The perils and merits (or the Good, the Bad and the Ugly) of DNA barcoding of sponges – a controversial discussion

Antonio M. Solé-Cava[(1*)], Gert Wörheide[(2)]

[(1)] Molecular Biodiversity Laboratory - Genetics Department – Instituto de Biologia – Universidade Federal do Rio de Janeiro. Bloco A – CCS – Ilha do Fundão. 21941-490 – Rio de Janeiro, RJ, Brazil. sole@biologia.ufrj.br
[(2)] Abteilung Geobiologie, Geowissenschaftliches Zentrum, Georg-August Universität Göttingen, Goldschmidtstr. 3, D-37077 Göttingen, Germany. gert.woerheide@geo.uni-goettingen.de

Abstract: DNA barcodes are defined as signature sequences used to identify unknown specimens. They have been proposed as a means to quickly solve the taxonomical impediment, through an organised and highly structured effort, using the same part of the mitochondrial sub-unit I of the Cytochrome c oxidase gene to identify all species of the planet. There has been much debate about the uses and misuses of DNA for taxonomy, and the radical proposition of DNA barcodes has heated the debate. In this paper we present two contrasting views of how DNA barcodes may or may not help sponge taxonomy.

Keywords: Barcodes, critical analysis, cytochrome oxidase, molecular systematics, Porifera

Introduction

"…a classification founded on any single character, however important that may be, has always failed"

Charles Darwin, 1859

The use of genetic markers to detect cryptic species and formulate phylogenetic hypotheses has revolutionised systematics and taxonomy in the last thirty years. From the early studies with allozymes (reviewed in Thorpe and Solé-Cava 1994) to the recent analyses of DNA sequences (reviewed in Avise 2004), molecular systematics have mostly corroborated classic taxonomy. However, the use of molecular markers has also challenged many long-held beliefs in taxonomy, such as the cosmopolitanism of many marine invertebrate species (Klautau *et al.* 1999, Knowlton 2000), the closer relationship of Nematoda to Arthropoda (forming the Ecdysozoa Aguinaldo *et al.* 1997) than to other worms (Halanych 2004, Mallatt and Giribet 2006) or the recent hypothesis about the phylogenetic position of Placozoa (Dellaporta *et al.* 2006). Over 20 years have passed since the first paper on molecular systematics of sponges was published (Solé-Cava and Thorpe 1986), and much progress has been made in technical and analytical approaches, which led to amazing discoveries, like the close affinity between some chondrosids and aplysinids (Borchiellini *et al.* 2004, Nichols 2005), the polyphyletism of *Axinella* (whose species seem to be scattered among different orders; Borchiellini *et al.* 2004) and the deconstruction of the Ceractinomorpha and Tetractinomorpha sub-classes of Demospongiae (Borchiellini *et al.* 2001, Boury-Esnault 2006). Clearly, taxonomy and systematics have benefited immensely from these new approaches, and will continue to do so.

The continuous advances in DNA sequencing technology have recently led to the proposition of using short (about 650 bp) mitochondrial DNA sequences (more specifically, of the Cytochrome Oxidase c subunit I gene, CO1 or cox1) to identify all living species (Hebert *et al.* 2003a). Those sequences would ideally function as species-specific signature sequences (so-called "DNA barcodes"), which would allow quick (in a few minutes) and unambiguous identification of any organism straight in the field. The proposed development and use of very small and cheap hand-held CO1 sequencers would obviate the need for field guides or taxonomists to identify samples at some point in the future (Hebert *et al.* 2003a). Obviously, such ambitious claims have achieved much attention and publicity (and some funding), to what its proponents would like to turn into something like the Human Genome project.

The essence of the Consortium Barcodes of Life (CBoL) initiative has produced a heated debate. In this paper we discuss the merits (good) and perils (bad and ugly) aspects of DNA barcodes applied to sponge taxonomy. Rather than produce positive or negative conclusions about the utility of DNA barcodes for sponges, we expect to foment the discussion, and help the readers see both sides of this contentious issue. Since this paper reports on the debate between the two authors during the Buzios Sponge Symposium, each author will present his point of view on the subject. AMSC will present his personal view on the bad and the ugly aspects of DNA barcodes in general as well as a more specific view for sponges, and GW will present why he believes DNA barcodes (or better a DNA-assisted taxonomy) might aid sponge taxonomists in species

description and discovery. But first we must define what DNA barcodes are and what they are not.

What are DNA barcodes?

A "DNA barcode" is a DNA signature sequence that allows the **identification** of a specimen to a known species. The stated goal of the Consortium for the Barcodes of Life (CBoL) is that "anyone, anywhere, anytime be able to identify quickly and accurately the species of a specimen whatever its condition" (http://phe.rockefeller.edu/barcode/). To be useful, that sequence must be short, ubiquitous, and easily amplifiable and sequenceable using universal primers. It must be conserved enough to allow the identification of higher taxonomical ranks, like phyla down to genera, variable enough to distinguish even highly similar species, but not too variable, so that levels of intraspecific variability do not add up too much noise to species identification.

The choice of a mitochondrial gene as the barcoding marker was based on three facts: 1) mitochondrial genes are relatively easy to amplify by PCR, because mitochondria are abundant in the cells, making DNA extraction straightforward even from degraded samples; 2) the mitochondrial genome is haploid, allowing for direct sequencing of PCR products without phasing alleles by e.g. cloning or SSCP as usually is the case for nuclear, diploid, genes and 3) levels of recombination on the mitochondrial genome are very low, reducing problems of paralogy. The mitochondrial gene chosen was the Cytochrome Oxidase c subunit I, whose choice was based primarily on the availability of a large number of sequences in GenBank, the existence of universal primers that allow amplification of this fragment from most phyla (Folmer *et al.* 1994), and the claim that different taxonomic levels could be resolved with the marker in most organisms (Hebert *et al.* 2003b).

