DNA barcoding as a molecular tool for the identification of the different taxa, the strength of coherence between a classical morphologically-based identification (when available) and the barcode(s) genetic divergence for each species was measured. Intraspecific, interspecific, overall mean K2P distances and relative standard errors were calculated with MEGA 4.1 (Kumar et al., 2008) – options = Kimura 2-parameters, pairwise deletion. In the case of filarioid nematodes (case study B1), K2P distances were calculated with two different applications, and the gaps were treated in two different ways: MEGA [29] – options = Kimura 2-parameters, both pairwise deletion and complete deletion were set in separate runs – and TREECON (Van de Peer & Wachter, 1994) – options = Kimura 2-parameters, both 'not take into account' and 'take into account' were set in separate runs. In all of the other cases, gaps were treated as pairwise deletion as suggested by CBOL standards.

DNA barcoding *sensu stricto* (see chapter 1.6) was performed with two different approaches. They have been differentially adopted depending on dataset characteristics (e.g. availability or lack of a morphological identification) and research objectives. In both cases, the use of more computationally intensive methods was avoided in favour of programs that could be executed in real time (e.g. MEGA).

The first approach was based on Neighbour-Joining (NJ) clustering and has been used to group barcode sequences in MOTUs when any additional taxonomic detail about samples was available (e.g. in case study D1) or when sampling coverage was insufficient to represent the extant intra and interspecific variability (e.g. in case studies B1, C2, C3). NJ trees using the (K2P) model were constructed with MEGA 4.1 (Kumar et al., 2008). More sophisticated tree-building methods exist, but since we are concerned about terminal branches, not deeper branching patterns, this method is sufficient (Kerr et al., 2009). Support for monophyletic clusters was determined using 500 bootstrap replicates. MOTUs were identified as they

comprised the smallest diagnosable cluster of barcode sequences with greater than 95% bootstrap support (Felsenstein, 1985; Figure 3.2).

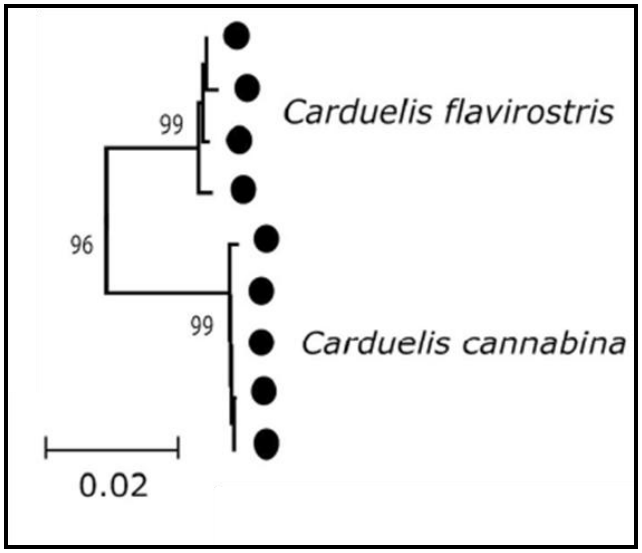


Figure 3.2: Example of divergence pattern illustrated in a putative NJ tree. The two bird species investigated with a DNA barcoding approach based on *coxI* are monophlyletic with > 95% bootstrap support in the NJ reconstruction therefore forming two distinct clusters. [Image modified from Kerr et al., 2009]

The second approach was based on the use of distance thresholds. In particular, in attempt to work with an integrated approach to taxonomy, an 'Optimum Threshold' (OT) of molecular divergence was calculated for each one of the selected datasets and allowed to trace the boundaries of IOTUs and MOTUs (see chapter 1.7). OT is a value of molecular divergence, directly deriving from the whole range of molecular variability in the dataset (see Figure 3.3, *a*). OT maximizes the coherence between the morphologically-based identification and the molecular variability in the barcode region minimizing, at the same time, the total amount of identification mismatches that could occur when data obtained with the two approaches are compared. Identification mismatches could include Type I errors (i.e.: false positives: when molecular variability values higher than OT are found among conspecific individuals) and Type II errors (i.e. false negatives: when congeneric species identified by morphology, show values of molecular variability lower than OT). The lower is OT, the higher is the probability to deal with Type I errors while high values of OT generally correspond to an high percentage of Type II errors (see Wiemers et al., 2007; Figure 3.3, *b*). As stated in Ferri et al. (2009) the sum of both false positive and false negative cases represents the so-called 'cumulative error' (CE), and when the minimum cumulative error value (MCE) is reached the Optimum Threshold is found (see Fig. 3.3, *b*).

Optimum threshold values were calculated on reference datasets of case studies A1, A3, B1, B2 and C1 by using a PERL script developed by Ferri and colleagues (Ferri et al., 2009).

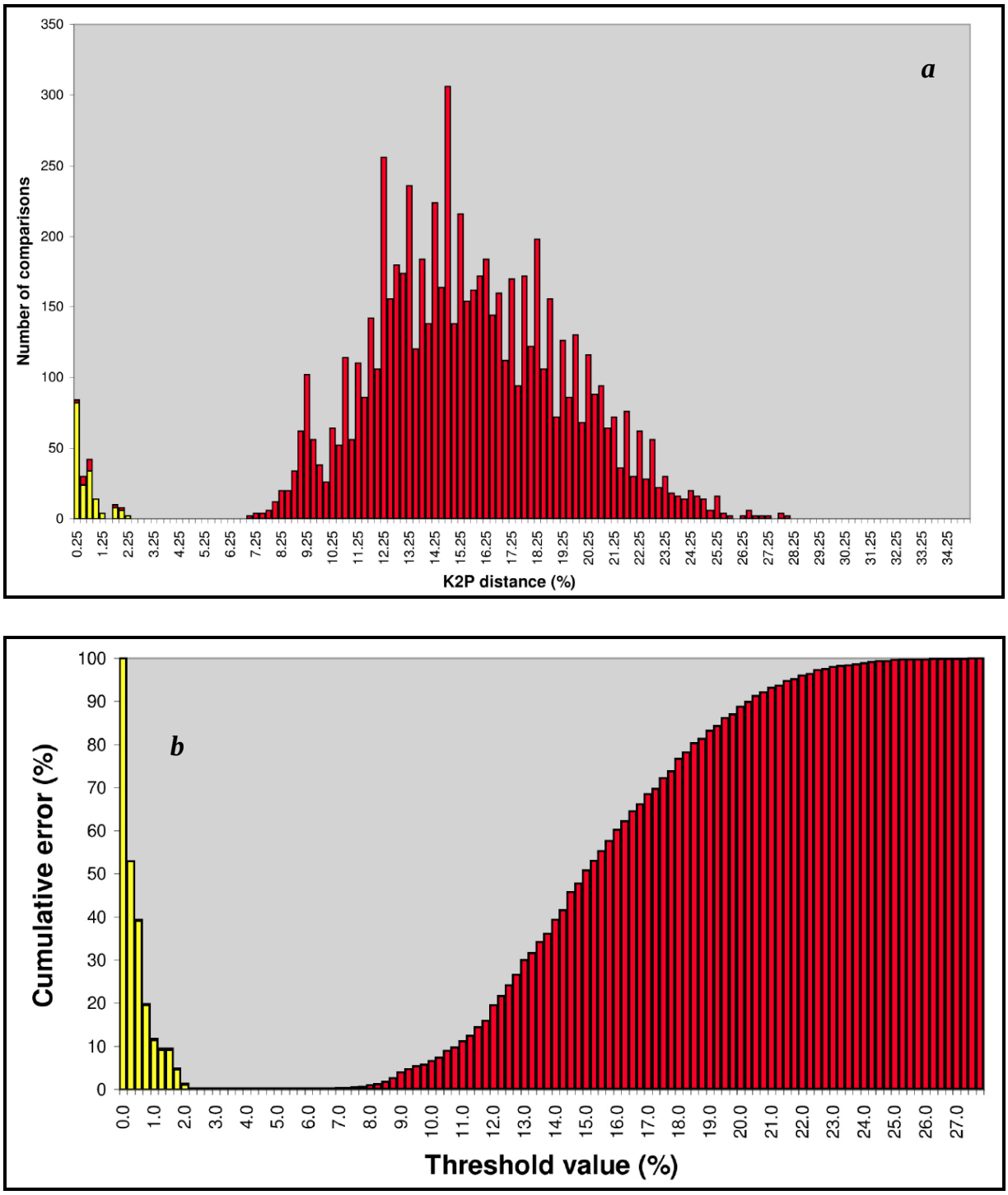


Figure 3.3: (*a*) Frequency distribution of intraspecific (yellow bars) and interspecific (red bars) genetic K2P divergences in the reference dataset encompassing *coxI* sequences of morphologically identified filarioid nematodes (case study B1). Graph shows 877 intraspecific and 21775 interspecific comparisons across 46 filarioid species. **(*b*)** Cumulative error plot based on the same dataset. Type I (yellow) and type II (red) errors obtained with different thresholds for *coxI* sequences of 46 spirurida species. The Optimum Threshold value (i.e. 4.8 %) correspond to the Minimum Cumulative Error (i.e 0.62 %) due to the occurrence of type II errors only.

In B1 (filarioid nematodes), K2P distance graphs and cumulative error plots were also produced for the sixteen combination of software/parameters tested on the two homologous datasets (i.e. those datasets including *coxI* and 12S rDNA barcode sequences for the same specimens; data not shown). A schematic representation of the efficacy for each different combination used is provided in Table 4.6.

Finally, in the context of case study D1 and in particular for the *coxI* sub-datasets relative to the clone sequences of nematodes, rotifers and tardigrads, MOTUs were alternatively identified by using BLASTClust (v. 2.2.1.6; Altschul et al., 1997 - http://toolkit.tuebingen.mpg.de/blastclust). BLASTClust is a program within the standalone BLAST package used to cluster either protein or nucleotide sequences. The program begins with pairwise matches and places a sequence in a cluster if the sequence matches at least one entry already presents in the cluster. BLASTClust accepts a number of parameters that can be used to control the stringency of clustering including thresholds for score density, percent identity, and alignment length. In the case of D1 dataset, sequences were grouped by considering the entire length of the alignment (i.e. 522bp - 'sequence length covered = 100%') and two different threshold of similarity: 99% and 97.5%. The former threshold value (i.e. 1%) was chosen according to the limit fixed by the BOLD system for species identification (Ratnasingham & Hebert, 2007). A less conservative threshold, representing an intermediate value of molecular divergence among species, was chosen according to recent literature. Indeed, 3% K2P threshold has been revealed to be efficient in distinguishing the most part of species for *coxI* in fishes (Wong & Hanner, 2008), tardigrads (Cesari et al., 2009), amphibians (Fouquet et al., 2007), anellids (Huang et al., 2007), birds (Hebert et al., 2004b), lepidopterans (Emery et al., 2009), dipters (Meiklejohn et al., 2009) and hymenopters (Smith et al., 2005). However, a 2% K2P threshold has also been successfully tested in order to identify other birds species (Johnsen et al., 2010), aracnids (Barrett & Hebert, 2005), and several arthropods (Zhou et al., 2009). Hence, an intermediate threshold value of 2.5% was chosen as the second cutoff to analyse MOTUs composition in case study D1.

3.8 MULTI-APPROACH IDENTIFICATION OF BLIND SAMPLES

Concerning case studies A1 (Italian bats) and C1 (shark seafood products), the OT values calculated on the respective reference datasets have been used to assess a species rank

to the unidentified samples, testing for the congruence with previously described IOTUs or MOTUs. In the case of Italian bats, the aim of this approach was the identification of those individuals belonging to cryptic taxa (e.g. genera *Myotis, Plecotus, Pipistrellus*) as reported in Appendix Table I.1. Blind samples collected in the context of case study C1 refer to slices or fillets of sharks sold as 'Palombo' (i.e. *Mustelus mustelus* or *M. asterias*) in Northern Italian markets (Figure 2.6).

In both cases, identification results were also compared with the Identification Engine tool (IDS) on BOLD (Barcode of Life Data System, http://www.barcodinglife.org/-views/login.php; Species Level Barcode Records database), which assumes a correct species identification of the queried barcode sequences for genetic distance values up to 99% of specimen similarity (Ratnasingham & Hebert, 2007). Similarities and differences inferred from this comparison are also reported in Appendix Tables II.4 and II.5). Moreover, for case study C1, the top species matches (highest percentage of similarity) obtained with these two approaches were compared to the labeled name recorded at the market in order to determine the percentage of species substitution in the acquiring of 'palombo' slices.

3.9 IDENTIFICATION OF MOLECULAR LINEAGES AND ADDITIONAL BIOINFORMATIC ANALYSES

Concerning comprehensive datasets, the resulting K2P distance matrixes have been used: *i*) to infer MOTUs delimited by OT; *ii*) to analyse the MOTU's composition testing for the congruence with previously described species (DNA barcoding); *iii*) to perform prediction and classification of potentially new taxa (DNA taxonomy). As a general assumption, when barcodes belonging to the same species could be divided into two or more well-supported clusters using the methods described above, they were flagged as potentially cryptic taxa deserving further investigations (e.g. *Myotis nattereri* in case study A1 or *Taenia taeniaformis*, in case study B2; see chapter 4).

Additional analyses have been performed on the new lineages revealed by OT in the case of Italian bats. In particular, following the observations of Clare et al. (2007) and Francis et al. (2010), we investigated the occurrence of new intraspecific molecular lineages for those species showing two or more distinct molecular clusters separated by a mean K2P distance greater than 2% or exceeding the species boundaries inferred by OT. A list of these taxa is

provided in Table 4.8. Two additionally NJ trees based on *cyt b* and *ND1* sequences belonging to *Myotis nattereri* and related taxa (e.g. *Myotis escalerai*) were produced (see Figure 4.5 and Appendix Figure III.1). Part of the sequences included in these datasets derives from previous phylogeographic studies on *M. nattereri* (Ibañez et al., 2006; García-Mudarra et al., 2009; see Appendix Table II.3) and have been retrieved from GenBank while the other sequences were directly obtained from samples analysed in the context of this study and refer to the lineages discovered in the Italian Peninsula by our preliminary DNA barcoding analyses (chapter 4.6). Following Kerr et al. (2009), a NJ clustering approach was used to identify different lineages and flag potentially cryptic taxa within the complex *M. nattereri* (see Figure 4.5 and Appendix Figure III.1). GenBank sequences DQ120800 (*ND1*) and AF246241 (*cyt b*) relative to the species *Myotis myotis* were used as outgroup.

In the case of Parrotbills further analyses were conducted in order to investigate the molecular differentiation between Chinese and the naturalized Italian populations and to trace the provenance of the founders released in the Palude Brabbia Natural Reserve (Varese, Italy). In particular, haplotype network reconstructions (Figure 4.7; Appendix Figure III.2), nucleotide diversity and uncorrected p-distances calculations have been performed. Details on softwares and conditions used in these analyses are provided in Appendix IV.1.

4. RESULTS AND DISCUSSION

4.1 DNA EXTRACTION, AMPLIFICATION AND SEQUENCING SUCCESS

A total of about 500 biological samples belonging to animals, plants and an environmental matrix were used for DNA extraction (Table 3.1 and Appendix I.1-I.8). High DNA quality and good yield (generally from 30 up to 50 ng/µl) were obtained from all the analysed samples, with the exclusion of oldest museum specimens (e.g. some bats), smallest organisms (e.g. microfilariae) and some industrially processed spices samples (e.g. mints and one basil samples). In these cases, electrophoretic analysis showed partially degraded DNA in the 100–1000 bp range and low yield of DNA extraction (less than 20 ng/µl) (data not shown).

Concerning animals (including the hydrobiont fauna isolated from the environmental sample of case study D1), the results of our tests showed good amplification and sequencing success among the wide range of investigated taxa (see Table 3.1). All of the analysed barcode loci exhibited high PCR success with standard primers also for the additional markers examined in some case studies (see Tab 3.3). Moreover, the cloning analysis conducted on the four *coxI* amplicons relative to the respective pools of organisms extracted with Baermann method (case study D1; see chapters 3.2 and 3.5), allowed to obtain a total of 526 sequences out of the 768 sequenced clones (i.e. 96 clones were sequenced for each plate): mean number of clone sequences per Baermann's aliquot: 131.5; S.E.: 16.7; range: 111 – 151; mean number of sequences per cloning plate: 65.8; S.E.: 11.8; range: 53 – 90; See also Figure 4.1).

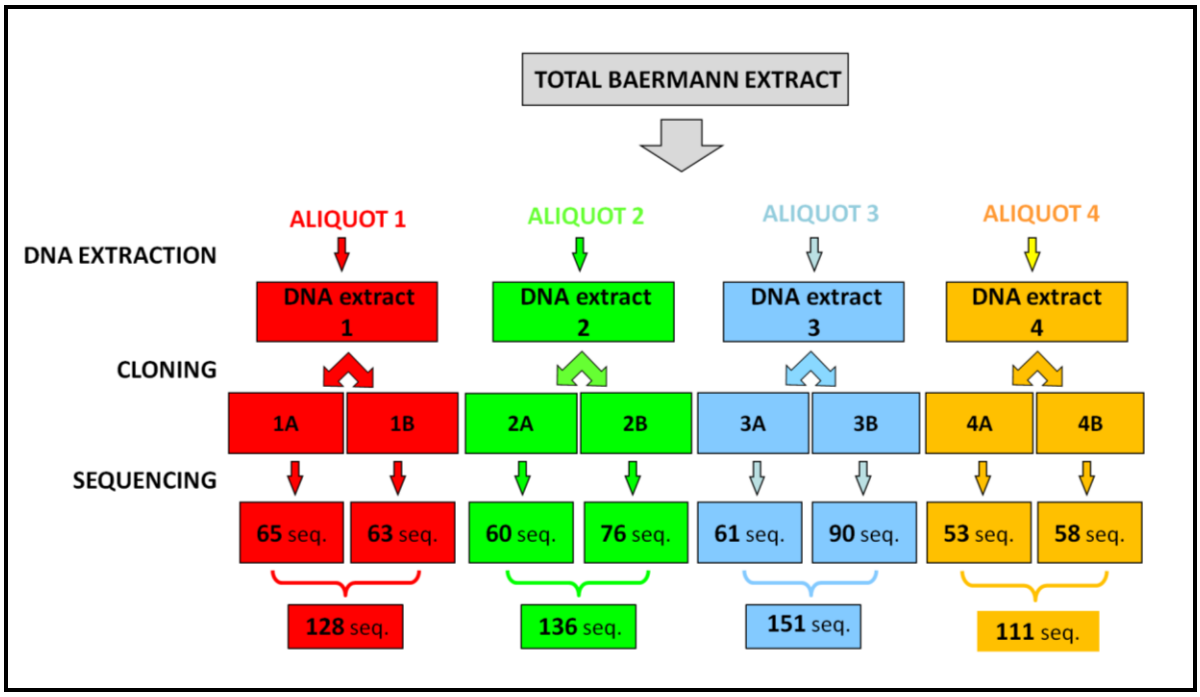


Figure 4.1: Schematic representation of the PCR-cloning approach performed on the hydrobiont moss extract of case study D1. The number of *coxI* sequences obtained for each cloning plate and the total number of *coxI* sequences for each aliquot of the Baermann extract is provided.

Some amplification or sequencing failures occurred in plant samples. In particular, *Group I* of spices dataset (case study C2) showed some problems in the amplification of *matK* and *rpoB* genes for the commercial processed mint MEC3 and *Mentha acquatica* L. samples respectively (Appendix Table I.7). Despite the amplification step was repeated three times, starting from different amounts of DNA (i.e. 1, 10 and 20 ng) and at low stringency conditions (see chapter 3.3), results were negative in all of the cases. Sequencing problems were also encountered with the trnH-psbA for some Basil cultivar samples (i.e. OBcv2, OBcv4, OBcv7; see Appendix I.7) and were in part attributable to the high frequency of mononucleotide repeats disrupting individual sequencing reads. Such a problem was solved with a strong and careful manual editing of the obtained sequences. These preliminary results indicate that in almost all the cases considered, the industrial processes which the commercial spices samples collected have been subjected to (e.g. crumbling, drying) did not affect the success rate of DNA extraction, amplification and sequencing. On the contrary, results of our DNA barcoding analysis across the three selected plant groups of case study C3 (i.e. poisonous plants) showed a conspicuous difference among the five tested loci with respect to amplification success, PCR products lengths and sequences quality (Table 4.1).

Tabella 4.1: Summary of the proportion of individuals successfully amplified and sequenced for the five candidate barcode regions in three analysed groups of poisonous plants (case study C3). NA = Not Alignable.

	Group	trnH-psbA	*matK*	*rpoB*	*sqd1*	*At103*
Amplification success	I	34/37	36/37	36/37	17/37	29/37
	IIa	3/4	4/4	3/4	2/4	2/4
	IIb	3/3	3/3	3/3	1/3	3/3
	IIIa	5/5	5/5	5/5	5/5	5/5
	IIIb	6/6	6/6	6/6	6/6	5/6
Sequencing success	I	33/37	35/37	36/37	16/37	27/37
	IIa	3/4	4/4	3/4	2/4	2/4
	IIb	3/3	3/3	3/3	1/3	3/3
	IIIa	5/5	5/5	5/5	5/5	5/5
	IIIb	6/6	6/6	6/6	6/6	5/6
Sequence length (bp)	I	210-472	788-850	416-479	237-262	299-730
	IIa	218-238	781-791	477	237	302-364
	IIb	423-454	771-788	417	_	305-478
	IIIa	470-486	837-846	471	262	326-351
	IIIb	358-364	850	474	255	482
Alignment success	I	NA	35/37	36/37	16/37	28/37
	IIa	3/4	4/4	3/4	2/4	2/4
	IIb	3/3	3/3	3/3	1/3	3/3
	IIIa	5/5	5/5	5/5	5/5	5/5
	IIIb	6/6	6/6	6/6	6/6	5/6

As showed in the table, all the three plastidial markers (i.e. *matK*, *rpoB*, trnH-psbA) exhibited highest PCR and sequencing success with standard primers with 10 ng of DNA as template for DNA amplification. In the case of nuclear markers, only the *At103* was successfully amplified in standard PCR conditions for a large number of plant species, while *sqd1* showed amplification problems for several samples belonging to Groups I and IIb also even starting from 1, 10 and 25 ng of DNA. All the PCR products were easily sequenced, with exclusion of some samples (Table 4.1). The greatest problem in obtaining good quality in bidirectional sequences were encountered with the *At103* sequences due to the high frequency of mononucleotide repeats.

Despite some problems in amplification and sequencing, these first results highlight the efficacy of DNA barcoding in working on a wide range of toxic plants species and parts of them. Moreover, it is important to underline that DNA was extracted from plant samples collected in the field or in private apartments and stored at room temperature without specific

methods to preserve DNA quality. DNA amplification and sequence analyses (Table 4.1) suggested that plant materials used for DNA extraction did not affect the efficacy of a DNA barcoding survey. These aspects should be useful in order to apply the method in poisonous centers, where toxic plants are usually identified starting from different portions of plant material (fruits, seed, leaves, etc.).

Finally, as a general remark, it is important to underline that several barcode sequences produced in the context of this research project represented the first entries in GenBank database for many animal and plant taxa (e.g. 23 out of the 31 species of Italian bats, 4 shark species, almost all the barcodes of case study C3 etc.).

4.2 NUMTs AND *COXI*- LIKE SEQUENCES INTERFERENCE

Almost any evidence for the occurrence of numts interference was revealed within the molecular barcode datasets and sub-datasets here analysed. However, some exceptions occurred in certain metazoan taxa and in the environmental datasets. In the context of case study A1 (i.e. Italian bats) all of the three processed *Plecotus sardus* samples showed the inadvertent amplification of a *coxI*-like sequence when the primer pair VF1d-VR1d was used in the amplification step. In particular, the obtained sequences contained one stop-codon and composition biases with respect to other congeneric taxa. As a further remark, the amplification of two additional mitochondrial markers for these samples (i.e. *ND1* and *cyt b*; data not shown), confirmed the validity of the taxonomical identification therefore excluding the opportunity to deal with a putative undescribed molecular-divergent taxon. A new PCR amplification with the other primer pair (i.e. LCO1490 – HCO2198) allowed to obtain the correct DNA barcode sequences. A similar situation occurred in the case of taeniids dataset (i.e. case study B2) in which numts screening procedures revealed the presence of *coxI*-like data among sequences retrieved from GenBank: a.n. AF096242, AF429314 (*T. asiatica*), AF096241, U45988 (*T. saginata*), AF360866, EU747650 (*T. solium*) and AF096243 (*T. taeniaeformis*). These sequences were removed from our molecular datasets.

Surprisingly, at least three out of the 47 *coxI* sequences obtained from terrestrial isopods of case study A3, have been assigned to the phylum Nematoda by a preliminary BLAST comparison within GenBank database (i.e. samples MIB:ZPL:01425; 01426; 01428

belonging to the genera *Armadillidium* and *Philoscia*). Although uncharacterized in GenBank (the nearest sequences belonged to *Rhabdias okuensis*) it is possible that these sequences belong to parasitic nematodes affecting isopods, and further analyses are ongoing. Indeed, the occurrence of endoparasitic nematodes in terrestrial isopods is largely documented (Poinar, 1981; Moore & Lasswell, 1986). As a further remark, the dissection of other individuals belonging to the same genera mentioned above showed the frequent occurrence of nematodes in the hemocoel region (see Figure 4.2). The universal primer used for *coxI* amplification probably fitted better with these parasitic nematodes therefore resulting in a over-amplification and consequently sequencing of their mtDNA. To avoid this problem DNA extractions of terrestrial isopods samples have been performed starting from internal organs and not from the entire individual.

Figura 4.2: Dissection of a *Porcellio siculoccidentalis* (MIB:ZPL:01602) hosting an endoparasite nematode. [images provided by Montesanto G.]

Finally, some of the *coxI* clones obtained for case study D1 revealed high BLAST matches with sequences of bacteria belonging to the *Wolbachia* group and were then removed from the molecular dataset. This situation is not uncommon when universal primers are used and it has been previously documented in lepidopterans and other insects (see for example Linares et al., 2009). Such a situation can open new perspectives on knowledge about the complicated symbiont-host interactions (e.g. the possible occurrence of lateral gene transfer events) that deserve to be further investigated.

4.3 ALIGNMENT CHARACTERISTICS

Alignment characteristics of DNA barcoding datasets and sub-datasets for each case study are resumed in Table 4.2. Sequences belonging to other molecular markers (e.g. in the case of parrotbills and Italian bats), blind samples or unknown organisms produced in this study or collected from GenBank have not been taken into account.

Table 4.2: Alignment characteristics of each molecular dataset and sub-dataset produced for the case studies investigated in this research project (C.S.). Alignment length, the occurrence of *indel* regions, empirical base frequencies and the number of variable sites (V) are also provided. n.c.: not computable. In the case of filarioid nematodes a sub-dataset of *coxI* barcode sequences (*coxI**) has been assembled.

C.S	DATASET	Barcode region	Sub-dataset	Lenght (bp)	indels	Average base frequencies				V
						π_T	π_C	π_A	π_G	
A1	ITALIAN BATS	*coxI*	_	556	no	0.328	0.250	0.256	0.166	226
A2	PARROTBILLS	*coxI*	_	663	no	0.249	0.337	0.245	0.170	22
A3	TERRESTRIAL ISOPODS	*coxI*	_	504	no	0.379	0.168	0.241	0.212	253
B1	FILARIOID NEMATODES	*coxI*	_	627	yes	0.471	0.127	0.183	0.219	342
			*coxI**	627	yes	0.472	0.127	0.187	0.214	329
		12S rDNA	_	643	yes	0.506	0.070	0.267	0.157	392
B2	TAENIIDS	*coxI*	_	380	no	0.449	0.088	0.227	0.236	161
C1	SHARK SEAFOOD PRODUCTS	*coxI*	_	550	no	0.338	0.248	0.260	0.154	241
C2	SPICES	*matK*	*Group I*	810	no	0.365	0.181	0.287	0.166	9
			Group II	801	no	0.369	0.177	0.284	0.170	43
			Group III	810	no	0.362	0.187	0.290	0.161	0
			Group IV	810	no	0.366	0.184	0.284	0.166	44
			Group V	810	no	0.363	0.186	0.288	0.163	0
			Group VI	804	no	0.364	0.182	0.281	0.173	0
		trnH-psbA	*Group I*	406	yes	0.348	0.139	0.351	0.163	15
			Group II	395	yes	0.359	0.139	0.341	0.161	21
			Group III	437	yes	0.353	0.136	0.356	0.155	51
			Group IV	421	yes	0.351	0.133	0.363	0.153	32
			Group V	416	yes	0.361	0.132	0.353	0.154	0
			Group VI	337	no	0.329	0.139	0.359	0.172	0
		rpoB	*Group I*	491	no	0.277	0.169	0.312	0.242	0
			Group II	491	no	0.279	0.173	0.312	0.236	0
			Group III	491	no	0.277	0.167	0.312	0.244	0
			Group IV	491	no	0.278	0.170	0.314	0.237	9
			Group V	491	no	0.277	0.167	0.312	0.244	0
			Group VI	491	no	0.279	0.171	0.314	0.236	0
		rbcL	*Group I*	551	no	0.287	0.214	0.262	0.238	1
			Group II	551	no	0.287	0.214	0.263	0.236	4
			Group III	551	no	0.287	0.214	0.260	0.239	2
			Group IV	551	no	0.287	0.215	0.263	0.236	14
			Group V	551	no	0.289	0.212	0.263	0.236	0
			Group VI	551	no	0.289	0.212	0.263	0.236	0

C.S	DATASET	Barcode region	Sub-dataset	Lenght (bp)	indels	Average base frequencies				V
						π_T	π_C	π_A	π_G	
C3	POISONOUS PLANTS	*matK*	Group I	825	yes	0.370	0.182	0.292	0.156	525
			Group IIa	787	no	0.384	0.176	0.297	0.143	328
			Group IIb	767	no	0.359	0.187	0.284	0.170	2
			Group IIIa	846	yes	0.365	0.186	0.299	0.150	25
			Group IIIb	850	no	0.369	0.182	0.294	0.156	24
		trnH-psbA	Group I	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
			Group IIa	240	yes	0.426	0.110	0.262	0.203	22
			Group IIb	454	yes	0.407	0.096	0.333	0.165	44
			Group IIIa	501	yes	0.409	0.121	0.319	0.151	64
			Group IIIb	380	yes	0.414	0.114	0.318	0.154	65
		rpoB	Group I	423	yes	0.284	0.177	0.339	0.201	176
			Group IIa	477	no	0.274	0.188	0.307	0.232	2
			Group IIb	417	no	0.273	0.194	0.331	0.202	1
			Group IIIa	471	no	0.285	0.169	0.322	0.223	13
			Group IIIb	474	no	0.269	0.186	0.322	0.223	60
		Sqd1	Group I	237	no	0.266	0.183	0.285	0.265	96
			Group IIa	237	no	0.275	0.172	0.289	0.264	6
			Group IIb	n.c.	_	n.c.	n.c.	n.c.	n.c.	n.c.
			Group IIIa	262	no	0.285	0.188	0.305	0.223	17
			Group IIIb	255	no	0.256	0.192	0.307	0.244	47
		At103	Group I	234	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
			Group IIa	300	no	0.320	0.186	0.283	0.211	16
			Group IIb	166	yes	0.313	0.227	0.276	0.184	146
			Group IIIa	230	yes	0.325	0.216	0.276	0.183	123
			Group IIIb	482	no	0.320	0.217	0.287	0.176	27
D1	MEIOFAUNAL MOSS BIODIVERSITY	*coxI*	_	522	yes	0.425	0.137	0.257	0.181	421

Table 4.2. *continued.*

As showed in the table, alignment lengths for the same barcode region varied significantly among animal and plant taxa mainly due to slightly differences in the position of the annealing sites or depending on the amplification of unconventionally regions for some barcode markers (e.g. the primer pair JB3-JB4.5 in taeniids). In animal datasets, alignment lengths for *cox*I ranged between 380 bp (taeniids dataset) and 663 bp (parrotbills), while in plants larger differences were observed for some markers in different sub-datasets even characterized by species belonging to the same family. As an example, in case study C2 (i.e. spices belonging to the family Lamiaceae) the same sequence length was observed in all analysed group for *matK* (810 bp), *rbcL* (551 bp) and *rpoB* (491 bp), but substantial differences were observed in the trnH-psbA alignment length among the different groups (i.e. from 337 to 437 bp). Similarly, considerable differences in sequence and alignments lengths

were observed within and among sub-datasets of case study C3 (i.e. poisonous plants; see Table 4.1), even impeding the alignment of trnH-psbA sequences in the case of GroupI.

4.4 EXTENT AND DISTRIBUTION OF K2P MOLECULAR VARIABILITY

Extent of genetic differences (in terms of K2P values transformed into percent) between and within species for each case study (and related sub-datasets) and for each barcode region is provided in Table 4.3. These kinds of values are reported in a separate table for case study C2 in which three combinations of the best performing markers (obtained by merging in the same alignment homologous barcode sequences for each sample) have been also investigated (i.e.: matK+ trnH-psbA; matK + rbcL; trnH-psbA +rbcL; seeTable 4.4).

As showed in Table 4.3, according to the key-points fixed by CBOL, at least three barcode sequences per species have been produced for the most part of those case studies including animal samples. Related high values for standard deviation in some molecular datasets are due to the over-representation of specimens for some species (i.e. 28 *coxI* sequences of *Squalus acanthias* in C1; 38 *coxI* sequences of *Taenia solium* in B2). However, a mean of one barcode sequence per species was produced for the most part of plant species therefore impeding the inference of an Optimum Threshold value useful to delimitate species, IOTUs and MOTUs boundaries (see chapter 3.7).

As a general condition, the mean K2P variability among species was several times greater than the intraspecific one for almost all the tested barcode markers. The only exception occurred for trnH-psbA and *rbcL* barcode sequences belonging to the sub-dataset *Group III* (*Origanum* samples) of case study C2 in which the highest levels of molecular variability correspond to the intraspecific comparisons (see Table 4.4). A possible explanation could be achieved by taking into account the occurrence of hybridization events between different commercial species belonging to the Oregano group (as already reported by Gounaris et al., 2002; see also DeMattia et al., *in press*).

Table 4.3: Extent of K2P molecular variability for each dataset and sub-dataset belonging to the case studies analysed in this research project (C.S). The average K2P distance (transformed into percent), the standard error (S.E.) and the range of K2P variation are given for each tested marker. For each group, the number of species (n) and the mean number of barcode sequences per species (m.n.) is provided. Both the intraspecific and the interspecific comparisons for each group are considered. n.c.: not computable.

C.S.	DATASET	Barcode region	Sub-dataset	n	m.n. (S.E.)	Mean K2P variability % (S.E.%)				O. M. % (S.E. %)
						Within species	Range %	Between species	Range %	
A1	ITALIAN BATS	*coxI*	_	30	5.8 (4.4)	0.33 (0.53)	0 - 3.92	21.22 (3.51)	0 - 28.64	19.82 (1.35)
A2	PARROTBILLS	*coxI*	_	2	12.5 (6.4)	0.54 (0.47)	0 - 2.00	0.68 (0.15)	0 - 2.15	0.72 (0.16)
A3	TERRESTRIAL ISOPODS	*coxI*	_	26	3.7 (5.1)	1 .00 (0.20)	0 - 5.40	26.00 (2.30)	9.4 - 35	23.20 (1.20)
B1	FILARIOID NEMATODES	*coxI*	_	46	3.3 (3.4)	0.50 (0.60)	0 - 2.40	16.20 (3.70)	0 - 27.80	16.00 (1.00)
			coxI*	44	2.0 (1.5)	0.50 (0.50)	0 - 2.00	15.50 (3.70)	0.20 - 27.80	15.00 (1.00)
		12S rDNA	_	44	2.0 (1.5)	2.20 (1.70)	0 - 6.00	17.40 (4.20)	0.20 - 34.50	17.00 (1.10)
B2	TAENIIDS	*coxI*	_	14	9.5 (9.8)	0.60 (0.40)	0 - 3.50	14.10 (3.70)	1.90 - 20.30	12.40 (1.20)
C1	SHARK SEAFOOD PRODUCTS	*coxI*	_	110	4.5 (3.8)	0.19 (0.34)	0 - 2.60	19.55 (5.62)	0.40 - 30	19.20 (1.40)
C3	POISONOUS PLANTS	*matK*	Group I	35	1 (0)	n.c.	n.c.	22.80 (1.07)	0 - 39.88	22.80 (1.07)
			Group IIa	4	1 (0)	n.c.	n.c.	1.40 (0.70)	0.10 - 1.90	1.40 (0.70)
			Group IIb	3	1 (0)	n.c.	n.c.	0.20 (0.17)	0 - 0.30	0.20 (0.17)
			Group IIIa	5	1 (0)	n.c.	n.c.	1.38 (0.49)	0.50 - 2	1.38 (0.49)
			Group IIIb	6	1 (0)	n.c.	n.c.	1.07 (0.50)	0.40 - 2.20	1.07 (0.50)
		trnH-psbA	Group I	33	1 (0)	n.c.	n.c.	n.c.	n.c.	n.c.
			Group IIa	3	1 (0)	n.c.	n.c.	6.33 (2.96)	2.40 - 10.30	6.33 (2.96)
			Group IIb	3	1 (0)	n.c.	n.c.	7.57 (3.28)	3.80 - 9.80	7.57 (3.28)
			Group IIIa	5	1 (0)	n.c.	n.c.	6.35 (2.66)	2.60 - 10.30	6.35 (2.66)
			Group IIIb	6	1 (0)	n.c.	n.c.	6.75 (7.19)	0.60 - 17.80	6.75 (7.19)
		rpoB	Group I	36	1 (0)	n.c.	n.c.	11.80 (1.10)	0 - 18.67	11.80 (1.10)
			Group IIa	3	1 (0)	n.c.	n.c.	0.23 (0.15)	0 - 0.40	0.23 (0.15)
			Group IIb	3	1 (0)	n.c.	n.c.	0.13 (0.12)	0 - 0.20	0.13 (0.12)
			Group IIIa	5	1 (0)	n.c.	n.c.	1.53 (0.71)	0 - 2.40	1.53 (0.71)
			Group IIIb	6	1 (0)	n.c.	n.c.	4.85 (6.15)	0.20 - 13.60	4.85 (6.15)
		Sqd1	Group I	16	1 (0)	n.c.	n.c.	19.70 (2.06)	0 - 30.59	19.70 (2.06)
			Group IIa	2	1 (0)	n.c.	n.c.	1.26 (1.12)	1- 3.40	1.26 (1.12)
			Group IIb	1	1 (0)	n.c.	n.c.	n.c.	n.c.	n.c.
			Group IIIa	5	1 (0)	n.c.	n.c.	3.30 (1.29)	1.20 - 5.20	3.30 (1.29)
			Group IIIb	6	1 (0)	n.c.	n.c.	7.54 (9.20)	0 - 21.00	7.54 (9.20)

C.S.	DATASET	Barcode region	Sub-dataset	n	m.n. (S.E.)	Mean K2P variability % (S.E.%)				O. M. % (S.E. %)
						Within species	Range %	Between species	Range %	
C3	POISONOUS PLANTS	*At103*	Group I	27	1 (0)	n.c.	n.c.	17.80	0 -	17.80
			Group IIa	2	1 (0)	n.c.	n.c.	2.20 (0.96)	1 - 3.40	2.20 (0.96)
			Group IIb	3	1 (0)	n.c.	n.c.	17.43 (3.55)	13.6 - 20.6	17.43 (3.55)
			Group IIIa	5	1 (0)	n.c.	n.c.	9.87 (8.16)	2.20 - 20.1	9.87 (8.16)
			Group IIIb	5	1 (0)	n.c.	n.c.	2.35 (1.23)	0.60 - 4.30	2.35 (1.23)
D1	MEIOFAUNAL MOSS BIODIVERSITY	*coxI*	_	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	37.1 (2.11)

Table 4.3. *continued.*

Table 4.4*a*: Comparative performances and variability of different DNA barcoding markers tested for spices identification (case study C2). For each group (G), the average K2P distance (transformed into percent), the standard error (S.E.) and the range of K2P variation are given for each tested marker. For each group, the number of species (N) and the mean number of Barcode sequences per species (MN) is provided. Both the intraspecific and the interspecific comparisons for each group are considered. n.c.: not calculable.

G	N	MN (S.E.)	Comparison	Single markers							
				matK		trnH-psbA		*rpoB*		*rbcL*	
				Mean % variation (S.E.%)	range %	Mean % variation (S.E.%)	range %	Mean % variation (S.E.%)	range %	Mean % variation (S.E.%)	range %
I	3	2.3(0.6)	between sp.	0.72(0.60)	0.08-1.08	2.31(1.70)	0.30-3.33	0 (0)	n.c.	0.04(0.03)	0-0.06
			within sp.	0.03(0.05)	0-0.08	0.20(0.18)	0-0.34	0 (0)	n.c.	0.04(0.07)	0-0.12
II	3	4.3(5.8)	between sp.	1.76(0.40)	1.26-2.09	2.75(0.70)	2.12-3.51	0(0)	n.c.	0.48(0.20)	0.36-0.73
			within sp.	1.05(0)	n.c.	0.72(0)	n.c.	0(0)	n.c.	0(0)	n.c.
III	4	2.8(2.1)	between sp.	0(0)	n.c.	1.05(0.53)	0-1.37	0(0)	n.c.	0.05(0.03)	0-0.08
			within sp.	0(0)	n.c.	1.37(0.55)	0.98-1.76	0(0)	n.c.	0.10(0.01)	0.09-0.11
IV	4	1.5(1)	between sp.	3.39(1.70)	0.74-4.86	4.90(2.47)	1.03-7.05	1.03(0.43)	0.41-1.65	1.54(0.85)	0.36-2.41
			within sp.	0(0)	n.c.	0(0)	n.c.	0(0)	n.c.	0(0)	n.c.
V	1	3 (n.c.)	between sp.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
			within sp.	0(0)	_	0(0)	n.c.	0(0)	n.c.	0(0)	n.c.
VI	1	3 (n.c.)	between sp.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
			within sp.	0(0)	n.c.	0(0)	n.c.	0(0)	n.c.	0(0)	n.c.

Table 4.4*b*: Comparative performances and variability of different DNA barcoding markers tested for spices identification (case study C2). For each group (G), the average K2P distance (transformed into percent), the standard error (S.E.) and the range of K2P variation are given for each tested combination of barcode markers. For each group, the number of species (N) and the mean number of barcode sequences per species (MN) is provided. Both the intraspecific and the interspecific comparisons for each group are considered. n.c.: not calculable.

G	N	MN (S.E.)	Comparison	Marker combinations					
				matK+ trnH-psbA		*matK* + *rbcL*		trnH-psbA +*rbcL*	
				Mean % variation (S.E.%)	range %	Mean % variation (S.E.%)	range %	Mean % variation (S.E.%)	range %
I	3	2.3(0.6)	between sp.	2.31(1.74)	0.3-3.33	0.44(0.33)	0.07-0.67	0.97(0.70)	0.16-1.40
			within sp.	0.20(0.18)	0-0.34	0.03(0.06)	0-0.1	0.11(0.11)	0-0.21
II	3	4.3(5.8)	between sp.	2.07(0.50)	1.54-2.54	1.23(0.30)	0.89-1.43	1.40(0.29)	1.08-1.63
			within sp.	0.95(0)	n.c.	0.62(0)	n.c.	0.30(0)	n.c.
III	4	2.8(2.1)	between sp.	0.35(0.18)	0-0.46	0.02(0.01)	0-0.03	0.47(0.24)	0-0.63
			within sp.	0.46(0.18)	0.33-0.59	0.04(0)	n.c.	0.64(0.24)	0.47-0.81
IV	4	1.5(1)	between sp.	3.88(1.94)	0.84-5.40	2.63(1.32)	0.59-3.55	2.91(1.45)	0.64-4.01
			within sp.	0(0)	n.c.	0(0)	n.c.	0(0)	n.c.
V	1	3 (n.c.)	between sp.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
			within sp.	0(0)	n.c.	0(0)	n.c.	0(0)	n.c.
VI	1	3 (n.c.)	between sp.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
			within sp.	0(0)	n.c.	0(0)	n.c.	0(0)	n.c.

In other cases, reduced values of interspecific K2P divergence or slightly differences between the two components of the molecular variability were observed such as in the case of parrotbills and for some barcode markers in plant datasets (see Tables 4.3 and 4.4 *a*,*b*). A brief description of these cases is here provided:

- **Parrotbills (case study A2)**

  Preliminary investigations on *coxI* molecular variability for the two parrotbills species (i.e. *Paradoxornis webbianus* and *P. alphonsianus*) morphologically identified and collected from both Italian and Chinese localities resulted as follow: mean K2P distance within species 0.54%, mean K2P distance between species 0.68% (see also

Table 4.3). These amounts of genetic distances are low in comparison to the values showed in the others animal datasets and are far from being close to the *coxI* barcode thresholds fixed for species identification in birds (i.e. > 3.00%; see Aliabadian et al., 2009 for more details). This situation can be also observed in the NJ reconstruction provided in Fig. 4.3.

