Four years of DNA barcoding: current advances and prospects.

Lise Frézal, Raphael Leblois

4 years of DNA barcoding : current advances and prospects

Authors : Lise Frézal [1], Raphael Leblois[2]*

1 Laboratoire de Biologie intégrative des populations, Ecole Pratique des Hautes Etudes, Paris, France

2 Unité Origine, Structure et Evolution de la Biodiversité UMR 5202 CNRS/MNHN, Muséum National d'Histoire Naturelle, Paris, France

* Corresponding author (Phone: +33 (0)1 40 79 33 49; Fax: +33 (0)1 40 79 33 42; Email: leblois@mnhn.fr)

Abstract

L. Frézal and R. Leblois – 4 years of DNA barcoding: current advances and prospects - Infection, Genetics and Evolution

Research using cytochrome *c* oxidase barcoding techniques on zoological specimens was initiated by Hebert et al. (2003). By March 2004, the Consortium for the Barcode of Life started to promote the use of a standardized DNA barcoding approach, consisting of identifying a specimen as belonging to a certain animal species based on a single universal marker: the DNA barcode sequence. Over the last four years, this approach has become increasingly popular and advances as well as limitations have clearly emerged as increasing amounts of organisms have been studied. Our purpose is to briefly expose DNA Barcode of Life principles, pros and cons, relevance and universality. The initially proposed Barcode of life framework has greatly evolved, giving rise to a flexible description of DNA barcoding and a larger range of applications.

Index Key words: **DNA barcode, cytochrome *c* oxidase, COI, DNA taxonomy, species identification, international species databank, BOLD**

1- **Introduction**

Species identification and classification have traditionally been the specialist domain of taxonomists, providing a nomenclatural backbone and a key prerequisite for numerous biological studies. Indeed, today's society has to resolve many crucial biological issues, among which are the need to maintain biodiversity, to ensure bio-security, to protect species and to avoid pandemics. The achievement of such goals and the success of subsequent action programs require efficient global networks and rely on our capacity to identify any described species. As Dayrat (2005) clearly expressed, 'delineating species boundaries correctly – and also identifying species – are crucial to the discovery of life's diversity because it determines whether different individual organisms are members of the same entity or not'. The identification of species depends on the knowledge held by taxonomists whose work cannot cover all taxon identification requested by non specialists. To deal with these difficulties, the 'DNA Barcode of Life' project aims to develop a standardized, rapid and inexpensive species identification method accessible to non-specialists (i.e. non-taxonomists).

The idea of a standardized molecular identification system emerged progressively during the 1990s with the development of PCR-based approaches for species identification. Molecular identification has largely been applied to bacterial studies, microbial biodiversity surveys (e.g. Woese, 1996; Zhou et al., 1997) and routine pathogenic strains diagnoses (e.g. Maiden et al., 1998, Sugita et al., 1998; Wirth et al., 2006) due to a need for culture-independent identification systems. PCR-based methods have also been frequently used in fields related to taxonomy, food and forensic molecular identification (Teletchea et al., 2005) and for identification of eukaryotic pathogens and vectors (e.g. Walton et al., 1999). Several universal systems for molecular-based identification have been used for lower taxa (e.g. nematodes, Floyd et al., 2002) but were not successfully implemented for broader scopes. The Barcode of Life project soon after became that attempt, aiming to create a universal system for a eukaryotic species inventory based on a standard molecular approach. It was initiated in 2003 by researchers at the University of Guelph in Ontario, Canada (http://www.barcoding.si.edu) and promoted in 2004 by the international initiative 'Consortium for the Barcode of Life' (CBOL). By then, it had more than 150 member organizations from 45 countries including natural history museums, zoos, herbaria, botanical gardens, university departments as well as private companies and governmental organizations. The DNA barcode project does not have the ambition to build the tree of life nor to perform molecular taxonomy (Erbach and Holdrege, 2005; Gregory, 2005), but rather to produce a simple diagnostic tool based on strong taxonomic knowledge that is collated in the DNA barcode reference library (Schindel and Miller, 2005). The DNA Barcode of Life Data System (BOLD, http://www.boldsystems.org) has

progressively been developed since 2004 and was officially established in 2007 (Ratnasingham and Hebert, 2007). This data system enables the acquisition, storage, analysis and publication of DNA barcode records.

In the present paper we briefly review the current state of DNA barcode advances, trends and pitfalls. The main methods of the DNA barcoding approach are given. The feasibility of a universal barcoding approach and interest in the DNA barcoding approach for microbial studies are discussed.

2- The DNA Barcoding approach: definitions and objectives

2-1 DNA barcode definition and primary objectives

The DNA barcode project was initially conceived as a standard system for fast and accurate identification of animal species. Its scope is now that of all eukaryotic species (Hebert et al., 2003; Miller, 2007). The DNA barcode itself consists of a 648 bp region 58-705 from the 5'-end of the cytochrome *c* oxidase 1 (COI) gene using the mouse mitochondrial genome as a reference. It is based on the postulate 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) and that genetic variation between species exceeds variation within species (Hebert et al., 2003; Hebert, 2004a). The two main ambitions of DNA barcoding are to (i) assign unknown specimens to species and (ii) enhance the discovery of new species and facilitate identification, particularly in cryptic, microscopic and other organisms with complex or inaccessible morphology (Hebert et al., 2003).

2-2 When is the DNA barcode useful?

Access to a public reference database of taxa allowing identification of a wide range of species will be beneficial whenever accurate taxonomic identifications are required. The DNA barcode can in this way be of great support to numerous scientific domains (e.g. ecology, biomedicine, epidemiology, evolutionary biology, biogeography and conservation biology) and in bio-industry. The cost and time-effectiveness of DNA barcoding enables automated species identification, which is particularly useful in large sampling campaigns (e.g. Craig Venter's Global Ocean sampling team, Rusch et al., 2007). In this way, DNA barcoding could also improve large surveys aiming at unknown species detection and identification of pathogenic species with medical, ecological and

agronomical significance (Armstrong and Ball, 2005; Ball and Armstrong, 2006). Besides, it is important to be able to recognize, detect and trace dispersal of patented organisms in agro-biotechnology, either to certify the source organism (e.g. truffles, Rastogi et al., 2007) or secure intellectual property rights for bioresources (Gressel and Ehrlich, 2002; Kress and Erickson, 2007; Taberlet et al., 2007).

One obvious advantage of DNA barcoding comes from the rapid acquisition of molecular data. As a contrast, morphological data gathering can be time consuming, in some cases totally confusing and in others, almost impossible (e.g. Dinoflagellate taxonomy, Litaker et al., 2007; *diatomea*, Evans et al., 2007; earthworms, Huang et al., 2007). Furthermore, in three important situations, relevant species identification must necessarily be molecular-based. First, in determining the taxonomic identity of damaged organisms or fragments of (e.g. goods, food and stomach extracts). The DNA barcoding tool is thus potentially useful in the food industry, diet analyses, forensic sciences and in preventing illegal trade and poaching of endangered species (e.g. fisheries, trees, bushmeat). Second, molecular-based identification is necessary when there are no obvious means to match adults with immature specimens (e.g. fish larvae, Pegg et al., 2006; amphibians, Randrianiaina et al., 2007; coleoptera, Caterino and Tishechkin, 2006; Ahrens et al., 2007; fungal sexual stage, Shenoy et al., 2007). The third case is when morphological traits do not clearly discriminate species (e.g. red algal species, Saunders, 2005; fungal species, Jaklitsch et al., 2006; and field-collected mosquito specimens, Kumar et al., 2007), especially when size precludes visual identification (i.e. 'unseeable animals', Blaxter et al., 2005; Webb et al., 2006) or if species have polymorphic life cycles and/or exhibit pronounced phenotypic plasticity (e.g. Lamilariales, Lane et al., 2007).

2-3 DNA Barcoding as a driving force in biological sciences

More than being a species identification tool for non specialists, DNA barcoding is also of interest to specialists. To achieve the CBOL objectives, species have to be taxonomically described before their deposit in BOLD, which leads researchers to resolve analytical, technical and fundamental issues beforehand. It also brings together (and complements) taxonomy, molecular phylogenetics and population genetics (Hajibabaei et al., 2007b). According to Rubinoff and Holland (2005), DNA barcoding can be regarded as a 'tremendous tool' to accelerate species discovery and initiate new species descriptions (DeSalle et al., 2005; DeSalle, 2006). Moreover, it re-opens the debate on species concepts (Fitzhugh, 2006; Rubinoff et al., 2006b; Balakrishnan, 2007; Miller, 2007; Vogler and Monaghan, 2007). Unlike other well-known sequence libraries (e.g.

NCBI), BOLD is an interactive interface where deposited sequences can be revised and taxonomically reassigned. The compiling of sequences, from one or few common loci improves synergic studies at large geographic scales and across numerous genera (Hajibabaei et al., 2007b). Such information on the global distribution of species, their genetic diversity and structure will enhance the speed and effectiveness of local population studies.

3- Advances in Barcoding

3-1 State of the art

By March 2008, the total available DNA Barcode records were at 363,584 sequences (50,039 species), of which 136,338 sequences (13,761 species) satisfied DNA barcoding criteria (i.e. minimum sequence length of 500 bp and more than 3 individuals per species). At this date, more than 65% of all barcoded specimens had been collected in the last five years. The majority of the specimens (over 98%) are from the animal kingdom with more than 65% representing Insecta. The International Barcode of Life project (iBOL) is now under development by the new Canadian International Consortium Initiative (ICI). Researchers from 25 countries will be involved in this large-scale and collaborative program, which aims at building a comprehensive DNA barcode registry for eukaryotic life. The program's starting date is tentatively set at January 2009 and within the first 5-year period there are plans to acquire DNA barcode records for 5 million specimens representing 500,000 species (out of more than an estimated 10 million species to be discovered).

