

DNA barcoding as a tool for coral reef conservation

J. Neigel · A. Domingo · J. Stake

Received: 9 January 2007 / Accepted: 4 May 2007
© Springer-Verlag 2007

Abstract DNA Barcoding (DBC) is a method for taxonomic identification of animals that is based entirely on the 5' portion of the mitochondrial gene, *cytochrome oxidase subunit I (COI-5)*. It can be especially useful for identification of larval forms or incomplete specimens lacking diagnostic morphological characters. DBC can also facilitate the discovery of species and in defining “molecular taxonomic units” in problematic groups. However, DBC is not a panacea for coral reef taxonomy. In two of the most ecologically important groups on coral reefs, the Anthozoa and Porifera, *COI-5* sequences have diverged too little to be diagnostic for all species. Other problems for DBC include paraphyly in mitochondrial gene trees and lack of differentiation between hybrids and their maternal ancestors. DBC also depends on the availability of databases of *COI-5* sequences, which are still in early stages of development. A global effort to barcode all fish species has demonstrated the importance of large-scale coordination and is yielding

promising results. Whether or not *COI-5* by itself is sufficient for species assignments has become a contentious question; it is generally advantageous to use sequences from multiple loci.

Keywords DNA barcoding · DNA taxonomy · Cytochrome oxidase I · Environmental sampling · Microarrays · Real-time PCR

Introduction

Coral reef communities around the world have been in decline since humans began to exploit them (Pandolfi et al. 2003), but the degradation of reefs has rapidly accelerated over the past several decades (Knowlton 2001; McClanahan 2002; Gardner et al. 2003; Cote et al. 2005). This crisis has made the development of new and more effective approaches to coral reef conservation a high priority (Bellwood et al. 2004). However, our ability to assess and respond to changes in coral reef communities is limited by the current state of taxonomy and systematics of coral reef species. Coral reefs are the most diverse of marine communities and many of their species remain undescribed. For some groups, such as the Porifera and Scleractinia, traditional approaches based on morphology alone have proven unreliable (e.g., Romano and Palumbi 1996; Lazoski et al. 2001). For other groups, there are simply too many undescribed species for the few systematists who are trained to describe them. Even for species that have been adequately described, field keys are often incomplete or unreliable. Larval forms of fish and invertebrates are especially problematic; often they can only be identified to the level of family or genus. These taxonomic problems introduce extreme biases in surveys of biodiversity and community

Communicated by Biology Editor M. van Oppen.

J. Neigel (✉) · A. Domingo · J. Stake
Department of Biology, University of Louisiana at Lafayette,
Box 42451, Lafayette, LA 70504, USA
e-mail: jneigel@louisiana.edu

A. Domingo
Smithsonian Tropical Research Institute,
Unit 0948, APO AA 34002-0948, Miami, USA
e-mail: alanadon@gmail.com

J. Stake
Division of Science and Math,
University of the Virgin Islands,
St. Thomas, VI 00802, USA
e-mail: jstake@uvi.edu

structure, which favor groups and life-stages that are relatively easy to identify in the field (Mikkelsen and Cracraft 2001).

Recent trends in the science of marine reserve design have emphasized connectivity within reserve networks (Roberts 1997; Cowen et al. 2006) and the need to consider all life stages (St Mary et al. 2000; Gaylord et al. 2005), with particular emphases on dispersal stages (Gaines et al. 2003). These considerations are often based on familiar fish and invertebrate species with larvae that are planktonic for weeks or longer. However, this may again reflect our bias towards large, conspicuous organisms. Small body size can be correlated with limited dispersal and narrower geographic range (Reaka-Kudla 2000). Very small animals (mm in length) and those that live hidden within sediments often lack planktonic dispersal (Grantham et al. 2003). A conceptual bias that favors large-bodied organisms and geographically widespread species may have also contributed to a commonly held belief that species-area relationships are unimportant in the marine realm, despite evidence to the contrary (Neigel 2003).

DNA barcoding (DBC) is an alternative to traditional taxonomic methods that could become a useful tool for coral reef conservation. The potential value of DBC in marine biology has been commented on elsewhere (Schander and Willassen 2005), and the Census of Marine Life has included DBC among its recommended methods (O'Dor and Gallardo 2005). Two main purposes have been proposed for DBC, (1) assignment of specimens to known species, and (2) discovery of new species. Neither is a novel use of DNA sequences; each has precedents in molecular systematics (Avice 1994). What is new about DBC is a set of tools and protocols that can be widely applied to the majority of metazoan animals. The protocols include standards for the quality of DNA sequence references and the curation of voucher specimens that represent significant improvements over those currently in practice for public sequence databases such as GenBank (Hanner 2005). The most distinctive and controversial assumption of DBC is that divergence in the 5' portion of the mitochondrial gene *cytochrome oxidase I* (*COI-5*) is sufficient to reliably distinguish species. Perhaps not surprisingly, the strongest proponents of DBC have been field ecologists who are desperate to solve the problem of specimen identification while its most outspoken critics have been systematists who have raised both philosophical and practical objections.

Here we review the current status of DBC with specific consideration of its application to coral reef conservation biology. We begin with the original concept of DBC, as it was first proposed by Paul Hebert and colleagues. Next, we consider some of the different ways that DBC can be used, especially in the context of coral reef systems. We then

consider some of the limitations of DBC along with potential alternatives. Finally, we summarize the current status of DBC and its possible future.

The concept of DNA barcoding

In 2003, Hebert and co-workers proposed that DNA sequences should be used as “taxon barcodes” to circumvent the “limitations inherent in morphology-based identification systems” and the problem of the “dwindling pool of taxonomists” (Hebert et al. 2003a). The term “DNA barcode” was intended to suggest that DNA sequences can identify species in much the same way as 11-digit Universal Product Codes (UPCs) identify retail products. Central to this concept is the use of a sequence standard that corresponds to a single homologous gene region, can be amplified by the polymerase chain reaction (PCR) with “universal primers”, and distinguishes species across a broad range of taxa. At present, only mitochondrial sequences come close to fulfilling these requirements and *COI-5* in particular is a logical choice. The basic gene content of metazoan mitochondrial genomes is mostly conserved so that identification of orthologous sequences is straightforward. Within *COI* gene sequences there are stretches that are highly conserved at the amino acid level, which has made it possible to design “universal” PCR primers that amplify a ~700 base pair (bp) region of the 5' end of the gene across a broad range of metazoan phyla (Folmer et al. 1994). In contrast to these conserved sequences, the overall rate of sequence evolution for *COI-5* is relatively high, especially at degenerate codon positions. With some important exceptions that are discussed below, the rate of evolution is high enough to result in sequence divergence between most species as well as varying levels of sequence polymorphism within species. There are some additional advantages offered by mitochondrial sequences. Mitochondria are haploid and in most cases, maternally inherited, so the problem of sequencing heterozygous templates is avoided. Mitochondrial genomes are also relatively unaffected by recombination, which simplifies phylogenetic analysis. Finally, sequences for *COI* and other mitochondrial genes have already been determined for a large number of metazoan species and deposited in public databases. For all these reasons it difficult to argue that a sequence region other than *COI-5* would have been a much better choice as a standard DNA barcode.