What DNA barcodes are not

There is much confusion between the concepts of DNA barcodes (a Technique) and Molecular Systematics (a Science). In order to be really useful for quick identification of specimens in the field, as said before, DNA barcode systems must be ubiquitous and use universal primers (precluding the need for prior identification of the phylum or class of the organism by the user). Therefore, as stated by the Barcodes of Life (BoL) consortium (http://barcoding.si.edu/), DNA barcoding is a technique to identify specimens. Nothing else. DNA barcoding is not phylogenetic analysis. If the system requires a critical phylogenetic analysis to work, it is not useful for barcoding purposes. DNA barcoding is not molecular systematics (Moritz and Cicero 2004). It cannot rely on extensive geographic sampling to decide the taxonomic status of each unknown specimen.

Perils and pitfalls: the Bad about DNA barcoding

The aim of the Barcodes of Life project is to develop a methodology that will allow the quick, on the fly, identification of specimens in the field, using portable sequencing devices ("barcoders") connected by satellite to large databanks (http://www.dnabarcoding.ca/barcode_initiative.php). This means that the methodology must be robust enough to be used by laypersons, under varied circumstances. Consequently, many problems commonly faced and properly handled by scientists working with molecular systematics, like contamination, paralogy and identification errors can become important sources of error. The problems summarised below have all been handled appropriately by molecular systematists through careful, case-by-case, analysis of the data. Therefore, for molecular systematics, those problems are just a source of noise/homoplasy, which can be made explicit and be solved through the critical analysis of the data. They become important and very bad, however, when the middle-man (the biologist) is removed from the process, and the communication is made directly by the DNA sequencer and the databanks, as envisaged by CBoL.

Contamination

Many organisms live in intimate associations with other species. In those cases, contamination can be an important problem for sequencing using universal (i.e. not phylum specific) primers. Sponges harbour enormous amounts of other organisms, and direct sequencing of PCR products will often produce misleading results (see Erpenbeck *et al.* 2002 for a good analysis of this problem in sponges). This can be circumvented by scientists with a careful analysis of the produced sequences using phylogenetic methods, but it may be an important problem if a "barcoder" is to be trusted by people in the field to identify single specimens (Hurst and Jiggins 2005).

Paralogy

One of the advantages of using CO1 was that recombination was rare, and the haploid nature of mtDNA made it easy to assume homology of the analysed sequences. However, copies of parts of the mitochondrial genome are often found in the nuclear genome (Mourier *et al.* 2001). In most cases, direct sequencing of PCR products of mitochondrial genes will produce the true mitochondrial sequence. However, sometimes the nuclear, pseudo-mitochondrial, copies will be preferentially amplified and sequenced instead of the true mtDNA (Williams and Knowlton 2001, Thalmann *et al.* 2004). In those cases, the produced sequences will be paralogous to those present in the databases. If their transfer to the nucleus is old enough, those sequences will have diverged over the threshold of 2.5% divergence used by the CBoL to exclude identification to known species, and the result will be the wrong identification of the samples. Another source of paralogy is incomplete lineage sorting (Wahlberg *et al.* 2003). Using a single gene sequence to identify species will miss much of their evolutionary history, and taxa recently diverged may all too easily be overlooked (Choat 2006).

Horizontal gene transfer and introgression

In plants and protists, horizontal gene transfer can be a relatively common phenomenon (Bergthorsson *et al.* 2003), while in animals it is considered to be rare (Kurland *et al.* 2003). We do not know how common horizontal gene transfer may be in sponges, but there are some evidences indicating

that this may have happened in *Tetilla* (Rot *et al.* 2006). Another source of polyphyletism on mtDNA is introgression, a phenomenon that is not uncommon in animals (Moritz 1987, Quesada *et al.* 1995). When those processes occur, an immediate result is that mitochondrial gene trees will not be adequate representations of the species´ phylogenies. Nuclear DNA determines about 100% of the phenotype of each organism, the way it looks and adapts to the environment. An individual from one species that has a mitochondrial DNA from another will still belong to the former species, but it would be wrongly identified, if we used the CO1 sequence as the sole parameter to identify it, as the one whence its mtDNA came from.

Identification errors in the database

Any database is only as good as the data put in it. GenBank is riddled with errors, which are often dismissed by many authors using their data for their own research. These errors include sequencing errors (Karlin *et al.* 2001, Foster 2003) but, more importantly, identification errors. For example, an *ad hoc* identification analysis of fungi species whose sequences had been deposited in GenBank revealed that over 20% of them had been wrongly identified (Bridge *et al.* 2003). This problem has been efficiently handled by CBoL through the establishment of quality standards for the submitted sequences, and the requirement of voucher museum specimens for each sequence entered into their database. However, the sheer volume of specimens deposited into the *museii* will inevitably mean that most specimens will not have their taxonomic identification verified after they have received their first name. Once the name has been tied to the sequence in the database, the error may be perpetuated in subsequent identifications, leading to a cascade of taxonomic errors.

Most results presented by barcoding advocates are of groups with well resolved taxonomy, where the system is more likely to work well. However, a recent, large-scale (over 2,000 individuals belonging to 263 taxa) evaluation of the barcoding approach to a marine invertebrate group (Gastropoda) found that when no representatives were present in the database (simulating what would happen when using barcodes to unveil new species), barcodes failed to identify species recognised by taxonomy over 20% of the time (Meyer and Paulay 2005). Errors included the lumping of different species as single entities, and considering conspecific specimens as belonging to different species (Meyer and Paulay 2005). This lack of correlation between identification by barcodes and by conventional taxonomy may, in fact, indicate that conventional taxonomy is wrong, and that levels of paraphyly and polyphyly observed all resulted from oversplitting and overlumping real biological species (Funk and Omland 2003, Meyer and Paulay 2005). However, even if 100% of the mismatches between species identified by taxonomists and by barcodes were due to taxonomical errors, this would still be a major drawback to the BoL initiative, since it would mean that even the initial database, built on species identified by experts in the field, would be liable to be wrong. Consequently, unless only holotypes were used for building the database, sequences could not be reliably attached to species names. For example, we know, now, that *Chondrilla nucula*, formerly considered to be a cosmopolitan species is, in fact, a species complex (Klautau *et al.* 1999, Usher *et al.* 2004). If that information was not available, during the building of the CBoL database of "known sponge species" what sequence would definitely represent *C. nucula* would depend on where the sample had been collected. It could be argued that, since all *Chondrilla* specimens from the Mediterranean analysed to date formed a monophyletic, low divergent cluster, the sequence from a Mediterranean specimen (the type locality of *C. nucula*) would adequately represent that species. But what would happen, then, with the "cosmopolitan" *Oscarella lobularis,* that aggregates two sibling species (Boury-Esnault *et al.* 1992, Loukaci *et al.* 2004) within the Mediterranean, where it was originally described?

