Göttinger Zentrum für Biodiversitätsforschung und Ökologie

– Göttingen Centre for Biodiversity and Ecology –

Molecular Evolution in non-bilaterian Metazoa, with Emphasis on Phylum Porifera

Dissertation zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultäten der
Georg-August-Universität zu Göttingen

vorgelegt von

Oliver Voigt

aus

Hannover

Göttingen, August 2009

Referent: Prof. Dr. Gert Wörheide

Korreferent: Prof. Dr. Allen G. Collins

Contents

Chapter 3

Evolution of Calcareous Sponges inferred from SSU and LSU ribosomal RNA genes – new insights and remaining problems

Chapter 4

Mitochondrial diversity of early branching Metazoa is revealed by the complete mt genome of a haplosclerid demosponge

Chapter 5

A fragmented metazoan organellar genome: the two mitochondrial chromosomes of Hydra magnipapillata

Summary of results and conclusions 97

Introduction

General Introduction

The relationships at the base of the metazoan tree are still highly controversial. Especially the relationships between, but also within, the non-bilaterian phyla remain uncertain (Minelli 2009). The application of phylogenetic inference based upon the analyses of DNA sequence is especially useful in these relatively character-poor (compared to Bilateria) taxa. The study of the molecular evolution and systematics therefore has contributed much to our current understanding of the phylogeny of early diverging Metazoa.

This work explored the molecular evolution and systematics of non-bilaterian Metazoa on two different levels. I first examined the special features of ribosomal RNA genes and their evolution within sponges (Phylum Porifera). I hereby focused on the special traits of ribosomal RNA genes as phylogenetic markers. This enabled me to address questions regarding the evolution of rRNA itself, and the phylogenetic relationships of taxonomically difficult sponge taxa. In a second stage, I investigated the evolution of complete mitochondrial (mt) genomes in non-bilaterian Metazoa. This was achieved by analyzing the gene content and arrangement of a demosponge mt genome and the very specially organized mt genome of *Hydra magnipapillata* (Cnidaria, Hydrozoa).

Non-bilaterian Metazoa - an overview

Metazoa represent a monophyletic group within the tree of life, with choanoflagellates as their sister group. This has been confirmed by the analysis of DNA sequence data (Wainright et al., 1993; Lang et al., 2002; Cavalier-Smith and Chao, 2003), but was already suspected because of the similarity between these collar-flagellated protists and the choanocytes of sponges (Porifera) (e.g.,James-Clark, 1866). The non-bilaterian phyla in the Metazoa split off before the occurrence of the last common ancestor of the Bilateria, and comprise the phyla Porifera**,** Placozoa, Cnidaria and Ctenophora, which shall be introduced below.

Porifera are sessile suspension feeders with an aquiferous system, in which special ciliated cells (the choanocytes) create a water flow (Hooper et al., 2002). Food particles and dissolved nutrients are taken up by pinocytosis or phagocytosis, respectively. Porifera has about 8,300 described species (The World Porifera database, http://www.marinespecies.org/porifera, consulted in August 2009, van Soest et al. 2009) and will be described in greater detail after the other non-bilaterian groups have been introduced.

Placozoans are small, benthic animals that resemble a flat ciliated disk and posses a distinguishable upper and lower side. Placozoans have the simplest organization of all metazoan phyla, with just four readily distinguishable somatic cell types (Grell and Ruthmann, 1991). However, it seems likely that the organizational simplicity is not an ancestral state maintained from the last common ancestor of Metazoa (Siddall et al., 1995; Cavalier-Smith et al., 1996; Collins, 1998; Srivastava et al., 2008; but see also Dellaporta et al., 2006; Schierwater et al., 2009). The unicellular layer of

the upper side of placozoans contains flat, ciliated cells and larger spherical structures, the shiny spheres, which may have a defensive function (Pearse and Voigt, 2007). The lower unicellular layer has flask-shaped ciliated cells and gland cells and is usually facing the substrate, over which the animals move in an amoeboid, shape-shifting fashion. Both, the upper and the lower cell layer lack basal laminae, or any recognizable extracellular matrix. Interestingly, however, genes for putative extracellular matrix proteins are present in the recently sequenced placozoan genome, including collagen IV and others (Srivastava et al., 2008). Gland cells of the lower side probably excrete digestive enzymes, and nutrients are taken up from the lower side via endocytosis (Grell and Ruthmann, 1991). Between the upper and the lower layer, a loose meshwork of syncytially organized, contractile fiber cells are present. Asexual reproduction occurs in forms of binary fission, budding or fragmentation. Despite the lack of observation of their development, there is evidence for sexual reproduction in Placozoa (Grell, 1972; Signorovitch et al., 2005). Placozoans are not a very diverse group: indeed, it has long been assumed that they were represented by a sole species, *Trichoplax adhaerens*. According to molecular data however, the diversity is somewhat higher, and there are at least four deeply diverged genetic clades, representing morphologically indistinguishable species (Voigt et al., 2004; Pearse and Voigt, 2007).

Porifera and Placozoa lack a tissue grade of organization, i.e. their epithelia do not have basal laminae (with the exception of homoscleromorph sponges, Boury-Esnault et al., 2003), and they lack organs e.g., nervous systems or sensory organs. Cnidaria and Ctenophora on the other hand are eumetazoans. They have tissues with basal laminae, a nervous network (and in many cases sensory organs), and a gastrovascular system.

Cnidarians are very diverse and include over 11.000 described species (Daly et al., 2007). They are characterized by cnidae (which can be present in form of nematocysts, ptychocysts or spirocysts). Their body is formed by two unicellular epithelia: the ectoderm on the outside, and the endoderm, delimiting the gastrocoel. Muscles are present in form of epithelio-muscular cells. Between the ectoderm and the endoderm lies the mesoglea, primarily containing extracellular matrix. Additionally, cells from the ectoderm can invade into the mesoglea (Storch and Welsch, 1997). In their lifecycles many cnidarians have an alternation of generations from a sessile, asexually reproducing polyp stage to a pelagic, sexually reproducing medusa stage. In the class Anthozoa the medusa stage is missing. The other cnidarian classes that generally have a medusa stage are referred to as Medusozoa and comprise the classes Cubozoa, Hydrozoa, Scyphozoa and Staurozoa (Daly et al., 2007). The lifecycle of Medusozoa is modified in many taxa and both polyp or medusa stage can be missing (Storch and Welsch, 1997). E.g., polyps of the freshwater genus *Hydra* (Hydrozoa) reproduce asexually and sexually without a medusa stage (Pearse et al., 1987).

Ctenophores are a relatively small group of marine animals with about 90 to 100 described species (Storch and Welsch, 1997). Usually eight rows of ciliary combs radiate over the ctenophore body. For locomotion, the combs of each row beat successively in a wavelike fashion (Pearse et al., 1987). The body is organized in three layers. The ectoderm is a unicellular layer covering the outside and the pharynx; the endoderm, also a unicellular layer, lines the gastrovascular cavity. A

mesoglea between these layers supports the body. It contains a gelatinous mass with extracellular fibers, long muscle cells and amoeboid cells (Pearse et al., 1987). Ctenophores have two long retractable tentacles with special adhesive cells, the colloblasts. Small organisms and other food particles are captured with the tentacles, and moved into the mouth and the gastrovascular cavity. Some ctenophores have specialized in feeding on larger prey and lack tentacles. Most species are free-swimming, but some taxa are benthic (Storch and Welsch, 1997). Ctenophores have no alternation of generations in their life cycle, and reproduction is sexual with a few exceptions.

The relationships among the non-bilaterian phyla are still controversial. From cladistic analyses of morphological characters the relationships described in the following were suggested e.g. by Ax (1995) and Nielsen et al. (1996) and are commonly found in biological textbooks (e.g.,Westheide and Rieger, 1996; Nielsen, 2001). According to this phylogeny, Porifera is a sister taxon to the rest of Metazoa, which are characterized by the presence of belt desmosomes between cells, and therefore referred to as “Epithelizoa” (Ax, 1995). Subsequently, Placozoa is the sister taxon to “Eumetazoa”. In the latter, Cnidaria is the sister taxon to the clade containing Ctenophora and Bilateria (“Acrosoma”, ” Ax, 1995). Molecular analyses provided an additional independent source of information about the relationships of non-bilaterian Metazoa. However, instead of resolving the deep nodes in the metazoan tree, molecular analyses regularly bring forward different, and sometimes – from a morphological point of view – unexpected hypotheses, such as Ctenophora as the first diverging metazoan phylum (Dunn et al., 2008), or a sister group relationship between Bilateria and a clade comprising Porifera, Placozoa, Cnidaria and Ctenophora (Dellaporta et al., 2006; Schierwater et al., 2009). Other authors have found no supported resolution at these deep nodes, and interpreted this as the result of a rapid radiation of metazoan phyla that makes the recovery of a phylogenetic signal from random 'noise' in the DNA data impossible (Rokas et al., 2005; Rokas and Carroll, 2006). However, the sequence data to address these questions is steadily increasing, and many studies today apply phylogenomic approaches via sequencing of expressed sequence tags (ESTs), or whole genomes (e.g., Dunn et al., 2008; Srivastava et al., 2008; Philippe et al., 2009). Some recent work suggested that analyses of microRNAs (noncoding regulatory RNA molecules) might be an additional marker that can be applied to solve phylogenetic questions at the phylum level (e.g., Sempere et al., 2006; Sempere et al., 2007; Wheeler et al., 2009) but a comprehensive dataset for all non-bilaterian phyla is not yet available.

The additional data and markers may prove to be sufficient to answer the question if the deep metazoan nodes can be resolved at all by means of DNA sequence analysis.

Phylum Porifera

Because large parts of this work are dealing with phylum Porifera, a more detailed introduction is appropriate. Porifera are characterized by their aquiferous system, consisting of inhalant canals, choanocyte chambers and exhalent canals. Water enters the sponge through numerous pores and leaves it through larger openings, the osculae (either singular or plural). The sponge body has three organizational layers, the pinacoderm, the choanoderm and the mesohyl, each characterized by specific cell types.

The pinacoderm covers the outer surface, and sometimes parts of the internal cavities of a sponge, and is characterized by flat cells. The choanoderm delimits all or parts of the internal cavities of the sponge and is formed by a special cell type, the choanocytes. These are cells possessing a microvilli collar with a large central flagellum. The beating flagella of the choanoderm create a water-current through the sponge. Dissolved nutrients and small food particles are taken up from this water via phagocytosis or pinocytosis. The mesohyl is situated between the other two layers and consists of extracellular matrix in which collagenous fibers, skeletal spicules and numerous cell types are present. Sponges posses a population of highly mobile cells. Many of these cell types, especially the so-called archaeocytes, are capable of transforming into other cell types (van Soest, 1996).

The mineral skeleton of sponges consists of calcareous or siliceous spicules; some sponges are mainly supported by spongin fibers. The shape of the spicules is important for sponge taxonomy. In some species, a rigid skeleton can be formed by the fusion of spicules. Some other sponges are characterized by a hyper-calcified basal skeleton (Storch and Welsch, 1997; Hooper and van Soest, 2002; Hooper et al., 2002; Manuel et al., 2002; Reiswig, 2002).

Classically, three classes were distinguished in Porifera: Demospongiae Sollas, 1885, Hexactinellida Schmidt, 1870 and Calcarea Bowerbak, 1864. In Hexactinellida the spicules are characterized by a triaxonic symmetry, and the living tissue forms a syncytium. In Demospongiae, the symmetry of the spicules is monaxonic or tetraxonic. In the class Calcarea the skeleton consists of calcareous spicules, which in contrast to the siliceous spicules of the other sponge classes are extracellular formations (van Soest, 1996; Manuel et al., 2003; Manuel, 2006). The diversity of Demospongiae, with about 6.000 described species, is much higher than in Hexactinellida and Calcarea, which each comprises ca. 500 described species (Hooper and van Soest, 2002; Manuel et al., 2002; Reiswig, 2002).

Implications of from molecular phylogenetic studies on the relationships within sponges

The knowledge about relationships between and within the sponge classes still remains unclear (Hooper et al., 2002), although molecular data has recently provided additional insights, e.g., the sponges of the subclass Homoscleromorpha Lévi, 1973 were found not to be included in Demospongiae. Thus Porifera should be distinguished into four groups, Demospongiae *sensu stricto* (Borchiellini et al., 2004), Homoscleromorpha, Hexactinellida and Calcarea.

The question of sponge paraphyly

Several molecular sudies also suggested that the phylum Porifera itself is not a monophyletic group (e.g., Lafay et al., 1992; Collins, 1998; Zrzavy et al., 1998; Cavalier-Smith et al., 1996), e.g., with either Calcarea or Homoscleromorpha being more closely related to Eumetazoa than to other classes of Porifera (Borchiellini et al., 2001; Sperling et al. 2007) (for an overview, see Erpenbeck and Wörheide, 2007). Sponge paraphyly recieved quite some attention, because of its implication that the last common ancestor of Metazoa would have been a sponge-like organism (Halanych, 2004; Peterson and Butterfield, 2005; Nielsen 2008). However, the bootstrap support in

molecular phylogenies supporting paraphyletic sponges remained low. Very recently, a comprehensive phylogenomic study with a dataset of 128 protein coding genes found monophyletic Porifera with high support, and suggested the results from the previous studies may have suffered from insufficient taxon- and data-sampling (Philippe et al., 2009). The study also provided insight into inter-class relationships of Porifera, with Demospongiae *sensu stricto* and Hexactinellida forming a sister clade to Calcarea and Homoscleromorpha (Philippe et al., 2009).

Examples for implications of DNA studies on the taxonomy of sponge classes

Molecular analyses contributed to our understanding of the phylogenetic relationships within poriferan classes. In Hexactinellida, phylogenetic trees are in most parts concordant with the morphology-based taxonomic system (Dohrmann et al., 2008; Dohrmann et al., 2009). In Demospongiae *sensu stricto*, several classically recognized taxa were found to be artificial based on the analyses of molecular data, e.g., the subdivision of Demospongiae into the subclasses Ceractinomorpha and Tetractinomorpha (Borchiellini et al., 2004), or the order Haplosclerida and many families and genera within (McCormack et al., 2002; Erpenbeck et al., 2004; Redmond et al., 2007), to give just two examples (for a more detailed overview, see Erpenbeck and Wörheide, 2007). Several additional taxa of demosponges have yet to be resolved by molecular analyses.

The classification of the class Calcarea is the most challenging amongst the Porifera (Manuel et al., 2003). Here, the taxonomy is largely based upon typological observations and untested hypotheses about the evolution of certain morphological traits (Borojevic et al., 1990; Borojevic et al., 2000; Manuel, 2006). Moreover, phylogenetic analyses of morphological data showed very little resolution and suggested a high amount of homoplasy (Manuel et al., 2002). Analyses of ribosomal RNA genes on the other hand, while supporting the subdivision into the subclasses Calcinea and Calcaronea, questioned the monophyly of several taxa on order, family and genus level (Manuel et al., 2002; Manuel et al., 2004; Manuel, 2006).

Molecular evolution of rRNA in selected taxa of Porifera

With the first part of this thesis, I aim to gain further inside of how DNA analyses can be used to resolve the evolution of difficult groups of Porifera. For this purpose, I chose to investigate how the analyses of the most commonly used phylogenetic marker – the nuclear encoded ribosomal RNA (rRNA) genes (rDNA) – can contribute to adress phylogenetic questions in these taxa and how these analyses can be optimized. Because the nature and evolution of rRNA genes have consequences to their application in phylogenetic analyses, their traits require more detailed considerations.

Ribosomal RNA: form and function

As part of the ribosome, rRNAs are involved in the translation of mRNAs into polypeptides (Lafontaine and Tollervey, 2001). The translation process is pivotal for all life forms, even viruses that lack this translational machinery are dependent on their host's capability to form functional proteins. Hence, ribosomes are present in Archaea, Eubacteria and Eukaryota. In Eukaryota, the

organelles that were gained by endosymbiosis sometimes still carry parts of the translational toolkit from the former endosymbiont (see below), e.g., we find rRNA and a set of tRNA genes in mitochondrial genomes (Lang et al., 1999).

The ribosome has two subunits, each consisting of rRNA(s) and a number of ribosomal proteins (Lafontaine and Tollervey, 2001). The rRNA provides a structural backbone for the ribosomal proteins, but also form the main structures of the functional sites of the ribosome. The ribosome therefore is a ribozyme (Nissen et al., 2000; Noller, 2005). The number of rRNAs and ribosomal proteins in the ribosome differ, e.g., between eukaryotes and Eubacteria. In eukaryotes, the nuclear (nc) small subunit (SSU) contains the Small Subunit (SSU) rRNA (also called 18S rRNA), and, in *Saccharomyces cervisiae*, 21 ribosomal proteins (Lafontaine and Tollervey, 2001). The nc large subunit (LSU) in eukaryotes is formed from the large subunit (LSU) rRNA (also called 28S rRNA), and the smaller 5S and 5.8S rRNAs and a number of ribosomal proteins (46 in *S. cervisiae*, Lafontaine and Tollervey, 2001). In the translational processing, the SSU is involved in decoding the messenger RNA, while the LSU catalyses the formation of peptide bonds (Lafontaine and Tollervey, 2001).

The structure of the rRNA molecules is crucial for their function in the ribosome. The three dimensional shape of RNA is mainly maintained by hydrogen bonds of nucleotides of complementary strands forming helices. Hairpin-loops are found at the end of such helices. Unpaired insertions on one or both strand of a helix are called bulges, and junctions connect different helices (Smit et al., 2007). The secondary structures of rRNAs have first been inferred by comparative approaches (e.g., Woese et al., 1980; Noller et al., 1981; Woese et al., 1983; Gutell, 1993; Wuyts et al., 2000; Wuyts et al., 2001), and were later confirmed by X-ray crystallography of the whole ribosome, allowing very detailed three dimensional visualizations of the ribosome and its components (Clemons et al., 1999; Ban et al., 2000; Wimberly et al., 2000; Schluenzen et al., 2000).

Organization and concerted evolution of rRNA genes

In eukaryotes, nuclear rRNAs are encoded as tandem repeats on one or more chromosomes in a region called the nucleolar organizer region (Elder and Turner, 1995), with exception of the 5S rRNA, which is encoded elsewhere. In this nucleolar organizer region, the rRNA genes are co-linearly arranged in transcriptional units or cistrons. Each cistron consists in the following order of 5' external transcribed spacer (ETS), SSU (18S) rDNA, internal transcribed spacer 1 (ITS1), 5.8S rDNA, internal transcribed spacer 2 (ITS2), LSU (28S) rDNA and the 5' ETS (Fig 0.1). This complete

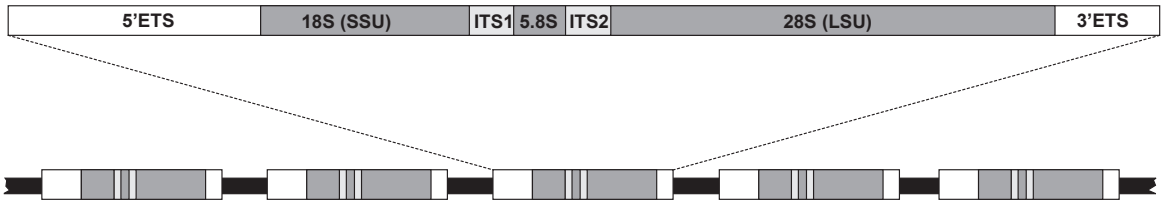


Figure 0.1: Organization of the nuclear RNA cistrons and their occurrence as tandem repeats on the nucleolar organizer region (genes not drawn to scale). See text for further details.

sequence is transcribed as a whole to a pre-rRNA, from which the mature rRNA molecules are derived by splicing of the external and internal spacer regions (Lafontaine and Tollervey, 2001).

The tandem repeats of this rRNA cistron do not evolve independently from each other. Instead the copies in a genome are homogenized, and differences between species are higher than within a species or individual, where the copies are almost identical (Liao, 1999). As mechanism for such a concerted evolution of the multi-copy rRNA, unequal crossing-over and gene conversion have been proposed (Elder and Turner, 1995). In the case of rDNA, concerted evolution provides a 'quality control' for functional, i.e. almost identical rRNAs, that is necessary to guarantee consistent functionality of the ribosomes (Liao, 1999), thereby inhibiting the creation of paralogous copies of rDNA. It can be assumed that such suboptimal versions of rRNA would interfere in the translational process, therefore being deleterious and selected against. Without the homogenization of the multiple copies of rRNA genes in genomes, these genes could not be used in phylogenetic studies.

Ribosomal RNA genes as phylogenetic marker

rRNA genes have very conserved as well as highly variable regions (see e.g., Ben Ali et al., 1999). The more conserved regions have been used in phylogenetic analyses to address questions on how the kingdoms of life are related to each other (which was an impossible task by means of morphology), while more variable parts still provide enough phylogenetic signal to study closely related species (Hillis and Dixon, 1991). The high abundance of rRNA in cells even made direct RNA sequencing possible (Hillis and Dixon, 1991), but today PCR provides the most efficient method to amplify and sequence these genes. The numerous, almost identical copies of rRNA genes allow easy PCR amplification with primers in the conserved regions of the gene, and are one reason for the popularity of rRNA as phylogenetic maker. Additional benefit for molecular systematics with rRNA genes comes from the concerted evolution: it lowers the variability of rDNA within species, in comparison to other, single copy genes, in which the intraspecific variability can be considerably higher (Hillis and Dixon, 1991).

Coevolution of paired sites: implications for phylogenetics

Because the secondary structure of rRNAs is formed by hydrogen bonds between nucleotides, it is directly dependent on the primary RNA sequence. Nonetheless, the primary sequence can vary, as long as the secondary structure and the function of the mature rRNA is not affected. Indeed, the secondary structure of rRNA is much more conserved than its primary sequence (Higgs, 2000). In RNA, the secondary structure is mainly maintained by the canonical standard Watson-Crick (CG, GC, UA, AU) and canonical (GU, UG) base pairs (Higgs, 2000). Paired nucleotides in helices coevolve to maintain the secondary structure. Often, so-called compensatory base changes in RNA helices can be observed (Dixon and Hillis, 1993), e.g., the change of a CG pair in one sequence to UA in another sequence. This implies a double substitution: C to G at the 5' position, and G to A at the 3' position. However, it has been suggested that this double substitution is really a two-step process through a slightly deleterious (because of its weaker hydrogen bond) intermediate doublet. Such intermediates are supposed to be short lived and occur with a low frequency in the

population of rRNAs and therefore are only rarely observed in real data (Rousset et al., 1991; Higgs, 2000).

The co-evolution of the bases in a doublet violates the assumption of independent evolution of sites that is made by most phylogenetic methods (Hancock et al., 1988; Higgs, 2000; Savill et al., 2001; Galtier, 2004). Ignoring this assumption can lead to suboptimal tree topologies, and yield misleading support values (Galtier, 2004; Telford et al., 2005; Erpenbeck et al., 2007a). A solution to this problem exists in form of special doublet models of nucleotide substitution. Here, two paired nucleotides (a doublet) are treated as one character rather than single nucleotides as in standard 4x4 models. A number of such doublet models have been described (Schöniger and von Haeseler, 1994; Muse, 1995; Tillier and Collins, 1995; Tillier and Collins, 1998; Savill et al., 2001).

Although standard Watson-Crick and GU pairs are the strongest pairs in RNA and have the highest frequencies, weaker hydrogen bonds can be formed between all possible nucleotide combinations (for an overview, see Lee and Gutell, 2004). Because these non-standard pairs can be present, there are 16 possible doublets in RNAs, resulting in a 16x16 matrix to calculate the likelihood in phylogenetic inference. When all doublet frequencies and doublet substitution rates are assumed to be independent, this results in a very parameter-rich model (15 free frequency and 119 free rate parameters), which, due to its computational demands, is unpractical to use in phylogenetic inference (Savill et al., 2001). Therefore, restrictions have been proposed by reducing the rate or frequency parameters by pooling some of them into classes or disallowing e.g., double substitutions (i.e. changes of both nucleotides in a doublet). Moreover, the less stable non-standard base pairs (other than GC, CG, AU, UA, GU, UG) are much less frequent in real data. In so-called 7-state models, such rare doublets are therefore pooled into a single class (referred to as mismatches, MM). In 6-state models, the rare doublets are ignored completely. The comparison of the performance between these three groups of doublet models has been impossible or were unclear for real data (Savill et al., 2001).

Molecular evolution of mitochondrial genomes

In the second part of this thesis, the evolution of mitochondrial (mt) genomes in Porifera and the special genome organization of *Hydra magnipapillata* (Cnidaria, Hydrozoa) were studied to obtain further insight into the evolution of organellar genomes in the non-bilaterian taxa.

Mitochondrial genome evolution in non-bilaterian Metazoa

Mitochondria are organelles in the eukaryotic cell responsible for the oxidative phosphorylation of adenosine di-phosphate to adenosine tri-phosphate, the main energy source to sustain the biochemical functions of cells. Mitochondria originate from an α-proteobacterial ancestor, which was an endosymbiont in pre-eukaryotic cells (Gray et al., 1999; Burger et al., 2003b). Today mitochondria still carry their own, but strongly reduced genome: many of the former symbionts genes have been transferred to the host's nucleus (Henze and Martin, 2001; Timmis et al., 2004). The size of the mt genome is very variable and not always an indication for the gene content, because of non-

coding regions that can be present. Most eukaryote mt genomes are about 15 to 60 kb and encode 5 to 100 genes in different organisms. All mt genomes include genes for proteins involved in respiratory and/or oxidative phosphorylation (Burger et al., 2003b). Other genes code for proteins or RNAs, which play a role in mt translation, transcription, RNA maturation or protein import from the cytoplasm (Burger et al., 2003b). Typically, animal mt genomes have been expected to comprise a ca. 16 kb circular DNA molecule, with genes for 13 proteins, two rRNAs and 22 tRNAs (Boore, 1999). This, however, mainly refers to bilaterian mt genomes, and recently more non-bilaterian mt genome sequences were published and shed a new light on the evolution of animal mt genomes (Lavrov, 2007). Today, mt genomes from the non-bilaterian phyla Porifera (Haen et al., 2007; Lavrov and Lang, 2005; Lavrov et al., 2005; Lukić-Bilela et al., 2008; Rosengarten et al., 2008; Wang and Lavrov, 2007; Wang and Lavrov, 2008), Cnidaria (Brugler and France, 2008; Sinniger et al., 2007; Kayal and Lavrov, 2008; Brugler and France, 2007; Shao et al., 2006; Medina et al., 2006; Beagley et al., 1998) and Placozoa (Dellaporta et al., 2006; Srivastava et al., 2008) are available, but no mt genome from phylum Ctenophora. Several differences to bilaterian mt genomes have been discovered, e.g., the occasional occurrence of additional genes of known or unknown function (e.g., Pont-Kingdon et al., 1998; Pont-Kingdon et al., 1995; Dellaporta et al., 2006; Srivastava et al., 2008), a large amount of non-coding sequence (e.g., in Placozoa, Dellaporta et al., 2006; Srivastava et al., 2008), reduction of tRNAs in Cnidaria (e.g., Beagley et al., 1998; Medina et al., 2006). In Cnidaria, or more specifically in Medusozoa, the mt genome is not a circular DNA molecule as usually in Metazoa, but linear and additionally may be subdivided into two or presumably even more mt chromosomes (Warrior, 1987; Pont-Kingdon et al., 2000; Bridge et al., 1992; Ender and Schierwater, 2003).