Identification by DBC is based on matching an unknown specimen's barcode sequence to one or more sequences from specimens that have been positively identified by other means. The effectiveness of this approach clearly depends on the availability of extensive databases of *COI-5* barcode sequence standards. There are now initiatives to

assemble such databases for particular groups of organisms and to provide a central repository for all barcode data and analytical tools for barcode data users (see “[Barcoding initiatives](#)”). *COI* sequences can also be found in general sequence databases, such as GenBank, although entries into these databases are generally not screened for taxonomic accuracy and do not require that sequences be associated with voucher specimens. However, new requirements for barcode submissions to GenBank that are identified by the keyword BARCODE are under development. DNA barcodes are not perfectly analogous to UPCs because every species is not characterized by a single, unique barcode sequence. Typically, *COI-5* sequences are polymorphic and the frequencies of haplotypes (sequence variants) often vary geographically. Thus the ideal barcode database would include an extensive set of *COI-5* sequences that represented the full range of variation for every species.

In many cases, a barcode sequence from an unidentified specimen will not be an exact match with any database sequence. This could reflect either the absence of sequences from the relevant species in the database or simply the absence of a particular haplotype for that species. Criteria are therefore needed to decide when a match is “close enough” to be considered a positive for species identification. The approach advocated by Hebert et al. (2003b) is to examine the ranges of sequence divergence within and between species, and set a threshold of divergence for species assignments. In some cases the ranges will overlap, so that some values of divergence will be ambiguous. As noted by Hebert et al. (2003b), ranges of sequence variation differ among taxa, so that thresholds should be estimated separately for each taxonomic group.

In addition to species identification, DBC can lead to the discovery of species. This has occurred as a result of the collection of barcode sequences; occasionally sequences that are as divergent as those from recognized species-pairs are found in specimens of the same nominal species (Hebert et al. 2004b; Holland et al. 2004). This would be expected if the nominal species is actually a group of morphologically cryptic species, a situation that appears to be common in some marine taxa (Knowlton 1993). Such findings by themselves may not justify taxonomic revision; but they can at least lead to further investigation that might confirm the existence of previously unrecognized species (e.g., Gómez et al. 2007). Species discovery can also be accelerated by the efficiency of DBC in comparison with traditional means of specimen identification. If all of the species in a group (e.g., a family) are well represented in a barcode sequence database it will be easier to spot outliers that could belong to an unrecognized species; the specimens in question can then be more carefully examined by a taxonomic specialist.

Barcoding initiatives

There are now about a half-dozen well-organized efforts to develop community databases of *COI-5* sequences and other barcoding resources. The international umbrella organization for all of these is the Consortium for the Barcode of Life (CBOL), which is hosted at the Smithsonian Institution in Washington, DC, and at the time of writing listed over 130 member organizations (<http://www.barcoding.si.edu/>). The Barcode of Life Data Systems (BOLD) (<http://www.boldsystems.org>) is a web-based workbench for coordinating the data collection activities of a barcoding project and performing data analysis with barcode data (Ratnasingham and Hebert 2007). BOLD has three components: (1) A database of barcode records (*COI-5* sequences and specimen information) with analytical tools to explore them (MAS); (2) an identification engine (IDS) that attempts to find a species-level match between a *COI-5* sequence from an unknown specimen and a database record; and (3) tools for developing new data analysis modules that connect with the databases (ECS). At present there are two BOLD databases that can be searched. The full database of all available barcode records (at present 165,048 sequences representing 19,163 species) includes species represented by fewer than three specimens and sequences that have not been validated. The reference barcode database is a subset (59,885 sequences representing 6,476 species) of validated records for species represented by three or more individuals and that show less than 2% sequence divergence within species.

There are also several taxon-specific DBC initiatives underway of special relevance to coral reef conservation. One of the most ambitious and successful to date is the Fish Barcode of Life Initiative (FISH-BOL), a coordinated global effort to barcode all 29,112 recognized fish species (<http://www.fishbol.org/>). At the time of writing (29 November 2006) FISH-BOL includes 11,813 sequences from 2,705 species, or 9% of the total number of recognized species. We investigated the progress of FISH-BOL for reef fishes by comparing records for reef-associated species in FishBase (Froese and Pauly 2000) with barcode records in FISH-BOL. As of now, FishBase lists 4,263 species as reef-associated. Of these, 957 (22.4%) are represented by at least one sequence in FISH-BOL, and 478 (11.2%) were represented by three or more sequences (Fig. 1). The reef-associated species in FishBase are assigned to 929 genera. For 129 (15.0%) of those genera, every reef-associated species listed in the genus is represented by at least one sequence in FISH-BOL and for 56 (6.0%) every reef-associated species in a genus is represented by at least three sequences.

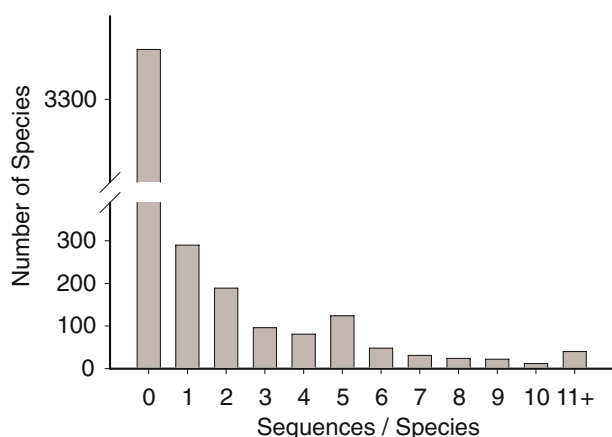


Fig. 1 Number of 5' cytochrome oxidase subunit I sequences per species for reef associated fish species in the FISH-BOL database

DNA barcoding for coral reef biology

Surveys of biodiversity inevitably require tradeoffs between taxonomic coverage and spatiotemporal coverage. It is possible for a SCUBA diver surveying a transect on a Caribbean reef to identify all of the scleractinian corals and fish encountered. However the diver would have considerable difficulty identifying all of the hydroids, bryozoans, and sponges seen. Virtually impossible would be identification of cryptic organisms, such as fish and arthropods living within the cavities of sponges and corals, or the nematodes, copepods, gastropods and other meiofauna living in sediments. For identification of all but the most conspicuous organisms, it would be necessary to collect specimens, bring them to the laboratory, and spend many hours working through numerous keys. In the time it would take to identify all of the meiofauna present in a few hundred cubic centimeters of sediment, it might be possible to survey coral and fish species over several kilometers of transects. Confined by these limitations, most of what we know about large-scale patterns and trends in the biodiversity of marine benthos has been based on the most conspicuous and easily field-identified taxa (Mikkelsen and Cracraft 2001). DBC and related approaches offer the potential for significant advances in the study of coral reef biodiversity and ecology. Next we consider several of these possibilities.

Identification of individual specimens

Assignment of specimens to species is the core of field taxonomy and the most straightforward application of DBC. *COI-5* sequences are determined for unidentified specimens from the field and the sequences are (hopefully) matched to

those in a database of reference sequences. Specimens can be whole organisms, parts of organisms, or even the remains of organisms found in the guts of predators (e.g., Smith et al. 2005). The techniques involved are sample preservation, DNA extraction, PCR, sequencing, and database analysis. Although it is common to preserve specimens in ethanol or other preservatives for transport back to a central laboratory for processing, it has been our experience that better yields and quality of DNA can be obtained if extractions are performed on freshly collected tissue. Fortunately, DNA extractions and even PCR can be performed at field stations or in hotel rooms. All of the essential supplies and equipment can be packed into a suitcase or two and DNA samples are much easier to transport than preserved specimens. It can also be good insurance to complete some PCR amplifications at field sites, since this is the step that is most likely to fail if there is a problem with DNA quality. A few test amplifications can indicate if additional DNA extractions or even additional specimens are needed.

