ALG11 – A new variable DNA marker for sponge phylogeny: Comparison of phylogenetic performances with the 18S rDNA and the COI gene

Frida Belinky a,1, Amir Szitenberg a,1, Itay Goldfarb a,1, Tamar Feldstein a, Gert Wörheide b,c, Micha Ilana a, Dorothee Huchon a,d,*

a Department of Zoology, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel
b Department of Earth- and Environmental Sciences & GeobiolCenterLMU, Ludwig-Maximilians-Universität München, Richard-Wagner-Str. 10, 80333 München, Germany
c Bavarian State Collections of Palaeontology and Geology, Richard-Wagner-Str. 10, 80333 München, Germany
d National Evolutionary Synthesis Center, 2024 W. Main Str., Suite A200 Durham, NC 27705-4667, USA

1 These authors contributed equally to this work.

E-mail address: huchond@post.tau.ac.il (D. Huchon).

A R T I C L E   I N F O

Article history:
Received 21 March 2011
Revised 13 February 2012
Accepted 14 February 2012
Available online 25 February 2012

Keywords:
ALG11
COI
18S rDNA
Porifera
Cnidaria
Molecular phylogeny
Saturation

A B S T R A C T

Phylogenetic relationships within sponge classes are highly debated. The low phylogenetic signal observed with some current molecular data can be attributed to the use of few markers, usually slowly-evolving, such as the nuclear rDNA genes and the mitochondrial COI gene. In this study, we conducted a bioinformatics search for a new molecular marker. We sought a marker that (1) is likely to have no paralogs; (2) evolves under a fast evolutionary rate; (3) is part of a continuous exonic region; and (4) is flanked by conserved regions. Our search suggested the nuclear ALG11 as a potential suitable marker. We next demonstrated that this marker can indeed be used for solving phylogenetic relationships within sponges. Specifically, we successfully amplified the ALG11 gene from DNA samples of representatives from all four sponge classes as well as from several cnidarian classes. We also amplified the 18S rDNA and the COI gene for these species. Finally, we analyzed the phylogenetic performance of ALG11 to solve sponge relationships compared to and in combination with the nuclear 18S rDNA and the COI mtDNA genes.interestingly, the ALG11 marker seems to be superior to the widely-used COI marker. Our work thus indicates that the ALG11 marker is a relevant marker which can complement and corroborate the phylogenetic inferences observed with nuclear ribosomal genes. This marker is also expected to contribute to resolving evolutionary relationships of other apparently slow-evolving animal phyla, such as cnidarians.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Sponges are highly diversified with an estimated 15,000 living species (Hooper and Van Soest, 2002). Because sponges are both morphologically simple and diverse, sponge morphological systematics is known to be exceedingly difficult (Boury-Esnault, 2006; Hooper and Van Soest, 2002). As a case in point, numerous sponge orders that had been defined based on morphological similarities (e.g., Poecilosclerida, Halichondrida and Haplosclerida) have since been recovered as polyphyletic in molecular phylogenetics (Erpenbeck et al., 2005a, 2007a,b; Nichols, 2005; Voigt et al., 2008). Both nuclear rDNAs and the COI gene have their advantages. For example, the rDNA genes possess both conserved stems and variable loop regions which provide signals for different levels of phylogenetic inference (Voigt et al., 2008). Moreover, their amplification is facilitated by their high copy number in the genome and suitable PCR primers are available for a variety of taxa. Consequently, rDNA genes, and the 18S rDNA in particular, have been widely used to infer the evolutionary relationships among Metazoa (Aguinaldo et al., 1997; Collins, 1998; Medina et al., 2001; Peterson and Eernisse, 2001; Voigt et al., 2008; Zrzavy et al., 1998). Similar to the rDNA genes, mitochondrial markers have the advantage of being present in high copy number in each cell. They also present an independent source of information when compared to the nuclear genome. Among the 13 mitochondrial protein-coding genes shared by all animal mitochondrial genomes, the COI gene is the most

Previous sponge phylogenetic studies beneath the class level considered relatively large sponge samplings, but relied on only a few markers (usually the nuclear rDNA genes: 18S and 28S rDNA, or the mitochondrial COI gene) (Cárdenas et al., 2010; Erpenbeck et al., 2005a, 2007b; Nichols, 2005; Voigt et al., 2008). Both nuclear rDNAs and the COI gene have their advantages. For example, the rDNA genes possess both conserved stems and variable loop regions which provide signals for different levels of phylogenetic inference (Voigt et al., 2008). Moreover, their amplification is facilitated by their high copy number in the genome and suitable PCR primers are available for a variety of taxa. Consequently, rDNA genes, and the 18S rDNA in particular, have been widely used to infer the evolutionary relationships among Metazoa (Aguinaldo et al., 1997; Collins, 1998; Medina et al., 2001; Peterson and Eernisse, 2001; Voigt et al., 2008; Zrzavy et al., 1998). Similar to the rDNA genes, mitochondrial markers have the advantage of being present in high copy number in each cell. They also present an independent source of information when compared to the nuclear genome. Among the 13 mitochondrial protein-coding genes shared by all animal mitochondrial genomes, the COI gene is the most...
sequenced, as indicated by the number of entries in the nucleotide database of NCBI (Sayers et al., 2011). Utilization of this gene was promoted by the existence of universal primers, which facilitate its amplification in most metazoan species (Folmer et al., 1994). Additionally, the low sequence variability observed within species when compared to the higher variability among species has made it the marker of choice for the barcoding of animals in general (Hebert et al., 2003, 2004) and sponges in particular (Pöppe et al., 2010; Solé-Cava and Wörheide, 2007; Wörheide and Erpenbeck, 2007) but see (Erpenbeck et al., 2006; Huang et al., 2008). Given the rather large, and constantly increasing, number of sequences available for COI, this gene is often included as a marker in phylogenetic analyses of sponge evolutionary relationships (Addis and Peterson, 2005; Cárdenas et al., 2010, 2011; Erpenbeck et al., 2007a, 2008; Nichols, 2005).