Goals

Three case studies were conducted to address the question of how incorporating the aforementioned particularities of rRNAs can contribute to the inference of phylogenies in taxonomically challenging groups of Porifera. In **chapter 1**, I determined the secondary structure of SSU rRNA for all available poriferan sequences, as this is a prerequisite for the application of doublet models in phylogenetic inferences. Special attention was paid to the evolution of marine Haplosclerida. In this taxon, large discrepancies with the classical system had been uncovered through molecular analyses (McCormack et al., 2002; Raleigh et al., 2007; Redmond et al., 2007). Additionally, several of the species of this group were found to possess longer, hyper-variable insertions in the SSU rRNA gene (Redmond et al., 2007), which had diverged too strongly to be unambiguous aligned. These insertions had therefore been neglected in standard phylogenetic inferences. However, because the secondary structure of rRNA is generally more conserved than the primary sequence, we studied their evolution and evaluated the phylogenetic signal present in the secondary structures of such hyper-variable insertions.

In **chapters 2 and 3**, doublet models were applied in analyzes of SSU and LSU rRNA data to clarify relationships within the class Calcarea. In **chapter 3**, I compared the performance of different doublet models in order to find out what model fits the rRNA data best, and to consider the con-

sequences of suboptimal model choice on node support and tree topology.

In order to understand the evolution of mt genomes in non-bilaterian Metazoa, two complete mt genome sequences were determined in this work. The mt genome of *Amphimedon queenslandica* (Demospongiae, Haplosclerida) contributed to our understanding of mt genome diversity in the phylum Porifera (**chapter 4**). The nature of the mt genome of *Hydra magnipapillata* – with linear mt chromosomes of roughly the same size – was known from electrophoresis experiments (Warrior, 1987; Pont-Kingdon et al., 2000). However, a detailed understanding of the consequences of this subdivision on the mt chromosome organization was lacking. By determining the complete sequence of both mt chromosomes (**chapter 5**), I aimed to provide further insight into the nature and evolution of such fragmented organellar genomes.

Chapter 1

Molecular evolution of rDNA in early diverging Metazoa: First comparative analysis and phylogenetic application of complete SSU rRNA secondary structures in Porifera

Oliver Voigt[1], Dirk Erpenbeck[1] and Gert Wörheide[1,*]

[1]Dept. of Geobiology, Geoscience Centre Göttingen, University of Göttingen, D-37077 Göttingen, Germany

* Corresponding author

This version of the article was published in BMC Evolutionary Biology [24(1):19–22. 2007, (doi:10.1093/molbev/msl154)] and as Advance Access on October 19, 2006.

Abstract

Background

The cytoplasmic ribosomal small subunit (SSU, 18S) ribosomal RNA (rRNA) is the most frequently-used gene for molecular phylogenetic studies. However, information regarding its secondary structure is neglected in most phylogenetic analyses. Incorporation of this information is essential in order to apply specific rRNA evolutionary models to overcome the problem of co-evolution of paired sites, which violates the basic assumption of the independent evolution of sites made by most phylogenetic methods. Information about secondary structure also supports the process of aligning rRNA sequences across taxa. Both aspects have been shown to increase the accuracy of phylogenetic reconstructions within various taxa.

Here, we explore SSU rRNA secondary structures from the three extant classes of Phylum Porifera (Grant, 1836), a pivotal, but largely unresolved taxon of early branching Metazoa. This is the first phylogenetic study of poriferan SSU rRNA data to date that includes detailed comparative secondary structure information for all three sponge classes.

Results

We found base compositional and structural differences in SSU rRNA among Demospongiae, Hexactinellida (glass sponges) and Calcarea (calcareous sponges). We showed that analyses of primary rRNA sequences, including secondary structure-specific evolutionary models, in combination with reconstruction of the evolution of unusual structural features, reveal a substantial amount of additional information. Of special note was the finding that the gene tree topologies of marine haplosclerid demosponges, which are inconsistent with the current morphology-based classification, are supported by our reconstructed evolution of secondary structure features. Therefore, these features can provide alternative support for sequence-based topologies and give insights into the evolution of the molecule itself. To

encourage and facilitate the application of rRNA models in phylogenetics of early metazoans, we present 52 SSU rRNA secondary structures over the taxonomic range of Porifera in a database, along with some basic tools for relevant format-conversion.

Conclusions

We demonstrated that sophisticated secondary structure analyses can increase the potential phylogenetic information of already available rDNA sequences currently accessible in databases and conclude that the importance of SSU rRNA secondary structure information for phylogenetic reconstruction is still generally underestimated, at least among certain early branching metazoans.

Background

Tens of thousands of sequences of the small subunit ribosomal RNA (SSU rRNA, 18S) gene of eukaryotes have accumulated in public databases such as NCBI GenBank (http://www.ncbi.nlm.nih.gov/), making this gene one of the first and most frequently used markers for molecular phylogenetics. Its popularity is due to a high degree of conservation in some regions of the molecule, in combination with a considerable amount of variability in others. These features enable phylogenetic questions to be addressed between relatively closely related taxa, as well as between different domains of life (Higgs, 2000). Therefore, analyses of SSU rRNA sequences have a long history, and new sequences are still being continuously generated.

SSU rRNA molecules fold into a specific secondary structure, which is essential for maintenance of their three dimensional structure and their function within the ribosome (Green and Noller, 1997), but which also has consequences for the use of rRNA molecules in phylogenetics. The secondary structure of rRNAs is maintained by hydrogen bonds between RNA nucleotides, which form helices (or stems). These helices are interleaved by regions consisting of unpaired nucleotides, forming loops at the end of a helix and bulges within different helices. Secondary structure of RNAs is generally much more conserved than their primary sequence (Higgs, 2000). Therefore, considering this structure during multiple sequence alignment can greatly improve the assignment of homologous positions, consequently resulting in more probable phylogeny estimations (e.g., Kjer, 1995; Hickson et al., 2000; Gillespie et al., 2005b). Furthermore, paired nucleotides (= doublets) frequently co-evolve in order to maintain rRNA structure and function. The co-evolution of doublets violates the assumption of independent evolution of sites made by most phylogenetic methods (Dixon and Hillis, 1993). Consequently, specific evolutionary models have been proposed for paired sites and have been shown to outperform standard (4 x 4) nucleotide models (Schöniger and von Haeseler, 1994; Tillier and Collins, 1995; Muse, 1995; Tillier and Collins, 1998; Telford et al., 2005; Dohrmann et al., 2006; Erpenbeck et al., 2007a). However, secondary structure models are still rarely used in phylogenetic analyses, presumably because establishing a secondary structure for a new sequence is still a time-consuming exercise even for the conserved core structure of SSU rRNA, and very few software packages allow the simultaneous analysis of paired and unpaired rRNA regions. Some rRNA databases (as the Comparative RNA Web Site and Project

http://www.rna.ccbb.utexas.edu, Cannone et al., 2002; and the The European ribosomal RNA database, http://bioinformatics.psb.ugent.be/webtools/rRNA/, Wuyts et al., 2004) provide secondary structure information for a number of organisms, but their records are far from complete and structures of hypervariable insertions are usually not presented, or are only presented to a certain extent. In particular, the lower Metazoa, which are pivotal for the understanding of animal evolution, are still under-represented in databases.

One key taxon for early metazoan evolution is Phylum Porifera (sponges), in which the relationships are unresolved at all taxonomical levels, even between the three extant sponge classes Demospongiae, Calcarea (calcareous sponges) and Hexactinellida (glass sponges). Within sponge classes, the results of molecular phylogenies are often incongruent with morphological expectations (e.g., Manuel et al., 2003; Borchiellini et al., 2004; Nichols, 2005; Erpenbeck et al., 2006; Dohrmann et al., 2006). In this study, we performed the first comprehensive survey of the complete SSU rRNA secondary structures of representatives of the main lineages of phylum Porifera, and evaluated how secondary structure information and features other than the primary sequence can contribute to improve phylogenetic reconstructions. For these purposes, we considered all available SSU rRNA sequences of Porifera, inferred their secondary structures (a selection of which we are presenting in a new database), and analyzed base compositions and sequence lengths. We reconstructed a phylogeny with partitioned phylogenetic analyses using specific rRNA models of nucleotide evolution for paired sites. Using this backbone, we assessed the phylogenetic value of secondary structures of unique insertions found in a specific demosponge clade (Order Haplosclerida), which would usually be disregarded as 'unalignable sites' and thus excluded from standard phylogenetic analyses.

Methods

Sequence acquisition, analyses and inference of secondary structures

We analyzed all 170 published full or nearly full-length SSU rRNA sequences of Porifera (see Appendix 1, Table A1.1 for a complete listing). For taxonomy of the taxa included in our study we followed Systema Porifera (Hooper and van Soest, 2002) and the World Porifera Database (http://www.marinespecies.org/porifera/), where also the species authorities are available. The SSU rRNA sequence of *Amphimedon queenslandica* was reconstructed by performing a local Blast search (Altschul et al., 1997) against data from GenBank's trace archive. Traces from significant hits (see Appendix 1, Fig. A1.1) were downloaded and assembled in CodonCode Aligner 1.6.3 (http://www.codoncode.com/). This resultant sequence can be downloaded from our database of SSU rRNA secondary structures of Porifera (http://www.palaeontologie.geo.lmu.de/molpal/RRNA/fasta/Amphimedon_queenslandica.fasta). For Class Hexactinellida, only limited data was available in GenBank: All three full-length SSU rRNA sequences belong to Subclass Hexasterophora. Two additional hexactinellid sequences were provided by Martin Dohrmann ahead of their publication in a comprehensive phylogenetic study of Hexactinellida (Dohrmann et al., 2008): *Semperella schulzei* (subclass Amphidiscophora) and *Aphrocallistes vastus* (Subclass Hexasterophora).

All sequences were initially aligned with CLUSTAL W 1.83 (Thompson et al., 1994) and the preliminary alignments were manually improved in SeaView (Galtier et al., 1996). Gblocks 0.91b (Castresana, 2000) was used to identify and isolate the conserved sites of the alignment before clustering similar sequences using the Neighbor Joining (NJ) algorithm in PAUP* 4.0b10 (Swofford, 2003). Secondary structures for resulting clades were established for certain representatives of the clade by aligning to known structures from the European RNA Database (Wuyts et al., 2002; Wuyts et al., 2004) in separate alignments for each clade and considering compensatory base changes. SSU rRNA clade-alignments were then further refined according to secondary structure information.

The unusual structures of marine Haplosclerida (=Order Haplosclerida excluding Suborder Spongillina) and Hexactinellida (including conserved flanking regions with known structure) were initially examined under minimum free energy predictions from the mfold-server (http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi). In most cases, only one structure was predicted by the algorithm. If multiple structures were predicted, we chose the structure with either the minimal free energy or with the best compatibility to similar sequences.

A comparative approach (see e.g., Gillespie et al., 2005a) was chosen if permitted by an appropriate level of sequence divergence. For this approach, we used the alifold server (http://rna.tbi.univie.ac.at/cgi-bin/alifold.cgi) to infer secondary structures of the insertions. Alifold infers secondary structures by considering both minimum free folding algorithms and compensatory base changes, and therefore includes additional information that provides hints for secondary structural motifs. Since this method requires a correct alignment, it could only be used if sequences were not too divergent from each other, such as with a subset of marine Haplosclerida (Demospongiae) and the insertions of Hexasterophora (Hexactinellida) (Appendix 1, Fig. A1.2). However, secondary structures inferred with both methods were identical, or only differed in a few positions (Appendix 1, Fig. A1.2). Therefore, while the comparative method is preferred, we still found that minimum free energy based predictions performed adequately to be used in cases where unambiguous alignments or missing comparative data does not allow inference of secondary structures based on compensatory base exchanges. For taxa that were suitable for a comparative approach, compensatory base exchanges are presented together with the corresponding alignments in Appendix 1, Fig. A1.2.

We visualized selected structures by converting the sequence and structure information to a ct-format with a Perl-script. This format can be displayed in RNAviz 2 (De Rijk et al., 2003). Helix names correspond to Wuyts et al. (2000), with the exception of helices E23_1 and E23_2, which together are referred to as E23_1. Insertions are designated by the name of the conserved helix in which they occur, and a period plus the number of the additional helix is added: Parts of conserved helices separated by insertions are named after the original helix followed by a letter (e.g., one helical insertion within E23_1 will be called E23_1.1, the 5' part of the helix before the insertion will be called E23_1a, the 3' part after the insertion E23_1b).

Base compositions and the lengths of the secondary structure features were calculated with a custom-made Perl script. To avoid biases introduced by missing data from the published se-

quences, we used a fragment (corresponding to ca. 95% of SSU rRNA) spanning from helix 5 until 2bp before helix 50 (i.e. positions 48-1896 in *Amphimedon queenslandica*), and only considered the 123 sequences without data missing within this region (listed in Appendix 1, Table A1.2). Representative poriferan secondary structures are available as *.fasta-format (with bracket-dot annotation) and in *.ct-format from our database for SSU rRNA secondary structures of Porifera (http://www.palaeontologie.geo.lmu.de/molpal/RRNA/index.htm). Furthermore, several Perl scripts (written for Mac OS X/Linux) for format conversion are provided (along with other scripts: Tools for conversion from annotated alignments to ct-format and vice versa, and from alignments to MrBayes or PHASE data-files containing the secondary structure information are included).

Phylogenetic analyses

The secondary structure information from the previous step was used to generate a new alignment in SeaView. We generated a taxon-set comprising of 78 taxa (for accession numbers see http://www.biomedcentral.com/content/supplementary/1471-2148-8-69-s5.mase) and focused on relationships of haplosclerid demosponges, in a similar way to Redmond et al. (2007). The SSU rRNAs from this diverse taxon have been found to possess numerous insertions and extensions and our aim was to unravel their evolution.

Sites with uncertain homology even after considering secondary structure were excluded from the phylogenetic analyses. This was achieved by assigning sites to two groups and discarding those sites that were regarded as ambiguously aligned by the following criteria:

1. Unpaired sites: with length polymorphism and sequence divergence too high to identify homologous positions for all sequences. (Bulge after 3' helix 8; loops of helices 6, 10, E10_1,11,17, E23_12, 29, 44, 49)

2. Paired sites: with length polymorphisms in helices and/or structural homologies that could not be unambiguously assigned (e.g., in cases of elongation of helices, parts of helices 10, E10_1, E23_1/E23_2, 49).

Furthermore, taxon-specific insertions within helices (found in some marine Haplosclerida), as well as nucleotide insertions found only in single sequences were excluded.

Doublet positions were only regarded as pairings in the consensus secondary structure if the two involved nucleotides formed a Watson-Crick (G-C, A-U) or G-U wobble pairing in at least five sequences within the alignment. Corresponding sites falling below this five-sequence threshold were treated as unpaired. For phylogenetic reconstructions, sites were allocated to one of the following two partitions: Partition 'stem' (= paired sites) or partition 'loop' (= unpaired sites). We used MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) and PHASE 2.0 (http://www.bioinf.manchester.ac.uk/resources/phase/index.html) for the phylogenetic analyses, as both programs allow the simultaneous analysis of a partitioned dataset with both rRNA models for paired sites and standard models for unpaired sites.

MrBayes only allows the usage of a doublet model corresponding to the SH model (Schöniger and von Haeseler, 1994). This is a 16 state-RNA model, which considers all possible doublets as characters and assumes that compensatory base exchanges result from at least two substitution events. A GTR + G + I model (Tavaré, 1986) was assigned for the loop partition. The Markov chain Monte Carlo (MCMC) analysis comprised two runs (eight chains each) for 12.142 million generations, with the sample frequency set to 100 and the temperature for the heated chains set to 0.2. Sampled trees were summarized using the *sumt* command in MrBayes with a burn-in set to the first 2 million generations. Sufficient convergence of chains for the MrBayes runs was monitored by observing log-likelihood values, the standard derivation of split frequencies (> 0.008), and diagnostics provided by AWTY (http://king2.scs.fsu.edu/CEBProjects/awty/awty_start.php; Nylander et al., 2008).

In PHASE, we applied the RNA7A model (Higgs, 2000) and RNA7D model (Dixon and Hillis, 1993) for stem regions in independent runs. RNA7A is the most general 7-state RNA model. RNA7D (seven frequencies, four rate parameters) is a simplification of RNA7A (7 frequencies, 21 rate parameters). The 7-state RNA models treat all mismatches as one single state. This simplification increases the risk of loss of phylogenetic information, but the occurrence of mismatch-pairs in rRNA data is small, therefore, an estimation of mismatch substitution parameters from the data is probably not accurate (Savill et al., 2001). Furthermore, by pooling mismatches into a single character, the number of parameters to be estimated in a phylogenetic analysis, and consequently the computational demands are significantly decreased. For loop regions, the REV model (Tavaré, 1986) was chosen. In addition, a gamma distribution accounting for rate heterogeneity among sites and a proportion of invariant sites were assigned to each model for both partitions. Independent runs were performed in PHASE 2: Two runs with the RNA7A model (40 million generations) and one run (5 million generations) with the RNA7D model for stem positions. Every 100th generation a sample was taken from the MCMC chains (after a burn-in-phase of 1 million generations).

Tracer v1.4 (http://tree.bio.ed.ac.uk/software/tracer/) was used to monitor sufficient parameter stabilization. To create readable input files for Tracer from the PHASE runs, we used a slightly modified version of the perl script ‘phase2tracer.pl’ (originally programmed by Matt Yoder, http://hymenoptera.tamu.edu/rna/download.php), which is available upon request.

The presented tree topology is based upon one of the 40 million PHASE runs with the RNA7A model for stem partition (loop model as mentioned above). To obtain branch-lengths for the tree, we conducted an additional analysis (4 million generations) under the same models, and tree topology was fixed to the consensus tree from the original 40 million generation analysis as suggested in the PHASE manual (all other parameters unchanged).

Results

SSU rRNA length differences and base composition

To avoid biases due to missing data, we analyzed base composition and sequence length for a fragment of SSU rRNA that covers about 95 % of the gene (see Methods). Base composition and fragment length differed considerably among the 123 poriferan sequences (Fig. 1.1). The GC content varied between 45.5 and 56.3 %.

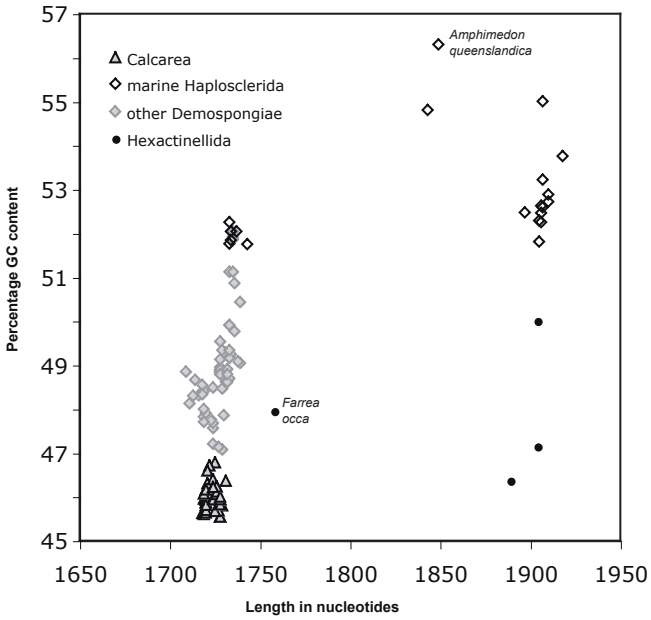


Figure 1.1: GC content against SSU rRNA fragment length (Fragment corresponds to *A. queenslandica* positions 48-1896). A ca. 95% -fragment of SSU rRNA was used for analysis and only sequences with sequence information over the whole range of this fragment were considered (n=123). Note that *Farrea occa* (Hexactinellida, [GenBank: AF159623]) is an incomplete potential pseudogene sequence.

Calcarea posses the lowest GC contents with a modest variation from 45.5 to 46.8 %. In this aspect they are clearly separated from demosponges, which display significantly higher GC contents, since the lowest demosponge value (47.1%) still exceeds the highest GC content (46.8%) of Calcarea (Fig. 1.1). Most demosponge SSU rRNAs show modest length variations in a range comparable to those of Calcarea. Notable exceptions are the extraordinary large rRNA molecules found in several marine haplosclerids. The highest GC contents of Porifera are also found within this group (with a maximum of 56.3% in *Amphimedon queenslandica*). The high GC pattern is independent of the presence of insertions in these large molecules, since members of marine Haplosclerida with smaller rRNA molecules also possess similar GC contents (Fig. 1.1). SSU rRNAs of the few available hexactinellid sponges are approximately equal in length to large molecules of several haplosclerids (with the exception of *Farrea occa* [GenBank: AF159623], see below). In contrast to haplosclerids, hexactinellid sponges have lower GC contents, with base compositions in the range of those of Demospongiae and Calcarea.

Secondary structure

Porifera have the typical eukaryotic core SSU rRNA structure (see Figs. 1.2-1.4). The moderate length variation between Calcarea and most demosponges is primarily caused by insertions in unpaired regions or by elongation of helices 10, E10_1 and 43 (Table 1.1). In Hexactinellida, on average, these three helices are largely elongated compared to Calcarea and Demospongiae (Fig. 1.3), but the lengths of the E10_1 helices of some demosponge sequences fall into the same range.

In addition, we observed extra-helical insertions in Hexactinellida and in several marine haplosclerid demosponges that are not part of the eukaryote core structure. In marine haplosclerids,

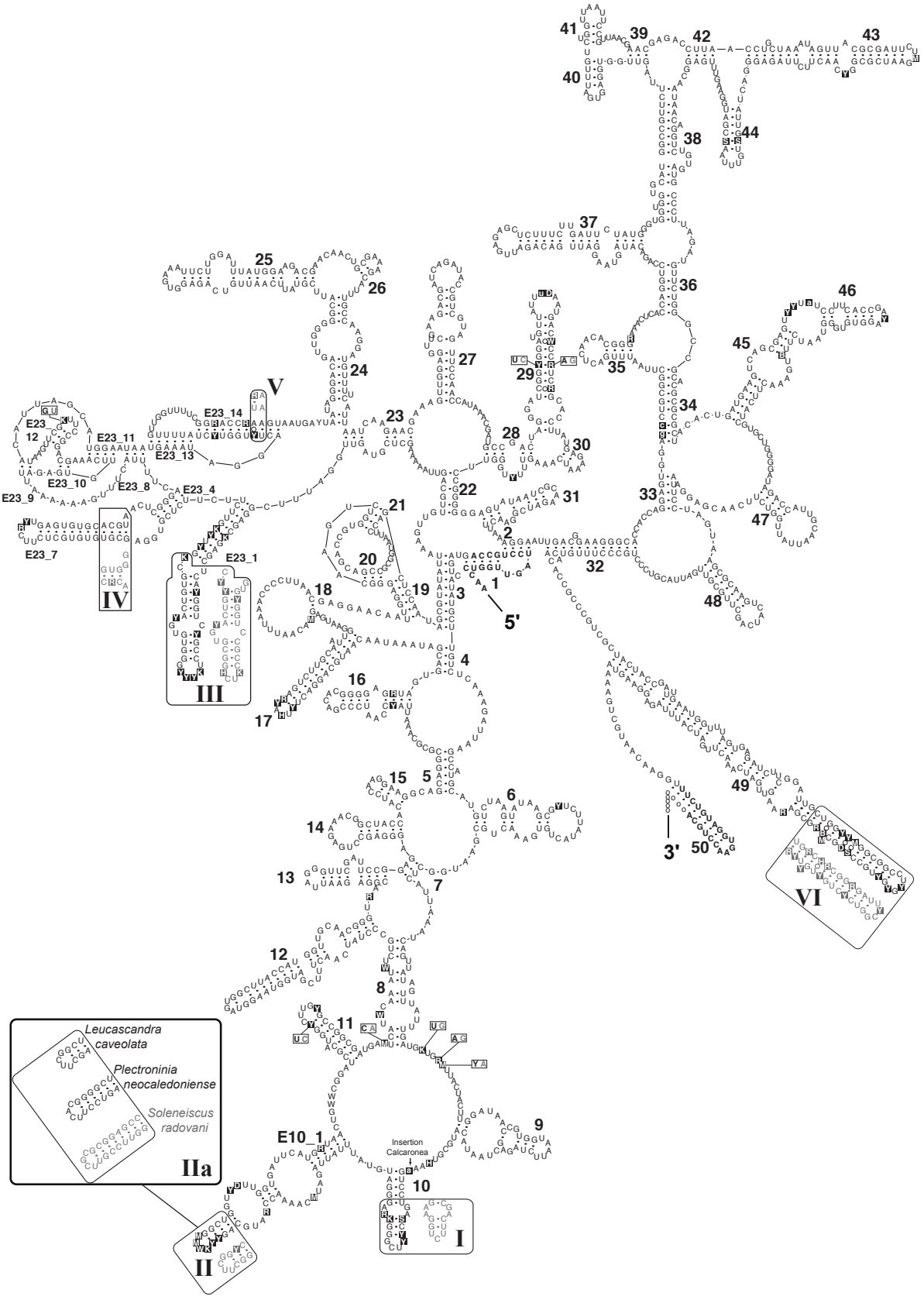


Figure 1.2: SSU rRNA secondary structure for Calcarea.Sequence is given as 90% consensus with variable positions in black boxes. Lower case indicates deletions at the site for some sequences, according to the consensus level. Differences in helices between Calcaronea and Calcinea are in frames (Calcaronea=black, Calcinea=grey). Synapomorphies for each subclass are shown in boxes with the same color code. Primer positions are bold at the 5' and 3' end, respectively. Open circles instead of dots mark positions where mismatches occur in some sequences. Inset: Shortening and elongations in the boxed part of Helix E10_1 for two calcaronean sequences and one calcinean sequence.

Table 1.1: Mean and range of the length of the most variable helices within the three sponge classes.

	Calcarea (n=48)		**Demospongiae** (n= 109*)		**Hexactinellid**a (n=5**)	
Helix	**bp (range)**		**bp (range)**		**bp (range)**	
10	22.8	(21-26)	21.9	(20-30)	36.0	(35-37)
E10_1	55.4	(54-64)	61.6	(55-74)	75.4	(67-80)
43	49.1	(49-50)	49.5	(48-99)	91.5	(72-101)

*Marine haplosclerids not included in Helix 43; **Farrea occa not included for helix 43.

these extra sequences were inserted within helices E23_1, E23_14 and 43, and in Hexactinellida, the insertions only occurred within helix E23_1 at a different position than in marine haplosclerids (Figs. 1.3, 1.4). All of the helices where sequence elongations and/or insertions occur belong to regions that are known to be highly variable within eukaryotes (see e.g., Wuyts et al., 2000; Wuyts et al., 2001).