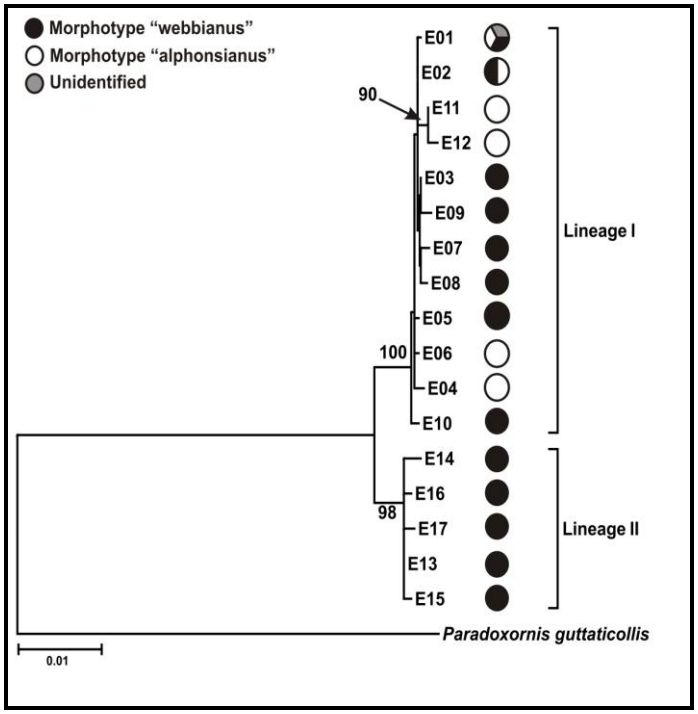


Figure 4.3: NJ tree based on the analysis of the parrotbills *coxI* dataset (case study A2) using K2P correction, 1000 boostrap replicates and a chimera sequence of *P. guttaticollis* as outgroup. The percentage of *coxI* haplotypes from individuals assigned to one or none of the two taxa is identified by the presence of a differently filled pie. Black: vinous-throated, white: ashy-throated, grey: unassigned individual. Bootstrap values lower than 70 not shown.

Interestingly, the NJ tree identifies two distinct and well-supported clades here tentatively named as Lineage I and Lineage II. Lineage I includes haplotypes belonging to both the morphospecies here investigated (i.e. *Paradoxornis webbianus* and *P. alphonsianus*) and also undetermined individuals sampled in Italy (i.e. showing intermediate phenotypic characters between the two species). Lineage II includes only *coxI* haplotypes belonging to *P. webbianus*. These lineages are separated for a mean K2P distance value of 1.41% (standard deviation: 0.26%; range: 0.98% – 2.16%) which is in line with the genetic divergences generally observed at the subspecies level in birds.

These results supported the synonymisation of *P. alphonsianus* with *P. webbianus* in spite of the existence of two different morphotypes. However, further investigations on the genetic differentiation between the two molecularly-identified lineages are largely required in order to shed more light on the taxonomic status of these birds. In this context a phylogeographic approach has been performed (see chapter 4.7.2).

- **Spices (case study C2)**

As showed in Table 4.4 *a*, the trnH-psbA barcode region ranked first in divergence values among species, followed by *matK*. On the contrary, *rpoB* sequences showed the lower K2P divergence values. A negligible genetic diversity based on 6 single base mutations for trnH-psba and only one for *matK* was revealed between *M. spicata* and *M. piperita* (*Group I*). Analyses carried out on samples belonging to *Group II* (genus *Ocimum*) provided evidence that trnH-psbA and *matK* barcode sequences also show appreciable differences among the analyzed basil cultivars with mean K2P distance values of 0.82% (s.e. 0.5%) for trnH-psbA and 1.21% (s.e. 0.6%) for *matK*. In both cases the observed differences were attributable to SNPs - Single Nucleotide Polymorphisms (10 and 27 SNP for trnH-psbA and *matK*, respectively). Interestingly, while *matK* and *rpoB* barcodes did not show any sequence polymorphism on *Origanum* samples (*Group III*), sequences of trnH-psbA and *rbcL* marker revealed several genetic differences among samples with the intraspecific genetic diversity higher than the interspecific one. DNA barcoding also worked well in the case of *Group IV*, allowing to recognize all the species belonging to the genus *Salvia*.

Moreover, in order to evaluate potential benefits of multilocus barcodes over a single-marker, we examined multiple combinations of the three plastidial barcode loci that showed appreciable genetic diversity levels in the previous analyses: trnH-psbA, *matK* and *rbcL*. As showed in Table 4.4 *b*, the most suitable combination was the *matK* + trhH-psbA; although, as expected, the observed K2P distance values were very similar to the performance of the best performing locus (i.e. trnH-psbA) for all of the analysed groups.

Since an ideal DNA barcode should be applied to a large number of species with standard PCR conditions (Hebert et al., 2003; Chase et al., 2007), our research showed that all the tested markers generated positive results among the analysed groups with a single step of PCR conditions. However, despite their universality, *rpoB* and *rbcL* barcode regions should be considered unsuitable for the identification of aromatic plants due to the relatively low discrimination power at the species level. On the contrary, based on the patterns of genetic diversity detected in the tested spices groups and given the comparative performances of the analysed markers (see Table 4.4 *a*), we concur with Kress & Erickson (2007) that trnH-psbA is the most suitable marker for DNA barcoding of plants. In particular, the presence of highly conserved PCR priming sites, sites combined with a non-coding region that exhibits high numbers of substitutions, makes the trnH-psbA spacer highly suitable as a plant barcode. The *matK* gene showed easy amplification and alignment in the investigated taxa, although a good level of discrimination based on this marker was observed only in some spices groups (*Group II* and *Group IV*; see Table 4.4 *a*).

Our tests also showed that in some cases spices are characterized by consistent traceability problems. As an example, in the case of mint group (*Group I*), trnH-psbA and *matk* barcode markers clearly distinguished *M. aquatica* L. from the other two *Mentha* species. However, the same markers showed low genetic differences or complete genetic identity between *M. piperita* L. and *M. spicata* L.. It should be acknowledged that the genus *Mentha* is characterized by a large number of species and hybrids (Gobert et al., 2002) and that peppermint (*M. piperita* L.) is a sterile hybrid of *M. aquatica* L. x *M. spicata* L. (Tucker, 1992). This situation leads to large genetic diversity and to several taxonomic problems, further complicated by polyploidy and vegetative propagation phenomena. In particular, these events may have generated the genetic differences detected by trnH-psbA among different analysed mints. The chloroplast uniparental markers used in this study, confirmed that *M. spicata* L. is the maternal parental of *M. piperita* L. because both species showed the same plastidial DNA profile. However, in order to prove the hybrid origin of *M. piperita* L. and identify both parental species, co-dominant markers should be considered (Bruni et al., 2010; Trindade, 2010). A similar situation was also observed within the *Origanum* L. group (*Group III*).

Based on these data we can therefore conclude that a DNA barcoding approach based on plastidial markers cannot be considered as a good traceability tool for mint and oregano groups, because it is not able to distinguish different hybrids and these from their parental lineages. In all the other spices groups, DNA barcoding successfully identified species and even cultivars.

- **Poisonous plants (case study C3)**

  As observed in case study C2, the trnH-psbA barcode region also ranked first in molecular K2P divergence values in the sub-datasets encompassing poisonous plant species. However, the high variability of this DNA spacer did not allow to align properly the sequences related to species of the comprehensive Group I and to define genetic diversity values (see Tables 4.1 and 4.3). Plastidial *matK* showed consistent levels of genetic variability among samples belonging to groups I and IIa; however, the divergence among plant species of Group IIb was very low with this barcode marker (i.e. *Sambucus ebulus* L. and *Sambucus nigra* L. showed the same *matK* sequences). This marker is also able to distinguish edible from poisonous plants of Group III. Nuclear markers showed consistent genetic variability among all the analysed groups; however, it should be underlined that *sqd1* was amplified in a limited number of species and in Group IIa, sequences of this gene showed a large conserved trait. On the contrary, *At103* showed a satisfactory level of amplification and a wide genetic variability. This marker also seems to allow a good distinction among congeneric species, including toxic and edible species included in groups IIIa and IIIb. However, preliminary results from data-mining sequences in GenBank (performed with BLAST) indicated that while *matK* and trnH-psbA were successful at returning a correct match for 24 and 22 samples respectively (data not shown) nuclear markers had a few too many sequences in GenBank (*sqd1* showed three correct matches, while no corresponding sequences for the *At103* were found) to be routinely amplified as useful barcodes for poisonous plants identification. On the other hand, nuclear markers can provide a more reliable assessment of hybridization than uniparentally inherited plastid DNA regions (Chase et al., 2005; Grassi et al., 2006). This is a very important characteristic when dealing with plants among which hybridization events are largely

diffused. For example, *Sambucus* and *Prunus* are two important genera considered in our work showing different natural and commercial hybrids (Liu et al., 2007; Simonovik et al., 2007). Breeding events between poisonous and non-poisonous plants could also result in individuals showing different degrees of toxic substances (Hammouda et al., 1988) and only by the combination of plastidial and nuclear markers it would be possible to successfully distinguish hybrids from parental species.

Concerning the other case studies here analysed, Table 4.3 also shows that although mean interspecific K2P distances were usually higher than the intraspecific comparisons, their ranges often overlap. This situation is clearly illustrated in Figures 3.3 (*a*) and 4.4 (*a*-*d*) which shows the graphs of frequency distribution of intraspecific (yellow bars) and interspecific (red bars) K2P distances for the mitochondrial *coxI* in reference datasets of case studies A1, A3, B1, B2 and C1 (all belonging to metazoans).

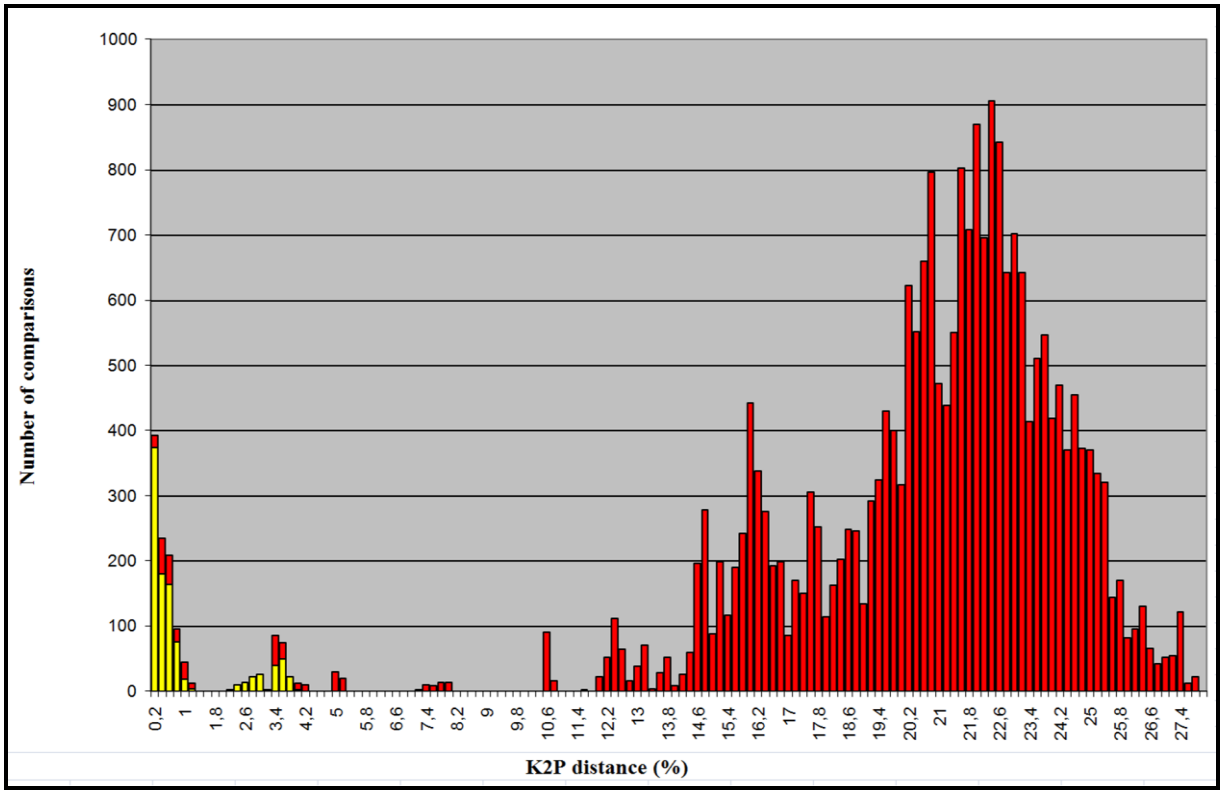


Figure 4.4 *a*: K2P distances distribution graph of case study A1 (Italian bats).

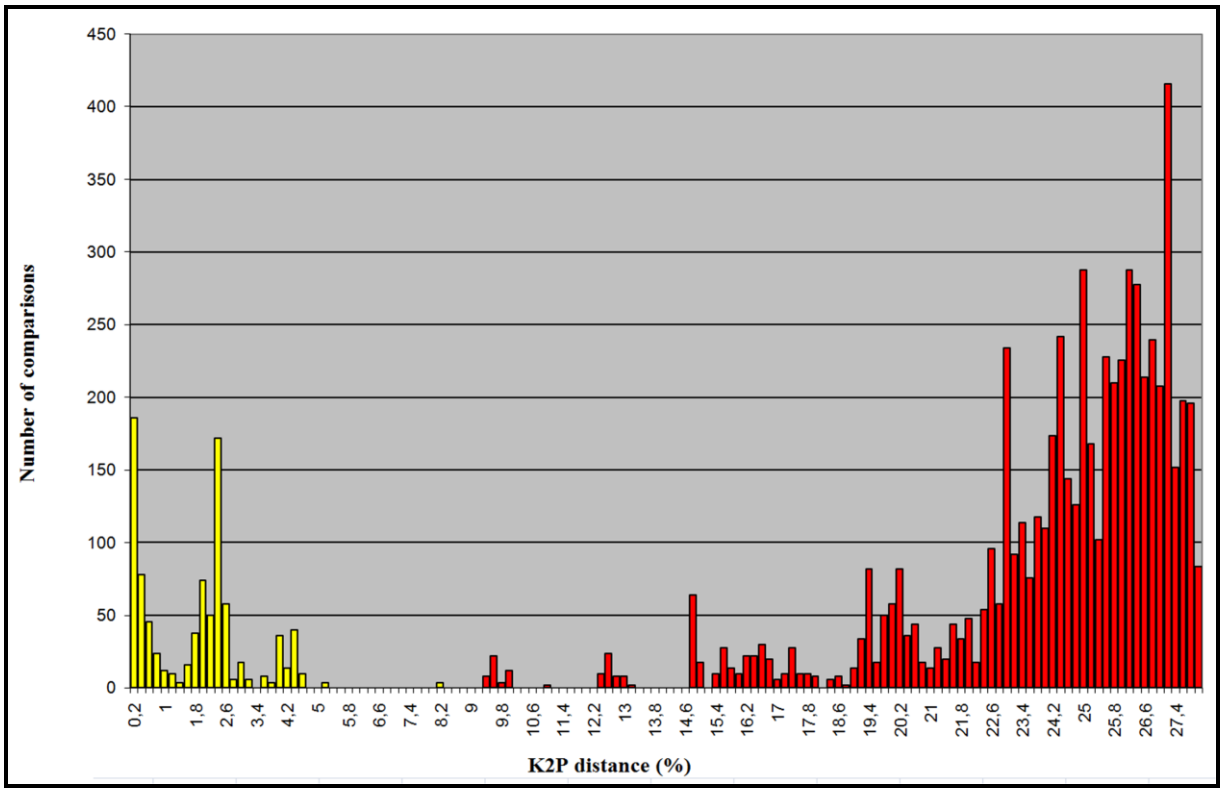


Figure 4.4 *b*: K2P distances distribution graph of case study A3 (terrestrial isopods).

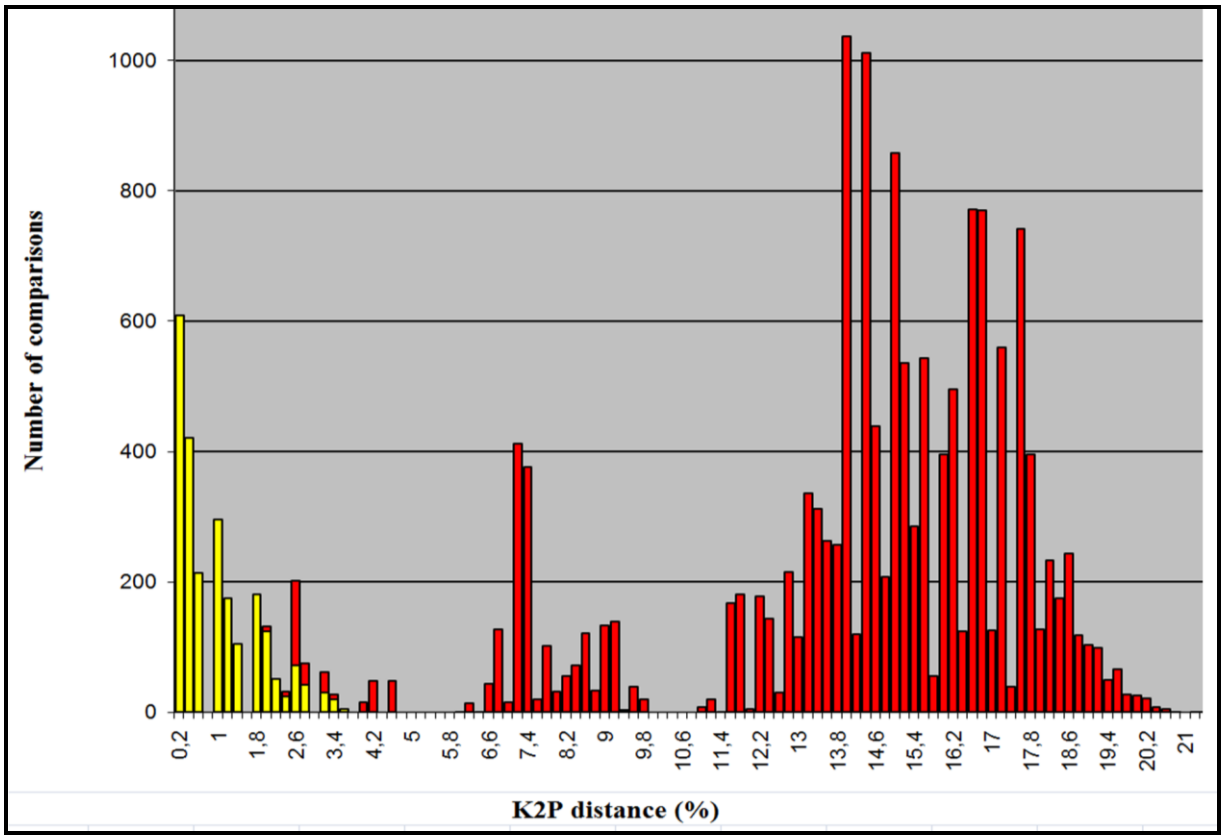


Figure 4.4 *c*: K2P distances distribution graph of case study B2 (taeniids).

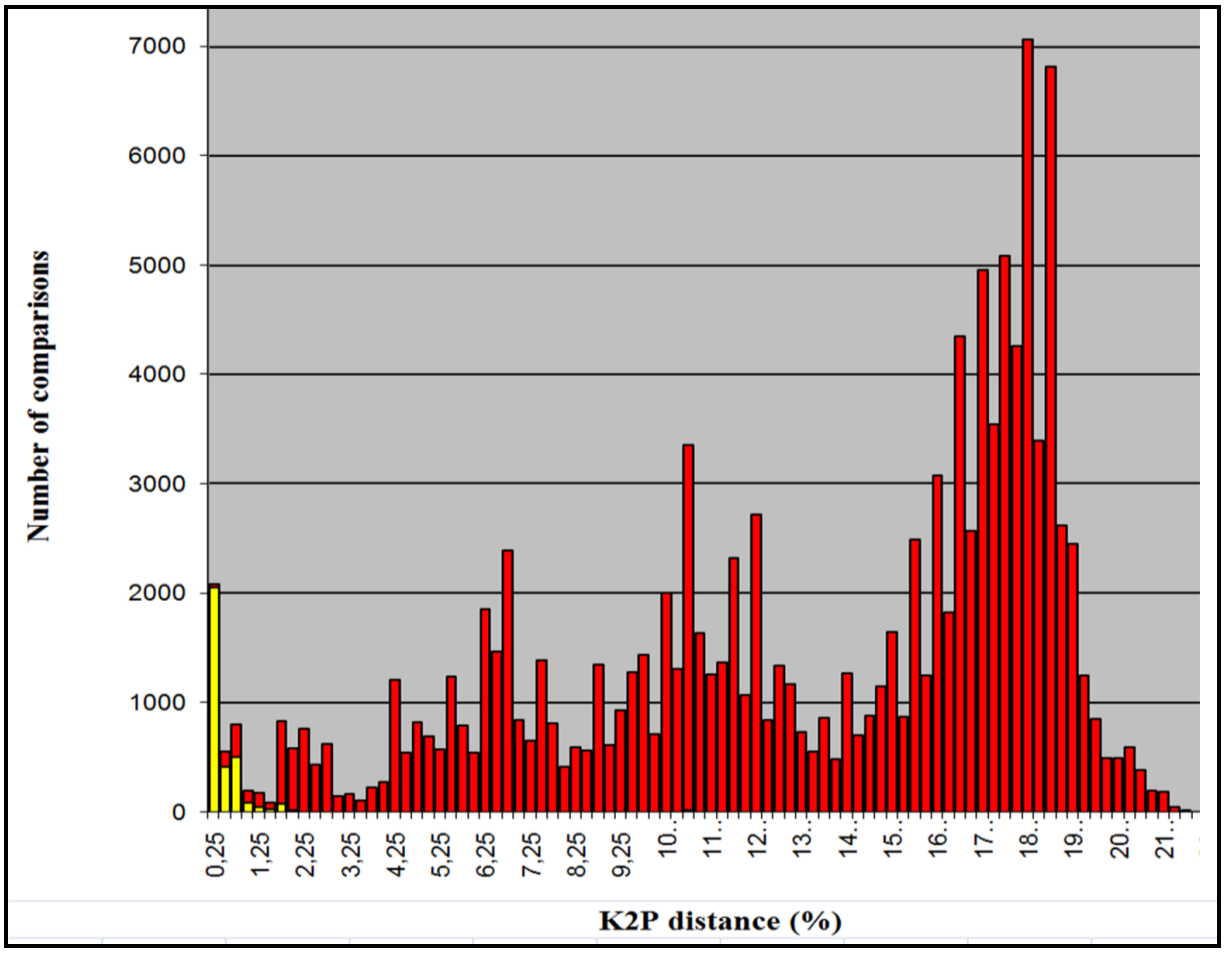


Figure 4.4 *d*: K2P distances distribution graph of case study C1 (shark seafood products).

Except for the case of terrestrial isopods (case study A3) which shows no overlap between the two distribution curves (a perfect 'barcoding gap' occurs; see Figures 1.3 and 4.4 *b*), in all of the others barcode datasets a partial overlap at K2P distance values minor than a certain threshold is observable (e.g. 2.0% in filarioid nematodes; 4.0% in Italian bats; see Figures 3.3 *a* and 4.4 *a* respectively). This pattern was due to the occurrence of false negative and false positive cases. These kinds of mismatches can emerge when identification of species obtained through classical taxonomy (e.g. morphology, ecology, host specificity etc.) is compared with molecular variability of the barcode region (see chapter 3.7).

4.5 OPTIMUM THRESHOLDS ANALYSES

4.5.1 OT VALUES AND IDENTIFICATION MISMATCHES

As stated previously, an Optimum Threshold (OT) value of molecular divergence has been calculated for those molecular barcode datasets characterized by the availability of a certain taxonomical assignment for the most part of samples and a good sampling coverage at both intra and interspecific level (see Material and Methods; 3.7). Table 4.5 provides a list of OT values and related parameters (MCE, number of IOTUs, etc.) for each dataset analysed with this method.

Table 4.5: For each case study listed in the table, Optimum Threshold (OT) and relative Minimum Cumulative Errror (MCE) calculated on reference datasets are provided with also reference to cumulative error composition (False negatives and False positives) and efficacy of OT in recognizing the investigated species. The number of IOTUs and MOTUs calculated on the basis of OT is also reported.

C.S.	DATASET	Barcode region	n	OT (%)	IDENTIFICATION MISMATCHES			Efficacy %	n IOTUs	n MOTUs
					False negatives %	False positives %	MCE %			
A1	ITALIAN BATS	*coxI*	30	4.4	1.08	0	1.08	87	26	28
A3	TERRESTRIAL ISOPODS	*coxI*	26	8.8	0	0	0	100	26	26
B1	FILARIOID NEMATODES	*coxI*	46	4.8	0.62	0	0.62	96	42	44
B2	TAENIIDS	*coxI*	14	3.6	1.45	0	1.45	86	12	13
C1	SHARK SEAFOOD PRODUCTS	*coxI*	110	2.0	1.65	0.5	1.70	72	79	96

As showed in K2P distance graphs (Figures 3.3 *a* and 4.4 *a-d*), using OT as a molecular cutoff, no overlap of intraspecific and interspecific nucleotide divergence occurs at distance values greater than the threshold. This situation allows to exclude the occurrence of Type I errors (false positives) for all the dataset considered except in the case of case study C1 (i.e shark seafood products; see Fig. 4.4 *d*). On the contrary, since some interspecific divergences values are as low as 0%, Type II errors (false negatives) occur at K2P distance values lower than OT in all of the cases considered with the only exception of terrestrial isopods. Indeed, as showed by the cumulative error analysis (Figures 3.3 *b* and Table 4.5) false negatives represent almost the overall amount of cumulative error except in the case of case study A3 in which MCE is 0%.

The highest frequency of false negatives could have different explanations depending on the taxonomic criticisms, biogeography and phylogeny of the group of taxa under examination (see for example DeSalle et al., 2005; Kerr et al., 2009; Petit et al., 2009; Padial et al., 2010). Several biological and evolutionary processes such as past- or still ongoing hybridization events (e.g. Berthier et al., 2006), incomplete lineage sorting (Funk & Omland, 2003) or recent speciation (Ferri et al., 2009) could also lead to the occurrence of false negatives.

As showed in Table 4.5, the values of MCE reached in the calculation of OTs are low (i.e. ranging between 0% and 1.70%) for all the case studies analysed with this kind of approach. These results clearly demonstrate the occurrence of a high strength of coherence between classical taxonomy and *coxI* molecular variability in reference datasets. This condition allows to group specimens belonging to the same species into single IOTUs (see 1.7) for almost all the taxa considered. However, it should be also underlined that some exceptions in the correspondence between species identified with classical taxonomy approaches and IOTUs definition (mediated by OT) were encountered in all reference datasets with the only exclusion of the terrestrial isopods case (case study A3). Indeed, as showed in Table 4.5, in almost all the reference datasets considered, there is not a complete accordance between the number of species identified with classical taxonomy and the related number of IOTUs. Hereafter, some of the causes leading to the emergence of these discrepancies are briefly discussed.

Concerning the Italian bats reference dataset (i.e. case study A1), interspecific K2P distances were almost invariably higher than OT, ranging from a minimum of 4.98% (*Rhinolophus mehelyi – R. euryale*) to a maximum of 28.64% (*Tadarida teniotis – Pipistrellus kuhlii*), even among groups characterized by cryptic species (e.g. 7.63% *Pipistrellus pipistrellus* - *P. pygmaeus*; 14.90% *Plecotus auritus* - *P. macrobullaris*). A schematic representation of these relationships can be observed in the NJ reconstruction of Appendix Figure III.3. False negative identification mismatches were detected for two pairs of closely related congeneric taxa: *i*) *Myotis myotis* and *M. blythii* (mean K2P distance between species: 1.56%, standard deviation: 0.31%, range: 0% – 4.11%) and *ii*) *Eptesicus serotinus* and *E. nilsonii* (mean K2P distance between species: 0.91%, standard deviation: 0.38%, range: 0.72% – 1.09%). These species pairs were included in the same MOTU, therefore generating the false negative cases responsible of the total amount of MCE. In particular, the two

Eptesicus species are characterized by an extreme similarity in their barcode sequences, as already observed for other mitochondrial markers in Western European populations (Mayer et al., 2001; 2007; Artyushin et al., 2009). This pattern of reduced mtDNA variation has been recently explained by Artyushin et al. (2009) as they detected the presence of a mtDNA haplotype group specific to Eastern European populations of *E. serotinus* that is highly divergent from the haplotype group encompassing populations belonging to both *E. serotinus* and *E. nilsonii* from Western Europe. The authors justified this situation by hypothesizing that Russian *E. serotinus* populations preserved their original specific mitochondrial lineage which was lost in others western populations due to complete introgression of mtDNA of *E. nilsonii*.

A similar pattern of reduced interspecific mtDNA genetic variability has been reported for European populations of two sibling species belonging to the mouse-eared bats complex (i.e. *Myotis myotis* and *Myotis blythii*) that occur in sympatry over wide areas of Southern and Central Europe (Mayer et al., 2001, 2007; Ruedi et al., 2001; Berthier et al., 2006). Although slightly different in appearance, phylogenetic analyses showed the frequent sharing of mtDNA haplotypes between *M. myotis* and *M. blythii* in sympatric areas among which Alpine and Northern Italian regions (Castella, 2000; Mayer, 2001). According to these studies, our DNA barcoding approach based on OT also revealed the occurrence of a false negative misidentification for these two taxa (See Appendix Figure III.3). Analyses of both mitochondrial and nuclear markers conducted by Berthier et al. (2006) confirmed that individuals morphologically identified as *M. blythii* shared the same mtDNA haplotypes with *M. myotis* in several European populations whereas haplotypes of *M. blythii* from the original Asian populations were significantly different. Other analyses based on nuclear markers, supported the occurrence of an asymmetric pattern of hybridization (that is still ongoing) between the two species in the areas of sympatry (among which Northern Italian regions), therefore suggesting a progressive replacement in Europe of the mitochondrial genome of *M. blythii* by that of *M. myotis*. Our DNA barcoding results revealed, for the first time, an almost complete genetic similarity in mtDNA between *M. myotis* and *M. blythii* belonging to Central and Southern Italian populations that were previously uncharacterized by other authors. Obviously, the collection of further samples based on a finer geographical coverage and the use of nuclear markers are required to confirm the occurrence of hybridization events also in these localities.

In the reference dataset encompassing barcode sequences of filarioid nematodes (i.e. case study B1) the OT- MCE analysis revealed a really good discrimination power in the identification of species (and related IOTUs) but also the occurrence of false negatives (Figures 3.3 *b* and Appendix Figure III.4). In particular, these are generated by two couples of congeneric species which interspecific divergence is lower than the optimal threshold (i.e. they are included in the same MOTU): *i*) *Onchocerca. volvulus* and *O. ochengi* (mean interspecific K2P distance 1.9%) and *ii*) *Cercopithifilaria bulboidea* and *C. longa* (mean interspecific K2P distance 0.2%). The formers species are identifiable in a relatively easily way based on morphology and host specificity and it has been hypothesized that they could derive from a recent speciation event (Morales-Hojas et al., 2006). This event could explain the scarce resolution power of DNA barcoding for this species pair. Another putative recent speciation has been proposed for two species of *Cercopithifilaria* genus (*C. longa* and *C. bulboidea*), causing the other false negative case detected in our reference dataset (Uni et al., 2001; Agatsuma et al., 2005).

Also in the case of taeniids (i.e. case study B2) our results revealed that the identifications of species based on *coxI* variability is highly consistent with identifications based on classical taxonomic approaches (Table 4.5). However, the application of OT (3.6%) on the reference dataset was biased by the occurrence of a false negative case concerning the species pair *Taenia asiatica* and *T. saginata*. All the *coxI* sequences of these two morphological species diverged for a genetic distance lower than OT (i.e. mean K2P distance between species: 2.5%, s.e: 0.8%, range: 1.9% – 3.2%) and were consequently included in the same MOTU. Despite our results, there are many clear evidences that support the division of *T. asiatica* and *T. saginata* in two distinct species, such as intermediate host specificity, overall molecular variability (Jeon et al., 2006) and morphological characteristics of both adult worms and metacestodes (Eom & Rim, 1993).

Concerning sharks reference dataset (i.e. case study C1), the OT value showed a rather high strength of coherence between morphological and molecular identification of the analysed species even if, some inconsistencies were detected (see Table 4.5). In fact, 31 species (belonging to six different families) out of the 110 considered can be marked as ‘problematic’. In almost all these cases false negatives occurred (e.g. *Squalus albifrons – S. montalbani*; *S. edmunsi – S. hemipinnis*), while only in the case of *Pristiophorus nudipinnis* a

false positive was detected. Most of these situations were already reported in other studies (Ward et al., 2005, 2008) and explained by different causes like species misidentification (e.g. due to the presence of cryptic species or uncertain taxonomy), wrong labeling of the specimens or mistakes during sequences submission to GenBank.

Moreover, concerning the main objectives of this case study (see chapter 2.3), five 'problematic' cases related to species belonging to genus *Mustelus* were detected. In particular, the *coxI*-based OT allowed the identification of *M. mustelus* only, while *M. asterias* was included in a mixed MOTU together with *M. antarticus*, *M. schmitti*, *M. lenticulatus*, and *M. manazo*. Such a situation had been previously discussed by Lopez et al. (2006) in a phylogenetic study based on mitochondrial and nuclear genes. The authors highlighted a strong correlation between genetics and bioecological aspects for several *Mustelus* species, leading to the identification of two species complexes named 'asterias clade' and 'mustelus clade' respectively. In particular, low levels of genetic divergence among species were detected for the two clades and the authors pointed out for hybridization and/or incomplete lineage sorting phenomena as causes. Given these assumptions, the false negative cases identified in our dataset confirm the need for a thorough taxonomical revision of this species complex.

It should be underlined that false negative and false positive cases described above rely on evolutionary dynamic processes that are often difficult to identify (see for example Dorris et al., 2002; Petit et al., 2009). Moreover, these effects are particularly evident in mitochondrial gene trees, and can represent a serious problem for DNA barcoding. As a consequence, the usage of a tree-based method alone for species identification could be dangerous and deceptive (Ferri et al., 2009). A clear demonstration of this criticism can be obtained if the number of molecular entities inferred using OT (IOTUs or MOTUs) is compared with the number of MOTUs identified with a NJ clustering analysis. As an example, in the case of Italian bats (case study A1) on the 30 microchiropteran species included in the reference dataset, 26 IOTUs and 28MOTUs were identified using OT (see Appendix Figure III.3) while 31 MOTUs can be found with a NJ clustering approach (data not shown).

4.5.2 OT AS A MEASURE OF DNA BARCODING PERFORMANCES

As previously stated, a proper data management (from choice of the alignment software to the gap treatment) is fundamental in the context of DNA barcoding analyses (see for instance Britten et al., 2003). The relevance of a DNA barcoding approach based on multiple markers had also been underlined by different authors (e.g. Vences et al., 2005; Lefébure et al., 2006). In order to better investigate these topics, performances of DNA barcoding under different combinations of data handling and with different DNA barcodes (i.e. *coxI* and 12S rDNA) have been evaluated for case study B1 (filarioid nematodes; see also chapter 3.6). This comparison was based on the measure of the strength of correlation between classical taxonomy and molecular variability of each barcode region by evaluating OT and MCE values for each combination (i.e. the lesser the MCE, the better is the performance). The results of this comparative analysis are provided in Table 4.6.

Table 4.6: Minimum cumulative errors (MCE) relative to optimum threshold (OT) for different markers and different data handling procedures. NTA is for not taken into account, TA is for taken into account; PD is for pairwise deletion, CD is for complete deletion.

Barcode region	alignment	distance calculation	gap treatment	OT (%)	MCE (%)
coxI	MUSCLE	TREECON	TA	4.5	0.3
coxI	MUSCLE	TREECON	NTA	4.5	0.3
coxI	MUSCLE	MEGA	PD	3.9	0.3
coxI	MUSCLE	MEGA	CD	4.5	0.3
coxI	Clustal	TREECON	TA	4.5	0.3
coxI	Clustal	TREECON	NTA	4.5	0.3
coxI	Clustal	MEGA	PD	3.9	0.3
coxI	Clustal	MEGA	CD	4.5	0.3
12S rDNA	MUSCLE	TREECON	TA	9.0	0.4
12S rDNA	MUSCLE	TREECON	NTA	6.4	0.3
12S rDNA	MUSCLE	MEGA	PD	6.7	0.3
12S rDNA	MUSCLE	MEGA	CD	5.8	1.9
12S rDNA	Clustal	TREECON	TA	7.2	0.4
12S rDNA	Clustal	TREECON	NTA	5.8	0.4
12S rDNA	Clustal	MEGA	PD	5.8	0.4
12S rDNA	Clustal	MEGA	CD	4.4	1.1

Based on MCE-OT rate comparison, the two DNA barcodes tested as barcodes showed different performances (mean MCE for *coxI* is 0.3% and mean MCE for 12S rDNA is 0.7%).

Also for 12S rDNA dataset, the identification mismatches of DNA barcoding performed with the better data handling are all attributable to false negatives and concern the species pairs *O. volvulus* - *O. ochengi* and *C. bulboidea* - *C. longa* previously detected with *coxI*. Using the marker *coxI*, the eight different combinations of data handling returned the same value of MCE therefore indicating the no-susceptibility of method performances to the tested data handlings. Differently, using 12S rDNA, the eight combinations showed in Table 4.6 the rates of MCE were remarkably different thus indicating that the performance of DNA barcoding with 12S rDNA is very susceptible to different data handling. Interestingly, the two markers show very different manageability: *coxI* has revealed to be less susceptible than 12S rDNA to changes in alignment algorithm software used for distance estimation and gap treatment. The lower manageability observed for the marker 12S rDNA is certainly caused by the presence of numerous indel regions. When 12S rDNA sequences are processed with MUSCLE and MEGA (pairwise deletion), DNA barcoding performs better than using the combination MUSCLE and MEGA (complete deletion) or ClustalX as alignment software. This is a quite relevant observation: the generation of a reliable alignment is a major impediment limiting the use of 12S rDNA gene sequences for DNA barcoding purposes (Chu et al., 2006).

4.5.3 IDENTIFICATION OF UNRECOGNIZED SPECIMENS AND BLIND SAMPLES WITH OT

OT values calculated for each one of the case studies listed in Table 4.5 were used to perform a DNA barcoding analysis on the unidentified or blind specimens collected in the field (see chapter 3.6 and Appendix Tables I.1, I.4, I.5 and I.6). In particular this approach has been used to reach two different purposes *i*) DNA barcoding (i.e. blind specimens were checked to correspond with previously identified species, MOTUs or IOTUs, constituting reference datasets) and *ii*) DNA taxonomy (i.e. which allows to identify potentially new molecular lineages and even species for undetermined samples).

The DNA barcoding approach performed on the comprehensive dataset of case study A1 (i.e. Italian bats) using the BOLD-IDS tool (see chapter 3.8) allowed to uniquely recognize at the species level 14 out of the 41 unidentified vespertilionid samples collected in the field (see Appendix Table II.4). Only samples belonging to the genus *Plecotus* were

uniquely assigned to a given species (i.e. *P. auritus* and *P. macrobullaris*), while several indecision cases between congenerics occurred for the genera *Myotis* (13 cases out of 17) and *Pipistrellus* (14 cases out of 17). On the contrary, as shown in Appendix Table II.4, OT approach allowed to identify all of the 41 blind specimens considered (100%) (see also Appendix Figure III.3). On the whole, six different bat species belonging to the most important cyptic groups in Italy were unequivocally recognized. It is important to underline that any of these blind samples has been included in the two mixed MOTUs interested by the occurrence of false negative cases (see chapter 4.5.1 and Appendix Figure III.3). Moreover, the OT approach allowed the identification of specimens as *M. mystacinus* and *P. pygmaeus* that BOLD-IDS tool was unable to identify (see Appendix Table II.4). This inadequacy is probably due to the occurrence in BOLD database of congeneric reference barcode specimens characterized by taxonomic uncertainties (e.g. the doubtful taxon *M. aurascens*) and /or relying on interspecific genetic distances lower than the 1% K2P threshold fixed by the system.

In a context of food traceability, a similar approach has been performed on the blind specimens of case study C1 (i.e. shark seafood products) in order to assess a species rank to the samples labeled as 'palombo' (i.e. *Mustelus mustelus* and *M. asterias*) collected in Italian markets. DNA barcoding performed with the BOLD-IDS allowed to recognize at the species level 34 blind samples out of 45 (75.6%) belonging to species representative of five shark families (see Appendix Table II.5 and Table 4.7 for more details). Six cases of indecision among species of the genus *Mustelus* and another case of doubtful identification relative to two species of the genus *Squalus* were found. Additionally, four blind samples did not reach any match with IDS (see Appendix Table II.5). The DNA barcoding approach performed with OT method allowed to identify shark species in 37 cases out of 45 (82.2%). Six different species belonging to five families were unequivocally recognized, and, similarly to the BOLD-IDS search, we found the same seven indecision cases (concerning the genera *Mustelus* and *Squalus*). For one blind sample (MIB:ZPL:00058; Appendix Table II.5), no species correspondence for a genetic K2P divergence value lower than OT was found. It is important to underline that the OT approach allowed the identification of three specimens as *M. mustelus* that IDS was unable to identify due to the absence of *coxI* sequences for this species in the BOLD Reference Database (Appendix Table II.5). On the whole, the two approaches provided a coherent identification at the species level in 34 cases out of 45.

Moreover, both methods cannot discriminate in six cases out of 45 among three different species of *Mustelus* (*Mustelus antarticus*, *M. asterias* and *Mustelus lenticulatus*). In summary, both IDS and OT approaches revealed that out of the 45 blind specimens analyzed, only three (6.7%) can be unequivocally assigned to 'palombo' and in particular to the species *M. mustelus* (see Table 4.7 for details).

Table 4.7: Summary of the identification of blind samples at the species level through IDS and OT approach. The prices of species, relative to Mercato Ittico di Milano, are reported.

OT/ IDS	Number of specimens	Family	Italian regulation	Price €/kg
Mustelus mustelus	3	Triakidae	included	7,26
Squalus acanthias	23	Squalidae	included	3,90
Prionace glauca	3	Carcharhinidae	included	2,99
Galeorhinus galeus	1	Triakidae	included	3,00
Alopias superciliosus	1	Alopiidae	not included	-
Isurus oxyrinchus	6	Lamnidae	included	5,50

As showed in the table, 35 cases of species substitution out of 45 (77.8%) collected blind specimens were detected. Surprisingly, among these samples, one has been identified as *Alopias supeciliousus*: a species not included in the Italian fish-trade regulation. Given these results, some consideration is needed about the following economical aspects: 'palombo' is valued by Italian fisheries at 7.26 €/kg (data relative to Mercato ittico Milano, July 2010). As expected, it is usually substituted with less valuable species, for example *S. acanthias* (i.e. more than 50% of substitution cases in our blind specimens sampling are indeed relative to this species) which price is fixed at 3.90 €/kg. The prices discrepancy is evident, and the economic impact on sellers and consumers is clear. DNA barcoding can thus represent a valuable tool for addressing these topics.

Also in the case of spices sub-datasets (i.e. case study C2), some processed samples (e.g. powdered or shredded leaves), have been tested for their correct labeling. In particular, a comparison between their barcode sequences and the reference molecular datasets was performed through a NJ clustering approach on the two best performing markers (i.e. trnH-

psbA and *matK*; data not shown). Only spices belonging to the groups *II*, *IV*, *V* and *VI* (corresponding to basil, sage, thyme and rosemary respectively) can be successfully identified with DNA barcoding. On the contrary, all the tested markers failed in the identification of some mint and oregano species (*Group I* and *Group III* respectively) due to the inadequacy of DNA barcoding in identifying different hybrids and these from their parents (see chapter 4.4). Such a situation impedes an efficient traceability of commercial-processed or edible samples belonging to these taxa.

In the case of filarioid nematodes (case study B1), OT allowed to identify at least five additional MOTUs (i.e. MOTU 1-5; see Appendix Table I.4 and Appendix Figure III.4) within the comprehensive *coxI* dataset. These MOTUs encompass specimens collected from avian and mammal hosts and are characterized by not morphologically identified organisms (potentially non described species) of filarioid nematodes. This is a typical case where molecular analysis can help to discover new species (the so called 'DNA taxonomy'). It must be underlined that DNA taxonomy performed with simple molecular data can only suggest the presence of potential new species, whose real existence must be corroborated by further identification approaches (DeSalle et al., 2005; Padial et al., 2010).