There is an expectation, or at least a hope, that advances in technology will soon culminate in a hand-held barcoding device that can identify specimens in the field without the need for a laboratory at all (Janzen 2004). While this vision is tantalizing, the necessary technology has yet to be developed and the market demand for such devices might be more comparable to that for pocket microscopes than for mobile phones. A more likely development that we can foresee in the near future would be a bench-top system that could be transported and set-up at field stations and would automate the entire DNA barcoding process at high rates of throughput and low unit costs.

An important issue that is often raised in discussions of DBC is the value of voucher specimens. We strongly agree that properly curated voucher specimens should be available for barcode reference standards. However, it will often be impractical to maintain vouchers for every specimen that is identified by DBC. For very small organisms (i.e., dimensions of several millimeters or less) it is difficult to obtain a DNA sample without destroying the specimen. For very large organisms (e.g., large invertebrate colonies or marine mammals) it is usually impractical, unethical or illegal to kill and collect whole specimens. Even for moderate-sized organisms the value of individual specimens as vouchers must be weighed against the costs of transporting and storing them by the thousands. A practical compromise is to keep subsamples of specimens as representative vouchers and rely on photographic records for the remainder. In contrast, there is little cost associated with the transport and storage of quantities of extracted DNA beyond what are needed for barcode identifications, but there is great potential value. As discussed below, there are concerns about the sufficiency of *COI-5* alone for species

assignment that could be addressed by generating data for additional sequences from archived DNA samples.

Identification of early life stages

One of the most exciting applications of DBC is identification of larval or juvenile individuals that lack the morphological characters that are the basis of traditional species descriptions and keys. It is possible to obtain sufficient DNA from larvae as small as ~1 mm in length for amplification of *COI-5* (Baldwin et al. 1996; Hare et al. 2000; Schander and Willassen 2005; Pegg et al. 2006; Richardson et al. 2006; Webb et al. 2006). Identification of larvae or recently settled benthic juveniles should open up new possibilities for the measurement of recruitment success (e.g., Shearer and Coffroth 2006), a parameter of great importance in the dynamics of marine populations (Caley et al. 1996). Surveys of planktonic larvae could also provide a sample of local biodiversity that would complement benthic surveys. During spawning periods, planktonic larvae outnumber other life stages by orders of magnitude, and are easily collected with minimal impact on local populations. A compelling example of the potential for this approach to reveal previously hidden biodiversity comes from a recent survey of larval stomatopods on Indo-Pacific coral reefs (Barber and Boyce 2006). *COI-5* sequences from unidentified stomatopod larvae were grouped into operational taxonomic units (OTUs) defined by a threshold of 3% sequence divergence and compared with those of known adults. Although Indo-Pacific stomatopods are a well-studied group, at least 50% of the larval OTUs represented new, undescribed species.

Environmental sampling

Conventional application of DBC (and most other methods of identification) requires that each specimen is identified individually. However for organisms that are very small and numerous, this one-by-one approach can easily be overwhelmed by large numbers of specimens. A single plankton or meiofaunal sample can contain so many organisms that it is impractical to count them all, let alone perform DNA extractions on each individual. Small subsamples will fail to represent the less abundant components of the original sample. For environmental samples filled with organisms that are small and numerous it can be more effective to analyze the bulk composition of the sample as whole. Rather than separating individual organisms, DNA is extracted from entire sample, which yields a complex mixture of sequences from different species. The mixture can then be analyzed by methods that have been developed

for genomics research. This approach is now being used to characterize the “metagenomes” of microbial communities (Venter et al. 2004; DeLong 2005; Tringe and Rubin 2005). Remarkably, even cell-free DNA that is bound to beach sand can reveal both prokaryotic and eukaryotic components of a marine community (Naviaux et al. 2005).

Conventional PCR amplification and sequencing with universal barcode primers is not suitable for environmental samples. The DNA from a mixture of different organisms would produce an unintelligible signal, with the sequences of different species superimposed. Single molecules can be cloned and sequenced individually, but this approach leads to the same problems of scale as one-by-one analysis of individuals (Markmann and Tautz 2005). The individual sequences of thousands of different DNA molecules can be determined in parallel by methods such as pyrosequencing (e.g., Edwards et al. 2006). Unfortunately, routine application of massive sequencing to characterize environmental samples is not yet practical, at least in terms of cost. A more economical, although less comprehensive approach is to avoid DNA sequencing altogether and instead use the specificity of DNA hybridization probes to detect species-specific sequence “signatures” within mixtures (e.g., Goffredi et al. 2006). The tradeoff is that only those species that are specifically probed can be detected, and species for which no sequence information is available cannot be “discovered” by this approach. *COI-5* sequences are one possible target for species-specific probes. DNA microarrays and real-time PCR (RT-PCR) with fluorescent probes are two DNA hybridization technologies that have proven feasible for species identification.

DNA microarrays

DNA microarrays can be used to detect and roughly quantify DNA molecules with specific sequences in mixtures. The simplest form of a DNA microarray is a glass slide upon which microscopic spots of different DNA hybridization probes are deposited in a rectangular array. If the microarray were to be used for DBC, the mixture of PCR products from a whole-sample amplification of *COI-5* would be fluorescently labeled and added in a small volume of hybridization buffer to the slide. After the labeled PCR products bound to specific hybridization probes, unbound products would be washed off the slide. Fluorescent spots on the slide would then indicate the presence of PCR products that hybridized to specific probes and the locations of the spots would indicate which probes. Unfortunately, this simple concept for a microarray barcoder is not practical because there are tradeoffs among the sensitivity, specificity and sequence coverage of DNA hybridization probes. A probe complementary to the entire 600+ bp of *COI-5* would be sensitive enough to detect small amounts of DNA and

would provide complete coverage of the barcode sequence, but would lack the specificity needed to distinguish species. Hybridization between molecules that are hundreds of nucleotides in length typically require only 85–90% sequence identity (Call 2005), which would allow sequences from different species (with typical levels of intraspecific divergence) to hybridize. Although the amount of hybridization, and thus fluorescence, is reduced when sequences are only approximate matches, this signal is confounded with the effects of the concentration of each sequence in the mixture, which would be uncontrolled in environmental samples. Short oligonucleotide probes of about 25 nucleotides can provide the specificity required to detect single-nucleotide differences when compared to mismatch controls (e.g., Religio et al. 2002), but with coverage that would be limited to a small portion of the entire *COI-5* sequence. Combinations of short probes that provide overlapping coverage of longer sequences can be used to identify DNA from a single specimen (Angelov et al. 2004; Summerbell et al. 2005), however it remains to be seen whether this approach would work with signals generated by complex mixtures that represent many species. The power of DNA microarrays is their ability to simultaneously determine the presence or absence of thousands of different sequence targets. A logical way to use this power for species identification is to target not just one sequence, such as *COI-5*, but many different sequences to generate a detailed composite of species-distinguishing features (Borneman 2001).