Calcarea

The SSU rRNA of this sponge class comprises all of the typical eukaryote helices and lacks unusual structural features. A calcarean SSU rRNA consensus sequence and structure is shown in Fig. 1.Fig. 1.2. Several synapomorphies for the two Calcinea and Calcaronea subclasses were detected in the secondary structure. In Calcinea, helices 10 and E23_1 are shorter by at least one base pair when compared to Calcaronea (Fig. 1.2, insets I & III). In helix E10_1, Calcaronea typically have three pairs at the helix end, whereas Calcaronea dominantly possess four pairs (Fig. 1.2, inset II). However, independent elongations of this helix can be found in both subclasses (Fig. 1.2, inset IIa: Calcaronea: *Plectroninia neocaledoniense*; Calcinea: *Soleneiscus radovani*). These elongations are homoplasies as is evident when considering the subclass-specific compensatory base change (Calcaronea: A-U; Calcinea: G-C) at the beginning of inset II (Fig. 1.2): The A-U pair in the corresponding structure of the calcaronean *Leucascandra caveolata* supports a secondary loss of a pair compared to other Calcaronea. Differences in helix nucleotides between both subclasses occur in helices 11, E23_7, E23_14 and 29 (Fig. 1.2, and insets IV and V). Most of these changes maintain the helix-relevant pairings (e.g., in 11 or E23_7), but a few cause mismatches in at least some sequences (in E23_7, E23_14, 29 and 49). Base changes and insertions in unpaired regions are also specific for the Calcinea-Calcaronea split. This is indicated in Fig. 1.2 for three bases in the bulge between helices 8 and 9, one base within the loop of E23_12, and a calcaronean-specific insertion of one adenosine between helix 9 and 10.

Hexactinellida

The SSU secondary structure of *Acanthascus dawsoni* is presented in Fig. 1.3 representatively for Hexactinellida. In all hexactinellid sequences, specific insertions were observed (Fig. 1.3, inset). As mentioned previously, hexactinellid SSU rRNA sequences are considerably longer than in other

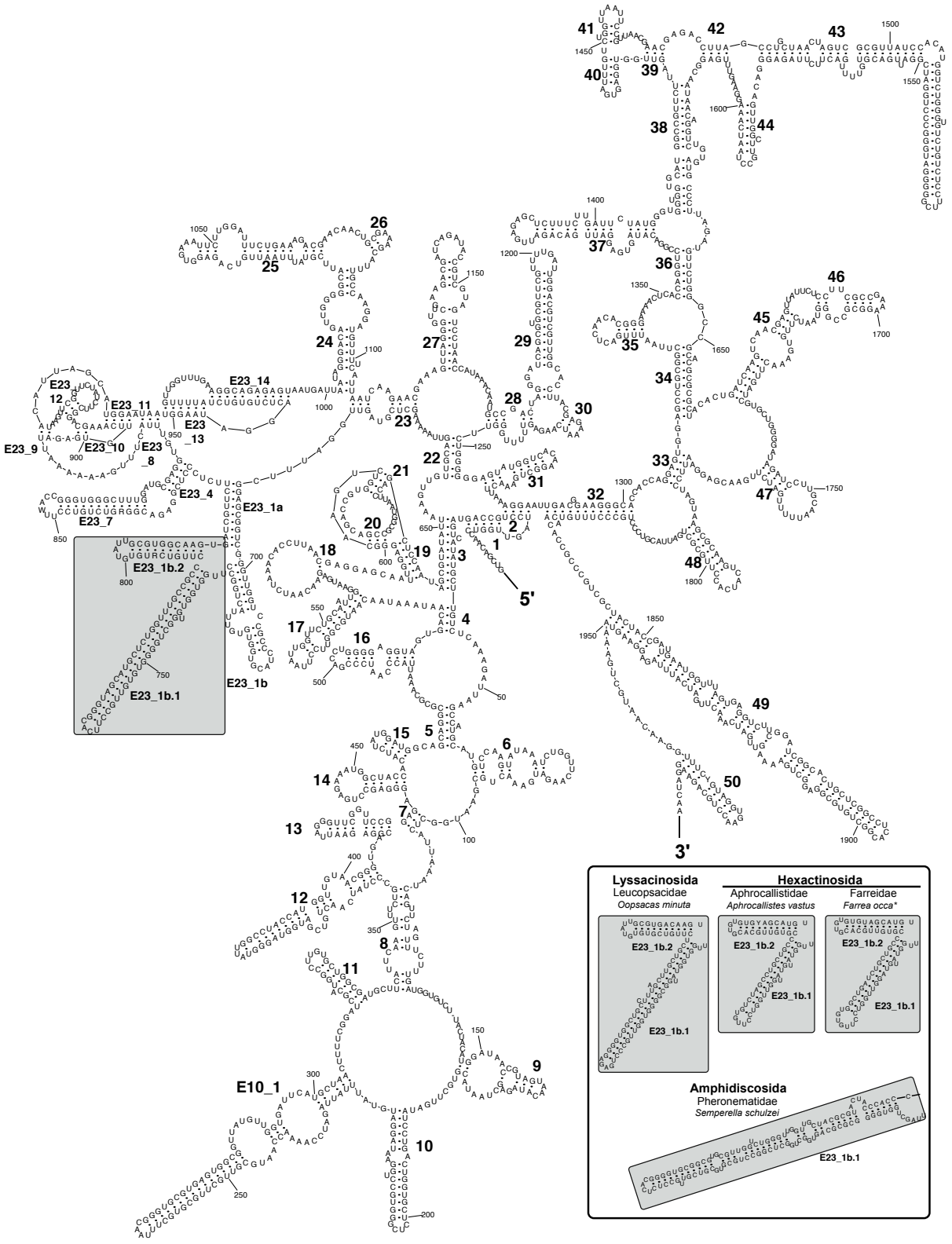


Figure 1.3: SSU rRNA secondary structure of *Acanthascus dawsoni* [GenBank: AF100949] (Lyssacinosida, Rossellidae). Hexactinellid-specific helical insertions within E23_1 are shown in a box. Inset: Prediction of secondary structure insertions in E23_1 within other Hexactinellida. The insertions are predicted to form two helices in Hexasterophora (Lyssacinosida + Hexactinosida), and one helix in Amphidiscophora (*Semperella schulzei*). *Note that *Farrea occa* (AF159623) represents an (in other than the displayed part) incomplete, potential pseudogene molecule.

poriferans, except in some marine haplosclerid demosponges (Fig. 1.1). The additional nucleotides occur in extensive elongations of common helices (10, E10_1, and 43; Table 1.1, Fig. 1.3), and a helical insertion in helix E23_1. The insertions in helix E23_1 occur at a unique position among sponges and may form two helices (assigned the names E23_1b.1 and E23_1b.2 in Fig. 1.3) in all the studied SSU rRNA molecules of subclass Hexasterophora. Helix E23_1b.2 contains 10 doublets and is much more conserved within Hexasterophora than Helix E23_1b.1, which varies in length from 37 to 55 bp. In contrast to Hexasterophora, *Semperella schulzei* (Subclass Amphidiscophora) has a helical insertion of 107 bp within E23_1, which is predicted to form a single helix E23_1b.1 (Fig. 1.3 inset).

Within the sequence of *Farrea occa* [GenBank: AF159623], we found deletions in conserved regions. Helices 13 and 15 are missing completely, as are the 3' strand of helix 7, parts of helix 43 and the 5' strand of helix 15 (compare Fig. 1.3). Such complete or partial deletion of conserved helices has been shown to be typical for non-functional rRNA pseudogenes (Pons and Vogler, 2005). Potential paralogs like this one are not necessarily subject to concerted evolution, and are therefore not suitable for phylogenetic inference. In this context, the consideration of secondary structure is crucial for identification of such non-functional sequences, and prevents biases in phylogenetic reconstruction due to potentially misleading data. Nonetheless, predictions of insertions for this sequence are displayed in Fig. 1.3 (inset), since no suspicious modifications were found within this part of the molecule and no other sequence of Farreidae was available. However, the results for this species should be treated with caution.

Demospongiae

Most demosponges possess a SSU rRNA molecule with the common metazoan secondary structure. Remarkable exceptions are only found within the marine Haplosclerida (Figs. 1.4-1.6), which possess insertions that are long enough to be predicted to form additional helices. Those helices are found within known variable regions for eukaryotes and appear in the 5’ strand of Helix E23_1/2, the 5’strand within Helix E23_14 and the 3’ strand of helix 43.

Phylogenetic analyses

We inferred the phylogeny of marine haplosclerids to compare the evolutionary history of helical insertions found in this group of Demospongiae (see section "Successive evolution of additional helices in marine haplosclerids"). Results from the PHASE- and MrBayes analyses of 78 taxa are shown in Fig. 1.5. Although more general 7-state models have been shown to result in higher likelihood values for phylogenies than less parameter-rich models for real rRNA data (Savill et al., 2001), our analyses with PHASE with the RNA7A model and the less complex RNA7D model yielded identical tree topologies (with almost identical support values). Independent runs in PHASE and MrBayes resulted in similar, almost identical topologies, and differences in demosponge relationships were only observed in the positions of clades with weak support values. Namely these are the relationship of Dictyoceratida to the Myxospongiae (sensu Borchiellini et al. 2004 [= clade Verongida +*Chondrosia reniformis*]), the position of *Scopalina ruetzleri* and relationships within freshwater sponges (where branch lengths were short, Fig. 1.5). Additionally, differ-

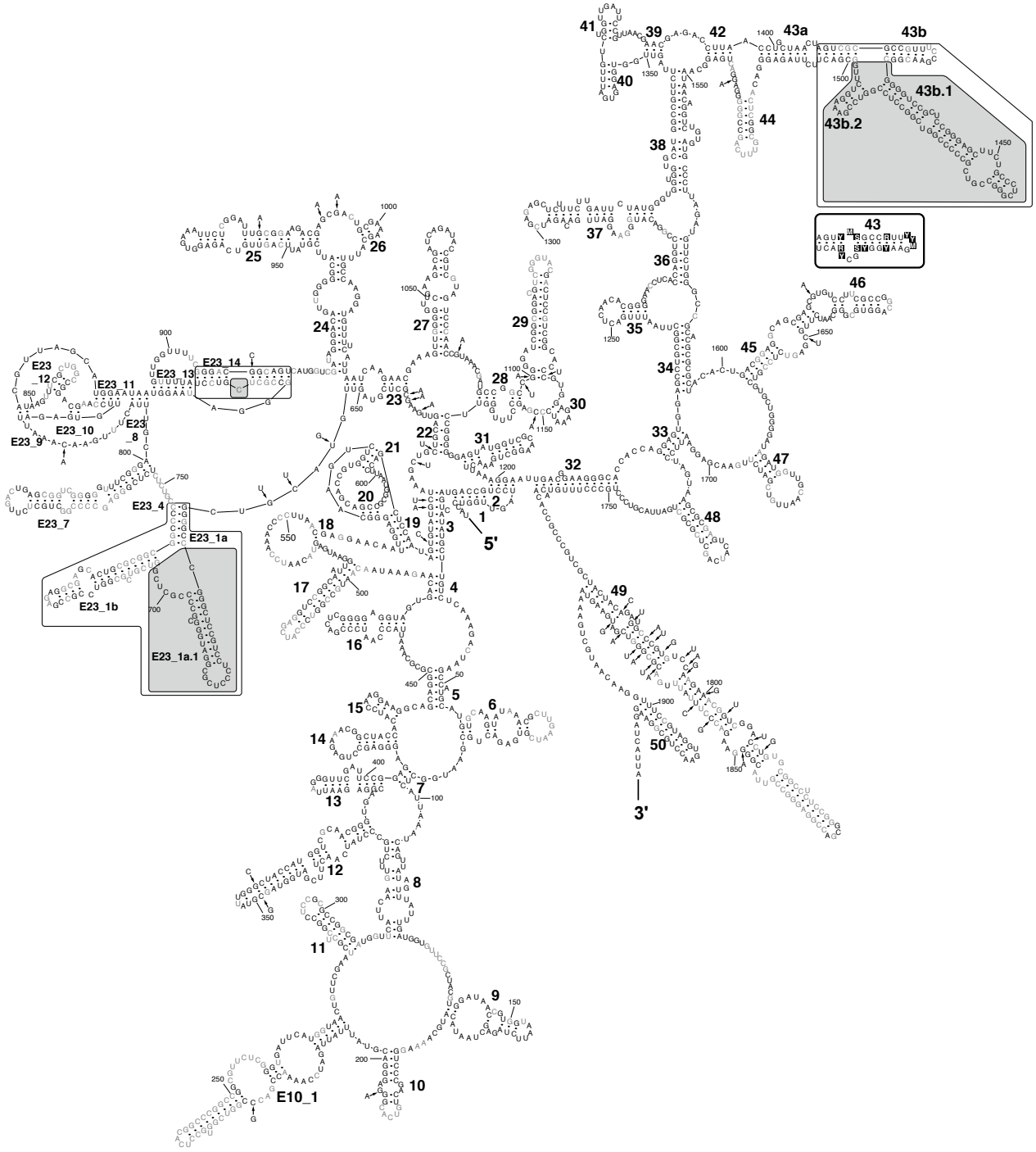


Figure 1.4: SSU rRNA secondary structure of the demosponge *Amphimedon queenslandica* (Haplosclerida). Nucleotides conserved in Demospongiae at the 90% level are shown in black, other nucleotides are in grey. Nucleotides at positions that are present in demosponges above the 90% consensus level but differ from *A. queenslandica* nucleotides are shown with an arrow pointing to their corresponding position. Specific insertions for *A. queenslandica* that are atypical for demosponges are displayed in shaded frames. Outlined frames highlight the regions of insertion within Haplosclerida that are displayed as sketches in Fig. 1.6. Inset: 90% consensus sequence and structure of partial helix 43 for 76 demosponges that do not belong to the marine haplosclerids.

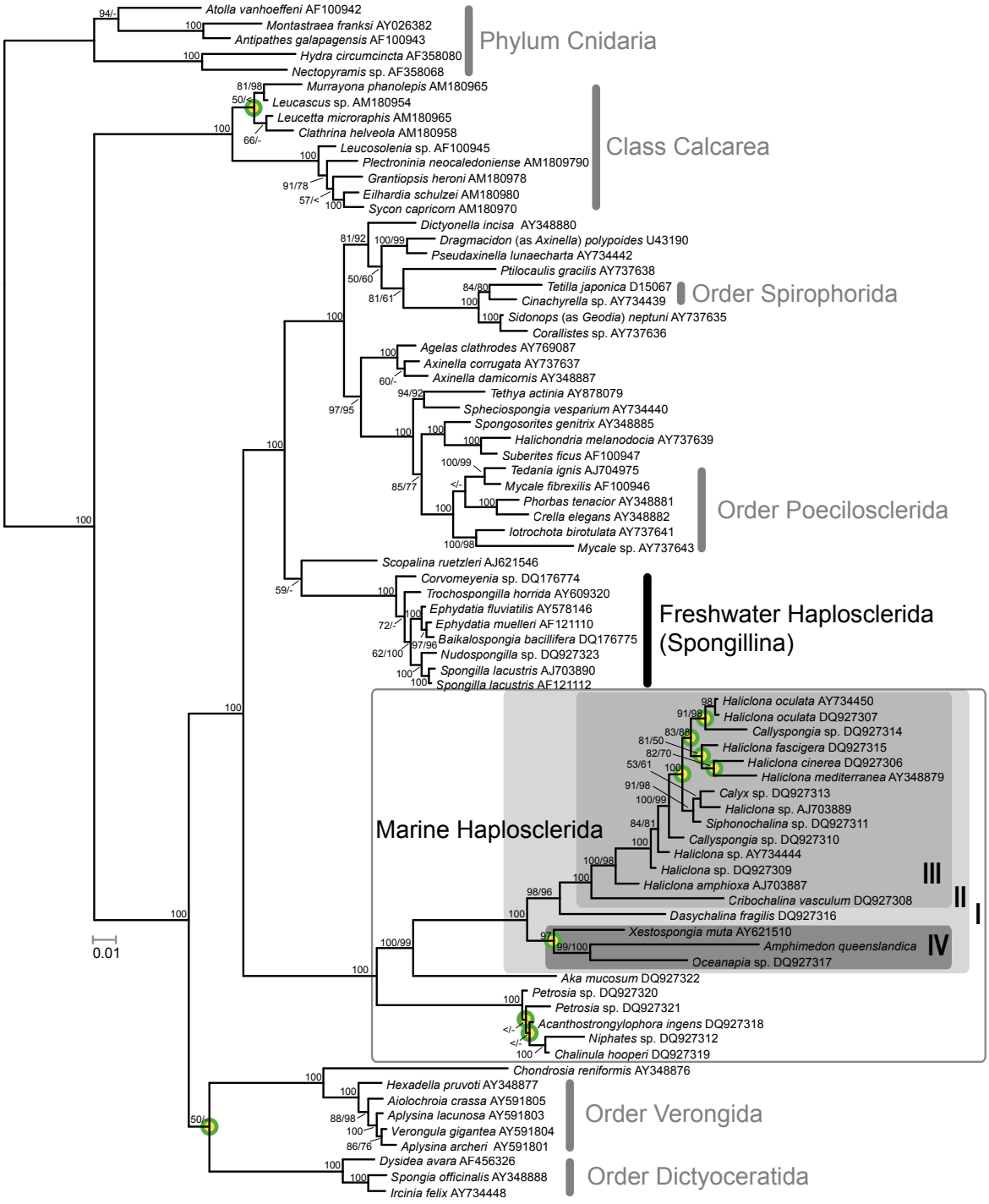


Figure 1.5: Phylogeny inferred with PHASE. Nodes that differ from the topology published by Redmond et al. (2007) are encircled. The boxed clades correspond to the excerpt displayed in Fig. 1.6. Support values are given at, or close to the corresponding node (values from analyses with PHASE/MrBayes; where the same support values were found in both analyses, only one number is shown; '<'= support values below 50; '-' = node not recovered in MrBayes analysis.). Monophyletic higher taxa are assigned.

ences were observed in Calcarea and Cnidaria.

The order Haplosclerida was not resolved as monophyletic, since Suborder Spongillina (freshwater sponges) fell into other distantly related demosponge clades, rather than into marine Haplo-

sclerida. The two suborders, Haplosclerina (families Callyspongiidae, Chalinidae and Niphatidae) and Petrosina (represented here by the families Petrosiidae and Phloeodictyidae) were not supported as monophyla (Fig. 1.6). These results are congruent with results from former analyses of SSU rRNA, 28S rRNA and cytochrome oxidase subunit I (Borchiellini et al., 2004; Erpenbeck et al., 2004; Nichols, 2005).

According to our analysis, *Amphimedon queenslandica* (Family Niphatidae) is most closely related to *Oceanapia* sp. (Family Phloeodictyidae), with *Xestospongia muta* (Family Petrosiidae) as sister taxon, and both nodes in the tree are very highly supported by posterior probability (PP) values in both Bayesian analyses. Other species of the family Niphatidae (*Niphates* sp. and *Dasychalina fragilis*) are not closely related to each other or to *Amphimedon queenslandica* (Fig. 1.6). In addition, other members of families Petrosiidae and Phloeodictyidae are not found in a closer relationship to the three species clade. Our results were mostly concordant with Redmond et al. (2007), but with higher support values in several clades. We did not find any monophyletic haplosclerid families or genera in our taxon set. Differences between our results and the previous study are highlighted at the nodes in Fig. 1.5. We could not recover monophyletic Petrosiidae in clade I, and relationships of several clade III taxa differed. Furthermore, *Xestospongia muta* and *Oceanapia* sp. cluster in one (clade IV, including *Amphimedon queenslandica*).

Successive evolution of additional helices in marine haplosclerids

Within haplosclerids, the evolution of additional helices can be reconstructed by plotting structures to the well-supported phylogenetic backbone (Fig. 1.6). Primary sequences of these motifs were not included in the tree construction (Fig. 1.5) due to ambiguous alignment, but can be regarded as additional phylogenetic characters. The helical insertions apparently evolved in at least two steps, which fits the findings of the SSU rRNA gene tree strikingly well. The relationships within marine haplosclerids can be described as four well-supported (PP> 0.97) nested clades I-IV that display different stages of secondary structure evolution (see above and Figs. 1.5, 1.6). Clade I contains all marine Haplosclerida. The basal diverging taxa lack any large insertions that are typical for other marine Haplosclerida. However, the predicted structure within helix 43 differs from the standard structure in this region found in other Porifera (compare outgroup in Fig. 1.6) and displays a larger bulge of unpaired bases at the insertion point of the larger helical structures found within all taxa in Clade II. This bulge may be the precursor for the extensions at this position observed in Clade II. Within basal diverging taxa of Clade II (i.e. Clade II without Clade III), a similar bulge is found for *Xestospongia muta* in helix E23_14 at the insertion-site of subsequent extensions in Clade III, but not the other sequences lacking E23_14.1.

Larger insertions appeared in helices E23_1 and 43 'simultaneously' (according to phylogenetic resolution recovered by our analyses) in the common ancestor of Clade II taxa. The three taxa of Clade IV according to our minimum free energy calculations share an additional helix 43b.2 as synapomorphy (Figs. 1.4, 1.5, 1.6).

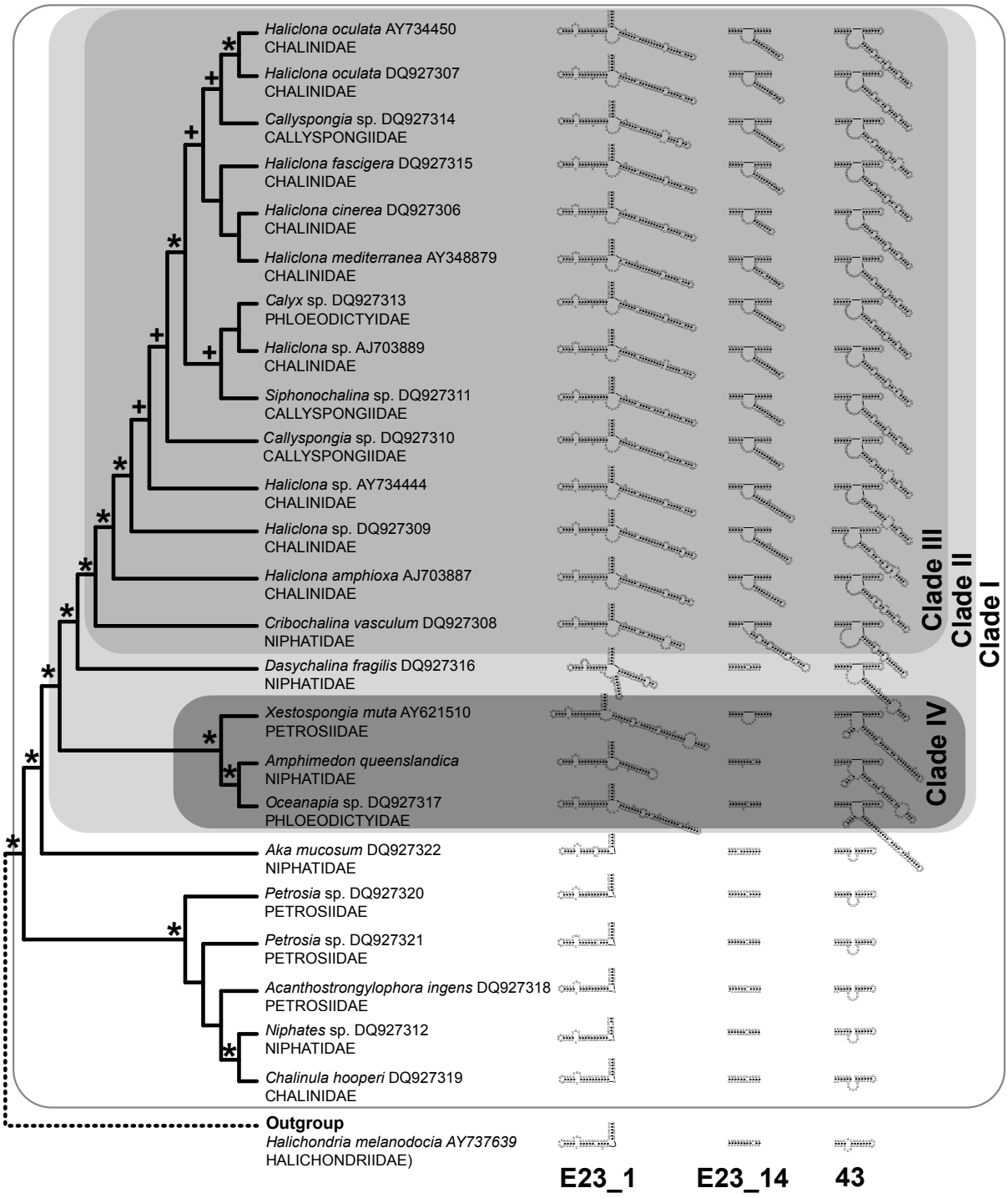


Figure 1.6: Relationships of marine Haplosclerida (excerpt from larger phylogenetic analyses shown in Fig. 1.5) and evolution of extension regions. Sketches of predicted secondary structures for extensions and conserved flanking regions correspond to outlined boxes in Fig. 1.4. Asterisks mark nodes that were found in at least 96% of sampled trees after burn-in in both Bayesian analyses (PHASE and MrBayes, see Material and Methods for details); plus signs mark nodes that appeared in lower frequencies, but still above 84% in one, or both of the analyses. For each species, the family is shown below the sequence name.

After the introduction of helical insertions in helices E23_1 and 43, a long extension evolved within helix E23_14 as a synapomorphy in Clade III. An autapomorphy for *Dasychalina fragilis* is an additional helix formed b insertions within E23_1. Within the complete taxon of marine haplosclerids, no loss of formerly gained additional helical insertions has been documented, therefore, no SSU rRNA molecule from a descendent of a taxon with extraordinary features has returned to the ancestral basic metazoan core structure.

Discussion

Unusual patterns within poriferan SSU rRNA secondary structure

We reported the secondary structures of a variety of poriferan SSU rRNA sequences, and suggest structure predictions for secondary structure motifs that are specific for some lineages, i.e. marine Haplosclerida (= Haplosclerida with the exception of members of the Spongillina) and Hexactinellida. Such additional helical insertions occur in a variety of eukaryotes and are known to be homoplasies, because they occur in several, not closely related taxa (Wuyts et al., 2001). Our data shows that such structures are also present in early diverging Metazoa (sponges).

Insertions in helix E23_1 evolved independently in Hexactinellida and the marine Haplosclerida (Clade II), which is evident (a) from our phylogenetic analyses that captured 'snapshots' of the evolution of helices within the marine Haplosclerida (Fig. 1.6) and (b) from the observations that insertions appear at different positions within Helix E23_1 (compare Fig. 1.3 with Fig. 1.4). Although additional helical insertions are present within the E23-extension fragment in various eukaryotic taxa, to our knowledge, none have been reported within helix E23_14, which is therefore synapomorphic for Clade III haplosclerids. Interestingly, helical insertions within haplosclerids first appeared in the typical regions for such insertions, namely within helices E23_1 and 43, before they evolved within E23_14 (see Fig. 1.6). Therefore, the evolution of extensions at more common insertion sites might be a prerequisite for the evolution of additional helical structures within E23_14.