There are nonetheless several disadvantages to these two markers. It is generally assumed that all rDNA copies are homogenized. However, several cases of divergent 18S rDNA copies within the same genome have been reported in Metazoa (Carranza et al., 1996; Krieger and Fuerst, 2002; Papillon et al., 2006), thus rendering the rDNA marker less reliable than previously thought. In sponges, ribosomal intragenomic variation has been shown to affect phylogenetic inferences based on internal transcribed spacer (ITS) sequences (Alvarez et al., 2007; Redmond and McCormack, 2009; Wörheide et al., 2004) but not 18S rDNA or 28S rDNA inferences. Additionally, although the 18S rDNA provides high statistical support for many metazoan clades, being a slow-evolving marker it may fail to provide enough phylogenetic signal to resolve relationships at the family level among slow-evolving members of phyla such as Porifera and Cnidaria (Bernstein et al., 2001; Dohrmann et al., 2006). The COI gene is also a slow evolving marker in sponges and anthozoans (Huang et al., 2008) and, indeed, COI sometimes fails to robustly resolve sponge relationships when used as a phylogenetic marker (Addis and Peterson, 2005; Erpenbeck et al., 2007a; Nichols, 2005). Generally, the phylogenetic relevance of COI should be evaluated for sponges, in particular since hexactinellid sponges possess a different mitochondrial genetic code than demosponges, which could affect COI-based phylogenetic reconstructions (Haen et al., 2007). Additionally, the presence of a mitochondrial intron inserted at the same location as the reverse barcoding primer complicates its amplification in sponges (Rot et al., 2006; Sizenberg et al., 2010). Finally, evolutionary relationships inferred from a single gene marker may be affected by various types of phylogenetic artifacts. However, not all genes are expected to be affected by the same biases. For example, one marker may be sensitive to the long-branch attraction artifact (Felsenstein, 1978), while a second marker may be much less sensitive to this bias. This can happen, for example, when a specific marker is fast-evolving within a specific clade because of gene-specific relaxation of selection. The above arguments suggest that more than one phylogenetic marker should be used to infer a species tree, and incongruences among markers should be studied for potential biases. Indeed, the combination of several gene markers generally improves the inference of evolutionary relationships when compared to the analysis of separate markers (Huchon et al., 2002; Mitchell et al., 2000; Russo et al., 1996).

Three approaches can be used to increase the number of gene markers: (1) large-scale sequencing approaches, such as EST projects (e.g., Dunn et al., 2008; Philippe et al., 2009; Pick et al., 2010); (2) the sequencing of specific genomic DNA markers (e.g., Bortchielini et al., 1998; Erpenbeck et al., 2005b); and (3) the sequencing of specific cDNA markers (Sperling et al., 2007, 2009). In the first approach, data originate from high throughput genomic or EST sequencing. Indeed, recent studies aiming to clarify the position of sponges among metazoans, as well as the relationships among the four sponge classes (Hexactinellida, Calcarea, Demospongiae and Homoscleromorpha), have considered numerous genes, albeit from a relatively small sampling of the sponge diversity (usually less than 10 sponge species (Belinsky et al., 2010; Dunn et al., 2008; Philippe et al., 2009; Pick et al., 2010; Rokas et al., 2005; Schierwater et al., 2009)). While such approaches are clearly powerful, they are usually more costly than classical genomic DNA sequencing, and thus they often come at the expense of large taxonomic sampling. In addition, these high throughput approaches cannot always be applied to all samples as they require both high quality and high quantity of DNA or RNA samples. In particular, these methods are often inapplicable for most museum specimens, in which RNA, and to a lesser extent DNA, is often degraded.

The second and third approaches are to sequence independently specific markers that, in combination with existing markers, will improve the resolution of sponge phylogeny. Identification of a novel phylogenetic marker for sponges is not an easy task. Phylogenetic markers should be long enough (minimum 500 bp), additionally, they should neither be too conserved nor too variable. Since intronic sequences and intergenic regions are usually too variable to be used as informative markers, except at the intra-specific level (Bentlage and Wörheide, 2007; Wörheide et al., 2008), the second approach is to sequence long exonic regions for phylogenetic reconstructions above the species level. An alternative by which to avoid intronic sequence is that of the third approach: cDNA sequencing (Sperling et al., 2007, 2009). However, similar to the large-scale sequencing approach, cDNA sequencing is not possible for most museum samples for which RNA is not preserved.

In all three methods, the markers considered should be unlikely to have paralogous copies, since erroneous inference of orthology can mislead phylogenetic inference (Martin and Burg, 2002). Finally, for the two last approaches the marker should also be flanked by conserved regions to allow the design of universal sponge primers. Searching for such markers is a challenging task given the limited current knowledge regarding the sponge nuclear genome, apart from Amphemidon queenslandica (Srivastava et al., 2010). Notably, while databases of orthologous genes exist for representatives of the mammalian diversity (Ranwez et al., 2007), no such tool exists for sponges. As a case in point, no sponge species have yet been included in major comparative genome databases such as HomoloGene (Sayers et al., 2011), Ensembl (Flicek et al., 2011) or the UCSC Genome Browser (Fujita et al., 2011).