4.6 IDENTIFICATION OF INTRASPECIFIC AND CRYPTIC LINEAGES

In some cases, DNA barcoding data can be used to infer preliminary investigations at the intraspecific level allowing to uncover the presence of different molecular lineages among conspecifics. Such kinds of differences could be related to different factors like geographical provenance or host specificity. Moreover, as previously explained, samples belonging to some species of Italian bats and taeniids (case studies A1 and B2 respectively) were tentatively removed from the respective reference datasets in order to avoid computational biases during OT calculation due to their uncertain taxonomic status (e.g. *Myotis nattereri* and *T. taeniaeformis*). These samples have been however identified with classical taxonomic approaches and their barcode sequences have been used to assemble comprehensive datasets. OT was then used aiming to perform a DNA taxonomy approach on these morphospecies in attempt to shed more light on their taxonomic condition.

An interesting situation dealing with these aspects could be found in the case of Italian bats. As showed in Table 4.8, we detected species that split into two or more distinct lineages with high bootstrap support (see also Appendix Figure III.3). Of the 31 morphologically identified species with multiple specimens barcoded, at least 5 of them, revealed the existence of two or more divergent molecular lineages differing from each other by more than 2% of K2P distance (and even more than OT as well as in the case of *Myotis nattereri*).

Table 4.8: Five bat species from Northern (NI), Central (CI), Southern (SI) Italy and Ireland (IR) with large sequence divergence (K2P > 2%) between lineages at *coxI*. Bootstrap replicates: 500; See also Appendix Figure III.3.

Species	Number of lineages exceeding 2% K2P	Geographical localization of lineages	Mean K2P divergence %	Bootstrap values between lineages
Myotis myotis	2	(NI, CI, SI); (NI)	3.51	100 / 99
Myotis blythii	2	(NI, CI, SI); (NI)	3.40	100 / 99
Myotis nattereri	3	(IR); (NI); (CI, SI)	9.47, 9.34, 5.60	93 / 100 /100
Myotis bechsteinii	2	(SI); (IR, NI, CI)	2.52	99 / 100
Plecotus auritus	3	(NI, CI); (NI); (CI, SI)	2.52, 2.56, 2.62	100 / 98 / 91

As showed in the table, interesting patterns of K2P variability have been found within the same IOTUs for the species *Plecotus auritus* and *Myotis bechesteinii* (see also Appendix Figure III.3), suggesting the occurrence of geographic lineages some of which exclusive of certain Italian group localities (e.g. *M. bechsteinii* samples from Southern Italy), while others occurred sympatrically in the same geographic regions (e.g. *P. auritus* samples from Northern or Central Italy). Moreover, barcode sequences of morphologically identified vespertilionids *Myotis myotis* and *M. blythii* (grouped into the same MOTU according to OT) could be divided into two molecular clusters separated for an averaging K2P distance lower than OT (i.e. 3.51%, S.E. 0.77%) and containing sequences belonging to both taxa. One group is exclusive of Northern Italian regions while the other includes individuals belonging to all of the Italian geographic areas considered (i.e. Northern, Central and Southern Italy). Noteworthy, individuals of *M. myotis* and *M. blythii* of both lineages occur sympatrically in Northern Italy and even in the same colonies (See Appendix Table I.1 and Appendix Figure III.3). A similar pattern of genetic variability (i.e. the occurrence in *M. myotis* of divergent molecular mtDNA lineages within the Italian Peninsula) has been recently described by Ruedi

et al. (2008). In particular, through the analysis of a portion of the mitochondrial *d-loop*, they observed that mouse-eared bat populations from Apennines are characterized by a complex mixture of several endemic lineages, which evolved *in situ* with others resulting from a recent colonization. As stated by Ruedi et al. (2008), it seems plausible to assume that similar processes occurred for several other bats species as a consequence of the climatic fluctuations of Pleistocene, therefore influencing the evolution of the divergent coxI molecular lineages.

Finally, high levels of variability have been detected in *Myotis nattereri*, characterized by the occurrence of three distinct *coxI* lineages (corresponding to three different MOTUs) isolated by mean K2P distance values higher than OT (see Table 4.8 and Appendix Figure III.3). In particular, at least for specimens collected in the Italian Peninsula they showed a highly divergent pattern of genetic structure that did not correspond to the currently recognized species based on morphology. Of these two lineages, one is exclusive of Northern Italy, while the other encompasses individuals from both Central and Southern Italian localities. The remaining lineage refers to a single barcode sequence belonging to an Irish individual retrieved from GenBank (a.n. GU270561). Aware of the fact that the removal from reference dataset of *coxI* sequences belonging to this taxon could be interpreted as a sort of *ad hoc* strategy, these results and the clear evidences provided by recently published studies conducted on European *M. nattereri* populations supported our choice (e.g. García-Mudarra et al., 2009). Moreover, if *coxI* sequences of Italian *M. nattereri* were considered, we would have obtained an OT value greater than the one calculated with the current reference dataset (i.e.: 8.6%, versus the current 4.4%) and a higher MCE value (3.0% versus the current 1.08%) causing a lower support for the identification of IOTUs, MOTUs as well as a significant decreasing of DNA barcoding efficacy. A possible explanation to the occurrence of this molecular scenario is the presence of undescribed cryptic lineages related to this taxon such as the case of *Myotis escalerai*: a recently discovered species morphologically similar to *M. nattereri* that was found to strongly diverge at the both the molecular and ecological level (Evin, 2009). Unfortunately, morphological and ecological differences between both Italian mitochondrial lineages could not be further investigated, because sampled bats were immediately released after sampling and standard measurements (further analyses are ongoing).

Concerning case study B2, all of the cestode samples found in wild and domestic cats (see Appendix Table I.5) have been unambiguously identified by morphologic survey and measurements as *Taenia taeniaeformis*, also according to the descriptions provided by Iwaki et al. (1994) and Azuma et al. (1995). The application of OT value on the comprehensive dataset revealed the presence of at least 18 different MOTUs out of the 16 morphological identified species considered (Table 4.5). This discrepancy is due to the occurrence of false positive cases belonging to the morphospecies *T. polyacantha* and *T. taeniaeformis* provisionally removed from the reference dataset due to the well-documented uncertainties on their taxonomic status provided by Lavikainen et al. (2008). Although *T. polyacantha* is not directly pertaining with the host species we analyzed (*Felis silvestris* ssp.), the case of *T. taeniaeformis* deserves to be discussed more in depth.

In particular, on the whole panel of 77 *T. taeniaeformis coxI* sequences (belonging to GenBank, Okamoto et al., 1995 and our sampling) three well-defined MOTUs here named as Lineage 1, Lineage 2 and Lineage 3 have been detected (Appendix Table I.5). Lineage 1 encompasses 61 barcode sequences exclusive of Italian wild and domestic cat parasites, and it represents the 79% of the comprehensive panel of *T. taeniaeformis* sequences included in the molecular dataset. Lineage 2 represents the 5% of *T. taeniaeformis* dataset and encompasses four *coxI* sequences: two collected from GenBank (a.n. EU861478 and EU544596), one found in Okamoto et al. (1995) (*T. taeniaeformis* isolate ACR) and one found in a wildcat individual collected in Italy (MIB:zpl:01385; a.n. FN547850). Finally, Lineage 3 represents 16% of the comprehensive panel of *T. taeniaeformis* molecular data and encompasses 12 *coxI* sequences: four collected from GenBank (a.n. FJ939135, EF090612, AB221484 and EU544597) and 8 found in Okamoto et al. (1995) (*T. taeniaeformis* isolates: Nop, BMM, Cha, KRN, Mar, Tom, KaRN, KaAA. The mean K2P distances within each *T. taeniaeformis* molecular lineage were substantially low (0.3%, 1% and 1.4% for Lineage 1, Lineage 2 and Lineage 3 respectively), while the mean K2P distances among the three lineages were higher than OT: 5.7% (Lineage 1 – Lineage 2); 10.4% (Lineage 1 – Lineage 3); 9.7% (Lineage 2 – Lineage 3). These values are in line (or greater) with the interspecific K2P values found between other species of the genus *Taenia* considered in this case study (data not shown). No significant correlations were found in host specificity between each molecular *T. taeniaeformis* lineage and the host subspecies in which it was recovered. Despite its relative high abundance in the dataset and its syntopy with Lineage 2, Lineage 1 was so far

uncharacterized. On the contrary, Lineage 2 and 3 have been already described both morphologically and molecularly by other authors (Iwaki et al., 1994; Azuma et al., 1995; Okamoto et al., 1995; Lavikainen et al., 2008). However, despite the molecular variability of *coxI* sequences in *T. taeniaeformis* has been previously studied by other authors (e.g. Lavikainen et al., 2008) they did not treat their data in the light of a strictly standardized DNA barcoding approach. In this context, our DNA barcoding analyses strongly support the occurrence of highly divergent molecular lineages and also provide clear evidences for the existence of a third potential cryptic species (i.e. Lineage 1). Lineage I is actually undescribed and seems to be exclusive of the Italian Peninsula.

Further investigations at a fine-scale are required to clarify its status and distribution in the peninsula. In fact, it is important to underline that wildcat populations also occur in Northeastern Italian regions and Sicily that were not included in our preliminary sampling. Moreover, additional analyses could confirm whether or not the single individual morphologically recognized as *T. taeniaeformis* collected in Italy and belonging to Lineage 2 has to be considered as a rare molecular lineage or the result of an accidental introduction. In conclusion, this pattern of molecular variability provides clear evidences for the presence of at least three putative Unconfirmed Candidate Species (UCS *sensu* Padial et al., 2010) within *T. taeniaeformis*.

4.7 ADDITIONAL ANALYSES ON SOME DIVERGENT LINEAGES

The application of a DNA barcoding approach on the case studies here analysed revealed interesting patterns of molecular variability that could be related to the occurrence of isolated geographical lineages, subspecies and even new cryptic species (here named as UCS *sensu* Padial et al., 2010). Two cases have been further investigated with the analysis of additional molecular markers and data management approaches in order to shed more light on these new putative taxonomic entities.

4.7.1 THE CASE OF MYOTIS NATTERERI *ITALIAN LINEAGES*

The first case concerns the surprisingly high levels of molecular variability detected among samples collected in the Italian Peninsula and belonging to individuals morphologically identified as *Myotis nattereri* (see chapter 4.6). In particular, the genetic structure of Italian *M. nattereri* has been here preliminary investigated with the analysis of two additionally mitochondrial molecular markers. As stated in chapter 3, we assembled two additional molecular dataset encompassing aligned sequences of *cyt b* (21 sequences, 768bp long) and *ND1* (17 sequences, 605bp long); see Appendix Table II.3. Both mtDNA NJ trees (Figure 4.5, *ND1* and Appendix Figure III.1, *cyt b*) provide clear evidences for the occurrence of five major lineages: one localized in Northern Africa (named as *Myotis* sp. B) and four European (named as *Myotis nattereri*, *Myotis escalerai*, *Myotis* sp. A and *Myotis* sp. C), with at least 7.8% and 7.6% K2P distances between all groups for *ND1* and *cyt b* respectively.

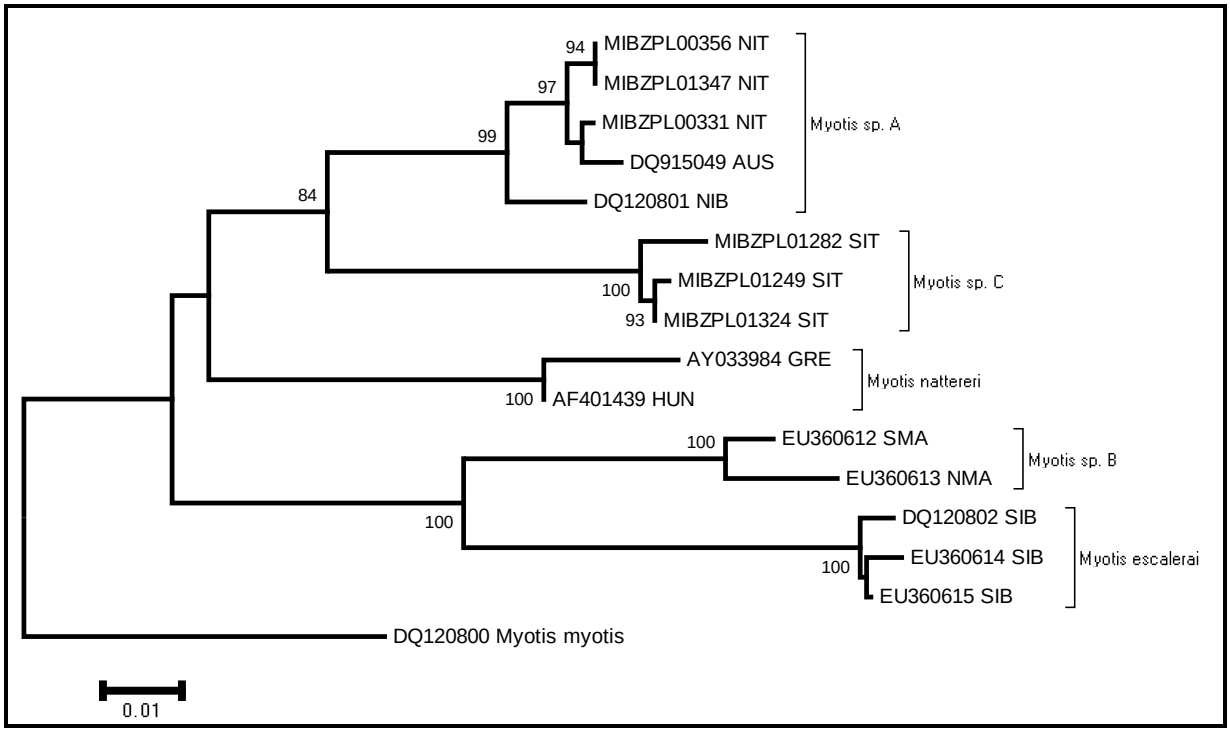


Figure 4.5: NJ tree based on the analysis of the *Myotis nattereri ND1* dataset using K2P correction, 1000 boostrap replicates and a sequence of *Myotis myotis* as outgroup. Further details about the samples included in the dataset can be retrieved from Appendix Table II.3. Lineages has been identify within square brackets. Bootstrap values lower than 70 not shown.

Samples belonging to Northern Italian localities were included in the lineage that also encompasses haplotypes of *M. nattereri* sampled in Northern Iberia and Austria (*Myotis* sp. A). Surprisingly, both *ND1* and *cyt b* NJ trees, revealed that samples collected in Central and Southern Italy are grouped into a new and previously undescribed lineage (*Myotis* sp. C). Both Italian lineages are consistently different from samples belonging to the *Myotis nattereri* lineage. They have also been found to significantly differ from the lineage encompassing North African samples and to the recent separated taxon *Mytois escalerai* (*Myotis* sp. B and *Myotis escalerai* lineages respectively).

The high values of molecular variability detected among lineages, if compared with data from previous genetic studies on mammals, and more peculiarly on microchiropterans, are in line with the thresholds that generally indicate the occurrence of different species (Mayer et al. 2007, García-Mudarra et al. 2009), therefore suggesting the occurrence of at least two distinct UCS entities morphologically similar to the nominate *M. nattereri* in the Italian Peninsula. These preliminary results support the conclusions of García-Mudarra et al. (2009) that those individuals belonging to the lineage *Myotis* sp A should be ranked at the species level as well as the new lineage, *Myotis* sp C, first detected by our DNA barcoding approach. It is important to underline that in order to prove a full species rank for both taxa, an intensive sampling in the currently unknown contact zone in the Italian Peninsula is essential for investigating the question of whether or not both molecular lineages occur in sympatry and interbreed. Moreover, an integrative taxonomic approach based on accurate morphological and ecological data, the use of fast-evolving and nuclear markers would allow a better characterization of taxonomic status of the *M. nattereri* complex.

4.7.2 THE CASE OF ITALIAN AND CHINESE PARROTBILLS

Additional investigations have been conducted on parrotbills samples of case study A2 in order to investigate the nature of the two MOTUs identified with the NJ clustering analysis performed on the *coxI* dataset (see chapter 4.4 and Figure 4.3).

Genetic coincidence of Paradoxornis webbianus *and* P. alphonsianus *and relationships between molecularly-identified lineages*

The haplotype network reconstruction (Figure 4.6; A) revealed two divergent groups of haplotypes (Figure 4.6; A, I and II). To link these groups, a minimum of twenty-two mutational steps is required. Interestingly, the two identified lineages do not correspond neither to the morphotypes nor to any of the subspecific attributions.

The first haplogroup (Figure 4.6; A, I) is made of 12 haplotypes of both taxa with individuals collected from all the five sampling localities, while the second group (Figure 4.6; A, II) is made of five haplotypes, all belonging to the vinous-throated taxon and coming from two Chinese localities (Tianjin and Shanghai); haplotype E17 is found in Shanghai, while the other haplotypes are found in Tianjin. The first haplogroup (Figure 4.6; A, I) shows higher genetic variability, but no clear pattern of geographical substructure can be identified.

In Henan, Sichuan and Varese (Italy) we found four, five and one haplotypes respectively, all belonging to the first haplogroup (Figure 4.6; A, I), whereas in Tianjin and in Shanghai we found respectively six and two haplotypes, belonging to both haplogroups (Figure 4.6; A). All individuals of the Italian population share the same haplotype (E01), even if they were morphologically assigned to both different taxa. Similarly, haplotype E02 is shared by both taxa and was found at two rather distant localities.

A similar pattern can be also inferred from the analysis of only *cyt b* sequences (Appendix Figure III.2). Indeed, in this reconstruction, haplotype A11 is shared by individuals of both taxa from five different localities out of the seven considered. This observation, and the fact that haplotypes E02 (Figure 4.6; A, I) and A11 (Appendix Figure III.2; A, I) occupy an internal position within the network, suggests that these could be rather widespread and ancestral haplotypes (Posada and Crandall, 2001).

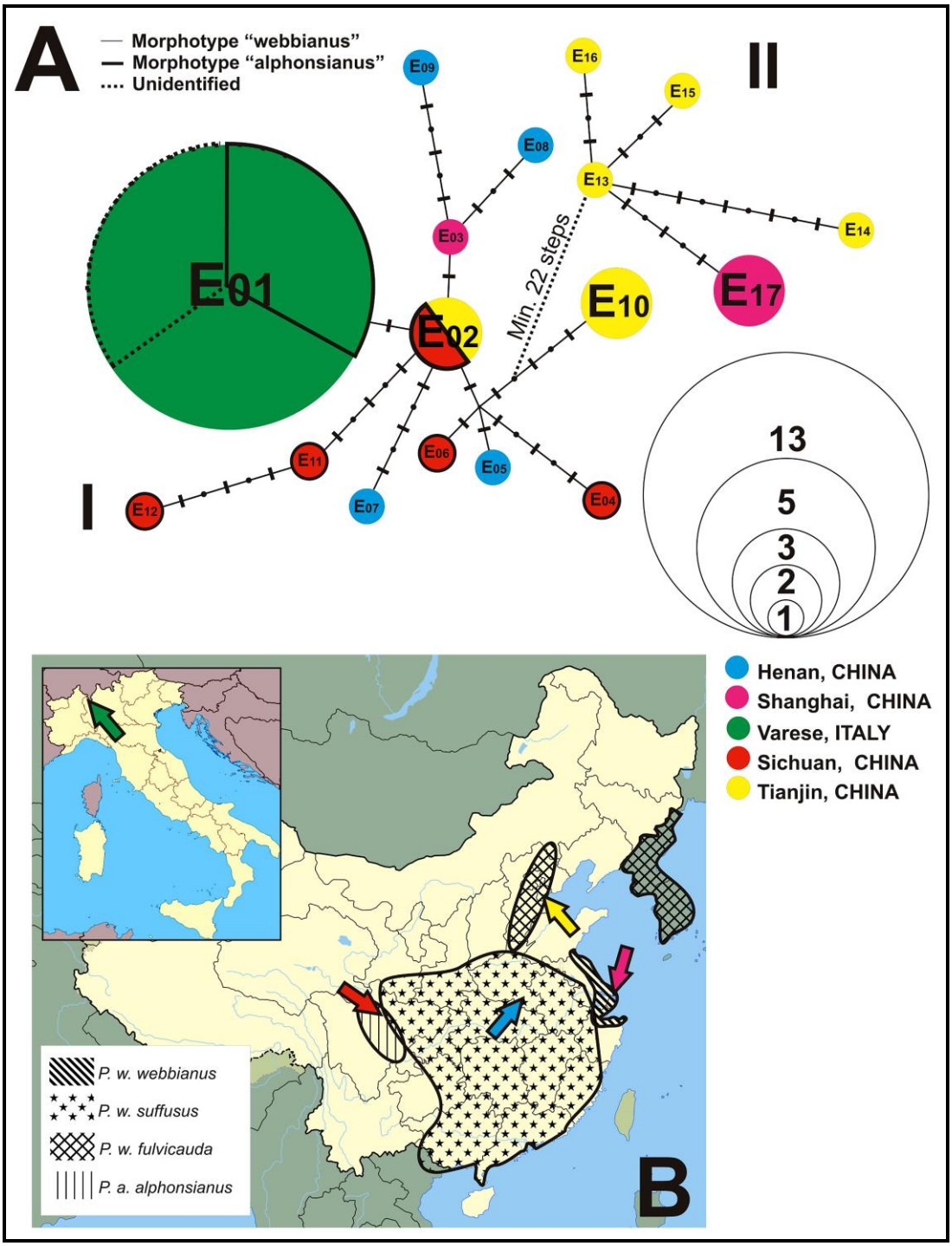


Figure 4.6: Haplotype network reconstruction based on a fragment of 2463 bp resulting from the concatenation of four mitochondrial gene fragments: *cyt* b, *coxI*, 16S rDNA and 12S rDNA (A). Size of circles is proportional to the number of individuals sharing a given haplotype. Black bars represent single nucleotide substitutions and each haplotype pie shows the percentage of sequences belonging either to the different sampling localities (indicated with different colours). The outline pattern of each pie indicates the presence and percentage of each of the two investigated taxa, or the occurrence of unassigned individuals. A map showing the geographical distribution of Paradoxornis subspecies used in this study [following Robson (2007) and Penhallurick and Robson (2009)] is also provided (B); coloured arrows indicate the sampling localities reported in Appendix Table I.2.

The ΠΙs of Nei (1987) (Table 4.9) of the vinous-throated taxon (*P.webbianus*) are higher than the ΠΙs of the ashy-throated taxon (*P. alphonsianus*) and the two molecularly-identified lineages showed similar ΠΙs values. Due to the occurrence of individuals of both networks, the ΠΙs in Shanghai and Tianjin are higher than in other localities.

Table 4.9: Nucleotide diversity values in vinous-throated and ashy-throated parrotbills and in the two lineages identified in this study (Lineage I and Lineage II). Π = Nei (1987) nucleotide diversity; sd (Π) = standard deviation of the Nei's nucleotide diversity. Values are given separately for the different taxa, molecular lineages and sampling localities.

Morphotype		**Shanghai**	**Henan**	**Tianjin**	**Sichuan**	**Italy**	**Semi-tot**	**TOTAL**
webbianus	**Π**	0.01056	0.00217	0.00609	_	0	0.00652	0.00586 Π
	sd(Π)	0.00528	0.00042	0.00148	_	0	0.00073	
alphonsianus	**Π**	_	_	_	0.00211	0	0.00195	0.00088 sd(Π)
	sd(Π)	_	_	_	0.00048	0	0.00042	

Lineage		**Shanghai**	**Henan**	**Tianjin**	**Sichuan**	**Italy**	**Semi-tot**	**TOTAL**
Lineage I	**Π**	**0**	0.00217	0.00162	0.00211	**0**	0.00207	0.00586 Π
	sd(Π)	**0**	0.00042	0.00081	0.00048	**0**	0.00030	
Lineage II	**Π**	**0**	_	0.00183	_	_	0.00195	0.00088 sd(Π)
	sd(Π)	**0**	_	0.00053	_	_	0.00045	

The p-distance values among the four species of *Paradoxornis* currently considered plus the two molecular lineages identified in this study ranges between: 0.02 % (Lineage I – Lineage II) and 7.29 % (*P. gularis – P. nipalensis*) in 12s rDNA; 0.05 % (Lineage I – Lineage II) and 5.53 % (*P. gularis* – Lineage II) in 16s rDNA; 1.94 % (Lineage I – Lineage II) and 15.12% (Lineage II – *P. nipalensis*) in c*yt b*; and 1.41% (Lineage I – Lineage II) and 12.48% (*P. guttaticollis* – Lineage I) in *coxI*. All p-distance values for each gene and for each comparison are provided in Appendix Table II.6. The p-distance comparisons showed that: *i*) the genetic distances between the molecularly-identified lineages are higher than the genetic distances between the morphotypes (Appendix Table II.6, a vs. Appendix Table II.6, b); *ii*) the extent of the genetic diversity between the two lineages is lower than the genetic diversity between different *Paraxodornis* species.

On the whole, the supplementary mitochondrial data here provided confirm an almost absence of genetic differentiation between the two morphotypes and reveal the occurrence of two partially sympatric lineages.

Genetic diversity between and within Chinese and the naturalized Italian populations

All naturalized Italian individuals from the Palude Brabbia Natural Reserve (Varese, Italy) previously assigned to both taxa have been here assigned to the Lineage I (Figure 4.6; A, I) showing a high degree of genetic homogeneity. In the People's Republic of China almost all haplotypes have been found in only one or two individuals at each locality, suggesting that the genetic diversity within these taxa might be rather high. On the contrary, the parrotbills accidentally introduced in Italy are genetically identical at the four analyzed loci. Therefore, it is reasonable to confirm the occurrence of a strong founder effect. Moreover, considering the relatively high genetic divergence (in term of different identified haplotypes) found at each locality in their original range of distribution, we presumed that the Italian population raised from an original stock, most likely derived from one single locality (and/or a few individuals), instead that from different ones, as was previously suggested by Baratelli et al. (2008).

Given these assumptions, we can hypothesize that the few founder individuals of the Italian population were caught in a locality where the two taxa occur in sympatry. Due to the limited genetic differentiation between the Italian parrotbills and some individuals from Tianjin and Sichuan (1 substitution; see Figure 4.6; A), and due to the reported sympatry of both morphotypes in Sichuan (Robson, 2007; Penhallurick & Robson, 2009), we suggest the likely provenance of the individuals that founded the Italian population from this province, although new distributional data might become available calling this hypothesis into question.

Taxonomic conclusions

The results here provided support the synonymisation of *P. alphonsianus* with *P. webbianus* as previously hypothesized with the DNA barcoding approach based on *coxI* sequences only. In spite of the existence of two morphologically identifiable taxa, four mitochondrial markers consistently identify a low genetic distance between them suggesting to associate the morphological differences (i.e. plumage colour) with clinal variations through their distributional area. An integrative taxonomic approach based on accurate morphological

and ecological data and including the use of fast-evolving and nuclear markers, a wider geographical coverage and a sampling activity performed in both the breeding and the wintering season would allow a better characterization of the relationships among these taxa. Indeed, alternative scenarios that might be supported in the future are: (1) the maintenance of the two morphotypes as two separate species advocating a possible event of incomplete lineage sorting, even if evidences of hybridizations and records of intermediate morphs (Robson, 2007) call this hypothesis into question; or (2) the subspecific validation of the two identified evolutionary lineages.

4.8 ENVIRONMENTAL DNA BARCODING

In the context of case study D1 (i.e meiofaunal moss biodiversity), moss samples have been selected as environmental matrix in order to evaluate the performances of a massive characterization of biodiversity with DNA barcoding (see also chapter 2.4). The Baermann extraction (chapter 3.1) provided a successful isolation of the most part of the hydrobiont fauna biodiversity usually reported for moss. In particular, a preliminary microscopic survey on the total extract revealed the presence of the typical taxa inhabiting this kind of soil matrix. As expected, the most part of the individuals collected have been assigned to the phyla Rotifera, Nematoda and Tardigrada as previously reported in Ramazzotti (1958); Zullini (1970); Barbuto & Zullini (2006) (see also Figure 2.9). A precise taxonomic assignment for rotifers and tardigrads has not been performed due to the lack of morphologists experienced in these groups. However, in the case of nematodes, at least eight putative morpho-groups have been identified on the basis of some 'external diagnostic characters' allowing to assign some individuals to the order Monhysterida and to the genera *Plectus* and *Tobrilus*. Unfortunately, it was not possible to create a library of *coxI* reference sequences for these organisms (especially concerning the three phyla cited above). Apart from the shortcomings emerged during the roughly discrimination based on morphotypes, other problems (too expensive in terms of time and resources to be easily overcome) such as the achievement of a reliable DNA extraction and *coxI* sequencing from single individuals were encountered.

As briefly discussed in chapter 4.1, the cloning analysis performed on *coxI* amplicons belonging to the four aliquots (and relative DNA extracts) of Baermann extraction was successful in providing a great number of barcode sequences for the environmental sample investigated (see also Figure 4.1). It is important to remark that out of the 768 clones selected for DNA sequencing, only 526 *coxI* sequences (about 70 %) reached a sufficient length and quality to be successfully included in the comprehensive alignment. Moreover, the 'interference' caused by inadvertent amplification of *coxI*-like sequences or bacterial genomic regions also contributed to reduce the total number of available sequences to be included in the comprehensive molecular dataset (see chapter 4.2).

A preliminary analysis of haplotype composition on this cloning library conducted with the DNAcollapser tool in FaBOX (http://www.birc.au.dk/~biopv/php/fabox/dnacollapser.php#)

resulted as follow. Number of haplotypes: 326; mean number of *coxI* sequences per haplotype: 1.6; range 1-47; standard deviation: 3.1. On the whole, these data and the high value of overall mean diversity showed in Table 4.3 indicates a good performance of the universal primer pair used (i.e. LCO1490 – HCO2198) that are not biased by some sample/s. . Indeed, concerning the total alignment, our amplification approach seems to be not interested by the preferential amplification of some organisms or taxa (if so, we would have obtained a reduced number of haplotypes in spite of a high number of sequences per haplotype). However, more detailed analyses on MOTUs composition and *in silico* testing of primer specificity with reference barcode sequences belonging to the different taxa encompassed in the moss sample are required to better address this issue.

A preliminary data-mining on the 526 selected *coxI* clone sequences performed with BLAST did not return any complete match (i.e. 100% of maximum similarity) or maximum similarity score higher than 95% (i.e. the interval of similarity usually found in animals between conspecifics for *coxI*; see chapter 3.3) with homologous records deposited in GenBank. Such a situation highlights the substantial lack of reference sequences for the most part of the meiofaunal community biodiversity. Apart from the still unknown biodiversity for these taxa (Giller 1996; André et al., 2001; 2002; Blaxter, 2003; see also chapter 1.1), this condition is mainly due to the fact that almost all the studies published so far on the molecular taxonomy (or phylogeny) of rotifers, tardigrads and nematodes have not been based on *coxI* but considered the mitochondrial 16S rDNA or the nuclear 18S rDNA and 28S rDNA regions (see for example Blaxter et al., 2004; 2005; Porazinksa et al., 2009). As an example, it should be considered that in the case of nematodes, only 1,300 *coxI* sequences are available in GenBank against the 8,000 relative to 18S rDNA. Moreover, the most part of the nematodes *coxI* sequences refer to parasitic nematodes (e.g. belonging to the order Spirurida) and very few sequences are available for free-living nematodes which are the most abundant organisms in soil.

Data-mining performed with BLAST on the clone library revealed high similarity results (maximum similarity >90%) for 36 *coxI* sequences (see Appendix Table II.7). These entries have been included in the molecular dataset and were used as references in order to reasonably assign each clone sequence at the phylum level.

The NJ reconstruction based on the resulting molecular dataset (i.e. clone sequences + reference GenBank sequences; see Figure 4.7) shows that the identification approach based on BLAST similarity is coherent with the occurrence of monophyletic clades corresponding to the three main hydrobiont taxa inhabiting moss (i.e. rotifers, tardigrads and nematodes).

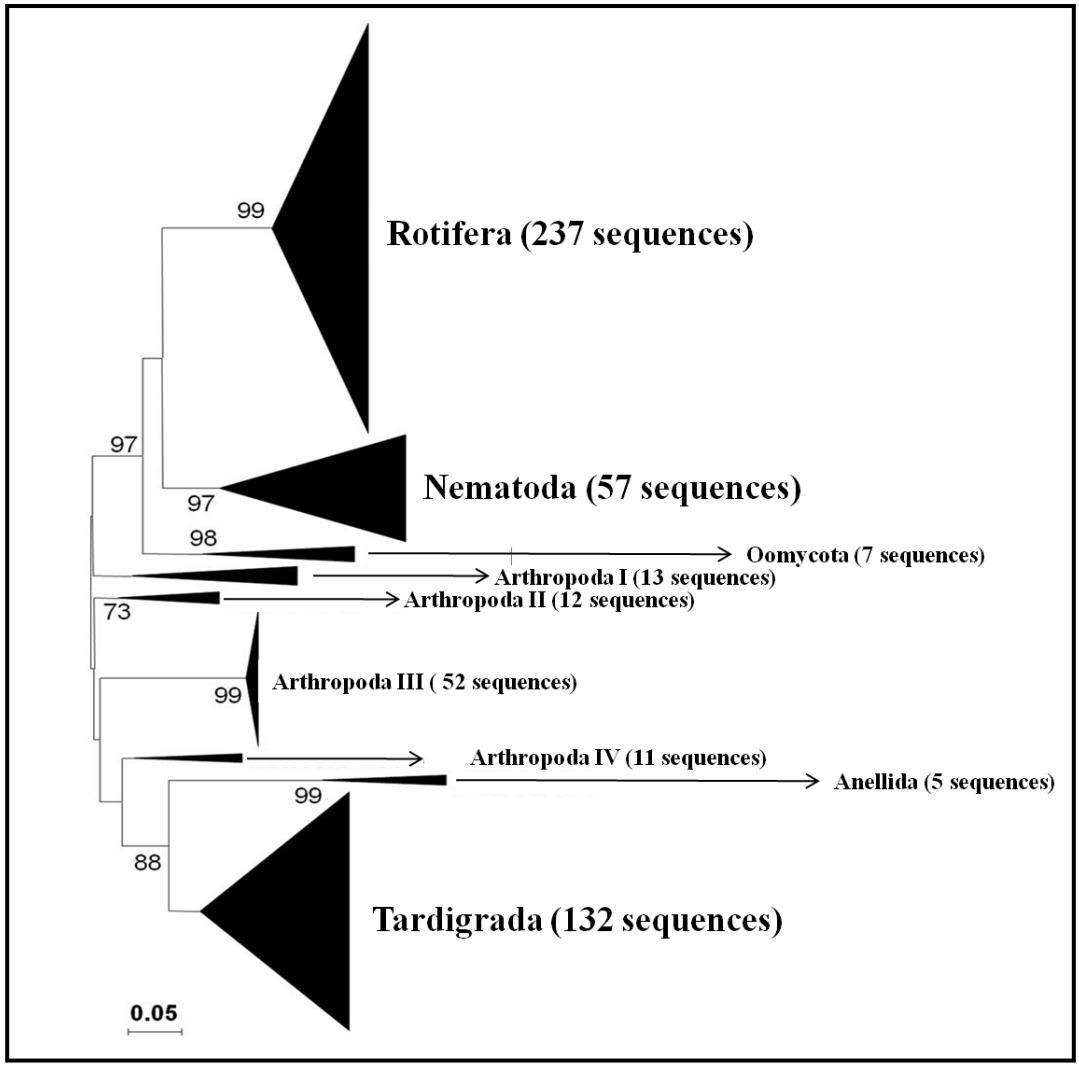


Figure 4.7: Condensed NJ tree representing the main phyla belonging to the hydrobiont fauna of moss (case study D1) identified on the basis of BLAST similarity matches of *coxI* clone sequences with GenBank reference entries. The number of clone sequences found for each phylum is shown within brackets.

Moreover, as showed in the NJ tree, groups of *coxI* sequences belonging to the phyla Arthropoda, Anellida and Oomycota were found in the comprehensive clone library of moss sample. Although these three phyla have not been directly observed during the microscopical

survey conducted on the Baermann extract, their occurrence in the related molecular dataset is not surprising due to the fact that the combination of PCR and cloning steps can results in a increased sensitivity therefore allowing the amplification of traces of other organisms (e.g. eggs, faeces and parts of individuals). Interestingly, all of the barcode sequences belonging to Nematoda, Rotifera, Tardigrada, Oomycota and Anellida were grouped in single clusters while *coxI* sequences of Arthropoda result grouped in at least four different clusters. This is not surprising, since insects as well as many arthropod taxa related to them (e.g. Collembola, Protura, Diplura that are nowadays considered independent from insects) are present in the soil and moss. Despite the fact that the meaning of the NJ reconstruction here provided would not be intended in a strict phylogenetic context (see also chapter 1.9), this pattern could be related to the paraphyletic relationships discovered within the phylum Arthropoda (e.g. Lake, 1990; Turbeville et al., 1991; Carapelli et al., 2007).

Although this kind of experimental system cannot be easily interpreted under a quantitative approach, due to some theoretical and practical biases (see below), a provisional survey on the relative abundances of the taxa represented by the *coxI* sequences is provided (see Table 4.10 and Figure 4.8).

Table 4.10: taxonomic assignment at phylum level of the *coxI* sequences obtained from the cloning analysis performed on the four aliquots of Baermann extraction. The numbers of clones assigned to each phylum for each aliquot and for the comprehensive sample are provided.

BAERMANN ALIQUOT	phylum						n° SEQ
	ROTIFERA	TARDIGRADA	NEMATODA	ARTHROPODA	OOMYCOTA	ANELLIDA	
1	86	18	13	3	7	1	128
2	92	28	13	3	_	_	136
3	43	9	13	82	_	4	151
4	16	77	18	_	_	_	111
n° SEQ	**237**	**132**	**57**	**88**	**7**	**5**	**526**

As showed in Figure 4.8 and Table 4.10, the most part of the *coxI* sequences obtained (81%) belong to the three main phyla typical of the hydrobiont fauna of moss: Rotifera (45%), Tardigrada (25%) and Nematoda (11%). The remaining 19% of clones correspond to the others phyla that can occasionally occur in moss samples (e.g. arthropods; see for example Sayre and Brunson, 1971; Barbuto & Zullini, 2006).

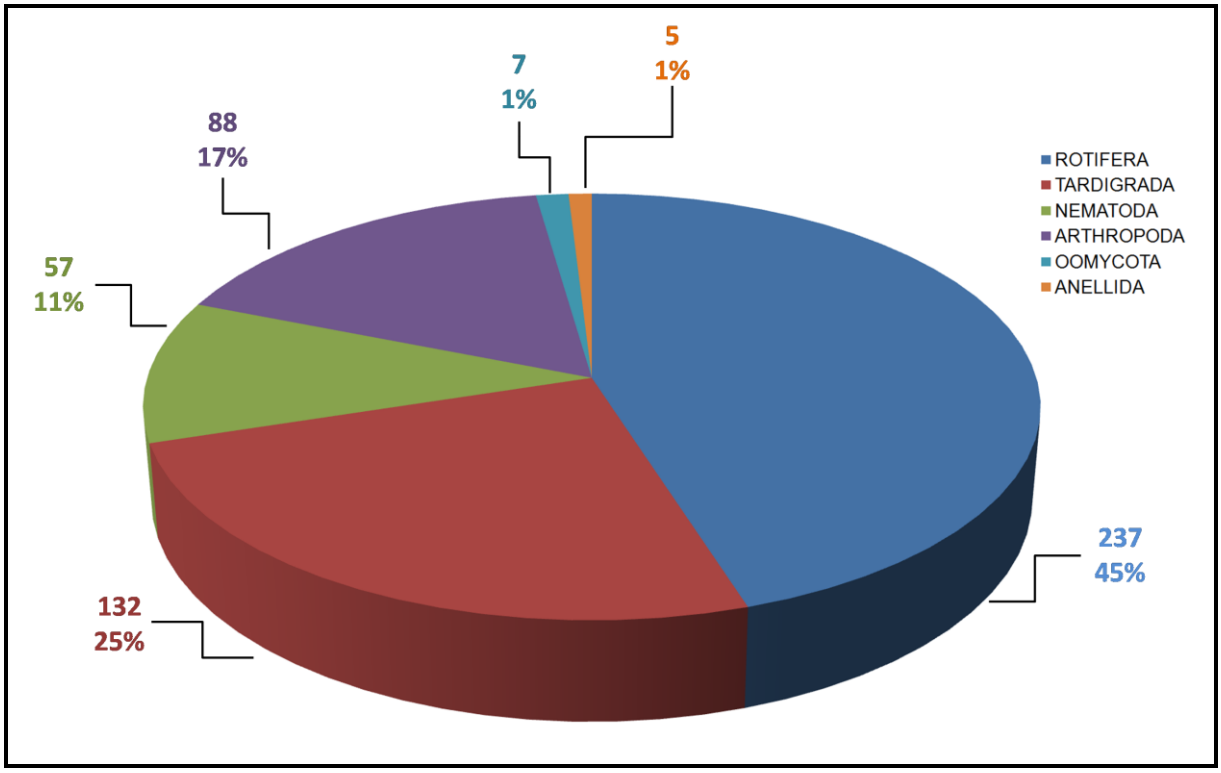


Figure 4.8: Schematic representation of the number of *coxI* sequences (also expressed into percent) of hydrobiont fauna of moss assigned to each phyla.

Similar observations can be inferred considering the relative abundances of the phyla listed above for each one of the four aliquots of the Baermann extract performed on the moss sample (Table 4.10). Although sequences belonging to rotifers, tardigrads and nematodes have been obtained for each aliquot, proportions among these organisms (intended as the number of *coxI* sequences representative of each phylum) was not constant. In particular, the first two aliquots (i.e. aliquot 1 and aliquot 2; Table 4.9) showed an almost identical number of *coxI* sequences for nematodes (10% in both aliquots) and rotifers (67% and 68% respectively), while slightly differences were observed in the relative abundance of barcode sequences for tardigrads (14% and 21% respectively). On the contrary, aliquots 3 and 4 showed a different pattern of abundance characterized by a clear prevalence of arthropods (54%) and tardigrads (69%) respectively.

All of the *coxI* sequences obtained from the cloning analysis have been subsequently partitioned in three sub-datasets encompassing only those entries belonging to rotifers (237

sequences), tardigrads (132 sequences) and nematodes (57 sequences) respectively. As described in chapter 3.7, for each sub-dataset, barcode sequences were assigned to different MOTUs by using BLASTClust (v. 2.2.1.6; Altschul et al., 1997) with 99% and 97.5% similarities. The results of this approach are showed in Table 4.11.

Table 4.11: Number of MOTUs resulting from the BLASTClust analyses performed on the clone library of *coxI* sequences by using two similarity thresholds. For each phylum, the mean number of sequences for each MOTU (with standard error and range) is also provided.

	phylum								
	ROTIFERA			**TARDIGRADA**			**NEMATODA**		
SIMILARITY THRESHOLD	MOTUs	Mean n of sequences per MOTU (s.e.)	range	MOTUs	Mean n of sequences per MOTU (s.e.)	range	MOTUs	Mean n of sequences per MOTU (s.e.)	range
1.0%	45	5.27 (12.18)	1-60	11	12 (30.84)	1-104	13	4.38 (4.41)	1-14
2.5%	24	9.88 (17.21)	1-65	5	26.6 (47.82)	1-111	9	6.33 (4.92)	1-15

As showed in the table, the number of MOTUs in each sub-dataset is nearly two-fold greater when passing from 1.0% to 2.5% threshold. Values of standard deviation are considerably high (reaching a maximum of 47.82 in the case of tardigrads) as well as the ranges of the sequences abundance for each MOTU. Although a NJ clustering analysis revealed a number of MOTUs for each sub-dataset roughly similar to the values obtained with a 2.5% BLASTClust threshold (data not shown), the absence of reference sequences for the putative species encompassed by each phylum, did not allow to obtain a significant estimation of the real number of species present in the environmental sample. However, these preliminary results confirmed our assumption that the massive approach based on DNA barcoding is successful for the characterization of biodiversity of moss hydrobiont community although some aspects causing bias in the analysis deserve to be further discussed.

The recent development of next-generation sequencing technologies (e.g. 454 pyrosequencing) could help overcoming the criticisms discussed above due to the possibility of obtaining thousands of sequences (e.g. > 400,000) for each analysed environmental sample. This astounding amount of data would allow an efficient qualitative interpretation of environmental surveys based on DNA barcoding (i.e. which species are present in the sampled environmental matrix).