Higher Metazoa with unusual SSU rRNA structures also contain unusual motifs in their large ribosomal subunit (LSU, 28S), e.g., in branchiopod Crustacea (Crease and Taylor, 1998). In sponges, additional motifs in a fragment of the LSU have previously been reported for Hexactinellida and marine haplosclerid demosponges (Erpenbeck et al., 2004), but not for non-haplosclerid demosponges or Calcarea. This is in striking accordance with our SSU rRNA findings, and encourages further studies of the complete LSU secondary structure of these taxa. Since both rRNA units are encoded in one translational unit, the same mechanisms may be responsible for the formation of extra helical features in both rRNA molecules.

Remarkably, not only the nuclear rRNAs display unusual secondary structure motifs in marine Haplosclerida. In the recently published mitochondrial genome of the haplosclerid *Amphimedon queenslandica*, both of the mitochondrial (mt) rRNA genes (12S and 16S rRNA) also contain additional helices that are not found in other demosponges (Erpenbeck et al., 2007b). Although this

may be a coincidental observation and needs to be verified by data from additional haplosclerid mt rRNA sequences, it is possible that the same selection mechanisms act on the nuclear and mt rRNA in this taxon. However, such correlation do not exist in all taxa, since the recently sequenced mitochondrial genomes of Hexactinellida (Haen et al., 2007) contain extremely short rRNAs (compared to the ones found in Demospongiae), in contrast to the large insertions in the hexactinellid nuclear rRNAs (e.g., see Fig. 1.3).

For the nuclear SSU rRNA, the fact that extra-helical structures are found in the E23-extension region and helix 43 in various taxa indicates that these regions are under less functional constraints than are the core regions of SSU rRNA; Wuyts et al. (2001) showed by considering the tertiary structure of rRNA that nucleotide variability increases with distance from the ribosome centre. Eukaryotic insertion sites for additional helices are therefore located in the same, or similar regions at the (3D-) periphery of rRNA molecules. The authors concluded that these insertions do not interfere with the ribosomal function of the ribosome and can therefore arise independently within different lineages, similar to our observations in Hexactinellida and marine haplosclerids.

Phylogenetic value of rRNA features

We demonstrated different applications of SSU rRNA features for phylogenetic analyses:

Base composition and synapomorphic base exchanges.

Base compositions of SSU rRNA differ strikingly between Demospongiae and Calcarea. The GC contents of the (much more diverse) Demospongiae are always higher and show a wider range of variability than the ones of Calcarea. For Hexactinellida, only five sequences were available (of which one is probably a non-functional copy), therefore general conclusions regarding their GC contents should be interpreted with care. However, for the few sequences available the GC contents fell into the ranges observed in Calcarea and Demospongiae.

Several apomorphic positions identified in calcarean SSU rRNA allow to unambiguously distinguish between the two subclasses Calcinea and Calcaronea, thus supporting other morphological and molecular data (Manuel et al., 2003; Manuel et al., 2004; Manuel, 2006; Dohrmann et al., 2006).

RNA models for phylogeny estimation and evolution of additional helical structures as evolutionary markers.

The Bayesian phylogenetic reconstructions using structure-defined partitions with different rRNA models for doublets in MrBayes 3.1.2 and PHASE 2 yielded very similar tree topologies with increased support for several nodes compared to the Maximum Parsimony (MP) and Maximum Likelihood (ML) analyses presented by Redmond et al. (2007) (Fig. 1.5). Partitioned analyses using rRNA models as applied in our analysis, have been reported to result in better-supported topologies (for sponges: Dohrmann et al., 2006; Erpenbeck et al., 2007a). However, other factors may have contributed to our findings, e.g., it is known that Bayesian posterior probability values are often higher than corresponding nonparametric bootstrap values and may even provide support for the 'wrong' clades in studies with simulated data (Erixon et al., 2003). The relevant important

haplosclerid clades were supported with very high PP values (>97%). These high values should overcome eventual problems of support overestimation. Also, support for 'wrong' clades is unlikely to be a problem in our results, since the topologies from Redmond et al. (2007) are mostly concordant with ours. Regarding the general differences in bootstrap and PP values, the different software packages used and the difference in the data set (taxon sampling, alignment and included sites) may have contributed to the higher support found for several clades.

Although standard models of nucleotide evolution violate the assumption of independent evolution of all sites when also applied to paired sites as done by Redmond et al. (2007), this seems to have little impact on nodes with high support values in the case of the demosponge dataset studied (compare Fig. 1 in Redmond et al., 2007, with our Fig. 1.5). This suggests that the biases introduced by the use of less well fitting 'standard' rRNA models may have a higher impact on clades that are difficult to resolve (e.g., due to noisy data), whereas a strong phylogenetic signal will be recovered even if a sub-optimal evolutionary model is used for analyses.

Amphimedon queenslandica, the target of the Sponge Genome Sequencing project, did not cluster with any other representative of its Family Niphatidae. Likewise, neither the other haplosclerid families (Callyspongiidae, Chalinidae, Petrosiidae, Phloeodictyidae), nor the genera (*Callyspongia*, *Haliclona*, *Petrosia*) could be recovered as monophyletic, (besides of the genus *Petrosia*) in accordance with the SSU-rRNA based findings of Redmond et al. (2007). Strikingly, these inferred relationships are supported by the presence or absence of secondary structure motifs within Haplosclerida: different members of the families Niphatidae, Petrosiidae and Phloeodictyidae show a different number of specific insertions that are congruent with the phylogenetic relationships that we previously inferred without the inclusion of these extended regions (see Figs. 1.5, 1.6). The presence and absence of such helices are therefore good phylogenetic indicators for these relatively closely related taxa, even though alignment of the primary sequence of these helices (and cladistic or phenetic sequence analysis) between all taxa is difficult due to high evolutionary rates. Homology inference of sites in hypervariable regions according to their secondary structure is known to be problematic. Functional constraints are probably more relaxed in these regions, and the observed evolution of insertions is driven by unknown mutational mechanisms, which might tend to produce similar motifs by homoplasy (Hancock and Vogler, 2000). In contrast, within marine haplosclerids, no loss of helical insertions that arose at some earlier point of their evolutionary history occurred (Fig. 1.6). Furthermore, no independent homoplasic helical insertions ever appeared in the same positions within SSU rRNA. Considering our findings, the presence and absence of large helical insertions appears to provide strong phylogenetic information at selected taxonomical levels.

A similar phylogenetic information value of additional helical structures may be present for insertions in Helix E23_1 of Hexactinellida, although generalized conclusions are limited by the small sample size. Nonetheless, the insertions in Hexasterophora (*Acanthascus*, *Oopsacas*, *Aphrocallistes*, and *Farrea*) are predicted to possess two additional helices, while the insertion found in the only considered amphidiscophoran, *Semperella schulzei*, forms only one helix (Fig.

1.3). It is evident that integration of secondary structure information in sequence alignment and analyses (in the form of rRNA substitution models) will optimize rRNA phylogenies considerably.

Conclusions

The SSU rRNA provides far more valuable phylogenetic information than just its primary sequence. Even simple features like base composition already bear enough information to distinguish between the two higher sponge taxa Calcarea and Demospongiae (Fig. 1.1). Unusual secondary structures can lend further support to results from independent phylogenetic inferences, as we showed for helical insertions in marine haplosclerid demosponges (Fig. 1.6) and for a small number of hexactinellid sponges (Fig. 1.3). In this way, otherwise neglected hypervariable insertions can yield further support to a given topology. Although we only explored additional structures of SSU rRNA in sponges, our results should encourage further studies. Especially the study of LSU rRNA structure seems promising in this regard, since this gene is more variable than SSU rRNA and contains large extension regions that strongly differ among higher taxa (Schnare et al., 1996). On the intraspecific level, the even more variable internal transcribed spacer regions (ITS 1 and ITS 2) can provide secondary structure features of phylogenetic value (see e.g., Sánchez et al., 2007). Relating secondary structure to sequence information will allow the phylogenetic signal of the huge numbers of rRNA sequences currently available in Genbank to be considerably increased. Our newly generated database of Porifera SSU rRNA secondary structures will facilitate the inclusion of secondary structure information in phylogenetic analyses.

Authors' contributions

O.V. contributed to the conception and design of the study, acquired, analyzed (including development of analytical Perl scripts) and interpreted the data, and wrote the manuscript. D.E. and G.W. contributed to the conception and design of the study, and critical revision of the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

This work was financially supported by the German Research Foundation (DFG, Project Wo896/3-1, 3) and is also a contribution from the DFG priority program SPP1174 'Deep Metazoan Phylogeny' (Project Wo896/6-1). We thank Martin Dohrmann for providing hexactinellid sequences ahead of publication and for useful comments on the manuscript and Catherine Vogler for commenting on the manuscript and for very productive discussions. O.V. would like to thank and the organizers and participants of the 2006 Workshop on Molecular Evolution in Woods Hole for inspiring discussions and the support of the Boehringer Ingelheim Fonds to cover travel expenses and tuition fees. D.E. acknowledges financial support from the European Union under a Marie-Curie outgoing fellowship (MOIF-CT-2004 Contract No 2882), GW acknowledges financial support from the Marie-Curie EST HOTSPOTS. The Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen (GWDG) is acknowledged for access to their computer facilities. We thank three anonymous reviewers for their constructive comments, especially reviewer 1.

Supplementary information

The alignment with secondary structure information and character set is available at:
http://www.biomedcentral.com/content/supplementary/1471-2148-8-69-s5.mase

Appendix 1:

Table A1.1: Taxa and their GenBank Accession numbers for poriferan SSU sequences.

Figure A1.1: : Trace IDs (TI) from GenBank trace archive used to assemble the *Amphimedon queenslandica* SSU rDNA sequence

Figure A1.2: Compensatory base changes and alignments of predicted secondary structures

Table A1.2: Base composition and fragment lengths of examined poriferan SSU rRNAs

Chapter 2

Non-monophyly of most supraspecific taxa of calcareous sponges (Porifera, Calcarea) revealed by increased taxon sampling and partitioned Bayesian analysis of ribosomal DNA

Martin Dohrmann[1], Oliver Voigt[1], Dirk Erpenbeck [1,2], Gert Wörheide[1,*]

[1] Department of Geobiology, Geoscience Centre Göttingen, Goldschmidtstr. 3, D-37077 Göttingen, Germany

[2] Queensland Museum, South Brisbane, Qld., Australia

* Corresponding author.

This version of the article was published in Molecular Phylogenetics and Evolution on 30 April 2006 (Molecular Phylogenetics and Evolution 40 (2006) 830–843; doi:10.1016/j.ympev.2006.04.016).

Abstract

Calcareous sponges (Porifera, Calcarea) play an important role for our understanding of early metazoan evolution, since several molecular studies suggested their closer relationship to Eumetazoa than to the other two sponge 'classes,' Demospongiae and Hexactinellida. The division of Calcarea into the subtaxa Calcinea and Calcaronea is well established by now, but their internal relationships remain largely unresolved. Here, we estimate phylogenetic relationships within Calcarea in a Bayesian framework, using full-length 18S and partial 28S ribosomal DNA sequences. Both genes were analyzed separately and in combination and were further partitioned by stem and loop regions, the former being modelled to take non-independence of paired sites into account. By substantially increasing taxon sampling, we show that most of the traditionally recognized supraspecifc taxa within Calcinea and Calcaronea are not monophyletic, challenging the existing classification system, while monophyly of Calcinea and Calcaronea is again highly supported.

Introduction

Sponges (Porifera Grant, 1836) are sessile, aquatic filter feeders that are considered to be the earliest branching metazoans (e.g., Ax, 1995). Monophyly of Porifera has been questioned by a number of molecular studies (e.g., Adams et al., 1999; Borchiellini et al., 2001; Cavalier-Smith et al., 1996; Collins, 1998; Kruse et al., 1998; Zrzavy et al., 1998; Lafay et al., 1992; Medina et al., 2001) —albeit usually with low statistical support—with the calcareous sponges (Calcarea Bowerbank, 1864) being more closely related to eumetazoans than to the other two classically recognized major sponge lineages Demospongiae Sollas, 1885 and Hexactinellida Schmidt, 1870, which are commonly grouped together as Silicispongia or Silicea. As this would imply that the last recent common ancestor of (Eu)metazoa was a sponge-like organism or, alternatively, the sponge bauplan evolved twice, Calcarea play an important role in the reconstruction of early animal evolution, making a well-resolved and supported phylogeny of this group clearly desirable.

The calcareous sponges are represented by about 500, exclusively marine species distributed in all oceans (Manuel et al., 2002). While the mineral skeleton of Demospongiae and Hexactinellida consists of intracellularly formed siliceous spicules, Calcarea is characterized by the intercellular formation of spicules composed of calcium carbonate, which is an autapomorphic character of this group (Ax, 1995; Böger, 1988; Manuel et al., 2002; Manuel, 2006). The monophyly of calcareous sponges is also supported by ribosomal DNA (rDNA) data (Borchiellini et al., 2001; Manuel et al., 2003; Manuel et al., 2004).

Cytological and embryological characters and features of spicule morphology strongly suggest a division of the Calcarea into the subtaxa Calcinea and Calcaronea (Bidder, 1898; Borojevic et al., 1990; Borojevic et al., 2000; Manuel et al., 2002; Manuel, 2006). Another character distinguishing these two groups is the ratio of different carbon isotopes that are incorporated into the spicules during biomineralisation (Reitner, 1992; Wörheide and Hooper, 1999). Although the Calcinea and Calcaronea are very well characterized by these features, there still remains the possibility that some character states in one of the groups represent symplesiomorphies, rendering the respective group paraphyletic with regard to the other (Manuel et al., 2002; but see Manuel, 2006). As rDNA studies (Borchiellini et al., 2001; Manuel et al., 2003; Manuel et al., 2004) do support monophyly of Calcinea and Calcaronea, this scenario seems rather unlikely, however.

In contrast, phylogenetic relationships within Calcinea and Calcaronea remain largely unclear, because the existing classification of calcareous sponges (Borojevic et al., 1990; Borojevic et al., 2000; Borojevic et al., 2002b; Borojevic et al., 2002c; Borojevic et al., 2002a; Vacelet et al., 2002b; Vacelet et al., 2002a) is primarily typologic and a phylogenetic system of this group has not been proposed so far (but see Reitner, 1992). Because of the apparent high level of morphological homoplasy (Manuel et al., 2003), such a system would be difficult or impossible to base on the available morphological data alone. Therefore, molecular data provide the most promising means to resolve this branch of the tree of life.

So far, only two studies (Manuel et al., 2003; Manuel et al., 2004) explicitly addressed the question of phylogenetic relationships within Calcarea, applying maximum parsimony (MP) and maximum likelihood (ML) methods to infer trees from 18S and 28S rDNA sequences and morphological character data of 17 calcareous sponge species, representing 15 ‘genera,’ 13 ‘families’ and three out of five ‘orders.’ An important result of these studies was the placement of *Petrobiona massiliana* Vacelet and Lévi, 1958 in Baerida Borojevicet al., 2000 instead of Lithonida Vacelet, 1981, which is also supported by some spiculation features such as the occurrence of microdiactines and pugioles (dagger-shaped tetractines). Furthermore, monophyly of Leucosolenida Hartman, 1958, Grantiidae Dendy, 1892, and *Sycon* Risso,1826, was not supported. However, taxon sampling was still too sparse, especially with respect to Calcinea, to make further inferences about higher-level relationships within the two major groups of calcareous sponges.

With this study, we extend the set of available calcarean18S and 28S rDNA sequences to 44 (mostly Indo-Pacifc) species, representing 27 ‘genera’, 18 ‘families’ and all We currently recognized ‘orders’ of Calcarea. Taxon sampling of Calcinea is increased from four (Manuel et al., 2003;

Manuel et al., 2004) to 20 species. From 31 species we also sequenced ~750 additional base pairs (bp) of the 28S rRNA gene. We analyzed both genes separately and in combination in a Bayesian framework that accounts for different evolutionary constraints of stem and loop regions and non-independence of paired sites, thereby representing a modeling scheme that is biologically more realistic than standard models commonly applied today and leads to statistically more robust estimations of phylogeny (Telford et al., 2005; Erpenbeck et al., 2007a). The aims of this study were to evaluate the validity of classically recognized calcinean and calcaronean supraspecifc taxa, for most of which no clear statements about potential morphological apomorphies can be found in the literature, and to re-evaluate earlier findings (Manuel et al., 2003; Manuel et al., 2004) in the light of substantially increased taxon sampling and a more flexible approach of inferring phylogenies. While distinction of the classically recognized 'subclasses' Calcinea and Calcaronea is highly supported by our analyses, our results suggest that the majority of 'orders' and 'families', as well as some 'genera', such as the species-rich *Clathrina* and *Leucandra,* are not monophyletic.

Materials and methods

Species, collection sites, sample-numbers of the Queensland Museum (QM), South Brisbane (Australia), where most vouchers are deposited, and GenBank accession numbers of the sequences generated in this study, as well as those retrieved from GenBank (http:// www.ncbi.nlm.nih.gov/), are given in Table 2.1.

DNA-extraction, -amplifcation, and –sequencing

Genomic DNA was extracted from ethanol-preserved or silica-dried samples with the DNEasy Tissue Kit of Qiagen (Hilden, Germany), following the manufacturer's protocol. To avoid contamination with epibiontic organisms, tissue from the interior of the sponges was used whenever possible. Full-length 18S rDNA was amplified by polymerase chain reaction (PCR) with primers 18S1 and 18S2 (Manuel et al., 2003, see Appendix 2, Table A2.1) (2 min/94 °C; 34 cycles [1 min/94 °C; 1min/ 50–58 °C; 2min/ 72 °C]; 7min/72 °C). Partial 28S rDNA (domain D2 to helix 36; nomenclature of Michot et al., 1990) was amplified with primers from Medina et al. (2001) and Nichols (2005) (see Appendix 2, Table A2.1) (10min/95 °C; 34 cycles [1 min/95 °C; 1 min/50–58 °C; 1–4 min/72 °C]; 7min/ 72 °C). Reaction mixes contained 2.5 µl of 10× NH4 PCR buffer (Bioline, Luckenwalde, Germany), 1.0–1.5 µl MgCl2 (50 mM), 1 µl of each primer (10 µM), 0.5 µl dNTPs (10 mM each), 0.05 µl *Taq*-DNA-Polymerase (5 u/µl; Bioline, Luckenwalde, Germany) and 0.5–5 µl template. Bands of expected size were cut out from agarose gels and purified following Boyle and Lew (1995). Both strands of the amplicons were sequenced directly with BigDye Terminator 3.1 chemistry and an ABI Prism 3100 Genetic Analyser (Applied Biosystems). Sequencing primers are given in Appendix 2, Table A2.1. Intragenomic length variation did not allow direct sequencing of *Eilhardia schulzei* and *Plectroninia neocaledoniense*, so PCR products were cloned with the TOPO Cloning Kit for Sequencing (Invitrogen, Karlsruhe) and up to three clones were sequenced. Because the intragenomic indels appeared in regions that were not included in the phylogenetic analyses (see below), only one sequence of each Species was used. Sequences were assembled and edited with the program CodonCode Aligner (http://www.codoncode.com), and validated via BLAST searches

T**able 2.1:** Species used in this study with accession numbers of the corresponding sequences, as well as collection sites and QM specimen numbers of the species for which new sequences have been generated

Taxon	Collection site	QM-No.	Acc-No. 18S	Acc-No. 28S
Calcinea				
Clathrina wistariensis (Clathrinida, Clathrinidae)	Wistari Reef (GBR)	G313663	AM180961	AM180990
Clathrina adusta (Clathrinida, Clathrinidae)	Wistari Reef (GBR)	G313665	AM180961	AM180991
Clathrina helveola (Clathrinida, Clathrinidae)	Heron Reef (GBR)	G313680	AM180958	AM180987
Clathrina luteoculcitella (Clathrinida, Clathrinidae)	Heron Island/Wistari Reef	G313684	AM180959	AM180988
Clathrina sp. (Clathrinida, Clathrinidae)	Yonge Reef (GBR)	G313693	AM180960	AM180989
*Clathrina cerebrum** (Clathrinida, Clathrinidae)	—	—	U42452	AY563541
Clathrina aff .'*cerebrum*' (Clathrinida, Clathrinidae)	Hook Reef (GBR)	G313824	AM180957	AM180986
Guancha sp. (Clathrinida, Clathrinidae)	Rene's Nook (GBR)	G316033	AM180963	AM180992
Soleneiscus radovani (Clathrinida, Soleneiscidae)	Wistari Reef (GBR)	G313661	AF452017	AM180982
Soleneiscus stolonifer (Clathrinida, Soleneiscidae)	Wistari Reef (GBR)	G313668	AM180955	AM180983
Levinella prolifera (Clathrinida, Levinellidae)	Hook Reef (GBR)	G313818	AM180956	AM180984
Leucaltis clathria (Clathrinida, Leucaltidae)	DJ's Reef (GBR)	G316022	AF452016	AM180985
Leucascus sp. (Clathrinida, Leucascidae)	GBR	G316051	AM180954	AM180981
Leucetta sp. (Clathrinida, Leucettidae)	Yonge Reef (GBR)	G313691	AM180964	AM180993
Leucetta chagosensis (Clathrinida, Leucettidae)	Osprey Reef (Coral Sea)	G316279	AF182190	AM180994
Leucetta microraphis (Clathrinida, Leucettidae)	Wistari Reef (GBR)	G313659	AM180965	AM180995
Leucetta villosa (Clathrinida, Leucettidae)	Wistari Reef (GBR)	G313662	AM180966	AM180996
Pericharax heteroraphis (Clathrinida, Leucettidae)	Holmes Reef (Coral Sea)	G316295	AM180967	AM180997
Murrayona phanolepis (Murrayonida, Murrayonidae)	Bougainville Reef (Coral Sea)	G316290	—	AM180998
Murrayona phanolepis (Murrayonida, Murrayonidae)	Osprey Reef (Coral Sea)	G313992	AM180968	—
Lelapiella incrustans (Murrayonida, Lelapiellidae)	Vanuatu (SW Pacific)	G313914	AM180969	AM180999
Calcaronea				
Leucosolenia sp. (Leucosolenida, Leucosoleniidae)	—	—	AF100945	AY026372
Sycon capricorn (Leucosolenida, Sycettidae)	Ribbon Reef (GBR)	G316187	AM180970	AM181000
*Sycon raphanus** (Leucosolenida, Sycettidae)	—	—	AF452024	AY563537
*Sycon ciliatum** (Leucosolenida, Sycettidae)	—	—	L10827	AY563532
*Sycon calcaravis** (Leucosolenida, Sycettidae)	—	—	D15066	—
*Grantia compressa** (Leucosolenida, Grantiidae)	—	—	AF452021	AY563538
Ute ampullacea (Leucosolenida, Grantiidae)	Wistari Reef (GBR)	G313669	AM180972	AM181002
Aphroceras sp. (Leucosolenida, Grantiidae)	Osprey Reef (Coral Sea)	G316285	AM180971	AM181001
Leucandra nicolae (Leucosolenida, Grantiidae)	Wistari Reef (GBR)	G313672	AM180974	AM181003
*Leucandra aspera** (Leucosolenida, Grantiidae)	—	—	AF452022	AY563535
Leucascandra caveolata (Leucosolenida, Jenkinidae)	Hardline (GBR)	G316057	AM180973	AM181004
*Anamixilla torresi** (Leucosolenida, Jenkinidae)	—	—	AF452020	AY563536
Vosmaeropsis sp.* (Leucosolenida, Heteropiidae)	—	—	AF452018	AY563531
Syconessa panicula (Leucosolenida, Heteropiidae)	Wistari Reef (GBR)	G313671	AM180976	AM181007
Sycettusa tenuis (Leucosolenida, Heteropiidae)	Heron Reef (GBR)	G313685	AM180975	AM181006
Sycettusa sp.* (Leucosolenida, Heteropiidae)	—	—	AF452025	AY563530
Paraleucilla magna (Leucosolenida, Amphoriscidae)	South Atlantic	—	—	AM181005
Paraleucilla sp.* (Leucosolenida, Amphoriscidae)	—	—	AF452023	—
Grantiopsis sp. (Leucosolenida, Lelapiidae)	GBR	G313969	AM180977	AM181008
Grantiopsis heroni (Leucosolenida, Lelapiidae)	Wistari Reef (GBR)	G313670	AM180978	AM181009
*Leuconia nivea** (Baerida, Baeriidae)	—	—	AF182191	AY463534
Eilhardia schulzei (Baerida, Baeriidae)	Mac's Reef (GBR)	G316071	AM180980	AM181010
*Petrobiona massiliana** (Baerida, Petrobionidae)	—	—	AF452026	AY563533
Plectroninia neocaledoniense (Lithonida, Minchinellidae)	Holmes Reef (Coral Sea)	G316300	AM180979	AM181011
Outgroups				
Suberites ficus (Demospongiae)	—	—	AF100947	AY026381
Mycale fibrexilis (Demospongiae)	—	—	AF100946	AY026376
Acanthascus (*Rhabdocalyptus*) *dawsoni*	(Hexactinellida) —	—	AF100949	AY026379
Antipathes galapagensis (Cnidaria, Anthozoa)	—	—	AF100943	AY026365
Atolla vanhoeffeni (Cnidaria, Scyphozoa)	—	—	AF100942	AY026368
Saccharomyces cerevisiae (Fungi, Ascomycota)	—	—	V01335	U53879

Table 2.1, continued

Classification of Calcarea after Borojevic et al. (2002a,b,c); Vacelet et al. (2002a,b) and Manuel et al. (2003). GBR, Great Barrier Reef (Australia). Accession numbers of new sequences are given in boldface. Asterisks indicate ingroup-species for which no genomic DNA or complete 28S rDNA sequences from GenBank were available.

[a] Note: The specimen with QM-number G313824 shows clear affnities to *Clathrina cerebrum* and *C. brasiliensis* Solé-Cava et al., 1991, because it shares spines on the apical actines of tetractines with these two species, a trait that is known from no other *Clathrina* species (see Klautau and Valentine, 2003). *C. brasiliensis* was described solely from Brazil, and a cosmopolitan distribution of *C. cerebrum* is not considered valid by Klautau and Valentine (2003, 15–16), who restrict the species to the Mediterranean and Adriatic seas. However, *Clathrina cerebrum* possibly constitutes a complex of morphologically similar species (Klautau and Valentine, 2003, 15), and distinction between *C. cerebrum* and *C. brasiliensis* is mainly based on genetical differences (Klautau and Valentine, 2003; Solé-Cava et al., 1991, 11–12). Because G313824 was collected from the Great Barrier Reef (Australia), we give it here the preliminary name *Clathrina* aff. '*cerebrum*', indicating that it might belong to a putative *C. cerebrum*/*C. brasiliensis* species complex.

(http://www.ncbi.nlm.nih.gov/BLAST/; Altschul et al., 1990) against the GenBank nucleotide database.