In this study, we conducted a bioinformatics search for a novel sponge nuclear marker. Three markers were selected, however, successful PCR amplification could only be achieved for the asparagine-linked glycosylation 11 protein (ALG11, also known as the alpha-1, 2-mannosyltransferase in yeast). We here compare the phylogenetic performance of the ALG11 gene with those of the 18S rDNA and COI gene.

2. Materials and Methods

2.1. Screen for a novel gene marker

In order to identify a fast-evolving nuclear exon, we selected homology groups from the National Center for Biotechnology Information (NCBI) HomoloGene database build 46 that fitted two criteria. First, the group had to include exactly one representative from Homo sapiens, Drosophila melanogaster, Caenorhabditis elegans and Saccharomyces cerevisiae. This criterion provides a rough means by which to eliminate duplicated gene families. Second, the nucleotide coding sequences of H. sapiens, D. melanogaster and C. elegans had to include at least one exon longer than 500 bp. Since we aimed to amplify an exonic region, this second criterion was used to filter genes harboring only short exons. After this initial search, 29 homology groups remained.
We furthered narrowed the list of candidate markers by keeping only genes that are present in *Nematostella vectensis* and harbor a long exon (this step was performed before publication of the *A. queenslandica* genome). To this end, BLASTp searches were conducted against all *N. vectensis* predicted protein sequences available in StellaBase (Sullivan et al., 2006) using the *H. sapiens* sequence of each of the 29 homology groups as query. Following Hall (2004, p. 16), only sequences with an E-value $\leq 10^{-5}$ were assigned to a HomoloGene group index according to the best hit it received in the similarity search. A reciprocal BLASTp search, against all human protein sequences present in the NCBI protein database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein), was then performed using as query the 29 *N. vectensis* sequences selected in the previous step. This step ensured that no other human sequence was more closely related to one of the 29 *N. vectensis* sequences than the one in the HomoloGene group.

To determine if the genes selected could possess a long intronless region in Porifera, protein sequences of each homology group were aligned. Intron positions were extracted from StellaBase and from NCBI for all other species, and converted into positions in the aligned protein sequences using Perl scripts (see Supplementary material). The protein alignments were examined manually. Conserved proteins, highly variable proteins, and proteins for which no intronless region was shared among species were eliminated. Three candidate markers were selected at the end of this search: the asparagine-linked glycosylation 11 homolog (ALG11), the asparagine-linked glycosylation 2 homolog (ALG2), and the solute carrier family 33 (acetyl-CoA transporter), member 1 (SLC33A1).

For each candidate a BLASTp search was conducted using the Human sequence as query and a phylogenetic tree was built to confirm that the markers were indeed single copy in all animals. The result of such BLASTp search is presented, for the ALG11 gene, in Supplementary material (Fig. S1). PCR primers were designed for each marker based on the multiple sequence alignment of fungi, choanoflagellate and animal sequences (see Supplementary material, Fig. S2 for the ALG11 gene). Only the ALG11 gene could be successfully amplified in species representatives of the sponge diversity (see below).

### 2.2. DNA extraction, amplification, and sequencing

The source of the DNA and tissue samples used in this work are listed in Table 1. Tissue samples, cleaned of visible epibionts, were reduced to powder in a mortar with liquid nitrogen. Genomic DNA was extracted from 95% ethanol-preserved tissue samples following the procedure of Fulton et al. (1995).

Amplifications of the ALG11 sequences were performed in two steps. A first amplification was performed with the primers ALG11-D1 and ALG11-R1, followed by a re-amplification of the initial PCR