Real-time PCR

RT-PCR provides greater sensitivity and accuracy in detection and quantification of specific DNA sequences than microarrays. The increase in concentration of an amplicon is quantified (in “real-time”) after each cycle by the accumulation of a fluorescent product. Quantification of the original template is based on how many cycles are required for the concentration of the product to cross a standard threshold of detection. The fluorescent product can be the amplicon itself, labeled with an intercalating dye such as Sybr Green, although specificity then depends entirely on the PCR primers. Fluorescence can also be generated by hybridization probes that bind specifically to the amplicon, which provides another level of specificity. A dual-labeled fluorescent probe (DLFP) consists of a single-stranded DNA with two dye molecules (Livak et al. 1995). When the probe is intact, one of the dyes (the quencher) prevents the other (the reporter) from fluorescing. However, if the probe hybridizes to a complementary amplicon it is degraded by the exonuclease activity of DNA polymerase, the reporter dye and quencher are separated and the reporter dye fluoresces (Heid et al. 1996). RT-PCR with DLFPs has been

used to identify the eggs and larvae of the Japanese eel (*Anguilla japonica*) (Watanabe et al. 2004) and to quantify laboratory-hatched abalone larvae (*Haliotis kamtschaticana*) (Vadopalas et al. 2006).

In our laboratory, we have developed RT-PCR assays with DLFPs for species-specific detection of fish and invertebrate larvae. The design of primers and probes for these RT-PCR is constrained with respect to the GC content, length, melting temperature and other sequence characteristics (Livak 1999). In addition to these thermodynamic requirements, if the probes are to be species-specific they must bind to a target within the amplicon that is constant within a species but divergent among species. Lack of complete probe specificity can often be compensated by primers that also confer specificity. However, with all of these constraints we have found that it is not always feasible to find a suitable target sequence within *COI-5*; we often must use other mitochondrial genes.

An important limitation of all of the above methods for detection of specific DNA sequences within mixtures is that they cannot resolve multi-locus genotypes. Once DNA molecules from different individuals are mixed, it is no longer possible to determine whether a particular multi-locus combination occurs in one individual. This is not to say that these methods are limited to single-gene identifications, but that it will be a challenge to analyze multi-locus data from environmental samples. We expect that this limitation will increase the popularity of single-gene approaches, such as DBC, in the analysis of complex environmental samples.

Describing species with DNA sequences

For the assignment of unknown specimens to recognized species, DBC is limited to species that have been formally described. This severely limits the scope of DBC identification for groups such as nematodes in which the vast majority of species have not been described. However, this limitation can be circumvented by assigning specimens to molecular operational taxonomic units (MOTUs) that are defined by DNA sequences but are intended to correspond to biological species. Perhaps it is not surprising that a nematode biologist has been a strong proponent of this approach to taxonomy (Blaxter 2004), which has proven its value in surveys of terrestrial meiofauna (Blaxter et al. 2005). The use of DBC to define species has been highly controversial because it is based on two practices that are considered unreliable: inference of species trees from gene trees and the use of a phenetic (distance) measure to delineate taxa (Lipscomb et al. 2003; Will and Rubinoff 2004; Brower 2006). These are valid concerns that should discourage the use of DBC to define species when more

suitable methods are available. Nevertheless for applications such as characterization of environmental samples that are rich in undescribed species, MOTUs are a justifiable alternative to ignorance.

DBC problems and alternatives

DBC has been subject to a fair amount of criticism (e.g., Will and Rubinoff 2004; Wheeler 2005; Will et al. 2005; Rubinoff 2006). While many of these criticisms are valid, most do not apply to all uses of DBC. For example, criticisms of the use of DBC to define MOTUs do not apply to the use of DBC to assign specimens to known taxa. The value of a DBC approach should be assessed for each application in terms of its expected cost and performance relative to alternatives. For general estimates of biodiversity, 90% accuracy for species assignments might compare favorably to that attained with morphological keys, while for documentation of new species it might be too low. Below we consider some of the limitations of DBC and the situations in which they become most problematic.

Problems with *COI* as the barcode standard

COI sequences cannot serve as a truly universal barcode because in some taxa mitochondrial sequences evolve too slowly to distinguish closely related species (Hebert et al. 2003a). Rates of mitochondrial sequence evolution that are an order of magnitude slower than those observed in most metazoans have been documented in higher plants (Wolfe et al. 1987), sponges (Erpenbeck et al. 2006), and anthozoans (McFadden et al. 2000; Shearer et al. 2002; Hellberg 2006). These taxa include individuals with very long lifespans (hundreds of years or longer), which may have required retention of ancestral DNA repair mechanisms to prevent the accumulation of deleterious mitochondrial mutations in somatic cell lineages (Hellberg 2006). Slow rates of mitochondrial sequence evolution are of concern in coral reef conservation, because two of the most ecologically important groups on coral reefs, the Scleractinia and the Demospongia are included in these groups. Interestingly, *COI* appears to have evolved at a rapid enough pace to distinguish species in the Rhodophyta (Saunders 2005; Robba et al. 2006) the Scyphozoa (Holland et al. 2004) and the Hydrozoa (Govindarajan et al. 2005).

Another limitation of taxonomy based on a single nuclear gene or mitochondrial sequences is the frequent lack of congruence between gene-trees and species-trees, sometimes called the “paraphyly problem”. Gene lineages can be polyphyletic or paraphyletic with respect to true species boundaries because of incomplete sorting of common

ancestral lineages (Neigel and Avise 1986) or introgressive hybridization (e.g., Shaw 2002). As a result, mitochondrial sequences from the same species may be less related to each other than to sequences from other species. In a survey of 584 published animal mitochondrial DNA studies (Funk and Omland 2003), paraphyly or polyphyly was found in 23.1% of the 2,319 species represented. Thus paraphyly appears to be a widespread phenomenon. However, as the authors of this survey pointed out, some unknown proportion of these apparent cases undoubtedly represent errors in phylogenetic inference, imperfect taxonomy, or gene duplications. True cases of paraphyly cannot be resolved with data from additional mitochondrial genes, because the entire mitochondrial genome is inherited as a single non-recombining unit. Only with independently segregating nuclear loci is it possible to obtain an “average” of individual gene trees that would faithfully represent the true species tree (Avise and Ball 1990).

Both the problem of slow mitochondrial sequence evolution in some taxa and the paraphyly problem were acknowledged by Hebert and colleagues from the start (Hebert et al. 2003b). They argued that these exceptions did not void the benefits of a single gene standard for specimen identification. Nevertheless, critics of DBC continue to point out these problems.

Alternatives to *COI*

The use of the 5' portion of the *COI* gene as a universal standard is the key feature that distinguishes the original concept of DBC from other approaches to molecular taxonomy, and so “barcoding” with any other gene would not really be DBC. Nevertheless, alternatives to *COI*-5 should be considered. As reviewed above, *COI* will not work for some groups because it evolves too slowly, but other sequences might be suitable. Erpenbeck et al. (2006) found that the 3' portion of the *COI* gene was more useful for species identification in the Porifera and Anthozoa than *COI*-5. In our laboratory, we have explored the use of anonymous single-copy nuclear sequences for taxonomy and systematics of scleractinian corals. These sequences readily differentiate species that cannot be distinguished by *COI*-5 sequences. For example, pairs of species in the genus *Porites* that have identical *COI*-5 sequences differ by at least 1% at an anonymous nuclear locus (Fig. 2). Even when *COI*-5 is suitable for DBC, there are other sequences that have already become standards for species identification in some groups. Fungal taxonomists use a combination of the internal transcribed spacers of the nuclear ribosomal gene cluster, variable domains of the 28S ribosomal gene, and the nuclear gene that encodes translation elongation factor 1 α (Summerbell