Alignments

Published calcarean sequences and outgroup-sequences were downloaded from GenBank (Table 1.1) and automatically aligned together with our new sequences with ClustalX 1.81 (Thompson et al., 1997), followed by manual adjustment using SeaView (Galtier et al., 1996) and Mac Clade 4.08 (Maddison and Maddison, 2002). For some of the species (indicated by asterisks in Table 2.1) 28S rDNA sequences deposited in GenBank only ranged from domain D2 to helix 26, and no genomic DNA was available. Manual adjustments were done according to secondary structural information that was used to define partitions and paired bases for phylogenetic analyses (see below). 28S rRNA secondary structure was assessed using Hancock et al. (1988); Michot et al. (1990); Schnare et al. (1996); and Erpenbeck et al. (2004) as references. For domains D2, D6, and D7, no unambiguous predictions of paired sites could be made for a consensus structure, so these regions were effectively treated as loops. Secondary structure predictions for 18S rRNA were developed using information on the structure of *Saccharomyces cerevisiae* from the European ribosomal RNA database (http://www.psb.ugent.be/rRNA/; Wuyts et al., 2002) and the structure suggested by Wuyts et al. (2000). For variable regions of the 18S rRNA, predictions from the secondary structure algorithm implemented in RNA structure 4.1 (Mathews et al., 2004), as well as compensatory base changes between sequences of closely related taxa, were taken into account. In regions of the 28S rDNA alignment where ambiguity was caused solely by outgroup taxa, the corresponding nucleotides of these taxa were recoded as missing data, because a large proportion of sites (mainly in the D2 domain) was affected in this way, and total exclusion of these sites would have led to the loss of many phylogenetically informative sites for the ingroup. This approach allowed us to keep as much of the available phylogenetic information as possible in the alignment, while minimizing the potentially misleading effects of uncertain assessments of positional homology. In both the 18S and the 28S rDNA alignment, positions that could not be aligned unambiguously for all taxa, and insertions comprising only one or two species or only outgroup taxa, were excluded from all analyses. For the combined analysis, the 28S rDNA sequence of *Sycon calcaravis*, which was not available, was coded as missing data, and the 18S rDNA sequence of *Paraleucilla* sp. was concatenated with the 28S rDNA sequence of *Paraleucilla magna*, because these two species ap-

peared at the same positions in the topologies of the separate analyses. Alignments and corresponding trees are deposited in TreeBASE (http://www.treebase.org; study number: S1520).

Phylogenetic analyses

Phylogenies were estimated with MrBayes 3.1.1 (Ronquist and Huelsenbeck, 2003) under default priors from the 18S rDNA alignment, the 28S rDNA alignment, and a combined matrix. *S. cerevisiae* was used as the outgroup taxon. ML tree searches and non-parametric bootstrap analyses (Felsenstein, 1985) were also conducted, using the web server of the heterogeneous distributed computing system MultiPhyl (http://www.cs.nuim.ie/distributed/multiphyl.php; see also Keane et al., 2005) with SPR tree search and 1000 bootstrap replicates. However, because the modelling scheme described in the next section could not be implemented in the ML analyses, the results of the two methods were not directly comparable (see Discussion). Given that bootstrap proportions (BP values) are a conservative measure of clade support (e.g., Hillis and Bull, 1993), and Bayesian posterior probabilities (PP values) might be overestimations (e.g., Suzuki et al., 2002; but see Huelsenbeck and Ronquist, 2005; Huelsenbeck and Rannala, 2004), PP values >95% and BP values >75% were interpreted as giving strong support to the respective clade.

Partitioning and model choice

Stem and loop regions of folded RNA molecules are subjected to different evolutionary constraints (e.g., Dixon and Hillis, 1993; Wheeler and Honeycutt, 1988), and thus require different models of nucleotide substitution. Furthermore, the assumption of independence of sites is clearly violated when stem regions are analyzed like unpaired characters, because paired sites evolve together in order to maintain secondary structure (Dixon and Hillis, 1993; Hillis and Dixon, 1991). The Bayesian Markov chain Monte Carlo (MCMC) technique (see Huelsenbeck et al., 2002 and references therein) makes it possible to combine different datasets in a single analysis and to partition single datasets into potentially differently evolving subsets, while allowing each partition to be modelled independently (Huelsenbeck and Ronquist, 2005; Ronquist and Huelsenbeck, 2003). In addition, the great computational efficiency of the method (Larget and Simon, 1999) allows large datasets to be analyzed within a reasonable time, even under complex models (e.g., Nylander et al., 2004). Although models have been developed to account for non-independence of nucleotide sites (Jow et al., 2002; Muse, 1995; Schöniger and von Haeseler, 1994; Tillier and Collins, 1995; Tillier and Collins, 1998), it has not yet become common practice to use such models in phylogenetic analyses of rDNA sequences.

In this study, alignments were partitioned into stem and loop regions, and stem regions were analyzed under the Doublet model, which is based on the SH model (see Schöniger and von Haeseler, 1994 and Huelsenbeck and Ronquist, 2005, for details). In both stem and loop regions, all six substitution types were allowed to have different probabilities (nst=6), which corresponds to the General Time Reversible model of nucleotide substitution (GTR; Tavaré, 1986). Loop regions and regions where paired sites could not be defined unambiguously (see above) were analyzed under the GTR model alone. This most parameter-rich model of the time reversible family of models (see Swofford et al., 1996) was chosen because Bayesian inference has been shown to be much more

robust to over- than to underparameterization (Huelsenbeck and Rannala, 2004; Lemmon and Moriarty, 2004). The partitioned Doublet+GTR approach was also tested against a GTR-only approach (no partitioning into stems and loops, no consideration of paired sites) by use of the Bayes factor (Kass and Raftery, 1995, see below), to assess if the Doublet+GTR model could explain our data significantly better. In all analyses, among-site rate variation was modelled with a Γ-distribution with four rate categories, allowing a proportion of sites to be invariant (I+G; Gu et al., 1995). Values for the individual model parameters were estimated by MrBayes from the data. Data partitions (18S stems, 18S loops, 28S stems, 28S loops) were unlinked for all parameters except topology and branch lengths. ML model search was performed with MultiPhyl (see above) under the Akaike Information Criterion (AIC; Akaike, 1974) and the Bayesian Information Criterion (BIC; Schwarz, 1978).

MCMC settings

Two independent runs with one cold and seven heated Markov chains each per analysis were performed simultaneously until the average standard deviation of split frequencies between the two runs dropped below 0.005,lowered from the default stop value of 0.01 to improve convergence of chains. Analyses were run twice to check for consistency of results. A longer run of the combined dataset (>8×10^6 generations) was also performed to check if running the Markov chains for more generations could additionally improve convergence. To improve mixing, the temperature-values of the heated chains were lowered from the default (0.20) to 0.01. Trees were sampled every 100 generations. Topology and branch-length information was summarized in 50% majority rule consensus trees with the 'sumt' command; samples obtained before stationarity of ln-likelihoods against generations had been reached were discarded as burn-in. Analyses were carried out with the MPI-enabled parallel version of MrBayes (Altekar et al., 2004) on a 64-node Linux cluster at the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG; www.gwdg.de), requesting one processor for each of the six-teen Markov chains per analysis. The longer analysis of the combined matrix was run on an Apple Power Mac G5 Dual computer. Batch files are available upon request.

Testing hypotheses of monophyly

To test whether non-monophyly of traditionally recognized supraspecifc taxa was statistically significant, we enforced constraints on the topology-priors, making the affected taxa monophyletic a priori. Phylogenetic analysis of the combined dataset was then repeated for each constraint as described above, and the difference between the harmonic means of the likelihood values sampled by the MCMC procedure of the constrained (null hypothesis, H_0) and the unconstrained (alternative hypothesis, H_1) analysis was calculated. A Bayes factor (B_{10}) is equal to the ratio of the marginal likelihoods of H_1 and H_0; as these are difficult to calculate analytically, one can use the harmonic means as a valid approximation (Newton and Raftery, 1994). Harmonic means were obtained using the 'sump' command; the first 25% of the samples were discarded as burn-in. It is possible that trees sampled during the unconstrained analysis accidentally contain the constraint that was used in the constrained analysis, there by potentially biasing subsequent calculations.

Therefore, we filtered the post-burn-in samples of the unconstrained analysis for those trees, using PAUP* 4.0b10 (Swofford, 2003). If such topologies were present, we corrected the harmonic mean (hm) of the likelihood values of the unconstrained analysis (H_1) by multiplying it with n/($n+n_{cons}$), where n is the number of trees sampled, and n_{cons} is the number of trees containing the constraint. The formula for calculating Bayes factors then became 2 ln(B_{10})= hm (H_1) (n/($n+n_{cons}$)) - hm (H_0). Bayes factors were interpreted according to the table of Kass and Raftery (1995; reproduced in Table 2.2).

Table 2.2: Interpretation of Bayes factors according to Kass and Raftery (1995)

2 ln(B_{10})	**Evidence against H_0**
0-2	Not worth more than bare mentioning
2-6	Positive
6-10	Strong
>10	Very strong

Results

Model comparison

According to the Bayes factor, the partitioned Doublet+GTR model could explain our data significantly better than the GTR-only approach; evidence against the latter was 'very strong' in both the separate and the combined analyses (Table 2.3). For the ML analyses, both AIC and BIC chose the Tamura–Nei model (TrN; Tamura and Nei, 1993) with a proportion of invariant sites and a Γ-distribution of the variable sites (I+G).

18S rDNA

The two independent Bayesian analyses produced identical topologies, and differences in PP values, where present, were minimal. The tree of the first analysis is shown in Fig. 2.1 (results of second analysis not shown). Monophyly of Calcarea, Calcinea, Calcaronea, Silicea, Demospongiae, and Cnidaria was strongly supported. Porifera was recovered as paraphyletic: cnidarians (as representatives of the Eumetazoa) formed a clade with the siliceous sponges; however, with poor support (PP=64). In the ML tree (Appendix 2, Fig. A2.1), Cnidaria weakly grouped with Calcarea

Table 2.3: Harmonic means (hm) of the sampled likelihood values of phylogenies obtained with two different modelling schemes, and the respective Bayes factors.

	18S		**28S**		**18S+28S**	
Model (+I+G)	hm	2 ln(B_{10})	hm	2 ln(B_{10})	hm	2 ln(B_{10})
GTR	– 8,403.77	1,887.62	– 14,645,45	5,562.30	– 23,130.49	7,664,04
Doublet + GTR	– 7,459,96		– 11,864,30		– 19,298,47	

Bayes factors were calculated as 2 ln(B_{10}) D2(hm (L_1)¡hm (L_0)), where L_1, likelihood values of H_1 (i.e., Doublet+GTR; stem/loop partitioned) and L_0, likelihood values of H_0 (GTR only; no stem/loop partitioning). See Table 2.2 for interpretation.

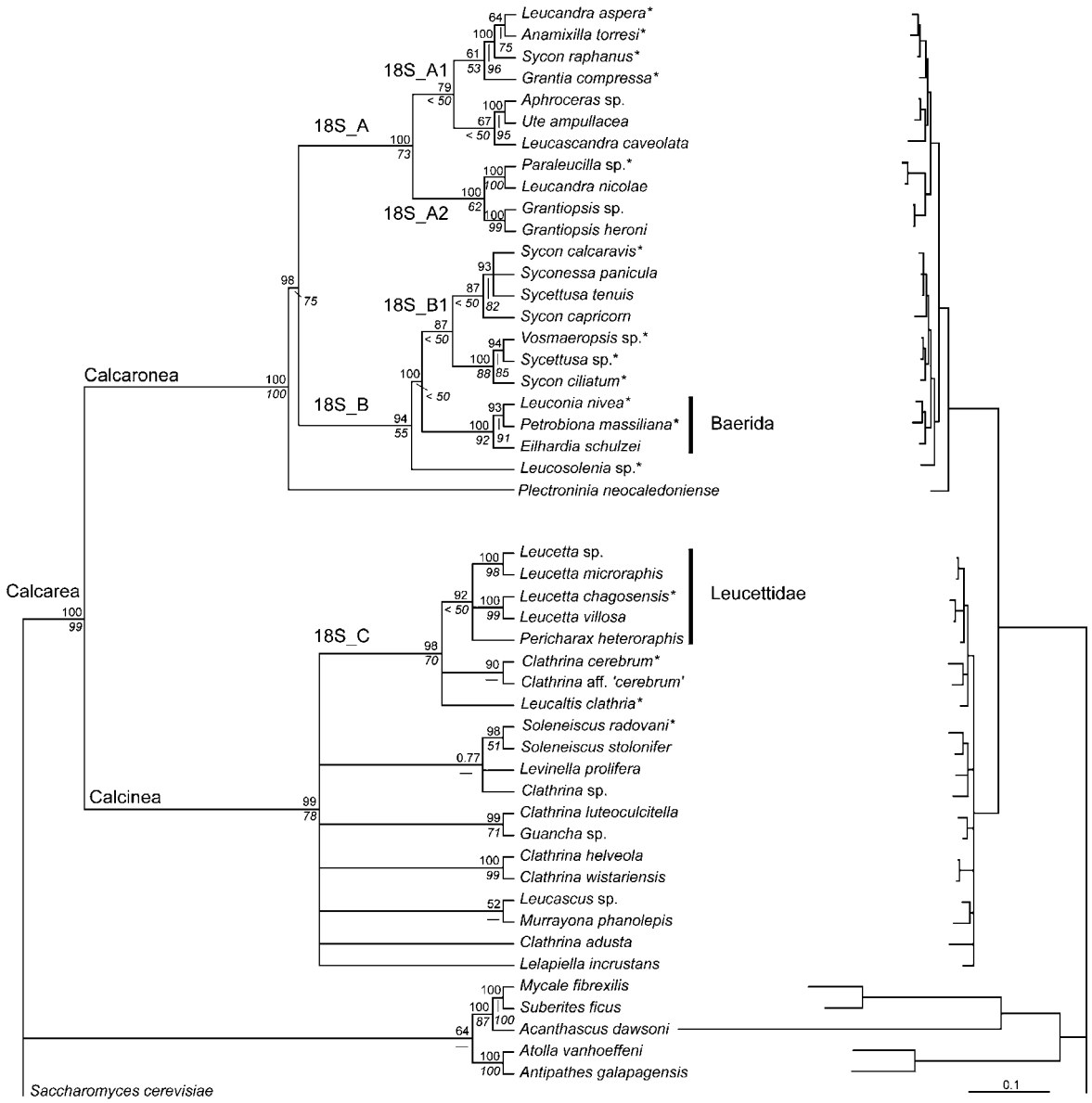


Figure 2.1: Bayesian 50% majority rule consensus tree (19,650 trees sampled; burn-in = 1500 trees) inferred from the 18S rDNA alignment under the partitioned Doublet+(GTR+I+G) model. Asterisks indicate previously published ingroup sequences. Bayesian posterior probabilities (%) are given above branches. ML bootstrap proportions (%) calculated under the TrN+I+G model are given below branches (—, clade not included in ML tree). Branch lengths (shown on the right; scale bar, expected number of substitutions per site) are proportional to the mean of the posterior probabilities of the branch lengths of the sampled trees (Huelsenbeck and Ronquist, 2005).

(BP<50). Branches within Calcinea and Calcaronea were extremely short in comparison with those of the outgroup taxa and the branches leading to the Calcarea and its two subclades.

Calcaronea 18S rDNA

Among Calcaronea, *Plectroninia neocaledoniense* (Minchinellidae, Lithonida) was the sister taxon to a well-supported (PP=98; BP=75) clade consisting of all other calcaronean species, which split into the subclades named 18S_A and 18S_B in Fig. 2.1. The Baerida (*Petrobiona massiliana*, *Leuconia nivea*, *Eilhardia schulzei*) were monophyletic but belonged to 18S_B (PP=94; BP=55), render-

ing Leucosolenida paraphyletic. They formed the sister group to 18S_B1 (PP=87; BP<50), which contained all members of Heteropiidae (*Sycettusa tenuis*, *Syconessa panicula*, *Vosmaeropsis* sp., *Sycettusa* sp.) and all but one *Sycon* species. Heteropiidae and *Sycettusa*, as well as *Sycon* (and therefore Sycettidae), were not monophyletic. *Leucosolenia* sp. was the sister taxon of 18S_B1/ Baerida (PP=100; BP<50). 18S_A (PP=100; BP=73) contained all members of Grantiidae (*Leucandra aspera*, *L. nicolae*, *Grantia compressa*, *Ute ampullacea*, *Aphroceras* sp.) and Jenkinidae (*Anamixilla torresi, Leucascandra caveolata*), as well as *Sycon raphanus*, *Paraleucilla* sp. (Amphoriscidae), and the two *Grantiopsis* species (Lelapiidae). In 18S_A1 (PP=79; BP<50), *Ute ampullacea* and *Aphroceras* sp. (both Grantiidae) grouped together and formed a clade with *Leucascandra caveolata* that was the sister taxon to the remaining species of 18S_A1 [(((*L. aspera*/*A. torresi*) *S. raphanus*) *G. compressa*)]. The positions of *L. caveolata* and *Grantia compressa* within 18S_A1 were not well supported. 18S_A2 (PP=100; BP=62) consisted of the clade *Paraleucilla* sp. *Leucandra nicolae* and a monophyletic *Grantiopsis*. The topology of 18S_A indicates non-monophyly of Grantiidae, *Leucandra*, *Sycon*, and Jenkinidae.

Calcinea 18S rDNA

The topology of Calcinea was poorly resolved by the 18S rDNA data; it contained only one well-supported clade with more than two species (18S_C in Fig. 2.1; PP=98; BP=70), which included a monophyletic Leucettidae (PP=92; BP<50), *Leucaltis clathria* (Leucaltidae), and *Clathrina cerebrum* and *C*. aff. '*cerebrum*.' The latter two species grouped together (as expected; see footnote of Table 2.1) in the Bayesian tree (Fig. 2.1), but in the ML tree (Appendix 2, Fig. A2.1), they were successive sister groups to Leucettidae. Their position and that of *L. clathria* within 18S_C was not resolved in the Bayesian tree. The same holds true for the position of *Pericharax heteroraphis* within Leucettidae; monophyly of L*eucetta* therefore remained unclear. *Soleneiscus* (Soleneiscidae) was monophyletic (PP=98; BP=51); it was associated with *Levinella prolifera* (Levinellidae) and *Clathrina* sp., however with low support. The position of this clade was not resolved, as were the positions of the remaining species. Among these, only a close relationship between *C. luteoculcitella* and *Guancha* sp., and *C. helveola* and *C. wistariensis*, respectively, was inferred. *Leucascus* sp. (Leucascidae) and *Murrayona phanolepis* (Murrayonida) formed a poorly supported clade to the exclusion of *Lelapiella incrustans* (Murrayonida). In the ML tree (Appendix 2, Fig. A2.1), *Murrayona* and *Lelapiella* only weakly grouped together (BP<50). The question of monophyly of Murrayonida and Clathrinida therefore remained open. Monophyly of Leucettidae was relatively well supported by the Bayesian analysis, whereas monophyly of *Clathrina* and Clathrinidae was not recovered by both the Bayesian (Fig. 2.1) and the ML analysis (Appendix 2, Fig. A2.1).

28S rDNA

Differences in PP values of the two independent Bayesian analyses were, where present, minimal, and topologies were identical; the tree of the first analysis is shown in Fig. 2.2 (results of second analysis not shown). Monophyly of Calcarea, Calcinea, and Calcaronea was recovered, but Calcinea received less support (PP=93; BP<50) than in the 18S rDNA tree. Silicea, Demospongiae, Porifera, and Cnidaria were also monophyletic, albeit Bayesian support for the latter two was rather

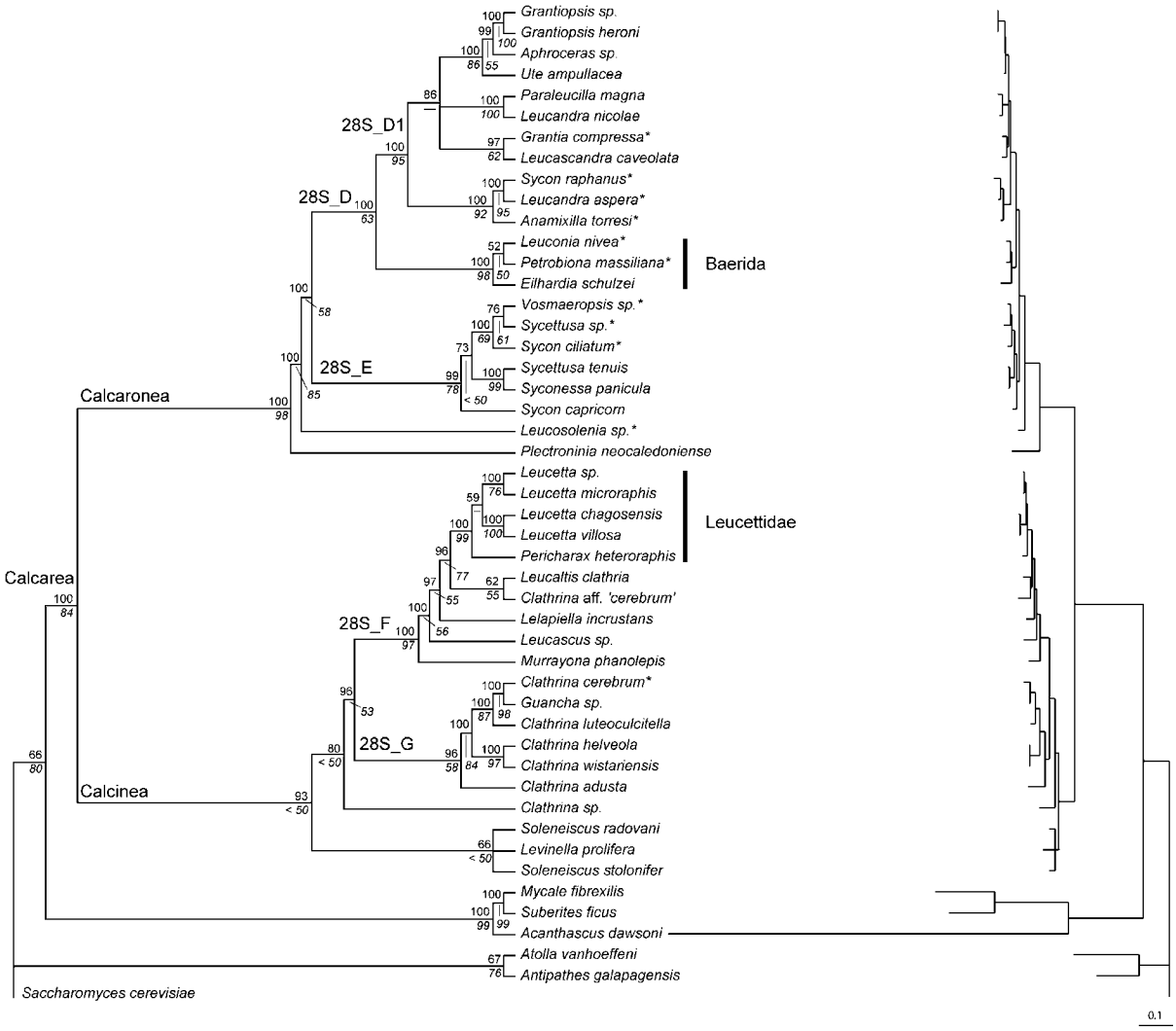


Figure 2.2. Bayesian 50% majority rule consensus tree (12,980 trees sampled; burn-in = 600 trees) inferred from the 28S rDNA alignment under the partitioned Doublet+(GTR+I+G) model. Asterisks indicate previously published ingroup sequences. Bayesian posterior probabilities (%) are given above branches. ML bootstrap proportions (%) calculated under the TrN+I+G model are given below branches (—, clade not included in ML tree). Branch lengths (shown on the right; scale bar, expected number of substitutions per site) are proportional to the mean of the posterior probabilities of the branch lengths of the sampled trees (Huelsenbeck and Ronquist, 2005).

low (PP=66 and 67, respectively). In contrast, bootstrap proportions for Porifera and Cnidaria were relatively high (BP=80 and 76, respectively). Relative branch lengths were similar to those of the 18S rDNA tree.

Calcaronea 28S rDNA

Like in the 18S rDNA tree, *P. neocaledoniense* was the sister taxon to the rest of the calcaroneans. The remaining topology differed in some respects, however: Although 28S_E in Fig. 2.2 corresponds to 18S_B1 in Fig. 2.1, and 28S_D1 corresponds to 18S_A, relationships within these clades were different. In 28S_E, *Sycon capricorn* was the sister taxon to the remaining species; in 28S_D1, *L. caveolata* and *Grantia compressa* grouped together, *Ute ampullacea* and *Aphroceras* sp. were

successive sister groups to *Grantiopsis*, and *S. raphanus* (instead of *Anamixilla torresi*) was more closely related to *Leucandra aspera*. Major differences were the placement of Baerida, which was more closely related to 28S_D1 than to 28S_E (compare with Fig. 2.1),and *Leucosolenia* sp., which was the sister-taxon to 28S_D/28S_E. Implications for (non-) monophyly of supraspecific taxa are the same as in the 18S rDNA analyses.

Calcinea 28S rDNA

Resolution within Calcinea was increased here com-pared to the 18S rDNA tree. The two *Soleneiscus* species and *L. prolifera* formed a clade that was the sister group of the remaining calcineans. The clade was poorly supported (PP=66; BP<50), and relationships between the three species were unclear, however, thereby questioning monophyly of *Soleneiscus*. *Murrayona phanolepis*, *Leucascus* sp., *Lelapiella incrustans*, and a poorly supported clade consisting of *Leucaltis clathria* and *Clathrina* aff. '*cerebrum*' were successive sister groups to Leucettidae (28S_F in Fig. 2.2; PP=100; BP=97). *Leucetta* was recovered as monophyletic by the Bayesian analysis, but with poor support (PP=59); in the ML tree (Appendix 2, Fig. A2.2), *Pericharax heteroraphis* weakly grouped with *Leucetta* sp./ *Leucetta microraphis* (BP=58). 28S_G, the sister group of 28S_F, showed a very well supported topology (except the bootstrap value for inclusion of *Clathrina adusta*; BP=58). It contained most of the *Clathrina* species, with *Guancha* sp. nested within them. Surprisingly, it also contained *C. cerebrum* (sister-taxon to *Guancha* sp.), thereby questioning a close relationship with *C*. aff. '*cerebrum*' (see above and Fig. 2.1). *Clathrina* sp. was the sister taxon to 28S_F/28S_G, but this was not well supported (PP=80; BP<50). Except the unclear status of *Soleneiscus* and a higher support for Leucettidae (PP=100; BP=99; compare with Fig. 2.1), implications are the same as in the 18S rDNA analyses. However, monophyly of Murrayonida and Clathrinida was clearly rejected (see placement of *Murrayona* and *Lelapiella* in Fig. 2.2).