### Table 1

<table>
<thead>
<tr>
<th>Phylum (class)/order</th>
<th>Species</th>
<th>Tissue number/voucher number</th>
<th>Origin</th>
<th>Donor/collector</th>
<th>Tissue sample quality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cnidaria (Anthozoa)</strong></td>
<td>Stylophora pistillata</td>
<td>DH G33</td>
<td>Red Sea, Eilat, Israel</td>
<td>Alain Daniel</td>
<td>DNA sample</td>
</tr>
<tr>
<td><strong>Cnidaria (Hydrozoa)</strong></td>
<td>Hydractinia echinata</td>
<td>DH 5188</td>
<td>Lab culture</td>
<td>Sharon Gild</td>
<td>DNA sample</td>
</tr>
<tr>
<td><strong>Cnidaria (Scyphozoa)</strong></td>
<td>Pelagia noctiluca</td>
<td>TAU CO35436*</td>
<td>Mediterranean Sea, Ashdod, Israel</td>
<td>Chana Feldstein</td>
<td>Fresh tissue sample</td>
</tr>
<tr>
<td><strong>Porifera (Demospongiae)</strong></td>
<td>Rhopilema nomadica</td>
<td>DH 5143</td>
<td>Mediterranean Sea, Akko, Israel</td>
<td>Tamar Feldstein</td>
<td>Fresh tissue sample</td>
</tr>
<tr>
<td><strong>Porifera (Demospongiae)</strong></td>
<td>Chondrosia reniformis</td>
<td>TAU 25193*</td>
<td>Mediterranean Sea, Achziv, Israel</td>
<td>Chagai Rot</td>
<td>Old tissue sample</td>
</tr>
<tr>
<td><strong>Porifera (Demospongiae)</strong></td>
<td>Chondrilla sp. (Cf. nucula)</td>
<td>DH 5013</td>
<td>Mediterranean Sea, Achziv, Israel</td>
<td>Chagai Rot</td>
<td>Old tissue sample</td>
</tr>
<tr>
<td><strong>Porifera (Demospongiae)</strong></td>
<td>Ircinia sp.</td>
<td>TAU 25496*</td>
<td>Mediterranean Sea, Neve-Yam, Israel</td>
<td>Tamar Feldstein</td>
<td>Fresh tissue sample</td>
</tr>
<tr>
<td><strong>Porifera (Demospongiae)</strong></td>
<td>Sarcoptosinus sp.</td>
<td>TAU 25501*</td>
<td>Mediterranean Sea, Neve-Yam, Israel</td>
<td>Tamar Feldstein</td>
<td>Fresh tissue sample</td>
</tr>
<tr>
<td><strong>Porifera (Demospongiae)</strong></td>
<td>Desmosammina anthemistora</td>
<td>UCMP WC1660</td>
<td>Bocas del Toro, Panama</td>
<td>Scott A. Nichols</td>
<td>Museum sample</td>
</tr>
<tr>
<td><strong>Porifera (Demospongiae)</strong></td>
<td>Negombata magnifica</td>
<td>TAU 25198*</td>
<td>Red Sea, Eilat, Israel</td>
<td>Chagai Rot</td>
<td>Old tissue sample</td>
</tr>
<tr>
<td><strong>Porifera (Demospongiae)</strong></td>
<td>Biominia fistulosa</td>
<td>TAU 25567*</td>
<td>Bahamas</td>
<td>Chagai Rot</td>
<td>DNA sample</td>
</tr>
<tr>
<td><strong>Porifera (Demospongiae)</strong></td>
<td>Ectyphosia ferox</td>
<td>UCMP WC970*</td>
<td>Galeta, Panama</td>
<td>Scott A. Nichols</td>
<td>Museum sample</td>
</tr>
<tr>
<td><strong>Porifera (Demospongiae)</strong></td>
<td>Cinachyrella levantinensis</td>
<td>TAU 25445*</td>
<td>Mediterranean Sea, Achziv, Israel</td>
<td>Chagai Rot</td>
<td>Fresh tissue sample</td>
</tr>
<tr>
<td><strong>Porifera (Demospongiae)</strong></td>
<td>Cniella sp.</td>
<td>QM G316372*</td>
<td>South Norfolk Ridge (Norfanz cruise)</td>
<td>John N.A. Hooper</td>
<td>Museum sample</td>
</tr>
</tbody>
</table>


Origin of the tissue samples. Fresh tissue samples are samples that were kept in ethanol at –80°C and extracted few months after collection. Old tissue samples are samples that were kept in ethanol at –80°C and extracted 2 or 4 years after collection. Museum samples are samples from museum origin. In our case, these samples had been collected and kept frozen in ethanol or DMSO at least a few years before receiving them. All museum samples were shipped by mail oversea at room temperature, and placed at –80°C upon reception. Poecilosclerida is a polyphyletic order, *Biemma* and *Ectyphosia* belong to families that are not related to other Poecilosclerida families (Erpenbeck et al., 2007a,b). Voucher specimens are indicated by a *C*. BELUM, Ulster Museum (Belfast, Northern Ireland); DH, Dorothee Huchon lab collection; GW, Gert Wörheide lab collection; QM, Queensland Museum (Brisbane, Australia); SMF, Senckenberg Museum (Frankfurt am Main, Germany); TAU, Zoological Museum, Tel Aviv University (Tel-Aviv, Israel); UCMP, University of California Museum of Paleontology (Berkeley, USA).
product using the nested primers ALG11-D2 and ALG11-R2. Primer sequences are indicated in Table 2. There were, however, two exceptions to this protocol. For Craniella sp. 3878 (Demospongiae), the first amplification was performed with the primers ALG11-D1 and ALG11-R1 but the re-amplification was done with the primers ALG11-D2-new and ALG11-R2-new. For Calcarea species the first amplification was performed with the primers ALG11-D1 and ALG11-R2(t) followed by a re-amplification of the initial PCR product into two overlapping fragments using the primers ALG11-D2 and ALG11-R3 for the first fragment and the primers ALG11-F3 and ALG11-R2(t) for the second fragment.

The 18S rDNA gene was PCR-amplified using the primer sets 18S1/18S2 (Borchiellini et al., 2001). Since the quantity of DNA obtained was usually insufficient for direct sequencing, the PCR products obtained were re-amplified in two overlapping fragments of approximately 1 kb and 1.2 kb using the primer-pairs 18S1/18S6 and 18S7/18S2 (Table 2). For Craniella sp. 3878, the 18S rDNA was amplified with the primers 18S-poe-D and 18S-poe-R (Table 2), yielding a slightly shorter fragment.