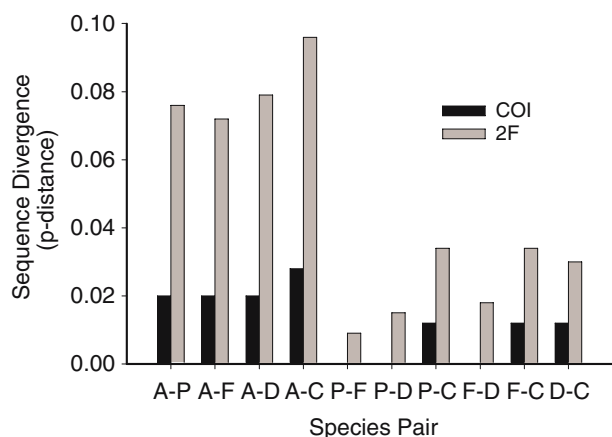


Fig. 2 Sequence divergence between DNA barcoding sequences from representatives of different species in the scleractinian coral genus, *Porites* in comparison with sequence divergence at an anonymous non-coding single copy nuclear locus for the same specimens. Some pairs of species have identical barcoding sequences, but all are distinguished by the anonymous nuclear sequence

et al. 2005). Nematode taxonomists favor the use of 18S and 28S ribosomal genes (Floyd et al. 2002; Powers 2004). These choices of genes are not arbitrary, but reflect their proven value for particular taxonomic groups. For example, in a study that compared the usefulness of several nuclear and mitochondrial genes for identification of nematodes, the ribosomal 18S gene was found to be the most consistently amplifiable and assigned over 97% of specimens to the correct species (Bhadury et al. 2006).

Detection of species-specific sequences in mixed environmental samples is especially challenging because standard methods based on DNA hybridization cannot distinguish among all *COI-5* variants or detect multi-locus combinations of markers (see above). For at least some species, non-coding middle repetitive sequences (NCMRS) can be used as diagnostic species-tags that do not suffer from these limitations. For reasons that are not completely understood (Georgiev et al. 1982) some NCMRS sequences are highly characteristic of individual species. NCMRS that are represented by thousands of copies per genome are easy to detect. A PCR assay for a species-specific NCMRS from the crab *Sesarma reticulatum* was capable of detecting an amount of genomic DNA less than would be present in a single cell (Bilodeau et al. 1999). Variation among copies of a NCMRS within each individual's genome provides a "broad target" for hybridization probes and PCR primers; there is thus less concern that an individual will lack a detectable target sequence. The downside of the use of NCMRS for species identification is that they must be developed by an extremely *ad hoc* procedure. First they must be isolated from genomic libraries (although because they are repetitive, the libraries can be very small), and then

their specificity must be validated by extensive tests against DNA from other species (Bilodeau et al. 1999; MaKinster et al. 1999).

Problems with thresholds

In principle, we might not expect that a threshold of sequence divergence for *COI* or any other gene would delineate species. Pairs of sister species vary in how long they have been separated (Stanley 1998) and rates of mitochondrial sequence evolution range several-fold or more among metazoan lineages (Vawter and Brown 1986; Martin and Palumbi 1993, Shearer et al. 2002). Nevertheless, clusters of divergence times for recent species-pairs might be expected from synchronization of processes or events that promote speciation, such as changes in sea level (McManus 1985) or the rise of the Central American isthmus (Collins et al. 1996). The question of whether there are thresholds of divergence that work reasonably well for species identification is debated. It has been reported that thresholds work for Australian fishes (Ward et al. 2005), crustaceans (Lefebure et al. 2006), North American birds (Hebert et al. 2004a) tropical lepidopterans (Hajibabaei et al. 2006) and cave-dwelling spiders (Paquin and Hedin 2004). However, studies that have included closely related sister-species have found thresholds are not always reliable. In a study that included *COI-5* sequences from over 2,000 cowries (the gastropod family Cypraeidae) that represented 263 taxa, or over 93% of the species in the family, considerable overlap was found between intra- and interspecific divergence (Meyer and Paulay 2005). Moritz and Cicero (2004) re-examined the findings by Hebert et al. (2004a), who found *COI-5* divergence between species of North American birds ranges from 7.05 to 7.93% while divergence within species ranges from 0.27 to 0.43%. The wide zone of separation between these ranges implies that a threshold placed anywhere between 0.44 and 7.05% would unambiguously delineate species. These findings were contrasted with those of another study of North American birds (Johnson and Cicero 2004), in which divergence within species ranged from 0.0 to 8.2%; this range entirely spans both the intra- and interspecific ranges reported by Hebert et al. (2004a). Other counter examples come from a study that examined pairs of closely related species from seven families of insects (Cognato 2006). Intraspecific divergence in *COI-5* between these closely related species ranged from 0.04 to 26%, which broadly overlapped the 1.0–30.7% range of divergence found between species, among species-pairs, ranges overlapped in 28 of 62 comparisons.

The debate over whether or not thresholds of sequence divergence are a sound approach to species identification concerns what is meant by a species. In one sense, species

are simply labels of convenience, but in another important sense we intend these labels to correspond to coherent evolutionary units. From the latter sense, when two arthropod specimens key out to the same species but differ by 20–30% in their *COI-5* sequences, we can legitimately ask first whether they were correctly identified, and second, whether the taxonomy of the genus is in need of revision. The same questions could again be asked when specimens key out to different species, but differ by much less than 1% in *COI-5* divergence. We do not mean to imply that such cases can be dismissed as errors in taxonomy, but only that errors do occur and can produce an inherent bias against the resolution of thresholds. The finding that the apparent incidence of paraphyly (which can be due to errors in identification or taxonomy) is lower in better-studied groups suggests this bias is significant (Funk and Omland 2003). The challenge is therefore to distinguish the true exceptions from those due to error.

An ideal species-delimiting threshold would require that the range of sequence divergence within species does not overlap with the range between species. Much of the controversy in the DBC literature concerns whether or not this ideal is fulfilled, and there is a tendency to conclude that DBC should not be used if it is not. In reality, we should expect that as data accumulate for more species we will find additional cases of overlap. In these situations it will be necessary to adjust thresholds to control the probabilities of errors, as is done with other tests of statistical hypotheses. Failures to assign sequences to the correct species when divergence exceeds the threshold are Type I errors and assignments to incorrect species when divergence is below the threshold are Type II errors. Methods that estimate and control these statistical errors are needed to provide objectivity in the evaluation and application of DBC.

Alternatives to thresholds

It is a common practice to generate phylogenetic trees from *COI-5* sequences and examine the placement of sequences from unknown specimens on those trees. If the sequence of an unknown specimen is placed within a group of sequences from a single species, this suggests the specimen should be assigned to that species (Hebert et al. 2004a, b). These trees, typically generated from Kimura two-parameter distances and the Neighbor-Joining algorithm, are often used as an adjunct to threshold-based species assignments. They provide a graphical representation of all divergence values within and between species and can reveal potential problems such as paraphyly and cryptic species. Measures of tree support, such as the non-parametric bootstrap, can be used to assess the robustness of sequence clusters that correspond to species and higher taxa (Brower 2006).

Phylogenetic placement of *COI-5* sequences could become the preferred tool for species assignments, but thresholds have one important advantage. Threshold values can be validated for groups (i.e., genera) in which multiple sequences are available for every member species, and then applied more broadly to include cases in which some species are represented by single sequences and some not at all. In contrast, phylogenetic placement requires an effectively complete representation of gene genealogical relationships within and between all relevant species. Just how many sequences are needed from each species to achieve accurate phylogenetic placement is a difficult question; it depends on genealogical structure, which varies greatly among species. A minimum of three sequences per species has become a common goal for barcoding initiatives, although simple considerations suggest that such a small sample will often fail to represent every major branch in a mitochondrial genealogy.