Reification of species

One of the things that made the CBoL so attractive was their clear aim and the promise of unambiguously identifying, in a short time, all species of the planet. To identify a specimen to a species, it is important, above all, to know what a species is. There is an enormous ongoing debate about what a species may be, with over 22 species definitions used by different authors (Mallet 1996). CBoL does not try to define what a species is. They follow the pragmatic approach of verifying how much divergence in CO1 sequence exists between species acknowledged as different by taxonomists, and use the average divergence as a rule of thumb threshold above which specimens are considered to belong to different species. The currently accepted threshold for the CBoL consortium is a p distance of 0.025 (Hebert *et al.* 2003b). This means that, if the sequence of an unknown specimen is less than 2.5% divergent from a sequence present in the database, it will be identified as belonging to that species. A species, then, is reified by barcoders as a group of organisms that is over 2.5% different from any other groups. It is as simple as that. No doubts, no grey zones. That is the advantage of relying on a single character to identify species. However attractive that can be to ecologists, pharmaceutical companies or other users of taxonomic identifications, it is at least naïve, and at worst very dangerous. Anyone with some experience in taxonomy knows how this simplistic approach to identification is prone to error and can seriously go wrong.

The definition of the threshold value above which sequences are considered to belong to different species is also very important: setting a high threshold value means that false positives (= incorrectly deciding that a given sequence belongs to a different species from that in the database, which would correspond to a type I statistical error) will be more rare, but it will also mean that many different species will be considered as belonging to the same "genetic species" (= false negatives, which would correspond to type II statistical errors). Conversely, setting a low threshold value will increase the number of species likely to be detected, but it will also mean giving species status to what may be simply intraspecific varieties (see e.g. Bradley and Baker 2001). This question was addressed by the analysis of a large dataset by Meyer and Paulay (2005). They found that, using a carefully built phylogeny for all the 263 evolutionary significant

units (ESUs) sampled (through the use of molecular, morphological, ecological and reproductive data), if they chose a threshold value of 2% they would have between 11% and 20% (depending on number of individuals sampled per ESU) false positives (oversplitting), and 8% false negatives (overlumping). Increasing the threshold value to 3% would diminish the number of false positives considerably, to 2% to 3%, but it would also increase the proportion of false negatives to 16%. For the gastropods studied, Meyer and Paulay found that the threshold value that would produce the smallest number of false positives and false negatives would be 2.6%. However, even at that best threshold level total error rates were still as high as 17% (Meyer and Paulay 2005). A big problem with having to decide on threshold distance values as a basis for taking taxonomic decisions is that they are reductionist and bound to lead to artificial taxonomic entities. Taxonomy took a long time to incorporate the conceptual and analytical advances of cladistics and evolutionary biology. It would be sad, now, to return to phenetic, distance-based approaches (de Queiroz and Good 1997), abandoning critical character-based thinking. Furthermore, even if we were to accept the overall idea of a distance-based taxonomy, we would have to deal with the probably insurmountable problem of the inexistence of a precise evolutionary clock. Evolutionary rates can vary enormously not only between different genes or *taxa*, but also between different parts of the same genes (Stevens and Schofield 2003). There are analytical ways to deal with this problem (Aris-Brosou and Yang 2002, Thorne and Kishino 2005) but, again, they depend on a case-by-case analysis incompatible with the idea of automatic identification.

In sponges, there are few works using CO1 sequences for species-level taxonomy (Schroder *et al.* 2003, Duran *et al.* 2004, Nichols and Barnes 2005, Wörheide 2006), but it appears that the barcoding region of CO1 may be too conserved in sponges (Wörheide *et al.* 2004, Erpenbeck *et al.* 2006b). For example, several species of *Chondrilla* that could be identified through allozymes, ribosomal sequences and conventional taxonomy would all be clustered into a single species if we used a 2% CO1 divergence threshold to separate them (Zilberberg, personal communication). It is clear that at this point in time, we do not have sufficient amount of data at hand to decide on any threshold, should there be a universal one for sponges.

The Ugly

"Your work, Sir, is both new and good, but what's new is not good and what's good is not new"

Samuel Johnson, XVIII century
(cited by Will *et al.* 2005)

Bad as they may currently be, the technical problems of molecular barcodes may eventually be circumvented through technological developments and rigorous methodological approaches. However, there are more serious, deeper philosophical and political problems with the idea of molecular barcodes, particularly in relation to their ultimate end of identifying all species of the planet. The ugly aspects of the BoL initiative are related to philosophical issues, like the return to a 19th century typological thinking and the idea that scientific knowledge can be crystallised. But they also include serious political questions, like the brain-drain of young students and scientists from taxonomic work into the band-wagon of methodologically easy, well funded, highly publishable but scientifically empty barcode programs.

Brain-drain from classical systematics

The recognition of Science as a major source of National wealth resulted in increased levels of funding and a strong sustained growth of Graduate programs and Research Institutes. This has led to deep changes in the way scientists are funded and evaluated, with much weight being put into publication and impact factors, and public accountability of the work done. Those are welcomed changes, since Science largely relies on public funding, and it is natural that scientists should be evaluated in relation to the way they perform their work. However, because the evaluation system is still being constructed, there are large distortions, which favour scientists working in the more fashionable areas of Genetics and Biotechnology, in detriment of more slow producing fields like Zoology and Ecology. A consequence of this distortion has been a brain-drain, with graduate students and young scientists migrating from the slower to faster publishing fields. A program that puts even more emphasis on DNA for systematics will only make matters worse (Ebach and Holdrege 2005), to a point where we may irreversibly lose expertise, as the best sponge taxonomists will fail to train students interested on identifying and describing sponge species before they retire from their field. This problem is particularly serious because taxonomy expertise takes years to build.