Combined analysis

Differences in PP values of the two shorter independent Bayesian analyses and those of the longer run (burn-in=20,000 trees; not shown) were, where present, minimal. Topologies were identical, except of an unresolved position of *L. prolifera* within Calcinea in one of the shorter analyses (not shown). The tree of the other analysis is shown in Fig. 2.3. Monophyly of Calcarea, Calcinea, Calcaronea, Silicea, Demospongiae and Cnidaria was highly supported, but interrelationships of Calcarea, Silicea and Cnidaria (Eumetazoa) remained unclear according to the Bayesian analysis. In the ML topology (Appendix 2, Fig. A2.3), Silicea and Calcarea weakly grouped together (BP=59).

Calcaronea 18S/28S rDNA

Consistent with the results from the single-gene analyses (Figs. 2.1, 2.2), *P. neocaledoniense* was the sister taxon to the remaining calcaroneans. The position of *Leucosolenia* sp. was the same as in the 28S rDNA topology. The remaining species were distributed on two clades (Clade_H and Clade_I in Fig. 2.3). Clade_H corresponds to 18S_A in Fig. 2.1 and 28S_D1 in Fig. 2.2. Its topology more closely resembled the 18S rDNA topology, but Clade_H1 and Clade_H2 received less support (PP=69 and 70, respectively) than 18S_A1 and 2 (see Fig. 2.1) and were not contained in the ML

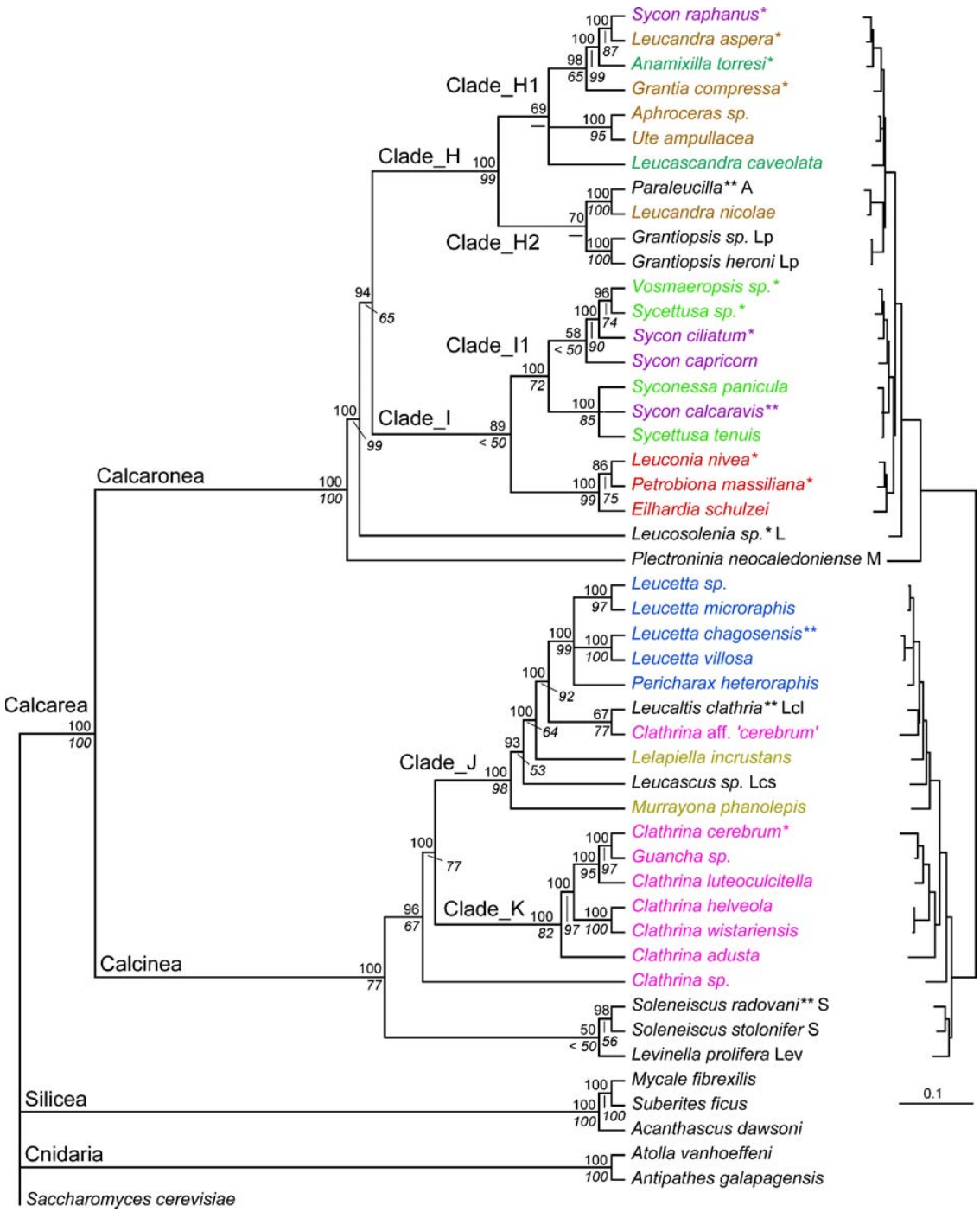


Figure 2.3: Bayesian 50% majority rule consensus tree (36,990 trees sampled; burn-in = 1000 trees) inferred from the combined 18S/28S rDNA alignment under the partitioned Doublet+(GTR+I+G)-model. Bayesian posterior probabilities (%) are given above branches. ML bootstrap proportions (%) calculated under the TrN + I+G model are given below branches (—, clade not included in ML tree). Branch lengths (shown on the right; scale bar, expected number of substitutions per site, outgroups omitted for clarity) are proportional to the mean of the posterior probabilities of the branch lengths of the sampled trees (Huelsenbeck and Ronquist, 2005). Selected species are colored according to their assignment to classically recognized supraspecific taxa; 'families' of the other species are given as abbreviations after the species names. Blue, Leucettidae; brown, Grantiidae; green, Heteropiidae; olive, Murrayonida; pink, Clathrinidae; purple, *Sycon*; red, Baerida; and turquoise, Jenkinidae. A, Amphoriscidae; L, Leucosoleniidae; Lcl, Leucaltidae; Lcs, Leucascidae; Lev, Levinellidae; Lp, Lelapiidae; M, Minchinellidae (Lithonida sensu Manuel et al., 2003); S, Soleneiscidae. * Both sequences from GenBank; **, one sequence from GenBank (see Table 2.1).

topology, where the two *Grantiopsis* species grouped with *Ute ampullacea*/*Aphroceras* sp. (Supplementary Fig. 2.3). The relationships between *S. raphanus*, *L. aspera* and *A. torresi* were identical to those recovered from the 28S rDNA analysis. The topology of Clade_I was almost identical to 18S_B excl. *Leucosolenia* sp., the only difference being the position of *S. capricorn*, which was very poorly supported, however.

Calcinea 18S/28S rDNA

The topology of Calcinea was largely identical to that of the 28S rDNA analysis, but it was generally more robust in terms of clade support. Exceptions were the resolution within Leucettidae and the monophyly of *Soleneiscus*, which correspond to the 18S rDNA tree (Fig. 2.1).

Table 2.4: Results of the comparison of constrained analyses vs. the unconstrained analysis of the combined matrix using the Bayes factor (2 ln (B_{10})).

Taxon constrained to be monophyletic	2 ln (B_{10})
Leucosolenida	31.76
Grantiidae	449.30
Heteropiidae	61.44
Jenkinidae	115.82
Sycon	414.94
Leucandra	838.48
Sycettusa	158.14
Clathrinida	160.66
Murrayonida	27.60
Clathrinidae	216.66

See Table 2.2 for interpretation.

Hypothesis testing

Evidence against monophyly of all taxa found in our analysis as non-monophyletic was 'very strong' (Table 2.4). Trees in the samples of the unconstrained analysis containing the respective constraint were only found in the cases of Murrayonida and Leucosolenida. Given their small numbers (3 and 9, respectively, out of 35,990), correcting for those topologies did not change the outcome of the calculations.

Discussion

Calcarea are notorious for being taxonomically difficult. Except from the major split into the two 'subclasses' Calcinea and Calcaronea, phylogenetic relationships of calcareous sponges have remained enigmatic for the most part, and classification schemes currently in use do not rest upon well-supported hypotheses about the underlying phylogeny. Due to limited taxon sampling, the molecular studies conducted so far provided only few detailed insights into relationships within the two 'subclasses.' With the present study, we have substantially increased taxonomic sampling of 18S and 28S rDNA for calcareous sponges and provide a much more comprehensive picture of their phylogeny. Monophyly of Calcarea and its subtaxa Calcaronea and Calcinea was strongly confirmed. In contrast, most of the 'orders', 'families' and 'genera' with more than one species sampled did not represent monophyla. Notable exceptions were the Leucettidae (Calcinea) and the Baerida (Calcaronea), the monophyly of both of which was highly supported.

Bayesian vs. ML analyses

With some exceptions (e.g. monophyly of Porifera in the 28S rDNA analyses), bootstrap proportions were generally lower than Bayesian posterior probabilities, sometimes considerably so. Especially striking was the very low bootstrap support for monophyly of Calcinea in the 28S rDNA

analysis. Also, there were some topological differences, such as the position of *Grantiopsis* in the trees of the combined analyses. However, as already mentioned, outcomes of ML and Bayesian analyses in this study were not directly comparable due to differences in the underlying evolutionary models. When compared to the Bayesian GTR-only trees that we obtained from the model testing (Appendix 2, Fig. A2.4, A2.5, and A2.6), the differences in clade support and topology were much less striking in most cases. For example, support for monophyly of Calcinea was only 69% in the Bayesian 28S rDNA GTR-only tree (Appendix 2, Fig. A2.5). This indicates that the differences between Bayesian and ML analyses in our study were largely due to suboptimal modelling in the latter and did not stem from flaws in one or the other inference method. Therefore, we consider the outcomes of our Bayesian analyses as the more reliable estimates of calcarean phylogeny. For in-depths discussions of posterior probabilities vs. bootstrap proportions, we refer the reader to Alfaro et al. (2003, and references therein) and Huelsenbeck and Rannala (2004).

Branch-lengths

Branches within Calcinea and Calcaronea were much shorter than branches outside calcareans and branches leading to the two subtaxa. This indicates that they might have undergone a relatively recent radiation, as has been proposed earlier (Borojevic, 1979; Manuel et al., 2003). Alternatively, evolutionary rates might have slowed down in the Calcinea and Calcaronea after the two lineages split. Unfortunately, there is not enough palaeontological data yet to elucidate this issue: the fossil record of modern non-hypercalcified Calcarea is generally very sparse (see Pickett, 2002), and isolated spicules cannot be assigned with certainty to one of the subgroups in most cases (Reitner, 1992).

Phylogeny of Calcaronea

The most remarkable result concerning the phylogeny of Calcaronea is probably the early-branching position of *Plectroninia neocaledoniense*. This species belongs to the Minchinellidae (Lithonida), a group that is characterized by the formation of a rigid basal skeleton composed of fused spicules (Borojevic et al., 1990; Vacelet et al., 2002a). Calcarea with rigid basal skeletons are often regarded as relicts of otherwise extinct groups of calcareous sponges that survived in cryptic habitats (Reitner, 1992; Vacelet, 1991). Such forms include not only the Minchinellidae, but also *Petrobiona massiliana* (now placed in Baerida; see Introduction) and three species of Calcinea (see next section), of which the basal skeletons are structurally very different, however (Vacelet, 1991). The position of *Plectroninia* in our inferred trees might suggest that a rigid basal skeleton composed of fused spicules is a ground-plan character of Calcaronea that got lost in the lineage leading to the 'Leucosolenida'/Baerida-clade. Alternatively, it might be a highly derived (possibly synapomorphic) character of taxa assigned to Minchinellidae. Decision between these two hypotheses depends primarily on the question whether the Minchinellidae are monophyletic or not, which could not be answered here. Since *Plectroninia* has a leuconoid aquiferous system, its non-nested position also implies that the type of aquiferous system in the most recent common ancestor of Calcarea was not necessarily asconoid, as reconstructed by Manuel et al. (2004): When mapped on the tree of the combined analysis with MacClade 4.06 (Maddison and Maddison, 2002), the

ancestral state of Calcarea was in fact equivocal (results not shown). A sistergroup relationship of Lithonida (excl. *Petrobiona*; i.e., Minchinellidae) and Baerida, as proposed by Manuel et al.(2003, Fig. 8) on the grounds of a combined morphological/18S rDNA- analysis, is not well supported in our view, because their analysis included no molecular characters of Minchinellidae, and the proposed synapomorphies (absence of an atrial cavity and no axial symmetry of the architecture of the skeleton along the body axis) can easily be interpreted as convergent losses. The remaining Calcaronea formed a well-supported monophyletic group, with Leucosolenia sp. being the sistertaxon of the rest of the species in the 28S rDNA and combined trees. The nested position of Baerida within 'Leucosolenida,' rendering the latter paraphyletic, is in agreement with earlier studies (Manuel et al., 2003; Manuel et al., 2004). There was, however, some amount of uncertainty regarding the exact placement of Baerida, given that the 18S rDNA and the 28S rDNA alignments contained conflicting signal reflected by lowered clade support in the combined analysis, so additional data is needed to resolve this issue. There were some interesting trends concerning the other supraspecific taxa classically assigned to Leucosolenida (compare Manuel, 2006, Fig. 2.8): Heteropiidae and most species of *Sycon* (Sycettidae) fell into one clade, although both groups were not recovered as monophyletic. Polyphyly of *Sycon* had already been suggested by Manuel (2001) on the basis of morphological evidence, which was later confirmed with molecular data (Manuel et al., 2003; Manuel et al., 2004). *Sycon* is a very large, cosmopolitan group and might be regarded as a kind of 'taxonomic waste bin', so this result was not surprising. Heteropiidae was found to be monophyletic by Manuel et al. (2003; 2004), which appears to be a chance result: *Sycettusa* sp. and *Vosmaeropsis* sp. were the only sampled species, and they indeed seem to be closely related, as our results confirmed. Inclusion of only two more species of Heteropiidae here led to the hypothesis of non-monophyly of Heteropiidae and *Sycettusa*. The Heteropiidae are characterized by the presence of a "sub-cortical layer of pseudosagittal triactines" (Borojevic et al., 2000; Borojevic et al., 2002c), which could be interpreted as an autapomorphy of this group. However, isolated pseudosagittal spicules also occur in other calcaroneans (e.g., *Sycon ensiferum* Dendy and Row, 1913), so this character might not be as strong an evidence for delimiting the Heteropiidae as was originally thought (see Borojevic et al., 2000, pp.234–235). The second major calcaronean clade contained all members of Grantiidae, the representatives of Jenkinidae, Amphoriscidae and Lelapiidae, as well as *S. raphanus*. Neither *Leucandra* nor Grantiidae were monophyletic, which is comprehensible, given that -like *Sycon*- both are large groups, in which a number of unspecialized, phonetically similar calcaroneans are merged. The 'family' Jenkinidae was erected by Borojevic et al. (2000) for thin-walled Calcaronea with an inarticulate choanoskeleton; in the light of our results this growth form appears to have originated several times independently instead of being due to common ancestry. A close relationship between *Aphroceras* and *Ute*, as recovered from the 18S rDNA and the combined analysis, had already been suggested by Borojevic (1966); both taxa are characterized by the presence of cortical giant longitudinal diactines (Borojevic et al., 2000; Borojevic et al., 2002c). This character also occurs in other grantiid 'genera' not included in the present study (e.g., *Sycute* Dendy and Row, 1913) and might be a synapomorphy of these taxa.

Phylogeny of Calcinea

The 18S rRNA gene apparently contains little phylogenetic information for relationships within Calcinea. Because this gene is thought to be more conserved than the 28S rRNA gene (Hillis and Dixon, 1991), this finding might indicate a more recent radiation of extant Calcinea that could only be fully resolved with the more variable 28S rRNA gene. This conclusion is supported by the fact that the branch leading to Calcinea was shorter than the branch leading to Calcaronea. Unfortunately, this hypothesis can-not be tested with palaeontological data at the moment, given the sparse fossil record of unequivocally identifiable Calcarea (see above). A split of Calcinea into Murrayonida and Clathrinida (Borojevic et al., 1990; Borojevic et al., 2002b; Vacelet et al., 2002b), and thus the idea that the former are relicts of an ancient radiation and representatives of the latter are the product of amore recent radiation (Borojevic et al., 1990; Vacelet, 1991; see also Reitner, 1992) was rejected, because *Murrayona* and *Lelapiella* were nested at different positions within 'Clathrinida.' Inclusion of *Lelapiella* in Murrayonida in the current classification is somewhat uncertain (see Vacelet et al., 2002b), and Clathrinida are defined solely by the absence of rigid basal skeletons (see Borojevic et al., 1990; Borojevic et al., 2002b), so paraphyly of the two 'orders' of Calcinea is not particularly surprising. Interestingly, all species of Clade_J in Fig. 2.3 (except *C.* aff. '*cerebrum*', see below) possess a cortex. This clearly differentiated external layer of spicules is not present in the other species, so it might be an autapomorphy of this clade. In addition, Clade_J contains all syconoid (*Leucaltis clathria*, *Leucascus* sp.) and leuconoid (*Leucettidae*, *Murrayona*, *Lelapiella*) calcinean species from our dataset, whereas the other species all have an asconoid (i.e., the most simple form of) aquiferous system. The more nested position of Clade_J is therefore in good agreement with the notion that the evolution of Calcinea progressed from simple to complex forms (Borojevic et al., 1990; see also Manuel, 2006). In all analyses, *Levinella* seemed to be somehow associated with *Soleneiscus*, albeit with weak support. The monophyly of *Soleneiscus* was recovered from the 18S rDNA and the combined analysis, but the 28S rDNA alignment contained ambiguous signal. Apart from Soleneiscidae, we were able to include more than one species from only two 'families': Leucettidae and Clathrinidae. The Leucettidae were recovered as monophyletic with high support, but internal relationships of that group were poorly resolved, and the phylogenetic status of *Leucetta* awaits further investigation (see Wörheide et al., 2004). Clathrinidae (*Clathrina*+*Guancha*) was not recovered as a monophylum, but the majority of species did form a well-supported clade. Paraphyly of *Clathrina* with respect to *Guancha* is easily comprehensible from a morphological perspective: The latter is distinguished only by possession of a peduncle (stalk) from the former, whereas all characters that are ascribed to *Clathrina* also apply to Clathrinidae (see Borojevic et al., 1990; Borojevic et al., 2002b). The positions of *Clathrina* sp. and *Clathrina* aff. '*cerebrum*' indicate non-monophyly of Clathrinidae. The placement of the latter species implies secondary morphological simplification, because it is the only asconoid species, and the only species without a cortex, in Clade_J. The possession of spines on the apical actines of tetractines links *C.* aff. '*cerebrum*' to *C. cerebrum*. Since the 18S rDNA tree is in agreement with this, *C.* aff. '*cerebrum*' appears at the same position in both single-gene trees, and repetition of extraction, amplification and sequencing resulted in the same sequences for *C.* aff. '*cerebrum*,'

we suspect that the 28S rDNA sequence of *C. cerebrum*, which was retrieved from GenBank, might have come from another *Clathrina* species.

Conclusion and outlook

Our study is by far the most comprehensive molecular phylogenetic analysis of Calcarea conducted to date, demonstrating that the existing 'order'- to 'genus'- level classification of calcareous sponges is probably largely artificial and does not reflect the phylogeny of the group. However, to assess the phylogenetic status of still underrepresented taxa (e.g., Amphoriscidae, Lelapiidae, Soleneiscidae), and to place pivotal taxa, such as *Paramurrayona* Vacelet, 1967, or those assigned to Sycanthidae Lendenfeld, 1891, it is crucial to further broaden taxonomic sampling in future studies. Furthermore, our results await corroboration by analyses of nuclear and/or mitochondrial protein-coding genes.

Authors' contributions

M.D. generated 18S and 28S sequence data, performed the sequence analyses and wrote the manuscript. G.W. collected specimens and together with M.D. designed the study and determined specimens. O.V. generated data for an additional 5' fragment of 28S, performed second strand sequencing of several 18S sequences, provided the secondary structure information for the 18S and critically reviewed the manuscript. D.E. provided secondary structure information for the 28S fragment and critically reviewed the manuscript.

Acknowledgments

We thank the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) for providing computer power, Laura Epp and Eilika Wülfing for help in the lab, and Fredrik Ronquist and Paul van der Mark for helpful discussions at the 2005 Workshop on Molecular Evolution in Woodshole, MA. Two anonymous reviewers and the editor contributed to the improvement of an earlier draft of this manuscript. This work was financially supported by the German Research Foundation (DFG, Project Wo896/3-1). Collection of most samples was facilitated by fellowship of the University-Special-Program III of the Federal Republic of Germany through the DAAD (German Academic Exchange Service) to G.W. and a research grant to G.W. and John N. A. Hooper from the Australian Biological Resources Study (ABRS), as well as additional funding from AstraZeneca R&D Griffith University, Brisbane. D.E. acknowledges financial support of the European Union under a Marie-Curie outgoing fellowship (MOIF-CT-2004 Contract No:2882). G.W. also acknowledges financial support through the European Marie Curie project HOTSPOTS (contract MEST-CT-2005-020561). We would like to thank the Great Barrier Reef Marine Park Authority for permitting the fieldwork (Permit Nos.: G98/142, G98/022).

Supplementary information

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2006.04.016.

Appendix 2

Table A2:1: Primers used for PCR and sequencing

Figure A2.1: Maximum likelihood tree inferred from the 18S rDNA alignment

Figure A2.2: Maximum likelihood tree inferred from the 28S rDNA alignment

Figure A2.3: Maximum likelihood tree inferred from the combined 18S/28S rDNA alignment

Figure A2.4: Bayesian 50% majority rule consensus tree inferred from the 18S rDNA alignment

Figure A2.5: Bayesian 50% majority rule consensus tree inferred from the 28S rDNA alignment

Figure A2.6: Bayesian 50% majority rule consensus tree inferred from the combined 18S/28S rDNA alignment

Chapter 3

Evolution of Calcareous Sponges inferred from SSU and LSU ribosomal RNA genes –new insights and remaining problems

Oliver Voigt, Eilika Wülfing and Gert Wörheide

Palaeontology & Geobiology, Department of Earth and Environmental Sciences, Ludwig-Maximilians-University Munich, D-80333 München

Abstract

Background

The class Calcarea (Phylum Porifera) is taxonomically difficult. Phylogenetic studies using ribosomal RNA genes (rDNA) have revealed many discrepancies with classically recognized taxa and the observed relationships on order, family and genus level. While several previous hypotheses for the evolution within this sponge class were questioned by these results, our knowledge about alternative scenarios is still very fragmentary. We therefore extended the available taxa and character set by sequencing the complete Small Subunit (SSU) rDNA and almost the complete Large Subunit (LSU) rDNA of additional specimens. In our phylogenetic analyses we applied and compared RNA specific models of substitution that take the special substitution patterns of rDNA into account.

Results

Methodologically, we could confirm that doublet models should be preferred to standard GTR models for paired sites in rDNA data. Six-state models are to be preferred to 7-state and 16-state models, and among tested models RNA6A is the best. Similar results were obtained when applying the tests to an alternative dataset, suggesting that they may be valid for many rRNA datasets. With our phylogenetic analyses we found several additional taxa of Calcarea to be paraphyletic: In Calcinea the families Leucettidae and Leucaltidae, and the genus *Leucetta*; in Calcaronea the family Amphoriscidae and the genus *Ute*. Several unexpected relationships were discovered. We found some species of Grantiidae with giant diactines in their cortex to be closely related the paraphyletic Heteropiidae. Another important finding was a clade of *Sycon* cf. *carteri* (Sycettidae) and *Leucascandra caveoltata* (Jenkinidae), which at first glance seem to differ essentially in their organization. Evolutionary patterns contradicting the classical taxonomy are clarified in Calcinea. We confirm that within this class, simple asconoid sponges are ancestral. Derived from such forms we find more complex forms with a cormus of branching and anastomosing tubes, and even more derived Calcinea, which are characterized by possessing a cortex and, except for one exception,

more complex aquiferous systems. We can therefore reject ideas about several parallel evolutionary lineages that led to more complex Calcinea, as was proposed before.

Conclusions

We obtained new insights into the evolution of Calcarea, especially Calcinea. However, this taxonomically difficult sponge class needs thorough revision, a task that cannot be fulfilled by morphology alone. A much broader taxon sampling is necessary to untangle the relationships and understand the evolution within this sponge group.

Background

Among the extant Porifera the three classes Demospongiae Sollas, 1885, Hexactinellida Schmidt, 1870 and Calcarea Bowerbank 1864 can readily be distinguished. Recent molecular studies suggest, that Homoscleromorpha Lévi 1973 –traditionally placed in Demospongiae– form an additional group separate from Demospongiae *sensu stricto* (Borchiellini et al., 2004). We will use the term Demospongiae following this concept. Yet, even with molecular data, the relationship between these sponge groups have long remained uncertain, especially the question of sponge paraphyly (e.g., a closer relationship of Calcarea to Eumetazoa than to the other sponge classes, Borchiellini et al., 2001 but also see Erpenbeck and Wörheide, 2007 for an overview) with its possible implications for our understanding of the evolution of Metazoa (Sperling et al., 2007). A recently published phylogenomic study contributed much to unravel the relationships of the poriferan groups and proposed the monophyly of Porifera with Homoscleromorpha and Calcarea as a sister clade to Hexactinellida and Demospongiae (Philippe et al., 2009). Calcareous sponges are characterized by the synapomorphy of calcite spicules (Manuel et al., 2002), while the skeleton of other sponges is formed by silicious spicules or spongin fibers. Still, the internal relationships of Calcarea remain largely unresolved. Calcareous sponges are taxonomically difficult (Manuel et al., 2002), and their morphological features carry a high amount of homoplasy (Manuel et al., 2003; Manuel et al., 2004; Manuel, 2006).