So far, the COI gene has proved to be suitable for the identification of a large range of animal taxa, including gastropods (Remigio and Hebert, 2003), springtails (Hogg and Hebert, 2004), butterflies (Hebert et al., 2004a; Hajibabaei et al., 2006a), birds (Hebert et al., 2004b; Kerr et al., 2007), mayflies (Ball et al., 2005), spiders (Greenstone et al., 2005), fish (Ward et al., 2005), ants (Smith et al., 2005), Crustacea (Costa et al., 2007) and recently, *diatomea* and Protista (Evans et al., 2007). Hajibabaei et al. (2006a) showed that 97.9% of 521 described species of Lepidoptera possess distinct DNA barcodes and furthermore that the few instances of sequence overlap of different species involve very similar ones.

The efficiency of DNA barcoding has been reported in the detection and description of new cryptic species (Hansfield and Hansfield, 2006; Smith et al., 2006b; Anker et al., 2007; Bucklin et al., 2007; Gomez et al., 2007; Pfenninger et al., 2007; Tavares and Baker, 2008) and of sibling species (Hogg and Hebert, 2004; Amaral et al., 2007; Van Velzen et al., 2007). This identification tool can clearly give support to improve classifications and to critically examine the precision of

morphological traits commonly used in taxonomy. Indeed, several studies have already illustrated the advances provided by the iterative processes between morphological- and DNA barcode- based studies in taxonomy (Hebert et al., 2004a; Hebert and Gregory, 2005b; Page et al., 2005; Carlini et al., 2006; Smith et al., 2006a; Smith et al., 2007; Van Velzen et al., 2007).

3-2 New insights into ecology and species biology

New insights into ecology and species biology have already emerged from the DNA barcoding project. For example, the identification of organisms contained in stomach extracts allows the elucidation of wild animal diets, especially when behavioural studies are not feasible (e.g. Krill diets, Passmore et al., 2006; affirmation of polyphagy of the moth *Homona mermerodes*, Hulcr et al., 2007; *Xenoturbella bocki* diet, Bourlat et al., 2008). DNA barcoding could also become an efficient tool to clarify host-parasite and symbiotic relationships (Besansky et al., 2003) and in turn give new insights on host spectra, as well as on the geographical distributions of species (host, parasites and/or endangered species). Moreover, the tool is suitable to elucidate the symbiont and parasite transmission pathways from one host generation to the next as illustrated in the interaction of beetles (*Lecythidaceae*) with their endosymbiotic yeasts (*Candida spp.* clades and other undescribed yeast species) (Berkov et al., 2007). Molecular dating of symbiotic relationships can also be deduced using barcoding tools (Anker et al., 2007).

3-3 Technical advances in barcoding

The purpose of the DNA barcoding project is to rapidly assemble a precise and representative reference library. Thus it is based on conventional and inexpensive protocols for DNA extraction, amplification and sequencing. With time, the reference library will become increasingly useful, enabling the rapid identification of low taxonomic level taxa with specific short-DNA sequences (i.e. mini-barcode 100bp, Hajibabaei et al., 2006c; 300 bp, Min and Hickey, 2007.). It has been shown that species identity can be validated or inferred from a small number of polymorphic positions within the COI-barcode (‘microcoding’ of 25bp, Summerbell et al., 2005; DNA arrays-based identification, Hajibabaei et al., 2007; SNP-based discrimination, Xiao et al., 2007). Other new molecular technologies used in bioengineering (e.g. silicon-based microarrays, nylon membrane-based macroarrays, etc.) are becoming cheaper and may be integrated into the ‘second step of DNA barcoding’ (Summerbell et al., 2005). Furthermore, new sequencing techniques such as

pyrosequencing (454, Solexa, SOLID) enable rapid and representative analyses of mixed samples (e.g. stomach contents, food, blood or water columns. Largely used in the emerging field of metagenomics, this advance could be promising for future DNA barcoding initiatives.

DNA barcoding could also be used as a technical enhancer. Indeed, one condition for data submission to BOLD is the conservation of entire morphological reference for species (voucher). Indeed, new techniques of non-destructive DNA extraction from recently collected specimens have already been developed (Pook and McEwing, 2005; Hunter et al., 2007; Rowley et al., 2007) and additional improvements in specimen conservation may arise. One major drawback of molecular-based studies as for example DNA barcoding is our incapacity to extract DNA from specimens conserved in formalin. Indeed, museum collections of animals represent the major part of voucher specimens from which species have been described and most of these are conserved in formalin. The ultimate challenge is to find the appropriate ways to extract DNA from formalin-conserved specimens and harvest DNA barcodes from them.

4- What can be learnt from the limitations of DNA barcoding?

Despite the promises of the global barcoding initiative, some crucial pitfalls must be mentioned. We believe that these limitations should be clearly identified and resolved in the library construction phase, otherwise the BOLD database will not ever become universally relevant.

4-1 The under-described part of biodiversity

The sampling shortage across taxa can sometimes lead to ‘barcoding gaps’ (Meyer and Paulay, 2005), which highlights the care that must be accorded to sampling quality during the database construction phase (Wiemer and Fiedler, 2007). The individuals chosen to represent each taxon in the reference database should cover the major part of the existing diversity. Indeed, in the interrogation of BOLD, identification difficulties arise when the unknown specimens come from a currently under-described part of biodiversity (Rubinoff, 2006; Rubinoff et al., 2006). Meyer and Paulay (2005) estimated the error rates for specimen assignment in well-characterized phylogenies and in partially known groups. They showed that the DNA barcode exclusively promises robust specimen assignment in clades for which the taxonomy is well understood and the representative

241 specimens are thoroughly sampled. Their conclusions are totally concordant with the example of the
242 *Muntjac* described in DeSalle et al. (2005).
243
244 *4-2 Inherent risks due to mitochondrial inheritance*
245
246 The diversity of mitochondrial DNA (mtDNA) is strongly linked to the female genetic
247 structure due to maternal inheritance. The use of mitochondrial loci can thus lead to overestimate
248 sample divergence and render conclusions on species status unclear. For instance, in *Homona*
249 *mermerodes* (Lepidoptera) mtDNA polymorphism is structured according to the host plants on
250 which females feed, and the two clades produced by phylogenetic analyses are artefacts of female
251 nutritional choice (Hulcr et al., 2007). Heteroplasmy and dual uniparental mitochondrial inheritance
252 (e.g. Mussels, Terranova et al., 2007) are further misleading processes for mitochondrion-based
253 phylogenetic studies.
254 The mitochondrial inheritance within species can also be confounded by symbiont infection.
255 Firstly, indirect selection on mitochondrial DNA arises from linkage disequilibria with
256 endosymbionts, either obligate beneficial micro-organisms, parasitically- or maternally- inherited
257 symbionts (Funk et al., 2000; Whitworth et al., 2007). Such symbionts are very common in
258 arthropods (e.g. *Wolbachia* infects at least 20% of Insecta and 50% of spiders, Hurst and Jiggins,
259 2005; *Cardinium* infects around 7% of arthopods, Weeks et al., 2007) and are probably widespread
260 in many other Metazoa. Secondly, interspecific hybridization and endosymbiont infections can
261 generate transfer of mitochondrial genes outside an individual's evolutionary group (Dasmahapatra
262 and Mallet, 2006). Examples are the cross-generic mitochondrial DNA introgression observed
263 between *Acreae* (Lepidoptera) and *Drosophila* (Diptera) coming from the vertically transmitted
264 *Wolbachia* (Hurst and Jiggins, 2005), or the cross-kingdom horizontal mtgene transfer detected
265 between sponges and their putative fungal symbionts (Rot et al., 2006). Finally, one host species can
266 bear different symbionts (e.g. european populations of *Adalia* bearing three symbionts, *Spiroplasma*,
267 *Rickettsia* and *Wolbachia,* Hurst et al., 1999), leading to intraspecific (i.e. inter-population) variation
268 in mtDNA sequences.
269 In all these cases, nuclear loci are required to resolve phylogenetic relationships and may
270 serve as a validating tool during the database construction stage. Besides, special care must be
271 accorded to the compilation of reference sequences (i.e. DNA barcode), especially for species with
272 already known disturbed mitochondrial inheritance. The presence of such potentially misleading
273 effects should be explicitly indicated in the BOLD. However, unknown endosymbionts or exclusive

274 causes of mtDNA inheritance disturbance could also be revealed during the DNA barcoding
275 database filing.
276
277
278 *4-3 Nuclear copies of COI (NUMTs)*
279
280 Nuclear mitochondrial DNAs (NUMTs) are nuclear copies of mitochondrial DNA sequences
281 that have been translocated into the nuclear genome (Williams and Knowlton, 2001). In eukaryotes,
282 the number and the size of NUMTs are variable, ranging from none or few in *Anopheles*,
283 *Caenorhabditis* and *Plasmodium*, to more than 500 in humans, rice and *Arabidopsis* (Richly and
284 Leister, 2004). As reported by Ann Bucklin (Oral comm., the 3rd international Cons. Gen.
285 symposium, New York, 2007) using DNA barcoding in investigations on marine zooplankton, and
286 by Lorenz et al. (2005) performing primate DNA barcoding, nuclear COI copies can sometimes
287 greatly complicate the straightforward collection of mitochondrial COI sequences. Disturbance due
288 to NUMTs must be seriously considered, in both DNA barcode library construction and further
289 specimen identification. Owing to their particular codon structure, non-synonymous mutations,
290 premature stop codons and insertion-deletions (Strugnell and Lindgren, 2007), NUMTs can be
291 recognized in the sequence and in the amino acid alignments. In the sequence acquisition stage,
292 NUMTs can be detected by the sequence checking process proposed in BOLD (i.e. rejection of
293 inconsistent amino acid alignment), and in such cases, their occurrence should be referenced in
294 BOLD. Only recently integrated NUMTs that are difficult to detect (Thalmann et al., 2004), could be
295 ignored. Although it is more difficult, it is nevertheless possible to get the true mtCOI sequence of
296 voucher specimens with the reverse transcription (Collura et al., 1996). In the diagnostic stage, there
297 may be cases where NUMT occurrence is unknown, which highlights the care that should be taken
298 in DNA barcode alignments.
299
300
301 *4-4 Rate of evolution in COI*
302
303 The rate of genome evolution (mitochondrial or nuclear) is not equal for all living species.
304 Notably, molluscs have a higher evolutionary rate than other bilateral metazoans (Strugnell and
305 Lindgren, 2007). In contrast, diploblast sponges and cnidarians have an evolutionary rate 10-20
306 times slower than in their bilaterian counterparts, a consequence of which is the lack of COI-
307 sequence variation that prevents distinction below the family level (Erpenbeck et al., 2006). The rate

of evolution can even differ at the ordinal level, as shown between six dermapteran (Insect) species (Wirth et al., 1999). In the same way, the level of variation in mitochondrial sequences in the plant kingdom excludes species identification based on COI sequence polymorphism (Kress et al., 2005).