Following Rot et al. (2006), the COI gene from sponges was amplified with the primer LCO1490 and COX1R1, with a few exceptions. Irccidin COI was amplified with the newly-designed primers Cox_Irc_D1 and Cox_Calc_R1 followed by a re-amplification of the initial PCR product using the nested primers Cox_Irc_L-CO and Irc_Cox1R1 (Table 2). The Walteriu (Hexactinellida) sequence was amplified with the primers HEX-COX1D1 and Cox_Calc_R1. The PCR products obtained were then re-amplified in two semi-nested reactions with the internal primer set HEX-COX1D1/HEX-COX1D3 and Cox_Calc_R1/Cox_Calc_R2 (Table 2). Unfortunately, no COI sequences could be obtained for any Calcarea samples. Amplifications of Cnidaria COI were performed in two steps. A first amplification was performed with the primers Clath_COX_F1 and Cnidr_R1-COX, followed by a re-amplification of the initial PCR product using the nested primers Cnidaria_LCO and Cnidr_R2_COX (or with the reverse primer Clath_COX_R1 for Craniella sp. 3878). PCR products were purified using polyethylene glycol (PEG) in saline (NaCl). Most of the products were directly sequenced using Big Dye Terminator v1.1 (Applied Biosystems) on an ABI 310 sequencer. In a few cases, PCR products were ligated into the pSC-A vector. Ligation products were then transformed into StrataClone™ SoloPack™ competent cells (Stratagene). In such cases, five clones per species were sequenced. All genes were completely sequenced on both strands. The 52 new sequences (Table 3) have been deposited in the European Nucleotide Archive of EMBL under Accession Numbers FR819667–FR819693 and HE591446–HE591470.

2.3. Alignment generations

In addition to the 52 sequences obtained as described above, 18S rDNA, COI and ALG11 sequences of species representing the major animal clades were retrieved from GenBank and trace data (Table 3). Protein sequences of the COI and ALG11 genes and DNA sequences of the 18S rDNA gene were aligned with MAFFT under the L-INS-I strategy (Katoh et al., 2005) using the GUIDANCE web-server (Penn et al., 2010). Unreliable positions (i.e., position with a GUIDANCE threshold below 0.93) were excluded from the alignment as well as positions with 50% or more of missing data. The DNA sequences of COI and ALG11 were then aligned according to the protein sequence alignments using the program PAL2NAL (Suyama et al., 2008). For the ALG11 and the COI genes two datasets were considered: the first dataset includes all codon positions, in the second dataset the third codon positions were excluded using PAUP* v4.0b10 (Swofford, 2000). Sequence alignments are available as Supplementary material and in the Dryad repository: doi:10.5061/dryad.20617b5g.

2.4. Phylogenetic analyses

Phylogenetic analyses were conducted on each gene separately (with and without third codon positions for the ALG11 and COI genes). Phylogenetic trees were inferred using maximum likelihood (ML) in RAxML (Stamatakis, 2006) and Bayesian inference (BI) in MrBayes (Huelsenbeck et al., 2001) for the individual genes and the concatenated dataset, as well as the analyses that included the outgroups. The gene datasets were analyzed individually using RAxML and MrBayes. A phylogenetic tree was obtained for each gene, and a consensus tree was generated for the concatenated dataset. The trees obtained were then compared and the results were used to infer the evolutionary relationships of the different taxa.
Discarding 20% of the cycles as burnin. For each chain, the total package. The sampling was performed every ten cycles after verified using the bpcomp and tracecomp programs of the PhyloBayes depending on the dataset. Convergence of the two chains was verified by two independent chains were run for 154,296–628,402 cycles on DNA sequences. For each dataset (separated and combined), all analyses were based cause the mixture models available in PhyloBayes do not allow the site-heterogeneous CAT–GTR model (Lartillot et al., 2009). Bayesian inference.

Bayesian analyses were performed using PhyloBayes v 3.2e with the site-heterogeneous CAT–GTR model (Lartillot et al., 2009). Because the mixture models available in PhyloBayes do not allow combining DNA and protein sequences, all analyses were based on DNA sequences. For each dataset (separated and combined), two independent chains were run for 154,296–628,402 cycles depending on the dataset. Convergence of the two chains was verified using the bpcomp and tracecomp programs of the PhyloBayes package. The sampling was performed every ten cycles after discarding 20% of the cycles as burnin. For each chain, the total number of trees considered was higher than 8000. As recommended by the PhyloBayes manual, we verified that for each run the parameters “maxdiff” were less than 0.1, “rel_diff” were less than 0.1 and “effsize” were higher than 100. To verify that the CAT model would better explain the evolution of the 18S rDNA k sequences. (2) COI Plakortis angulospiculatus; AlG11, 18S rDNA Plakinasterella sp. (3) COI Oscarella carmelle; AlG11, 18S rDNA Oscarella sp. Sequences in bold are newly obtained sequences.

Table 3
Accession numbers of the sequences. (1) COI Nematostella vectensis (2) COI Plakortis angulospiculatus; AlG11, 18S rDNA Plakinasterella sp. (3) COI Oscarella carmelle; AlG11, 18S rDNA Oscarella sp. Sequences in bold are newly obtained sequences.
were compared based on the same species sampling. Phylogenetic area were excluded from the PABA analysis so that all datasets Because COI sequences of Calcarea could not be determined, Calc- tations were repeated for each marker and for 12 nodes of interest. calculate the difference in BP support values resulting from the addition of the ALG11 dataset to the 18S rDNA dataset, and once to calculate the difference in BP support values resulting from the second marker, the same computation was performed twice: once between partitions. To compute the average the support of a node while a negative value indicates a conflict be-

Maximum likelihood analyses of the separated and concate- nated datasets were performed using RAxML 7.2.6 (Stamatakis, 2006) with the GTR + 1/4 model. Bootstrap percentages (BPs) were computed for each dataset based on 100 replicates. For the ALG11 and the COI genes each codon position was treated as a different partition with different shape parameters, GTR-rates, and empirical base frequencies. In the combined analysis, similarly, different parameters were assumed for the 18S rDNA partition and for each codon position of the ALG11 and COI genes.