The diversity of organization is however relatively high. In Calcarea, all grades of organization of the aquiferous systems can be found. In asconoid species, the whole atrium is lined by choanocytes (Fig. 3.1, A). Such sponges are also homocoel, i.e. all internal cavities are lined by choanoderm, while all other organization forms are heterocoel, with parts of the internal cavities lined by pinacocytes (Poléjaeff, 1883). In syconoid Calcarea, choaonocyte chambers are radially arranged around the central atrium; water enters the sponge via inhalant canals and the choanocyte chambers via pores. The choanocyte chamber open to the atrium, which generally is lined by pinacocytes (Fig. 3.1,B). In sylleibid sponges, radially arranged choanocyte chambers do not open directly into the atrium. Instead, several choanocyte chambers open into a cavity lined by pinacoderm, which itself has an opening to the atrium (Fig. 3.1, C). The most complex organization of the aquiferous system is the leuconoid grade. Here, water enters the sponge through a system of inhalant canals leading to numerous, more or less spherical choanocyte chambers. These open to exhalant canals, through which the water reaches the atrium (Fig. 3.1, D). Other prominent features

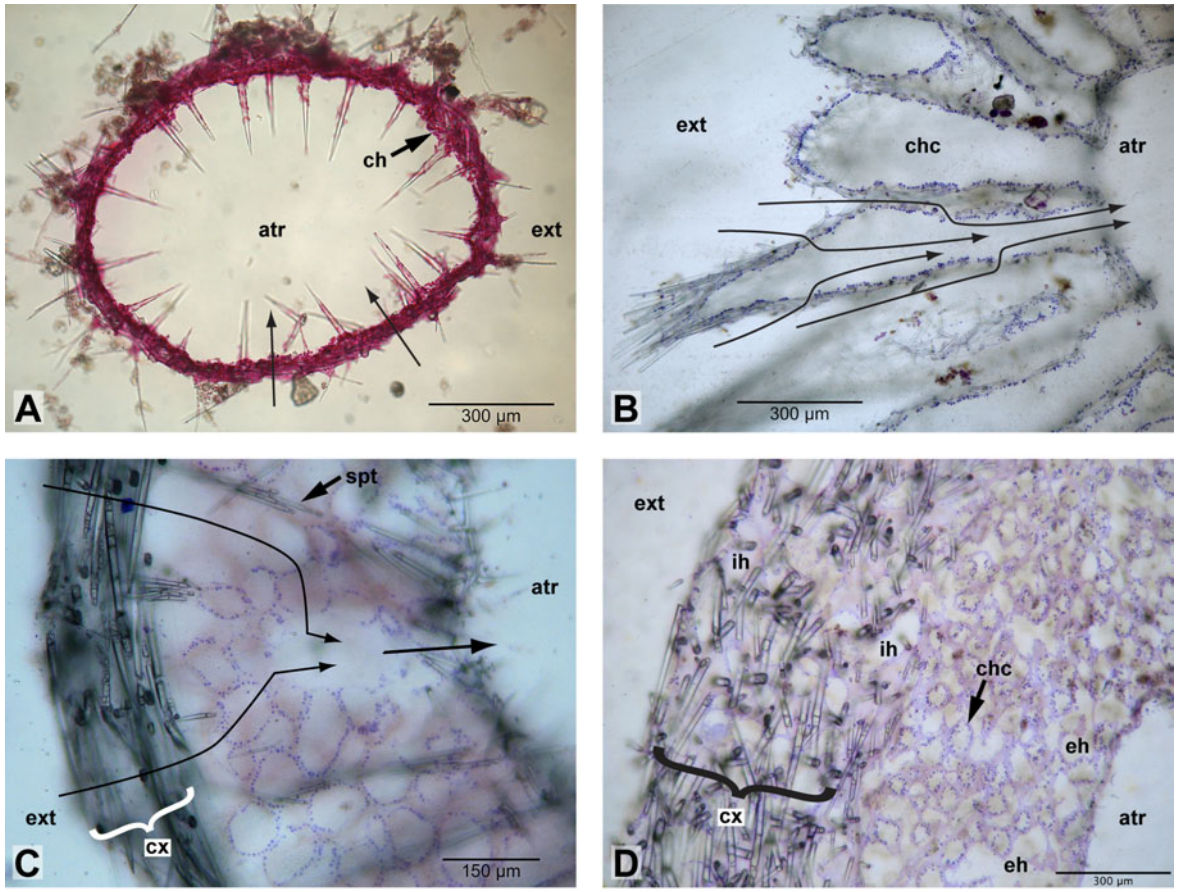


Figure 3.1: Different organizations of the aquiferous system in Calcarea.. A: asconoid (*Soleneiscus radovani*); B: syconoid (*Sycon cilatum*, collected on Helgoland, Germany); C: sylleibid (*Grantiopsis* aff. *cylindrica*); D: leuconoid (*Leucettusa* sp. 1). Thin arrows show the water flow in A,B and C. atr= atrium; ch=choanocytes; chc=choanocyte chambers; eh= exhalant channel; ext: exterior of the sponge;ih=inhalant channel.

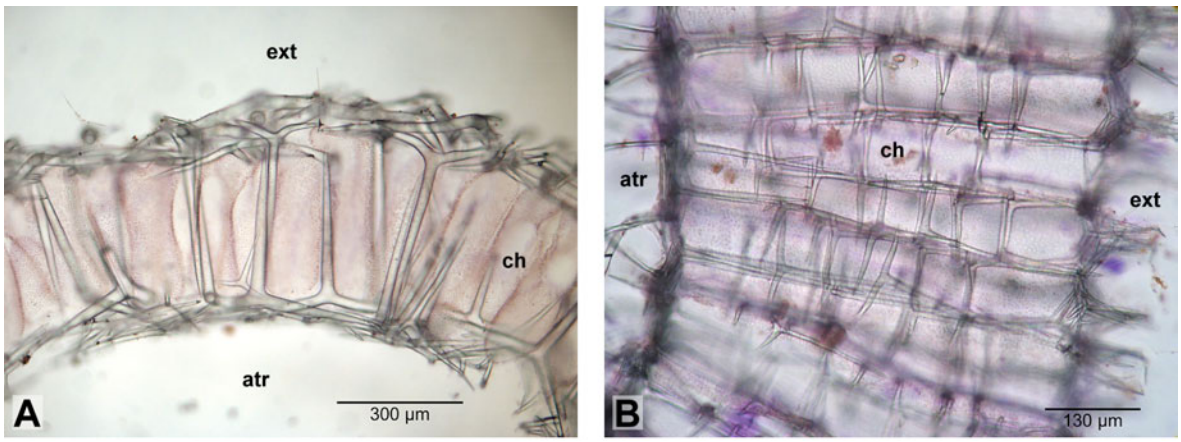


Figure 3.2: A: inarticulated choanoskeleton (*Sycettusa* aff. *hastifera*); **B:** articulated choanoskeleton (*Grantessa* sp. GW974).

of Calcarea are found in their skeleton, especially in the arrangement of their calcite spicules. In the simplest sponges, the skeleton consist of only one spicule type, which supports the ectoderm on the outside, and the choanoderm on the inside of the sponge (e.g., Fig. 3.1.A). More complex skeletons can be divided into an atrial skeleton (delimiting the atrial cavity), and the choanoskeleton (supporting the choanosome). In sponges with thin walls, the choanosome is only supported

by unpaired actines of subatrial spicules and, depending on the sponge, also from actines of (sub-) cortical spicules; such choanoskeltons are referred to as inarticulated choanoskeltons (Fig 3.2, A). So-called articulated choanoskeletons are build from several, more or less parallel, rows of similar spicules, usually sagittal triactines, with the unpaired actine pointing to the outside of the sponge. With this arrangement, the sponge can build thick walls (Fig. 3.2, B). Both forms are typical for heterocoel Calcaronea Bidder, 1898 of the Order Leucosolenida Hartman, 1958. The choanosome of thick walled sponges can also be supported by numerous spicules, without apparent order (e.g. in Leucettidae), or by spicular tracts of modified triactine (Fig. 3.1, C). Reinforced skeletons can be formed by fused (sometimes modified) spicules or an aspicular calcite mass. A tangential layer of spicules that covers the external surface of the sponge is called cortex (Manuel et al., 2003). It can be thin, formed by a single layer of spicules, or thick, sometimes primarily sustaining the sponge wall (Fig. 3.1, C&D).

The above named features are important diagnostic characters for the taxonomy of Calcarea. However, we know little about the evolution of these characters. As a consequence, almost 140 years after Haeckel's first attempt to establish a natural system for this group (Haeckel, 1872c; Haeckel, 1872a; Haeckel, 1872b), most of the classification systems that have been proposed in the meantime remain highly speculative (see Manuel, 2006). Haeckel's conception of the system of Calcarea distinguished between three 'families', "Ascones", "Sycones" and "Leucones", according to the organization of the aquiferous system. However, his 'natural' system was soon found to be artificial by subsequent taxonomists. Poléjaeff (1883) suggested a different scheme in which he separated Calcarea into the two orders Homocoela and Heterocoela, but again the system was soon questioned. Finally Bidder (1898) generalized a concept that had been used by Minchin (1896) to separate asconoid sponges, and divided Calcarea into the subclasses Calcinea and Calcaronea based upon the position of the nucleus in the choanocytes (Calcinea: nucleus basal, not linked to the flagellum, vs. Calcaronea: nucleus apical and linked to the flagellum). Independent support for this subclass division comes from different larvae types and their development in both subclasses [coenoblastula in Calcinea, amphiblastula in Calcaronea, see (Hartman, 1958)], different ratios of isotopes incorporated into the spicules during bio-mineralization (Wörheide and Hooper, 1999) and the analyses of small subunit (SSU) and partial large subunit (LSU) ribosomal RNA genes (rDNA) (Manuel et al., 2003; Manuel et al., 2004; Dohrmann et al., 2006).

According to the latest revisions of supraspecific calcarean taxonomy (Borojevic et al., 1990; Borojevic et al., 2000; Hooper and van Soest, 2002), the subclass Calcinea is divided into the orders Clathrinida Hartman, 1958 and Murrayonida Vacelet, 1981, wheras the subclass Calcaronea contains the orders Leucosolenida Hartman, 1958, Baerida Borojevic, Boury-Esnault & Vacelet, 2000 and Lithonida Vacelet,1981. The allocation of genera to families is based on several debatable ideas about the evolution of certain morphological traits (Borojevic et al., 1990; Borojevic et al., 2000; Hooper and van Soest, 2002; visualized in Manuel, 2006). Phylogenetic analyses with morphological data have shown that little phylogenetic information is present in these characters, and suggests a high level of morphological homoplasy (Manuel et al., 2003). The finding of rDNA studies that many of the classically recognized taxa are not monophyletic (Manuel et al., 2003; Manuel

et al., 2004; Dohrmann et al., 2006) is therefore not surprising. However, no convincing alternative hypothesis for the evolution of calcareous sponges explaining the evolution of skeletal arrangements and other morphological features is at hand, and studies are hampered by uncertainties in the taxonomy at order, family and species level (Dohrmann et al., 2006). In an attempt to clarify the evolution of this group, we included several new taxa to our analyses and sequenced complete SSU and almost complete LSU rDNA. Additional LSU rDNA data was generated for taxa of a previous study (Dohrmann et al., 2006).

Special care was taken in our study to analyze the data under appropriate models of nucleotide substitutions. Most phylogenetic methods assume that characters in a data matrix evolve independently from each other This assumption is clearly violated in helices of rRNA, as here nucleotides forming a pair coevolve, driven by the selection pressure to maintain the secondary structure pivotal for their function within the ribosome (Wheeler and Honeycutt, 1988; Hancock et al., 1988; Hillis and Dixon, 1991; Higgs, 2000; Savill et al., 2001; Dixon and Hillis, 1993). By neglecting these coevolutionary processes, phylogenetic inferences can be biased and result in suboptimal tree topologies (e.g., Wheeler and Honeycutt, 1988; Telford et al., 2005; Erpenbeck et al., 2007a). A solution to this problem are special evolutionary models, which instead of single bases consider the two paired bases of helices, the so-called doublet, as single characters. Such models have been shown to outperform standard 4x4 models of nucleotide evolution in analyses of rDNA data (Dixon and Hillis, 1993; Schöniger and von Haeseler, 1994; Muse, 1995; Tillier and Collins, 1995; Tillier and Collins, 1998; Telford et al., 2005; Erpenbeck et al., 2007a). Several doublet models that make different assumptions on the evolution of doublets are available (a comprehensive overview is given by Savill et al., 2001). We analyzed our data with 17 different doublet models in partitioned phylogenetic analyses and compared their performance (i.e., their fit to the actual data), to improve our understanding of calcareous sponge taxonomy.

Material and Methods

Sample collection and species identification

Calcareous sponge specimens were collected on the Great Barrier Reef near Lizard Island and in the Red Sea (Gulf of Aqaba) in 2006, additional specimens were obtained from museum collections (Table 3.1). To determine the sponges we examined the skeletal arrangements and nature of the aquiferous system in thin sections, which were prepared as follows:

Parts of the sponges preserved in ethanol (EtOH) were gradually transfered to 30% EtOH in water over a dilution series (70%, 50%, 30% EtOH). Tissue was then stained overnight in a 30% EtOH-Fuchsine solution. The stained tissue was dehydrated in a dilution series (50%, 70%, 90%, 99% EtOH-Fuchsine-solution). For embedding, the EtOH-Fuchsine solution was gradually replaced with LRwhite resin (in dilution steps of 33%, 50%, 66%, 100% LRwhite, all at 4°C to prevent polymerization and the last step with overnight incubation). For final embedding, LRwhite was exchanged and after one hour incubation at 45°C, polymerization was induced at 60°C overnight. From the resulting block, we took sections of suitable thickness (10-500 µm; starting with a 200 µm section)

Table 3.1: Included specimens of Calcarea, their sample localities and GenBank accession numbers for published sequences. New species are bold. See Dohrmann et al. (2006) for accession numbers of the already published, smaller LSU fragments. * Holotype; #SSU sequence comes from another individual.

from the block with a Leica 1600 saw microtome (Leica, Nußloch, Germany). To stain the cells and nuclei on the surface of the section, we suspended it for 1:30 min to a 30% EtOH-Touledein blue

Species	Family	Voucher	Locality	SSU	LSU
CALCINEA					
Clathrinida					
*Clathrina adusta**	Clathrinidae	QM G313665	GBR, Wisatri Reef	AM180962	extended
Clathrina cerebrum	Clathrinidae	–	–	U42452	AY563541
*Clathrina helveola**	Clathrinidae	QM G313680	GBR, Heron Reef	AM180958	extended
*Clathrina luteoculcitella**	Clathrinidae	QM G313684	GBR, Channel Wistari/Heron Reef	AM180959	extended
Clathrina sp. GW45	Clathrinidae	QM G313693	GBR, Yonge Reef	AM180960	extended
***Calthrina* sp. GW957**	Clathrinidae	GW 975	GBR, Mac's Reef	new seq.	new seq.
Clathrina wistariensis	Clathrinidae	QM G313663	GBR, Wistari Reef	AM180961	extended
Guancha sp.	Clathrinidae	QM G316033	GBR, Rene's Nook	AM180963	extended
*Soleneiscus radovani**	Soleneiscidae	QM G313661#	GBR, Wistari Reef	AF452017	extended
Soleneiscus stolonifer	Soleneiscidae	QM G313668	GBR, Wistari Reef	AM180955	extended
Levinella prolifera	Levinellidae	QM G313818	GBR, Hook Reef	AM180956	extended
***Ascandra* sp.**	Leucaltidae	QM G323326	Tasmania, King Island Canyons	new seq.	new seq.
Leucaltis clathria	Leucaltidae	QM G316022#	GBR, DJ's reef	AF452016	extended
***Leucettusa* sp. 1**	Leucaltidae	QM G323232	Tasmania, Ling Hole	new seq.	new seq.
***Leucettusa* sp. 2**	Leucaltidae	QM G323283	Tasmania, Ling Hole	new seq.	new seq.
***Leucettusa* sp. 2**	Leucaltidae	QM G323253	Tasmania, King Island Canyons	new seq.	new seq.
Ascaltis sp.	Leucascidae	QM G313824	South Pacific, Pitcairn Islands	AM180957	extended
Leucascus sp.	Leucascidae	QM G316051	GBR, Hook Reef	AM180954	extended
indet. Calcinea	?	QM G323250	Tasmania, King Island Canyons	new seq.	new seq.
Leucetta chagosensis	Leucettidae	QM G316279#	Coral Sea, Osprey Reef	AF182190	extended
Leucetta microraphis	Leucettidae	QM G313659	GBR, Wistari Reef	AM180965	extended
Leucetta sp.	Leucettidae	QM G313691	GBR, Yonge Reef	AM180964	extended
*Leucetta villosa**	Leucettidae	QM G313662	GBR, Wistari Reef	AM180966	extended
Pericharax heteroraphis	Leucettidae	QM G316295	Coral Sea, Holmes Reef	AM180967	extended
Murrayonida					
Murrayona phanolepis	Murrayonidae	QM G313992	Coral Sea, Osprey Reef	AM180968	extended
Lelapiella incrustans	Lelapiellidae	QM G313914	Vanuatu	AM180969	extended
CALCARONEA					
Baerida					
Petrobiona massiliana	Petrobionidae	–	Mediterranean, Marseille	AF452026	new seq.
Eilhardia schulzei	Baeridae	QM G316071	GBR, Mac's reef	AM180980	extended
Leuconia nivea	Baeridae	–	–	AF182191	extended
Lithonida					
Plectroninia neocaledoniense	Minchinellidae	QM G316300	Coral Sea, Holmes Reef	AM180979	extended
Leucosolenida		–	–		
Leucosolenia sp.	Leucosolenidae	–	–	AF100945	AY026372
Sycon capricorn	Sycettidae	QM G316187	GBR, Ribbon Reef 3	AM180970	extended
Sycon* cf. *carteri	Sycettidae	SAM PS 0142	Australia, Ulladulla	new seq.	new seq.
Sycon ciliatum	Sycettidae	–	–	AJ627187	AY563532
Sycon raphanus	Sycettidae	–	–	AF452024	AY563537
Grantia compressa	Grantiidae	–	–	AF452021	AY563538
Teichonopsis labyrinthica	Grantiidae	SAM PS 0228	South Australia, Kangaroo Island	new seq.	new seq.
*Ute amupllacea**	Grantiidae	QM G313669	GBR, Wistari Reef	AM180972	extended
Ute* aff. *syconoides 1	Grantiidae	QM G323233	Tasmania, King Island Canyons	new seq.	new seq.
Ute* aff. *syconoides 2	Grantiidae	QM G313694	GBR, Yonge Reef	new seq.	new seq.
Ute* aff. *syconoides 3	Grantiidae	GW 975	GBR, Lizard Island	new seq.	new seq.

Table 3.1, continued.

Species	Family	Voucher	Locality	SSU	LSU
Synute pulchella	Grantiidae	WAM Z1404	West Australia, Reru Island	new seq.	new seq.
Leucandra aspera	Grantiidae	–	–	AF452022	AY563535
*Leucandra nicolae**	Grantiidae	QM G313672	GBR, Wistari Reef	AM180974	extended
Leucandra sp.	Grantiidae	QM G316285	Coral Sea, Osprey Reef	AM180971	extended
***Aphroceras* sp.**	Grantiidae	SAM PS 0349	Tasmania, Waterfall Bay	new seq.	new seq.
Leucascandra caveolata	Jenkinidae	QM G316057	GBR	AM180973	extended
Anamixilla toressi	Jenkinidae	–	–	AF452020	AY563536
Syconessa panicula	Heteropiidae	–	–	AM180976	extended
Sycettusa* aff. *hastifera	Heteropiidae	GW 893	Red Sea, Gulf of Aqaba	new seq.	new seq.
Sycettusa* cf. *simplex	Heteropiidae	ZMA POR11566	Western Indian Ocean, Amirantes	new seq.	new seq.
Sycettusa tenuis	Heteropiidae	QM G313685	GBR, Heron Reef	AM180975	extended
Sycettusa sp.	Heteropiidae	–	–	AF452025	AY563530
Vosmaeropsis sp.	Heteropiidae	–	–	AF452018	AY563531
***Grantessa* sp. 1**	Heteropiidae	GW 974	GBR, Lizard Island	new seq.	new seq.
***Grantessa* sp. 2**	Heteropiidae	GW 979	GBR, Lizard Island	new seq.	new seq.
Leucilla sp.	Amphoriscidae	ZMA POR5381	Caribbean, Netherlands Antilles	new seq.	new seq.
Paraleucilla magna	Amphoriscidae	GW 824#	Brazil, Arailal de Cobo	AF452023	extended
Grantiopsis cf. cylindrica	Lelapiidae	GW 973	GBR, Lizard Island	new seq.	new seq.
*Grantiopsis heroni**	Lelapiidae	QM G313670	GBR, Wisatri Reef	AM180975	extended
Grantiopsis sp.	Lelapiidae	QM G313969	Coral Sea, Osprey Reef	AM180977	extended

and 30% basic fuchsine solution, then immediately washing off the extant dye with water. The dried and stained sections were mounted on microscopic slides with Eukitt (Fluka). Spicules were obtained either from the lysis of the DNA extract (see below) or by dissolution of tissue with sodium hypochlorite. Obtained spicules were washed twice with water and transferred to a microscopic slide, dried, and mounted on microscope slides.

Sections and spicule preparations were observed and documented on a Zeiss Axiolab Microscope with a mounted Canon PowerShot G2 digital camera. Images were imported to the Macnification software (http://www.macnification.com/) and calibrated with images taken from scale slide (Leitz). Scale bars were generated in Macnification.

Identification of calcarean genera followed available keys (Hooper and van Soest, 2002). Species were identified when possible by comparing original descriptions to our specimens. However, the descriptions in many cases were not detailed enough to allow unambiguous species identification. For species identification of specimens that were only identified to the genus level, it will be necessary to compare holotypes with the available material. Short descriptions and comments on the determination is given in Appendix 3, Supplementary information 3.1.

DNA extraction, PCR, sequencing and alignment

DNA was extracted with the QIAgen DNeasy tissue kit or by standard phenol-chloroform extraction. Template DNA was used in dilutions of 1:1 to 1:500 in PCR reactions, depending on DNA quantity and quality. Because many samples from museum collections yielded only highly degraded DNA, it was necessary to amplify SSU rDNA in two and LSU rDNA in up to five smaller fragments. PCRs were carried out with the BioTaq (BioLine) as described before (Dohrmann et al., 2006), with different combinations of the primers given in Appendix 3, Table A3.1. The purified PCR products were sequenced on an ABI 3100 capillary sequencer (Applied Biosystems). Consen-

sus sequences were created in CodonCode Aligner(http://codoncode.com). Sometimes it was not possible to amplify all SSU or LSU fragments for a given sample or the sequences of different fragments did not overlap. In such cases, we combined the sequences by aligning them to the most similar full 28S sequence, and recoding the missing parts as gaps.

Additional SSU rDNA and LSU rDNA sequences from Calcarea and outgroup taxa were downloaded from GenBank (http://www.ncbi.nlm.nih.gov/) (Table 3.1 for Calcarea and Appendix 3, Fig. A3.2 for outgroup taxa). Outgroup sequences were only included, when both SSU and LSU sequences were available in almost full length (with exception of hexactinellid 28S sequences due to limited availability). We aligned the sequences in Seaview (Galtier et al., 1996), taking into account secondary structure information (SSU: Voigt et al., 2008; LSU :Schnare et al., 1996). The C-Domain in LSU was excluded for the outgroup taxa for our analyses and was coded as 'gaps' in the alignment, because homology of sites between all taxa could not be established with certainty. By doing so, it was possible to keep the calcarean sites of this most variable region in the analyses, and here alignment was straightforward. Further sites of uncertain homology were removed from our alignment, and custom made perl scripts (Voigt et al., 2008) were used to generate input files including secondary structure information suitable for PHASE (www.bioinf.manchester.ac.uk /resources/phase/index.html). An additional script was developed to calculate the frequencies of doublets in each sequence (available upon request).

Phylogenetic analyses

In contrast to standard 4x4 models of nucleotide substitution, in doublet models the paired nucleotides in an RNA helix are the single characters. Three families of doublet models can be distinguished by the number of recognized doublets (Savill et al., 2001). In 16-state models, all possible pairs are considered. The likelihood is calculated in a 16x16 matrix, resulting that a general reversible model includes 119 free rate parameters and 15 free frequency parameters. Such a high number of parameters make general reversible 16-state models impractical to use (Savill et al., 2001). Moreover, because mismatch base pairs (MM), i.e. other pairs than Watson-Crick pairs and GU/UG pairs, are rare in real RNA data, these states are pooled into one class (MM) in 7-sate models, or ignored completely in 6-state models. Each model family has a number of different models, which through restrictions and assumptions reduce the number of parameters compared to the most general model (Table 3.2). In a previous study with a five taxon set Savill et al.(2001), comparisons within each model family suggested that the most general models are to be preferred over restricted ones. However, these results did not allow concluding which family of RNA models (16-state, 7-state or 6-state models) describes the evolution of RNA better. We aimed to test which of the models listed in Table 3.2 is best fitting our dataset. To test if some of our findings are transferable to other real world datasets, we re-analyzed an independent SSU dataset (Voigt et al., 2008) with the most general model of each model family.

We decided to only use a concatenated dataset of SSU and LSU rDNA in our analyses. Furthermore, we partitioned the combined dataset only into two partitions called 'stem' (paired sites)

and 'loop (unpaired sites), but did not account for the different genes (SSU or LSU). This was done for the following reasons:

1. Combining SSU and LSU data of the same organism makes sense from a biological point of view. Both these RNA genes are not independent phylogenetic markers; they are part of one cistron and in transcription forming one pre-rRNA before the splicing of the internal transcribed spacer regions (ITS). Also, because both genes contain variable as well as conserved parts, the substitution rates can be more similar between certain helices (or loops) of SSU and LSU, than between helices or loops within the same gene. We applied to each model for stem and loop partition two additional free parameters: a proportion of invariant sites and a gamma distribution to account for rate heterogeneity in the data.
2. For our testing of the doublet models, we would have had to test all possible combination of models for SSU and LSU models. Although it would have been possible to analyze the genes separately and evaluate the best fitting model for each, the tree topologies of such single gene trees would have been different, and possibly affecting the optimal model for the stem positions in a combined analysis. By combing SSU and LSU stems or loops, respectively, the number of analyses were kept feasible.
3. Mismatch states in real data are known to be rare and therefore it can be difficult to satisfactorily estimate the frequency and rate parameters for the MM states (Savill et al., 2001). Especially for 16-state models a larger number of MM characters in the dataset is desirable. A combined SSU and LSU stem partition has therefore to be preferred.

Table 3.2: Tested doublet models. In our analyzes, additional parameters were included, i.e. the REV model parameter for partition loop, and parameters for proportion of invariant sites and gamma distribution to account for rate heterogeneity for both partitions.