On the contrary, methods and data handling tools able to perform a realistic quantitative analysis (i.e. which is the relative abundance of each taxa in the investigated environment) are still lacking due to the occurrence of several biases that could affect such a kind of inferences. In fact, despite its attractiveness, this perspective is far to be significantly supported and currently, any empiric demonstration has been provided in order to accomplish it (Valentini et al., 2008). Recent studies suggest being cautious when formulating quantitative assumptions based on massive DNA barcoding approaches (Valentini et al., 2008; Deagle et al., 2009; Soininen et al., 2009). It is indeed fundamental to consider that when dealing with molecular approaches, several factors could be responsible of the introduction of experimental biases that could significantly alter the molecular representation of the extant biodiversity.

One of the most important factors affecting the quantitative analysis of species composition in an environmental sample is the amount of DNA used as a template. This varies a lot depending on the taxa under examination and in particular on dimensions and relative abundance of each species. In the case study here analysed, although organisms dimensions were roughly comparable, data on the relative abundance of the different taxa recovered with Baermann method were lacking. Techniques used for the isolation of organisms from the environmental matrix and those methods used to extract total DNA, can also have a differential efficacy depending on the taxa considered. In particular, the latter aspect can be influenced in presence of species with exoskeleton, thick cuticle or other anatomical characteristics that could impede an efficient fragmentation of the specimen. Another category of theoretical biases concerns the PCR amplification of the target genomic region(s) to be used as barcode(s). In particular, as clearly stated in Suzuki & Giovannoni (1996), a significant estimate of the proportions occurring among taxa could be successfully obtained from the amplification of a target genomic region (e.g. a DNA barcode) only if four main assumptions are satisfied: *i*) all of the molecules of DNA would be equally accessible in the step of primers annealing; *ii*) efficiency in the formation of primer-template hybrids would be constant for all the taxa; *iii*) DNA polymerase would guarantee the same efficiency for all templates; *iv*) limitations in the availability of substrates would equally interest all the available templates.

A further level of bias that is introducible by the analysis is the so-called ‘emulsion PCR’ to be performed on the amplified samples before being loaded on the massive sequencer.

This PCR allows to link the amplified DNA samples on the beads for machine reading. However, this step can introduce a further source of differential linking among different sequences (as similarly occur with the two-steps PCR performed in our cloning approach).

Concerning primers, our data clearly confirmed the universality of the Folmer pair LCO1490 – HCO2198 (Folmer et al., 1994) in amplifying all the main phyla identified in the preliminary morphological inspection on the hydrobiont fauna extracted from moss. This condition perfectly agrees with the standards of a DNA barcoding approach (see chapter 1) but, on the other hand, did not allow a precise evaluation on the extent of mismatch cases between primer and templates. Such a situation can represent another source of bias in the comprehensive amplification of DNA barcode regions on environmental samples (Sipos et al., 2007). To cope with this problem, a *in silico* approach should be performed on a molecular datasets of known composition in order to test primer affinity for each taxon as described in Porazinska et al. (2009).

Finally, as extensively discussed above, the most important theoretical bias affecting the DNA barcoding-based characterization of biodiversity in environmental samples is the absence of a reference barcode library. In fact, the availability of a set sequences belonging to at least some representatives of the most abundant taxa inhabiting an environment (e.g. the moss) is crucial to define species boundaries and provide a realistic estimate of total biodiversity.

5. CONCLUSIONS AND FUTURE PERSPECTIVES

The main objective of this research project was to investigate the degree of effectiveness reached in combining a standardized molecular methodology (i.e. DNA barcoding) with classical biological information (e.g. morphology, ecology, host specificity), toward the synthesis of an integrated approach to taxonomy (see chapter 1.10). To cope with this requirement, nine case studies encompassing a wide panel of taxa (belonging to animal, plant and environmental samples) and some applicative fields of a diagnostic molecular tool (e.g. from pure taxonomy to the characterization of unknown biodiversity, from medical diagnostics to food traceability) have been analysed in the light of a DNA barcoding approach (see chapter 2). Standardization in the collection and processing of biological samples, as well as in the bioinformatic approaches used to manage and analyse molecular data has been a fundamental point in the experimental workflow we adopted (see chapter 3). A critical comparison between our data and already available molecular entries (e.g. deposited in GenBank, BOLD) was also performed, allowing in some cases to efficiently combine these resources toward the generation of reference datasets (useful in delimiting species boundaries) or in order to support identification results (e.g. in the context of food traceability).

The application of DNA barcoding in taxonomical investigations (e.g. case studies A1-A3, B1) revealed that the calculation of an optimum threshold of molecular divergence (OT) coupled with a cumulative error analysis (CE) is useful to synergically integrate molecular variability and classical taxonomic aspects. Such an analysis allowed to evaluate which is the extant strength of coherence between the two identification approaches for each considered taxon (e.g. is maximum in the case of terrestrial isopods; case study A2) therefore acting as an informative 'flag' for the occurrence of taxonomical criticisms (e.g. in the case of *Myotis myotis-M. blythii* complex; case study A1, see chapter 4.5.1). In this context, molecular barcodes were associated to different entities (i.e. IOTUs and MOTUs) on the basis of OT. These entities showed different attitudes in reducing the degree of uncertainty in a taxonomic scenario. In particular, we introduced here the concept of Integrated Operational Taxonomic Units (IOTUs; see chapter 1.7), to better define the entities existing between the ranks of species and of some other levels of variability.

In such a context, our results revealed that MOTUs and IOTUs are clearly different leading to a sort of increasing confidence in the taxonomic assignment starting from MOTUs (i.e. defined on the basis of one approach only, the molecular), passing through IOTUs (i.e. defined on the basis of at least two taxonomic approach, one of which is molecular) and arriving to species. However, it is important to underline that species level is always difficult to reach and caution should be placed in its definition. For this reason, following Padial et al. (2010) we here adopted the concept of Unconfirmed Candidate Species (UCSs) when highly divergent molecular lineages within conspecific individuals were identified with OT, but any other biological characteristic was so far available so far (e.g. *M. nattereri*, see case study A1; *T. taeniaeformis*, see case study B2). To be promoted to CCS (Confirmed Candidate Species), such divergent molecular entities deserve to be further investigated with additional approaches based on both molecular (e.g. nuclear markers, microsatellites) and classical taxonomy data (e.g. morphology, biogeography, behaviors) as partially discussed for case studies A1 and A3.

In conclusion, our results clearly show that DNA barcoding represents a powerful tool for taxonomy, but without the integration of traditional approaches could become a simple collection of MOTUs. Given this assumption, it is reasonable to affirm that the establishment, improvement and maintenance of DNA barcoding as a taxonomic tool cannot prescind from a long-lasting interaction between traditional taxonomy and DNA-based approaches.

Integrative approach also permitted to infer some generalization concerning the marker(s) chosen as molecular barcode(s) and different kinds of data handling. In this context, while the mitochondrial *coxI* (and in particular the Folmer region) has revealed to be successfully in the characterization of a wide range of animal taxa, some doubts raised on the effectiveness for molecular markers adopted in plants identification (i.e. case studies C2 and C3). In fact, although the CBOL suggests the use of a combination of the two plastidial *matK* and *rbcL* as a standard barcode, our results clearly showed that an alternative combination based on *matK* and the intergenic spacer trnH-psbA (and even nuclear markers) provide a better resolution. The most performing data handling approach is the one based on OT-MCE calculation (as discussed above), but NJ clustering can often represent the best choice (in terms of operating speed) when a taxonomical identification is lacking or in case of insufficient sampling. As widely discussed in chapter 1.9, several alternative bioinformatical methods exist to cope with these situations (e.g. character based analysis), but an accurate

evaluation of the comparative performances of these techniques and a substantial decreasing of computational times and requirements is desirable before they will be routinely adopted.

Currently available barcode data deposited in the most important international archives (e.g. GenBank and BOLD) partially agreed with our results in assigning a species rank for undetermined samples (e.g. 'blind samples' of case studies A1 and B1; see also chapters 4.5.3). However, such a comparison also showed several problems concerning the biodiversity coverage for some taxa (both intra and interspecic) and the occurrence of identification mistakes in the deposited barcode sequences.

The results reached in case studies B1 and B2 (i.e. filarioid nematodes and taeniids respectively) confirmed the suitability of a DNA barcoding approach in the field of medical and veterinary diagnostics. Moreover, despite the databases here reported did not encompass all of the endoparasite species of medical (or veterinary) interest, they represent a useful starting point for rapid identification of these parasites and for applications such as epidemiological surveys and populational dynamics.

Similar relevant results have been obtained in the context of food and toxic plants traceability (i.e. case studies C1-C3). In particular, our data confirmed the reliability of DNA barcoding in recognizing commercial frauds in the trade of fishes products (e.g. the case of 'palombo', *Mustelus* spp.), in spices traceability (with the exception of some groups because of the occurrence of hybridization events) and in the fast identification of poisonous plants starting from plant fragments, seeds or fruits (parts that are frequently found in stomach continent of patients). These results demonstrated that DNA barcoding could have a wide implication as a supporting tool in the survey of fish or edible plant species which commerce is regulated by local and international directives or as a diagnostic application useful in poison control centers.

Finally, preliminary results obtained for the characterization of environmental biodiversity (i.e. case study D1) showed that a DNA barcoding approach, based on the use of universal primers, is successful in acquiring qualitative information on the fauna inhabiting a certain environmental matrix (i.e. moss in our study; see chapter 4.8). However, the creation of reference libraries and an accurate *in silico* investigation on the potential biases affecting the large-scale applicability of the method is strongly required to standardize this approach. Once further data will be available, the next step of the analysis will focus on the employment

of next-generation sequencing (e.g. 454 pyrosequencing) in order to obtain a broader and realistic coverage of the extant biodiversity in environmental samples.

As a final conclusion, it is possible to affirm that this research project highlighted the role of DNA barcoding as a catholic method to discriminate biological entities. Although clear limitations arise from the incomplete coverage of the existing diversity, the inherent characteristics of the mitochondrial or plastidial DNA and the single-locus strategy initially proposed by CBOL, the method showed to be more flexible than expected. Moreover, even if developed beyond standard approaches based on the existing taxonomic knowledge, DNA barcoding can meet all the requirements to enhance communication between different scientific disciplines, including taxonomy, phylogenetics and population genetics.

6. REFERENCES

Abdo Z, Golding GB: A step toward barcoding life: a model-based, decision-theoretic method to assign genes to preexisting species groups. *Systematic Biology* 2007; 56:44–56.

Adams DC, Chelsea M, Kozak KH, Wiens JJ: Are rates of species diversification correlated with rates of morphological evolution? *Proceedings of the Royal Society B: Biological Science* 2009; 276:2729-2738.

Agatsuma T, Iwagami M, Uni S, Takaoka H, Katsumi A, Kimura E, Bain O: Molecular phylogenetic relationships among seven Japanese species of *Cercopithifilaria*. *Parasitology International* 2005; 54:195-199.

Agnelli P, Martinoli A, Patriarca E, Russo D, Scaravelli D, Genovesi P: Guidelines for bat monitoring: methods for the study and conservation of bats in Italy. *Quaderni di Conservazione della Natura* 19 bis 2006; Min. Ambiente – Ist. Naz. Fauna Selvatica, Rome and Ozzano dell'Emilia (Bologna), Italy.

Aliabadian M, Kaboli M, Nijman V, Vences M: Molecular identification of birds: performance of distance-based DNA barcoding in three genes to delimit parapatric species. *PLoS ONE* 2009; 4,: e4119.

Alström P, Rasmussen PC, Olsson U, Sundberg P: Species delimitation based on multiple criteria: the Spotted Bush Warbler *Bradypterus thoracicus* complex (Aves:Megaluridae). *Zoological Journal of the Linnean Society* 2008; 154:291-307.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *Journal of Molecular Biology* 1990; 215:403–10.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman, DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 1997; 25: 3389-3402.

André HM, Ducarme X, Anderson JM et al.: Skilled eyes are needed to go on studying the richness of the soil. *Nature* 2001; 409: 761.

André HM, Ducarme X, Lebrun P: Soil biodiversity: myth, reality or conning? *Oikos* 2002; 96: 3–24.

Artyushin IV, Bannikova AA, Lebedev VS, Kruskops SV: Mitochondrial DNA relationships among North Palaearctic *Eptesicus* (Vespertilionidae, Chiroptera) and past hybridization between Common Serotine and Northern Bat. *Zootaxa* 2009; 2262: 40-52.

Austerlitz F, David O, Schaeffer B, Bleakley K, Olteanu M, Leblois R, Veulle M., Laredo C: DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. *BMC Bioinformatics* 2009; 10 (Suppl 14): S10.

Azuma H, Okamoto M, Oku Y, Kamiya M: Intraspecific variation of *Taenia taeniaeformis* as determined by various criteria. *Parasitolology Research* 1995; 81: 103–108.

Baker CS, Palumbi SR: Which whales are hunted – a molecular genetic approach to monitoring whaling. *Science* 1994; 265: 1538–1539.

Ball SL, Hebert PDN, Burian SK, Webb JM: Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *Journal of the North American Benthological Society* 2005; 24: 508–524.

Baratelli D, Casiraghi M, Celada C, Gagliardi A: Le specie alloctone in Italia: censimenti, invasività e piani di azione. *Memorie della Società Italiana di Scienze Naturali e del Museo Civico di Storia Naturale di Milano* 2008; 36: 23-24.

Barbuto M, Zullini A: Moss inhabiting nematodes: influence of the moss substratum and geographical distribution in Europe. *Nematology* 2006; 8: 575-582.

Barbuto M, Galimberti A, Ferri E, Labra M, Malandra R, Galli P, Casiraghi M: DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of ''palombo''(*Mustelus* spp.). *Food Research International* 2010; 43: 376–381.

Barrett RDH, Hebert PDN: Identifying spiders through DNA barcodes. *Canadian Journal of Zoology* 2005; 83: 481–491.

Begerow D, Nilsson H, Unterseher M, Maier W: Current state and perspectives of fungal DNA barcoding and rapid identification procedures. *Applied Microbiology and Biotechnology* 2010; 87(1): 99-108.

Berthier P, Excoffier L, Ruedi M. Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. *Proceedings of the Royal Society B: Biological Sciences* 2006; 273: 3101–3109.

Besansky NJ, Severson DW, Ferdig MT: DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. *Trends in Parasitology* 2003; 19:545-546.

Bensasson D, Zhang D, Hartl DL, Hewitt GM: Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution* 2001; 16:314-321.

Bhadury P, Austen MC, Bilton DT, Lambshead PJD, Rogers AD, Smerdon GR: Development and evaluation of a DNA barcoding approach for the rapid identification of nematodes. *Marine Ecology Progress Series* 2006; 320:1–9.

Bisby FA, Shimura J, Ruggiero M, et al.: Taxonomy at the click of a mouse. *Nature* 2002; 418:367.

Blanco M, Pérez-Martìn RI, Sotelo CG: Identification of shark species in seafood products by forensically informative nucleotide sequencing (FINS). *Journal of Agricultural and Food Chemistry* 2008; 56: 9868–9874.

Blaxter M.: Molecular systematics: counting angels with DNA. *Nature* 2003; 421:122–4.

Blaxter M: The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2004; 359:669–79.

Blaxter M, Mann J, Chapman T, et al.: Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2005; 1462:1935–43.

Borisenko AV, Lim BK, Ivanova NV, Hanner RH, Hebert PDN: DNA barcoding in surveys of small mammal communities: a field study in Suriname. *Molecular Ecology Resources* 2008; 8: 471–479.

Bouzid W, Ŝtefka J, Hypŝa V, Lek S, Scholz T, Legal L, Hassine OKB, Loot G: Geography and host specificity: Two forces behind the genetic structure of the freshwater fish parasite *Ligula intestinalis* (Cestoda: Diphyllobothriidae). *International Journal for Parasitology* 2008; 38: 1465–1479.

Brambilla M, Vitulano S, Spina F, et al.: A molecular phylogeny of the *Sylvia cantillans* complex: Cryptic species within the Mediterranean basin. *Molecular Phylogenetics and Evolution* 2008; 48: 461-472.

Britten RJ, Rowen L, Williams J, Cameron RA: Majority of divergence between closely related DNA samples is due to indels. *Proc Natl Acad Sci Unit States Am* 2003; 100:4661-4665.

Bruni I, De Mattia F, Galimberti A, Galasso G, Banfi E, Casiraghi M, Labra M: Identification of poisonous plants by DNA barcoding approach. *International Journal of Legal Medicine* 2010; 124(6): 595-603.

Buhay JE: ‘‘COI-like’’ sequences are becoming problematic in molecular systematic and DNA barcoding studies. *Journal of Crustaceans Biology* 2009; 29:96–110.

Burrows GE, Tyrl RJ: *Handbook of Toxic Plants of North America*. Blackwell Publishing Professional, Ames, IA 2006; pp: 55-307.

Calvete C, Lucientes J, Castillo JA, Estrada R, Gracia MJ, Peribanez MA, Ferrer M: Gastrointestinal helminth parasites in stray cats from the mid-Ebro Valley, Spain. *Veterinary Parasitology* 1998; 75: 235–240.

Carapelli A, Liò P, Nardi F, van der Wath E, Frati F: Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. *BMC Evolutionary Biology* 2007; 7(2): S8.

Cardoso A, Serrano A, Vogler AP: Morphological and molecular variation in tiger beetles of the Cicindela hybrida complex: is an 'integrative taxonomy' possible? *Molecular Ecology* 2009; 18: 648-664.

Casiraghi M, Labra M, Ferri E, Galimberti A., De Mattia F: DNA barcoding: a six-question tour to improve users' awareness about the method. *Briefings in Bioinformatics* 2010; 11(4): 440-453.

Castella VM, Ruedi L, Excoffier C, Ibáyez R, Arlettaz, Hausser J: Is the Cryptic diversity in Iberian bats 289Gibraltar Strait a barrier to gene flow for the bat *Myotis myotis* (Chiroptera: Vespertilionidae)? *Molecular Ecology* 2000; 9: 1761–1772.

Caterino MS, Tishechkin AK: DNA identification and morphological description of the first confirmed larvae of Hetaeriinae (Coleoptera: Histeridae). *Systematic Entomology* 2006; 31: 405–418.

CBol Plant working group: A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* 2009; 106: 12794–7.

Cesari M, Bertolani R, Rebecchi L, Guidetti R: DNA barccoding in Tardigrada: the first case study on *Macrobiotus macrocalix* Bertolani & Rebecchi 1993 (Eutardigrada, Macrobiotidae). *Molecular Ecology Resources* 2009; doi: 10.1111/j.1755-0998.2009.02538.x.

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V: Land plants and DNA barcodes: short-term and longterm goals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2005; 360: 1889–1895.

Chase MW, Salamin N, Wilkinson M et al.: Land plants and DNA barcodes: short-term and long-term goals. *Philos Trans R Soc London [Biol]* 2005; 360:1889–1895.

Chu KH, Li CP, Qi J: Ribosomal RNA as molecular barcodes: a simple correlation analysis without sequence alignment. *Bioinformatics* 2006; 22: 1690–701.

Chu KH, Xu M, Li CP: Rapid DNA barcoding analysis of large datasets using the composition vector method. *BMC Bioinformatics* 2009; 10(Suppl 14):S8.

Clare EL, Lim BK, Engstrom MD, Eger JL, Hebert PDN: DNA barcoding of Neotropical bats: species identification and discovery within Guyana. *Molecular Ecology Notes* 2007; 7: 184-190.

Clement M, Posada D, Crandall KA: TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 2000; 9: 1657–1659.

Clements JF: *The Clements Checklist of Birds of the World*, 6th ed. Ithaca: Cornell Univ. Press; 2007.

Dayrat B: Toward integrative taxonomy. *Biological Journal of the Linnean Society* 2005; 85: 407-415.

DasGupta B, Konwar KM, Mandoiu II, et al.: DNA-BAR: distinguisher selection for DNA barcoding. *Bioinformatics* 2005; 21: 3424–6.

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

De Maddalena A, Piscitelli L: Analisi preliminare dei selaci registrati presso il mercato ittico di Milano (aprile-settembre 2000). *Bollettino del Museo Civico di Storia Naturale di Venezia* 2001; 52: 129–145.

Deagle BE, Kirkwood R, Jarman SN: Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology* 2009; 18: 2022-2038.

Degnan J, Rosenberg N: Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* 2009; 24: 332-340.

De Ley P, De Ley IT, Morris K, et al: An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2005; 360:1945–58.

De Mattia F, Bruni I, Galimberti A, Cattaneo F, Casiraghi M, Labra M: A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea. *Food Research International* in press.

DeSalle R, Birstein VJ: PCR identification of black caviar. *Nature* 1996; 381:197–198.

DeSalle R, Egan MG, Siddall M: The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2005; 360: 1905–16.

DeSalle R: Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. *Conservation Biology* 2006; 20: 1545–1547.

Dhanya K, Sasikumar B: Molecular marker based adulteration detection in traded food and agricultural commodities of plant origin with special reference to spices. *Current Trends in Biotechnology and Pharmacy* 2010; 4: 454-489.

Dickinson EC: *The Howard and Moore Complete Checklist of the Birds of the World*, third ed. Christopher Helm, London; 2003.

Dietz C, Von Helversen O. Illustrated identification key to the bats of Europe. Version 1.0. *Electronic Publication, Tuebingen & Erlangen*; 2004.

Dorris M, Viney ME, Blaxter ML: Molecular phylogenetic analysis of the genus Strongyloides and related nematodes. *Int J Parasitol* 2002; 32:1507-1517.

Dunning L, Savolainen V: Broad-scale amplification of matK for DNA barcoding plants, a technical note. *Botanical Journal of the Linnean Society* 2010; 164(1): 1-9.

Eddy SR: Profile hidden Markov models. *Bioinformatics* 1998; 14: 755–763.

Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; 32:1792-1797.

Ekrem T, Willassen E, Stur E: A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution* 2007; 43: 530–42.

Eldredge N, Cracraft J: Phylogenetic Patterns and The Evolutionary Process: Method and Theory in Comparative Biology. New York: Columbia University Press, 1980.

Elias M, Hill RI, Willmott KR, et al: Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B: Biological Sciences* 2007; 274: 2881–9.

Emery VJ, Landry J.F., Eckert CG: Combining DNA barcoding and morphological analysis to identify specialist floral parasites (Lepidoptera: Coleophoridae: Momphinae: Mompha). *Mol.Ecol.Res* 2009; 9(1): 217-223.

Eom KS, Rim HJ: Morphologic descriptions of *Taenia asiatica* sp. n. *Korean Journal of Parasitology* 1993; 31: 1–6.

Evans KM, Wortley AH, Mann DG: An assessment of potential diatom ''barcode'' genes (*cox1*, *rbcL*, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta). *Protist* 2007; 158: 349–364.

Evin A, Lecoq V, Durand MO, Tillon L, Pons JM: A new species for the French bat list: *Myotis escalerai* (Chiroptera: Vespertilionidae). *Mammalia* 2009; 73: 142-144.

Farrell ED, Clarke MW, Mariani S: A simple genetic identification method for Northeast Atlantic smoothhound sharks (*Mustelus* spp.). *ICES Journal of Marine Science* 2009; 66: 561–565.

Fazekas AJ, Burgess KS, Kesanakurti PR, et al.: Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One* 2008; 3: e2802.

Fazekas AJ, Kesanakurti PR, Burges KS, et al.: Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources* 2009; 9:130–9.

Felsenstein J: Confidence limits on phylogenies - an approach using the bootstrap. *Evolution* 1985; 39:783-791.

Ferguson JWH: On the use of genetic divergence for identifying species. *Biological Journal of the Linnean Society* 2002; 75: 509–16.

Ferri E, Barbuto M, Bain O, et al.: Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda). *Frontiers in Zoology* 2009; 6:1.

Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessière J, Taberlet P, Pompanon F: An *In silico* approach for the evaluation of DNA barcodes. *BMC Genomics* 2010; 11:434.

Floyd R, Abebe E, Papert A, et al.: Molecular barcodes for soil nematode identification. *Molecular Ecology* 2002; 11: 839–50.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R: DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 1994; 3: 294–299.

Fouquet A, Gilles A, Vences M, Marty C, Blanc M, Gemmell NJ: Underestimation of Species Richness in Neotropical Frogs Revealed by mtDNA Analyses. *PloS One* 2007; 2(10): e1109.

Fox GE, Stackebrandt E, Hespell RB, Gibson J, Maniloff J, Dyer TA, Wolfe RS, Balch WE, Tanner RS, Magrum LJ, Zablen LB, Blakemore R, Gupta R, Bonen L, Lewis BJ, Stahl DA, Luehrsen KR, Chen KN, Woese CR: The phylogeny of prokaryotes. *Science* 1980; 209: 457–463.

Francis CM, Borisenko AV, Ivanova NV, et al.: The Role of DNA Barcodes in Understanding and Conservation of Mammal Diversity in Southeast Asia. *PloS one* 2010; 5(9):e12575.

Frézal L, Leblois R: Four years of DNA barcoding: current advances and prospects. *Infection, Genetics and Evolution* 2008; 8:727–36.

Funk DJ, Omland KE: Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insight from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics* 2003; 34: 397–423.

Gardner DR, Pfister JA: Toxic alkaloid concentrations in *Delphinium nuttallianum*, *Delphinium andersonii*, and *Delphinium geyeri* in the Intermountain Region. *Rangeland Ecology & Management* 2007; 60: 441–446.

García-Mudarra JL, Ibáñez C, Juste J: The Straits of Gibraltar: barrier or bridge to IberoMoroccan bat diversity? *Biological Journal of the Linnean Society* 2009; 96: 434–450.

Gatti F, Parisi V: Composizione delle faunule muscinali in ambienti urbani: ricerche nella città di Parma. In Ecologia. *Atti del XIV Congresso Nazionale della Società Italiana di Ecologia* (Siena, 4-6 ottobre 2004).

Geiger DL: Stretch coding and block coding: two new strategies to represent questionably aligned DNA sequences. *Journal of Molecular Evolution* 2002; 54: 191–9.

Gibbs J: Integrative taxonomy identifies new (and old) species in the *Lasioglossum* (*Dialictus*) *tegulare* (Robertson) species group (Hymenoptera, Halictidae). *Zootaxa* 2009; 2032: 1-38.

Giller PS: The diversity of soil communities, the 'poor man's tropical rainforest'. *Biodiversity Conservation* 1996; 5: 135–168.

Glaw F, Köhler J, De la Riva I, Vieites DR, Vences M: Integrative taxonomy of Malagasy treefrogs: combination of molecular genetics, bioacoustics and comparative morphology reveals twelve additional species of Boophis. *Zootaxa* 2010; 283: 1-82.

Godfray HCJ: Challenges for taxonomy. *Nature* 2002; 417: 17–19.

Goldstein PZ, DeSalle R: Calibrating phylogenetic species formation in a threatened insect using

Gomez-Alvarez V, King GM, Nüsslein K: Comparative bacterial diversity in recent Hawaiian volcanic deposits of different ages. *FEMS Microbiology Ecology* 2007; 60: 60–73.

Gobert V, Moja S, Colson M, Taberlet P: Hybridization in the section Mentha (Lamiaceae) inferred from AFLP markers. *American Journal of Botany* 2002; 89: 2017–2023.

Gounaris Y, Skoula M, Fournaraki C, Drakakaki G, Makris A: Comparison of essential oils and genetic relationship of Origanum x intercedens to its parental taxa in the island of Crete. *Biochemical Systematics and Ecology* 2002; 30:249-258.

Grassi F, Labra M, Minuto L, Casazza G, Sala F: Natural hybridization in *Saxifraga callosa Sm. Plant Biol* 2006; 8:243–252.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN: DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America* 2006; 103: 968–971.

Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp* 1999; 41:95–98.

Hammouda FM, Rizk AM, El-Missiry MM, et al.: Poisonous plants contaminating edible ones and toxic substances in plant foods phytochemistry and toxicity of Lolium temulentum. *Pharm Biol* 1988; 26:240–245.

Han L: A taxonomic study on Rufous-headed Crowtit in China. *Zoological Research* 1991; 12: 117-124 (English translation).

Hebert PDN, Ratnasingham S, deWaard JR: Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Science* 2003; 270(Suppl): S96–9.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W: Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences of the United States of America* 2004a; 101: 14812–14817.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM: Identification of birds through DNA barcodes. *Plos Biology* 2004b; 2: 1657–1663.

Helbig AJ, Knox AG, Parkin DT, et al.: Guidelines for assigning species rank. *Ibis* 2002; 144: 518–25.

Herrera A, Héry M, Stach JEM, Jaffré T, Normand P, Navarro E: Species richness and phylogenetic diversity comparisons of soil microbial communities affected by nickel-mining and revegetation efforts in New Caledonia. *European Journal of Soil Biology* 2007; 43: 130–139.

Hewitt GM: Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 1996; 58: 247–276.

Hewitt GM: The genetic legacy of the Quaternary ice ages. *Nature* 2000; 405: 907–913.

Hogg ID, Hebert PDN:Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. *Canadian Journal of Zoology* 2004; 82: 749–754.

Hollingsworth ML, Clark A, Forrest LL, et al.: Selecting barcoding loci for plants: evaluation of seven candidate loci with species level sampling in three divergent groups of land plants. *Molecular Ecology Resources* 2009; 9: 439–57.

Holmes BH, Steinke D, Ward RD: Identification of shark and ray fins using DNA barcoding. *Fisheries Research* 2009; 95: 280–288.

Hooper, DJ. Extraction and processing of plant and soil nematodes. In: *Plant Parasitic Nematodes in Subtropical and Tropical Agriculture* 1990; Luc, M., Sikora, R.A. and Bridge, J. (Eds.), pp. 45-68. CAB International, Wallingford.

Huang J, Xu Q, Sun ZJ, Tang GL, Su ZY: Identifying earthworms through DNA barcodes. *Pedobiologia* 2007; 51:301-309.

Huelsenbeck JP: The performance of phylogenetic methods in simulation. *Systematic Biology* 1995; 44: 17–48.

Hurst GDD, Jiggins FM: Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Science* 2005; 272: 1525–34.

Ibáñez C, García-Mudarra JL, Ruedi M, Stadelmann B, Juste J: The Iberian contribution to cryptic diversity in European bats. *Acta Chiropterologica* 2006; 8(2): 277-297.

Isaac NJ, Mallet J, Mace GM: Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology and Evolution* 2004; 19 :464-469.

Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN: Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes* 2007; 7: 544–548.

Iwaki T, Nonaka N, Okamoto M, Oku Y, Kamiya M: Developmental and morphological characteristics of *Taenia taeniaeformis* (Batsch, 1786) in *Clethrionomys rufocanus bedfordiae* and *Rattus norvegicus* from different geographical locations. *J. Parasitol.*1994; 80(3):461 – 467.

Jaarola M, Searle JB: A highly divergent mitochondrial DNA lineage of *Microtus agrestis* in southern Europe. *Heredity* 2004; 92: 228–234.

Jaklitsch WM, Komon M, Kubicek CP, Druzhinina IS: *Hypocrea crystalligena* sp nov., a common European species with a white-spored Trichoderma anamorph. *Mycologia* 2006; 98: 499–513.

Jeon HK, Eom KS: *Taenia asiatica* and *Taenia saginata*: Genetic divergence estimated from their mitochondrial genomes. *Experimental Parasitology* 2006; 113:58–61.

Johnsen A, Rindal E, Ericson GP, Zuccon D, Kerr KCR, Stoeckle MY, Lifjeld JT: DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. *J. Ornithol.* in press.

Jones G: Acoustic signals and speciation: the roles of natural and sexual selection in the evolution of cryptic species. *Advances in the Study of Behavior* 1997; 26: 317-354.

Kelly RP, Sarkar IN, Eernisse DJ, DeSalle R: DNA barcoding using chitons (genus *Mopalia*). *Molecular Ecology Notes* 2007; 7: 177–183.

Kent WJ: BLAT - The BLAST-like alignment tool. *Genome Research* 2002; 12: 656–64.

Kerr KCR, Birks SM, Kalyakin MV, Red'Kin YA, Koblik EA, Hebert PDN: Filling the gap – COI barcode resolution in eastern Palearctic birds. *Frontiers in Zoology* 2009; 6: 29.

Kerr CRK: A cryptic, intergeneric cytochrome c oxidase I pseudogene in tyrant flycatchers (family: Tyrannidae). *Genome* 2010; 53:1103-1109.

Klossa-Kilia E, Kilias G, Sfenthourakis S: Increased genetic divergency in Greek populations of the genus *Ligiditum* (Crustacea: Isopoda: Oniscidea) revealed by RFLP analysis of mt-DNA segments. *Contribution to Zoology* 2006; 74: (3/4).

Knowles LL, Carstens BC: Delimiting species without monophyletic gene trees. *Systematic Biology* 2007; 56: 887–95.

Köhler J, Vieites DR, Bonett RM, García FH, Glaw F, Steinke D, Vences M: Boost in species discoveries in a highly endangered vertebrate group: new amphibians and global conservation. *BioScience* 2005; 55: 693-696.

Kolaczkowski B, Thornton JW: Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature* 2004; 431 :980–4.

Koski LB, Golding GB: The closest BLAST hit is often not the nearest neighbor. *Journal of Molecular Evolution* 2001; 52: 540–2.

Kress WJ, Wurdack KJ, Zimmer EA, et al.: Use of DNA barcodes to identify flowering plants. PNAS 2005; 102: 8369–74.

Kress WJ, Erickson DL: A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. *PLoSone* 2007; 2(6): e508.

Kress WJ, Erickson DL: DNA barcodes: genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences of the United States of America* 2008; 105: 2761–2.

Krone O, Guminsky O, Meinig H, Herrmann M, Trinzen M, Wibbelt G: Endoparasite spectrum of wild cats (*Felis silvestris* Schreber 1777) and domestic cats (*Felis catus* L.) from the Eifel, Pfalz region and Saarland, Germany. *European Journal of Wildlife Research* 2008; 54: 95–100.

Kumar S, Dudley J, Nei M, Tamura K: MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform*. 2008; 9: 299-306.

Kuksa P, Pavlovic V: Efficient alignment-free DNA barcode analytics. *BMC Bioinformatics* 2009; 10(Suppl 14):S9.

Lahaye R, van der Bank M, Bogarin D, et al.: DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences of the United States of America* 2008; 105: 2923–8.

Lake JA: Origin of the Metazoa. *Proc. Natl. Acad. Sci. U.S.A.* 1990; 87:763-766.

Lavikainen A, Haukisalmi V, Lehtinen MJ, Henttonen H, Oksanen A, Meri S: A phylogeny of members of the family Taeniidae based on the mitochondrial *cox1* and *nad1* gene data. *Parasitology* 2008; 135: 1457–1467.

Lawrence RA: Poisonous plants: when they are a threat to children. *Pediatrics in Review* 1997; 18:162–168.

Lee MSY: The molecularisation of taxonomy. *Invertebrate Systematics* 2004; 18: 1–6.

Lefébure T, Douady CJ, Gouy M, et al.: Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phyogenetics and Evolution* 2006; 40: 435–47.

Li M, Wunder J, Bissoli G, Scarponi E, Gazzani S, Barbaro E, Saedler H, Varotto C: Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species. *Cladistics* 2008; 24:727–745.

Librado P, Rozas J: DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009; 25: 1451-1452.

Linares MC et al.: High mitochondrial diversity in geographically widespread butterflies of Madagascar: a test of the DNA barcoding approach. *Mol. Phylogenet. Evol.* 2009; 50:485–495.

Litaker RW, Vandersea MW, Kibler SR, Reece KS, Stokes NA, Lutzoni FM, Yonish BA, West MA, Black MND, Tester PA: Recognizing dinoflagellate species using ITS rDNA sequences. *Journal of Phycology* 2007; 43: 344–355.

Little DP, Stevenson DWM: A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. *Cladistics* 2007; 23: 1–21.

Lopez JA, Ryburn JA, Fedrigo O, Naylor GJP: Phylogeny of sharks of the family Triakidae (Carcharhiniformes) and its implications for the evolution of carcharhiniform placental viviparity. *Molecular Phylogenetics and Evolution* 2006; 40:50–60.

Magnacca KN, Brown MJF: Tissue segregation of mitochondrial haplotypes in heteroplasmic Hawaiian bees: implications for DNA barcoding. *Molecular Ecology Resources* 2010; 10: 60–68.

Maiden, MCJ, Bygraves JA, Feil E, et al.: Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* 1996; 95(6): 3140–3145.

Mallet J, Willmott K: Taxonomy: renaissance or Tower of Babel? *Trends in Ecology and Evolution* 2003; 18: 57–59.

Marko PB, Lee SC, Rice AM, Gramling JM, Fitzhenry TM., McAlister JS, et al.: Fisheries: Mislabelling of a depleted reef fish. *Nature* 2004; 430: 309–310.

Margam VM: A simplified arthropod genomic-DNA extraction protocol for polymerase chain reaction (PCR)-based specimen identification through barcoding. *Molecular Biology Reports* 2010; 37(7): 3631.

Matz MV, Nielsen R: A likelihood ratio test for species membership based on DNA sequence data. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2005; 360: 1969–74.

May RR, Harvey PH: Species uncertainties. *Science* 2009; 323: 687.

Mayer F, von Helversen O: Cryptic diversity in European bats. *Proceedings of the Royal Society of London Series B, Biological Sciences* 2001; 268:1825–1832.

Mayer F, Dietz C, Kiefer A: Molecular species identification boosts bat diversity. *Frontiers in Zoology* 2007; 4: 4.

McIntire MS, Guest JR, Porterfield JF: Philodendron - an infant death. *Journal of Toxicology – Clinical Toxicology* 1990; 28:177–183.

Meier R, Zhang G, Ali F: The use of mean instead of smallest interspecific distances exaggerates the size of the ‘‘Barcoding Gap’’ and leads to misidentification. *Systematic Biology* 2008; 57: 809–13.

Meiri S, Mace GM: New taxonomy and the origin of species. *PLoS Biology* 2007; 5: e194.

Meyer CP, Paulay G: DNA Barcoding: error rates based on comprehensive sampling. *PLoS Biology* 2005; 3:e422.

Min XJ, Hickey D: Assessing the effect of varying sequence length on DNA barcoding of fungi. *Molecular Ecology Notes* 2007; 7: 365–73.

Monaghan MT, Wild R, Elliot M, et al.: Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology* 2009; 58: 298–311.

Moore J, Lasswell J: Altered behavior in isopods (*Armadillidium vulgare*) infected with the nematode Dispharynx nasuta. *J. Parasitol.*1986; 72(1):186-189.

Morales-Hojas R, Cheke RA, Post RJ: Molecular systematics of five Onchocerca species (Nematoda: Filarioidea) including the human parasite, O. volvulus, suggest sympatric speciation. *J Helminthol* 2006; 80:281-290.

Moreira D, López-Garciá P: The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends in Microbiology* 2002; 10: 31–8.

Moritz C, Cicero C: DNA barcoding: promise and pitfalls. *PLoS Biology* 2004; 2: e354.

Mower J, Touzet P, Gummow J, et al.: Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evolutionary Biology* 2007; 7:135.

Mrvos R, Krenzelok EP, Jacobsen TD: Toxidromes associated with the most common plant ingestions. *Veterinary and Human Toxicology* 43: 366–369.

Naciri Y, Manen JF: PotentialDNA transfer fromthe chloroplast to the nucleus in *Eryngium alpinum*. *Molecular Ecology Resources* 2009; 10: 728-731.

Nei M, 1987. *Molecular Evolutionary Genetics*. Columbia Univ. Press, New York.

Nei M, Kumar S: *Molecular Evolution and Phylogenetics*. New York: Oxford University Press, 2000.

Neigel J, Domingo A, Stake J: DNA barcoding as a tool for coral reef conservation. *Coral Reefs* 2007; 26: 487–499.

Ngogang J, Nkongmeneck BA, Biyiti Bi Essam LF, Essame Oyono JL, Tsabang N, Zapfack L, Ngono Mballa R, Tamze V: Evaluation of acute and sub acute toxicity of four medicinal plants extracts used in Cameroon. *Toxicology Letters* 2008; 180: S185–S186.

Nichols R: Gene Trees and Species Trees Are Not the Same. *Trends in Ecology & Evolution* 2001; 16(7): 358-364.

Nielsen R, Matz: Statistical approaches for DNA barcoding. *Systematic Biology* 2006; 55:162–9.

Okamoto M, Bessho Y, Kamiya M, Kurosawa T, Horii T: Phylogenetic relationships within *Taenia taeniaeformis* variants and other taeniid cestodes inferred from the nucleotide sequence of the cytochrome *c* oxidase subunit I gene. *Parasitology Research* 1995; 81:451–458.

Oliver L, Beattie AJ: A possible method for the rapid assessment of biodiversity. *Conservation Biology* 1993; 7: 562–8.

Oliver PG, Meechan CJ: *Woodlice*; Linnean Society of London, Shrewsbery, U.K. 135 pp.1993.

Padial JM, De la Riva I: Taxonomic inflation and the stability of species lists: the perils of ostrich's behavior. *Systematic Biology* 2006; 55: 859-867.

Padial JM, Miralles A, De la Riva I, Vences M: The integrative future of taxonomy. *Frontiers in Zoology* 2010; 7:16.

Pearson WR, Lipman DJ: Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America* 1988; 85: 2444–8.

Peer Y Van de, Watcher R de: TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *CABIOS* 1994; 10: 569-570.

Pegg GG, Sinclair B, Briskey L, Aspden WJ: MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia. *Marine Science* 2006; 70: 7–12.

Penhallurick J, Robson C: The generic taxonomy of Parrotbills (Aves, Timaliidae). *Forktail* 2009; 25: 138-142.

Pennisi E: Modernizing the tree of life. *Science* 2003; 300: 1692–7.

Pennisi E: Wanted: a barocde for plants. *Science* 2007; 318(2): 190–191.

Petit RJ, Excoffier L: Gene flow and species delimitation. *Trends in Ecology and Evolution* 2009; 24(7): 386-393.

Pfenninger M, Schwenk K: Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology* 2007; 7:121.

Pignatti S: *Flora d'Italia* 1982; Italy, Bologna Ed. Agricole.

Poinar GO: Thaumamermiscosgrovei N-Gen,N-Sp(Mermithidae, Nematoda) parasitizing terrestrial isopods (Isopoda, Oniscoidea). *Syst. Parasitol*.1981; 2:261–266.

Polaszek A, Agosti D, Alonso-Zarazaga M, et al.: A universal register for animal names. *Nature* 2005; 437:477.

Pollock DD, Zwickl DJ, McGuire JA, et al.: Increased taxon sampling is advantageous for phylogenetic inference. *Syst Biol* 2002; 51:664–71.

Porazinska DL, Giblin-Davis RM, Faller L et al.: Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources* 2009; 9(6): 1439-1450.

Posada D, Crandall KA: Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol*. 2001; 16: 37-45.

Powers TO: Nematode molecular diagnostics: From bands to barcodes. *Annu Rev Phytopathology* 2004; 42:367-385.

Priemer J, Krone O, Schuster R: *Taenia krabbei* (Cestoda: Cyclophyllidea) in Germany and its delimitation from *T. ovis*. *Zoologischer Anzeiger* 2002; 241:333–337.

de Queiroz K: Species concepts and species delimitation. *Syst Biol* 2007; 56:879-886.

Rach J, DeSalle R, Sarkar IN, et al.: Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proc RSoc B* 2008; 275:237–47.

Ramazzotti G: Note sulle biocenosi dei muschi. *Mem. Ist. Ital. Idrobiol. "Dott Marco De Marchi"* 1958; 10:153-206.

Randrianiaina RD, Glaw F, Thomas M, Glos J, Raminosoa N, Vences M: Descriptions of the tadpoles of two species of Gephyromantis, with a discussion of the phylogenetic origin of direct development in mantellid frogs. *Zootaxa* 2007; 1401: 53–61.

Ratnasingham S, Hebert PDN: BOLD: The Barcode of Life Datasystem (www.barcodinglife.org). *Mol Ecol Notes* 2007; 7:355–64.

Richly E, Leister D: NUMTs in sequenced eukaryotic genomes. *MolBiolEvol* 2004; 21:1081–4.

Robson C: Family Paradoxornithidae (Parrotbills). In: Del Hoyo J, Elliot A, Christie D. (Eds.), *Handbook of Birds of the World* 2007; Lynx Edicions, Barcelona, Vol. 12. (Picathartes to Tits and Chickadees), pp. 292-321.

Rubinoff D, Cameron S,Will K: A genomic perspective on the shortcomings of mitochondrial DNA for ''barcoding'' identification. *JHered* 2006; 97:581–94.

Ruedi M, Mayer F: Molecular systematics of bats of the genus Myotis (Vespertilionidae) suggests deterministic ecomorphological convergences. *Molecular Phylogenetics and Evolution* 2001; 21: 436–448.