2.5. PABA analysis

To better evaluate the performance of each marker, we conducted a partition addition bootstrap alteration (PABA) analysis (Struck et al., 2006). For nodes of interest the PABA approach computes the alteration of BP (\( \delta \)) when a partition is added. In our case, three gene partitions were considered in the PABA analysis: the 18S rDNA, the ALG11 and the COI gene. To compute, for example, the \( \delta \) value of the ALG11 on the demosponge monophyly, when the ALG11 gene is added as third marker, we first computed the BP of this node when the 18S rDNA and COI are combined (in this case \( BP = 32 \)). We then computed the BP of demosponge mono-

While the COI gene shows a distribution of number of inferred profiles similar to that of the 18S rDNA, the ALG11 gene appears to be even more heterogeneous across sites (Fig. 1). A possible explana-

Maximum likelihood analyses of the separated and concate-

The ALG11 gene is a mannosyltransferase that belongs to the N- linked protein glycosylation pathway: a pathway that is highly conserved in eukaryotes (O’Reilly et al., 2006). In human (NG_028038) and fly (NT_037436 (Hoskins et al., 2007)), the ALG11 gene consists of four and five exons, respectively. However, in the cnidarians Nematosello vescens (NW_001834198 (Putnam et al., 2007)) and Hydra magnipapillata (NW_002161912 (Chapman et al., 2010)), as well as in the placozoan Trichoplax adhaerens (NW_002609497 (Srivastava et al., 2008)) and in the sponge A. queenslandica (trace data (Srivastava et al., 2010)), this gene is intronless, suggesting that the gain of introns occurred after the divergence of bilaterians. The region amplified in this work corre-

Our second goal – to determine a variable marker – was also achieved. After removing non-reliable characters from the align-

3. Results and Discussion

3.1. The ALG11 gene – a novel variable marker

Our motivation was to determine a new phylogenetic marker that could easily be amplified in all sponge classes and would also be more variable than the 18S rDNA and the COI genes. To identify such a marker, a high throughput bioinformatics search was con-

This region was chosen visually since its sequence variability seems to be at a level appropriate for resolving the desired level of sponge relationships and it is flanked by highly conserved re-

The ALG11 gene is a mannosyltransferase that belongs to the N- linked protein glycosylation pathway: a pathway that is highly conserved in eukaryotes (O’Reilly et al., 2006). In human (NG_028038) and fly (NT_037436 (Hoskins et al., 2007)), the ALG11 gene consists of four and five exons, respectively. However, in the cnidarians Nematosello vescens (NW_001834198 (Putnam et al., 2007)) and Hydra magnipapillata (NW_002161912 (Chapman et al., 2010)), as well as in the placozoan Trichoplax adhaerens (NW_002609497 (Srivastava et al., 2008)) and in the sponge A. queenslandica (trace data (Srivastava et al., 2010)), this gene is intronless, suggesting that the gain of introns occurred after the divergence of bilaterians. The region amplified in this work corre-

This region was chosen visually since its sequence variability seems to be at a level appropriate for resolving the desired level of sponge relationships and it is flanked by highly conserved re-

The marker was successfully amplified from genomic DNA of all sponge and cnidarian samples for which we could amplify the 18S rDNA and the COI gene, including museum samples that had been preserved for several years in ethanol (Table 1). The region amplified was about 930 bp and none of the sequences obtained were found to include an intron. This result confirms that the ALG11 marker can be used for museum samples from which cDNA cannot be obtained.

Our second goal – to determine a variable marker – was also achieved. After removing non-reliable characters from the align-

After removing non-reliable characters from the align-

This region was chosen visually since its sequence variability seems to be at a level appropriate for resolving the desired level of sponge relationships and it is flanked by highly conserved re-

The marker was successfully amplified from genomic DNA of all sponge and cnidarian samples for which we could amplify the 18S rDNA and the COI gene, including museum samples that had been preserved for several years in ethanol (Table 1). The region amplified was about 930 bp and none of the sequences obtained were found to include an intron. This result confirms that the ALG11 marker can be used for museum samples from which cDNA cannot be obtained.
differences between sequences. For example, among the 11 demosponge species sequenced, the \( p \)-distances ranged from 0.160 to 0.509 (140–449 substitutions, Supplementary Table S1), while for the 18S rDNA the \( p \)-distances ranged from 0.008 to 0.132 (12–209 substitutions, Supplementary Table S2) and for the COI they ranged from 0.047 to 0.303 (56–360 substitutions, Supplementary Table S3). Clearly, the ALG11 is more variable than the other two markers. It is worth noting that this was also true when the third codon positions were excluded. In this case the \( p \)-distances among demosponges ranged for the ALG11 gene from 0.078 to 0.416 (46–246 substitutions, Supplementary Table S4), while for the COI gene they ranged from 0.014 to 0.151 (11–117 substitutions, Supplementary Table S5).