Reductionism and pragmatism:

"Only through the ignorance of arrogance could one fail to learn the lessons of several centuries of comparative morphology. Single-character systems rarely work for even one truly diverse clade and never work for all clades"

Will *et al.* 2005

The barcoding of life project is not scientific. It has been successful in capturing the attention from the media and from some funding agencies because it makes huge promises and downplays the enormous difficulties associated with taxonomy. The CBoL site justifies the development and application of Barcodes saying that, in 250 years of existence, Zoologists have only described about 15% of animal diversity (www.dnabarcoding.ca/rationale.php). However, it is possible that taxonomic work has not been slow because zoologists haven't worked hard enough or because they lacked technology. Progress may have been slow because systematics is a complex science. If taxonomists were willing to make the same conceptual compromises as the CBoL proponents, by oversimplifying the complex task of delimiting biological species, they would have finished the description of biodiversity very quickly (and just as wrongly as CBoL would). If we are willing to accept that a species can be defined based on 650 bp of a mitochondrial gene (this represents less than 0.00000001% of the total genome

of species that have had their genome completely sequenced thus far), then we could, for example accept that sponge species could be described solely based on spicule types. We could even envisage a "Spiculometer" (fig 1), which could do an image recognition of spicule slides and, based on the different spicule combinations, give us a quick, reproducible and precise identification. The fact that that identification would be wrong (grouping, for example, most haliclonids as a single species), would be secondary to our objective of naming all of sponge biodiversity. By making false promises (like barcoding the whole biodiversity of the planet in ten years with 1 billion dollars; Hebert *et al.* 2003a) with very competent public relations and lobbying activists, CBoL has quickly attracted the attention of the media, which always welcome golden pill solutions to the problems of society.

Because evolutionary rates are not the same for the same genes across the taxonomical landscape, the relationships obtained from DNA sequence comparisons reflect only indirectly the evolutionary history of each group. Only through the sampling of several, different characters (including morphology, ecology and genetic data) can we begin to understand the limits between evolutionary lineages and, based on those, take informed decisions about species borders. A good example of how the use of multiple datasets has helped understanding taxonomic relationships in sponges can be found in Erpenbeck *et al.* (2006a).

The taxonomical impediment is real, but cutting corners in identifying species may only make matters worse, generating confusion and deviating resources from proper species descriptions into just discovering possible new biological entities. The rate at which systematists are describing new species is not limited by the number of new things to describe. The shelves in taxonomists' laboratories are already full of specimens waiting to be analysed, and the real limiting factor has been, and still is, the access to collections, bibliography and qualified personnel. In other words, taxonomists are already overburdened with new species to describe, and the bottleneck of species description may be made worse by the large amounts of putative new species found by barcoding. This unavoidable crisis may have a positive result to systematics, through the final realization that conventional taxonomy was, after all, what really needed support. However, this crisis may also have a different, less bright outcome to Systematics. Faced with huge numbers of species waiting to be named, it may become too tempting to simply replace formal taxonomic descriptions with some "barcode species" name (Baker and Bradley 2006) or "molecular Operational taxonomy unit" (Blaxter 2004) that will link specimens to gene sequences without further studies. We will have, then, a name (or a code) and a sequence, but will that be useful at all for biology? What is the difference between a conventional label in a collection jar, with data on time and local of collection and an arbitrary voucher number, and a similarly arbitrary number, linked to a DNA sequence? Without formal study by taxonomists how will those newly found "species" serve the biological community or society?

Crystalisation of knowledge

> **"***Moreover, the generation of COI profiles will provide a partial solution to the problem of the thinning ranks of morphological taxonomists by enabling a crystallization of their knowledge before they leave the field"*
>
> Hebert *et al.* 2003a

Barcodes for identifying things can be very good and useful, but only AFTER the taxonomic work has been done properly. For example, barcoding birds or whales can be very useful to control the illegal traffic of endangered species. However, the main argument used by the BoL initiative to justify its very large budget was the zoological impediment, which means that their ultimate promise is to identify the 85% species that have not been described by taxonomists. Barcoding things is essentially typological and, as Paul Hebert correctly puts it, could be a way to crystallise knowledge. It is true that stability in names is something important, but it should not be made arbitrarily (Knapp *et al.* 2004). There is a huge gap in taxonomy to be filled, and crystallising the current knowledge in a rapidly changing field, like sponge taxonomy, is bound to be a step backwards.

Merits and opportunities: the Good about DNA barcoding (or better a DNA-assisted taxonomy)

> *"Because what keeps on moving, is eternal"*
> *(Nam quod semper movetur, aeternum est)*
>
> M.T. Cicero (106-43 BCE): *De Re Publica* VI (27) (Scipio's Dream)

DNA barcoding provides exciting new means for quick species identification and discovery. The use of DNA signature sequences (aka DNA barcodes) in sponge taxonomy, supplementing conventional morphological characters, will revolutionize future ways in which we conduct taxonomic research to define and describe species. The fascinating idea of a universal DNA barcode for all organisms and a hand-held DNA barcode scanner, similar to the 'tricorder' in the now famous science fiction series *Star Trek* (www.startrek.com), that enables identification of any life form on our planet on the fly, might sound a bit too ambitious at present, but technological advances might enable such a system at some stage in the not too distant future. However, even nowadays, scientific research around DNA barcodes will provide multiple exciting opportunities for sponge research, e.g. to increase our knowledge and understanding about principles of molecular evolution, speciation processes, community ecology and species delimitation.

