Molecular Ecology Notes (2005) doi: 10.1111/j.1471-8286.2005.01259.x

TECHNICAL NOTE

CO1 phylogenies in diploblasts and the 'Barcoding of Life' — are we sequencing a suboptimal partition?

D. ERPENBECK,*† J. N. A. HOOPER* and G. WÖRHEIDE†
*Biodiversity Program, Queensland Museum, South Brisbane, Queensland, Australia, †Department of Geobiology, Geoscience Centre Göttingen, Göttingen, Germany

Abstract

Partitions of the cytochrome oxidase subunit 1 (*CO1*) gene, especially the 5′ end, are frequently recruited to infer lower level phylogenies in animals. In diploblasts, mitochondrial genes were found to evolve in a slower rate than their bilaterian counterparts. Therefore, diploblast *CO1* gene trees repeatedly remained unresolved, which also raises doubts on the suitability of *CO1* for DNA barcoding in these animals. The complete mitochondrial genome sequences from Anthozoa and recently from Porifera allow us to compare the resolution power of the 5′ partition, which has also been proposed as the standard marker for DNA barcoding, with a less frequently used partition further downstream. We report on the finding of significantly different substitution patterns of the downstream partition opposed to the 5′ partition. We discuss the consequences and potential in the light of diploblast phylogenetic reconstruction and DNA barcoding.

Keywords: barcoding, *CO1*, corals, partition, phylogenetic resolution, sponges

Received 08 October 2005; revision accepted 22 November 2005

The mitochondrial cytochrome oxidase subunit 1 (*CO1*) gene is among the most popular markers in shallow molecular systematics — partitions of this gene are frequently employed for higher metazoan phylogenies. In addition, *CO1* is currently in the focus of special interest: its 5′ partition is used for the 'Barcoding of Life' initiative (Hebert *et al*. 2003; Stoeckle 2003). The nucleotide sequence of this particular ~640 nucleotide region (Folmer *et al*. 1994) shall be the unique identification code for all species to facilitate the correct determination of specimens and the discovery of new species (Moritz & Cicero 2004). Such standardized identification techniques are particularly important for the classification of sponges (Porifera), as paucity of complex morphological characters and a high degree of homoplasy leave considerable parts of sponge classification ambiguous and unresolved (see Hooper & Van Soest 2002).

Molecular phylogenetics of sponges is not as well-established as for other animal groups, which is surprising given their pivotal position at the base of the Metazoa and the high interest in their biochemical compounds (e.g. Faulkner 2000). Mitochodrial markers have only been recently introduced to sponge systematics (*CO2*, Wörheide *et al*. 2000), and complete sponge mitochondrial genomes were among the last to be published of all major metazoan phyla (Lavrov *et al*. 2005).

However, the power of resolution of mitochondrial DNA (mtDNA) fragments in sponges and other diploblasts does not seem to correspond to their higher metazoan counterparts. While mtDNA fragments are successfully used for lower level phylogenies in higher Metazoa, diploblast mtDNA appears to be too conservative. The evolutionary rate of diploblast mtDNA is found to be 10–20 times lower than in Bilateria (Shearer *et al*. 2002). *CO1* shows a suitable resolution at the family level (Erpenbeck *et al*. 2002; see also Nichols 2005 and Erpenbeck *et al*., in press), but insufficient variability below this level: population-level analyses of the western Mediterranean and Atlantic sponge *Crambe crambe* demonstrate just one single polymorphic position over a geographical range of several thousand kilometres (Duran *et al*. 2004). In addition, four species from two different Lake Baikal freshwater sponge genera (*Lubomirskia* and *Baikalospongia*) bear identical sequences (Schröder *et al*. 2003), and cannot be distinguished by their *CO1* fragment data. Congruent with this outcome, new *CO1* data (Wörheide, in press) fails to separate

Correspondence: D. Erpenbeck, Fax: +61-7-3846-1226; E-mail: derpenb@gwdg.de

© 2005 Blackwell Publishing Ltd

three sibling coralline sponge species within the *Astrosclera 'willeyana'* species-complex, which are otherwise clearly defined by morphological and ITS sequence data (Wörheide *et al*. 2002). A similar picture emerged from anthozoan cnidarians (Shearer *et al*. 2002). All these results are based entirely on the particular *CO1* 5′ partition used for DNA barcoding (Folmer *et al*. 1994). They reveal not only the limited suitability of *CO1* on species and genus level phylogenies in diploblasts — they also raise serious doubts on the use of *CO1* for the 'Barcoding of Life' for the lower metazoans. The most important goal of DNA barcoding, i.e. distinction between species, will clearly be missed without any detectable sequence variation between species and even genera.