More generally, the lack of resolving power of COI-sequence reported for some taxa has led the CBOL to envisage the transition from the primary single-gene method (i.e. BARCODE) to a multi-region barcoding system, when it is justified (i.e. in cases where COI is not species specific, or for taxa with low mitochondrial evolutionary rates) taxon-specific reference regions (i.e. nuclear plus/or organelle genes), also called non-COI barcode (Bakker, Second International Barcode of Life Conference, TAIPEI, September 2007).

4-5 The intra-specific geographical structure should be taken into account

Geographical structure, if ignored, can blur and distort species delineation. Actually, high rates of intraspecific divergence can derive from geographically isolated populations (Hebert et al., 2003), and thus, must be considered in the setting up of the DNA barcode reference database. This point stresses a key challenge for the DNA barcoding initiative, from both the fundamental and analytical points of view. What is the boundary between a population and a species? Does it exist? To solve this issue, wide-ranging intra-specific sampling should be integrated in the reference database, and one must consider species boundaries not as a definitive but as a revisable concept. The relevance of the reference DNA barcode database depends on the exhaustiveness of intra-taxon sampling.

To prevent misleading results, the current data format for submission to BOLD should be complemented with new fields related to the limitations mentioned above (i.e. NUMT occurrence, known endosymbiont, available insight on molecular clock, genetic structure and geographical distribution). Besides the biological limitations, DNA barcoding raises analytical and statistical issues.

5- DNA-Sequence Analysis, a double trend: pure assignment vs delimitation of species

5-1 Query sequence assignment

342 The main and unambiguous objective of DNA barcoding analysis is to assign one query
343 sequence to a set of referenced tagged-specimen sequences extracted from BOLD. The method
344 currently used in BOLD combines similarity methods with distance tree reconstruction in the
345 following way: (i) First, the query sequence is aligned to the global alignment through a Hidden
346 Markov Model (HMM) profile of the COI protein (Eddy, 1998), followed by a linear search of the
347 reference library. The 100 best hits are selected as a pre-set of "closely related tagged-specimens";
348 (ii) Second, a Neighbor-Joining tree is reconstructed on this preset plus the query sequence to assess
349 the relationship between the query sequence and its neighboring referenced sequences (Kelly et al.,
350 2007). The query sequence is then assigned to the species name of its nearest-neighboring referenced
351 sequence, whatever the distance between the two sequences. This method is direct and rapid, but its
352 main shortcomings are high prevalence of sampling-dependent accuracy, high rates of false-positive
353 assignments (Koski and Golding, 2001) and the fact that there is no other way to infer the reliability
354 of the query assignment than computing percentages of similarity or genetic distances, two measures
355 that are known to be irrelevant for taxonomic relationship (Ferguson, 2002). The loss of character
356 information is also inherent in distance methods, as computing distances erase all character-based
357 information (DeSalle, 2006). Moreover, as both similarity and distance methods strongly depend on
358 the disparity between intra- and inter-specific variations, incomplete taxonomic sampling (i.e.
359 barcoding gaps) will artificially increase the accuracy of such methods.
360 Various alternative methods have been proposed to analyse DNA barcode data amongst which
361 we can distinguish four main categories of approaches: (i) similarity approaches, based solely on the
362 similarity between the total DNA barcode sequences or small parts of them (e.g. oligonucleotide
363 motifs, Dasgupta et al., 2005; Little et al., 2007); (ii) classical phylogenetic approaches, using either
364 genetic distances or maximum likelihood / Bayesian algorithms and assuming different mutational
365 models (e.g. Neighbor-Joining, phyML, MrBayes, Elias et al., 2007); (iii) multiple-character based
366 analysis (DeSalle et al., 2005) (vi) pure statistical approaches based on classification algorithms
367 without any biological models or assumptions (CAOS, Sarkar et al., 2002 a, b); and (v) genealogical
368 methods based on the coalescent theory using demo-genetic models and maximum likelihood /
369 Bayesian algorithms (Matz and Nielsen, 2005; Nielsen and Matz, 2006; Abdo and Golding, 2007).
370 The question here is whether it is worthwhile to adopt a biological, populational and/or phylogenetic
371 rationale for DNA barcode sequence analyses or, whether pure statistical approaches are more
372 efficient to assign a query sequence to a species name. Note that character-based methods (either
373 character-based phylogenetics, i.e. not distance-based, or statistical classification) are consistent with
374 the phylogenetic species concept (Goldstein and DeSalle, 2000), whereas distance-based methods
375 are not (Lipscomb et al., 2003). CAOS of Sarkar et al. (2002 a, b) is an example of character-based

376 analysis, in which the nucleotide sequence is considered as a chain of characters. In the same way,
377 DeSalle et al. (2005) proposed the combination morphological and molecular characters, which has
378 the advantage of bridging the gap between the classical taxonomy and 'molecular-taxonomy' and the
379 DNA barcoding approach.
380 At present, global comparisons between all these approaches are clearly missing. However,
381 few studies have already compared some of these algorithms (Elias et al., 2007, Ross et al., 2008).
382 For example, Austerlitz et al., (Second International Barcode of Life Conference TAIPEI, September
383 2007) compared phylogenetic tree reconstruction with various supervised classification methods
384 (CART and Random Forest, Support Vector Machines and Kernel methods, Breiman et al., 1984) on
385 both simulated and real data sets. Their main conclusions are: (i) maximum likelihood phylogenetic
386 (PhyML, Guindon and Gascuel, 2003) approaches always seem to be more accurate than distance-
387 based (Neighbor-Joining) phylogenetic inferences; (ii) computation times are much higher for
388 maximum likelihood phylogenetic reconstruction than for statistical classification; and (iii) the
389 accuracy of all the methods strongly depends on sample size and global variability of the taxa.
390 Supervised classification methods outperform phylogenetic analyses only when the reference sample
391 per species is large ($n \geq 10$).
392
393 Rigorous assignment relevancy depends on our capacity to estimate the probability of a false-
394 assignment event. False species assignments can be due to three types of errors (Nielsen and Matz,
395 2006): (i) the true species may not be represented in the database; (ii) the random coalescence of
396 lineages in populations and species may not necessarily lead the query sequence to be the most
397 closely related to the true species sequence; (iii) the random process at which mutations arise on
398 lineages may cause the sequence representing another species to be more similar to the query.
399 Population genetics theory, and more specifically coalescent theory, can help to assess the
400 probability of the occurrence of the last two events. Recently, model-based decision theory
401 framework based on the coalescence theory (Matz and Nielsen, 2005; Nielsen and Matz 2006; Abdo
402 and Golding, 2007) has been established, and should lead to greater accuracy in query sequence
403 assignment with an estimation of the degree of confidence with which this assignment can be made.
404 However, the major drawbacks of such model-based decision tools are high computation times and
405 the requirement of large data sets (e.g. more than 10 sequences per species) for enough genetic
406 information to perform accurate analyses. Moreover, the mitochondrial neutrality has recently being
407 put into question (Bazin et al., 2006), which may invalidate inferences using neutral coalescent
408 processes.

To conclude on the query assignment method, it would be advisable to adopt a sequential investigation. Firstly, to search the complete database with similarity methods thus reducing the total data set to the genus or family of the query sequence. Then, to use statistical classification and/or phylogenetic tools to more precisely assign the query sequence to a given species. If still no obvious assignment emerges, it should then be made using population genetic methods based on coalescence. However, even if the assignment with classification or phylogenetic methods seems unambiguous, coalescent-based methods running on the closest neighbours of the query sequence should give an idea of the degree of uncertainty associated with an identification.

5-2 Delimitation of species

The second and more controversial objective of DNA barcode analyses is to define clusters of individuals and consider them as species, in other words, to do molecular taxonomy on unidentified taxa. Unlike the approaches mentioned above, clustering is an unsupervised learning problem that involves identifying homogeneous groups in a data set. Beside all the well-justified discussions between taxonomists about the molecular delimitation of species, such a clustering approach is much more complicated than pure assignment to a pre-identified taxonomic group. Three main approaches have been put forward so far.

Hebert et al. (2004b) first proposed the use of a divergence-threshold to delimit species. The underlying idea was that intra-species divergence is lower than inter-species divergence. The standard divergence threshold value advised was of ten times the mean intraspecific variation ('10-fold rule') with the reciprocal monophyly. Despite the efficiency of the threshold approach reported for fishes (Ward et al., 2005), crustaceans (Lefebure et al., 2006), North American birds (Hebert et al., 2004b), tropical lepidopterans (Hajibabaei et al., 2006a) and cave-dwelling spiders (Paquin and Hedin, 2004), the use of thresholds in species delineation has been strongly discouraged. Indeed, the divergence-threshold methods lack strong biological support and undoubtedly could not become a universal criterion suited to animal species delineation (Meyer and Paulay, 2005; Hickerson et al., 2006; Wiemer and Fiedler, 2007). By their literature survey of mitochondrial DNA studies on low taxonomic-level animal phylogeny, Funk and Omland (2003) detected species-level paraphyly or polyphyly in 23% of 2,319 assayed species, demonstrating that NJ-tree analysis will fail to assign query sequences in a significant proportion of cases (Ross et al., 2008). The question is thus to clearly characterize the proportion of non-monophyletic species and the relationship between intra-

and inter-specific variability in various taxa to globally assess the relevance of such threshold approaches.