A DNA sequence-assisted taxonomic system for sponges, providing the means to quickly and unequivocally identify taxa, will significantly ease the workload of taxonomic service provided by the few experts in the field to pharmaceutical and ecological researchers, among others, who need to identify the taxa they encounter in their surveys or that show promising biochemical activities. DNA barcoding approaches will open up a new dimension and quality in biodiversity research and will become of vital importance for the survival and acknowledgement of sponge taxonomy and increase its

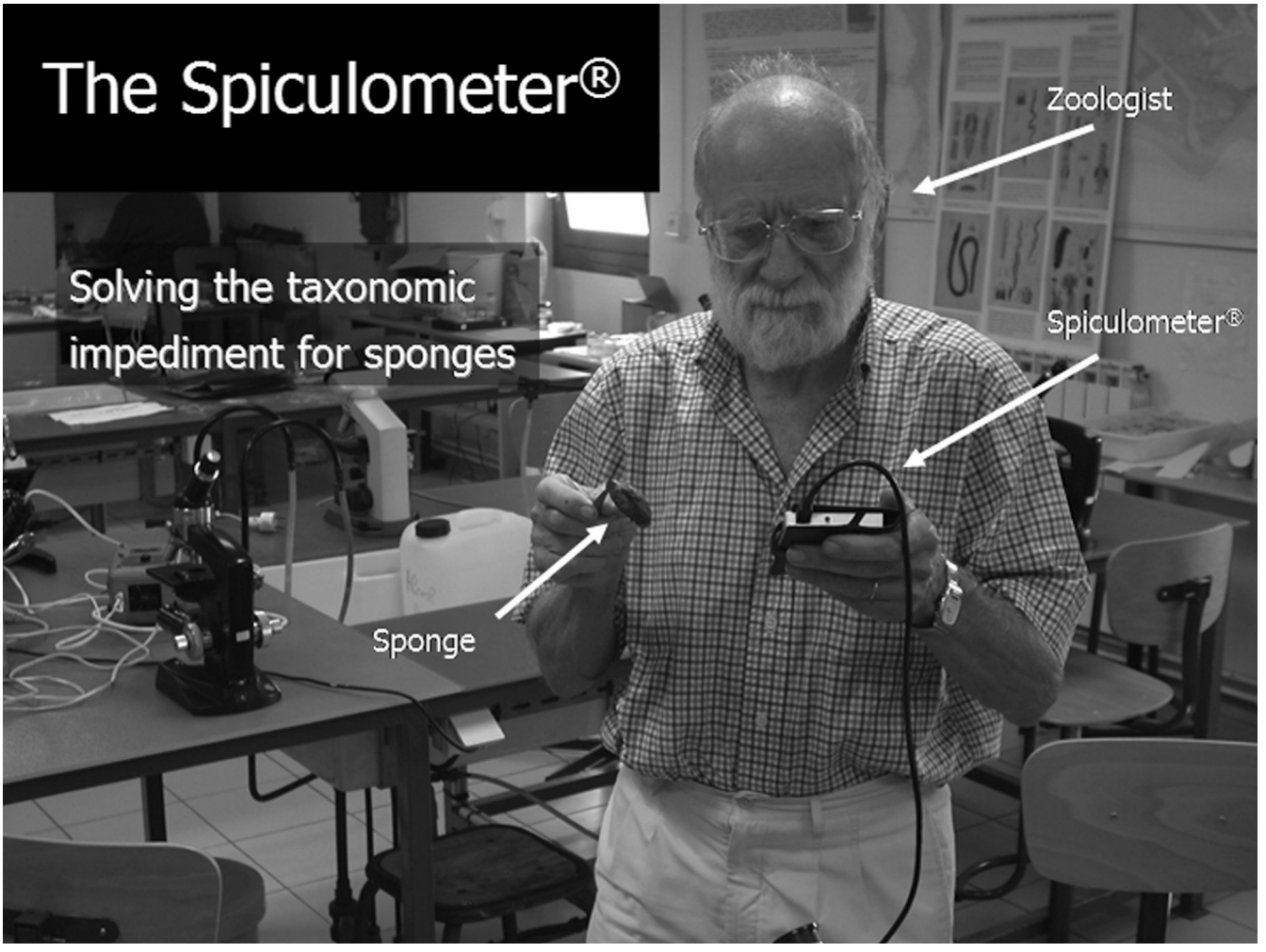

Fig. 1: The "Spiculometer". A parody on how morphological sponge taxonomy could join the fast lane.

reputation over the coming decades. It would be a serious disadvantage to disregard the opportunities that molecular (DNA barcoding) approaches bring to the field. We, as the community of scientists working on sponges, need to capitalize on (and not ignore) the new potential of scientific and financial opportunities and resources that the DNA barcoding movement creates and use them to our advantage, before others, who do not have the necessary taxonomic experience, do it. DNA barcoding resources will be vital to actually get the work done when attempting to identify taxa in large collections that exist in various museums around the world in a reasonable timeframe (i.e. before retirement and with a respectable publication list) – otherwise we will never create interest among young scientists to endeavour in sponge taxonomic research. A good example is the large collection of the Great Barrier Reef Seabed Biodiversity mapping project (www.reef.crc.org.au/resprogram/programC/seabed/index.htm), coordinated by the Australian Institute of Marine Science, which is attempting to document the sessile epibenthic fauna in the inter-reefal areas of the GBR. Thousands of samples have been collected, but without additional funding from DNA barcoding initiatives (or pharmaceutical companies for that matter), taxonomic work on such large collections will only proceed very, very slowly. Another yet unexplored aspect is the identification of the vast diversity of cryptic and/or small encrusting sponges (e.g. Richter *et al.* 2001), which then can be identified from tiny biopsies. This will open up a whole new dimension of sponge biodiversity, pivotal e.g. for our understanding of nutrient cycling and bentho-pelagic coupling in coral reefs (Lesser 2006).

Once funding for DNA barcoding is obtained, those new resources can, should and will be utilized to create also new positions for conventional taxonomic work and train a new generation of multidisciplinary taxonomists – ready for the challenges of an integrative taxonomy of the 21st century. Those new resources (monetary and human) will also create exciting new opportunities for international collaborations (see for example the Sponge Barcoding Project, introduced in this volume by Wörheide *et al.*) to tackle the many methodological and intellectual challenges that lie ahead. DNA barcoding of sponges will also change the society's appreciation of the taxonomic work done in "dusty" natural history museums and turn that into a picture of modern science that is methodologically up-to-date and ready for

future challenges. With world-wide declining funding for a not-so- terribly-sexy science like taxonomy, new resources from DNA barcoding might be pivotal for the survival of conventional taxonomy and will also enable research in natural history museums that goes beyond barcoding, i.e. do molecular systematics, phylogeography and molecular evolutionary research, to better understand the processes that shaped present-day biodiversity.