Ruedi M, Walter S, Fischer MC, et al.: Italy as a major Ice Age refuge area for the bat Myotis myotis (Chiroptera: Vespertilionidae) in Europe . *MolecularEcology* 2008; 17: 1801-1814.

Rusch DB, Halpern AL, Sutton G, et al.: The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *Plos Biol*. 2007; 5(3):e77, 398–431.

Sangster G: Increasing numbers of bird species result from taxonomic progress, not taxonomic inflation. *Proc R Soc Lond B* 2009; 276:3185-3191.

Sarkar IN, Joseph WT, Paul JP: An automated phylogenetic key for classifying homeoboxes. *MolPhylEvol* 2002; 24:388–99.

Sarkar IN, Paul JP, Bael TE, et al.: Characteristic attributes in cancer micro-arrays. *J Biomed Informat* 2002; 35:111–22.

Sarkar IN, Planet PJ, Desalle R: CAOS software for use in character-based DNA barcoding. *MolEcolResour* 2008;8: 1256–9.

Saunders GW:Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philos. Trans. R. Soc. Lond*. B 2005; 360: 1879–1888.

Savolainen V, Cowan RS, Vogler AP, et al.: Towards writing the encyclopedia of life: an introduction to DNA barcoding. *PhilTRoy Soc B* 2005; 360:1805–11.

Sayre RM, Brunson LK: Microfauna of moss habitats. *Am. Biol. Teach.* 1971; 33(2):100-102, 105.

Schmalfuss H: *World catalog of terrestrial isopods (Isopoda: Oniscidea)* 2004; Stuttgarter Beiträge zur Naturkunde, Serie A, Nr. 654: 341 pp.

Schlick-Steiner BC, Steiner FM, Moder K, Seifert B, Sanetra M, et al.: A multidisciplinary approach reveals cryptic diversity in Western Palearctic Tetramorium ants (Hymenoptera: Formicidae). *Mol. Phylogenet. Evol.* 2006; 40:259-273.

Schram F: The truly new systematics -- megascience in the information age. *Hydrobiologia* 2004; 519:1-7.

Seifert KA, Samson RA, deWaard JR, et al.: Prospects for fungus identification using CO1 DNA barcodes, with Penicillium as a test case. *Proc Natl Acad Sci USA* 2007;104: 3901–6.

Seutin G, White BN, Boag PT: Preservation of avian blood and tissue samples for DNA analysis. *Can. J. Zool.* 1991; 69:82–90.

Shaffer HB, Thomson RC: Delimiting species in recent radiations. *Syst Biol* 2007; 56:896-906.

Shearer TL, Coffroth MA: Barcoding corals: limited by interspecific divergence, not intraspecific variation. *Mol Ecol Resour* 2008; 8:247–255.

Shenoy BD, Jeewon R, Hyde KA: Impact of DNA sequence-data on the taxonomy of anamorphic fungi. *Fungal Div.*2007; 26: 1–54.

Simonovik B, Ivancic A, Jakse J, Bohanec B: Production and genetic evaluation of interspecific hybrids within the genus Sambucus. *Plant Breed* 2007; 126:628–633.

Sipos R, Szekely AJ, Palatinszky M, et al.: Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiol. Ecol* 2007; 60:341–350.

Smith MA, Fisher BL, Hebert PDN: DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.*2005; 360: 1825–1834.

Smith MA, Rodriguez JJ, Whitfield JB, et al.: Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105: 12359-12364.

Smagala JA, Dawson ED, Mehlmann M, et al.: ConFind: a robust tool for conserved sequence identification. *Bioinformatics* 2005; 21:4420–2.

Soininen EM, Valentini A, Coissac E, et al.: Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Front. Zool.* 2009; 6:16.

Song H, Buhay JE, Whiting MF, et al.: Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc Natl Acad SciUSA* 2008; 105:13486–91.

Song G, Qu Y, Yin Z, Li S, Liu N, Lei F: Phylogeography of the Alcippe morrisonia (Aves: Timaliidae): long population history beyond late Pleistocene glaciations. *BMC Evol. Biol*. 2009; 9:143.

Sugita T, Nishikawa A, Shinoda T: Identification of trichosporon asahii by PCR based on sequences of the internal transcribed spacer regions. *J. Clin. Microbiol.* 1998; 2742–2744.

Sungmin K, Hae-Seok E, Hyeyoung K, et al.: DNA Barcode-based molecular identification system for fish species.*Molcules and Cells* in press.

Sutton S: *Woodlice*, 1972; Invertebrate Types. Pergamon Press, Oxford. London, 144 pp.

Suzuki MT, Giovannoni SJ: Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol.* 1996; 62: 625–630.

Taberlet P, Fumagalli L, Wüst-Saucy AG, Cosson JF: Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 1998; 7:453–464.

Taberlet P., et al: Power and limitations of the chloroplast trnL(UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 2007; 35:e14.

Tautz D, Arctander P, Minelli A, et al.: A plea for DNA taxonomy. *Trends Ecol Evol* 2003; 18:70–4.

Templeton AR, Crandall KA, Sing CF: A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 1992; 132:619–633.

Thiele K, Yeates D: Tension arises from duality at the heart of taxonomy – names must both represent a volatile hypothesis and provide a key to lasting information. *Nature* 2002; 419:337–337.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997; 24:4876-4882.

Tinggi U: Essentiality and toxicity of selenium and its status in Australia: a review. *Toxicol Lett* 2003;137:103–110.

Trindade H.: Molecular biology of aromatic plants and spices. *Flavour and Fragrance Journal* 2010; DOI 10.1002/ffj .1974.

Turbeville JM, Pfeifer DM, Field KG, Raff RA: The phylogenetic status of arthropods, as inferred from 18s rRNA sequences. *Mol. Biol. Evol*. 1991; 8: 669-686.

Tucker AO: The truth about mints. *Herb Companion* 1992; 4: 51-52.

Uni S, Suzuki Y, Baba M, Mitani N, Takaoka H, Katsumi A, Bain O: Coexistence of five Cercopithifilaria species in the Japanese rupricaprine bovid, Capricornis crispus. *Parasite* 2001; 8:197-213.

Unwin R, Maiden MCJ: Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol* 2003; 11:479–87.

Valentini A, Miquel C, Nawaz MA, et al.: New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Mol. Ecol. Res*. 2008; 9:51–60.

Valentini A, Pompanon F, Taberlet P: DNA barcoding for ecologists. *Trends Ecol Evol* 2009a; 24:110–7.

Valentini A, Pegard A, Miquel C, et al.: Universal DNA-based methods for assessing the diet of grazing livestock and wildlife from feces. *J Agric Food Chem.* 2009; 57(13):5700-6.

Valentini A, Soinen EM, Coissac E, et al.: Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Frontiers in Zoology* 2009; 6:16.

Vences M, Thomas M, Bonett RM, Vieites DR: Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philos Trans Biol Sci* 2005; 360:1859-1868.

Vieites DR, Wollenberg KC, Andreone F, Köhler J, Glaw F, Vences M: Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proc Nat Acad Sci USA* 2009; 106:8267-8272.

Virgilio et al., 2010

Virgilio M, Backeljau T, Nevado B, et al.: RCesoeamrchp artaicrleative performances of DNA barcoding across insect orders *BMC Bioinformatics* 2010; 11:206.

Vogler AP, Monaghan MT: Recent advances in DNA taxonomy. *JZoolSystEvolRes* 2006; 45:1–10.

Ward RD, Zemlak TS, Innes BH, et al.: DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B* 2005; 360: 1847–1857.

Ward RD, Holmes BH, White W, Last PR: DNA barcoding Australasian chondrichthyans results and potential uses in conservation. *Marine and Freshwater Research* 2008; 59, 57–71.

Waugh J: DNA barcoding in animal species: progress, potential and pitfalls. BioEssays 2007;29:188–97.

Wiemers M, Fiedler K: Does the DNA barcoding gap exist? – A case study in blue butterflies (Lepidoptera: Lycaenidae). *Front Zool* 2005; 4:8.

Will KW, Mishler BD, Wheeler QD. The perils of DNA barcoding and the need for integrative taxonomy. *Syst Biol* 2005;54:844–51.

Will KW, Rubinoff D: Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 2004;20:47–55.

Wirth, T., Le Guellec, R., Veuille, M., 1999. Directional substitution and evolution of nucleotide content in the cytochrome oxidase II gene in earwigs (Dermapteran Insects). *Mol. Biol. Evol*.2005; 16 (12): 1645–1653.

Wong E H K, Hanner RH. DNA barcoding detects market substitution in North American seafood. *Food Research International* 2008; 41: 828–837.

Zhang AB, He LJ, Crozier RH, et al.: Estimating sample sizes for DNA barcoding. *Mol Phylogenet Evol* 2009, in press.

Zhou J, Davey ME, Figueras JB, et al.: Phylogenetic diversity of a bacterial community determined from Siberian tundra soil DNA. *Microbiology* 1997; 143, 3913–3919.

Zullini A: I nematodi muscicoli della Val Zebrù (Parco Nazionale dello Stelvio). *Istituto Lombardo di Scienze e Lettere (Rend. Sc.)* 1970; B 104: 88- 137.

Zwickl DJ, Hillis DM: Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol* 2002; 51:588–98.

APPENDIX I.1: List of Italian bats samples (case study A1) and *coxI* sequences retrieved from GenBank used in this study with reference to specimen voucher, previous species attribution (whenever possible), sampling locality (with locality Group as NI, CI, SI, Sar, IR indicating Northern, Central, Southern Italian regions, Sardinia and Ireland respectively). Samples included in the reference dataset are highlighted with bold characters.

Voucher	Species	Family	Locality	Locality group
MIB:ZPL:01485	*Barbastella barbastellus* (Schreber, 1774)	Vespertilionidae	Firenze (FI)	CI
MIB:ZPL:01264	*Barbastella barbastellus* (Schreber, 1774)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:01300	*Barbastella barbastellus* (Schreber, 1774)	Vespertilionidae	Riserva Naturale Monte Navegna (RI)	CI
MIB:ZPL:02257	*Eptesicus nilsonii* (Keyserling & Blasius, 1839)	Vespertilionidae	Chiareggio (SO)	NI
MIB:ZPL:02258	*Eptesicus nilsonii* (Keyserling & Blasius, 1839)	Vespertilionidae	Chiareggio (SO)	NI
MIB:ZPL:00312	*Eptesicus serotinus* (Schreber, 1774)	Vespertilionidae	Saronno (VA)	NI
MIB:ZPL:00333	*Eptesicus serotinus* (Schreber, 1774)	Vespertilionidae	Saronno (VA)	NI
MIB:ZPL:00349	*Eptesicus serotinus* (Schreber, 1774)	Vespertilionidae	Saronno (VA)	NI
MIB:ZPL:00369	*Eptesicus serotinus* (Schreber, 1774)	Vespertilionidae	Saronno (VA)	NI
MIB:ZPL:01342	*Eptesicus serotinus* (Schreber, 1774)	Vespertilionidae	Saronno (VA)	NI
MIB:ZPL:01232	*Hypsugo savii* (Bonaparte, 1837)	Vespertilionidae	Abruzzo	CI
MIB:ZPL:00254	*Hypsugo savii* (Bonaparte, 1837)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:00321	*Hypsugo savii* (Bonaparte, 1837)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:ZPL:01267	*Hypsugo savii* (Bonaparte, 1837)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01270	*Hypsugo savii* (Bonaparte, 1837)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01243	*Hypsugo savii* (Bonaparte, 1837)	Vespertilionidae	Duchessa (RI)	CI
MIB:ZPL:01295	*Hypsugo savii* (Bonaparte, 1837)	Vespertilionidae	Riserva Naturale Monte Navegna (RI)	CI
MIB:ZPL:00281	*Miniopterus schreibersii* (Kuhl, 1817)	Miniopteridae	Onferno (RN)	NI
MIB:ZPL:00286	*Miniopterus schreibersii* (Kuhl, 1817)	Miniopteridae	Onferno (RN)	NI
MIB:ZPL:00290	*Miniopterus schreibersii* (Kuhl, 1817)	Miniopteridae	Onferno (RN)	NI
MIB:ZPL:00292	*Miniopterus schreibersii* (Kuhl, 1817)	Miniopteridae	Onferno (RN)	NI
MIB:ZPL:01335	*Miniopterus schreibersii* (Kuhl, 1817)	Miniopteridae	Onferno (RN)	NI
GU270560*	*Myotis alcathoe* von Helversen & Heller, 2001	Vespertilionidae	_	Ireland
MIB:ZPL:01278	*Myotis alcathoe* von Helversen & Heller, 2001	Vespertilionidae	Cilento (SA)	SI
GU270562*	*Myotis bechsteinii* (Kuhl, 1817)	Vespertilionidae	_	Ireland
GU270563*	*Myotis bechsteinii* (Kuhl, 1817)	Vespertilionidae	_	Ireland
MIB:ZPL:00314	*Myotis bechsteinii* (Kuhl, 1817)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:00319	*Myotis bechsteinii* (Kuhl, 1817)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:00364	*Myotis bechsteinii* (Kuhl, 1817)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:ZPL:01271	*Myotis bechsteinii* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01283	*Myotis bechsteinii* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01327	*Myotis bechsteinii* (Kuhl, 1817)	Vespertilionidae	Umbria	CI
MIB:ZPL:01218	*Myotis blythii* (Tomes, 1857)	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01219	*Myotis blythii* (Tomes, 1857)	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01231	*Myotis blythii* (Tomes, 1857)	Vespertilionidae	Abruzzo	CI
MIB:ZPL:00499	*Myotis blythii* (Tomes, 1857)	Vespertilionidae	Buco del Frate (VA)	NI
MIB:ZPL:01178	*Myotis blythii* (Tomes, 1857)	Vespertilionidae	Buco del Frate (VA)	NI
MIB:ZPL:01246	*Myotis blythii* (Tomes, 1857)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:01263	*Myotis blythii* (Tomes, 1857)	Vespertilionidae	Matese (CE)	SI

Voucher	Species	Family	Locality	Locality group
MIB:ZPL:00285	*Myotis blythii* (Tomes, 1857)	Vespertilionidae	Onferno (RN)	NI
GU270564*	*Myotis brandtii* (Eversmann, 1845)	Vespertilionidae	_	Ireland
GU270565*	*Myotis brandtii* (Eversmann, 1845)	Vespertilionidae	_	Ireland
MIB:ZPL:01279	*Myotis brandtii* (Eversmann, 1845)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01176	*Myotis capaccinii* (Bonaparte, 1837)	Vespertilionidae	Buco del Frate (VA)	NI
MIB:ZPL:01179	*Myotis capaccinii* (Bonaparte, 1837)	Vespertilionidae	Buco del Frate (VA)	NI
MIB:ZPL:01181	*Myotis capaccinii* (Bonaparte, 1837)	Vespertilionidae	Buco del Frate (VA)	NI
GU270554*	*Myotis daubentonii* (Kuhl, 1817)	Vespertilionidae	_	Ireland
MIB:ZPL:00341	*Myotis daubentonii* (Kuhl, 1817)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:01344	*Myotis daubentonii* (Kuhl, 1817)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:00311	*Myotis daubentonii* (Kuhl, 1817)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:ZPL:00313	*Myotis daubentonii* (Kuhl, 1817)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:ZPL:00299	*Myotis daubentonii* (Kuhl, 1817)	Vespertilionidae	Valganna (VA)	NI
GU270553*	*Myotis emarginatus* (Geoffroy, 1806)	Vespertilionidae	_	Ireland
MIB:ZPL:00267	*Myotis emarginatus* (Geoffroy, 1806)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:00315	*Myotis emarginatus* (Geoffroy, 1806)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:ZPL:00322	*Myotis emarginatus* (Geoffroy, 1806)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:ZPL:01294	*Myotis emarginatus* (Geoffroy, 1806)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01242	*Myotis emarginatus* (Geoffroy, 1806)	Vespertilionidae	Duchessa (RI)	CI
MIB:ZPL:01247	*Myotis emarginatus* (Geoffroy, 1806)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:01325	*Myotis emarginatus* (Geoffroy, 1806)	Vespertilionidae	Umbria	CI
MIB:ZPL:01234	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01248	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:01255	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:01259	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:01265	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:00275	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Onferno (RN)	NI
MIB:ZPL:00280	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Onferno (RN)	NI
MIB:ZPL:00282	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Onferno (RN)	NI
MIB:ZPL:00284	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Onferno (RN)	NI
MIB:ZPL:00287	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Onferno (RN)	NI
MIB:ZPL:00300	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Onferno (RN)	NI
MIB:ZPL:00338	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Onferno (RN)	NI
MIB:ZPL:00377	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Onferno (RN)	NI
MIB:ZPL:01336	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Onferno (RN)	NI
MIB:ZPL:01339	*Myotis myotis* (Borkhausen, 1797)	Vespertilionidae	Onferno (RN)	NI
GU270555*	*Myotis mystacinus* (Kuhl, 1817)	Vespertilionidae	_	Ireland
GU270556*	*Myotis mystacinus* (Kuhl, 1817)	Vespertilionidae	_	Ireland
GU270557*	*Myotis mystacinus* (Kuhl, 1817)	Vespertilionidae	_	Ireland
GU270558*	*Myotis mystacinus* (Kuhl, 1817)	Vespertilionidae	_	Ireland
GU270559*	*Myotis mystacinus* (Kuhl, 1817)	Vespertilionidae	_	Ireland
MIB:ZPL:01240	*Myotis mystacinus* (Kuhl, 1817)	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01293	*Myotis mystacinus* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:02290	*Myotis mystacinus* (Kuhl, 1817)	Vespertilionidae	Coghinas (SS)	SAR
MIB:ZPL:02289	*Myotis mystacinus* (Kuhl, 1817)	Vespertilionidae	Rio Calaresu (NU)	SAR

Voucher	Species	Family	Locality	Locality group
MIB:ZPL:03775	*Myotis mystacinus* (Kuhl, 1817)	Vespertilionidae	Robecco sul Naviglio (MI)	NI
GU270561*	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	_	Ireland
MIB:ZPL:01233	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Abruzzo	CI
MIB:Zpl:00505	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Buco del Frate (BS)	NI
MIB:ZPL:00356	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:Zpl:00383	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:00318	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:ZPL:00326	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:ZPL:00331	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:Zpl:00384	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:ZPL:01347	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:Zpl:01273	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01282	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:Zpl:01284	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:Zpl:01285	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:Zpl:01286	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:Zpl:01291	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:Zpl:01292	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:Zpl:01308	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:Zpl:01309	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01249	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Matese (CE)	SI
MIB:Zpl:01266	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:01324	*Myotis nattereri* (Kuhl, 1817)	Vespertilionidae	Veio (Roma)	CI
MIB:ZPL:01566	*Myotis punicus* (Felten, 1977)	Vespertilionidae	Sardegna	SAR
MIB:ZPL:01567	*Myotis punicus* (Felten, 1977)	Vespertilionidae	Sardegna	SAR
MIB:ZPL:01568	*Myotis punicus* (Felten, 1977)	Vespertilionidae	Sardegna	SAR
MIB:ZPL:01211	*Myotis sp.*	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01214	*Myotis sp.*	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01216	*Myotis sp.*	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01221	*Myotis sp.*	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01222	*Myotis sp.*	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01223	*Myotis sp.*	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01228	*Myotis sp.*	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01230	*Myotis sp.*	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01235	*Myotis sp.*	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01281	*Myotis sp.*	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01287	*Myotis sp.*	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01289	*Myotis sp.*	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01301	*Myotis sp.*	Vespertilionidae	Duchessa (RI)	CI
MIB:ZPL:01302	*Myotis sp.*	Vespertilionidae	Duchessa (RI)	CI
MIB:ZPL:01303	*Myotis sp.*	Vespertilionidae	Duchessa (RI)	CI
MIB:ZPL:01256	*Myotis sp.*	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:01319	*Myotis sp.*	Vespertilionidae	Matese (CE)	SI
GU270566*	*Nyctalus leisleri* (Kuhl, 1817)	Vespertilionidae	_	Ireland?
MIB:ZPL:01207	*Nyctalus leisleri* (Kuhl, 1817)	Vespertilionidae	Abruzzo	CI

Voucher	Species	Family	Locality	Locality group
MIB:ZPL:01268	*Nyctalus leisleri* (Kuhl, 1817)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:00261	*Nyctalus leisleri* (Kuhl, 1817)	Vespertilionidae	Malcantone (CH)	SW
MIB:ZPL:00535	*Nyctalus leisleri* (Kuhl, 1817)	Vespertilionidae	Malcantone (CH)	SW
MIB:ZPL:00536	*Nyctalus leisleri* (Kuhl, 1817)	Vespertilionidae	Malcantone (CH)	SW
MIB:ZPL:01323	*Nyctalus leisleri* (Kuhl, 1817)	Vespertilionidae	Umbria	CI
MIB:ZPL:02259	*Nyctalus leisleri* (Kuhl, 1817)	Vespertilionidae	Val Bodengo (SO)	NI
MIB:ZPL:01476	*Nyctalus noctula* (Schreber, 1774)	Vespertilionidae	Cervia (RA)	NI
MIB:ZPL:00240	*Pipistrellus kuhlii* (Kuhl, 1817)	Vespertilionidae	_	NI
MIB:ZPL:00289	*Pipistrellus kuhlii* (Kuhl, 1817)	Vespertilionidae	_	NI
MIB:ZPL:00253	*Pipistrellus kuhlii* (Kuhl, 1817)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:01251	*Pipistrellus kuhlii* (Kuhl, 1817)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:01297	*Pipistrellus kuhlii* (Kuhl, 1817)	Vespertilionidae	Riserva Naturale Monte Navegna (RI)	CI
MIB:ZPL:01312	*Pipistrellus kuhlii* (Kuhl, 1817)	Vespertilionidae	Umbria	CI
MIB:ZPL:01322	*Pipistrellus kuhlii* (Kuhl, 1817)	Vespertilionidae	Veio (Roma)	CI
MIB:ZPL:01180	*Pipistrellus nathusii* (Keyserling & Blasius, 1839)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:01328	*Pipistrellus nathusii* (Keyserling & Blasius, 1839)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:01331	*Pipistrellus nathusii* (Keyserling & Blasius, 1839)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:01334	*Pipistrellus nathusii* (Keyserling & Blasius, 1839)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:01224	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01195	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Campo Franscia (SO)	NI
MIB:ZPL:00272	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Castelseprio (VA)	NI
MIB:ZPL:00283	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Castelseprio (VA)	NI
MIB:ZPL:00288	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Castelseprio (VA)	NI
MIB:ZPL:00291	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Castelseprio (VA)	NI
MIB:ZPL:01337	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Castelseprio (VA)	NI
MIB:ZPL:01290	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:02253	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Gravedona (CO)	NI
MIB:ZPL:02254	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Gravedona (CO)	NI
MIB:ZPL:02255	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Gravedona (CO)	NI
MIB:ZPL:01321	*Pipistrellus pipistrellus* (Schreber, 1774)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:02287	*Pipistrellus pygmaeus* (Leach, 1825)	Vespertilionidae	Seui (OG)	SAR
MIB:ZPL:02285	*Pipistrellus pygmaeus* (Leach, 1825)	Vespertilionidae	Sorso (SS)	SAR
MIB:ZPL:01239	*Pipistrellus sp.*	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01241	*Pipistrellus sp.*	Vespertilionidae	Abruzzo	CI
MIB:ZPL:02288	*Pipistrellus sp.*	Vespertilionidae	Calagonone (NU)	SAR
MIB:ZPL:03815	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03816	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03817	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03818	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03819	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03820	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03821	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03822	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03823	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03824	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI

Voucher	Species	Family	Locality	Locality group
MIB:ZPL:03825	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03826	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03827	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:03828	*Pipistrellus sp.*	Vespertilionidae	Bosco della Fontana (MN)	NI
MIB:ZPL:01206	*Plecotus auritus* (Linnaeus, 1758)	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01227	*Plecotus auritus* (Linnaeus, 1758)	Vespertilionidae	Abruzzo	CI
MIB:ZPL:01345	*Plecotus auritus* (Linnaeus, 1758)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:01346	*Plecotus auritus* (Linnaeus, 1758)	Vespertilionidae	Parco del Campo dei Fiori (VA)	NI
MIB:ZPL:01190	*Plecotus auritus* (Linnaeus, 1758)	Vespertilionidae	Cariadeghe (BS)	NI
MIB:ZPL:01269	*Plecotus auritus* (Linnaeus, 1758)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01276	*Plecotus auritus* (Linnaeus, 1758)	Vespertilionidae	Cilento (SA)	SI
MIB:ZPL:01253	*Plecotus auritus* (Linnaeus, 1758)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:00378	*Plecotus auritus* (Linnaeus, 1758)	Vespertilionidae	San Martino (LC)	NI
MIB:ZPL:01497	*Plecotus austriacus* (Fischer, 1829)	Vespertilionidae	Firenze (FI)	CI
MIB:ZPL:01252	*Plecotus austriacus* (Fischer, 1829)	Vespertilionidae	Matese (CE)	SI
MIB:ZPL:00256	*Plecotus macrobullaris* (Kusjakin, 1965)	Vespertilionidae	(TN)	NI
MIB:ZPL:01200	*Plecotus macrobullaris* (Kusjakin, 1965)	Vespertilionidae	Campo Moro (SO)	NI
MIB:ZPL:01572	*Plecotus sardus* Mucedda & Kiefer, 2002	Vespertilionidae	Sardegna	SAR
MIB:ZPL:01574	*Plecotus sardus* Mucedda & Kiefer, 2002	Vespertilionidae	Sardegna	SAR
MIB:ZPL:01575	*Plecotus sardus* Mucedda & Kiefer, 2002	Vespertilionidae	Sardegna	SAR
MIB:ZPL:03414	*Plecotus sp.*	Vespertilionidae	"Bocca di Lorenza" Santorso (VI)	NI
MIB:ZPL:00262	*Plecotus sp.*	Vespertilionidae	(TN)	NI
MIB:ZPL:01189	*Plecotus sp.*	Vespertilionidae	Campo Moro (SO)	NI
MIB:ZPL:00265	*Plecotus sp.*	Vespertilionidae	Monte Pravello (VA)	NI
MIB:ZPL:00268	*Plecotus sp.*	Vespertilionidae	Monte Pravello (VA)	NI
MIB:ZPL:00269	*Plecotus sp.*	Vespertilionidae	Monte San Martino (LC)	NI
MIB:ZPL:00270	*Plecotus sp.*	Vespertilionidae	Monte San Martino (LC)	NI
MIB:ZPL:00271	*Rhinolophus euryale* Blasius, 1853	Rhinolophidae	Onferno (RN)	NI
MIB:ZPL:00274	*Rhinolophus euryale* Blasius, 1853	Rhinolophidae	Onferno (RN)	NI
MIB:ZPL:00295	*Rhinolophus euryale* Blasius, 1853	Rhinolophidae	Onferno (RN)	NI
MIB:ZPL:00310	*Rhinolophus euryale* Blasius, 1853	Rhinolophidae	Onferno (RN)	NI
MIB:ZPL:01341	*Rhinolophus euryale* Blasius, 1853	Rhinolophidae	Onferno (RN)	NI
MIB:ZPL:00502	*Rhinolophus ferrumequinum* (Schreber, 1774)	Rhinolophidae	Buco del Frate (VA)	NI
MIB:ZPL:00316	*Rhinolophus ferrumequinum* (Schreber, 1774)	Rhinolophidae	Onferno (RN)	NI
MIB:ZPL:00354	*Rhinolophus ferrumequinum* (Schreber, 1774)	Rhinolophidae	Onferno (RN)	NI
MIB:ZPL:00376	*Rhinolophus ferrumequinum* (Schreber, 1774)	Rhinolophidae	Onferno (RN)	NI
MIB:ZPL:00393	*Rhinolophus ferrumequinum* (Schreber, 1774)	Rhinolophidae	Onferno (RN)	NI
MIB:ZPL:01204	*Rhinolophus ferrumequinum* (Schreber, 1774)	Rhinolophidae	San Cesario (MO)	NI
MIB:ZPL:00327	*Rhinolophus ferrumequinum* (Schreber, 1774)	Rhinolophidae	Saronno (VA)	NI
MIB:ZPL:00359	*Rhinolophus hipposideros* (Bechstein, 1800)	Rhinolophidae	Onferno (RN)	NI
MIB:ZPL:00387	*Rhinolophus hipposideros* (Bechstein, 1800)	Rhinolophidae	Onferno (RN)	NI
MIB:ZPL:01559	*Rhinolophus mehelyi* Matschie, 1901	Rhinolophidae	Sardegna	SAR
MIB:ZPL:01560	*Rhinolophus mehelyi* Matschie, 1901	Rhinolophidae	Sardegna	SAR
MIB:ZPL:01561	*Rhinolophus mehelyi* Matschie, 1901	Rhinolophidae	Sardegna	SAR
MIB:ZPL:01562	*Rhinolophus mehelyi* Matschie, 1901	Rhinolophidae	Sardegna	SAR

Voucher	Species	Family	Locality	Locality group
MIB:ZPL:01563	*Rhinolophus mehelyi* Matschie, 1901	Rhinolophidae	Sardegna	SAR
MIB:ZPL:01512	*Tadarida teniotis* (Rafinesque, 1814)	Molossidae	Pizzo, Calabria (VV)	SI
MIB:ZPL:01569	*Tadarida teniotis* (Rafinesque, 1814)	Molossidae	Sardegna	SAR
MIB:ZPL:01570	*Tadarida teniotis* (Rafinesque, 1814)	Molossidae	Sardegna	SAR

APPENDIX I.2: List of Paradoxornis samples used in this study with reference to specimen voucher (when available), previous morphotype and subspecies attribution, GenBank accession numbers of newly identified sequences, sampling locality (with latitude and longitude) and assigned group name. The haplotype name and lineage **m**embership in the network reconstruction are given for each sample.

Sample	Voucher	Morphotype	Subspecies	GenBank Accession No.				Haplotype	Lineage	Sampling Locality	Latitude	Longitude	Group name
				Cyt b	coxI	12S rDNA	16S rDNA						
VTP15	MIB:zpl:01141	*P. webbianus*	*suffusus*	FN552175	FN552143	FN552213	FN552245	E09	I	Dongzhai NR., Henan, China	31°56'30" N	114°10'56" E	Henan
VTP13	MIB:zpl:01137	*P. webbianus*	*suffusus*	FN552178	FN552146	FN552216	FN552248	E05	I	Dongzhai NR., Henan, China	31°56'30" N	114°10'56" E	Henan
VTP12	MIB:zpl:01139	*P. webbianus*	*suffusus*	FN552181	FN552149	FN552219	FN552251	E07	I	Dongzhai NR., Henan, China	31°56'30" N	114°10'56" E	Henan
VTP14	MIB:zpl:01140	*P. webbianus*	*suffusus*	FN552182	FN552150	FN552220	FN552252	E08	I	Dongzhai NR., Henan, China	31°56'30" N	114°10'56" E	Henan
AN13355_D10	MIB:zpl:01125	*P. webbianus*	_	FN552152	FN552120	FN552190	FN552222	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
AS22871_ID07	MIB:zpl:01126	*P. webbianus*	_	FN552153	FN552121	FN552191	FN552223	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
AS22878_IH16	MIB:zpl:01127	unidentified	_	FN552154	FN552122	FN552192	FN552224	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
AT27544_H17	MIB:zpl:01122	*P. alphonsianus*	_	FN552155	FN552123	FN552193	FN552225	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
AV73538_20	MIB:zpl:01123	*P. webbianus*	_	FN552156	FN552124	FN552194	FN552226	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
AV73544_H08	MIB:zpl:01124	*P. alphonsianus*	_	FN552157	FN552125	FN552195	FN552227	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
AV73547_37	MIB:zpl:01120	*P. webbianus*	_	FN552158	FN552126	FN552196	FN552228	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
AV73550_41	MIB:zpl:01121	unidentified	_	FN552159	FN552127	FN552197	FN552229	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
AV73635_H09	MIB:zpl:01128	*P. alphonsianus*	_	FN552160	FN552128	FN552198	FN552230	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
D01	MIB:zpl:01129	*P. alphonsianus*	_	FN552161	FN552129	FN552199	FN552231	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
ID05	MIB:zpl:01130	*P. webbianus*	_	FN552162	FN552130	FN552200	FN552232	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
ID06	MIB:zpl:01131	unidentified	_	FN552163	FN552131	FN552201	FN552233	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
ID08	MIB:zpl:01132	*P. alphonsianus*	_	FN552164	FN552132	FN552202	FN552234	E01	I	Palude Brabbia NR, Varese, NW Italy	45°47'24" N	8°42'34" E	Italy
web-s-3158	_	*P. webbianus*	*suffusus*	FN552186	_	_	_	_	_	Qinling Mt., Shaanxi, China	33°31'30" N	107°59'27" E	Shaanxi
web-s-3159	_	*P. webbianus*	*suffusus*	FN552187	_	_	_	_	_	Qinling Mt., Shaanxi, China	33°31'30" N	107°59'27" E	Shaanxi
web-s-834	_	*P. webbianus*	*suffusus*	FN552188	_	_	_	_	_	Qinling Mt., Shaanxi, China	33°31'30" N	107°59'27" E	Shaanxi
web-s-836	_	*P. webbianus*	*suffusus*	FN552189	_	_	_	_	_	Qinling Mt.,Shaanxi, China	33°31'30" N	107°59'27" E	Shaanxi
VTP4	MIB:zpl:01135	*P. webbianus*	*webbianus*	FN552174	FN552142	FN552212	FN552244	E03	I	Dongtan,Chongming Island,Shanghai, China	31°31'44" N	121°51'48" E	Shanghai
VTP1	MIB: zpl:02279	*P. webbianus*	*webbianus*	FN552168	FN552136	FN552206	FN552238	E17	II	Dongtan,Chongming Island,Shanghai, China	31°31'44" N	121°51'48" E	Shanghai
VTP2	MIB: zpl:02280	*P. webbianus*	*webbianus*	FN552169	FN552137	FN552207	FN552239	E17	II	Dongtan,Chongming Island,Shanghai, China	31°31'44" N	121°51'48" E	Shanghai
VTP22	MIB: zpl:02276	*P. alphonsianus*	*alphonsianus*	FN552165	FN552133	FN552203	FN552235	E12	I	Dunjiangyan, Sichuan, China	31°00'22" N	103°37'03" E	Sichuan
VTP21	MIB: zpl:01133	*P. alphonsianus*	*alphonsianus*	FN552172	FN552140	FN552210	FN552242	E02	I	Dunjiangyan, Sichuan, China	31°00'22" N	103°37'03" E	Sichuan
VTP24	MIB: zpl:01136	*P. alphonsianus*	*alphonsianus*	FN552176	FN552144	FN552214	FN552246	E04	I	Dunjiangyan, Sichuan, China	31°00'22" N	103°37'03" E	Sichuan
VTP25	MIB: zpl:02283	*P. alphonsianus*	*alphonsianus*	FN552177	FN552145	FN552215	FN552247	E11	I	Dunjiangyan, Sichuan, China	31°00'22" N	103°37'03" E	Sichuan
VTP23	MIB: zpl:01138	*P. alphonsianus*	*alphonsianus*	FN552183	FN552151	FN552221	FN552253	E06	I	Dunjiangyan, Sichuan, China	31°00'22" N	103°37'03" E	Sichuan
VTP5	MIB: zpl:01134	*P. webbianus*	*fulvicauda*	FN552173	FN552141	FN552211	FN552243	E02	I	East Tianjin, China	38°44'21" N	117°25'05" E	Tianjin
VTP11	MIB: zpl:02284	*P. webbianus*	*fulvicauda*	FN552179	FN552147	FN552217	FN552249	E10	I	East Tianjin, China	38°44'21" N	117°25'05" E	Tianjin
VTP6	MIB: zpl:01142	*P. webbianus*	*fulvicauda*	FN552180	FN552148	FN552218	FN552250	E10	I	East Tianjin, China	38°44'21" N	117°25'05" E	Tianjin
VTP9	MIB: zpl:02277	*P. webbianus*	*fulvicauda*	FN552166	FN552134	FN552204	FN552236	E15	II	East Tianjin, China	38°44'21" N	117°25'05" E	Tianjin
VTP8	MIB: zpl:02278	*P. webbianus*	*fulvicauda*	FN552167	FN552135	FN552205	FN552237	E13	II	East Tianjin, China	38°44'21" N	117°25'05" E	Tianjin
VTP7	MIB: zpl:02281	*P. webbianus*	*fulvicauda*	FN552170	FN552138	FN552208	FN552240	E16	II	East Tianjin, China	38°44'21" N	117°25'05" E	Tianjin
VTP10	MIB: zpl:02282	*P. webbianus*	*fulvicauda*	FN552171	FN552139	FN552209	FN552241	E14	II	East Tianjin, China	38°44'21" N	117°25'05" E	Tianjin
web-a-1652	_	*P. alphonsianus*	*yunnanensis*	FN552184	_	_	_	_	_	suburb of Kunming, Yunnan, China	25°7'4.7" N	102°31'0.5" E	Yunnan
web-a-1634	_	*P. alphonsianus*	*yunnanensis*	FN552185	_	_	_	_	_	suburb of Kunming, Yunnan, China	25°7'4.7" N	102°31'0.5" E	Yunnan

APPENDIX I.3: : List of terrestrial isopods samples (case study A3) and *coxI* sequences retrieved from GenBank used in this study with reference to specimen voucher and previous species attribution.

GENUS	SPECIES	ACCESSION NUMBER
Armadillidium	*lobocurvum*	EF027679; EF027680; EF027681; EF027682; EF027683; EF027684; EF027685; EF027686; EF027687; EF027688; EF027689; EF027692; EF027693; EF027694; EF027695; EF027696; EF027697; EF027698; EF027699; EF027700; EF027702; EF027704; EF027705; EF027706; EF027714; EF027715; EF027716
	nasatum	**MIB:Zpl:01612; MIB:Zpl:01613; MIB:Zpl:01614**
	vulgare	**MIB:Zpl:01626; MIB:Zpl:01627; MIB:Zpl:01628; MIB:Zpl:01629; MIB:Zpl:01630**
Armadillo	*officinalis*	**MIB:Zpl:01621; MIB:Zpl:01622; MIB:Zpl:01623; MIB:Zpl:01624; MIB:Zpl:01625**
Armadilloniscus	*ellipticus*	EU364630
Deto	*marina*	EU364625
Ligia	*italica*	DQ182858; DQ182859; DQ182860; DQ182861
	occidentalis	AF255780
	oceanica	DQ442914; NC008412
Laevophiloscia	*yalgooensis*	EU364629
Philoscia	*affinis*	**MIB:Zpl:01429**
Porcellio	*scaber*	DQ305142
	spinicornis	DQ889123
	baidensis	**MIB:Zpl:01607; MIB:Zpl:01608; MIB:Zpl:01609; MIB:Zpl:01610; MIB:Zpl:01611**
	hyblaeus	**MIB:Zpl:01592; MIB:Zpl:01593; MIB:Zpl:01595**
	imbutus	**MIB:Zpl:01422; MIB:Zpl:01423; MIB:Zpl:01433**
	laevis	**MIB:Zpl:01586; MIB:Zpl:01587; MIB:Zpl:01588; MIB:Zpl:01589; MIB:Zpl:01590**
	siculoccidentalis	**MIB:Zpl:01604**
Porcellionides	*myrmecophilus*	**MIB:Zpl:01615;MIB:Zpl:01616; MIB:Zpl:01617; MIB:Zpl:01618; MIB:Zpl:01619; MIB:Zpl:01620**
	pruinosus	**MIB:Zpl:01596; MIB:Zpl:01597; MIB:Zpl:01598; MIB:Zpl:01599; MIB:Zpl:01600; MIB:Zpl:01601**
Trichoniscus	*pusillus*	**MIB:Zpl:01431**
Tylos	*neozelanicus*	EU364624
	ponticus	EF027454; EF027455
Haloniscus	*anophthalmus*	EU364626
	longiantennatus	EU364578; EU364583; EU364584; EU364585
	searlei	EU364616; EU364617; EU364618; EU364619; EU364620; EU364621

APPENDIX I.4: List of species including biological data, accession numbers and sub-dataset assignment (A: reference *coxI* dataset; B: reference *coxI* sub-dtaset; C: reference 12S rDNA sub-dataset) of the filarioid nematodes (case study B1) included in this study. Where available, place of collection and hosts are indicated (n.d.: no data available). * The host species indicated as *Naemorhedus crispus* is synonymous to *Capricornis crispus*. ** Laboratory strain in European laboratories since 1970s. *** Collected from a patient travelling from Camerun. **** Collected from a patient travelling from India.