To evaluate the level of saturation of the ALG11 gene, following Philippe and Forterre (Philippe and Forterre, 1999), we plotted in Fig. 2 the average number of observed nucleotide substitutions per site (i.e., the \( p \)-distances) as a function of the average number of inferred nucleotide substitutions per site based on the Maximum Likelihood (ML) criterion (i.e., the patristic ML distances or the sum of lengths of all branches linking two sequences on the ML tree). Using this approach, mutational saturation is inferred when \( p \)-distances remain constant while ML distances increase (i.e., presence of a “plateau”) (Philippe and Forterre, 1999). For the ALG11 and COI genes the saturation plots were drawn either with the three codon positions (Fig. 2A and B) or excluding the third codon positions (Fig. 2C and D). The saturation plot involving all possible species pairs of our dataset, and third codon positions (Fig. 2A), shows that the ML tree distances are the largest for the COI gene and that \( p \)-distances above 0.3 are likely to be saturated for this marker. These results illustrates the extremely high mutation rate of the third codon positions in mitochondrial genes. By comparison, the ALG11 marker appears to be more variable and less saturated although \( p \)-distances above 0.5 are likely to be saturated. No saturation is noted for the 18S rDNA gene which is much less variable. When the third codon positions were removed (Fig. 2C) the tree distances were much reduced for both the ALG11 and the COI gene. The saturation plot suggests that \( p \)-distances above 0.4 are likely to be saturated for the ALG11 gene while \( p \)-distances above 0.2 are likely to be saturated for the COI gene. Again the COI gene appears to be less variable and more saturated than the ALG11 gene.

To evaluate the performance of each marker to solve sponge relationships, saturation plots were also drawn considering only distances among members of the same sponge class (Fig. 2 B and D). The ALG11 gene does not appear to be saturated, with or without the third codon positions (Fig. 2B). Consequently, third codon positions were conserved in subsequent phylogenetic analyses. The COI gene, conversely, seems less saturated, when third codon positions are excluded. However, the ML tree reconstructed based on the first and second codon positions of the COI gene supported less sponge clades than the tree reconstructed based on the three codon positions (Supplementary material, Fig. S3). Since our goal is to compare marker performances within sponge classes, all three codon positions of the COI gene were also considered in subsequent analyses.

More generally, the saturation plots confirm the high variability of the ALG11 marker when compared to the 18S rDNA and the COI gene, as the number of observed and inferred substitutions between sequence pairs is usually the highest for the ALG11 gene. This is particularly true when considering distances between two members of the same sponge class (Fig. 2B).

### 3.2. Separate phylogenetic analyses

Each DNA dataset described above was analyzed separately using the Bayesian inference scheme, with the CAT-GTR model of sequence evolution, and using the ML criterion with the
Fig. 3. Bayesian consensus trees reconstructed for each marker using the CAT–GTR model. A: Phylogenetic reconstruction based on 18S rDNA. B: Phylogenetic reconstruction based on ALG11 DNA sequences (all three codon positions). C: Phylogenetic reconstruction based on COI DNA sequences (all three codon positions). Only nodes supported by a posterior probability above 0.50 are presented. Posterior probabilities/ML bootstrap supports are given near the corresponding node. '-' indicates that the corresponding node is absent from the bootstrap consensus tree. Solid circles indicate branches with maximal support values (Bayesian PP = 1.0 and ML BP = 100); gray circles indicate branches with high support values (0.99 ≤ PP ≤ 1.0 and 98 ≤ BP ≤ 100).
Table 4
Alteration of bootstrap support δ to nodes with the addition of data partitions. NA – not applicable due to alteration from 100 to 100. Negative values are indicated with a gray background. The values of the genes that contributed the most to the support of a node are indicated in bold. G1, G2 and G4 are demosponge clades as defined by Borchiellini et al. (2004).

<table>
<thead>
<tr>
<th>Clades</th>
<th>BP in combined tree</th>
<th>18S rDNA</th>
<th>ALG11</th>
<th>COI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilateria</td>
<td>100</td>
<td>50.5</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Cnidaria</td>
<td>64</td>
<td>70</td>
<td>57</td>
<td>2</td>
</tr>
<tr>
<td>Placozoa + Cnidaria</td>
<td>53</td>
<td>66.5</td>
<td>30</td>
<td>22</td>
</tr>
<tr>
<td>Placozoa + Cnidaria + Bilateria</td>
<td>98</td>
<td>99</td>
<td>87</td>
<td>7</td>
</tr>
<tr>
<td>Hexactinellida</td>
<td>100</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Homoscleromorpha</td>
<td>100</td>
<td>63</td>
<td>19</td>
<td>29</td>
</tr>
<tr>
<td>Demospongiae</td>
<td>76</td>
<td>−7</td>
<td>1</td>
<td>34.5</td>
</tr>
<tr>
<td>Demospongiae + Hexactinellida</td>
<td>32</td>
<td>44.5</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>G1</td>
<td>100</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>G2</td>
<td>94</td>
<td>18</td>
<td>−6</td>
<td>46</td>
</tr>
<tr>
<td>G4</td>
<td>100</td>
<td>74</td>
<td>47</td>
<td>32</td>
</tr>
</tbody>
</table>

Fig. 4. Bayesian consensus tree reconstructed using the CAT–GTR model for the concatenated DNA sequences of 18S rDNA, ALG11, and COI. Posterior probabilities/ML bootstrap supports are given near the corresponding node. “−” indicates that the corresponding node is absent from the bootstrap consensus tree.
GTR + I4 model. The Bayesian trees are shown in Fig. 3. These trees were congruent with the ones obtained under the ML criterion. Only nodes with Bayesian posterior probability (PP) above 0.5 are shown. Similarly, given the low species sampling, relationships among Bilateria are not addressed since they have already been the subject of numerous studies (Dunn et al., 2008; Lartillot and Philippe, 2008; Philippe et al., 2005; Rokas et al., 2005).