The second approach to delimitate species has been developed by Pons et al. (2006) using a mixed model combining a coalescent population model with a Yule model of speciation. Their approach is based on the differences in branching rates at the level of species and populations. Such a model allows them to infer a time of branching regime change, from the coalescent rate to the speciation rate, and to define species as being the clusters for which all individuals are branched inside the coalescent time frame. Even if their approach is currently oversimplified because they consider a unique rate of coalescence (i.e. all population sizes are the same) and a unique shift from population to species processes, it is a promising step that combines the principles of population genetics and those of speciation processes.

The third methodology, which also uses principles of population genetics, is the extension of the coalescent-based models of Matz and Nielsen (2005), Nielsen and Matz (2006) and, Abdo and Golding (2007). This approach has not yet been fully developed but is suggested in the three above-mentioned papers. The underlying idea is that maximum likelihood or Bayesian inference, using coalescent models, should help to assess divergence times and/or presence or absence of gene flow between the clusters considered. Estimates of divergence times and gene flow can then be used to infer species status of clusters, based on the biological definition of species. As for the assignment methods, the main drawback of such coalescent-based methods are computation times and large intra-specific sampling requirements.

We emphasize that, whatever approach is used, every taxonomic decision using DNA Barcode data should be validated by other independent lines of evidence.

6- What level of universality can the DNA barcode reach?

6-1 The choice of the genome region(s)

The main difficulty of DNA barcoding is to find the ideal gene that discriminates any species in the animal kingdom. Hebert et al. (2003, 2004a,b) argue in favor of the mitochondrial 5' COI region (Folmer et al., 1994), a choice justified by its great resolving-power for birds, lepidopteran and dipteran species discrimination. Ideally, a single pair of universal primers (e.g. Folmer primers, Folmer et al., 1994) would amplify the DNA barcode locus in any animal species. The development of taxon-specific primers and their combinations are however sometimes necessary to obtain greater

intra-generic accuracy (e.g. coral reef, Neigel et al., 2007), as illustrated by the primer combinations and cocktails required to obtain DNA barcodes from fish species (Ward et al., 2005; Ivanova et al., 2007), or the primer sets needed to distinguish between primate genera (Lorenz, 2005). The COI amplification does not always ensure the success of the specimen identification. Indeed, the COI-based identification sometimes fails to distinguish closely related animal species, underlining the requirement of nuclear regions (e.g. C*yt*b and R*ho*d to identify all teleost fish species, Sevilla et al., 2007): the idea of a multi-locus DNA barcoding approach is progressively emerging.

The extension of DNA barcoding to other kingdoms is also progressing. The efficiency of COI-based barcoding has been documented for few groups of fungi (e.g. *Penicillium sp.*, Seifert et al., 2007), macroalgae (Rhodophyta, Saunders, 2005) and two ciliophoran protists genera (*Paramecium* and *Tetrahymena*s, Barth et al., 2006; Lynn and Strüder-Kypke, 2006; Chantangsi et al., 2007), suggesting that the DNA barcode standardization may be harder to reach than expected. It is now commonly accepted that the universality of the initial COI-based CBOL project is unlikely. Indeed, considering mitochondria solely would not solve problems of differential evolutionary rates among close genera, of inheritance discrepancy, of mtDNA introgression processes and of the intron-size variations that prevent COI-sequence alignment (e.g. fungi, plants). Besides, methods of sequence assignment based on a single-locus will often lack accuracy (Elias et al., 2007). The ineluctable future trend for species identification through DNA barcoding is to develop a multi-locus system, including COI-region or/and independent markers (Rubinoff and Holland, 2005; Dasmahapatra and Mallet 2006; Kress and Erickson, 2007; Smith et al., 2007; Sevilla et al., 2007). Additional molecular markers have already been proposed, among which the nuclear subunit ribosomal RNA genes are promising candidates because of their great abundance in the genome and their relatively conserved flanking regions. Moreover, the use of rRNA allows efficient species distinction (e.g. for amphibians, Vences et al., 2004, 2005; for truffles, El Karkouri et al., 2007), and can sometimes provide classifications into molecular taxonomic units, MOTU (e.g. for nematodes, Floyd et al., 2002; Blaxter et al., 2005).

In higher plants, the mitochondrial genome evolves much more slowly than in animals. The COI-region is thus inappropriate for plant species distinction (Rubinoff et al., 2006). The CBOL plant working group (PWG) agrees that plant barcoding will be multi-locus, with one ‘anchor’ (i.e. universal across the plant kingdom) and ‘identifiers’ to distinguish closely related species (Bakker, Second International Barcode of Life Conference TAIPEI, September 2007). Several combinations of DNA regions have been recently proposed (Kress et al., 2005; Chase et al., 2006; Kress and Erickson, 2007; Pennisi, 2007; Lee et al., Second International Barcode of Life Conference TAIPEI, September 2007). At present, there is still no consensus on which candidate markers are the best

plant DNA barcoding region (Pennisi, 2007). The future combination will certainly contain non-coding intergenic spacers (e.g. *trn*H-p*sb*A, Kress et al., 2005; Chase et al., 2006; Kress and Erickson, 2007) and plastidial coding sequences (e.g. *mat*K, Chase et al., 2007). Recently, Lahaye et al. (2008) working on a large representative sample (>1600 plants specimens) strongly converged with Chase et al.'s (2007) conclusion, and advocates the *mat*K locus as the best universal 'anchor' for DNA barcoding of plant taxa. However, they also agree with the need for extra loci (i.e. 'identifiers') to resolve lower taxon identification. In addition, Taberlet et al. (2007) focused on the feasibility of barcoding plants from highly degraded DNA that is of interest for ancient DNA studies (e.g. permafrost samples) and other applied fields (e.g. processed food, customs, medicinal plants). They promoted the chloroplast *trn*L (UAA) intron or a shorter fragment of this intron (the P6 loop, 10-143 bp), which, despite the relatively low resolution, can be amplified with highly conserved primers.

If the prior universality of a single locus and a single primer set remains utopian, the use of a few common loci is still a great advance for future diversity assessments within higher taxa. Steady common features of the DNA barcoding approach will remain, but will certainly evolve in kingdom- or even lower taxon- specific technical approaches.

6-2 The challenge: barcoding microscopic biodiversity

One of the greatest challenges for the Barcode of Life project is to account for the diversity of unicellular life (i.e. archea, bacteria, protists, and unicellular fungi). As a matter of fact, with an evolutionary history dating back to 3.5 billion years, microscopic life (<1mm) represents the largest part of biodiversity. Besides, microscopic species are the causal agents of numerous diseases and are keys to the functioning of trophic networks (Chantangsi et al., 2007). In oceans, microbial life is responsible for 98% of the primary production and the mediation all the biogeochemical cycles (Sogin et al., 2006). One of the striking characteristics of the microbiosphere is the unstable population size over short periods of time, one population can be dominant at a specific time or location, but rare, and thus difficult to survey at another time or location. The low-abundance populations (e.g. rare biosphere) that account for most of the phylogenetic diversity are masked by the dominant populations, leading to an underestimation of the diversity of microbial life (Sogin et al., 2006).

The microbiodiversity DNA barcoding has yet been poorly studied. The emerging community genomics (DeLong et al., 2006) and the metagenomics approaches promise great insights on prokaryote biodiversity and molecular evolution (DeLong, 2004; Tyson et al., 2004; Venter et al.,

2004; Tringe et al., 2005; Leclerc et al., 2007). But, for microscopic eukaryotes, the routine use of the metagenomics approach is not technically conceivable and too expensive today, highlighting the interest of DNA barcoding on pooled samples (i.e. where multiple species are present or communities) to assess eukaryotic microbiodiversity (Johnson and Slatkin, 2008).

Exploring the microscopic eukaryotic life diversity can be achieved by the COI-based barcode (Blaxter et al., 2005, Chantangsi et al., 2007; Evans et al., 2007). But as for the macroscopic species, a specific and multi-locus barcoding approach may be required. The 5'-end region (600 bp) of the small ribosomal subunit has already been reported to assign isolates to specific subtypes of the human parasite *Blastocyctis hominis* (Scicluna et al., 2006) and successfully reached types or species in the deep seas (Sogin et al., 2006).

6-3 Universality of the Barcode Of Life Data system (BOLD)

The universality of the Barcode of Life mainly resides in the synergic and standard approach for data acquisition and their compilation into BOLD, which is the central connection of the CBOL initiative. The current format for data submission to BOLD is composed of five fields for voucher specimen characterization: (i) the specimen identifier (the catalogue and collection codes, the institution responsible for providing the specimen samples); (ii) the taxonomic status; (iii) the specimen characteristics (sex, life stage, reproduction); (iv) the collection data (collector, collection date and location with GPS coordinates); (v) the DNA barcode sequence (gene name and location, trace file, alignment details, primers used to generate the amplicon). Next to this, pictures of vouchers and the trace of their DNA Barcode must be submitted. All the guidelines are specified on the BOLD website (http://www.barcodinglife.com/docs/boldtutorial.html).

Recently the international sequences depository (NCBI) have accepted to put in place a BARCODE keyword into their search facilities for entries that conform to the minimal CBOL requirements (i.e. traces file, collection location). The BOLD barcode Submission tool associates the 'Barcode submissions' with a further submission to NCBI using the 'My NCBI' user name. Note that for data that satisfies the BARCODE-keyword conditions, the NCBI taxonomy browser gives the direct link to the BOLD taxonomy browser. A keyword related to the concept of 'non-COI barcode' may be soon envisaged.