It is unrealistic to expect that highly-detailed genealogies will soon be available for phylogenetic placement of *COI-5* sequences. However, samples of sequences from *COI-5* genealogies can be used as the basis for probabilistic assignments to species that allow for uncertainties in genealogical structure. Matz and Nielsen (2005) developed a likelihood ratio test for the hypothesis that a sequence from an unknown specimen is from the same population (i.e., species) as a sample of sequences from a known population. Their test was based on a coalescent model, which they also used to explore the sensitivity of Type I error rates to the number of sequences in the sample and the population genetic parameter θ (two times the product of female effective population size and mutation rate). Nielsen and Matz (2006) used coalescent models to develop a Bayesian method to assign a sequence to either of two possible species represented by multiple sequences, or to indicate that the sequence shouldn't be assigned to either. In theory, these coalescent methods could deliver the greatest accuracy and statistical rigor of any the current approaches to DBC species assignments. In practice, they are still at the "proof-of-concept" stage; they require too many assumptions and are too computationally demanding to be used routinely. However, they have been useful in drawing attention to some important considerations for DBC, such as the need for greater numbers of sequences for each species and the limitations of any method based on a single gene.

Summary and conclusions

Both the naming of species and the assignment of specimens to species are of fundamental importance in conservation. The threatened or endangered status of an individual

species can only be determined if the species itself is recognized and identifiable. Conservation policy tools, such as the Convention on International Trade in Endangered Species of Wild Fauna and Flora, and the US Endangered Species Act, are based on lists of threatened or endangered species and for better or for worse, conservation priorities are often based on such lists (Mace 2004). Conservation priorities for tropical reefs (Roberts et al. 2002) and evaluation of marine reserves (Jones et al. 2004) have been based on numbers of species listed for different regions or habitats. However, the high species richness (Reaka-Kudla 1997) and endemism (Roberts et al. 2002) of coral reef communities along with the remote locations of many reefs make it likely that a large proportion of coral reef-associated species remain undescribed. The systematics of reef-building corals provide a prime example of how molecular methods have helped to resolve otherwise intractable questions concerning species boundaries and relationships (reviewed in van Oppen and Gates 2006). Molecular approaches to specimen identification follow naturally from molecular systematics, and have the potential to dramatically improve the quantity and quality of the data upon which coral reef conservation science is based.

DBC differs from other approaches to DNA-based taxonomy in its specification of a single sequence, *COI-5*, as a common standard for identification of most metazoan species. DBC could become an important tool in coral reef conservation if suitable databases of *COI-5* sequences became available. Accurate assignment to species by DBC requires that a sequence from an unknown specimen can be compared to multiple sequences from others of its own species as well as from closely related species. This suggests that the number of completed genera in a barcode database could be a more useful measure of database functionality than either the total number of sequences or the total number of species represented. Good progress has been made in assembling a comprehensive database of *COI-5* sequences for all fishes, although it is still too far from completion to support identification of most species of reef-associated fishes.

DBC can solve a number of problems that are otherwise intractable. At the very least, it can help us with specimens that are difficult to identify and as a side-benefit, it can facilitate the discovery of new species. These basic uses of DBC should allow us to include more species in surveys of reef biodiversity and reduce the present bias towards large and conspicuous organisms. Beyond identification of individual specimens from *COI-5* sequences, there are other applications for which *COI-5* should be useful. Assemblages of meiofauna, epifauna, plankton, and in general organisms that are small, numerous and diverse can be investigated by bulk analysis of their “metagenomes”. DNA sequences can serve as targets for hybridization

probes in assays that identify and quantify these organisms in mixtures. However, the requirements for species-level accuracy with methods such as RT-PCR and DNA microarrays can be difficult to meet with *COI-5* sequences alone. These methods are not well suited for distinguishing species-level differences in sequences that are as long as *COI-5*, but are highly specific for target sequences that are a few tens of nucleotides in length. This length constraint, along with other requirements, make it desirable to consider a wider array of potential targets than just *COI-5*.

For groups in which taxonomy based on morphology is highly problematic, DNA sequences can be used to define MOTUs at levels that should correspond to biological species. *COI-5* sequences can be useful for this purpose, but multiple sequences including some from the nuclear genome are needed to avoid artifacts from incomplete lineage sorting, hybridization, sequence paralogy and other problems. For some groups, sequences other than *COI-5* have already become proven standards. Sequences other than *COI-5* will certainly be needed to characterize species of Porifera, Anthozoa and other groups in which mitochondrial sequences evolve too slowly to differentiate species. However, even for groups in which *COI-5* does not provide species-specificity, it can still be useful for assigning specimens to genera or families. For some purposes, this might be sufficient. For purposes that require more precision, a rough identification with *COI-5* could serve as a starting point beyond which sequences from additional loci could be used to provide species-level resolution.

The idea of DBC has generated a great deal of enthusiasm because it promises to address some major shortcomings of conventional taxonomy for field biologists (Janzen 2004). What is most impressive is the development of a shared, global infrastructure to collect and manage barcode reference standards and advance the state-of-the art in every way possible. The BOLD system in particular, is a model for scientific workflow and workgroup integration. Although essentially every use proposed for DBC had been invented before 2003, DBC brought them all as a group to the forefront.

From this review it should be clear that the simple appeal of *COI-5* as a universal standard is also the most serious limitation of DBC. For coral reef biologists, a taxonomic method that does not reliably work for stony corals, sea anemones, zoanthids, sea fans, gorgonians, black corals, or sponges might not seem very powerful. At its best, DBC data will always be less informative than data from multiple segregating loci. The term DBC is now starting to be used for sequences other than *COI-5* and methods other than PCR. These are encouraging signs that suggest the enthusiasm for the concept of DNA taxonomy can be broadened to include a plurality of approaches. Now that

we have the infrastructure, it should be easy to make room for these other approaches.

Acknowledgments We thank Bob Hanner for allowing us to analyze the progress of FISH-BOL and for providing us with updates on CBOL projects, James Albert and Emily Capuli for providing us with a current list of reef-associated fish species, Bob Vrijenhoek and two anonymous reviewers for helpful comments, and the National Science Foundation for support (OCE 0326383). We also thank the organizers and participants of the Coral Reef Barcoding and Environmental Sampling Workshop held at the Smithsonian Tropical Research Institute in 2005 for their stimulating discussions.