Phylogenetic reconstructions show that, although it is the least variable, the 18S rDNA tree is the most resolved (Fig. 3A). Indeed the analysis of 18S rDNA sequences recovers the monophyly of: Bilateria (PP = 1.0; ML bootstrap percentage BP = 100), Cnidaria (PP = 0.82, BP = 92), Homoscleromorpha (PP = 1.0, BP = 100), Calcarea (PP = 1.0, BP = 100), Hexactinellida (PP = 1.0, BP = 100), Homoscleromorpha + Calcarea (PP = 0.99, BP = 85), Silicea (Hexactinellida + Demospongiae PP = 0.97, BP = 59), Metazoa (PP = 0.91, BP = 61), Chaoanoflagellida (PP = 1.0, BP = 100), and Fungi (PP = 1.0, BP = 99). Surprisingly, Demospongiae were not found to be monophyletic since Hexactinellida were placed as the sister clade of the Dictyoceratida species (Sarcotragus and Incria). However, Demosponge paraphyly is not supported (PP = 0.57, BP = 50).

The Bayesian ALG11 and COI trees are much less resolved. The ALG11 tree (Fig. 3B) only recovers the monophyly of the sponge clades Calcarea (PP = 0.99, BP = 99), Homoscleromorpha (PP = 0.56, BP = 49), Hexactinellida (PP = 0.99, BP = 100), Demospongiae (PP = 0.99, BP = 58), and Fungi (PP = 0.61, BP = 86) while the COI tree recovers the monophyly of Bilateria (PP = 1.0, BP = 100), Metazoa (PP = 0.98, BP = 80), Chaoanoflagellida (PP = 0.99, BP = 94), Fungi (PP = 0.99, BP = 96) and Hexactinellida (PP = 1.0, BP = 100).

The low support observed for Metazoa and Bilateria monophyly with the ALG11 marker can be explained by the fast evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolving ALG11 marker. The ALG11 dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene. This dataset is most probably saturated by the fast-evolutionary rate of this gene.

3.3. PABA analyses

In agreement with the phylogenetic results obtained on separate markers, the PABA analysis indicates that the 18S rDNA gene contributes the most to determining the animal phylogeny (Table 4). The 18S rDNA is the gene showing the highest alteration of bootstrap support in favor of the monophyly of Cnidaria, the grouping of Bilateria + Cnidaria + Placozoa, the monophyly of Homoscleromorpha, the grouping of Hexactinellida and Demospongiae, and the monophyly of the G4 clade sensu Bortchelli et al. (2004) (in our case Bienna, Cinachyrella, Cranilla, Desmapsamma, Ectyoplasia, and Negombata). Our results indicate that for relationships among sponges the ALG11 gene provides negative alterations of bootstrap support. Because these negative support alterations are rather weak (most $\delta$ values are above 10), these $\delta$ values most probably indicate lack of phylogenetic signal rather than genuine incongruence. Indeed, there is no reason to assume that the evolutionary history of the COI gene differed from that of the 18S rDNA. Preliminary analyses indicated that the removal of the third codon positions in the COI and ALG11 datasets does not alter these conclusions.

Although the ALG11 tree was not fully resolved, the PABA analysis indicates that, when combined with other markers, this gene improves branch support and thus tree resolution. Consequently, the ALG11 gene appears to be a good marker when combined with the 18S rDNA. Concerning the COI and 18S rDNA performances, our results agree with Struck et al. (2006), who show that for Eunicida (Annelida) the 18S rDNA has the largest positive influence on $\delta$ values, and COI has the least influence.

3.4. Combined analysis

A phylogenetic tree was reconstructed based on the combined dataset of the three markers. The resulting tree is presented in Fig. 4. It mostly agrees with the 18S rDNA tree, as it recovers with maximal support (PP = 1.0, BP = 100) the monophyly of Bilateria, Hexactinellida, Homoscleromorpha, Calcarea, Metazoa, Chaoanoflagellida, and Fungi. It also recovers Silicea (Hexactinellida + Demospongiae PP = 0.99, BP = 79) as well as Homoscleromorpha + Calcarea (PP = 0.98, BP = 68).

Similar to the 18S rDNA result Demospongiae were found to be paraphyletic, with Hexatinellida placed as sister clade of Dictyoceratida species. This weird positioning is most probably an artifact since most phylogenetic analysis support the monophyly of demosponges (Pick et al., 2010; Sperling et al., 2009). Additional species and markers are thus needed to establish the relationships among the four main demosponge lineages and Hexactinellida.

4. Conclusions

Our goal was to design a novel phylogenetic marker to investigate relationships within sponge classes. We successfully amplified a new variable exonic marker from the ALG11 gene for a few representatives of each sponge class. The ALG11 was found to be more informative regarding sponge relationships than the COI gene. The PABA analysis further shows that ALG11 increases the phylogenetic signal when combined with other markers. The ALG11 gene is thus a promising marker to improve our understanding of sponge relationships. Since no introns were found in the cnidarian species considered, it is probable that this marker can also be applied to solve debated nodes within the Cnidaria.

Acknowledgments

We would like to thank Naomi Paz for editing the text, Tal Pupko for his comments, and Nir Ezra for sequencing Desmapsamma anchorata. This work would not have been possible without the help of Alain Daniel, John Hooper and the Queensland Museum, Dorte Janussen and the Senckenberg Museum, Sally Leys, Scott Nicholls and the University of California Museum of Paleontology, Bernard Picton and the Ulster Museum, and Chagai Rot, who provided tissue samples. Revital Ben-David-Zaslow and Sigal Shefer provided invaluable help with the collections of the Zoological Museum at Tel Aviv University. We also would like to thank the Israeli Ministry of Science Culture & Sport for supporting the National Collections of Natural History at Tel-Aviv University. This