BOLD will provide an increasing amount of DNA barcode records (either COI- or non-COI barcodes) to clearly identify unknown specimens, which will enable accurate query assignments and will facilitate comparison between data obtained in geographically dispersed institutions. Recently, an increasing amount of initiatives for global data recording have been proposed to manage clinical and molecular information about infectious diseases (epiPATH, Amadoz and Gonzales-candelas,

2007), but also to focus on either human pathogenic bacteria (e.g. pathoMIPer, Thiyagarajan et al., 2006; pyloriBASE, Ahmed et al. 2007; VectorBase, Lawson et al., 2007; DengueInfo, Schreiber et al., 2007) or on plant pathogens (e.g. PhiBASE, Winnenburg et al., 2006). BOLD could serve as the universal starting point for species identification, which would convey users to refer to specialized databases (e.g. pathogenic strains, disease vector species and endangered species). The CBOL has already initiated the new International Network for the Barcoding of Invasive and Pest Species (INBIPS; www.barcoding.si.edu/INBIPS.htm) that will help to coordinate the collection of barcode data on pest species around the world (Ball and Armstrong, 2006).

7-Conclusion

After four years, the animal DNA barcoding approach has become less controversial, its relevancy is now supported by numerous successes and by the increasing amount of DNA barcoding projects, among which the barcoding of 500,000 species planned by the iBOL. However, clear limitations arise from the incomplete coverage of the existing diversity, the inherent characteristics of the mitochondrial DNA (evolutionary rate, inheritance, introns, neutrality) and the single-locus initial strategy. With its enlargement to all eukaryote taxa, the Barcode of life project has also evolved to a more flexible framework. The approach reveals to be more complex than the system projected by the CBOL initially. The multi-locus barcoding approach is now commonly accepted, particularly to discriminate between low level taxa and to increase the power of the sequence assignments.

The BOLD data system is central to the DNA barcoding approach. The specificities of BOLD are (i) to assemble standard information on voucher specimens using common description fields (DNA tag, specimen taxonomy, specimens details, collection information, voucher pictures), and, (ii) its dynamic status that allows taxonomic revisions and reassignment of the deposited sequences.

The final point concerning the Barcode of Life project is that, beyond the construction of a standard approach based on the existing taxonomic knowledge, it has enhanced communication between different scientific communities, including taxonomists, phylogeneticists and population geneticists.

Acknowledgments

We thank Kevin Bleakley, Pascale Chesselet, Olivier David, Mark Stevens, Michel Veuille and Thierry Wirth for constructive comments and usefull discussions on the manuscript. L.F. was financially supported by a ATER grant from the Ecole Pratique des Hautes Etudes.

References

Abdo, Z., Golding, G.B. 2007. A step toward barcoding life: A model-based, decision-theoretic method to assign genes to preexisting species groups. Syst. Biol. 56, 44-56.

Ahmed, N., Majeed, A.A., Ahmed, I., Hussain, M.A., Alvi, A., Devi, S.M., Rizwan, M., Ranjan, A., Sechi, L.A., Megraud, F. 2007. genoBASE *pylori*: A genotype search tool and database of the human gastric pathogen H*elicobacter pylori*. Infect. Genet. Evol. 7, 463-468.

Ahrens, D., Monaghan, M.T., Vogler, A.P. 2007. DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera *: Scarabaeidae*). Mol. Phylogenet. Evol. 44, 436-449.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402.

Amaral, A.R., Sequeira, M., Coelho, M.M. 2007. A first approach to the usefulness of cytochrome c oxidase I barcodes in the identification of closely related delphinid cetacean species. Mar. Freshwater Res. 58, 505-510.

Amadoz, A., Gonzales-candelas, F. 2007. epiPATH: an information system for the storage and management of molecular epidemiology data from infectious pathogens. BMC Infect. Dis. 7, 32. (doi:10.1186/1471-2334-7-32)

Anker, A., Hurt, C., Knowlton, N. 2007. Revision of the *Alpheus nuttingi* (Schmitt) species complex (Crustacea: Decapoda: Alpheidae), with description of a new species from the tropical eastern Pacific. Zootaxa 1577, 41-60.

Armstrong, K.F., Ball, S.L. 2005. DNA barcodes for biosecurity: invasive species identification. Phil. Trans. R. Soc. B. 360, 1813-1823.

Balakrishnan, R. 2007. Species concepts, species boundaries and species identification: a view from the tropics. Syst. Biol. 54(4):689-693.

Ball, S.L., Hebert, P.D.N., Burian, S.K., Webb, J.M. 2005. Biological identifications of mayflies (*Ephemeroptera*) using DNA barcodes. J. North Am. Benthol. Soc. 24, 508-524.

Ball, S.L., Armstrong, K.F. 2006. DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: *Lymantriidae*). Can. J. For. Res. 36: 337-350.

Barth, D., Krenek, S., Fokin, S.I., Berendonk, T.U. 2006. Intraspecific genetic variation in P*aramecium* revealed by mitochondrial cytochrome c oxidase 1 sequences. J. Eukaryot. Microbiol., 53: 20-25.

Bazin, E., Glémin, S., Galtier, N. 2006. Population size does mitochondrial genetic in animals. Science 312, 570-572.

Berkov, A., Feinstein, J., Small, J., Nkamany, M. 2007. Yeasts isolated from neotropical wood-boring beetles in SE Peru. Biotropica 39, 530-538.

652 Besansky, N.J., Severson, D.W., Ferdig, M.T. 2003. DNA barcoding of parasites and invertebrate
653 disease vectors: what you don't know can hurt you. Trends Parasitol. 19, 545-546.

654 Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., Abebe, E. 2005. Defining
655 operational taxonomic units using DNA barcode data. Phil. Trans. Roy. Soc. Lond. B 360,
656 1935-1943.

657 Blaxter, M. 2006. Will DNA barcoding advance efforts to conserve biodiversity more efficiently
658 than traditional taxonomic methods? Front. Ecol. Environ. 5, 269-270.

659 Bourlat, S.J., Nakano, H., Akerman, M., Telford, M.J., Thorndyke M.C., and Obst M. 2008. Feeding
660 ecology of Xenoturbella bocki (phylum *Xenoturbellida*) revealed by genetic barcoding Mol.
661 Ecol. Resources 8, 18-22

662 Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees.
663 Wadsworth, Belmont.

664 Bucklin, A., Wiebe, P.H., Smolenack, S.B., Copley, N.J., Beaudet, J.G., Bonner, K.G., Farber-
665 Lorda, J., Pierson, J.J. 2007. DNA barcodes for species identification of euphausiids
666 (*Euphausiacea, Crustacea*). J. Plankton Res. 29, 483-493.

667 Carlini, D.B., Kunkle, L.K., Vecchione, M. 2006. A molecular systematic evaluation of the squid
668 genus Illex (Cephalopoda: *Ommastrephidae*) in the North Atlantic Ocean and Mediterranean
669 Sea. Mol. Phylogenet. Evol. 41, 496-502.

670 Caterino, M.S., Tishechkin, A.K. 2006. DNA identification and morphological description of the
671 first confirmed larvae of *Hetaeriinae* (Coleoptera : *Histeridae*). Syst. Entomol. 31, 405-418.

672 Chantangsi, C., Lynn, D.H., Brandl, M.T., Cole, J.C., Hetrick, N., Ikonomi, P. 2007. Barcoding
673 ciliates: a comprehensive study of 75 isolates of the genus *Tetrahymena*. Int. J. Syst. Evol.
674 Microbiol. 57, 2412-2425.

675 Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haidar, N.,
676 Savolainen, V. 2005. Land plants and DNA barcodes: short-term and long-term goals. Phil.
677 Trans. R. Soc. B 360, 1889-1895

678 Chase, M.W., Cowan, R.S., Hollingsworth, P.M., van den Berg, C., Madrinan, S., Petersen, G.,
679 Seberg, O., Jorgsensen, T., Cameron, K.M., Carine, M., Pedersen, N., Hedderson, T.A.J.,
680 Conrad, F., Salazar, G.A., Richardson, J.E., Hollingsworth, M.L., Barraclough, T.G., Kelly, L.,
681 Wilkinson, M. 2007. A proposal for a standardised protocol to barcode all land plants. Taxon
682 56, 295-299.

683 Collura, R.V., Auerbach, M.R., Stewart, C-B. 1996. A quick, direct method that can differentiate
684 expressed mitochondrial genes from their nuclear pseudogenes. Curr. Biol. 6 (10), 1337-1339.

685 Costa, F.O., deWaard, J.R., Boutillier, J., Ratnasingham, S., Dooh, R.T., Hajibabaei, M., Hebert,
686 P.D.N. 2007. Biological identifications through DNA barcodes: the case of the Crustacea. Can.
687 J. Fish. Aquat. Sci. 64, 272-295.

688 Dasmahapatra, K.K., Mallet, J. 2006. DNA barcodes: recent successes and future prospects.
689 Heredity 97, 254-255.

690 DasGupa, B., Konwar, K.M., Mandoiu, I.I., Shvartsman, A.A. 2005. DNA-BAR: distinguisher
691 selection for DNA barcoding. Bioinformatics 21, 3424-3426.

692 Dayrat, B. 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85, 407-415.

693 DeLong, E.F. 2004. Microbial population genomics and ecology: the road ahead Environmental
694 Microbiology 6 (9), 875-878. (doi:10.1111/j.1462-2920.2004.00668.x)

695 DeLong, E.F., Preston, C.M., Mincer, T. , Rich, V., Hallam, S.J., Frigaard, N.-U., Martinez, A.,
696 Sullivan, M.B., Edwards, R., Brito, B.R., Chisholm, S.W., Karl, D.M. 2006. Community
697 genomics among stratified microbial assemblages in the ocean's interior. Science 311, 5760:
698 496-503.

699 DeSalle, R., Egan, M.G., Siddall, M. 2005. The unholy trinity: taxonomy, species delimitation and
700 DNA barcoding. Phil. Trans. R. Soc. B 360, 1905-1916.