Here, we report and discuss that *CO1* can be suitable for lower level phylogenies of sponges (and other diploblasts) and could provide sufficient information for the purposes of barcoding — but only if the right partition is utilized. Indeed, some phylogenetic *CO1* analyses of sponges have shown phylogenetic signals clearly below family level (Erpenbeck *et al*. 2002; Knowlton 2002; Erpenbeck *et al*. 2003). These analyses recruited a *CO1* partition of 453 base pairs spanning from the internal loop 3 (I3, Fig. 1) to the transmembrane domain 11 (M11) (subsequently referred to here as 'I3-M11'- partition). This I3-M11 partition is located 23 amino acids downstream of the 5′-'Folmer' partition, which originates near the amino-terminal and extends to the transmembrane domain 6 (M6, Fig. 1).

The recent publication of three complete demosponge mtDNA-genome sequences (*Tethya actinia* GenBank Accession no. NC006991, *Geodia neptuni* NC006990 and *Axinella corrugata* NC006894 (Lavrov *et al*. 2005; Lavrov & Lang 2005)) allows us to perform a direct comparison of the substitution behaviour of both *CO1* partitions. To analyse if our results are restricted to the Porifera or apply similarly to other diploblasts, we compared data from related anthozoans (*Sarcophyton glaucum* AF064823, *Montipora cactus* NC006902, *Anacropora matthai* NC006898, *Metridium senile* NC000933 and *Acropora tenuis* NC003522). The analyses of nucleotide and amino acid composition, frequency and substitution behaviour were performed with PAUP* 4.10b (Swofford 2002) and MEGA version 3.0 (Kumar *et al*. 2004).

We found the 5′ 'Folmer' and the I3-M11 — partitions congruent in nucleotide and amino acid variability (Table 1A–D). The third codon position, by far the most variable in the triplet (Table 1E–G), displays similarities in variability ratios (Table 1G) and relative substitution numbers (Table 1H) in both partitions. However, counting variable sites does not provide sufficient information on third codon position substitution patterns. Therefore, we plotted its transitions against the transversions, which revealed the following significant differences between the partitions: the transition/transversion (ti/tv) ratio in the I3-M11 fragment is clearly lower than in the 5′ partition; in the sponge taxon set, the 5′ partitions ratio exceeds by more than a third (1.01 against 0.63), and in the coral taxon set, by almost a fourth (1.26 against 0.98) (Table 1I), which is significantly different (t-test, $P < 0.05$). These different ti/tv ratios are unlikely to originate from amino acid sequence differences because the proportions of 0-, 2- and 4-fold degenerated codon sites are approximately equal (Table 1J–L).

Transversions are the dominant substitution types in the I3-M11 partition, which indicates a more progressive stage in character evolution in respect to the 5′ partition: apparently the I3-M11 partition substitutes at earlier stages of taxon evolution than the 5′ sites and provides informative