Species	Museum ID	Voucher	Accession number *coxI*	Accession number 12s rDNA	Dataset	Host	Locality	Produced in this study
***Acanthocheilonema reconditum* (Grassi, 1890)**	-	-	AJ544876	AJ544853	A,B,C	*Canis lupus familiaris*	-	-
***Acanthocheilonema viteae* (Krepkogorskaya, 1933)**	-	-	AJ272117	AJ544852	A,B,C	*Meriones libycus*	-	-
***Brugia malayi* (Brug, 1927)**	-	-	AJ271610	AJ544843	A,B,C	*Homo sapiens*	-	-
***Brugia malayi* (Brug, 1927)**	-	-	AF538716	AF538716	A,B,C	n.d.	-	-
***Brugia pahangi* (Buckley & Edeson, 1956)**	-	-	EF406112	-	A,B	*Homo sapiens*	-	-
***Brugia pahangi* (Buckley & Edeson, 1956)**	-	-	DQ977746	-	B	*Meriones unguiculatus*	-	-
***Brugia pahangi* (Buckley & Edeson, 1956)**	-	-	AJ271611	AJ544842	A,B,C	*Felis catus*	-	-
***Cercopithifilaria bulboidea* Uni & Bain, 2001**	C1-3	MIB:Zpl:00940	AM749247	AM779779	A,B,C	*Naemorhedus crispus* *	Japan, Gifu	YES
***Cercopithifilaria bulboidea* Uni & Bain, 2001**	C1-4A	MIB:Zpl:00942	AM749248	AM779780	A,B,C	*Naemorhedus crispus* *	Japan, Gifu	YES
***Cercopithifilaria bulboidea* Uni & Bain, 2001**	Isolate Gifu-111	-	AB178834	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria bulboidea* Uni & Bain, 2001**	Isolate Gifu-133	-	AB178835	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria bulboidea* Uni & Bain, 2001**	Isolate SW3-FL7	-	AB178836	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria bulboidea* Uni & Bain, 2001**	Isolate SW3-FL8	-	AB178837	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria bulboidea* Uni & Bain, 2001**	Isolate SW3-MB1	-	AB178838	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria bulboidea* Uni & Bain, 2001**	Isolate SW3-UA1	-	AB178839	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria crassa* Uni, Bain & Takaoka, 2002**	Isolate S15-097	-	AB178840	-	A,B	*Cervus nippon*	-	-
***Cercopithifilaria crassa* Uni, Bain & Takaoka, 2002**	Isolate S15-101	-	AB178841	-	A,B	*Cervus nippon*	-	-
***Cercopithifilaria crassa* Uni, Bain & Takaoka, 2002**	S51-PB6	MIB:Zpl:00925	AM749260	AM779791	A,B,C	*Cervus nippon*	Japan, Oita	YES
***Cercopithifilaria japonica* (Uni, 1983)**	BS6-2	MIB:Zpl:01156	AM749261	AM779794	A,B,C	*Ursus thibetanus*	Japan, Gifu	YES
***Cercopithifilaria japonica* (Uni, 1983)**	BS9-1	MIB:Zpl:00941	AM749262	AM779793	A,B,C	*Ursus thibetanus*	Japan, Gifu	YES
***Cercopithifilaria japonica* (Uni, 1983)**	BP5-1	MIB:Zpl:00939	AM749263	AM779792	A,B,C	*Ursus thibetanus*	Japan, Gifu	YES
***Cercopithifilaria longa* Uni, Bain & Takaoka, 2002**	S51-PB1	MIB:Zpl:00926	AM749243	AM779783	A,B,C	*Cervus nippon*	Japan, Oita	YES
***Cercopithifilaria longa* Uni, Bain & Takaoka, 2002**	AG1-10	MIB:Zpl:00931	AM749244	AM779782	A,B,C	*Cervus nippon*	Japan, Hyogo	YES
***Cercopithifilaria longa* Uni, Bain & Takaoka, 2002**	AG1-5	MIB:Zpl:00930	AM749245	AM779781	A,B,C	*Cervus nippon*	Japan, Hyogo	YES
***Cercopithifilaria longa* Uni, Bain & Takaoka, 2002**	Isolate S32-2	-	AB178842	-	A,B	*Cervus nippon*	-	-
***Cercopithifilaria longa* Uni, Bain & Takaoka, 2002**	Isolate S32-4	-	AB178843	-	A,B	*Cervus nippon*	-	-
***Cercopithifilaria longa* Uni, Bain & Takaoka, 2002**	Isolate S33-4	-	AB178844	-	A,B	*Cervus nippon*	-	-
***Cercopithifilaria longa* Uni, Bain & Takaoka, 2002**	Isolate S33-6	-	AB178845	-	A,B	*Cervus nippon*	-	-
***Cercopithifilaria longa* Uni, Bain & Takaoka, 2002**	S51-PB2	MIB:Zpl:00912	AM749246	AM779784	A,B,C	*Cervus nippon*	Japan, Oita	YES
***Cercopithifilaria minuta* Uni & Bain 2001**	C1-A4	MIB:Zpl:00905	AM749252	AM779785	A,B,C	*Naemorhedus crispus* *	Japan, Yamagata	YES
***Cercopithifilaria minuta* Uni & Bain 2001**	SW1-23	MIB:Zpl:00915	AM749253	AM779786	A,B,C	*Naemorhedus crispus* *	Japan, Yamagata	YES
***Cercopithifilaria minuta* Uni & Bain 2001**	Isolate SW3-FL3	-	AB178846	-	A,B	*Capricornis crispus*	-	-

Species	Museum ID	Voucher	Accession number *coxI*	Accession number 12s rDNA	Dataset	Host	Locality	Produced in this study
***Cercopithifilaria minuta* Uni & Bain 2001**	Isolate SW3-FL12	-	AB178847	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria multicauda* Uni & Bain, 2001**	G119	MIB:Zpl:00921	AM749255	AM779799	A,B,C	*Naemorhedus crispus* *	Japan, Gifu	YES
***Cercopithifilaria multicauda* Uni & Bain, 2001**	SW3-FL9	MIB:Zpl:00922	AM749254	AM779800	A,B,C	*Naemorhedus crispus* *	Japan, Gifu	YES
***Cercopithifilaria multicauda* Uni & Bain, 2001**	Isolate Gifu-39T	-	AB178848	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria multicauda* Uni & Bain, 2001**	Isolate Gifu-49C	-	AB178849	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria roussilhoni* Bain, Petit & Chabaud, 1986**	143 SE	MIB:Zpl:00959	AM749264	AM779798	A,B,C	*Atherurus africanus*	Gabon, Makokou Station	YES
***Cercopithifilaria shohoi* Uni, Suzuki & Katsumi, 1998**	C1-LB4	MIB:Zpl:00906	AM749249	AM779795	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Cercopithifilaria shohoi* Uni, Suzuki & Katsumi, 1998**	SW1-32	MIB:Zpl:00919	AM749250	AM779796	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Cercopithifilaria shohoi* Uni, Suzuki & Katsumi, 1998**	SW21-170	MIB:Zpl:00923	AM749251	AM779797	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Cercopithifilaria shohoi* Uni, Suzuki & Katsumi, 1998**	Isolate Gifu-07	-	AB178850	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria shohoi* Uni, Suzuki & Katsumi, 1998**	Isolate Gifu-14	-	AB178851	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria tumidicervicata* Uni & Bain, 2001**	C1-LB8	MIB:Zpl:00904	AM749256	AM779788	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Cercopithifilaria tumidicervicata* Uni & Bain, 2001**	SW5-119	MIB:Zpl:00914	AM749257	AM779790	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Cercopithifilaria tumidicervicata* Uni & Bain, 2001**	C1-LBB1	MIB:Zpl:00901	AM749258	AM779787	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Cercopithifilaria tumidicervicata* Uni & Bain, 2001**	SW1-9	MIB:Zpl:00918	AM749259	AM779789	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Cercopithifilaria tumidicervicata* Uni & Bain, 2001**	Isolate Gifu-91	-	AB178852	-	A,B	*Capricornis crispus*	-	-
***Cercopithifilaria tumidicervicata* Uni & Bain, 2001**	Isolate Gifu-132	-	AB178853	-	A,B	*Capricornis crispus*	-	-
***Dipetalonema gracile* (Rudolphi, 1809)**	-	-	AJ544877	AJ544854	A,B,C	*Cebus olivaceus*	-	-
***Dipetalonema gracile* (Rudolphi, 1809)**	15YU	MIB:Zpl:01175	AM749279	AM779824	A,B,C	*Cebus olivaceus*	Venezuela, Yutaje	YES
***Dirofilaria (Dirofilaria) immitis* (Leidy, 1856)**	-	-	DQ358815	-	B	*Canis lupus*	-	-
***Dirofilaria (Dirofilaria) immitis* (Leidy, 1856)**	-	-	AJ271613	AJ544831	A,B,C	*Canis lupus familiaris*	-	-
***Dirofilaria (Dirofilaria) immitis* (Leidy, 1856)**	-	-	AJ537512	AJ537512	A,B,C	*Canis lupus familiaris*	-	-
***Dirofilaria (Dirofilaria) immitis* (Leidy, 1856)**	CATR	MIB:Zpl:01165	AM749226	AM779769	A,B,C	*Felis catus*	Italy, Milan	YES
***Dirofilaria (Dirofilaria) immitis* (Leidy, 1856)**	GEN2	MIB:Zpl:01170	AM749227	AM779771	A,B,C	*Felis catus*	Italy, Milan	YES
***Dirofilaria (Dirofilaria) immitis* (Leidy, 1856)**	GEN3	MIB:Zpl:01167	AM749228	AM779770	A,B,C	*Canis lupus familiaris*	Italy, Milan	YES
***Dirofilaria (Dirofilaria) immitis* (Leidy, 1856)**	M1	MIB:Zpl:01157	AM749229	-	A,B	*Canis lupus familiaris*	Italy, Milan	YES
***Dirofilaria (Dirofilaria) immitis* (Leidy, 1856)**	-	-	EU159111	-	A,B	*Canis lupus familiaris*	-	-
***Dirofilaria (Nochtiella) repens* Railliet & Henry, 1911**	-	-	DQ358814	-	B	*Canis lupus*	-	-
***Dirofilaria (Nochtiella) repens* Railliet & Henry, 1911**	-	-	AJ271614	AJ544832	A,B,C	*Canis lupus familiaris*	-	-
***Dirofilaria (Nochtiella) repens* Railliet & Henry, 1911**	CATM	MIB:Zpl:01166	AM749232	AM779777	A,B,C	*Felis catus*	Italy, Milan	YES
***Dirofilaria (Nochtiella) repens* Railliet & Henry, 1911**	PAV	MIB:Zpl:01148	AM749233	AM779774	A,B,C	*Homo sapiens*	Italy, Pavia	YES
***Dirofilaria (Nochtiella) repens* Railliet & Henry, 1911**	POZ	MIB:Zpl:01163	AM749234	AM779778	A,B,C	*Homo sapiens*	Italy, Rome	YES
***Dirofilaria (Nochtiella) repens* Railliet & Henry, 1911**	CATG	MIB:Zpl:01168	AM749231	AM779773	A,B,C	*Felis catus*	Italy, Milan	YES
***Dirofilaria (Nochtiella) repens* Railliet & Henry, 1911**	GEN1	MIB:Zpl:01171	AM749230	AM779775	A,B,C	*Canis lupus familiaris*	Italy, Milan	YES
***Filaria martis* Gmelin, 1790**	-	-	AJ544880	AJ544855	A,B,C	*Martes foina*	-	-
***Foleyella furcata* (Linstow, 1899)**	-	-	AJ544879	AJ544841	A,B,C	*Chameleon (not determined)*	-	-
***Litomosa westi* (Gardner & Smith, 1986)**	-	-	AJ544871	AJ544851	A,B,C	*Geomys bursarius*	-	-

Species	Museum ID	Voucher	Accession number *coxI*	Accession number 12s rDNA	Dataset	Host	Locality	Produced in this study
***Litomosoides brasiliensis* Lins de Almeida, 1936**	-	-	AJ544867	AJ544850	A,B,C	*Carollia perspicillata*	-	-
***Litomosoides galizai* Bain, Petit, Diagne, 1989**	-	-	AJ544870	AJ544849	A,B,C	*Oecomys tr. tapajinus*	-	-
***Litomosoides hamletti* Sandground, 1934**	-	-	AJ544868	AJ544847	A,B,C	*Glossophaga soricina*	-	-
***Litomosoides scotti* Forrester & Kinsella, 1973**	-	-	EF661995	-	B	*n.d.*	-	-
***Litomosoides sigmodontis* Chandler, 1931**	-	-	AJ271615	AJ544848	A,B,C	*Sigmodon hispidus*	France, Paris **	-
***Litomosoides sigmodontis* Chandler, 1931**	1L	MIB:Zpl:01164	AM749286	AM779834	A,B,C	*Sigmodon hispidus*	Venezuela, Yutaje	YES
***Litomosoides yutajensis* Guerrero, Martin & Bain, 2003**	39 YU	MIB:Zpl:01155	AM749280	AM779825	A,B,C	*Pteronotus parnellii*	Venezuela, Yutaje	YES
***Litomosoides yutajensis* Guerrero, Martin & Bain, 2003**	-	-	AJ544869	AJ544846	A,B,C	*Pteronotus parnelli*	-	-
***Loa loa* (Cobbold, 1864)**	-	-	AJ544875	AJ544845	A,B,C	*Homo sapiens*	-	-
***Loxodontofilaria caprini* Uni & Bain, 2006**	YG2-25	MIB:Zpl:00928	AM749237	AM779817	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Loxodontofilaria caprini* Uni & Bain, 2006**	YG3-1	MIB:Zpl:00946	AM749238	AM779818	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Loxodontofilaria caprini* Uni & Bain, 2006**	C1-1A	MIB:Zpl:00903	AM749239	AM779820	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Loxodontofilaria caprini* Uni & Bain, 2006**	C1-SB10	MIB:Zpl:01151	AM749240	AM779821	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Loxodontofilaria caprini* Uni & Bain, 2006**	YG3-12	MIB:Zpl:01149	AM749241	AM779819	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Loxodontofilaria caprini* Uni & Bain, 2006**	C1-FFL1	MIB:Zpl:00902	AM749242	AM779822	A,B,C	*Naemorhedus crispus* *	Japan,Yamagata	YES
***Mansonella (Cutifilaria) perforata* Uni, Bain & Takaoka, 2004**	S51-PB9	MIB:Zpl:00911	AM749265	AM779803	A,B,C	*Cervus nippon*	Japan, Oita	YES
***Mansonella (Tetrapetalonema) atelensis amazonae* n. subsp. Bain & Guerrero, 2008**	15 YU	MIB:Zpl:00958	AM749278	AM779823	A,B,C	*Cebus olivaceus*	Venezuela, Yutaje	YES
Spirurida sp.MOTU1	CAM7	MIB:Zpl:01160	AM749287	-	B	*Redunca fulvorufula*	North Cameroon, Daoud Safari	YES
Spirurida sp.MOTU1	CAM8	MIB:Zpl:01158	AM749288	-	B	*Redunca fulvorufula*	North Cameroon, Daoud Safari	YES
Spirurida sp.MOTU2	NAM5	MIB:Zpl:01162	AM749289	-	B	*Equus zebra hartmannae*	Namibia, Ohorongo Safari Outjo	YES
Spirurida sp.MOTU2	NAM6	MIB:Zpl:01150	AM749290	-	B	*Equus zebra hartmannae*	Namibia, Ohorongo Safari Outjo	YES
Spirurida sp.MOTU3	SIT1	MIB:Zpl:00887	AM749291	-	B	*Sitta europea*	France, Pas-de-Calais, Dpt 62	YES
Spirurida sp.MOTU3	D06	MIB:Zpl:01153	AM749292	-	B	*Paradoxornis webbianus*	Itay, Palude Brabbia	YES
Spirurida sp.MOTU3	ID07	MIB:Zpl:01152	AM749293	-	B	*Paradoxornis webbianus*	Italy, Palude Brabbia	YES
Spirurida sp.MOTU4	DIP1	MIB:Zpl:00886	AM749294	-	B	*Sturnus vulgaris*	France, Roanne, Dpt 42	YES
Spirurida sp.MOTU4	DIP2	MIB:Zpl:00885	AM749295	-	B	*Sturnus vulgaris*	France,Firminy, Dpt 42	YES
Spirurida sp.MOTU4	DIP3	MIB:Zpl:01154	AM749296	-	B	*Sturnus vulgaris*	France, St Etienne, Dpt 42	YES
Spirurida sp.MOTU5	NAM7	MIB:Zpl:01161	AM749297	-	B	*Oryx gazella*	Namibia, Ohorongo Safari Outjo	YES
Ochoterenella sp. *sensu* Casiraghi et al., 2004	-	-	AJ544878	-	B	*Bufo marinus*	-	-
***Onchocerca dewittei japonica* Uni, Bain & Takaoka, 2001**	B61-7	MIB:Zpl:00913	AM749266	AM779816	A,B,C	*Sus scrofa leucomystax*	Japan, Oita	YES
***Onchocerca dewittei japonica* Uni, Bain & Takaoka, 2001**	B61-4	MIB:Zpl:00917	AM749267	AM779815	A,B,C	*Sus scrofa leucomystax*	Japan, Oita	YES
***Onchocerca eberhardi* Uni & Bain, 2007**	S51-9	MIB:Zpl:00956	AM749268	AM779810	A,B,C	*Cervus nippon*	Japan, Oita	YES
***Onchocerca gibsoni* (Cleland & Johnston, 1910)**	-	-	AJ271616	AJ544837	A,B,C	*Bos taurus*	-	-
***Onchocerca lupi* Rodonaja, 1967**	-	-	AJ415417	-	A,B	*Canis lupus familiaris*	-	-

Species	Museum ID	Voucher	Accession number *coxI*	Accession number 12s rDNA	Dataset	Host	Locality	Produced in this study
***Onchocerca lupi* Rodonaja, 1967**	-	-	EF521408	-	A,B	*Canis lupus familiaris*	-	-
***Onchocerca lupi* Rodonaja, 1967**	-	-	EF521409	-	A,B	*Canis lupus familiaris*	-	-
***Onchocerca lupi* Rodonaja, 1967**	-	-	EF521410	-	A,B	*Canis lupus familiaris*	-	-
***Onchocerca ochengi* Bwangamoi, 1969**	-	-	AJ271618	AJ544839	A,B,C	*Bos taurus*	-	-
***Onchocerca skrjabini* Ruklyadev, 1964**	S51-2	MIB:Zpl:00924	AM749269	AM779804	A,B,C	*Cervus nippon*	Japan, Oita	YES
***Onchocerca skrjabini* Ruklyadev, 1964**	S51-7	MIB:Zpl:00910	AM749270	AM779806	A,B,C	*Cervus nippon*	Japan, Oita	YES
***Onchocerca skrjabini* Ruklyadev, 1964**	S51-4	MIB:Zpl:00908	AM749271	AM779805	A,B,C	*Cervus nippon*	Japan, Oita	YES
***Onchocerca skrjabini* Ruklyadev, 1964**	G30	MIB:Zpl:00916	AM749272	AM779808	A,B,C	*Naemorhedus crispus* *	Japan, Gifu	YES
***Onchocerca skrjabini* Ruklyadev, 1964**	C1-FL5	MIB:Zpl:00909	AM749273	AM779807	A,B,C	*Naemorhedus crispus* *	Japan, Yamagata	YES
***Onchocerca skrjabini* Ruklyadev, 1964**	SW30-26	MIB:Zpl:00920	AM749274	AM779809	A,B,C	*Naemorhedus crispus* *	Japan, Yamagata	YES
***Onchocerca suzukii* Yagi, Bain & Shoho, 1994**	YG2-35	MIB:Zpl:00932	AM749275	AM779811	A,B,C	*Naemorhedus crispus* *	Japan, Yamagata	YES
***Onchocerca suzukii* Yagi, Bain & Shoho, 1994**	YG2-53	MIB:Zpl:00937	AM749276	AM779813	A,B,C	*Naemorhedus crispus* *	Japan, Yamagata	YES
***Onchocerca suzukii* Yagi, Bain & Shoho, 1994**	YG2-37	MIB:Zpl:00935	AM749277	AM779812	A,B,C	*Naemorhedus crispus* *	Japan, Yamagata	YES
***Onchocerca volvulus* (Leuckart, 1893)**	-	-	NC_001861	-	A,B	*Homo sapiens*	-	-
***Onchocerca volvulus* (Leuckart, 1893)**	M4	MIB:Zpl:01172	AM749285	AM779855	A,B,C	*Homo sapiens*	Italy, Rome ***	YES
***Onchocerca volvulus* (Leuckart, 1893)**	M3	MIB:Zpl:01173	AM749284	AM779854	A,B,C	*Homo sapiens*	Italy, Rome ***	YES
***Piratuba scaffi* Bain, 1974**	34 YU_1	MIB:Zpl:00955	AM749281	AM779831	A,B,C	*Ameiva ameiva*	Venezuela, Yutaje	YES
***Piratuba scaffi* Bain, 1974**	34 YU_2	MIB:Zpl:00956	AM749282	AM779832	A,B,C	*Ameiva ameiva*	Venezuela, Yutaje	YES
***Piratuba scaffi* Bain, 1974**	34 YU_3	MIB:Zpl:00957	AM749283	-	B	*Ameiva ameiva*	Venezuela, Yutaje	YES
***Setaria digitata* (Linstow, 1906)**	ST1	MIB:Zpl:00936	AM886173	AM779801	A,B,C	*Bos taurus*	Japan, Yamagata	YES
***Setaria digitata* (Linstow, 1906)**	Isolate SL/2005/K/Sd1	-	EF174428	-	A,B	*Bos taurus*	-	-
***Setaria digitata* (Linstow, 1906)**	Isolate SL/2005/K/Sd9	-	EF174427	-	A,B	*Bos taurus*	-	-
***Setaria digitata* (Linstow, 1906)**	Isolate SL/2005/K/Sd8	-	EF174426	-	A,B	*Bos taurus*	-	-
***Setaria digitata* (Linstow, 1906)**	Isolate SL/2005/K/Sd7	-	EF174425	-	A,B	*Bos taurus*	-	-
***Setaria digitata* (Linstow, 1906)**	Isolate SL/2005/K/Sd6	-	EF174424	-	A,B	*Bos taurus*	-	-
***Setaria digitata* (Linstow, 1906)**	Isolate SL/2005/K/Sd5	-	EF174423	-	A,B	*Bos taurus*	-	-
***Setaria equina* (Abildgaard, 1789)**	-	-	AJ544873	AJ544835	A,B,C	*Equus caballus*	-	-
***Setaria labiatopapillosa* (Alessandrini, 1848)**	-	-	AJ544872	AJ544833	A,B,C	*Bos taurus*	-	-
***Setaria tundra* Issaitshikoff & Rajewskaya, 1928**	SET1	MIB:Zpl:01159	AM749298	AM779848	A,B,C	*Capreolus capreolus*	France	YES
***Setaria tundra* Issaitshikoff & Rajewskaya, 1928**	-	-	AJ544874	AJ544834	A,B,C	*Capreolus capreolus*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 1	-	EF195132	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 2	-	EF394599	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 3	-	EF394600	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 4	-	EF394601	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 5	-	EF394602	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 6	-	EF394603	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 7	-	EF195133	-	A,B	*Canis lupus familiaris*	-	-

Species	Museum ID	Voucher	Accession number *coxI*	Accession number 12s rDNA	Dataset	Host	Locality	Produced in this study
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 8	-	EF394604	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 9	-	EF394605	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 10	-	EF394596	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 11	-	EF394597	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 12	-	EF394598	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 13	-	EF394606	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 14	-	EF394607	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 15	-	EF394608	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 16	-	EF394609	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 17	-	EF394610	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 18	-	EF394611	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 19	-	EF394612	-	A,B	*Canis lupus familiaris*	-	-
***Spirocerca lupi* (Rudolphi, 1809)**	Isolate 20	-	EF394613	-	A,B	*Canis lupus familiaris*	-	-
***Thelazia callipaeda* Railliet & Henry, 1910**	-	-	AJ544882	AJ544858	A,B,C	*Canis lupus familiaris*	-	-
***Thelazia callipaeda* Railliet & Henry, 1910**	Haplotype h1	-	AM042549	-	A,B	n.d.	-	-
***Thelazia callipaeda* Railliet & Henry, 1910**	Haplotype h2	-	AM042550	-	A,B	*Canis lupus familiaris*	-	-
***Thelazia callipaeda* Railliet & Henry, 1910**	Haplotype h3	-	AM042551	-	A,B	*Canis lupus familiaris*	-	-
***Thelazia callipaeda* Railliet & Henry, 1910**	Haplotype h4	-	AM042552	-	A,B	*Canis lupus familiaris*	-	-
***Thelazia callipaeda* Railliet & Henry, 1910**	Haplotype h5	-	AM042553	-	A,B	*Canis lupus familiaris*	-	-
***Thelazia callipaeda* Railliet & Henry, 1910**	Haplotype h6	-	AM042554	-	A,B	*Canis lupus familiaris*	-	-
***Thelazia callipaeda* Railliet & Henry, 1910**	Haplotype h7	-	AM042555	-	A,B	*Canis lupus familiaris*	-	-
***Thelazia callipaeda* Railliet & Henry, 1910**	Haplotype h8	-	AM042556	-	A,B	*Canis lupus familiaris*	-	-
***Thelazia gulosa* (Railliet & Henry, 1910)**	-	-	AJ544881	AJ544857	A,B,C	*Bos taurus*	-	-
***Thelazia lacrymalis* (Gurlt, 1831)**	-	-	AJ271619	AJ544856	A,B,C	*Equs caballus*	-	-
***Wuchereria bancrofti* (Cobbold, 1877)**	-	-	AJ271612	AJ544844	A,B,C	*Homo sapiens*	-	-
***Wuchereria bancrofti* (Cobbold, 1877)**	M005	MIB:Zpl:01169	AM749235	-	A,B	*Homo sapiens*	Italy, Milan ****	YES
***Wuchereria bancrofti* (Cobbold, 1877)**	M065	MIB:Zpl:01174	AM749235	-	A,B	*Homo sapiens*	Italy, Milan ****	YES

APPENDIX I.5: List of species of all the taeniid samples (case study B2) considered in this study, including biological data, voucher name and accession numbers (a.n.). Place of collection and host species are indicated. If no data concerning hosts are available commonest hosts are showed, marked with an asterisk (*). Accession numbers of complete mitochondrial sequences are also marked with an asterisk. In the last column, assignment to Lineage (1,2 or 3) of *Taenia taeniaeformis* as revealed by data analysis is reported. Samples included in the reference dataset are highlighted in bold.

Voucher/ a.n.	Species	Isolate	Host(s)	Sampling locality	a.n.	Lineage
MIB:Zpl:01356	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris* ssp.	_	FN547824	1
MIB:Zpl:01358	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Preci, Norcia (PG)	FN547825	1
MIB:Zpl:01359	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valnerina (PG)	FN547826	1
MIB:Zpl:01360	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valico di Fossato (PG)	FN547827	1
MIB:Zpl:01361	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Amelia (TR)	FN547828	1
MIB:Zpl:01362	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valnerina (PG)	FN547829	1
MIB:Zpl:01363	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Norcia (PG)	FN547830	1
MIB:Zpl:01364	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris* x *Felis silvestris catus*	Umbria, Colfiorito (PG)	FN547831	1
MIB:Zpl:01365	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris* ssp.	_	FN547832	1
MIB:Zpl:01366	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris* ssp.	_	FN547833	1
MIB:Zpl:01367	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Calabria, Sila (CS)	FN547834	1
MIB:Zpl:01370	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Abruzzo, Caramanico Terme (PE)	FN547835	1
MIB:Zpl:01371	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Marche, Visso (MC)	FN547836	1
MIB:Zpl:01372	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Lazio, Selva di Lamone (VT)	FN547837	1
MIB:Zpl:01373	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valnerina (PG)	FN547838	1
MIB:Zpl:01374	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Citta della Pieve (PG)	FN547839	1
MIB:Zpl:01375	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Marche, Camerino (MC)	FN547840	1
MIB:Zpl:01376	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valnerina (PG)	FN547841	1
MIB:Zpl:01377	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valnerina (PG)	FN547842	1
MIB:Zpl:01378	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Monti Martani (PG)	FN547843	1
MIB:Zpl:01379	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valnerina (PG)	FN547844	1
MIB:Zpl:01380	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris* x *Felis silvestris catus*	Toscana, Grosseto	FN547845	1
MIB:Zpl:01381	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Cascia (PG)	FN547846	1
MIB:Zpl:01382	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris* ssp.	_	FN547847	1
MIB:Zpl:01383	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris* ssp.	_	FN547848	1
MIB:Zpl:01384	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Cascia (PG)	FN547849	1
MIB:Zpl:01385	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valfabbrica (PG)	FN547850	2
MIB:Zpl:01386	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Norcia (PG)	FN547851	1
MIB:Zpl:01387	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Marche, Pieve Torina (MC)	FN547852	1
MIB:Zpl:01388	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Cascia (PG)	FN547853	1
MIB:Zpl:01389	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris* ssp.	_	FN547854	1
MIB:Zpl:01390	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris* x *Felis silvestris catus*	Umbria, Colfiorito (PG)	FN547855	1
MIB:Zpl:01392	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Monti Martani (PG)	FN547856	1
MIB:Zpl:01393	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valnerina (PG)	FN547857	1
MIB:Zpl:01394	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Abruzzo, Lago di Penne (PE)	FN547858	1
MIB:Zpl:01395	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Citta della Pieve (PG)	FN547859	1
MIB:Zpl:01396	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Citta della Pieve (PG)	FN547860	1
MIB:Zpl:01397	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris* x *Felis silvestris catus*	Marche, San Severino (MC)	FN547861	1
MIB:Zpl:01398	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Monti Martani (PG)	FN547862	1
MIB:Zpl:01399	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valnerina (PG)	FN547863	1
MIB:Zpl:01400	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Spoleto (PG)	FN547864	1
MIB:Zpl:01401	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Marche, San Severino (MC)	FN547865	1
MIB:Zpl:01402	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Norcia (PG)	FN547866	1
MIB:Zpl:01403	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Marche, Rustici (AP)	FN547867	1
MIB:Zpl:01404	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria,Orvieto (TR)	FN547868	1
MIB:Zpl:01405	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Monteleone di Spoleto (PG)	FN547869	1
MIB:Zpl:01406	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Lazio, Viterbo	FN547870	1
MIB:Zpl:01407	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris* ssp.	_	FN547871	1
MIB:Zpl:01408	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Ancarano, Norcia (PG)	FN547872	1
MIB:Zpl:01409	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris* ssp.	_	FN547873	1
MIB:Zpl:01410	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valnerina (PG)	FN547874	1
MIB:Zpl:01411	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Norcia (PG)	FN547875	1
MIB:Zpl:01412	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Abruzzo, Caramanico Terme (PE)	FN547876	1
MIB:Zpl:01413	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Norcia (PG)	FN547877	1
MIB:Zpl:01414	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Colfiorito (PG)	FN547878	1
MIB:Zpl:01415	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valnerina (PG)	FN547879	1
MIB:Zpl:01416	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Marche, Carpegna (PU)	FN547880	1
MIB:Zpl:01417	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Forca Canapine (PG)	FN547881	1
MIB:Zpl:01418	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Abruzzo, Lago di Penne (PE)	FN547882	1
MIB:Zpl:01419	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris silvestris*	Umbria, Valnerina (PG)	FN547883	1
MIB:Zpl:01638	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris catus*	Lombardia, Besana Brianza (MB)	FN547823	1
MIB:Zpl:01639	*Taenia* cf. *taeniaeformis*	_	*Felis silvestris catus*	Lombardia, Monticello Brianza (LC)	FN547884	1
NC_004826*	*Taenia asiatica*	_	*Sus scrofa domestica, Homo sapiens**	_	NC_004826*	_
AB066494	*Taenia asiatica*	_	*Sus scrofa domestica, Homo sapiens**	Taiwan	AB066494	_
AB107234	*Taenia asiatica*	_	*Sus scrofa domestica, Homo sapiens**	Taiwan	AB107234	_
AB107235	*Taenia asiatica*	_	*Sus scrofa domestica, Homo sapiens**	China	AB107235	_
AB107236	*Taenia asiatica*	_	*Sus scrofa domestica, Homo sapiens**	Indonesia	AB107236	_
NC_002547*	*Taenia crassiceps*	_	Rodentia, *Canis latrans, Vulpes* spp.*	_	NC_002547*	_
AB033411	*Taenia crassiceps*	_	Rodentia, *Canis latrans, Vulpes* spp.*	_	AB033411	_
EU544546	*Taenia crassiceps*	TcSv1	*Vulpes lagopus*	Norway: Svalbard	EU544546	_

Voucher/ a.n.	Species	Isolate	Host(s)	Sampling locality	a.n.	Lineage
EU544547	*Taenia crassiceps*	TcSv2	*Microtus levis*	Norway: Svalbard	EU544547	_
EU544548	*Taenia crassiceps*	TcYa	*Microtus gregalis*	Russia: Yamal Peninsula	EU544548	_
EU544549	*Taenia crassiceps*	TcBu	*Microtus fortis*	Russia: Buryatia	EU544549	_
EU544550	*Taenia crassiceps*	TcAl	*Microtus pennsylvanicus*	USA: Alaska	EU544550	_
NC_012896*	*Taenia hydatigena*	_	*Ovis aries*	China	NC_012896*	_
AB033410	*Taenia hydatigena*	_	*Ovis aries, Sus scrofa**	_	AB033410	_
AM503315	*Taenia hydatigena*	Thy29	*Canis lupus familiaris*	Kenya	AM503315	_
AM503316	*Taenia hydatigena*	Thy65	*Canis lupus familiaris*	Kenya	AM503316	_
AM503317	*Taenia hydatigena*	Thy124	*Canis lupus familiaris*	Kenya	AM503317	_
AM503318	*Taenia hydatigena*	Thy183	*Canis lupus familiaris*	Kenya	AM503318	_
DQ995656	*Taenia hydatigena*	_	*Ovis aries*	India	DQ995656	_
EU544551	*Taenia hydatigena*	ThFi1	*Ovis aries*	Finland	EU544551	_
EU544552	*Taenia hydatigena*	ThFi2	*Rangifer tarandus*	Finland	EU544552	_
EU544572	*Taenia krabbei*	TkSv1	*Vulpes lagopus*	Norway: Svalbard	EU544572	_
EU544573	*Taenia krabbei*	TkSv2	*Vulpes lagopus*	Norway: Svalbard	EU544573	_
EU544574	*Taenia krabbei*	TkSv3	*Vulpes lagopus*	Norway: Svalbard	EU544574	_
EU544575	*Taenia krabbei*	TkSv4	*Vulpes lagopus*	Norway: Svalbard	EU544575	_
EU544576	*Taenia krabbei*	TkSv5	*Vulpes lagopus*	Norway: Svalbard	EU544576	_
EU544577	*Taenia krabbei*	TkSv6	*Vulpes lagopus*	Norway: Svalbard	EU544577	_
EU544578	*Taenia krabbei*	TkSv7	*Vulpes lagopus*	Norway: Svalbard	EU544578	_
EU544579	*Taenia krabbei*	TkSv8	*Vulpes lagopus*	Norway: Svalbard	EU544579	_
AM503323	*Taenia madoquae*	Tma117	*Canis mesomelas*	Kenya	AM503323	_
AM503324	*Taenia madoquae*	Tma118	*Canis mesomelas*	Kenya	AM503324	_
AM503325	*Taenia madoquae*	Tma120	*Canis mesomelas*	Kenya	AM503325	_
AM503331	*Taenia madoquae*	Tma122	*Canis mesomelas*	Kenya	AM503331	_
EU544553	*Taenia martis*	TmaDe1	*Myodes glareolus*	Denmark	EU544553	_
EU544554	*Taenia martis*	TmaDe2	*Myodes glareolus*	Denmark	EU544554	_
EU544555	*Taenia martis*	TmaCr	*Myodes glareolus*	Croatia	EU544555	_
EU544556	*Taenia martis*	TmaTu	*Apodemus sylvaticus*	Turkey	EU544556	_
EU544557	*Taenia martis*	TmaBu	*Myodes rufocanus*	Russia: Buryatia	EU544557	_
EU544558	*Taenia martis*	TmaChi	*Myodes rufocanus*	China: Fenglin	EU544558	_
FN547885	*Taenia martis*	_	*Martes foina*	Italy	FN547885	_
FN547886	*Taenia martis*	_	*Martes foina*	Italy	FN547886	_
FN547887	*Taenia martis*	_	*Martes martes*	Italy	FN547887	_
NC_012894*	*Taenia multiceps*	_	*Canis lupus familiaris*	China	NC_012894*	_
DQ309767	*Taenia multiceps*	Tm1	*Ovis aries, Capra aegagrus hircus, Canis lupus**	Italy: Sardinia	DQ309767	_
DQ309768	*Taenia multiceps*	Tm2	*Ovis aries, Capra aegagrus hircus, Canis lupus**	Italy: Sardinia	DQ309768	_
DQ309769	*Taenia multiceps*	Tm3 12	*Ovis aries, Capra aegagrus hircus, Canis lupus**	Italy: Sardinia	DQ309769	_
DQ321830	*Taenia multiceps*	Tm1 - alfa1	*Ovis aries*	Italy: Sardinia	DQ321830	_
EF393620	*Taenia multiceps*	tmtr01	*Ovis aries*	Turkey: Nigde	EF393620	_
FJ744755	*Taenia multiceps*	_	*Ovis* spp.	Italy	FJ744755	_
EU544559	*Taenia mustelae*	TmuFi1	*Myodes glareolus*	Finland	EU544559	_
EU544560	*Taenia mustelae*	TmuFi2	*Myodes glareolus*	Finland	EU544560	_
EU544561	*Taenia mustelae*	TmuFi3	*Myodes glareolus*	Finland	EU544561	_
EU544562	*Taenia mustelae*	TmuFi4	*Myodes glareolus*	Finland	EU544562	_
EU544563	*Taenia mustelae*	TmuFi5	*Myodes glareolus*	Finland	EU544563	_
EU544564	*Taenia mustelae*	TmuFi6	*Myodes glareolus*	Finland	EU544564	_
EU544565	*Taenia mustelae*	TmuFi7	*Myodes glareolus*	Finland	EU544565	_
EU544566	*Taenia mustelae*	TmuFi8	*Myodes rufocanus*	Finland	EU544566	_
EU544567	*Taenia mustelae*	TmuFi9	*Myodes rutilus*	Finland	EU544567	_
EU544568	*Taenia mustelae*	TmuIr1	*Myodes rutilus*	Russia: Irkutsk region	EU544568	_
EU544569	*Taenia mustelae*	TmuIr2	*Myodes rutilus*	Russia: Irkutsk region	EU544569	_
EU544570	*Taenia mustelae*	TmuEv1	*Myodes rufocanus*	Russia: Evenkia	EU544570	_
EU544571	*Taenia mustelae*	TmuEv2	*Myopus schisticolor*	Russia: Evenkia	EU544571	_
EU544580	*Taenia parva*	TpaSp	*Apodemus sylvaticus*	Spain	EU544580	_
EU544581	*Taenia polyacantha*	TpoTu	*Microtus guentheri*	Turkey	EU544581	_
EU544582	*Taenia polyacantha*	TpoSc	*Myodes glareolus*	United Kingdom: Scotland	EU544582	_
EU544583	*Taenia polyacantha*	TpoDe	*Myodes glareolus*	Denmark	EU544583	_
EU544584	*Taenia polyacantha*	TpoFi1	*Myodes glareolus*	Finland	EU544584	_
EU544585	*Taenia polyacantha*	TpoFi2	*Myodes glareolus*	Finland	EU544585	_
EU544586	*Taenia polyacantha*	TpoFi3	*Microtus oeconomus*	Finland	EU544586	_
EU544587	*Taenia polyacantha*	TpoFi4	*Vulpes vulpes*	Finland	EU544587	_
EU544588	*Taenia polyacantha*	TpoFi5	*Vulpes vulpes*	Finland	EU544588	_
EU544589	*Taenia polyacantha*	TpoFi6	*Vulpes vulpes*	Finland	EU544589	_
EU544590	*Taenia polyacantha*	TpoSv1	*Vulpes lagopus*	Norway: Svalbard	EU544590	_
EU544591	*Taenia polyacantha*	TpoSv2	*Vulpes lagopus*	Norway: Svalbard	EU544591	_
EU544592	*Taenia polyacantha*	TpoSv3	*Vulpes lagopus*	Norway: Svalbard	EU544592	_
EU544593	*Taenia polyacantha*	TpoSv4	*Vulpes lagopus*	Norway: Svalbard	EU544593	_
EU544594	*Taenia polyacantha*	TpoGr	*Dicrostonyx groenlandicus*	Denmark: Greenland	EU544594	_
EU544595	*Taenia polyacantha*	TpoCa	*Lemmus trimucronatus*	Canada: Cape Bathurst	EU544595	_
FJ744756	*Taenia polyacantha*	_	*Vulpes* spp.	_	FJ744756	_
AM503328	*Taenia regis*	Tre152	*Panthera leo*	Kenya	AM503328	_
AM503329	*Taenia regis*	Tre154	*Panthera leo*	Kenya	AM503329	_
AM503330	*Taenia regis*	Tre157	*Panthera leo*	Kenya	AM503330	_
AY684274*	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	_	AY684274*	_
AB033409	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	_	AB033409	_
AB066495	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	China	AB066495	_
AB107237	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	Brazil	AB107237	_
AB107238	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	Ecuador	AB107238	_
AB107239	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	China	AB107239	_

Voucher/ a.n.	Species	Isolate	Host(s)	Sampling locality	a.n.	Lineage
AB107240	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	Indonesia: Bali	AB107240	_
AB107241	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	Ethiopia	AB107241	_
AB107242	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	Belgium	AB107242	_
AB107243	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	Nepal	AB107243	_
AB107244	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	Thailand: Bangkok	AB107244	_
AB107245	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	Thailand	AB107245	_
AB107246	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	Brazil: Mato Grosso do Sul	AB107246	_
AB107247	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	China: Yunnan	AB107247	_
AB271695	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	Mongolia	AB271695	_
AB275143	*Taenia saginata*	_	*Homo sapiens*	Cambodia	AB275143	_
AB494480	*Taenia saginata*	TS-090108	*Homo sapiens*	Japan: Osaka	AB494480	_
AM503326	*Taenia saginata*	Tsa136	*Homo sapiens*	Kenya	AM503326	_
AM503327	*Taenia saginata*	Tsa141	*Homo sapiens*	Kenya	AM503327	_
AY195858	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	_	AY195858	_
DQ768207	*Taenia saginata*	JDC2006	*Bos primigenius, Homo sapiens**	France	DQ768207	_
FJ744754	*Taenia saginata*	_	*Bos primigenius, Homo sapiens**	_	FJ744754	_
AJ239110	*Taenia serialis*	Ts1	*Vulpes* spp.	Australia: Victoria	AJ239110	_
AM503319	*Taenia serialis*	Tse79	*Canis aureus*	Kenya	AM503319	_
AM503320	*Taenia serialis*	Tse88	*Canis aureus*	Kenya	AM503320	_
AM503321	*Taenia serialis*	Tse91	*Canis aureus*	Kenya	AM503321	_
AM503322	*Taenia serialis*	Tse105	*Canis aureus*	Kenya	AM503322	_
DQ401138	*Taenia serialis*	_	*Homo sapiens*	France	DQ401138	_
AB086256*	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	China	AB086256*	_
AB033408	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	_	AB033408	_
AB066485	*Taenia solium*	CHI1	*Sus scrofa, Homo sapiens**	China	AB066485	_
AB066486	*Taenia solium*	CHI2	*Sus scrofa, Homo sapiens**	China	AB066486	_
AB066487	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	Thailand	AB066487	_
AB066488	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	Indonesia: Irian Jaya	AB066488	_
AB066489	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	India	AB066489	_
AB066490	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	Mexico	AB066490	_
AB066491	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	Ecuador	AB066491	_
AB066492	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	Brazil	AB066492	_
AB066493	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	Tanzania	AB066493	_
AB243755	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	Brazil: Piaui, Piracuruca, Cocal dos Alves	AB243755	_
AB271234	*Taenia solium*	_	*Homo sapiens*	Indonesia: Bali	AB271234	_
AF360865	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	Peru	AF360865	_
AF360867	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	Colombia	AF360867	_
AF360868	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	Mexico	AF360868	_
AF360869	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	India	AF360869	_
AF360870	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	Philippines	AF360870	_
AF360871	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	China	AF360871	_
AY211880	*Taenia solium*	_	*Sus scrofa*	Tanzania: Mbulu District	AY211880	_
AY395065	*Taenia solium*	_	*Homo sapiens*	Korea	AY395065	_
AY395066	*Taenia solium*	_	*Homo sapiens*	Korea	AY395066	_
DQ089663	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	_	DQ089663	_
DQ202385	*Taenia solium*	_	*Homo sapiens*	South Africa	DQ202385	_
DQ202386	*Taenia solium*	_	*Homo sapiens*	Indonesia: Papua	DQ202386	_
EF076752	*Taenia solium*	_	*Sus scrofa*	India	EF076752	_
EU078323	*Taenia solium*	TsCcp1	*Sus scrofa*	Mexico	EU078323	_
EU747651	*Taenia solium*	TsCcp3	*Sus scrofa*	_	EU747651	_
EU747652	*Taenia solium*	TsCcp4	*Sus scrofa*	_	EU747652	_
EU747653	*Taenia solium*	TsCcp5	*Sus scrofa*	_	EU747653	_
EU747654	*Taenia solium*	TsCcp6	*Sus scrofa*	_	EU747654	_
EU747655	*Taenia solium*	TsCcp7	*Sus scrofa*	_	EU747655	_
EU747657	*Taenia solium*	TsChb1	*Homo sapiens*	_	EU747657	_
EU747658	*Taenia solium*	TsChb2	*Homo sapiens*	_	EU747658	_
EU747659	*Taenia solium*	TsChb3	*Homo sapiens*	_	EU747659	_
EU747660	*Taenia solium*	TsChb4	*Homo sapiens*	_	EU747660	_
EU747661	*Taenia solium*	TsChb5	*Homo sapiens*	_	EU747661	_
S69013	*Taenia solium*	_	*Sus scrofa, Homo sapiens**	_	S69013	_
AB221484	*Taenia taeniaeformis*	TtSRN	*Rattus norvegicus*	Japan: Hokkaido, Sapporo	AB221484	3
EF090612	*Taenia taeniaeformis*	_	*Rattus rattus*	India	EF090612	3
EU544596	*Taenia taeniaeformis*	TtaTu	*Apodemus sylvaticus*	Turkey	EU544596	2
EU544597	*Taenia taeniaeformis*	TtaKa	*Apodemus sylvaticus*	Kazakhstan	EU544597	3
EU861478	*Taenia taeniaeformis*	TtaFi	*Felis silvestris catus*	Finland: Porvoo	EU861478	2
FJ939135	*Taenia taeniaeformis*	_	*Rattus rattus*	India	FJ939135	3
_	*Taenia taeniaeformis*	TtKRN	*Rattus norvegicus*	Malaysia: Kuala Lumpur	_	3
_	*Taenia taeniaeformis*	TtBMM	*Mus musculus*	Belgium	_	3
_	*Taenia taeniaeformis*	TtChi	*Mus musculus*	China	_	3
_	*Taenia taeniaeformis*	TtNop	*Apodemus argenteus*	Japan: Nopporo	_	3
_	*Taenia taeniaeformis*	TtACR	*Clethrionomys rufocanus bedfordiae*	Japan: Abuta	_	2
_	*Taenia taeniaeformis*	TtMar	*Rattus norvegicus*	Japan: Hokkaido, Sapporo	_	3
_	*Taenia taeniaeformis*	TtTom	*Rattus norvegicus*	Japan: Tomikawa	_	3
_	*Taenia taeniaeformis*	TtKaRN	*Rattus norvegicus*	Japan: Kamiiso	_	3
_	*Taenia taeniaeformis*	TtKaAA	*Apodemus argenteus*	Japan: Kamiiso	_	3
EU544598	*Taenia twitchelli*	TtwChu	*Gulo gulo*	Russia: Chukotka	EU544598	3

APPENDIX I.6: List of sharks species (case study C1; reference dataset) including taxonomical data (Order, family and genus), GenBank accession numbers of the sharks included in this study. GenBank accession numbers in bold character indicate *coxI* sequences produced in this study corresponding to samples collected and morphologically identified at the Milan fish market by experts.