701 DeSalle, R. 2006. Species discovery versus species identification in DNA barcoding efforts:
702 Response to Rubinoff. Cons. Biol. 20, 1545-1547.

703 Ebach, M.C., Holdrege, C. 2005. DNA barcoding is no substitute for taxonomy. Nature 434, 697-
704 697.

705 Eddy, S.R. 1998. Profile hidden Markov models. Bioinformatics 14, 755-763.

706 Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Andrew V.Z. Brower, A.V.Z., Mallet, J.,
707 Chris D., Jiggins, C.D. 2007 Limited performance of DNA barcoding in a diverse community
708 of tropical butterflies. Proc. Roy. Soc. Lond. B 274, 2881-2889.

709 El Karkouri, K., Murat, C., Zampieri, E., 2007. Identification of internal transcribed Spacer sequence
710 motifs in truffles: a first step toward their DNA barcoding. Appl. Environ. Microbiol. 73 (16),
711 5320-5330.

712 Erpenbeck, D., Hooper, J.N.A., Worheide, G. 2006. CO1 phylogenies in diploblasts and the
713 'Barcoding of Life' - are we sequencing a suboptimal partition? Mol. Ecol. Notes 6, 550-553.

714 Evans, K.M., Wortley, A.H., Mann, D.G. 2007. An assessment of potential diatom "barcode" genes
715 (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in
716 S*ellaphora* (*Bacillariophyta)*. Protist 158, 349-364.

717 Ferguson, J.W.H. 2002. On the use of genetic divergence for identifying species. Biol. J. Linn. Soc.
718 75, C509-C516.

719 Fitzhugh, K. 2006. DNA Barcoding: An Instance of Technology-driven Science? BioScience 56(6),
720 462-463.

721 Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R. 1994. DNA primers for amplification of
722 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar.
723 Biol. Biotech. 3, 294-299.

724 Floyd, R., Abebe, E., Papert, A., Blaxter, M. 2002. Molecular barcodes for soil nematode
725 identification. Mol. Ecol. 11, 839-850.

726 Funk, D.J., Helbling, L., Wernegreen, J.J., Moran, N.A. 2000. Intraspecific phylogenetic congruence
727 among multiple symbiont genomes. Proc. R. Soc. B. 267, 2517-2521.

728 Funk, DJ., Omland, K.E. 2003. Species-level paraphyly and polyphyly: frequency, causes, and
729 consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 34,
730 397-423

731 Guindon, S., Gascuel, O. 2003 A simple, fast, and accurate algorithm to estimate large phylogenies
732 by maximum likelihood. Syst. Biol. 52, 696-704.

733 Goldstein, P.Z., DeSalle, R. 2000. Phylogenetic species, nested hierarchies, and character fixation.
734 Cladistics 16, 364-384.

735 Gomez, A., Wright, P.J., Lunt, D.H., Cancino, J.M., Carvalho, G.R., Hughes, R.N. 2007. Mating
736 trials validate the use of DNA barcoding to reveal cryptic speciation of a marine bryozoan
737 taxon. Proc. Roy. Soc. Lond. B. 274, 199-207.

738 Greenstone, M.H., Rowley, D.L., Heimbach, U., Lundgren, J.G., Pfannenstiel, R.S., Rehner, S.A.
739 2005. Barcoding generalist predators by polymerase chain reaction: carabids and spiders. Mol.
740 Ecol. 14, 3247-3266.

741 Gregory, T.R., 2005. DNA barcoding does not compete with taxonomy. Nature 434, 1067-1067.

742 Gressel, J., Ehrlich, G. 2002. Universal inheritable barcodes for identifying organisms. Trends Plant
743 Sci. 7 (12), 542-544.

744 Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W., Hebert, P.D.N. 2006a. DNA barcodes
745 distinguish species of tropical Lepidoptera. Proc. Natl. Acad. Sci. U.S.A. 103, 968-971.

746 Hajibabaei, M., Singer, G.A.C., Hickey, D.A. 2006b. Benchmarking DNA barcodes: an assessment
747 using available primate sequences. Genome 49, 851-854.

748 Hajibabaei, M., Smith, M.A., Janzen, D.H., Rodriguez, J.J., Whitfield, J.B., Hebert, P.D.N. 2006c. A
749 minimalist barcode can identify a specimen whose DNA is degraded. Mol. Ecol. Notes 6, 959-
750 964.

751 Hajibabaei, M., Singer, G.A.C., Clare, E.L., Hebert, P.D.N. 2007a. Design and applicability of DNA
752 arrays and DNA barcodes in biodiversity monitoring. BMC Biol. 5 (24). (doi:10.1186/1741-
753 7007-5-24)

754 Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N., Hickey, D.A. 2007b. DNA barcoding: how it
755 complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. 23,
756 167-172.

757 Handfield, D., Handfield, L. 2006. A new species of *Plusia* (*Lepidoptera* : *Noctuidae*) from North
758 America. Can. Entomol. 138, 853-859.

759 Hebert, P.D.N., Ratnasingham, S., deWaard, J.R. 2003. Barcoding animal life: cytochrome c oxidase
760 subunit 1 divergences among closely related species. Proc. Roy. Soc. Lond. B. 270, S96-S99.

761 Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W. 2004a. Ten species in one:
762 DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes*
763 *fulgerator*. Proc. Natl. Acad. Sci. U.S.A. 101, 14812-14817.

764 Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M. 2004b. Identification of birds through
765 DNA barcodes. Plos Biol. 2, 1657-1663.

766 Hebert, P.D.N., Barrett, R.D.H. 2005a. Reply to the comment by L. Prendini on "Identifying spiders
767 through DNA barcodes". Can. J. Zool. 83, 505-506.

768 Hebert, P.D.N., Gregory, T.R. 2005b. The promise of DNA barcoding for taxonomy. Syst. Biol. 54,
769 852-859.

770 Hickerson, M.J., Meyer, C.P., Moritz, C. 2006. DNA barcoding will often fail to discover new
771 animal species over broad parameter space. Syst. Biol. 55, 729-739.

772 Hogg, I.D., Hebert, P.D.N. 2004. Biological identification of springtails (*Hexapoda : Collembola*)
773 from the Canadian Arctic, using mitochondrial DNA barcodes. Can. J. Zool. 82, 749-754.

774 Huang, J., Qin Xu, Q., Sun, Z.J., Tang, G.L., Su, Z.Y. 2007. Identifying earthworms through DNA
775 barcodes. Pedobiologia 51, 301-309.

776 Hudson, M.E., 2007. Sequencing breakthroughs for genomic ecology and evolutionary biology. Mol.
777 Ecol. Notes (doi:10.1111/j.1471-8286.2007.02019)

778 Hulcr, J., Miller, S.E., Setliff, G.P., Darrow, K., Mueller, N.D., Hebert, P.D.N., Weiblen, G.D. 2007.
779 DNA barcoding confirms polyphagy in a generalist moth, *Homona mermerodes* (*Lepidoptera* :
780 *Tortricidae*). Mol. Ecol. Notes 7, 549-557.

781 Hunter, S.J., Goodall, T.I., Walsh, K.A., Owen, R., Day, J.C. 2007. Nondestructive DNA extraction
782 from blackflies (*Diptera*: *Simuliidae*): retaining voucher specimens for DNA barcoding
783 projects. Mol. Ecol. Notes (doi: 10.1111/j.1471-8286.2007.01879)

784 Hurst, G.D.D., Jiggins, F.M., von der Schulenburg, J.H.G., Bertrand, D., West, S.A., Goriacheva,
785 I.I., Zakharov, I.A., Werren, J.H,. Stouthamer, R., Majerus, M.E.N. 1999. Male-killing
786 *Wolbachia* in two species of insect. Proc. Roy. Soc. B. 266, 735-740.
787 (doi:10.1098/rspb.1999.0698.)

788 Hurst, G.D., Jiggins, F.M. 2005. Problems with mitochondrial DNA as a marker in population,
789 phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proc. Biol. Sci.
790 272, 1525-1534.

791 Ivanova, N.V., Zemlak, T.S., Hanner, R.H., Hebert, P.D.N. 2007. Universal primer cocktails for fish
792 DNA barcoding. Mol. Ecol. Notes 7, 544-548.

793 Jaklitsch, W.M., Komon, M., Kubicek, C.P., Druzhinina, I.S. 2006. *Hypocrea crystalligena* sp nov.,
794 a common European species with a white-spored Trichoderma anamorph. Mycologia 98, 499-
795 513.

796 Johnson, P.L.F., Slatkin, M. 2006. Inference of population genetic parameters in metagenomics: A
797 clean look at messy data. Genome Res. 16: 1320-1327

798 Kelly, R.P., Sarkar, I.N., Eernisse, D.J., DeSalle, R. 2007. DNA barcoding using chitons (genus
799 *Mopalia*). Mol. Ecol. Notes 7, 177-183.

800 Kerr, K.C.R., Stoeckle, M.Y., Dove, C.J., Weigt, L.A., Francis, C.M., Hebert, P.D.N. 2007.
801 Comprehensive DNA barcode coverage of North American birds. Mol. Ecol. Notes 7, 535-
802 543.

803 Koski, L.B., Golding, G.B. 2001. The closest blast hit is often not the nearest neighbor. J. Mol. Ecol.
804 52, 540-542.

805 Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H. 2005. Use of DNA barcodes
806 to identify flowering plants. Proc. Natl. Acad. Sci. U.S.A. 102, 8369-8374.

807 Kress,W.J., Erickson, D.L. 2007. A two-locus global DNA barcode for land plants: the coding *rbc*L
808 gene complements the non-coding t*rn*H*psb*A spacer region. Plos one 2(6): e508.
809 (doi:10.1371/journal.pone.0000508)

810 Kumar, N.P., Rajavel, A.R., Natarajan, R., Jambulingam, P. 2007. DNA barcodes can distinguish
811 species of Indian mosquitoes (Diptera : *Culicidae*). J. Med. Entomol. 44, 1-7.