However, there are certainly some aspects of DNA barcoding that need careful consideration. First of all, especially in marine organisms harbouring numerous microbial and/or metazoan commensals or symbionts, contamination is definitely an issue. Designing sponge-specific primers for DNA-taxonomy markers should circumvent this issue, however, sequences obtained will have to be verified by phylogenetic tests in any case (this should be the usual procedure in any lab anyway). Paralogy, horizontal gene transfer and introgression on the other hand, can and will only be detected by phylogenetic tests once sufficient comparative data is accumulated – and we have to start doing so otherwise we will never get a deeper understanding of those issues. It is also clear that one mitochondrial marker will not be sufficient to establish a DNA taxonomic system for sponges that will aid species description and discovery; we will have to include at least one nuclear marker – an approach discussed in Wörheide *et al.* (2007). This can and will only be done together with taxonomic experts. Identification errors inherently occur in databases, but can be minimized by cross-verification by those taxonomic experts, an approach advocated in the Sponge Barcoding Project (SBP) (www.spongebarcoding.org; see also Wörheide *et al.* 2007).

A philosophical (and practical) problem certainly is the definition of what a (sponge) species actually is. (Sponge) taxonomists still mostly use fixed "diagnostic" characters (e.g. spicules and architecture) derived from comparative morphology to diagnose and separate species, not necessarily adhering to the biological species concept or any other than a typological one. While this has served reasonably well to catalogue diversity and is practical, it remains contentious whether it reflects the real biological diversity of sponges, considering that so-called 'cosmopolitan' sponge species, often only possessing a small number of morphological characters, are most likely a set of sibling (cryptic) species with different and divergent evolutionary histories, as uncovered by numerous genetic studies (e.g. Klautau *et al.* 1999). Existing morphological alpha-taxonomy of sponges is a rather artificial system solely based on morphological differences without considering evolutionary history and/or reproductive isolation. Furthermore, those morphological characters (spicules) used to define species differences have been shown to potentially vary with environmental conditions, i.e. the silica content of seawater has the potential to modulate the phenotypic expression of various spicule types (Maldonado *et al.* 1999). Quite disturbing.

The time has come to seriously consider additional characters, like DNA "signature" sequences, to corroborate taxonomic hypotheses. The argument that species identities will be reduced to single characters (a gene fragment) is not valid. Foremost, in a DNA sequence (how ever long it might be) each nucleotide position represents a separate character with four character states, in a protein sequence each amino acid represents one character, each with 20 character states (see textbooks like Page and Holmes 1998). So in e.g. the standard barcoding marker COI we have about 650 characters, some of them diagnostic, so it should be possible to quantify differences among "species" recognized by conventional taxonomy based on DNA sequences, preferably a combination of one mitochondrial and one nuclear marker. Inherent difficulties with species level analysis of DNA signature sequence are widely appreciated (e.g. Hickerson *et al.* 2006) and recent novel analytical approaches begin to tackle those problems at least in terrestrial organisms (Pons *et al.* 2006). Also the argument that a species can not be defined based on a single gene sequence is debatable, as it has been shown numerously that putative 'barrier genes' exist (most of them found in model species such as *Drosophila* spp.) that are associated with reproductive incompatibilities (see recent review by Noor and Feder 2006). Even if this makes only 0.00000001% of one species' genome, it certainly can make a difference. Further, the phenotype, solely recognized in conventional taxonomy, certainly is a reflection of the genotype, but only a reflection of a small fraction of this genotype.

However, in sponge taxonomy we are at the very beginning of establishing a system of DNA taxonomy and DNA barcodes that could aid in future species discovery and description, and currently we do not have data to decide on "thresholds" of genetic distances for species delimitation – nature certainly is not black and white in this regard and will not provide us with a simple solution, but we can only learn, develop and advance by gathering additional DNA sequence data and develop and apply novel analytical approaches to solve old problems where conventional taxonomy is at its limits. The DNA barcoding approach provides now novel ways to obtain funding to be able to do so and to rejuvenate taxonomy, and increase public awareness and appreciation of its new relevance.

The brain-drain from classical systematics can not be overcome by disregarding technological and analytical novelties and advances. It would be similar to argue against using email just because the guys in the post office might lose their jobs. The not-negligible brain-drain from classical taxonomy has deeper roots in the practice of how science is currently conducted and evaluated, as outlined above, and it would be naïve to believe that condemning DNA taxonomy/barcoding would solve this problem. Instead, we should take the opportunity and promote a taxonomic system that has a solid base both in conventional comparative morphology and DNA sequence analysis. We should not replace formal taxonomic (morphological) descriptions with DNA sequences (or worst barcodes), but try to unify both into one integrated system and use the potentially new funding sources from barcoding initiatives to support such a system.

"Where we go from there is a choice I leave to you"

(Neo, 1999 *The Matrix*)

The "gold nugget" of DNA barcoding

The new methodological approaches, intellectual challenges and potential funding resources provided by the DNA barcoding movement now puts us, as scientifically-conscious researchers, in the unique position to actually use to our advantage "the gold nugget dangling in front of us". We will be the ones who steer future sponge taxonomy in the right direction, get it out of the dust and make it ready for the multiple challenges of the 21st century. DNA barcoding will enable creation of new exciting positions for a new generation of integrative taxonomists, enable novel research directions and research collaborations. In a world of dwindling resources for taxonomy those new opportunities can not be dismissed and might prove pivotal for the future survival and appreciation of sponge taxonomy.

Concluding remarks

Molecular systematics is clearly a mature, growing science. It helps conventional taxonomy because it adds a new dimension to the analysis of species and their phylogenetic relationships. However, DNA barcodes are not synonymous with molecular systematics. Their sole stated aim is to provide a quick means to identify specimens to known species, in a similar way that supermarket barcodes serve for the identification of goods to the supermarket database. For that single end molecular barcodes are, indeed, very useful, and the error rates associated with the boundary between intraspecific and interspecific levels of CO1 sequence divergence (estimated at around 5% for the comparison of unknown samples with reference sequences of the species in the database; Meyer and Paulay 2005) are quite acceptable and possibly similar to error rates associated with the use of field guides and taxonomic keys. Problems with DNA barcodes become more prominent when they are used to identify new species, particularly in groups where conventional taxonomy is still far from being complete, as in the case of sponges. DNA barcodes can be seen as a welcome additional source of funds for *museii* and zoology departments, but they may also result in a brain-drain of young scientists away from conventional taxonomy which, ultimately, is what needs more support, since it represents the bottleneck in the description of the World's biodiversity. We believe that the only way the Barcodes of Life consortium will achieve its objective is through concurrent support of conventional taxonomy, and we propose that sponge barcoding projects should have that aim clearly stated, objectively allocating about 20% of all obtained resources specifically to the work of species descriptions by conventional taxonomists.