Order	Family	Genus	Species	Accession numbers
Carcharhiniformes	Carcharinidae	*Carcharhinus*	*albimarginatus*	EU398582; EU398583; EU398584; EU398585; EU398586
			altimus	EU398587; EU398588; EU398589
			amblyrhynchoides	EF609307; EU398590; EU398591; EU398592; EU398593
			amblyrhynchos	EF609308; EU398594; EU398595; EU398596; EU398597
			amboinensis	DQ885075; DQ885076; DQ885077; EU398599; EU398600
			brevipinna	EU398601; EU398602; EU398603
			cautus	EF609309; EU398605; EU398606
			dussumieri	DQ108301; DQ108302; DQ108303; DQ108304; EU398608; EU398609; EU398610
			falciformis	EU398611; EU398612; EU398613; EU398614
			fitzroyensis	EF609310; EU398615
			leucas	EF609311; EU398616; EU398617; EU398618; EU398619
			limbatus	EU398620; EU398621; EU398622; EU398623; EU398624; EU398625
			longimanus	EU398626; EU398627
			macloti	EF609312; EU398628; EU398629
			melanopterus	EF609313; EU398630; EU398631; EU398632; EU398633
			obscurus	DQ108291; DQ108306; EU398634; EU398635; EU398636; EU398637
			plumbeus	EU398638; EU39869
			sealei	EU398640; EU398641; EU398642; EU398643; EU398644
			sorrah	DQ108292; DQ108293; DQ108294; DQ108295; EU398645
			tilstoni	DQ108283; DQ108296; DQ108297; DQ108298
		Galeocerdo	*cuvier*	EU398785; EU398786; EU398787; EU398788; EU398789
		Lamiopsis	*temminckii*	EU398901
		Negaprion	*acutidens*	DQ108284; EU398935; EU398936; EU398937; EU398938; EU398939; EU398940
		Prionace	*glauca*	DQ108285; DQ108286; DQ108287; DQ108288; DQ108289; **FM164482**
		Rhizoprionodon	*acutus*	DQ108275; DQ108276; DQ108277; DQ108278; DQ108290
			taylori	EF609447; EU399000; EU399001; EU399003; EU399004; EU399005; EU399006; EU399002
		Triaenodon	*obesus*	EU399059; EU399060
	Hemigaleidae	*Hemigaleus*	*australiensis*	EU398810; EU398811; EU398812; EU398813; EU398814; EU398815; EU398816; EU398817; EU398818; EU398819
			microstoma	EF609372; EU398820; EU398821
		Hemipristis	*elongata*	EU398822; EU398823; EU398824; EU398825; EU398826; EU398827
	Scyliorhinidae	*Asymbolus*	*parvus*	EU398564; EU398565
			rubiginosus	EU398566; EU398567
		Atelomycterus	*baliensis*	EU398568; EU398569
			fasciatus	EU398570
			marmoratus	EU398571; EU398572
			marnkalha	EU398574; EU398575; EU398576; EU398577
		Aulohalaelurus	*labiosus*	EU398578; EU398579; EU398580; EU398581
		Cephaloscyllium	*laticeps*	DQ108322
		Galeus	*boardmani*	EU398789; EU398790; EU398791; EU398792; EU398793
		Scyliorhinus	*canicula*	Y16067; **FM164483**
			stellaris	**FM164484**
	Sphyrnidae	*Eusphyra*	*blochii*	EU398784
		Sphyrna	*lewini*	EU399011; EU399012; EU399013; EU399014
			mokarran	EU399015; EU399016; EU399017
			zygaena	EU399018;
	Triakidae	*Furgaleus*	*macki*	DQ108316; DQ108317; DQ108318; DQ108319
		Galeorhinus	*galeus*	FM164429; DQ108308; DQ108309; DQ108310; DQ108320; DQ108321
		Hemitriakis	*falcata*	EU398828
		Mustelus	*asterias*	**FM164477; FM164478**
			mustelus	**FM164479; FM164480; FM164481**
			antarcticus	DQ108311; DQ108312; DQ108313; DQ108314; DQ108315
			lenticulatus	DQ108299; DQ108300; DQ108307
			manazo	AB015962
			schmitti	EU074486; EU074487; EU074488
Heterodontiformes	Heterodontidae	*Heterodontus*	*francisci*	AJ310141
			galeatus	EU398829
			portusjackson	EU398830; EU398831; EU398832; EU398833; EU398834; EU398835; EU398836
Hexanchiformes	Hexanchidae	*Hexanchus*	*griseus*	EU398837; **FM164472**
		Notorynchus	*cepedianus*	DQ108326; DQ108327; DQ108332; DQ108333; DQ108334
Lamniformes	Alopiidae	*Alopias*	*pelagicus*	EU398513; EU398514; EU398515; EU398516; EU398517; EU398518
			superciliosus	EU398519; EU398520; EU398521; EU400162
			vulpinus	EU398522
	Lamnidae	*Carcharodon*	*carcharias*	DQ108328; EU398646
		Isurus	*oxyrhinchus*	EU398889; EU398890; EU398891; EU398892; EU398893; EU398894; EU398895; EU398896; EU398897; **FM164473**; **FM164474**; **FM164475**
			paucus	EU398899; EU398900
		Lamna	*nasus*	**FM164476**
	Megachasmidae	*Megachasma*	*pelagios*	EU398905
	Mitsukurinidae	*Mitsukurina*	*owstoni*	EU398906
	Pseudocarchariidae	*Pseudocarcharias*	*kamoharai*	EU398990
Orectolobiformes	Ginglymostomatidae	*Nebrius*	*ferrugineus*	EU398933; EU398934
	Hemiscylliidae	*Chiloscyllium*	*hasseltii*	EF609324; EU398677; EU398678; EU398679

Order	Family	Genus	Species	Accession numbers
			indicum	EF609325; EU398680; EU398681; EU398682; EU398683; EU398684; EU3986885; EU398686; EU398687; EU398688; EU398689; EU398690; EU398691; EU398692
			plagiosum	EU398693; EU398694; EU398695; EU398696
			punctatum	EF609326; EU398697; EU398698; EU398699; EU398700; EU398701; EU398702; EU398703; EU398704; EU398705; EU398706;EU398707
	Orectolobidae	*Orectolobus*	*halei*	EU398942; EU398943; EU398944
			hutchinsi	EU398945; EU398946; EU398947; EU398948; EU398949; EU398950; EU398951; EU398952
			maculatus	DQ108324; DQ108325; DQ108331; EU398953; EU398954; EU398955; EU398956; EU398957;EU398958
		Sutorectus	*tentaculatus*	EU399057
	Stegostomatidae	*Stegostoma*	*fasciatum*	EU399049; EU399050; EU399051; EU399052; EU399053
Pristiophoriformes	Pristiophoridae	*Pristiophorus*	*cirratus*	DQ108210; DQ108211; DQ108212; DQ108213; DQ108214; EU398979; EU398980; EU398981
			nudipinnis	DQ108204; DQ108205; DQ108206; DQ108207; DQ108215; EU398982; EU398983; EU398984; EU398985
Squaliformes	Centrophoridae	*Centrophorus*	*atromarginatus*	EU398647; EU398648; EU398649; EU398650;EU398651
			isodon	EU398652; EU398653; EU398654
			moluccensis	DQ108227; DQ108231; DQ108233; DQ108240; DQ108241; DQ108242; EU398655; EU398656; EU398657
			niaukang	DQ108228; DQ108229
			squamosus	DQ108230
			uyato	EU398658; EU398659; EU398660; EU398661; EU398662; EU398663; EU398664; EU398665; EU398666; EU398667; EU398668
		Deania	*calcea*	DQ108222; DQ108223; DQ108224; DQ108225
	Dalatiidae	*Centroscymnus*	*crepidater*	DQ108233; DQ108234
			fasciatum	EU398669
		Dalatias	*licha*	DQ108221; **FM164471**
		Etmopterus	*granulosus*	DQ108216; DQ108226
			pusillus	EU398783
		Zameus	*squamulosus*	DQ108208; DQ108209; DQ108217; DQ108218
	Echinorhinidae	*Echinorhinus*	*cookei*	EU398777
	Squalidae	*Cirrhigaleus*	*barbifer*	EU398719; EU398720
		Squalus	*edmundsi*	EU399025
			acanthias	DQ108267; DQ108279; DQ108280; DQ108281; DQ108282; EF539270; EF539271; EF539272; EF539273; EF539274; EF539275; EF539276; EF539277; EF539278; EF539279; EF539280; EF539281; EF539282; EF539283; EF539284; EF539285; EF539286; EF539287; EF539288; EF539289; EF539290; EF539291; Y18134
			albifrons	DQ108254; DQ108255; DQ108256; DQ108257; EF539292
			brevirostris	EF539293; EF539294; EF539295; EF539296; EF539297; EF539298; EF539299; EF539300
			chloroculus	EF539301
			crassispinus	DQ108247; DQ108248; EU399019; EU399020; EU399021; EU399023; EU399024
			edmundsi	DQ108243; DQ108244; DQ108245; DQ108246; DQ10825; DQ108251; DQ108252; DQ108265; DQ108266; EF539302; EF539303; EF539304; EF539305; EF539306; EU399026; EU399027
			grahami	DQ108235; DQ108236, DQ108238; DQ108239; EU399028
			hemipinnis	EF539307; EF539309; EF539310; EF539311; EF539312
			japonicus	EF539313; EF539314; EF539315; EF539316; EF539317
			megalops	DQ108268; DQ108269; DQ108270; DQ108271; EF539318; EU399029; EU399030; EU399031; EU399032
			montalbani	EF539319; EF539320; EF539321; EF539322; EU399033
			nasutus	DQ108249; DQ108250; EF539323; EF539324; EF539325; EF539326; EF539327; EF539328; EU399034; EU399035
Squatiniformes	Squatinidae	*Squatina*	*formosa*	EU399040; EU399041
			tergocellata	EU399048; DQ108194;DQ108195; DQ108197; EU399047
			australis	DQ108193; DQ108200; DQ108203; EU399036; EU399037; EU399038

APPENDIX I.7: List of analysed spices samples (case study C2) divided in different groups (Gr) according to their taxonomy and provenance. For each sample the voucher number (V.N.) the Reference Species name, the cultivar name or common name, for the commercial samples, are provided. The Accession Numbers correspondig of DNA sequences of the four analysed markers are also included. Freshes samples were collected (C) from greenhouse of Milan Botanical Garden (MBG), certify seed and garden center Ingegnoli (ING), garden center Biovivaio Grand Burrone (BGB) and garden center Viridea (VIR). Commercial samples were collected in four different European Commercial Companies and to ensure their anonymity each sample was marked with the following codes CP1, CP2, CP3 and CP4.

Gr	V. N.	Species	Cultivar	Common name	C	Code	Accession numbers			
							matK	psbA-trnH	*rpoB*	*rbcL*
Group I	MIB:Zpl:03291	Mentha piperita L.	piperita		MBG	MP1	FR719055	FR726096	FR720471	FR720529
	MIB:Zpl:03292	Mentha piperita L.	piperita		VIR	MP2	FR719056	FR726097	FR720472	FR720530
	MIB:Zpl:03780	Mentha piperita L.	piperita		ING	MP3	FR719057	FR726098	-	FR720531
	MIB:Zpl:03781	Mentha aquatica L.	gigante		ING	MA1	FR719058	FR726099	-	FR720532
	MIB:Zpl:03782	Mentha aquatica L.	-		BGB	MA2	FR719059	FR726100	-	FR720533
	MIB:Zpl:03783	Mentha spicata L.	Crispa		BGB	MS1	FR719060	FR726101	FR720473	FR720534
	MIB:Zpl:03784	Mentha spicata L.	maroccan		BGB	MS2	FR719061	FR726102	FR720474	FR720535
	MIB:Zpl:03301	_	-	Mint	CP2	MEC1	FR719062	FR726103	FR720475	FR720536
	MIB:Zpl:03306	_	-	Mint	CP3	MEC2	FR719063	FR726104	FR720476	FR720537
	MIB:Zpl:03785	_	-	Mint	CP4	MEC3	-	FR726105	FR720477	FR720538
Group II	MIB:Zpl:03288	Ocimum basilicum L.	Italian classic		MBG	OB1	FR719064	FR726106	FR720478	FR720539
	MIB:Zpl:03289	Ocimum basilicum L.	Italian classic		ING	OB2	FR719065	FR726107	FR720479	FR720540
	MIB:Zpl:03786	Ocimum basilicum L.	Italian classic		BGB	OB3	FR719066	FR726108	FR720480	FR720541
	MIB:Zpl:02997	Ocimum gratissimum L.	Vana tulsi		MBG	OG	FR719067	FR726109	FR720481	FR720542
	MIB:Zpl:02998	Ocimum tenuiflorum L.	Krisha Tulsi		MBG	OT	FR719068	FR726110	FR720482	FR720543
	MIB:Zpl:03299	_	-	Basil	CP2	BC1	FR719069	FR726111	FR720483	FR720544
	MIB:Zpl:03787	_	-	Basil	CP3	BC2	FR719070	FR726112	FR720484	FR720545
	MIB:Zpl:03788	_	-	Basil	CP4	BC3	FR719071	FR726113	FR720485	FR720546
	MIB:Zpl:02884	Ocimum basilicum L.	Mostruoso mammouth		ING	OBcv1	FR719072	FR726114	FR720486	FR720547
	MIB:Zpl:02885	Ocimum basilicum L.	Green leave		ING	OBcv1	FR719073	FR726115	FR720487	FR720548
	MIB:Zpl:02886	Ocimum basilicum L.	Gecom		ING	OBcv2	FR719074	FR726116	FR720488	FR720549
	MIB:Zpl:02887	Ocimum basilicum L.	Red leave		ING	OBcv3	FR719075	FR726117	FR720489	FR720550
	MIB:Zpl:02888	Ocimum basilicum L.	Verde a palla		ING	OBcv4	FR719076	FR726118	FR720490	FR720551
	MIB:Zpl:02889	Ocimum basilicum L.	Italiano		ING	OBcv5	FR719077	FR726119	FR720491	FR720552
	MIB:Zpl:02890	Ocimum basilicum L.	Napoletano		ING	OBcv6	FR719078	FR726120	FR720492	FR720553
	MIB:Zpl:02996	Ocimum basilicum L.	scernese		ING	OBcv7	FR719079	FR726121	FR720493	FR720554
Group III	MIB:Zpl:03290	Origanum majorana L.	Sweet		BGB	OM1	FR719080	FR726122	FR720494	FR720555
	MIB:Zpl:03789	Origanum majorana L.	Sweet		VIR	OM2	FR719081	FR726123	FR720495	FR720556
	MIB:Zpl:03790	Origanum majorana L.	Sweet		MBG	OM3	FR719082	FR726124	FR720496	FR720557
	MIB:Zpl:03791	Origanum majorana L.	Sweet		ING	OM4	FR719083	FR726125	-	FR720558
	MIB:Zpl:03293	Origanum vulgare L.	Aureum		MBG	OV1	FR719084	FR726129	FR720497	FR720559
	MIB:Zpl:03294	Origanum vulgare L.	Aureum		VIR	OV2	FR719085	FR726130	FR720498	FR720560
	MIB:Zpl:03792	Origanum vulgare L.	Aureum		BGB	OV3	FR719086	FR726131	FR720499	FR720561
	MIB:Zpl:03793	Origanum vulgare L.	Gigante		ING	OV4	FR719087	FR726132	-	FR720562
	MIB:Zpl:03794	Origanum vulgare L.	vulgaris		ING	OV5	FR719088	FR726133	-	FR720563
	MIB:Zpl:03795	Origanum pseudodictamnius Sieber	-		BGB	OP	FR719089	FR726137	FR720500	FR720564
	MIB:Zpl:03796	Origanum heracleoticum L.	-		BGB	OH	FR719090	FR726138	FR720501	FR720565
	MIB:Zpl:03797	_	-	Marjoram	CP1	MAC1	FR719091	FR726126	FR720502	FR720566
	MIB:Zpl:03300	_	-	Marjoram	CP2	MAC2	FR719092	FR726127	FR720503	FR720567
	MIB:Zpl:03798	_	-	Marjoram	CP4	MAC3	FR719093	FR726128	FR720504	FR720568
	MIB:Zpl:03302	_	-	Oregano	CP1	OC1	FR719094	FR726134	FR720505	FR720569
	MIB:Zpl:03799	_	-	Oregano	CP2	OC2	FR719095	FR726135	FR720506	FR720570
	MIB:Zpl:03302	_	-	Oregano	CP3	OC3	FR719096	FR726136	FR720507	FR720571

Gr	V. N.	Species	Cultivar	Common name	C	Code	Accession numbers			
							matK	psbA-trnH	*rpoB*	*rbcL*
Group IV	MIB:Zpl:03800	Salvia officinalis L.	albiflora		MBG	SO1	FR719097	FR726139	FR720508	FR720572
	MIB:Zpl:03801	Salvia officinalis L.	albiflora		BGB	SO2	FR719098	FR726140	FR720509	FR720573
	MIB:Zpl:03297	Salvia officinalis L.	albiflora		VIR	SO3	FR719099	FR726141	FR720510	FR720574
	MIB:Zpl:03802	Salvia rutilans	-		BGB	SR	FR719100	FR726142	FR720511	FR720575
	MIB:Zpl:03803	Salvia sclarea	-		BGB	SS	FR719101	FR726143	FR720512	FR720576
	MIB:Zpl:03804	Salvia uliginosa	-		BGB	SU	FR719102	FR726144	FR720513	FR720577
	MIB:Zpl:03304	_	-	Sage	CP2	SC1	FR719103	FR726145	FR720514	FR720578
	MIB:Zpl:03305	_	-	Sage	CP1	SC2	FR719104	FR726146	FR720515	FR720579
	MIB:Zpl:03306	_	-	Sage	CP3	SC3	FR719105	FR726147	FR720516	FR720580
Group V	MIB:Zpl:03307	Thymus vulgaris L.	vulgaris		MBG	TV1	FR719106	FR726148	FR720517	FR720581
	MIB:Zpl:03308	Thymus vulgaris L.	vulgaris		VIR	TV2	FR719107	FR726149	FR720518	FR720582
	MIB:Zpl:03309	Thymus vulgaris L.	vulgaris		ING	TV3	FR719108	FR726150	FR720519	FR720583
	MIB:Zpl:03298	_	-	Thyme	CP1	TC1	FR719109	FR726151	FR720520	FR720584
	MIB:Zpl:03305	_	-	Thyme	CP2	TC2	FR719110	FR726152	FR720521	FR720585
	MIB:Zpl:03810	_	-	Thyme	CP4	TC3	FR719111	FR726153	FR720522	FR720586
Group VI	MIB:Zpl:03295	Rosmarinus officinalis L.	Arp		MBG	RO1	FR719112	FR726154	FR720523	FR720587
	MIB:Zpl:03296	Rosmarinus officinalis L.	Arp		ING	RO2	FR719113	FR726155	FR720524	FR720588
	MIB:Zpl:03811	Rosmarinus officinalis L.	Arp		VIR	RO3	FR719114	FR726156	FR720525	FR720589
	MIB:Zpl:03812	_	-	Rosemary	CP1	RC1	FR719115	FR726157	FR720526	FR720590
	MIB:Zpl:03303	_	-	Rosemary	CP4	RC2	FR719116	FR726158	FR720527	FR720591
	MIB:Zpl:03813	__	-	Rosemary	CP3	RC3	FR719117	FR726159	FR720528	FR720592

APPENDIX I.8: List of analysed poisonous plants (case study C3) divided into different groups (Gr) (as discussed in Material and Methods, chapter 3). For each species, voucher number (V.N.), taxonomic assignment, common name, details on toxicity (poisonous organs and toxic substances) and information about material used as a source of DNA (Abbreviation: L=leaves; F=fruits) are provided.

N°	Gr	V. N.	Species	Family	Common name	Poisonous organ	Toxic sunbstances	DNA source
1	I	MIB:zpl:01657	*Nandina domestica* Thunb.	Berberidaceae	Sacred Bamboo	Fruit and other part.	Hydrocyanic acid and nandenine.	L- F
2	I	MIB:zpl:01658	*Ilex aquifolium* L.	Aquifoliaceae	Holly	Fruit, leaves and seed	Theobromine, alkaloid and glucoside.	L- F
3	I	MIB:zpl:01659	*Aucuba japonica* Thunb.	Garryaceae	Spotted-laurel	Fruit, leaves	Aucubin and differet glycosides.	L- F
4	I	MIB:zpl:01660	*Arum italicum* Mill.	Araceae	Lords-and-Ladies	All parts	Calcium oxalate crystals.	L
5	I	MIB:zpl:01661	*Arum maculatum* L.	Araceae	Lords-and-Ladies	All parts	Calcium oxalate crystals.	L
6	I	MIB:zpl:01662	*Convallaria majalis* L.	Ruscaceae	Lily-of-the-valley	All parts	Cardiac glycosides and saponins.	L
7	I	MIB:zpl:01664	*Ruscus aculeatus* L.	Ruscaceae	Butcher's-broom	Fruit	Unknown.	L
8	I	MIB:zpl:01665	*Hedera helix* L.	Araliaceae	Common Ivy	All parts	Triterpenoid saponins and polyacetylene.	L
9	I	MIB:zpl:01666	*Hedera hibernica* (G.Kirchn.) Bean.	Araliaceae	Irish Ivy	All parts	Triterpenoid saponins and polyacetylene.	L
10	I	MIB:zpl:01668	*Ligustrum vulgare* L.	Oleaceae	European Privet	Berries	Ligustrin, syringin and other glycosides.	L
11	I	MIB:zpl:01669	*Ligustrum lucidum* W.T.Aiton	Oleaceae	Glossy Privet	Berries	Ligustrin, syringin and other glycosides.	L
12	I	MIB:zpl:01673	*Ligustrum japonicum* Thunb.	Oleaceae	Japanese Privet	Berries	Ligustrin, syringin and other glycosides.	L
13	I	MIB:zpl:01674	*Phytolacca americana* L.	Phytolaccaceae	American Pokeweed	All parts	Phytolaccatoxin and related triterpene saponins, alkaloid and histamines.	L-F
14	I	MIB:zpl:01695	*Ficus benjamina* L.	Moraceae	Weeping Fig	Plant sap from all parts	Furocoumarins, psoralens, ficin.	L
15	I	MIB:zpl:01697	*Monstera deliciosa* Liebm.	Araceae	Mexican Breadfruit	All parts	Needle-like calcium oxalate crystals and other unidentified toxins.	L
16	I	MIB:zpl:01698	*Philodendron* sp.	Araceae	Philodendron	All parts	Calcium oxalate crystals and other toxins.	L
17	I	MIB:zpl:01669	*Dieffenbachia seguine* (Jacq.) Schott	Araceae	Dumb Cane	All parts	Calcium oxalate crystals, oxalic acid.	L
18	I	MIB:zpl:01706	*Spathiphyllum wallisii* Regel	Araceae	Peace-lily	Leaves	Calcium oxalate crystals.	L
19	I	MIB:zpl:01707	*Trachelospermum jasminoides* Lem.	Apocynaceae	Star Jasmine	Leaves	Unknown.	L
20	I	MIB:zpl:01708	*Schefflera arboricola* (Hayata) Merr.	Araliaceae	Schefflera	Leaves, plant sap	Oxalates.	L
21	I	MIB:zpl:01710	*Sansevieria trifasciata* Prain	Ruscaceae	Snake Plant	All parts	Possibly saponins and organic acids.	L
22	I	MIB:zpl:01711	*Hydrangea macrophylla* (Thunb.) Ser.	Hydrangeaceae	Hydrangea	All parts	Cyanogenic glycoside such as hydrangin.	L
23	I	MIB:zpl:01712	*Wisteria sinensis* (Sims) Sweet	Fabaceae	Chinese Wisteria	Seed, and other parts	Glycoside (i.e. wisterin) and resin	L
24	I	MIB:zpl:01713	*Nerium oleander* L.	Apocynaceae	Oleander	All parts, green or dry	Glycosides, nerioside, oleandroside, ecc.	L
25	I	MIB:zpl:01715	*Skimmia reevesiana* (Fortune) Fortune	Rutaceae	Skimmia	Fruit	Alkaloid called 'skimmianin'.	L
26	I	MIB:zpl:01716	*Kalanchoë daigremontiana* Hamet & Perrier	Crassulaceae	Mexican Hat Plant	Leaves, stems	Glycoside such as daigremontianin.	L
27	I	MIB:zpl:01717	*Anthurium andraeanum* Linden	Araceae	Tail Flower	All parts	Calcium oxalate crystals.	L
28	I	MIB:zpl:01719	*Veratrum lobelianum* Bernh.	Melanthiaceae	False-helleborine	All parts	Steroidal alkaloids.	L
29	I	MIB:zpl:01720	*Veratrum nigrum* L.	Melanthiaceae	Black False-helleborine	All parts	Steroidal alkaloids.	L
30	I	MIB:zpl:01722	*Lycianthes rantonnetii* (Carrière) Bitter	Solanaceae	Blue Potato-bush	All parts	Different alkaloids.	L
31	I	MIB:zpl:01663	*Atropa bella-donna* L.	Solanaceae	Deadly Nightshade	All parts, mainly berries	Tropane alkaloids, atropine and others.	L
32	I	MIB:zpl:01701	*Colchicum autumnale* L.	Colchicaceae	Meadow Saffron	All parts	Alkaloid colchicine.	L
33	I	MIB:zpl:01676	*Aconitum lycoctonum* L.	Ranunculaceae	Wolf's-bane	All parts	Alkaloids aconitine and others.	L
34	I-IIa	MIB:zpl:01675	*Aconitum napellus* L.	Ranunculaceae	Monk's-hood	All parts	Alkaloids aconitine and others.	L
35	IIa	MIB:zpl:01670	*Aconitum degenii* Gáyer subsp. *paniculatum* (Arcang.) Mucher	Ranunculaceae	Panicled Monk's-hood	All parts	Alkaloids aconitine and others.	L

N°	Gr	V. N.	Species	Family	Common name	Poisonous organ	Toxic sunbstances	DNA source
36	IIa	MIB:zpl:01700	*Aconitum anthora* L.	Ranunculaceae	Pyrenean Monk's-hood	All parts	Alkaloids aconitine and others.	L
37	I-IIb	MIB:zpl:01678	*Sambucus ebulus* L.	Adoxaceae	Dwarf Elder	Fruit; and other parts	Cyanogenic glycoside and others.	L + F
38	IIb	MIB:zpl:01679	*Sambucus racemosa* L.	Adoxaceae	Red-berried Elder	Edible fruit	Cyanogenic glycoside in vegetative parts.	L
39	IIb	MIB:zpl:01680	*Sambucus nigra* L.	Adoxaceae	Elder	Edible fruit	Cyanogenic glycoside in vegetative parts.	L + F
40	IIIa	MIB:zpl:01689	*Solanum dulcamara* L.	Solanaceae	Bittersweet	Berries	Solanine and other alkaloids.	L + F
41	IIIa	MIB:zpl:01690	*Solanum nigrum* L.	Solanaceae	Black Nightshade	Berries	Solanine and other alkaloids.	L
42	I-IIIa	MIB:zpl:01691	*Solanum pseudocapsicum* L.	Solanaceae	Jerusalem-cherry	Fruits, other parts	Alcaloid such solanocapsine.	L + F
43	IIIa	MIB:zpl:01693	*Solanum lycopersicum* L.	Solanaceae	Tomato	Edible fruit	Not toxic.	L
44	IIIa	MIB:zpl:01694	*Solanum tuberosum* L.	Solanaceae	Potato	Edible fruit	Not toxic.	L
45	IIIb	MIB:zpl:01681	*Prunus armeniaca* L.	Rosaceae	Apricot	Edible fruit	Seeds cointain cyanogenic glycoside.	L
46	IIIb	MIB:zpl:01682	*Prunus avium* L.	Rosaceae	Cherry	Edible fruit	Seeds cointain cyanogenic glycoside.	L
47	IIIb	MIB:zpl:01684	*Prunus cerasus* L.	Rosaceae	Sour Cherry	Edible fruit	Seeds cointain cyanogenic glycoside.	L
48	IIIb	MIB:zpl:01685	*Prunus domestica* L.	Rosaceae	Plum	Edible fruit	Seeds cointain cyanogenic glycoside.	L
49	I-IIIb	MIB:zpl:01687	*Prunus laurocerasus* L.	Rosaceae	Cherry Laurel	Vegetative parts, fruit and seed	Cyanogenic glycoside, amygdalin and other.	L
50	IIIb	MIB:zpl:01688	*Prunus persica* (L.) Batsch	Rosaceae	Peach	Edible fruit	Seeds cointain cyanogenic glycoside.	L

APPENDIX II.1

Tipology	Method(s)	Software / tool(s)	Resources	Reference
Threshold (distance)	similarity	blastall - BLASTn	ftp://ftp.ncbi.nih.gov/blast/	Altschul et al., 1990
	similarity	BLAT	http://genome-test.cse.ucsc.edu/~kent/exe/	Kent, 2002
	similarity	blastall - megaBLAST	ftp://ftp.ncbi.nih.gov/blast/	Little & Stevenson, 2007
	pairwise distance	TaxI	axel.meyer@uni-konstanz.de	Steinke et al., 2005
	pairwise distance	TaxonDNA	http://taxondna.sf.net/	Meier et al., 2006
	K2P distance	MUSCLE, MEGA	maurizio.casiraghi@unimib.it	Ferri et al., 2009
	K2P distance	BOLD-IDS	http://www.barcodinglife.org/views/idrequest.php	Ratnasingham & Hebert, 2007
	Patristic distance	MrBayes, PAUP, APE, Perl scripts	lefebure@univlyon1.fr	Lefébure et al., 2006
Phylogenetic	Neighbor Joining	MUSCLE, MEGA	marianne.elias@ed.ac.uk	Elias et al., 2007
	Parsimony	MUSCLE,TNT	dlittle@nybg.org	Little & Stevenson, 2007
	Maximum likelihood	MUSCLE, SPR1, PHYML2	http://atgc.lirmm.fr/spr/	Elias et al., 2007
	Bayesian inference	SAP	http://fisher.berkeley.edu/cteg/software/munch	Munch et al., 2008
	coalescent based	_	rasmus@binf.ku.dk	Matz & Nielsen, 2005
	coalescent based	_	rasmus@binf.ku.dk	Nielsen & Matz, 2006
	coalescent based	COALESCENCE, FLUCTUATE, PAUP, Seq-Gen	golding@mcmaster.ca	Abdo & Golding, 2007
	coalescent based	COAL, MESQUITE	knowlesl@umich.edu	Knowles & Carstens, 2005
	coalescent based	general mixed Yule-coalescent (GMYC) model	monaghan@igb-berlin.de	Monaghan et al., 2009
Character based	diagnostic	CAOS	http://www.genomecurator.org/CAOS/CAOSindex.html	Sarkar et al., 2008
	diagnostic	MATLAB, local perl scripts	drichardson@rsmas.miami.edu	Richardson et al., 2007
	diagnostic	DNA-BAR (degenbar)	http://dna.engr.uconn.edu/~software/DNA-BAR/	Dasgupta et al., 2005
	diagnostic	DOME ID (local perl scripts)	dlittle@nybg.org	Little & Stevenson, 2007
Combined	Yule model/coalescence	TCS, MEGA, Arlequin, PAUP, PAUPRat script, Phylip, r8s, R	http://www.imedea.uib.es/~jpons/JPWPhome.htm	Pons et al., 2006
	BLAST/parsimony ratchet	BLAST, MUSCLE, TNT	dlittle@nybg.org	Little & Stevenson, 2007
	BLAST/SPR	BLAST, MUSCLE, SPR	dlittle@nybg.org	Little & Stevenson, 2007
	BLAST/Neighbor Joining	BLAST, MUSCLE, neighbor	dlittle@nybg.org	Little & Stevenson, 2007
Alignment-free	tree-based	ATIM: TNT, local scripts	dlittle@nybg.org	Little & Stevenson, 2007
	component vector	CVTree alpha 1.0	http://cvtree.cbi.pku.edu.cn	Chu et al., 2006; 2009
	spectrum kernel method	Spectrum	vladimir@cs.rutgers.edu	Kuska et al., 2009
Web tool	_	Web browser	http://www.ibarcode.org	Singer et al., 2009
	_	Web browser	http://www.dnabarcodelinker.com/	Singer et al., 2009
	_	Web browser	http://www.asianbarcode.org/	Lim et al., 2009
Other	_	ConFind, Python	http://www.colorado.edu/chemistry/RGHP/software/	Smagala et al., 2005

APPENDIX II.2: Primers characteristics and conditions used in this research project.

Primer name	Genomic region	Sequence (5' --> 3')	Source	PCR conditions
VF1d	*coxI* (mitochondrial)	TTCTCAACCAACCACAAR GAYATYGG	Ivanova et al., 2006	94°C(60s), [94°C(30s), 50°C(40s), 72°C(60s) X 5], [94°C(30s), 55°C(40s), 72°C (60s) X 35], 72°C(10min)
VR1d	*coxI* (mitochondrial)	TAGACTTCTGGGTGGCCRAARAAYCA	Ivanova et al., 2006	
LCO1490	*coxI* (mitochondrial)	GGTCAACAAATCATAAAGATATTGG	Folmer et al., 1994	94°C(60s), [94°C(60s), 45°C(90s), 72°C(90s) X 5], [94°C(60s), 50°C(90s), 72°C (60s) X 35], 72°C(10min)
HCO2198	*coxI* (mitochondrial)	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al., 1994	
BIRDF1	*coxI* (mitochondrial)	TTCTCCAACCACAAAGACATTGGCAC	Hebert et al., 2002	94°C(60s), [94°C(60s), 45°C(90s), 72°C(90s) x 5], [94°C(60s), 51°C(90s), 72°C(90s) x 30], 72°C(5min)
BIRDR1	*coxI* (mitochondrial)	ACGTGGGAGATAATTCCAAATCCTG	Hebert et al., 2002	
COIintF	*coxI* (mitochondrial)	TGATTGGTGGTTTTGGTAA	Casiraghi et al., 2001	94°C(60s), [94°C(45s), 52°C(45s), 72°C(90s) x 40],72°C(5min)
COIintR	*coxI* (mitochondrial)	ATAAGTACGAGTATCAATATC	Casiraghi et al., 2001	
12SF	12S rDNA (mitochondrial)	GTTCCAGAATAATCGGCTA	Casiraghi et al., 2004	94°C(60s), [94°C(45s), 50°C(45s), 72°C(90s) x 40],72°C(5min)
12SR	12S rDNA (mitochondrial)	ATTGACGGATGRTTTGTACC	Casiraghi et al., 2004	
JB3	*coxI* (mitochondrial)	TTTTTTGGGCATCCTGAGGTTTAT	Bowles et al., 1993	94°C(60s), [94°C(30s), 55°C(30s), 72°C(30s) x 35],72°C(5min)
JB4.5	*coxI* (mitochondrial)	TAAAGAAAGAACATAATGAAAATG	Bowles et al., 1993	
Shark_int	*coxI* (mitochondrial)	ATCTTTGGTGCATGAGCAGGAATAGT	Ward et al., 2005	94°C(60s), [94°C(50s), 54°C(50s), 72°C(60s) x 35],72°C(5min)
FishR2	*coxI* (mitochondrial)	ACTTCAGGGTGACCGAAGAATCAGAA	Ward et al., 2005	
matK 390	*matK* (plastidial)	CGATCTATTCATTCAATATTC	Cuénod et al., 2002	94°C(7min), [94°C(45s), 48°C(30s), 72°C(60s) x 35],72°C(7min)
matK 1326	*matK* (plastidial)	TCTAGCACACGAAAGTCGAAGT	Cuénod et al., 2002	
psbA	psbA-trnH (plastidial)	GTTATGCATGAACGTAATGCTC	Newmaster et al., 2009	94°C(7min), [94°C(45s), 53°C(30s), 72°C(60s) x 35],72°C(7min)
psbA - trnH	psbA-trnH (plastidial)	CGCGCATGGTGGATTCACAATCC	Newmaster et al., 2009	
rpoB 1F	*rpoB* (plastidial)	AAGTGCATTGTTGGAACTGG	Fazekas et al., 2008	94°C(7min), [94°C(45s), 55°C(30s), 72°C(60s) x 35],72°C(7min)
rpoB4R	*rpoB* (plastidial)	GATCCCAGCATCACAATTCC	Fazekas et al., 2008	
rbcL 1F	*rbcL* (plastidial)	ATGTCACCACAAACAGAAAC	Fay et al., 1998	94°C(7min), [94°C(45s),48°C(30s), 72°C(60s) x 35],72°C(7min)
rbcL 724R	*rbcL* (plastidial)	TCGCATGTACCTGCAGTAGC	Fay et al., 1998	
Sqd1F	*sqd1* (nuclear)	CTTGGGACSATGGGTGARTATGG	Li et al., 2008	94°C(7min), [94°C(45s),63°C(30s), 72°C(60s) x 35],72°C(7min)
Sqd1R	*sqd1* (nuclear)	CCWACAGCAGCYTGMACACAGAACC	Li et al., 2008	
At103F	*At103* (nuclear)	CTTCAAGCCMAAGTTCATCTTCTA	Li et al., 2008	94°C(7min), [94°C(45s),55°C(30s), 72°C(60s) x 35],72°C(7min)
At103R	*At103* (nuclear)	TTGGCAATCATTGAGGTACATNGTMACATA	Li et al., 2008	
ER65	*ND1* (mitochondrial)	CCTCGATGTTGGATCAGG	Mayer et al., 2001	94°C(4min), [94°C(60s),60°C(30s), 72°C(90s) x 35],72°C(10min)
ER66	*ND1* (mitochondrial)	GTATGGGCCCGATAGCTT	Mayer et al., 2001	
Molcit_F	*cyt b* (mitochondrial)	AATGACATGAAAAATCACCGTTGT	Ibañez et al., 2006	94°C(4min), [94°C(60s),50°C(30s), 72°C(90s) x 35],72°C(10min)
MVZ-16	*cyt b* (mitochondrial)	AAATAGGAARTATCAYTCTGGTTTRAT	Smith & Patton, 1993	
L14990	*cyt b* (mitochondrial)	CATCCAACATCTCTGCTTGATGAAA	Cibois et al., 1999	94°C(60s), [94°C(45s),54°C(45s), 72°C(90s) x 38],72°C(10min)
H15916	*cyt b* (mitochondrial)	ATGAAGGGATGTTCTACTGGTTG	Edwards et al., 1991	
L1549	12S rDNA (mitochondrial)	GGGTTGGTAAATCCTGTGCCAGCCA	Cibois et al., 2002	92°C(4min), [93°C(30s),50°C(40s), 72°C(40s) x 35],72°C(7min)
H1991	12S rDNA (mitochondrial)	GCTATACCTTGACCTGTCTT	Cibois et al., 2002	
L3214	16S rDNA (mitochondrial)	CGCCTGTTTATCAAAAACAT	Hedges, 1994	92°C(4min), [93°C(30s),50°C(40s), 72°C(40s) x 35],72°C(7min)
H3783	16S rDNA (mitochondrial)	CCGGTCTGAACTCAGATCACGT	Hedges & Sibley, 1194	

APPENDIX II.3: List of *Myotis nattereri* samples (case study A1) and *cyt b* and *ND1* sequences retrieved from GenBank used in this study with reference to specimen voucher, previous species and lineage attribution (whenever possible), sampling locality (with locality Group within brackets). Further details on Italian samples can be retrieved in Appendix Table I.1

Species / Voucher	**Sampling Locality and abbreviation**	**Cyt b**	**ND1**	**Lineage**	**Reference**
Myotis nattereri MIB:ZPL:01249	Southern Italy (SIT)	_	_	Myotis sp. C	this study
Myotis nattereri MIB:ZPL:01282	Southern Italy (SIT)	_	_	Myotis sp. C	this study
Myotis nattereri MIB:ZPL:01324	Southern Italy (SIT)	_	_	Myotis sp. C	this study
Myotis nattereri sp. B	Southern Morocco (SMA)	EU360644	EU360612	Myotis sp. B	Garcia-Mudarra et al. 2009
Myotis nattereri sp. B	Central Morocco (CMA)	EU360645	_	Myotis sp. B	Garcia-Mudarra et al. 2009
Myotis nattereri sp. B	Central Morocco (CMA)	EU360646	_	Myotis sp. B	Garcia-Mudarra et al. 2009
Myotis nattereri sp. B	Northern Morocco/Central Morocco (NMA CMA)	EU360647	_	Myotis sp. B	Garcia-Mudarra et al. 2009
Myotis nattereri sp. B	Northern Morocco (NMA)	EU360648	EU360613	Myotis sp. B	Garcia-Mudarra et al. 2009
Myotis nattereri sp. A	Northern Iberia (NIB)	DQ120884	_	Myotis sp. A	Ibanez et al. 2006
Myotis nattereri sp. A	Northern Iberia (NIB)	DQ120886	DQ120801	Myotis sp. A	Ibanez et al. 2006
Myotis nattereri sp. A	Northern Iberia (NIB)	DQ120885	_	Myotis sp. A	Ibanez et al. 2006
Myotis nattereri sp. A	Austria (AUS)	_	DQ915049	Myotis sp. A	Mayer et al. 2007
Myotis nattereri MIB:ZPL:00331	Northern Italy (NIT)	_	_	Myotis sp. A	this study
Myotis nattereri MIB:ZPL:00356	Northern Italy (NIT)	_	_	Myotis sp. A	this study
Myotis nattereri MIB:ZPL:01347	Northern Italy (NIT)	_	_	Myotis sp. A	this study
Myotis nattereri	Switzerland (SWI)	DQ120892	_	Myotis nattereri	Ibanez et al. 2006
Myotis nattereri	Germany (GER)	DQ120893	_	Myotis nattereri	Ibanez et al. 2006
Myotis nattereri	Germany (GER)	DQ120894	_	Myotis nattereri	Ibanez et al. 2006
Myotis nattereri	Germany (GER)	DQ120895	_	Myotis nattereri	Ibanez et al. 2006
Myotis nattereri	Greece (GRE)	AF376863	AY033984	Myotis nattereri	Ruedi et al. 2001
Myotis nattereri	Hungary (HUN)	_	AF401439	Myotis nattereri	Mayer et al. 2001
Myotis *nattereri* (*escalerai* n.sp.)	Southern Iberia (SIB)	DQ120890	EU360614	Myotis escalerai	Garcia-Mudarra et al. 2009
Myotis *nattereri* (*escalerai* n.sp.)	Southern Iberia (SIB)	DQ120890	EU360615	Myotis escalerai	Garcia-Mudarra et al. 2009
Myotis *nattereri* (*escalerai* n.sp.)	Southern Iberia (SIB)	DQ120891	DQ120802	Myotis escalerai	Garcia-Mudarra et al. 2009
Myotis *nattereri* (*escalerai* n.sp.)	Southern Iberia (SIB)	EU360649	_	Myotis escalerai	Garcia-Mudarra et al. 2009
Myotis *nattereri* (*escalerai* n.sp.)	Northern Iberia (NIB)	DQ120887	_	Myotis escalerai	Ibanez et al. 2006
Myotis *nattereri* (*escalerai* n.sp.)	Northern Iberia (NIB)	DQ120888	_	Myotis escalerai	Ibanez et al. 2006
Myotis Myotis	Northern Iberia (NIB)	AF246241	DQ120800	_	Castella et al. 2000

APPENDIX II.4: : List of all identification results of 41 blind samples collected in Italy and belonging to cryptic vespertilionid species (see appendix I.1 for more details). The identifications were performed using the IDS (identification engine on BOLD System) and the OT (Ferri et al., 2009) approaches.

Voucher	Genus	OT identification	BOLD-IDS identification
MIB:ZPL:01211	Myotis sp.	*mystacinus*	*mystacinus - cf. aurascens*
MIB:ZPL:01214	Myotis sp.	*mystacinus*	*mystacinus - cf. aurascens*
MIB:ZPL:01216	Myotis sp.	*mystacinus*	*mystacinus - cf. aurascens*
MIB:ZPL:01221	Myotis sp.	*mystacinus*	*mystacinus - cf. aurascens*
MIB:ZPL:01222	Myotis sp.	*mystacinus*	*mystacinus - cf. aurascens*
MIB:ZPL:01223	Myotis sp.	*mystacinus*	*mystacinus - cf. aurascens*
MIB:ZPL:01228	Myotis sp.	*mystacinus*	*cf. aurascens - mystacinus*
MIB:ZPL:01230	Myotis sp.	*alcathoe*	*alcathoe*
MIB:ZPL:01235	Myotis sp.	*mystacinus*	*cf. aurascens - mystacinus*
MIB:ZPL:01281	Myotis sp.	*alcathoe*	*alcathoe*
MIB:ZPL:01287	Myotis sp.	*alcathoe*	*alcathoe*
MIB:ZPL:01289	Myotis sp.	*alcathoe*	*alcathoe*
MIB:ZPL:01301	Myotis sp.	*mystacinus*	*mystacinus - cf. aurascens*
MIB:ZPL:01302	Myotis sp.	*mystacinus*	*mystacinus - cf. aurascens*
MIB:ZPL:01303	Myotis sp.	*mystacinus*	*mystacinus - cf. aurascens*
MIB:ZPL:01256	Myotis sp.	*mystacinus*	*mystacinus - cf. aurascens*
MIB:ZPL:01319	Myotis sp.	*mystacinus*	*mystacinus - cf. aurascens*
MIB:ZPL:01239	Pipistrellus sp.	*pipistrellus*	*pipistrellus*
MIB:ZPL:01241	Pipistrellus sp.	*pipistrellus*	*pipistrellus*
MIB:ZPL:02288	Pipistrellus sp.	*pipistrellus*	*pipistrellus*
MIB:ZPL:03815	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03816	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03817	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03818	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03819	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03820	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03821	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03822	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03823	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03824	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03825	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03826	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03827	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03828	Pipistrellus sp.	*pygmaeus*	*pygmaeus - pipistrellus*
MIB:ZPL:03414	Plecotus sp.	*macrobullaris*	*macrobullaris*
MIB:ZPL:00262	Plecotus sp.	*macrobullaris*	*macrobullaris*
MIB:ZPL:01189	Plecotus sp.	*macrobullaris*	*macrobullaris*
MIB:ZPL:00265	Plecotus sp.	*auritus*	*auritus*
MIB:ZPL:00268	Plecotus sp.	*auritus*	*auritus*
MIB:ZPL:00269	Plecotus sp.	*auritus*	*auritus*
MIB:ZPL:00270	Plecotus sp.	*auritus*	*auritus*

APPENDIX II.5: List of all identification results of 45 blind samples collected in Italian market and labeled as ‘‘palombo” (case study C1). The identifications were performed using the IDS (identification engine on BOLD System) and the OT (Ferri et al., 2009) approaches.