812 Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit,
813 Barraclough, T.G., and Savolainen, V. 2008. DNA barcoding the floras of biodiversity
814 hotspots. Proc. Natl. Acad. Sci. U.S.A. 105 (8), 2923-2928.

815 Lane, C.E., Lindstrom, S.C., Saunders, G.W. 2007. A molecular assessment of northeast Pacific
816 *Alaria* species (Laminariales, *Phaeophyceae*) with reference to the utility of DNA barcoding.
817 Mol. Phylogenet. Evol. 44, 634-648.

818 Lawson, D., Arensburger, P., Atkinson, P., Besansky, N.J., Bruggner, R.V., Butler, R., Campbell,
819 K.S., Christophides, G.K., Christley, S., Dialynas, E., Emmert, D., Hammond,M., Hill, C.A.,
820 Kennedy, R.C., Lobo, N.F., MacCallum, M.R., Madey, G., Megy, K., Redmond, S., Russo, S.,
821 Severson, D.W., Stinson, E.O., Topalis, P., Zdobnov, E.M., Birney, E., Gelbart, W.M.,
822 Kafatos, F.C., Louis, C., Collins, F.H. 2007. VectorBase: a home for invertebrate vectors of
823 human pathogens. Nucleic Acids Res. 35, D503-505.

824 Lefebure, T., Douady, C.J., Gouy, M., Gibert, J. 2006. Relationship between morphological
825 taxonomy and molecular divergence within *Crustacea*: Proposal of a molecular threshold to
826 help species delimitation. Mol. Phylogenet. Evol. 40, 435-447.

827 Leclerc, M., Juste, C., Marteau, P., Nalin R., Blottiere H., Dore, J. 2007. Intestinal metagenomics
828 and nutrition. Annals Nutri. Metabol. 51 (13) (13 Suppl. 1).

829 Lipscomb, D., Platnick, N., Wheeler, Q. 2003. The intellectual content of taxonomy: a comment on
830 DNA taxonomy. Trends Ecol. Evol. 18, 65-66.

831 Litaker, R.W., Vandersea, M.W., Kibler, S.R., Reece, K.S., Stokes, N.A., Lutzoni, F.M., Yonish,
832 B.A., West, M.A., Black, M.N.D., Tester, P.A. 2007. Recognizing dinoflagellate species using
833 ITS rDNA sequences. J. Phycol. 43, 344-355.

834 Little, D.P., Stevenson, D.W. 2007. A comparison of algorithms for the identification of specimens
835 using DNA barcodes: examples from gymnosperms. Cladistics 23, 1-21.

836 Lorenz, J.G., Jackson, W.E., Beck, J.C., Hanner, R. 2005. The problems and promise of DNA
837 barcodes for species diagnosis of primate biomaterials. Phil. Trans. R. Soc. B 360, 1869-1877.

838 Lynn, D.H., Struder-Kypke, M.C. 2006. Species of *Tetrahymena* identical by small subunit rRNA
839 gene sequences are discriminated by mitochondrial cytochrome c oxidase I gene sequences. J.
840 Eukaryot. Microbiol. 53, 385-387.

841 Maiden, M.C.J., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R. Zhang, Q., Zhou, J.,
842 Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G. 1996. Multilocus sequence
843 typing: A portable approach to the identification of clones within populations of pathogenic
844 microorganisms. Proc. Natl. Acad. Sci. U.S.A. 95 (6), 3140-3145.

845 Matz, M.V., Nielsen, R. 2005. A likelihood ratio test for species membership based on DNA
846 sequence data. Phil. Trans. R. Soc. B 360, 1969-1974.

847 Meyer, C.P., Paulay, G. 2005. DNA barcoding: Error rates based on comprehensive sampling. Plos
848 Biology 3, 2229-2238.

849 Miller, S.E. 2007. DNA barcoding and the renaissance of taxonomy. Proc. Natl. Acad. Sci. U.S.A.
850 104, 4775-4776.

851 Min, X.J., Hickey, D.A. 2007. Assessing the effect of varying sequence length on DNA barcoding of
852 fungi. Mol. Ecol. Notes 7, 365-373.

853 Neigel, J., Domingo, A., Stake, J. 2007. DNA barcoding as a tool for coral reef conservation. Coral
854 Reefs 26, 487-499.

855 Nielsen, R., Matz, M. 2006. Statistical approaches for DNA barcoding. Syst. Biol. 55, 162-169.

856 Page, T.J., Choy, S.C., Hughes, J.M. 2005. The taxonomic feedback loop: symbiosis of morphology
857 and molecules. Biol. Lett.1, 139-142.

858 Paquin, P., Hedin, M. 2004. The power and perils of "molecular taxonomy": a case study of eyeless
859 and endangered Cicurina (*Araneae: Dictynidae*) from Texas caves. Mol. Ecol. 13: 3239-3255.

860 Passmore, A.J., Jarman, S.N., Swadling, K.M., Kawaguchi, S., McMinn, A., Nicol, S. 2006. DNA as
861 a Dietary Biomarker in Antarctic Krill, *Euphausia superba*. Mar. Biotech. 8, 686-696.

862 Pennisi, E. 2007 Wanted: A Barocde for Plants. Science 318(2): 190-191.

863 Pegg, G.G., Sinclair, B., Briskey, L., Aspden, W.J. 2006. MtDNA barcode identification of fish
864 larvae in the southern Great Barrier Reef, Australia. *Scientia Marina* 70, 7-12.

865 Pfenninger, M., Nowak, C., Kley, C., Steinke, D. Streit, B. 2007. Utility of DNA taxonomy and
866 barcoding for the inference of larval community structure in morphologically cryptic
867 *Chironomus* (*Diptera*) species. Mol. Ecol. 16, 1957-1968.

868 Pons, T., Gonzalez, B., Ceciliani, F., Galizzi, A. 2006. FlgM anti-sigma factors: identification of
869 novel members of the family, evolutionary analysis, homology modeling, and analysis of
870 sequence-structure-function relationships. J. Mol. Model. 12 (6), 973-983.

871 Pook, C.E., McEwing, R. 2005. Mitochondrial DNA sequences from dried snake venom: a DNA
872 barcoding approach to the identification of venom samples. Toxicon 46, 711-715.

873 Rastogi, G., Dharne, M.S., Walujkar, S., Kumar, A., Patole, M.S., Shouche, Y.S. 2007. Species
874 identification and authentication of tissues of animal origin using mitochondrial and nuclear
875 markers. Meat Sci. 76, 666-674.

876 Ratnasingham, S., Hebert, P.D.N. 2007. BOLD: The Barcode of Life Data System
877 (www.barcodinglife.org). Mol. Ecol. Notes 7, 355-364.

878 Remigio, E.A., Hebert, PDN. 2003. Testing the utility of partial COI sequences for phylogenetic
879 estimates of Gastropod relationships. Mol. Phylogenet. Evol. 29, 641-647.

880 Richly, E., Leister, D. 2004. NUMTs in sequenced eukaryotic genomes. Mol. Biol. Evol. 21, 1081-
881 1084.

882 Robba, L., Russell, S.J., Barker, G.L., Brodie, J. 2006. Assessing the use of the mitochondrial cox1
883 marker for use in DNA barcoding of red algae (*Rhodophyta*). Am. J. Bot. 93, 1101-1108.

884 Ross, H.A., Murugan, S., Li, W.L.S. 2008. Testing the reliability of genetic methods of species
885 identification via simulation. *Systematic Biology* 57:216-230.

886 Rot, C., Goldfarb, I., Ilan, M., Huchon, D. 2006 Putative cross-kingdom horizontal gene transfer in
887 sponge (*Porifera*) mitochondria. BMC Evol. Biol. 6(71). (doi:10.1186/1471-2148-6-71)

888 Rowley, D.L., Coddington, J., Norrbom, A., Ochoa, R., Vandenberg, N., Greenstone, M.H. 2007.
889 Vouchering specimens for documenting arthropod barcodes: a non-destructive method for
890 DNA extraction. Mol. Ecol. Notes 7, 915-924.

891 Rubinoff, D., Holland, B.S. 2005. Between two extremes: mitochondrial DNA is neither the panacea
892 nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 54,92-961.

893 Rubinoff, D. 2006a. Utility of mitochondrial DNA barcodes in species conservation. Cons. Biol. 20,
894 1026-1033.

895 Rubinoff, D. 2006b. DNA barcoding evolves into the familiar. Cons. Biol. 20, 1548-1549.

896 Rubinoff, D., Cameron, S. Will, K. 2006. Are plant DNA barcodes a search for the Holy Grail?
897 Trends Ecol. Evol. 21, 1-2.

898 Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D.,
899 Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson,H.,
900 Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J.E. , Li, K., Kravitz, S.,
901 Heidelberg, J.F., Utterback, T., Rogers, Y-H., Falcon, L.I., Souza, V., Bonilla-Rosso, G.,
902 Eguiarte, L.E., Karl, D.M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V.,
903 Tamayo-Castillo,G., Ferrari, M.R., Strausberg, R.L., Nealson, K., Friedman, R., Frazier, M.,
904 Venter, J.C. 2007. The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic
905 through eastern tropical Pacific. Plos Biol. 5(3) e77, 398-431.
906 (doi:10.1371/journal.pbio.0050077)

907 Sarkar, I.N., Joseph, W.T., Paul, J.P., David, H.F., Bernd, S., Rob., D. 2002a. An automated
908 phylogenetic key for classifying homeoboxes. Mol. Phylogenet. Evol. 24, 388-399.

909 Sarkar I.N., Paul, J.P., Bael, T.E., Stanley, S.E., Siddall, M., DeSalle, R., David, H.F. 2002b.
910 Characteristic attributes in cancer microarrays. J. Biomed. Info. 35, 111-122.

911 Saunders, G.W. 2005. Applying DNA barcoding to red *macroalgae*: a preliminary appraisal holds
912 promise for future applications. Phil. Trans. R. Soc. Lond. B 360, 1879-1888.