Ultimately, embracing or not the DNA barcodes program will be a personal decision, for which matters beyond scientific criteria may be important. With time, it will become clear if the promises of the Consortium Barcodes of Life will be fulfilled (original claim by Paul Hebert: all species barcoded until 2010 with a budget of about 1 billion dollars. Hebert *et al.* 2003a; revised targets: 2020 deadline and under 2 billion dollars budget. Paul Hebert, in Whitfield 2003). In any case, we believe that it is important that the choice be made in an informed, careful way, never losing sight of conventional taxonomy based on morphological characters.

Acknowledgements

We thank the organizers and the participants of the 7th Sponge Symposium for hosting the discussion on this important topic. AMSC was financed by the Brazilian (CNPq) and Rio de Janeiro (FAPERJ) Research Agencies. AMSC thanks Jean Vacelet for competently modelling the spiculometer photo, and the team from the Molecular Biodiversity Laboratory (LBDM) for suggestions and long debates. GW acknowledges funding from the German Research Foundation (DFG), from the sponsors of the symposium to fund his participation, and Dirk Erpenbeck as well as his group from the Molecular Marine Biodiversity Lab (MMBL) of the Dept. of Geobiology for fruitful discussions, Dirk Erpenbeck for the Cicero and Catherine Vogler for the Neo citation.

References

Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387: 489-493

Aris-Brosou S, Yang Z (2002) Effects of models of rate evolution on estimation of divergence dates with special reference to the Metazoan 18S ribosomal RNA phylogeny. *Syst Biol* 51: 703-714

Avise JC (2004) Molecular markers, natural history and evolution. Chapman & Hall, London

Baker RJ, Bradley RD (2006) Speciation in mammals and the genetic species concept. *J Mammal* 87: 643-662

Bergthorsson U, Adams KL, Thomason B, Palmer JD (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424: 197-201

Blaxter ML (2004) The promise of a DNA taxonomy. *Phil Trans Roy Soc Lond B* 359: 669-679

Borchiellini C, Chombard C, Manuel M, Alivon E, Vacelet J, Boury-Esnault N (2004) Molecular phylogeny of Demospongiae: implications for classification and scenarios of character evolution. *Mol Phylogenet Evol* 32: 823-837

Borchiellini C, Manuel M, Alivon E, Boury-Esnault N, Vacelet J, Le Parco Y (2001) Sponge paraphyly and the origin of Metazoa. *J Evol Biol* 14: 171-179

Boury-Esnault N (2006) Systematics and evolution of Demospongiae. *Can J Zool* 84: 205-224

Boury-Esnault N, Solé-Cava AM, Thorpe JP (1992) Genetic and cytological divergence between colour morphs of the Mediterranean sponge *Oscarella lobularis* Schmidt (Porifera, Demospongiae, Oscarellidae). *J Nat Hist* 26: 271-284

Bradley RD, Baker RJ (2001) A test of the genetic species concept. cytochrome b sequences and mammals. *J Mammal* 82: 960-973

Bridge PD, Roberts PJ, Spooner BM, Panchal G (2003) On the unreliability of published DNA sequences. *New Phytologist* 160: 43-48

Choat JH (2006) Phylogeography and reef fishes: bringing ecology back into the argument. *J Biogeogr* 33: 967-968

de Queiroz K, Good DA (1997) Phenetic clustering in biology: A critique. *Q Rev Biol* 72: 3-30

Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, Schierwater B (2006) Mitochondrial genome of *Trichoplax adhaerens* supports Placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci USA* 103(23): 8751-8756

Duran S, Pascual M, Turon X (2004) Low levels of genetic variation in mtDNA sequences over the western Mediterranean and Atlantic range of the sponge *Crambe crambe* (Poecilosclerida). *Mar Biol* 144: 31-35

Ebach MC, Holdrege C (2005) More taxonomy, not DNA barcoding. *BioScience* 55: 822-823

Erpenbeck D, Breeuwer JAJ, Parra-Velandia FJ, van Soest RWM (2006a) Speculation with spiculation? – Three independent gene fragments and biochemical characters versus morphology in demosponge higher classification. *Mol Phylogenet Evol* 38: 293-305

Erpenbeck D, Breeuwer JAJ, Velde HCvd, van Soest RWM (2002) Unravelling host and symbiont phylogenies of halichondrid sponges (Demospongiae, Porifera) using a mitochondrial marker. *Mar Biol* 141: 377-386

Erpenbeck D, Hooper JNA, Wörheide G (2006b) CO1 phylogenies in diploblasts and the 'Barcoding of Life'; are we sequencing a suboptimal partition? *Mol Ecol Notes* 6: 550-553

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

Foster P (2003) To Err is Human. *Ann Hum Genet* 67: 2-4

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Evol Syst* 34: 397-423

Halanych KM (2004) The new view of animal phylogeny. *Annu Rev Ecol Evol Syst* 35: 229-256

Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003a) Biological identifications through DNA barcodes. *Proc R Soc Lond B* 270: 313-321

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond B* 270: 96-99

Hickerson M, Meyer C, Moritz C (2006) DNA Barcoding will often fail to discover new animal species over broad parameter space. *Syst Biol* 55: 729-739

Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc R Soc Lond B* 272: 525-1534

Karlin S, Bergman A, Gentles AJ (2001) Genomics - Annotation of the *Drosophila* genome. *Nature* 411: 259-260

Klautau M, Russo CAM, Lazoski C, Boury-Esnault N, Thorpe JP, Solé-Cava AM (1999) Does cosmopolitanism result from overconservative systematics? A case study using the marine sponge *Chondrilla nucula*. *Evolution* 53: 1414-1422