Voucher	Accession number	Identified as (IDS)	Identified as (OT)
MIB:zpl:00004	FM164426	*Squalus brevirostris; Squalus cf. megalops*	*Squalus brevirostris* *Squalus megalops*
MIB:zpl:00005	FM164427	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00006	FM164428	*Prionace glauca*	*Prionace glauca*
MIB:zpl:00009	FM164429	*Galeorhinus galeus*	*Galeorhinus galeus*
MIB:zpl:00010	FM164430	*Mustelus antarticus; Mustelus asterias;* *Mustelus lenticulatus; Mustelus manazo;* *Mustelus schmitti; Mustelus stevensi; Mustelus sp.2*	*Mustelus antarticus* *Mustelus asterias* *Mustelus lenticulatus*
MIB:zpl:00012	FM164431	*Alopias superciliosus*	*Alopias superciliosus*
MIB:zpl:00013	FM164432	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00014	FM164433	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00015	FM164434	–	*Mustelus mustelus*
MIB:zpl:00016	FM164435	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00018	FM164436	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00019	FM164437	*Mustelus antarticus; Mustelus asterias;* *Mustelus lenticulatus; Mustelus manazo;* *Mustelus schmitti; Mustelus stevensi; Mustelus sp.2*	*Mustelus antarticus* *Mustelus asterias* *Mustelus lenticulatus*
MIB:zpl:00022	FM164438	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00023	FM164439	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00024	FM164440	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00025	FM164441	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00027	FM164442	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00028	FM164443	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00029	FM164444	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00030	FM164445	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00031	FM164446	*Mustelus antarticus; Mustelus asterias;* *Mustelus lenticulatus; Mustelus manazo;* *Mustelus schmitti; Mustelus stevensi; Mustelus sp.2*	*Mustelus antarticus* *Mustelus asterias* *Mustelus lenticulatus*
MIB:zpl:00032	FM164447	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00033	FM164448	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00034	FM164449	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00035	FM164450	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00036	FM164451	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00037	FM164452	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00038	FM164453	*Mustelus antarticus; Mustelus asterias;* *Mustelus lenticulatus; Mustelus manazo;* *Mustelus schmitti; Mustelus stevensi; Mustelus sp.2*	*Mustelus antarticus* *Mustelus asterias* *Mustelus lenticulatus*
MIB:zpl:00039	FM164454	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00040	FM164455	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00041	FM164456	*Mustelus antarticus; Mustelus asterias;* *Mustelus lenticulatus; Mustelus manazo;* *Mustelus schmitti; Mustelus stevensi; Mustelus sp.2*	*Mustelus antarticus* *Mustelus asterias* *Mustelus lenticulatus*
MIB:zpl:00042	FM164457	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00043	FM164458	*Prionace glauca*	*Prionace glauca*
MIB:zpl:00044	FM164459	*Prionace glauca*	*Prionace glauca*
MIB:zpl:00046	FM164460	–	*Mustelus mustelus*
MIB:zpl:00047	FM164461	–	*Mustelus mustelus*
MIB:zpl:00050	FM164462	*Isurus oxyrinchus*	*Isurus oxyrinchus*
MIB:zpl:00053	FM164463	*Isurus oxyrinchus*	*Isurus oxyrinchus*
MIB:zpl:00055	FM164464	*Squalus acanthias*	*Squalus acanthias*
MIB:zpl:00057	FM164465	*Isurus oxyrinchus*	*Isurus oxyrinchus*
MIB:zpl:00058	FM164466	–	–
MIB:zpl:00059	FM164467	*Isurus oxyrinchus*	*Isurus oxyrinchus*
MIB:zpl:00063	FM164468	*Isurus oxyrinchus*	*Isurus oxyrinchus*
MIB:zpl:00064	FM164469	*Mustelus antarticus; Mustelus asterias;* *Mustelus lenticulatus; Mustelus manazo;* *Mustelus schmitti; Mustelus stevensi; Mustelus sp.2*	*Mustelus antarticus* *Mustelus asterias* *Mustelus lenticulatus*
MIB:zpl:00065	FM164470	*Isurus oxyrinchus*	*Isurus oxyrinchus*

APPENDIX II.6: Among lineages and outgroups genetic divergence of four mitochondrial markers based on the pairwise distance calculation for the taxa considered in this study treated (a) as divided in the two molecular lineages identified in this study; (b) as the traditional nomenclature based on morphological plumage characters, therefore partitioned in *P. webbianus* and *P. alphonsianus*..

a

		P. guttaticollis	*P. nipalensis*	*P. davidianus*	*P. gularis*	Lineage I	Lineage II
cyt b	*P. guttaticollis*	_					
	P. nipalensis	14.42%	_				
	P. davidianus	12.92%	11.16%	_			
	P. gularis	12.95%	12.94%	13.19%	_		
	Lineage I	13.85%	14.46%	14.44%	13.17%	_	
	Lineage II	13.98%	15.12%	14.28%	12.28%	1.94%	_

		P. guttaticollis	*P. nipalensis*	*P. davidianus*	*P. gularis*	Lineage I	Lineage II
16S rDNA	*P. guttaticollis*	_					
	P. nipalensis	4.56%	_				
	P. davidianus	4.01%	3.63%	_			
	P. gularis	4.01%	4.41%	4.81%	_		
	Lineage I	3.76%	4.36%	3.24%	5.51%	_	
	Lineage II	3.79%	4.37%	3.24%	5.53%	0.05%	_

		P. guttaticollis	*P. nipalensis*	*P. davidianus*	*P. gularis*	Lineage I	Lineage II
12S rDNA	*P. guttaticollis*	_					
	P. nipalensis	7.04%	_				
	P. davidianus	5.78%	3.01%	_			
	P. gularis	4.02%	7.29%	6.78%	_		
	Lineage I	4.82%	4.82%	5.31%	7.05%	_	
	Lineage II	4.80%	4.80%	5.29%	7.07%	0.02%	_

		P. guttaticollis	Lineage I	Lineage II
cox I	*P. guttaticollis*	_		
	Lineage I	13.48%	_	
	Lineage II	13.36%	1.41%	_

b

		P. guttaticollis	*P. nipalensis*	*P. davidianus*	*P. gularis*	*P. webbianus*	*P. alphonsianus*
cyt b	*P. guttaticollis*	_					
	P. nipalensis	14.42%	_				
	P. davidianus	12.92%	11.16%	_			
	P. gularis	12.95%	12.94%	13.19%	_		
	Morphotype webbianus	13.92%	14.69%	14.31%	12.79%	_	
	Morphotype alphonsianus	13.83%	14.50%	14.54%	13.22%	0.92%	_

		P. guttaticollis	*P. nipalensis*	*P. davidianus*	*P. gularis*	*P. webbianus*	*P. alphonsianus*
16S rDNA	*P. guttaticollis*	_					
	P. nipalensis	4.56%	_				
	P. davidianus	4.01%	3.63%	_			
	P. gularis	4.01%	4.41%	4.81%	_		
	Morphotype webbianus	3.75%	4.36%	3.23%	5.50%	_	
	Morphotype alphonsianus	3.79%	4.37%	3.24%	5.53%	0.06%	_

		P. guttaticollis	*P. nipalensis*	*P. davidianus*	*P. gularis*	*P. webbianus*	*P. alphonsianus*
12S rDNA	*P. guttaticollis*	_					
	P. nipalensis	7.04%	_				
	P. davidianus	5.78%	3.01%	_			
	P. gularis	4.02%	7.29%	6.78%	_		
	Morphotype webbianus	4.86%	4.86%	5.35%	7.01%	_	
	Morphotype alphonsianus	4.80%	4.80%	5.29%	7.07%	0.06%	_

		P. guttaticollis	*P. webbianus*	*P. alphonsianus*
cox I	*P. guttaticollis*	_		
	Morphotype webbianus	13.43%	_	
	Morphotype alphonsianus	13.50%	0.67%	_

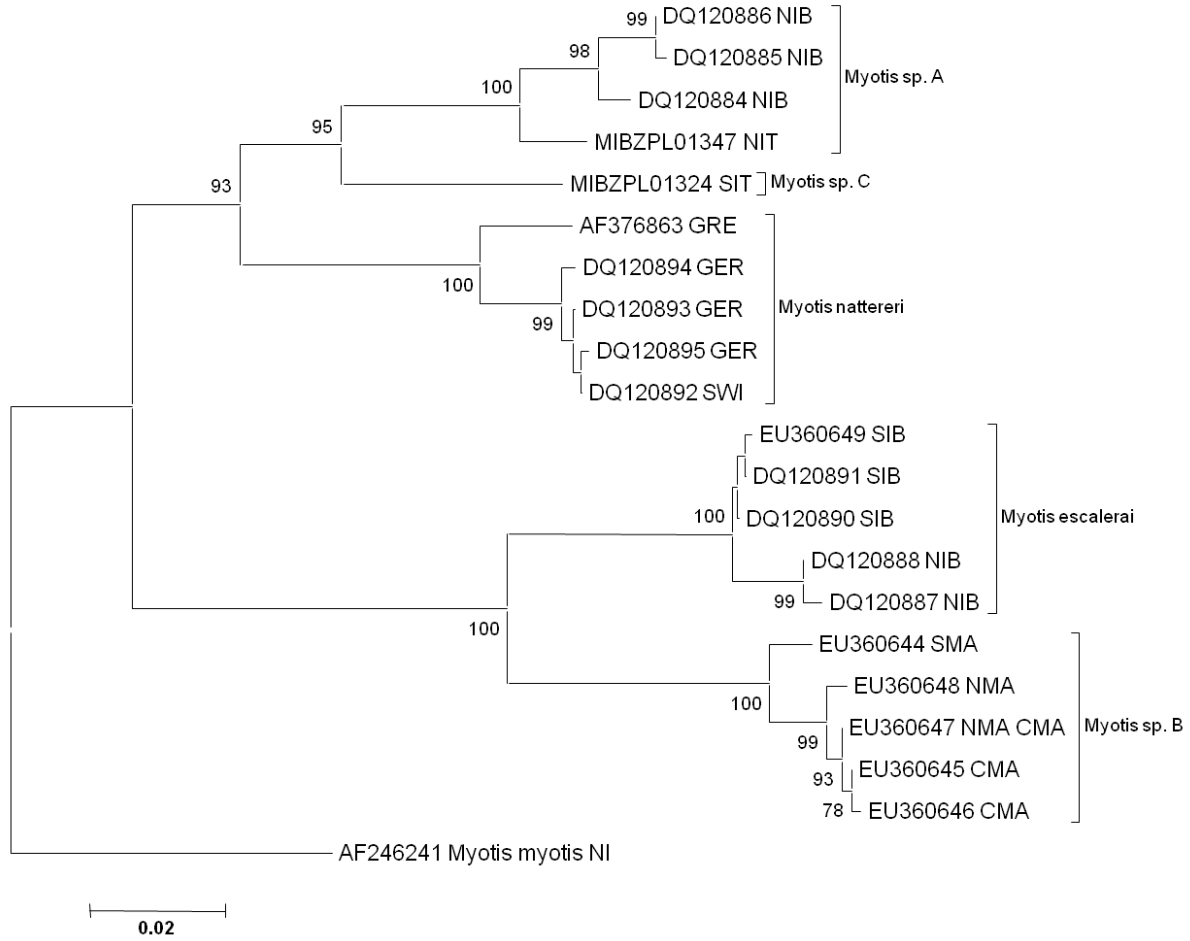
APPENDIX II.7: List of Paradoxornis samples included in the haplotype network reconstruction showed in Appendix Figure III.2 with reference to sample name, sampling locality, haplotype name and lineage membership in the network reconstruction. Additional details can be retrieved from Appendix Table I.2.

Sample	Haplotype	Lineage	Sampling locality
VTP15	A11	I	Dongzhai NR, Henan, China
VTP13	A19	I	Dongzhai NR, Henan, China
VTP12	A21	I	Dongzhai NR, Henan, China
VTP14	A21	I	Dongzhai NR, Henan, China
AN13355_D10	A01	I	Palude Brabbia NR, Varese, NW Italy
AS22871_ID07	A01	I	Palude Brabbia NR, Varese, NW Italy
AS22878_IH16	A01	I	Palude Brabbia NR, Varese, NW Italy
AT27544_H17	A01	I	Palude Brabbia NR, Varese, NW Italy
AV73538_20	A01	I	Palude Brabbia NR, Varese, NW Italy
AV73544_H08	A01	I	Palude Brabbia NR, Varese, NW Italy
AV73547_37	A01	I	Palude Brabbia NR, Varese, NW Italy
AV73550_41	A01	I	Palude Brabbia NR, Varese, NW Italy
AV73635_H09	A01	I	Palude Brabbia NR, Varese, NW Italy
D01	A01	I	Palude Brabbia NR, Varese, NW Italy
ID05	A01	I	Palude Brabbia NR, Varese, NW Italy
ID06	A01	I	Palude Brabbia NR, Varese, NW Italy
ID08	A01	I	Palude Brabbia NR, Varese, NW Italy
web-s-3158	A12	I	Qinling Mt., Shaanxi, China
web-s-3159	A16	I	Qinling Mt., Shaanxi, China
web-s-834	A23	I	Qinling Mt., Shaanxi, China
web-s-836	A23	I	Qinling Mt.,Shaanxi, China
VTP4	A11	I	Dongtan, Chongming Island, Shanghai, China
VTP1	A05	II	Dongtan, Chongming Island, Shanghai, China
VTP2	A05	II	Dongtan, Chongming Island, Shanghai, China
VTP22	A02	I	Dunjiangyan, Sichuan, China
VTP21	A11	I	Dunjiangyan, Sichuan, China
VTP24	A15	I	Dunjiangyan, Sichuan, China
VTP25	A17	I	Dunjiangyan, Sichuan, China
VTP23	A22	I	Dunjiangyan, Sichuan, China
VTP5	A11	I	East Tianjin, China
VTP11	A20	I	East Tianjin, China
VTP6	A20	I	East Tianjin, China
VTP9	A04	II	East Tianjin, China
VTP8	A05	II	East Tianjin, China
VTP7	A07	II	East Tianjin, China
VTP10	A08	II	East Tianjin, China
web-a-1652	A11	I	suburb of Kunming, Yunnan, China
web-a-1634	A18	I	suburb of Kunming, Yunnan, China

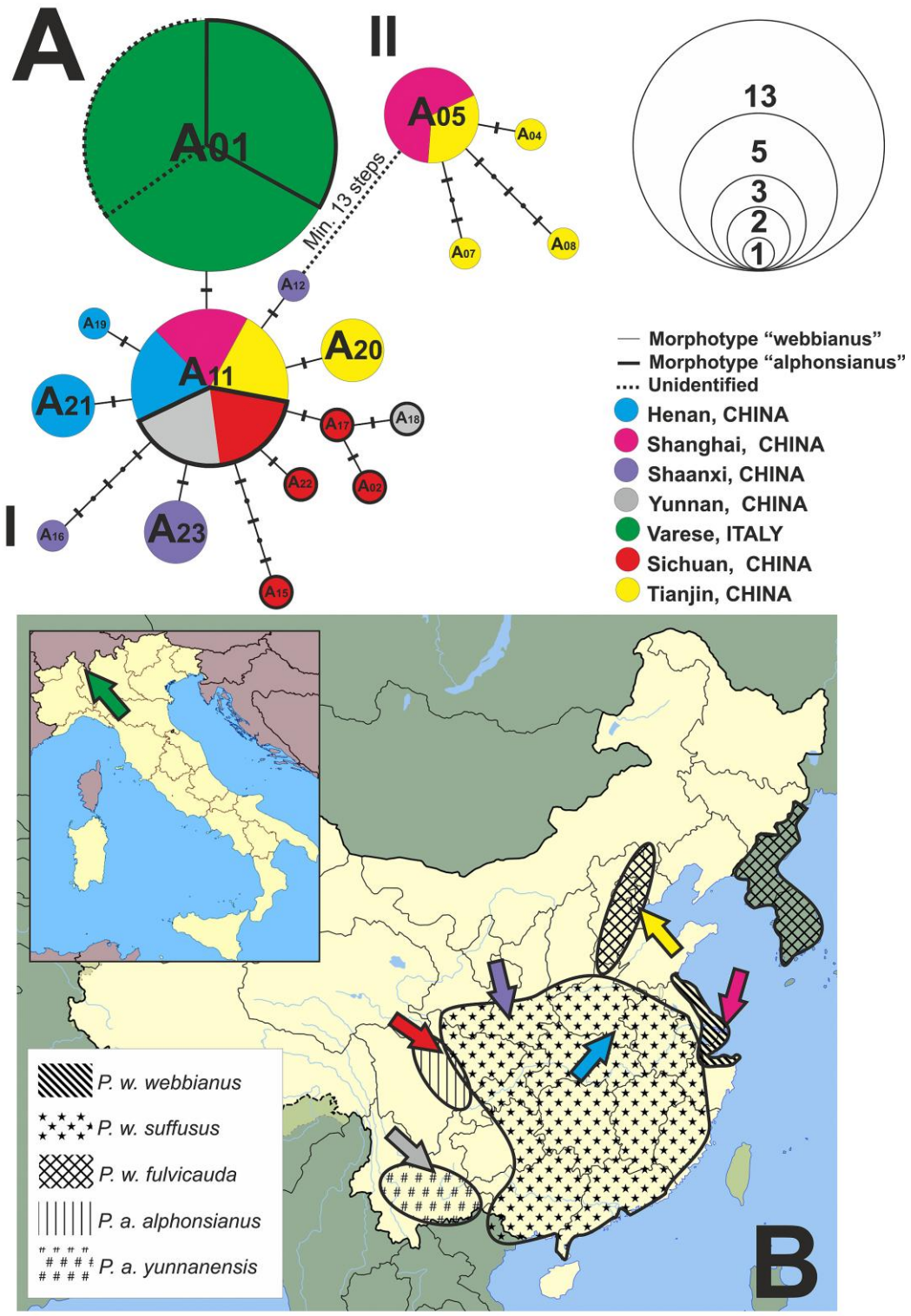
APPENDIX II.8: List of BLAST matches showing the highest maximum similarity values with *coxI* sequences obtained for dataset D1

Acc. Num.	Species	Phylum	Acc. Num.	Species	Phylum
EU244607	*Richtersius coronifer*	Tardigrada	EU652745	*Caenorhabditis brenneri*	Nematoda
EU244597	*Macrobiotus* sp.	Tardigrada	X54252	*Caenorhabditis elegans*	Nematoda
EU244598	*Macrobiotus* sp.	Tardigrada	AY649858	*Centruroides exilicauda*	Arthropoda
FJ176208	*Macrobiotus macrocalix*	Tardigrada	AY995831	*Centruroides sculpturatus*	Arthropoda
FJ176217	*Macrobiotus macrocalix*	Tardigrada	EU381075	*Centruroides vittatus*	Arthropoda
AY598773	*Macrobiotus* sp.	Tardigrada	DQ518421	*Cyrtarachne bufo*	Arthropoda
AY598775	*Macrobiotus terminalis*	Tardigrada	DQ518422	*Paraplectana tsushimensis*	Arthropoda
DQ890154	*Philodina flaviceps*	Rotifera	AY770785	*Agelenopsis aperta*	Arthropoda
EF650513	*Mniobia russeola*	Rotifera	AY788624	*Eresiae unice*	Arthropoda
EF650542	*Uncultured bdelloid rotifer*	Rotifera	EF050291	*Echinotheridion otlum*	Arthropoda
EF173261	*Adineta vaga*	Rotifera	EF127451	*Palpada ruficeps*	Arthropoda
FJ426422	*Macrotrachela quadricornifera*	Rotifera	EF537064	*Hypochilus pococki*	Arthropoda
AY129173	*Phytophthora colocasiae*	Oomycota	EU367539	*Phytomyza thalictrella*	Arthropoda
DQ092904	*Lumbricus rubellus*	Anellida	FJ525327	*Neoscona crucifera*	Arthropoda
EF043402	*Heterorhabditis bacteriophora*	Nematoda	FJ607553	*Araneus diadematus*	Arthropoda
DQ408633	*Heligmosomoides polygyrus*	Nematoda	FM877914	*Endomychus biguttatus*	Arthropoda
EU407780	*Caenorhabditis* sp.	Nematoda	GU109550	*Selenops bifurcatus*	Arthropoda
EU407790	*Caenorhabditis briggsae*	Nematoda	AY090205	*Euploea camaralzeman*	Arthropoda

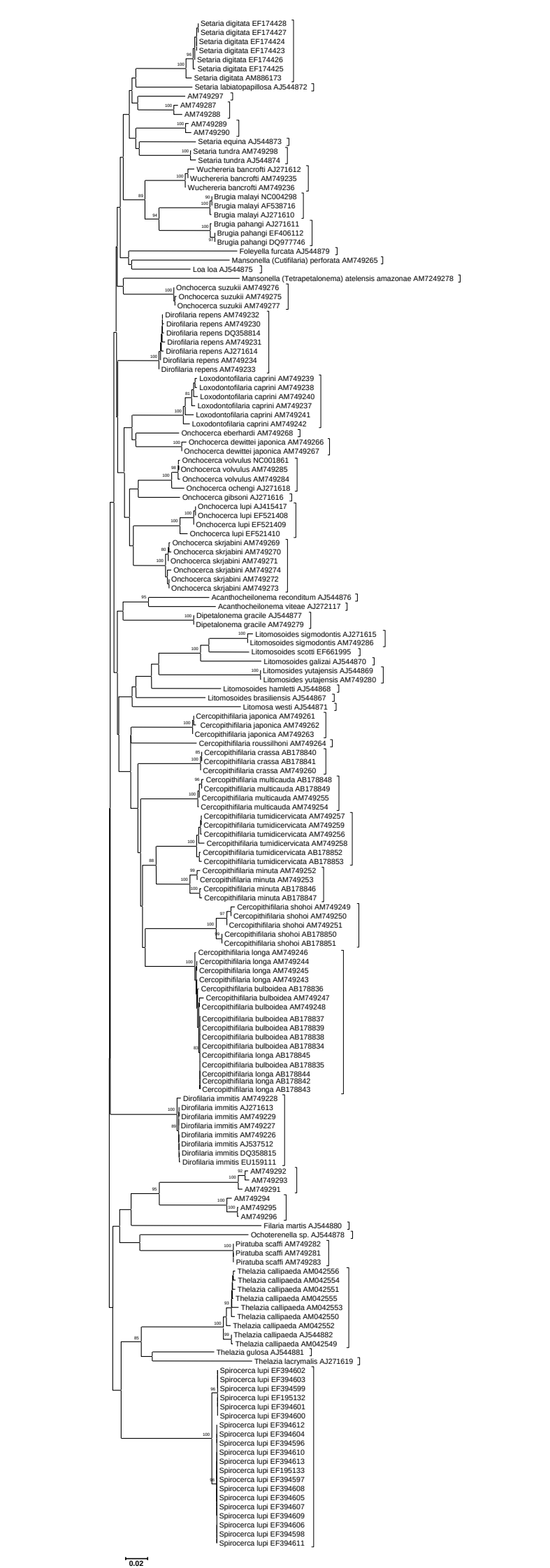
APPENDIX III.1



APPENDIX III.2



Myotis blythii MIBZPL01231 CI
Myotis myotis MIBZPL01234 CI
Myotis blythii MIBZPL01218 CI
Myotis myotis MIBZPL01255 SI
Myotis blythii MIBZPL01246 SI
Myotis myotis MIBZPL01265 SI
Myotis blythii MIBZPL01219 CI
Myotis myotis MIBZPL00287 NI
Myotis myotis MIBZPL00282 NI
Myotis myotis MIBZPL00377 NI
Myotis myotis MIBZPL00284 NI
Myotis myotis MIBZPL00338 NI
Myotis blythii MIBZPL00499 NI
Myotis blythii MIBZPL01178 NI
Myotis blythii MIBZPL01263 SI
Myotis myotis MIBZPL01259 SI
Myotis myotis MIBZPL01248 SI
Myotis blythii MIBZPL00285 NI
Myotis myotis MIBZPL01336 NI
Myotis myotis MIBZPL01339 NI
Myotis myotis MIBZPL00280 NI
Myotis myotis MIBZPL00275 NI
Myotis myotis MIBZPL00300 NI
Myotis nattereri GU270561 IR
Myotis nattereri MIBZPL00318 NI
Myotis nattereri MIBZPL00505 NI
Myotis nattereri MIBZPL00384 NI
Myotis nattereri MIBZPL01347 NI
Myotis nattereri MIBZPL00326 NI
Myotis nattereri MIBZPL00356 NI
Myotis nattereri MIBZPL00331 NI
Myotis nattereri MIBZPL00383 NI
Myotis nattereri MIBZPL01266 SI
Myotis nattereri MIBZPL01249 SI
Myotis nattereri MIBZPL01324 CI
Myotis nattereri MIBZPL01273 SI
Myotis nattereri MIBZPL01286 SI
Myotis nattereri MIBZPL01308 SI
Myotis nattereri MIBZPL01284 SI
Myotis nattereri MIBZPL01285 SI
Myotis nattereri MIBZPL01291 SI
Myotis nattereri MIBZPL01292 SI
Myotis nattereri MIBZPL01282 SI
Myotis nattereri MIBZPL01233 CI
Myotis nattereri MIBZPL01309 SI
Myotis punicus MIBZPL01566 SAR
Myotis punicus MIBZPL01567 SAR
Myotis punicus MIBZPL01568 SAR
Myotis daubentonii GU270554 IR
Myotis daubentonii MIBZPL00341 NI
Myotis daubentonii MIBZPL00299 NI
Myotis daubentonii MIBZPL00311 NI
Myotis daubentonii MIBZPL00313 NI
Myotis daubentonii MIBZPL01344 NI
Myotis bechsteinii MIBZPL01271 SI
Myotis bechsteinii MIBZPL01283 SI
Myotis bechsteinii GU270563 IR
Myotis bechsteinii MIBZPL00319 NI
Myotis bechsteinii MIBZPL01327 CI
Myotis bechsteinii MIBZPL00314 NI
Myotis bechsteinii GU270562 IR
Myotis bechsteinii MIBZPL00364 NI
Myotis alcathoe MIBZPL01278 SI
Myotis sp MIBZPL01289 SI
Myotis alcathoe GU270560 IR
Myotis sp MIBZPL01281 SI
Myotis sp MIBZPL01287 SI
Myotis sp MIBZPL1230 CI
Myotis emarginatus MIBZPL00322 NI
Myotis emarginatus MIBZPL01242 CI
Myotis emarginatus MIBZPL00267 NI
Myotis emarginatus MIBZPL01247 SI
Myotis emarginatus MIBZPL01325 CI
Myotis emarginatus MIBZPL01294 SI
Myotis emarginatus GU270553 IR
Myotis emarginatus MIBZPL00315 NI
Myotis brandtii GU270564 IR
Myotis brandtii GU270565 IR
Myotis brandtii MIBZPL01279 SI
Myotis capaccinii MIBZPL01179 NI
Myotis capaccinii MIBZPL01176 NI
Myotis capaccinii MIBZPL01181 NI
Myotis sp MIBZPL01211 CI
Myotis sp MIBZPL01301 CI
Myotis sp MIBZPL01221 CI
Myotis sp MIBZPL01222 CI
Myotis sp MIBZPL01319 SI
Myotis mystacinus MIBZPL01293 SI
Myotis sp MIBZPL01223 CI
Myotis sp MIBZPL01235 CI
Myotis sp MIBZPL01228 CI
Myotis sp MIBZPL01256 SI
Myotis sp MIBZPL01302 CI
Myotis sp MIBZPL01303 CI
Myotis mystacinus MIBZPL01240 CI
Myotis sp MIBZPL01214 CI
Myotis sp MIBZPL01216 CI
Myotis mystacinus MIBZPL03775 NI
Myotis mystacinus GU270557 IR
Myotis mystacinus MIBZPL02289 SAR
Myotis mystacinus MIBZPL02290 SAR
Myotis mystacinus GU270555 IR
Myotis mystacinus GU270556 IR
Myotis mystacinus GU270558 IR
Myotis mystacinus GU270559 IR
Barbastella barbastellus MIBZPL01300 CI
Barbastella barbastellus MIBZPL01264 SI
Barbastella barbastellus MIBZPL01485 CI
Miniopterus schreibersii MIBZPL00286 NI
Miniopterus schreibersii MIBZPL00290 NI
Miniopterus schreibersii MIBZPL01335 NI
Miniopterus schreibersii MIBZPL00281 NI
Miniopterus schreibersii MIBZPL00292 NI
Plecotus austriacus MIBZPL01252 SI
Plecotus austriacus MIBZPL01497 CI
Plecotus macrobullaris MIBZPL00256 NI
Plecotus sp MIBZPL00262 NI
Plecotus sp MIBZPL01189 NI
Plecotus macrobullaris MIBZPL01200 NI
Plecotus sp MIBZPL03414 NI
Plecotus sardus MIBZPL01574 SAR
Plecotus sardus MIBZPL01575 SAR
Plecotus sardus MIBZPL01572 SAR
Plecotus auritus MIBZPL01253 SI
Plecotus auritus MIBZPL01269 SI
Plecotus auritus MIBZPL01276 SI
Plecotus auritus MIBZPL01227 CI
Plecotus auritus MIBZPL01190 NI
Plecotus auritus MIBZPL01346 NI
Plecotus sp MIBZPL00265 NI
Plecotus auritus MIBZPL00378 NI
Plecotus auritus MIBZPL01206 CI
Plecotus sp MIBZPL00269 NI
Plecotus sp MIBZPL00268 NI
Plecotus auritus MIBZPL01345 NI
Plecotus sp MIBZPL00270 NI
Hypsugo savii MIBZPL00321 NI
Hypsugo savii MIBZPL01232 CI
Hypsugo savii MIBZPL00254 NI
Hypsugo savii MIBZPL01267 SI
Hypsugo savii MIBZPL01295 CI
Hypsugo savii MIBZPL01243 CI
Hypsugo savii MIBZPL01270 SI
Pipistrellus nathusii MIBZPL01328 NI
Pipistrellus nathusii MIBZPL01331 NI
Pipistrellus nathusii MIBZPL01180 NI
Pipistrellus nathusii MIBZPL01334 NI
Nyctalus leisleri GU270566 IR
Nyctalus leisleri MIBZPL02259 NI
Nyctalus leisleri MIBZPL00261 SW
Nyctalus leisleri MIBZPL00535 SW
Nyctalus leisleri MIBZPL01207 CI
Nyctalus leisleri MIBZPL00536 SW
Nyctalus leisleri MIBZPL01323 CI
Nyctalus leisleri MIBZPL01268 SI
Nyctalus noctula MIBZPL01476 NI
Pipistrellus kuhlii MIBZPL00289 NI
Pipistrellus kuhlii MIBZPL01322 CI
Pipistrellus kuhlii MIBZPL01312 CI
Pipistrellus kuhlii MIBZPL01251 SI
Pipistrellus kuhlii MIBZPL00240 NI
Pipistrellus kuhlii MIBZPL00253 NI
Pipistrellus kuhlii MIBZPL01297 CI
Pipistrellus pipistrellus MIBZPL00283 NI
Pipistrellus pipistrellus MIBZPL02254 NI
Pipistrellus pipistrellus MIBZPL00291 NI
Pipistrellus pipistrellus MIBZPL02255 NI
Pipistrellus pipistrellus MIBZPL00288 NI
Pipistrellus pipistrellus MIBZPL01195 NI
Pipistrellus pipistrellus MIBZPL01337 NI
Pipistrellus sp MIBZPL01241 CI
Pipistrellus pipistrellus MIBZPL01321 SI
Pipistrellus pipistrellus MIBZPL01224 CI
Pipistrellus pipistrellus MIBZPL01290 SI
Pipistrellus sp MIBZPL01239 CI
Pipistrellus sp MIBZPL02288 SAR
Pipistrellus pipistrellus MIBZPL00272 NI
Pipistrellus pipistrellus MIBZPL02253 NI
Pipistrellus sp MIBZPL03819 NI
Pipistrellus sp MIBZPL03824 NI
Pipistrellus sp MIBZPL03825 NI
Pipistrellus pygmaeus MIBZPL02285 SAR
Pipistrellus sp MIBZPL03818 NI
Pipistrellus sp MIBZPL03820 NI
Pipistrellus sp MIBZPL03828 NI
Pipistrellus sp MIBZPL03815 NI
Pipistrellus sp MIBZPL03827 NI
Pipistrellus sp MIBZPL03817 NI
Pipistrellus sp MIBZPL03826 NI
Pipistrellus sp MIBZPL03822 NI
Pipistrellus sp MIBZPL03816 NI
Pipistrellus sp MIBZPL03823 NI
Pipistrellus pygmaeus MIBZPL02287 SAR
Pipistrellus sp MIBZPL03821 NI
Eptesicus serotinus MIBZPL00333 NI
Eptesicus serotinus MIBZPL01342 NI
Eptesicus serotinus MIBZPL00349 NI
Eptesicus serotinus MIBZPL00369 NI
Eptesicus serotinus MIBZPL00312 NI
Eptesicus nilsonii MIBZPL02257 NI
Eptesicus nilsonii MIBZPL02258 NI
Tadarida teniotis MIBZPL01512 SI
Tadarida teniotis MIBZPL01570 SAR
Tadarida teniotis MIBZPL01569 SAR
Rhinolophus hipposideros MIBZPL00359 NI
Rhinolophus hipposideros MIBZPL00387 NI
Rhinolophus ferrumequinum MIBZPL00376 NI
Rhinolophus ferrumequinum MIBZPL00502 NI
Rhinolophus ferrumequinum MIBZPL00393 NI
Rhinolophus ferrumequinum MIBZPL00316 NI
Rhinolophus ferrumequinum MIBZPL01204 NI
Rhinolophus ferrumequinum MIBZPL00327 NI
Rhinolophus ferrumequinum MIBZPL00354 NI
Rhinolophus euryale MIBZPL00271 NI
Rhinolophus euryale MIBZPL00310 NI
Rhinolophus euryale MIBZPL00274 NI
Rhinolophus euryale MIBZPL01341 NI
Rhinolophus euryale MIBZPL00295 NI
Rhinolophus mehelyi MIBZPL01559 SAR
Rhinolophus mehelyi MIBZPL01562 SAR
Rhinolophus mehelyi MIBZPL01560 SAR
Rhinolophus mehelyi MIBZPL01561 SAR
Rhinolophus mehelyi MIBZPL01563 SAR
0.02


Setaria digitata EF174428
Setaria digitata EF174427
Setaria digitata EF174424
Setaria digitata EF174423
Setaria digitata EF174426
Setaria digitata EF174425
Setaria digitata AM886173
Setaria labiatopapillosa AJ544872
AM749297
AM749287
AM749288
AM749289
AM749290
Setaria equina AJ544873
Setaria tundra AM749298
Setaria tundra AJ544874
Wuchereria bancrofti AJ271612
Wuchereria bancrofti AM749235
Wuchereria bancrofti AM749236
Brugia malayi NC004298
Brugia malayi AF538716
Brugia malayi AJ271610
Brugia pahangi AJ271611
Brugia pahangi EF406112
Brugia pahangi DQ977746
Foleyella furcata AJ544879
Mansonella (Cutifilaria) perforata AM749265
Loa loa AJ544875
Mansonella (Tetrapetalonema) atelensis amazonae AM7249278
Onchocerca suzukii AM749276
Onchocerca suzukii AM749275
Onchocerca suzukii AM749277
Dirofilaria repens AM749232
Dirofilaria repens AM749230
Dirofilaria repens DQ358814
Dirofilaria repens AM749231
Dirofilaria repens AJ271614
Dirofilaria repens AM749234
Dirofilaria repens AM749233
Loxodontofilaria caprini AM749239
Loxodontofilaria caprini AM749238
Loxodontofilaria caprini AM749240
Loxodontofilaria caprini AM749237
Loxodontofilaria caprini AM749241
Loxodontofilaria caprini AM749242
Onchocerca eberhardi AM749268
Onchocerca dewittei japonica AM749266
Onchocerca dewittei japonica AM749267
Onchocerca volvulus NC001861
Onchocerca volvulus AM749285
Onchocerca volvulus AM749284
Onchocerca ochengi AJ271618
Onchocerca gibsoni AJ271616
Onchocerca lupi AJ415417
Onchocerca lupi EF521408
Onchocerca lupi EF521409
Onchocerca lupi EF521410
Onchocerca skrjabini AM749269
Onchocerca skrjabini AM749270
Onchocerca skrjabini AM749271
Onchocerca skrjabini AM749274
Onchocerca skrjabini AM749272
Onchocerca skrjabini AM749273
Acanthocheilonema reconditum AJ544876
Acanthocheilonema viteae AJ272117
Dipetalonema gracile AJ544877
Dipetalonema gracile AM749279
Litomosoides sigmodontis AJ271615
Litomosoides sigmodontis AM749286
Litomosoides scotti EF661995
Litomosoides galizai AJ544870
Litomosoides yutajensis AJ544869
Litomosides yutajensis AM749280
Litomosoides hamletti AJ544868
Litomosoides brasiliensis AJ544867
Litomosa westi AJ544871
Cercopithifilaria japonica AM749261
Cercopithifilaria japonica AM749262
Cercopithifilaria japonica AM749263
Cercopithifilaria roussilhoni AM749264
Cercopithifilaria crassa AB178840
Cercopithifilaria crassa AB178841
Cercopithifilaria crassa AM749260
Cercopithifilaria multicauda AB178848
Cercopithifilaria multicauda AB178849
Cercopithifilaria multicauda AM749255
Cercopithifilaria multicauda AM749254
Cercopithifilaria tumidicervicata AM749257
Cercopithifilaria tumidicervicata AM749259
Cercopithifilaria tumidicervicata AM749256
Cercopithifilaria tumidicervicata AM749258
Cercopithifilaria tumidicervicata AB178852
Cercopithifilaria tumidicervicata AB178853
Cercopithifilaria minuta AM749252
Cercopithifilaria minuta AM749253
Cercopithifilaria minuta AB178846
Cercopithifilaria minuta AB178847
Cercopithifilaria shohoi AM749249
Cercopithifilaria shohoi AM749250
Cercopithifilaria shohoi AM749251
Cercopithifilaria shohoi AB178850
Cercopithifilaria shohoi AB178851
Cercopithifilaria longa AM749246
Cercopithifilaria longa AM749244
Cercopithifilaria longa AM749245
Cercopithifilaria longa AM749243
Cercopithifilaria bulboidea AB178836
Cercopithifilaria bulboidea AM749247
Cercopithifilaria bulboidea AM749248
Cercopithifilaria bulboidea AB178837
Cercopithifilaria bulboidea AB178839
Cercopithifilaria bulboidea AB178838
Cercopithifilaria bulboidea AB178834
Cercopithifilaria longa AB178845
Cercopithifilaria bulboidea AB178835
Cercopithifilaria longa AB178844
Cercopithifilaria longa AB178842
Cercopithifilaria longa AB178843
Dirofilaria immitis AM749228
Dirofilaria immitis AJ271613
Dirofilaria immitis AM749229
Dirofilaria immitis AM749227
Dirofilaria immitis AM749226
Dirofilaria immitis AJ537512
Dirofilaria immitis DQ358815
Dirofilaria immitis EU159111
AM749292
AM749293
AM749291
AM749294
AM749295
AM749296
Filaria martis AJ544880
Ochoterenella sp. AJ544878
Piratuba scaffi AM749282
Piratuba scaffi AM749281
Piratuba scaffi AM749283
Thelazia callipaeda AM042556
Thelazia callipaeda AM042554
Thelazia callipaeda AM042551
Thelazia callipaeda AM042555
Thelazia callipaeda AM042553
Thelazia callipaeda AM042550
Thelazia callipaeda AM042552
Thelazia callipaeda AJ544882
Thelazia callipaeda AM042549
Thelazia gulosa AJ544881
Thelazia lacrymalis AJ271619
Spirocerca lupi EF394602
Spirocerca lupi EF394603
Spirocerca lupi EF394599
Spirocerca lupi EF195132
Spirocerca lupi EF394601
Spirocerca lupi EF394600
Spirocerca lupi EF394612
Spirocerca lupi EF394604
Spirocerca lupi EF394596
Spirocerca lupi EF394610
Spirocerca lupi EF394613
Spirocerca lupi EF195133
Spirocerca lupi EF394597
Spirocerca lupi EF394608
Spirocerca lupi EF394605
Spirocerca lupi EF394607
Spirocerca lupi EF394609
Spirocerca lupi EF394606
Spirocerca lupi EF394598
Spirocerca lupi EF394611
0.02

APPENDIX IV.1: ADDITIONAL ANALYSES CONDUCTED ON PARROTBILLS DATASET (CASE STUDY A2).

Alignments

Sequences of the four mitochondrial markers (*cyt* b 875 bp, *coxI* 663 bp, 16S rDNA 528 bp, 12S rDNA 397 bp) have been merged in a single alignment of 2463 bp. Only haplotypes were selected by using the online application DNAcollapser ver. 1.0 (http://www.birc.au.dk/fabox/dnacollapser.php) as reported in Appendix Table I.2. Additional *cyt b* sequences of four vinous-throated individuals from Shaanxi and of two ashy-throated individuals from Yunnan provinces (People's Republic of China) (See Appendix Tables I.2 and II.7), were provided by a Chinese researcher (Carol Yeung). These individual were not processed for the other genes.

Haplotype network reconstruction

Haplotype network reconstruction (Figure 4.6; A) was performed using the software TCS, version 1.21 (Clement et al., 2000). This software employs the method of Templeton et al. (1992) and it calculates the number of mutational steps by which pairwise haplotypes differ, computing the probability of parsimony for pairwise differences until the probability exceeds 0.95. The minimum number of mutational steps required to connect the different groups of haplotypes obtained using the Templeton et al. (1992) method was identified using the fix connection limit option, as implemented in TCS software. A second haplotype network reconstruction was performed using only *cyt b* sequences of the individuals included in the previous analysis and considering additional sequences of which only *cyt b* sequences were available. Data are provided in Appendix Tables I.2 and II.7

Nucleotide diversity and p-distances

To give a better characterization of the genetic variability within and between the two taxa, the nucleotide diversity (Π) of Nei (1987), the average number of nucleotide differences per site between sequences, was calculated for both taxa, for each sampling locality and for the two identified lineages (Table 4.9) using DnaSP software (Version 5.10.3; Librado and Rozas, 2009). Morphologically undetermined samples were excluded.

Uncorrected genetic divergences (p-distance, transformed into percent) of the four *Paradoxornis* species of which sequences were available in Genbank [*P. davidianus* (*cyt b*: AF484872; 16S rDNA: AF484378; 12S rDNA: AF484921), *P. gularis* (*cyt b*: AF484873; 16S rDNA: AF484379; 12S rDNA: AF484922), *P. guttaticollis* (*cyt b*: AF484874, DQ008488, EU447104; *cox* I: EU447059; 16S rDNA: AF484380; 12S rDNA: AF484923) and *P. nipalensis* (*cyt b*: AF484875; 16S rDNA: AF484381; 12S rDNA: AF484924)] were compared with the data obtained in this study (SM5). In particular we compared: *i*) p-distance values of *Paradoxornis* species versus the two molecular lineages as identified in this study (Appendix Table II.6, a); *ii*) p-distance values of *Paradoxornis* species versus the vinous-throated and the ashy-throated parrotbills as identified as two different taxa considering only the morphological plumage characters (Appendix Table II.6, b).