References

- Angelov S, Harb B, Kannan S, Khanna S, Kim J, Wang LS (2004) Genome identification and classification by short oligo arrays. In: Jonassen I, Kim J (eds) Algorithms in Bioinformatics, 4th International Workshop, WABI 2004, Bergen, Norway, September 17–21, 2004, Proceedings. Lecture Notes in Computer Science Vol 3240. Springer, Berlin, pp 400–411
- Avisé JC (1994) Molecular markers, natural history and evolution. Chapman and Hall, New York
- Avisé JC, Ball RM (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxf Surv Evol Biol* 7:45–67
- Baldwin BS, Black M, Sanjur O, Gustafson R, Lutz RA, Vrijenhoek RC (1996) A diagnostic molecular marker for zebra mussels (*Dreissena polymorpha*) and potentially co-occurring bivalves: Mitochondrial COI. *Mol Mar Biol Biotechnol* 5:9–14
- Barber P, Boyce SL (2006) Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. *Proc R Soc Lond B Biol Sci* 273:2053–2061
- Bellwood DR, Hughes TP, Folke C, Nystrom M (2004) Confronting the coral reef crisis. *Nature* 429:827–833
- Bhadury P, Austen MC, Bilton DT, Lamshead PJD, Rogers AD, Smerdon GR (2006) Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes. *Mar Ecol Prog Ser* 320:1–9
- Bilodeau AL, Lankford WS, Kim TJ, Felder DL, Neigel JE (1999) An ultrasensitive method for detection of single crab larvae (*Sesarma reticulatum*) using PCR amplification of a highly repetitive DNA sequence. *Mol Ecol* 8:683–684
- Blaxter ML (2004) The promise of a DNA taxonomy. *Proc R Soc Lond B Biol Sci* 359:669–679
- Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E (2005) Defining operational taxonomic units using DNA barcode data. *Proc R Soc Lond B Biol Sci* 360:1935–1943
- Borneman J (2001) Probe selection algorithms with applications in the analysis of microbial communities. *Bioinformatics* 17:1–9
- Brower AVZ (2006) Problems with DNA barcodes for species delimitation: ‘ten species’ of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperidae). *Systematics and Biodiversity* 4:127–132
- Caley MJ, Carr MH, Hixon MA, Hughes TP, Jones GP, Menge BA (1996) Recruitment and the local dynamics of open marine populations. *Annu Rev Ecol Syst* 27:477–500
- Call DR (2005) Challenges and opportunities for pathogen detection using DNA microarrays. *Crit Rev Microbiol* 31:91–99
- Cognato AI (2006) Standard percent DNA sequence difference for insects does not predict species boundaries. *J Econ Entomol* 99:1037–1045
- Collins LS, Budd AF, Coates AG (1996) Earliest evolution associated with closure of the Tropical American Seaway. *Proc Natl Acad Sci USA* 93:6069–6072
- Cote IM, Gill JA, Gardner TA, Watkinson AR (2005) Measuring coral reef decline through meta-analyses. *Proc R Soc Lond B Biol Sci* 360:385–395
- Cowen RK, Paris CB, Srinivasan A (2006) Scaling of connectivity in marine populations. *Science* 311:522–527
- DeLong EE (2005) Microbial community genomics in the ocean. *Nat Rev Microbiol* 3:459–469
- Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander EC, Rohwer F (2006) Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* 7:57
- Erpenbeck D, Hooper JNA, Worheide G (2006) CO1 phylogenies in diploblasts and the ‘Barcoding of Life’—are we sequencing a suboptimal partition? *Mol Ecol Notes* 6:550–553
- Floyd R, Abebe E, Papert A, Blaxter M (2002) Molecular barcodes for soil nematode identification. *Mol Ecol* 11:839–850
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3:294–299
- Froese R, Pauly D (2000) FishBase 2000: concepts, design and data sources. ICLARM, Los Baños
- Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: Frequency, causes, and consequences with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics* 34:397–423
- Gaines SD, Gaylord B, Largier JL (2003) Avoiding current oversights in marine reserve design. *Ecol Appl* 13:S32–S46
- Gardner TA, Cote IM, Gill JA, Grant A, Watkinson AR (2003) Long-term region-wide declines in Caribbean corals. *Science* 301:958–960
- Gaylord B, Gaines SD, Siegel DA, Carr MH (2005) Marine reserves exploit population structure and life history in potentially improving fisheries yields. *Ecol Appl* 15:2180–2191
- Georgiev GP, Kramerov DA, Ryskov AP, Skryabin KG, Lukanidin EM (1982) Dispersed repetitive sequences in eukaryotic genomes and their possible biological significance. *Cold Spring Harbor Symp Quant Biol* 47:1109–1121
- Goffredi SK, Jones WJ, Scholin CA, Marin R, Vrijenhoek RC (2006) Molecular detection of marine invertebrate larvae. *Mar Biotechnol* 8:149–160
- Gómez A, Wright PJ, Lunt DH, Cancino JM, Carvalho GR, Hughes RN (2007) Mating trials validate the use of DNA barcoding to reveal cryptic speciation of a marine bryozoan taxon. *Proc R Soc Lond B Biol Sci* 274:199–207
- Govindarajan AF, Halanych KK, Cunningham CW (2005) Mitochondrial evolution and phylogeography in the hydrozoan *Obelia geniculata* (Cnidaria). *Mar Biol* 146:213–222
- Grantham BA, Eckert GL, Shanks AL (2003) Dispersal potential of marine invertebrates in diverse habitats. *Ecol Appl* 13:S108–S116
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc Natl Acad Sci USA* 103:968–971
- Hanner R (2005) Proposed Standards for BARCODE Records in INS-SC (BRIs) Database Working Group, Consortium for the Barcode of Life, <http://www.barcoding.si.edu/>
- Hare MP, Palumbi SR, Butman CA (2000) Single-step species identification of bivalve larvae using multiplex polymerase chain reaction. *Mar Biol* 137:953–961
- Hebert PDN, Ratnasingham S, deWaard JR (2003a) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond B Biol Sci* 270:S96–S99
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003b) Biological identifications through DNA barcodes. *Proc R Soc Lond B Biol Sci* 270:313–321
- Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM (2004a) Identification of birds through DNA barcodes. *PLoS Biology* 2:1657–1663

- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004b) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci USA* 101:14812–14817
- Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. *Genome Res* 6:986–994
- Hellberg ME (2006) No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evol Biol* 6:24
- Holland BS, Dawson MN, Crow GL, Hofmann DK (2004) Global phylogeography of *Cassiopea* (Scyphozoa : Rhizostomeae): molecular evidence for cryptic species and multiple invasions of the Hawaiian Islands. *Mar Biol* 145:1119–1128
- Janzen DH (2004) Now is the time. *Proc R Soc Lond B Biol Sci* 359:731–732
- Johnson NK, Cicero C (2004) New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. *Evolution* 58:1122–1130
- Jones GP, McCormick MI, Srinivasan M, Eagle JV (2004) Coral decline threatens fish biodiversity in marine reserves. *Proc Natl Acad Sci USA* 101:8251–8253
- Knowlton N (1993) Sibling Species in the Sea. *Annu Rev Ecol Syst* 24:189–216
- Knowlton N (2001) The future of coral reefs. *Proc Natl Acad Sci USA* 98:5419–5425
- Lazoski C, Sole-Cava AM, Boury-Esnault N, Klautau M, Russo CAM (2001) Cryptic speciation in a high gene flow scenario in the oviparous marine sponge *Chondrosia reniformis*. *Mar Biol* 139:421–429
- Lefebvre T, Douady CJ, Gouy M, Gibert J (2006) Relationship between morphological taxonomy and molecular divergence within Crustacea: Proposal of a molecular threshold to help species delimitation. *Mol Phylogenet Evol* 40:435–447
- Lipscomb D, Platnick N, Wheeler Q (2003) The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends Ecol Evol* 18:65–66
- Livak KJ (1999) Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal Biomol Eng* 14:143–149
- Livak KJ, Flood SJA, Marmaro J, Giusti W, Deetz K (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods & Applications* 4:357–362
- Mace GM (2004) The role of taxonomy in species conservation. *Philos Trans R Soc Lond B* 359:711–719
- MaKinster JG, Roberts JE, Felder DL, Chlan CA, Bourdreaux M, Bilo-deau AL, Neigel JE (1999) PCR amplification of a middle repetitive element detects larval stone crabs (Crustacea: Decapoda: Menippidae) in estuarine plankton samples. *Mar Ecol Prog Ser* 188:161–168
- Markmann M, Tautz D (2005) Reverse taxonomy: an approach towards determining the diversity of meiobenthic organisms based on ribosomal RNA signature sequences. *Proc R Soc Lond B Biol Sci* 360:1917–1924
- Martin AP, Palumbi SR (1993) Body size, metabolic-rate, generation time, and the molecular clock. *Proc Natl Acad Sci USA* 90:4087–4091
- Matz MV, Nielsen R (2005) A likelihood ratio test for species membership based on DNA sequence data. *Proc R Soc Lond B Biol Sci* 360:1969–1974
- McClanahan TR (2002) The near future of coral reefs. *Environ Conserv* 29:460–483
- McFadden CS, Tullis I, Hutchinson MB, Winner K (2000) Rates of evolution of cnidarian mitochondrial genes. *Am Zool* 40:1124–1124
- McManus J (1985) Marine speciation, tectonics and sea-level changes in southeast Asia. *Proc 5th Int Coral Reef Congr* 4:133–138
- Meyer CP, Paulay G (2005) DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biology* 3:2229–2238
- Mikkelsen PM, Cracraft J (2001) Marine biodiversity and the need for systematic inventories. *Bull Mar Sci* 69:525–534
- Moritz C, Cicero C (2004) DNA barcoding: Promise and pitfalls. *PLoS Biology* 2:1529–1531
- Naviaux RK, Good B, McPherson JD, Steffen DL, Markusic D, Ransom B, Corbeil J (2005) Sand DNA - a genetic library of life at the water's edge. *Mar Ecol Prog Ser* 301:9–22
- Neigel JE (2003) Species-area relationships and marine conservation. *Ecol Appl* 13:S138–S145
- Neigel JE, Avise JC (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: Karlin S, Nevo E (eds) *Evolutionary Processes and Theory*. Academic Press, New York, pp 515–534
- Nielsen R, Matz M (2006) Statistical approaches for DNA barcoding. *Syst Biol* 55:162–169
- O'Dor R, Gallardo VA (2005) How to census marine life: ocean realm field projects. *Sci Mar* 69:181–199
- Pandolfi JM, Brabury RH, Sala E, Hughes TP, Bjorndal KA, Cooke RG, McArdle D, McClenahan L, Newman MJH, Paredes G, Warner RR, Jackson JBC (2003) Global trajectories of the long-term decline of coral reef ecosystems. *Science* 301:955–958
- Paquin P, Hedin M (2004) The power and perils of 'molecular taxonomy': a case study of eyeless and endangered Cicurina (Araneae : Dictynidae) from Texas caves. *Mol Ecol* 13:3239–3255
- Pegg G, Sinclair B, Briskey L, Aspden W (2006) MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia. *Sci Mar* 70S2:7–12
- Powers T (2004) Nematode molecular diagnostics: From bands to barcodes. *Annu Rev Phytopathol* 42:367–383
- Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data Systems. *Mol Ecol Notes* 7:355–364
- Reaka-Kudla ML (1997) The global biodiversity of coral reefs: A comparison with rain forests. In: Reaka-Kudla ML, Wilson DE, Wilson EO (eds) *Biodiversity II: understanding and protecting our biological resources*. Joseph Henry Press, Washington, pp 83–108
- Reaka-Kudla ML (2000) The evolution of endemism in insular Pacific faunas: coral-dwelling stomatopods. *Journal of Crustacean Biology* 20:56–70
- Relogio A, Schwager C, Richter A, Ansoerge W, Valcarcel J (2002) Optimization of oligonucleotide-based DNA microarrays. *Nucl Acids Res* 30:e51
- Richardson D, Vanwye J, Exum A, Cowen R, Crawford D (2006) High-throughput species identification: from DNA isolation to bioinformatics. *Mol Ecol Notes* "in press"
- Robba L, Russell SJ, Barker GL, Brodie J (2006) Assessing the use of the mitochondrial cox1 marker for use in DNA barcoding of red algae (Rhodophyta). *Am J Bot* 93:1101–1108
- Roberts CM (1997) Connectivity and management of Caribbean coral reefs. *Science* 278:1454–1457
- Roberts CM, McClean CJ, Veron JEN, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding M, Wells F, Vynne C, Werner TB (2002) Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* 295:1280–1284
- Romano SL, Palumbi SR (1996) Evolution of scleractinian corals inferred from molecular systematics. *Science* 271:640–642
- Rubinoff D (2006) DNA barcoding evolves into the familiar. *Conserv Biol* 20:1548–1549
- Saunders GW (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Proc R Soc Lond B Biol Sci* 360:1879–1888
- Schander C, Willassen E (2005) What can biological barcoding do for marine biology?. *Mar Biol Res* 1:79–83
- Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals

- and conceals about modes of speciation in Hawaiian crickets. *Proc Natl Acad Sci USA* 99:16122–16127
- Shearer TL, Coffroth MA (2006) Genetic identification of Caribbean scleractinian coral recruits at the Flower Garden Banks and the Florida Keys. *Mar Ecol Prog Ser* 306:133–142
- Shearer TL, Van Oppen MJH, Romano SL, Worheide G (2002) Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Mol Ecol* 11:2475–2487
- Smith PJ, McVeagh SM, Allain V, Sanchez C (2005) DNA identification of gut contents of large pelagic fishes. *J Fish Biol* 67:1178–1183
- St Mary CM, Osenberg CW, Frazer TK, Lindberg WJ (2000) Stage structure, density dependence and the efficacy of marine reserves. *Bull Mar Sci* 66:675–690
- Stanley SM (1998) *Macroevolution, Pattern and Process*. Johns Hopkins University Press, Baltimore
- Summerbell RC, Levesque CA, Seifert KA, Bovers M, Fell JW, Diaz MR, Boekhout T, de Hoog GS, Stalpers J, Crous PW (2005) Microcoding: the second step in DNA barcoding. *Proc R Soc Lond B Biol Sci* 360:1897–1903
- Tringe SG, Rubin EM (2005) Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* 6:805–814
- Vadopalas B, Bouma JV, Jackels CR, Friedman CS (2006) Application of real-time PCR for simultaneous identification and quantification of larval abalone. *J Exp Mar Biol Ecol* 334:219–228
- van Oppen MJH, Gates RD (2006) Conservation genetics and the resilience of reef-building corals. *Mol Ecol* 15:3863–3883
- Vawter L, Brown WM (1986) Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. *Science* 234:194–196
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu DY, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74
- Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Proc R Soc Lond B Biol Sci* 360:1847–1857
- Watanabe S, Minegishi Y, Yoshinaga T, Aoyama J, Tsukamoto K (2004) A quick method for species identification of Japanese eel (*Anguilla japonica*) using real-time PCR: An onboard application for use during sampling surveys. *Mar Biotechnol* 6:566–574
- Webb KE, Barnes DKA, Clark MS, Bowden DA (2006) DNA barcoding: A molecular tool to identify Antarctic marine larvae. *Deep-Sea Res Part II* 53:1053–1060
- Wheeler QD (2005) Losing the plot: DNA “barcodes” and taxonomy. *Cladistics* 21:405–407
- Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20:47–55
- Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Syst Biol* :54
- Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 84:9054–9058