Knapp S, Lamas G, Lughadha EN, Novarino G (2004) Stability or stasis in the names of organisms: the evolving codes of nomenclature. *Phil Trans R Soc Lond B* 359: 611-622

Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea. *In*: Solé-Cava AM, Russo CAM, Thorpe JP (eds). *Marine Genetics*. Kluwer Academic Publishers, Amsterdam, pp. 73-90

Kurland CG, Canback B, Berg OG (2003) Horizontal gene transfer: A critical view. *Proc Natl Acad. Sci USA* 100(17): 9658-9662

Lesser M, P. (2006) Benthic-pelagic coupling on coral reefs: Feeding and growth of Caribbean sponges. *J Exp Mar Biol Ecol* 328: 277-288

Loukaci A, Muricy G, Brouard JP, Guyot M, Vacelet J, Boury-Esnault N (2004) Chemical divergence between two sibling species of *Oscarella* (Porifera) from the Mediterranean Sea. *Biochem Systemat Ecol* 32: 893-899

Maldonado M, Carmona MC, Uriz MJ, Cruzado A (1999) Decline in Mesozoic reef-building sponges explained by silicon limitation. *Nature* 785-788

Mallatt J, Giribet G (2006) Further use of nearly complete, 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. *Mol Phylogenet Evol* 40: 772-794

Mallet J (1996) The genetics of biological diversity: from varieties to species. *In*: Gaston KJ (ed). *Biodiversity: a Biology of Numbers and Divergence*. Blackwell Science Inc., Malden, MA, USA, pp. 13-53

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

Moritz C, Cicero C (2004) DNA Barcoding: Promise and Pitfalls. *PLoS Biology* 2: 279-354

Moritz C, Dowling, T.E. and Brownn, W.M. (1987) Evolution of animal mtDNA: relevance for population biology and systematics. *Annu Rev Ecol Syst* 18: 269-292

Mourier T, Hansen AJ, Willerslev E, Arctander P (2001) The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol Biol Evol* 18: 1833-1837

Nichols SA (2005) An evaluation of support for order-level monophyly and interrelationships within the class Demospongiae using partial data from the large subunit rDNA and cytochrome oxidase subunit I. *Mol Phylogenet Evol* 34: 81-96

Nichols SA, Barnes PAG (2005) A molecular phylogeny and historical biogeography of the marine sponge genus *Placospongia* (Phylum Porifera) indicate low dispersal capabilities and widespread crypsis. *J Exp Mar Biol Ecol* 323: 1-15

Noor MAF, Feder JL (2006) Speciation genetics: evolving approaches. *Nat Rev Genet* 7: 851-861

Page RDM, Holmes EC (1998) *Molecular evolution: a phylogenetic approach*. Blackwell Science, Oxford

Pons J, Barraclough T, Gomez-Zurita J, Cardoso A, Duran D, Hazell S, Kamoun S, Sumlin W, Vogler A (2006) Sequence-Based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol* 55: 595-609

Quesada H, Wenne R, Skibinski DOF (1995) Differential introgression of mitochondrial DNA across species boundaries within the marine mussel genus *Mytilus*. *Proc R Soc Lond B* 262: 51-56

Richter C, Wunsch M, Rasheed M, Kotter I, Badran MI (2001) Endoscopic exploration of Red Sea coral reefs reveals dense populations of cavity-dwelling sponges. *Nature* 413: 726-730

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

Schröder HC, Efremova SM, Itskovich VB, Belikov S, Masuda Y, Krasko A, Müller IM, Müller WEG, Medina M, Weil E, Szmant AM (2003) Molecular phylogeny of the freshwater sponges in Lake Baikal. *J Zool Systemat Evol Res* 41: 80-86

Solé-Cava AM, Thorpe JP (1986) Genetic differentiation between morphotypes of the marine sponge *Suberites ficus* (Demospongiae: Hadromerida). *Mar Biol* 93: 247-253

Stevens JR, Schofield CJ (2003) Phylogenetics and sequence analysis - some problems for the unwary. *Trends Parasitol* 19: 582-588

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

Thorne JL, Kishino H (2005) Estimation of divergence times from molecular sequence data. *In*: Nielsen R (ed). *Statistical methods in molecular evolution*. Springer Verlag, London, pp. 235-256

Thorpe JP, Solé-Cava AM (1994) The use of allozyme electrophoresis in invertebrate systematics. *Zoolog Scripta* 23: 3-18

Usher KM, Sutton DC, Toze S, Kuo J, Fromont J (2004) Biogeography and phylogeny of *Chondrilla* species (Demospongiae) in Australia. *Mar Ecol Progr Ser* 270: 117-127

Wahlberg N, Oliveira R, Scott JA (2003) Phylogenetic relationships of *Phyciodes* butterfly species (Lepidoptera: Nymphalidae): complex mtDNA variation and species delimitations. *Syst Entomol* 28: 257-273

Whitfield J (2003) DNA barcodes catalogue animals. *News@nature* DOI:10.1038/news030512-7

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Syst Biol* 54: 844-851

Williams SL, Knowlton N (2001) Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*. *Mol Biol Evol* 18(8): 1484-1493

Wörheide G (2006) Low variation in partial cytochrome oxidase subunit I (COI) mitochondrial sequences in the coralline demosponge *Astrosclera willeyana* across the Indo-Pacific. *Mar Biol* 148: 907-912

Wörheide G, Erpenbeck D, Menke C (2007) The Sponge Barcoding Project: aiding in the identification and description of poriferan taxa. *In:* Custódio MR, Lôbo-Hajdu G, Hajdu E, Muricy G (eds). *Porifera research: biodiversity, innovation and sustainability*. Série Livros 28. Museu Nacional, Rio de Janeiro. pp. 123-128

Wörheide G, Fromont J, Solé-Cava AM (2004) Population genetics and phylogeography of sponges - a workshop synthesis. *In:* Pansini M, Pronzato R, Bavestrello G, Manconi R (eds). *Sponge science in the new millennium. Boll Mus Ist Biol Univ Genova* 68: 683-688