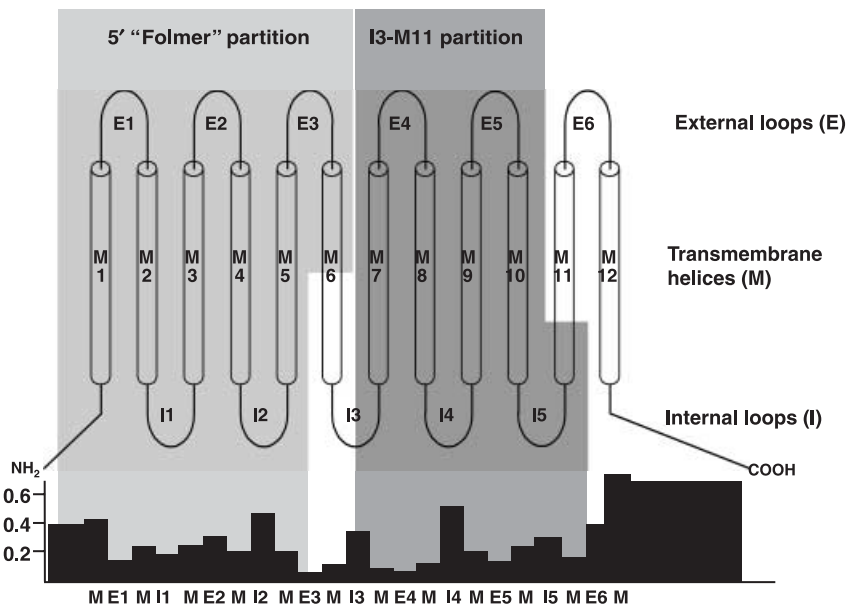

Fig. 1 Schematic view of *CO1*, which consists of approximately 520 amino acids forming 12 transmembrane helices (M), five internal (I) and six external (E) loops (cf. Lunt *et al*. 1996). The two partitions discussed are shaded in grey. The heights of the black columns below the domains represent their correspondent variability per amino acid site. Lunt *et al*. (1996) found the COOH-terminal significantly more variable than loops, helices, or the NH_2 terminal, which is congruent with the results we found for diploblasts.

Table 1 Comparison between the 5′ 'Folmer' and the I3-M11 partition for the Porifera and Anthozoa *CO1* data sets

	Taxon	Porifera		Anthozoa	
	Partition	5′	I3-M11	5′	I3-M11
A	nucleotide character sites	705	459	705	459
B	nucleotide variable sites ratio	0.28	0.25	0.32	0.29
C	aa character sites	235	153	235	153
D	aa variable sites ratio	0.12	0.12	0.15	0.12
E	variability 1st position	0.11	0.08	0.09	0.07
F	variability 2nd position	0.03	0.02	0.03	0.03
G	variability 3rd position	0.47	0.44	0.38	0.37
H	substitutions/taxon/3rd position	0.47	0.43	0.76	0.73
I	**3rd position ti/tv ratio**	**1.03**	**0.62**	**1.26**	**0.98**
J	0-fold degenerated sites ratio	0.62	0.64	0.62	0.64
K	twofold degenerated sites ratio	0.18	0.17	0.17	0.16
L	fourfold degenerated sites ratio	0.13	0.13	0.13	0.13

characters after shorter divergence times. The higher phylogenetic signal of the I3-M11 would remain undetected under the standard counting of polymorphic sites, which is mostly used to judge the suitability of a DNA partition for phylogenetic and other purposes. It demonstrates why the 5′ partition of *CO1* apparently is not suitable for lower level phylogenetics and barcoding of sponges (and other diploblasts) — and why a slightly longer fragment, including the I3-M1 partition apparently, is suitable.

For the successful application of *CO1* partitions in phylogenetics and DNA barcoding, we therefore suggest the inclusion of the I3-M11 partition. This measurement may be restricted to the application on sponge species but eventually be applied to other diploblastic and higher Metazoa. The 5′ partition alone may be useful to distinguish between sponge taxa down to genus level (satisfying the aims of many projects), but appending the I3-M11 partition might be needed to distinguish between species, without recruiting nucleotide information from a different gene. In addition, our analysis of amino acid variability in diploblast *CO1* domains (Fig. 1, lower part) indicates that the I3-M11 partition spans over a gene region which is distant from the indel-bearing COOH-terminal (Lunt *et al*. 1996) and not problematic for primer design.

Acknowledgements

D.E. acknowledges financial support of the European Union under a Marie-Curie outgoing fellowship (MOIF-CT-2004). G.W. acknowledges financial support by the German Research Foundation (DFG; Projects Wo896/3, 4, 5).