913 Schindel, D.E., Miller, S.E. 2005. DNA barcoding a useful tool for taxonomists. Nature 435, 17-17.

914 Schreiber, M.J., Ong, S.H., Holland, R.C.G., Hibberd, M.L., Vasudevan, S.G., Mitchell, W.P.,
915 Holmes, E.C. 2007. DengueInfo: A web portal to dengue information resources. Infect. Genet.
916 Evol. 540-541.

917 Scicluna, S.M., Tawari, B., Clark, C.G. 2006. DNA barcoding of *Blastopystis*. Protist 157, 77-85.

918 Seifert, K.A., Samson, R.A., Dewaard, J.R., Houbraken, J., Levesque, C.A., Moncalvo, J.M., Louis-
919 Seize, G., Hebert, P.D.N. 2007. Prospects for fungus identification using C01 DNA barcodes,
920 with *Penicillium* as a test case. Proc. Natl. Acad. Sci. U.S.A. 104, 3901-3906.

921 Sevilla, R.G., Diez, A., Noren, M., Mouchel, O., Jerome, M., Verrez-Bagnis, V., van Pelt, H., Favre-
922 Krey, L., Krey, G., Bautista, J.M. 2007. Primers and polymerase chain reaction conditions for
923 DNA barcoding teleost fish based on the mitochondrial cytochrome b and nuclear rhodopsin
924 genes. Mol. Ecol. Notes 7, 730-734.

925 Shenoy, B.D., Jeewon, R., Hyde, K.A. 2007. Impact of DNA sequence-data on the taxonomy of
926 anamorphic fungi. Fungal Div. 26, 1-54.

927 Smith, M.A., Fisher, B.L. Hebert, P.D.N. 2005. DNA barcoding for effective biodiversity
928 assessment of a hyperdiverse arthropod group: the ants of Madagascar. Phil. Trans. R. Soc. B
929 360, 1825-1834.

930 Smith, G.F., Roux, J.P., Tolley, K. Conrad, F. 2006a. Taxonomy and barcoding: conflict or
931 companions? S. Af. J. Sci. 102, 517-518.

932 Smith, M.A., Woodley, N.E., Janzen, D.H., Hallwachs, W., Hebert, P.D.N. 2006b. DNA barcodes
933 reveal cryptic host-specificity within the presumed polyphagous members of a genus of
934 parasitoid flies (*Diptera : Tachinidae*). Proc. Natl. Acad. Sci. U.S.A. 103, 3657-3662.

935 Smith, M.A., Wood, D.M., Janzen, D.H., Hallwachs, W., Hebert, P.D.N. 2007. DNA barcodes
936 affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, *Tachinidae*)
937 are not all generalists. Proc. Natl. Acad. Sci. U.S.A. 104(12), 4967-4972.

938 Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M.,
939 Herndl, G.J. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere".
940 Proc. Natl. Acad. Sci. U.S.A. 103, 12115-12120.

941 Strugnell, J.M., Lindgren, A.R. 2007. A barcode of life database for the C*ephalopoda*?
942 Considerations and concerns. Rev. Fish Biol. Fish. 17, 337-344.

943 Sugita, T., Nishikawa, A., Shinoda, T. 1998. Identification of *trichosporon asahii* by PCR based on
944 sequences of the internal transcribed spacer regions. J. of clinical microbiol. 2742–2744.

945 Summerbell, R.C., Levesque, C.A., Seifert, K.A., Bovers, M., Fell, J.W., Diaz, M.R., Boekhout, T.,
946 de Hoog, G.S., Stalpers, J., Crous, P.W. 2005. Microcoding: the second step in DNA
947 barcoding. Phil. Trans. R. Soc. B 360, 1897-1903.

948 Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier,
949 G., Brochmann, C., Willerslev, E. 2007. Power and limitations of the chloroplast *trn*L (UAA)
950 intron for plant DNA barcoding. Nucleic Acids Res. 35 (3), e14.

951 Tavares, E.S., and Baker, A.J. 2008. Single mitochondrial gene barcodes reliably identify sister-
952 species in diverse clades of birds. BMC Evol. Biol. 8: 81

953 Terranova, M.S., Brutto, S.L., M. Arculeo, M., Mitton, J.B. 2007. A mitochondrial phylogeography
954 of *Brachidontes variabilis* (*Bivalvia: Mytilidae*) reveals three cryptic species J. Zool. Syst.
955 Evol. Res. 45(4), 289-298.

956 Teletchea, F., Bernillon, J., Duffraisse, M., Laudet, V., Hänni, C. 2008 Molecular identification of
957 vertebrate species by oligonucleotide microarray in food and forensic samples. J. of App.
958 Ecol.(doi: 10.1111/j.1365-2664.2007.01415.x)

959 Thalmann, O., Hebler, J., Poinar, H.N., Paabo, S., Vigilant, L. 2004. Unreliable mtDNA data due to
960 nuclear insertions: a cautionary tale from analysis of humans and other great apes. Mol. Ecol.
961 13, 321-335.

962 Thiyagarajan, S., Karhanek, M., Akhras, M., Davis, R.W., Pourmand, N. 2006. PathogenMIPer: a
963 tool for the design of molecular inversion probes to detect multiple pathogens. BMC
964 Bioinform. 7. (doi:10.1186/1471-2105-7-500).

965 Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M.,
966 Short, J.M., Mathur, E.J., Detter, J.C., Bork, P., Hugenholtz, P., Rubin, E.M. 2005.
967 Comparative metagenomics of microbial communities. Science 308, 554-557.

968 Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev,
969 V.V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F. 2004. Community structure and metabolism
970 through reconstruction of microbial genomes from the environment. Nature 428, 37-43.

971 Van Velzen, R., Freek, T., Bakker, L., Van Loon, J.J.A. 2007. DNA barcoding reveals hidden
972 species diversity in Cymothoe (*Nymphalidae*). Proc. Neth. Entomol. Soc. Meet. 18, 95-103.

973 Vences, M., Thomas, M., Van der Meijden, A., Chiari, Y., Vieites, D.R. 2004. Performance of 16S
974 rRNA in DNA barcoding of amphibians. Integr. Comp. Biol. 44, 657-657.

975 Vences, M., Thomas, M., Bonett, R.M., Vieites, D.R. 2005. Deciphering amphibian diversity
976 through DNA barcoding: chances and challenges. Phil. Trans. Roy. Soc. Lond. B 360, 1859-
977 1868.

978 Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson
979 KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J,
980 Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers Y, Smith HO: Environmental
981 genome shotgun sequencing of the sargasso sea. Science 2004 304, 66-74.

982 Vogler, A.P., Monaghan, M.T. 2007. Recent advances in DNA taxonomy. J. Zool. Syst. Evol. Res.
983 45(1), 1-10.

984 Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D.N. 2005. DNA barcoding Australia's
985 fish species. Phil. Trans. Roy. Soc. Lond. B. 360, 1847-1857.

986 Webb, K.E., Barnes, D.K.A., Clark, M.S., Bowden, D.A. 2006. DNA barcoding: A molecular tool to
987 identify Antarctic marine larvae. Deep-Sea Res. Part II 53, 1053-1060.

988 Weeks, A.R., Turelli, M., Harcombe, W.R., Reynolds, K.T., Hoffmann, A.A. 2007. From parasite to
989 mutualist: Rapid evolution of *Wolbachia* in natural populations of D*rosophila*. Plos Biol 5(5),
990 e114. (doi:10.1371/journal.pbio.0050114)

991 Whitworth, T.L., Dawson, R.D., Magalon, H., Baudry, E. 2007. DNA barcoding cannot reliably
992 identify species of the blowfly genus *Protocalliphora* (*Diptera : Calliphoridae*). Proc. Roy.
993 Soc. Lond. B 274, 1731-1739.

994 Wiemer, M., Fiedler, K. 2007. Does the DNA barcoding gap exist? – a case study in blue butterflies
995 (*Lepidoptera: Lycaenidae*). Front. Zool. 4(8) (doi:10.1186/1742-9994-4-8).

996 Winnenburg, R., Baldwin1, T.K., Urban1, M., Rawlings, C. Kohler, J., Hammond-Kosack, K.E.
997 2006. Nucleic Acids Res. 34, D459-464 (doi:10.1093/nar/gkj047).

998 Willams, S.T., Knowlton, N. 2001. Mitochondrial pseudogenes are pervasive and often insidious in
999 the snapping shrimp genus *Alpheus*. Mol. Biol. Evol. 18,1484-1493

1000 Wilson, O.E. 2004. The encyclopedia of life. Trends Ecol. Evol. 18(2), 77-80.

1001 Wirth, T., Le Guellec, R., Veuille, M. 1999. Directional Substitution and Evolution of Nucleotide
1002 Content in the Cytochrome Oxidase II Gene in Earwigs (Dermapteran Insects). Mol. Biol.
1003 Evol. 16(12), 1645-1653.

1004 Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., Karch, H., Reeves, P.R., Maiden,
1005 M.C.J., Ochman, H., Achtman, M. 2006. Sex and virulence in *Escherichia coli*: an
1006 evolutionary perspective. Mol. Microbiol. 60 (5), 1136-1151.

1007 Woese, C.R. 1996. Phylogenetic trees: whither microbiology? Current biol. 1060-1063.

1008 Xiao, M., Gordon, M.P., Phong, A., Ha, C., Chan, T.F., Cai, D.M., Selvin, P.R., Kwok, P.Y. 2007.
1009 Determination of haplotypes from single DNA molecules: A method for single-molecule
1010 barcoding. Hum. Mut. 28, 913-921.

1011 Zhou, J., Davey M.E., Figueras, J.B., Rivkina, E., Gilichinsky, D., Tiedje, J.M.. 1997. Phylogenetic
1012 diversity of a bacterial community determined from Siberian tundra soil DNA. Microbiol. 143,
1013 3913-3919.

1014