References

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). *Marine Biology*, **144**, 31–35.

Erpenbeck D, Breeuwer JAJ, van der Velde HC, Van Soest RWM (2002) Unravelling host and symbiont phylogenies of halichondrid sponges (Demospongiae, Porifera) using a mitochondrial marker. *Marine Biology*, **141**, 377–386.

Erpenbeck D, Knowlton AL, Talbot SL, Highsmith RC, Van Soest RWM (2004) A molecular comparison of Alaskan and North East Atlantic *Halichondria panicea* (Pallas 1766) (Porifera: Demospongiae) populations. *Bolettino dei Musei Istituti Biologici Universita Genova*, **68**, 319–325.

Erpenbeck D, Parra-Velandia FJ, Breeuwer JAJ, Van Soest RWM (in press) Speculation with spiculation? – Three independent gene fragments and biochemical characters versus morphology in demosponge higher classification. *Molecular Phylogenetics and Evolution*, in press. doi: 10.1016/j.ympev.2005.11.001

Faulkner DJ (2000) Highlights of marine natural products chemistry (1972–99). *Natural Product Reports*, **17**, 1–6.

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. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 313–332.

Hooper JNA, Van Soest RWM (2002) *Systema Porifera, A Guide to the Classification of Sponges*. KluwerAcademic/Plenum Publishers, New York.

Knowlton AL (2002) *Sponges Dominant in the Alaska Intertidal: Biology, Ecology, and Genetic Diversity*. University of Alaska, Fairbanks, Alaska.

Kumar S, Tamura K, Nei M (2004) MEGA 3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.

Lavrov DV, Forget L, Kelly M, Lang BF (2005) Mitochondrial genomes of two demosponges provide insights into an early stage of animal evolution. *Molecular Biology and Evolution*, **22**, 1231–1239.

Lavrov DV, Lang BF (2005) Transfer RNA gene recruitment in mitochondrial DNA. *Trends in Genetics*, **21**, 129–133.

Lunt DH, Zhang DX, Szymura JM, Hewitt GM (1996) The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Molecular Biology*, **5**, 153–165.

Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *Public Library of Science Biology*, **2**, 1529–1531.

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. *Molecular Phylogenetics and Evolution*, **34**, 81–96.

Schröder HC, Efremova SM, Itskovich VB *et al.* (2003) Molecular phylogeny of the freshwater sponges in Lake Baikal. *Journal of Zoological Systematics and Evolutionary Research*, **41**, 80–86.

Shearer TL, Van Oppen MJH, Romano SL, Wörheide G (2002) Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Molecular Ecology*, **11(12)**, 2475–2487.

Stoeckle M (2003) Taxonomy, DNA and the bar code of life. *Bioscience*, **53**, 2–3.

Swofford DL (2002) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts.

Wörheide G (in press) Low variation in partial cytochrome oxidase subunit I (COI) mitochondrial sequences in the coralline demosponge *Astrosclera willeyana* across the Indo-Pacific. *Marine Biology*, in press. doi: 10.1007/s00227-005-0134-y

Wörheide G, Degnan BM, Hooper JNA (2000) Population phylogenetics of the common coral reef sponges *Leucetta* spp. and *Pericharax* spp. (Porifera: Calcarea) from the Great Barrier Reef and Vanuatu. In: *Ninth International Coral Reef Symposium* (eds Hopley D, Hopley P, Tamelander J, Done T), Volume Abstracts, p. 21. Bali, Indonesia.

Wörheide G, Degnan BM, Hooper JNA, Reitner J (2002) Biogeography and taxonomy of the Indo-Pacific reef cave dwelling coralline demosponge *Astrosclera 'willeyana'*: new data from nuclear internal transcribed spacer sequences. In: *Proceedings of the 9th International Coral Reef Symposium, Bali, Indonesia* (eds Moosa KM, Soemodihardjo S, Soegiarto A *et al.*), pp. 339–345. State Ministry for the Environment, Indonesia.

© 2005 Blackwell Publishing Ltd, *Molecular Ecology Notes*, 10.1111/j.1471-8286.2005.01259.x