Model	Frequency parameters	Rate parameters	Constraints	Free parameters	Restrictions/ characteristics	Reference
RNA6A	**6**	**15**	**2**	**19**	General reversible	Savill et al. (2001)
RNA6B	**6**	**3**	**2**	**7**	As RNA6A, plus α_1= single transition; α_2= double transversions	Savill et al. (2001)
RNA6C	**3**	**3**	**2**	**4**	As RNA6B, plus base-pair reversal symmetry in frequencies	Tillier (1994)
RNA6D	**3**	**2**	**2**	**3**	As RNA6C, plus no double transitions	Tillier (1994)
RNA6E					Nested in RNA6B, no double transitions	Savill et al. (2001)
RNA7A	**7**	**21**	**2**	**26**	General reversible	Higgs (2000)
RNA7B	**4**	**21**	**2**	**23**	As RNA7A, plus base-pair symmetry	Savill et al. (2001)
RNA7C	**7**	**10**	**2**	**15**	As RNA7A, plus no double substitutions	Savill et al. (2001)
RNA7D	**7**	**4**	**2**	**9**	As RNA7A, α_1= single transition; α_2= double transversions; α_3= substitution rate with state MM	Tillier and Collins (1998)
RNA7E	**7**	**2**	**2**	**7**	Restriction of RNA7C and RNA7D; α_1= substitution rate with state MM	Tillier and Collins (1998)
RNA7F	**4**	**4**	**2**	**6**	Restriction of RNA7C and RNA7D; base-pair frequency and substitution symmetry	Savill et al. (2001)
RNA16A	**10**	**5**	**2**	**19**	Restricted from a general reversible 16-state model: α_1= double transition; α_2= double transversions, α_3= single substitutions; α_4= MM-MM substitution	Savill et al. (2001)
RNA16B	**16**	**1**	**2**	**15**	As RNA16A, plus no double substitutions	Schöniger and von Haeseler (1994)
RNA16C	**7**	**5**	**2**	**10**	As RNA16A, plus only one frequency parameter for MM	Savill et al. (2001)
RNA16D	**4**	**4**	**2**	**6**	Generalization of RNA16E and RNA16F:4 frequency parameters for single nucleotides, α_1= transversion rate, λ_1, λ_2 to control fitness of GU/UG pairings.	Savill et al. (2001)
RNA16E	**4**	**3**	**2**	**5**	As RNA16E, plus λ_2=1 (GU/UG treated as MM)	Muse (1995)
RNA16F	**4**	**3**	**2**	**5**	As RNA16D, plus $\lambda_1 = \lambda_2$ (GU/UG treated as standard doublets)	Muse (1995)

4. Combined analyses with four instead of two partitions would have doubled the free parameters of the dataset, resulting in much higher computational time.
5. Previous studies with data from calcarean SSU and a smaller LSU-fragment have shown that the combination of both genes lead to a finer phylogenetic resolution, compared to single gene analyses (esp. with SSU DNA, Manuel et al., 2004; Dohrmann et al., 2006).

Phylogenetic analyses were performed in PHASE (www.bioinf.manchester.ac.uk/resources/phase/index.html). MrBayes (Ronquist and Huelsenbeck, 2003; Huelsenbeck and Ronquist, 2001) was used to verify the results, but here only one doublet model (SH) is implemented. In PHASE, we applied the most general 4x4 model REV (Tavaré, 1986) for the loop partition. In different phylogenetic analyses we used one of 17 different doublet models (Table 3.2) implemented in PHASE for the stem partition. In an additional run, the REV model was applied on the partition stem to compare it to the doublet models.

Each run was started with the command mcmcphase and had a burn-in phase of 1,000,000 generations, followed by 10,000,000 generations, in which every 200th tree was sampled. We used Tracer v1.4.1 (http://tree.bio.ed.ac.uk/software/tracer/) to monitor parameter sampling of each run and for the calculation of Bayes factors between runs. To transform the PHASE output files to a readable Tracer format we modified the Perl script phase2tracer.pl from Matt Yoder (available at http://hymenoptera.tamu.edu/rna/download.php) to handle larger PHASE2 output files (available upon request). Trees were visualized with FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

To test whether our results on model performance were applicable to other rDNA datasets, we re-analyzed an SSU rDNA dataset of a previous study (Voigt et al., 2008) with models RNA6A, RNA7A and RNA16A and calculated Bayes factors.

We also tested an alpha release of RAxML 7.1.1 (Stamatakis, pers. comm.), which contains many of the doublet models implemented in PHASE2. We ran analyses with the implemented models RNA6A (S6A), RNA7A (S7A) and RNA16A (S16A), with a GTRMIX model to the loop partition under the rapid bootstrap algorithm.

Results

Model comparison

The analyses with different models resulted in mostly congruent topologies. A strict consensus of the the trees from all 17 doublet model analyses is shown in Suppl. Fig. 3.2., where differences in tree topologies between the different runs appear as polytomies. The posterior probabilities for nodes recovered in all trees are also shown in this figure.

Bayes factor comparisons (Appendix 3, Table A3.2) suggested that the model RNA6A (in combination with the other model parameters and the corresponding tree topology) fits the data best. In this comparison, 6-state models perform a better fit to the data than 7-state models, which themselves have to be preferred to the use of 16-state models (Appendix 3, Table A3.2, Fig. 3.3). There

is 'very strong' (following Kass and Raftery, 1995) evidence against the standard 4 state model REV, when it is compared to any of the doublet models. Of each of the doublet model families, the most general and parameter rich models (RNA6A, RNA7A and RNA16A) performed better than the more restricted models. However, there is no direct correlation between performance of the model and the number of its parameters in general. E.g., RNA 7A has the most parameters among the tested models (because RNA16A is already a restricted 16-state model), but according to our Bayes factor comparison only ranks on position 6. Also, within each doublet model family such correlation does not exists, as models with less parameters in several cases perform better than models with more parameters (e.g., RNA6E, RNA7C and RNA16B, Fig. 3.3). When comparing all models, we come to following ranking:

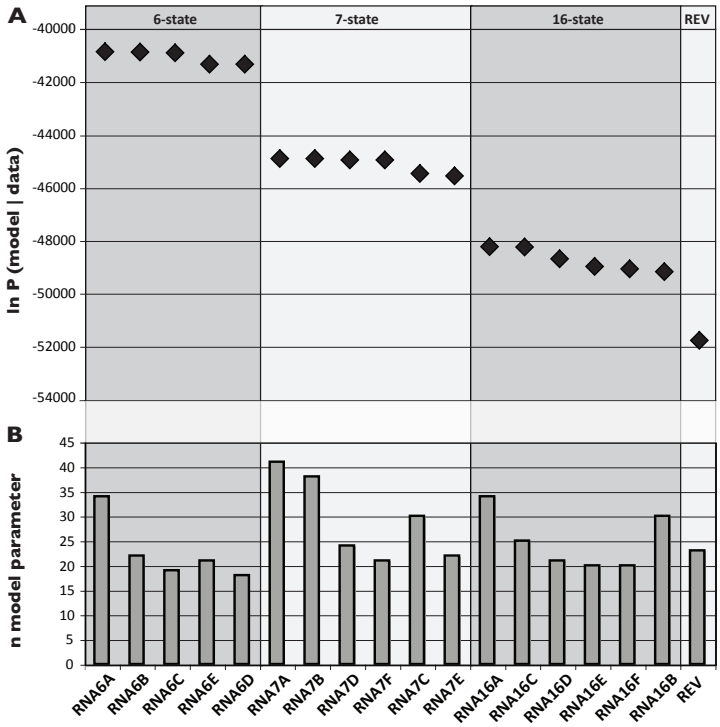


Figure 3.3: Comparison of models for partition stem, ordered by rank (according to Bayes factor comparison, Appendix 3, Suppl. Table 3.2) with the best fitting model on the left. A. Marginal likelihood estimates from Tracer. B. Number of free parameters (partition stem and loop).

RNA6A, RNA6B, RNA6C, RNA6E, RNA6D, RNA7B, RNA7D, RNA7F, RNA7C, RNA7E, RNA16A, RNA16C, RNA16D, RNA16E, RNA16F, RNA16B, REV, where evidence of the former model against the following one is "very strong" in all but the following cases: for RNA6B-RNA6C and RNA7A and RNA7b there exists "strong" evidence and in the comparison RNA7D-RNA7F "positive" (See Appendix 3, Table A3.2), using the interpretation of 2 $\log_e$ Bayes factors as proposed by Kass and Raftery (Kass and Raftery, 1995). With the likelihood analyses in RAxML, the best likelihood was recovered by the RNA6A, the lowest with the RNA16A (Table 3.3).

Table 3.3: Likelihood of best trees recovered in RAxML 7.1.1 under different doublet models for stem sites.

Doublet model	- ln likelihhod
RNA6A	-41584.848
RNA7A	-46035.954
RNA16A	-47824.480

Tree Topologies

Deep Metazoan relationships

All phyla but Porifera are monophyletic with high support by posterior probability values (PP) in all analyses (Fig. 3.4). By the application of different doublet models, the relationships among certain

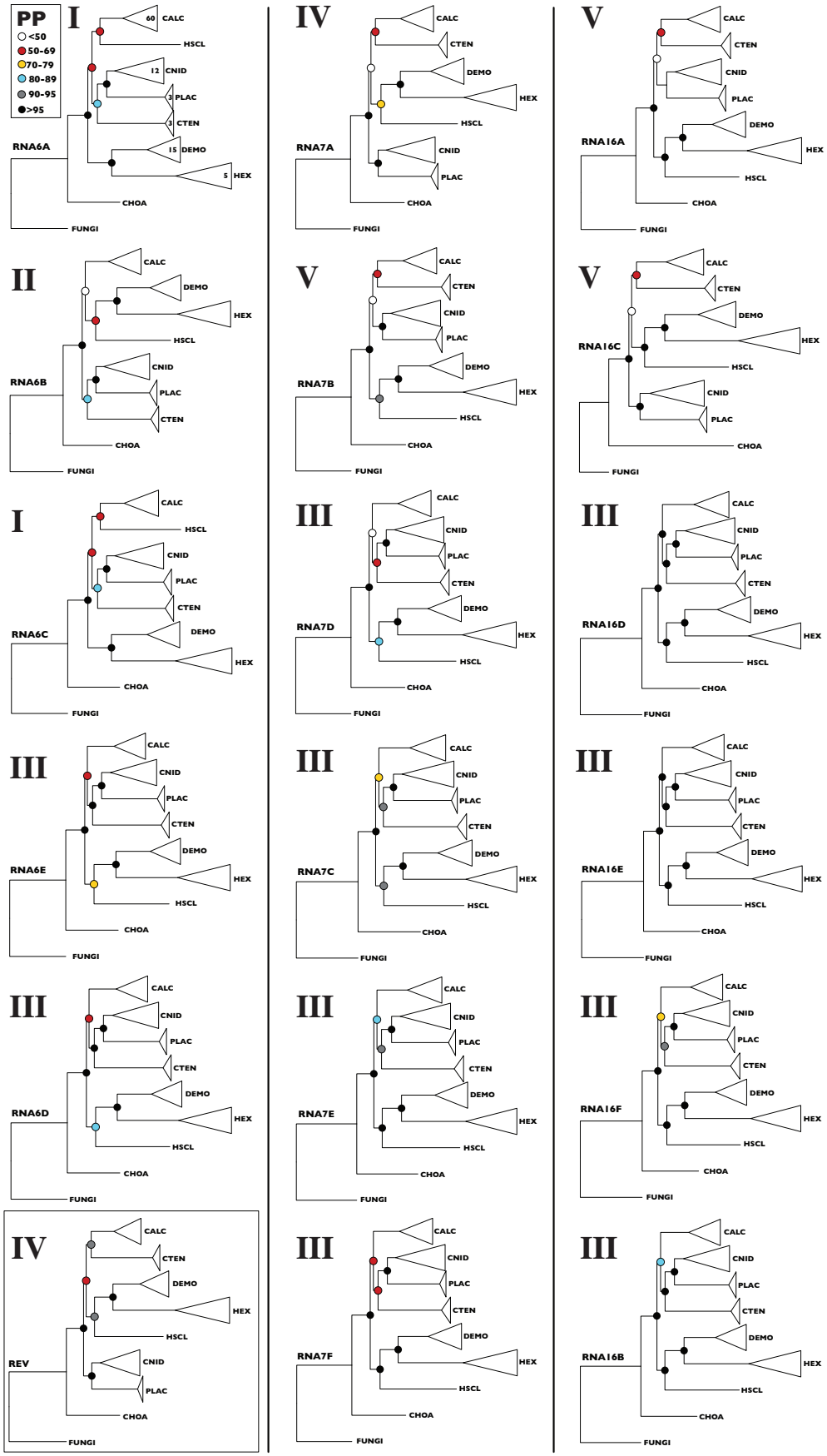


Figure 3.4: Relationships and PP support for deeper metazoan relationships. Numbers in collapsed branches in the RNA6A shows the number of included sequences. The different topologies are labeled with roman numbers. CALC=Calcarea; CHOA=Choanoflagellida; CNID=Cnidaria; CTEN=Ctenophora; DEMO=Demospongiae; FUNGI=Fungi; HEX=Hexactinellida; HSCL=Homoscleromorpha; PLAC=Placozoa.

phyla and higher taxa of Porifera, as well as the support for these different topologies varies considerably (Fig. 3.4). Nonetheless, some nodes remain constant with very high support in all trees: monophyletic Metazoa, Cnidaria and Placozoa and Silicea GRAY 1867 (Demospongiae *sensu stricto* + Hexactinellida) recieve high PP support in all analyses.

Other relationships between higher taxa are recovered differently depending on which RNA model was used in the analysis. In the analysis with the best fitting RNA6A model (Appendix 3, Table A3.2) Homoscleromorpha and Calcarea form a clade (with low support) which is the sister taxon to the Eumetazoa (which themselves are supported with 84 PP). In the latter, Ctenophora is sister to (Cnidaria + Placozoa). Therefore, Porifera are paraphyletic according to this result, with the highly supported Silicea being a sister group to a low supported (PP 65) clade of ((Calcarea+Homoscleromorpha) +Eumetazoa). When applying other models, the relationships of these less supported clades varied from the one described. In total, we recovered five different topologies with varying support (Fig. 3.4). Note that especially some of the 16-state models give very high support for clades, which are not recovered (or only with considerably lower PP support) by the more fitting 6-state and 7-state models. For example, the analyses with RNA16E and RNA16D result in high PP values for all clades, and suggest paraphyletic Porifera, with Homoscleromorpha+(Demospongiae+Hexactinellida) as sister group to Calcarea +Eumetazoa.

Relationships of Calcarea

Our inferred phylogeny (Fig. 3.5) confirms former results that most of the higher taxa of Calcarea below subclass level are not monophyletic, while the subclasses Calcinea and Calcaronea form two highly supported clades . Compared to Dohrmann et al. (2006), our topology shows a finer resolution in parts of the tree. Additionally several clades that had no high PP and BS in the former analyses were not recovered in the analyses of our extended taxon and character set (e.g., our topology does not contain Clade H1 and clade H2, in Fig. 3 in Dohrmann et al., 2006).

Relationships within Calcinea

In the subclass Calcinea, the order Murrayonida, represented by two *Murrayona phaneolepis* und *Lelapiella incrustans*, is not monophyletic (*Murrayona phanolepis* forming a low supported clade with *Leucascus* sp.). Both species are nested within Clathrinida.

At the base of Calcinea, the relationships presented in Fig. 3.5 do not find high support (below 86 PP) and therefore have to be considered as unresolved. The basal diverging taxa comprise two species of *Soleneiscus* (Soleneiscidae), *Ascandra* (Leucaltidae), *Levinella* and an undetermined *Clathrina* species (*Clathrina sp.* QM G313693). The family Clathrinidae and the genus *Clathrina* are not monophyletic, because the latter *Clathrina* species is not included in the clade that contains all other representatives of the family and the genus, respectively. *Clathrina* sp. GW957 was not included in former studies and forms a well supported clade with *Clathrina adusta.* Otherwise relationships in Clathrinidae I corresponds to clade K in Dohrmann et al. 2006 (Dohrmann et al., 2006).

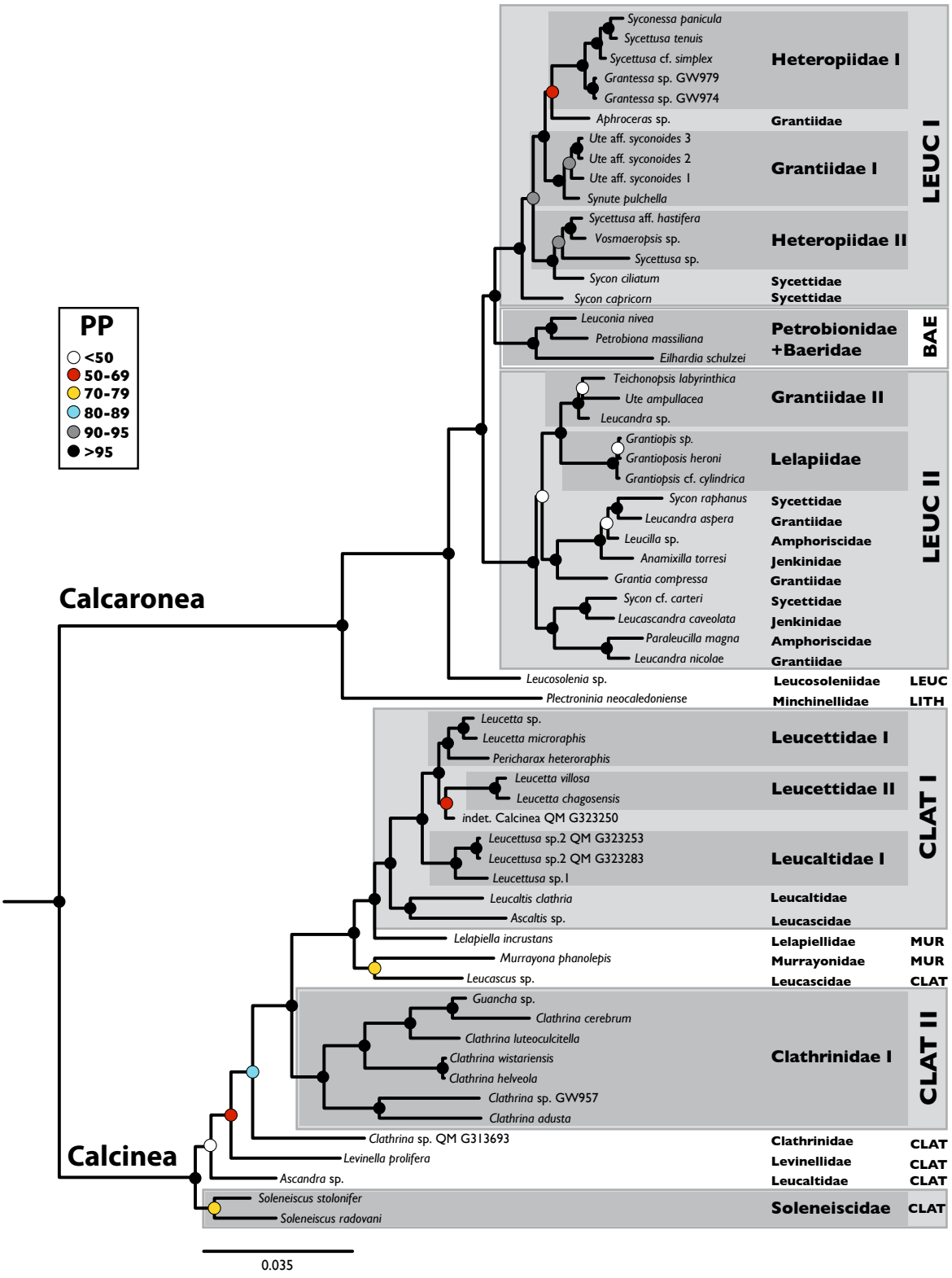


Figure 3.5: Phylogeny of Calcarea calculated with the RNA6A for partition stem. Outgroup taxa not shown (compare Suppl. Fig. 3.2 for details). Clades are shaded and numbered for taxa that are not not monophyletic. Order names are abbreviated: BAE= Baerida; CLAT= Clathrinida; LEUC= Leucosolenida; LITH= Lithonida; MUR=Murrayonida. PP= Posterior probability of clades.

We included additional taxa from two genera of the Family Leucaltidae (Order Clathrinida) compared to the previous study of Dohrmann et al. (2006): *Ascandra* sp. and three specimens representing two undetermined species of the genus *Leucettusa*. None of the genera are closely related to each other or to *Leucaltis clathria ,* the other included species of Leucaltidae. Instead, *Ascandra* is associated with So*leneiscus and Levinella,* thus closely related to other taxa with asconoid grade of organization*. The Leucettusa* species, in contrast form a sister group to Leucettidae (including a undetermined Calcinea sample) and together with these a sister clade to another clade formed by *Ascaltis* sp.[1] (Leucascidae) and *Leucaltis clathria* (Leucaltidae). Leucaltidae are therefore polyphyletic.

The undetermined calcinean specimen QM G323250 that falls into the clade of Leucettidae may represent a *Leucascus* species. It has a clear distinction between a cortex and the choanosome. Unfortunately the organization of the choanocyte chambers are not clearly recognizable in the section, but the choanoskeleton seems to be restricted to the choanosome and to form anastomosing and ramified tubes, therefore not representing a species of Leucettidae. Within Leucettidae, *Leucetta microraphis* and *Leucetta* sp. GW43 are more closely related to *Pericharax* than to *L. chagosensis* and *L. villosa.* Thus, the previously suggested monophyly of Leucettidae is doubtful, and our data does not support a monophyletic genus *Leucetta*.

With the presented relationships, the following classically recognized taxa of Calcinea are not monophyletic (citations are given, if the result was included in former studies):

Orders (with more than one sampled family):

Order Clathrinida (Dohrmann et al., 2006)

Order Murrayonida (Dohrmann et al., 2006)

Families (with more than one sampled genus):

Clathrinidae (Dohrmann et al., 2006, but see above for Ascaltis)

Leucettidae, but only supported by a yet undetermined calcinean specimen

Leucaltidae

Genera *(with more than one sampled genus):*

Clathrina (Dohrmann et al., 2006, but see above for *Ascaltis*)

Leucetta

Relationships within Calcaronea

In Calcaronea the only sampled Lithonida species *Plectroninia neocledoniense* is the sister group to a clade comprising the other sampled Calcaronea, of which *Leucosolenia* is diverging basally.

[1] Note that the specimen of *Ascaltis* sp. previously was referred to as "*Clathrina* aff. *cerebrum*" (Dohrmann et al., 2006). However, after re-examination we found it to represent an *Ascaltis* species (Appendix 3.1) of the family Leucascidae.

Order Baerida (clade BAER) is nested within Leucosolenida, forming a highly supported clade with LEUC I.

The clade LEUC I (Fig. 3.5) comprises two clades of Heteropiidae, several Grantiidae that are not a monophyletic group, and two *Sycon* species from family Sycettidae. Heteropiidae I contains two *Sycettusa* species which are paraphyletic with *Sycettusa panicula* more closely related to *Syconessa tenuis*, and, as sister group to all the latter species, two *Grantessa* specimen (that are probably conspecific). *Aphroceras* sp. (family Grantiidae) is diverging basal to Heteropiidae I, but this grouping has only very low PP support. Grantiidae I is the sister clade to (Heteropiidae I + *Aphroceras)* and contains specimens of *Synute pulchella* and *Ute* aff. *syconoides,* which like *Aphroceras* belong to the family Grantiidae and likewise have giant longitudinal diactines supporting their cortex (Fig. 3.6). The three specimens of *Ute* aff. *syconoides* may represent the same species. *Synute pulchella* has a very distinctive organization consists of fused syconoid units, each of which has a similar organization to the single tubes of *Ute* aff. *syconoides* (Fig. 3.6). This syconoid units are covered by a common cortex (Fig. 3.6).

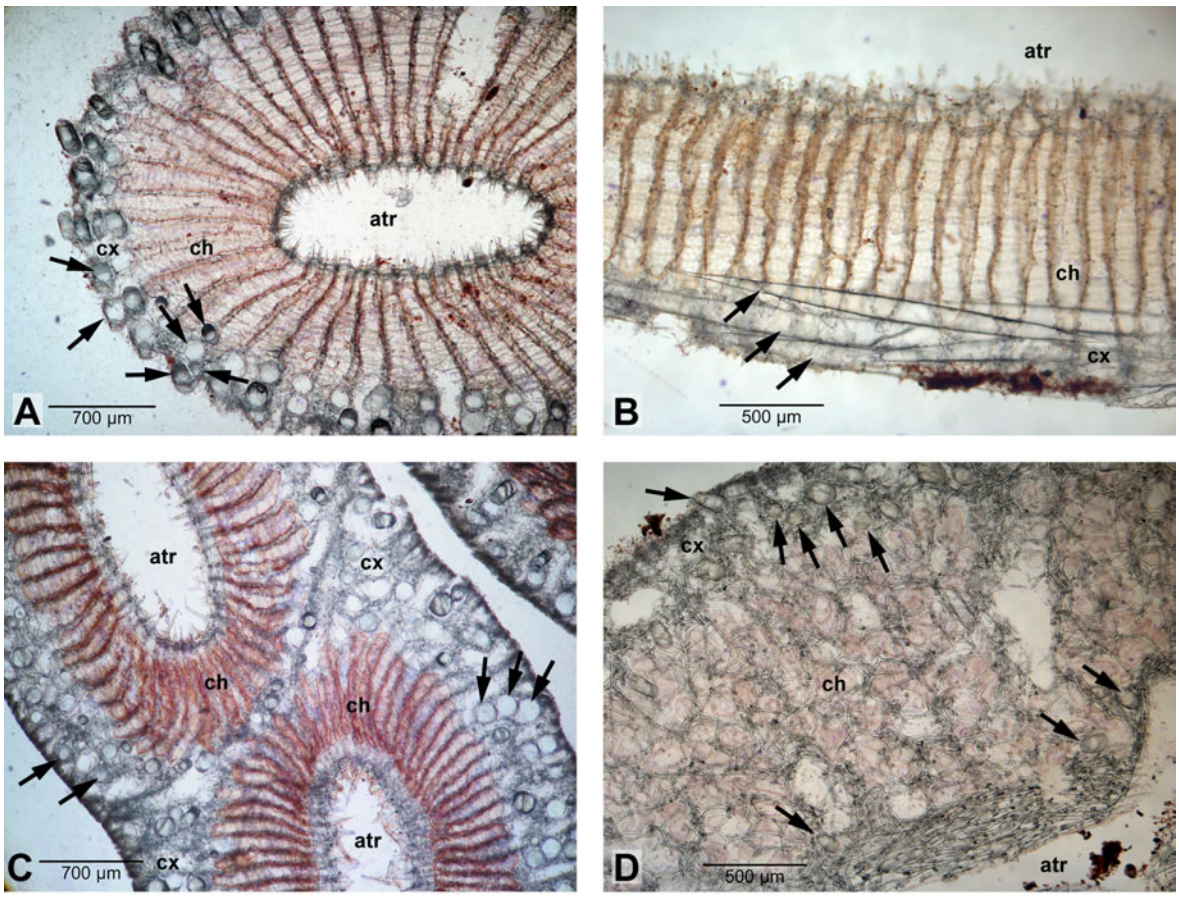


Figure 3.6: Skeletal organization of Grantiidae of clade LEUC 1: A,B: *Ute* aff. *syconoides* (GW975) in cross section (A) and longitudinal section (B); C: Cross section of *Synute pulchella*; D: Cross section of *Aphroceras* sp.. Arrows point to the giant longitudinal diactines. ch= choanosome; atr= atrium; cx=cortex.

The clade Heteropiidae II has high support and comprises two *Sycettusa* species, of which one (*S.* aff. *hastifera*) is more closely related to *Vosmaeropsis* than to the second *Sycettusa* species. Subsequently, *Sycon capricorn a*nd *Sycon ciliatum* (Sycettidae) diverge basally within LEUC I.

In Baerida, *Eilhardia schulzei* is sister taxon to (*Petrobiona massiliana* +*Leuconia nivea*), resulting in Baeridae being not monophyletic.

In LEUC II Grantiidae II and Lelapiidae are forming a clade with high support. Grantiidae II comprises *Teichonopsis*, *Ute ampullacea* and *Leucandra* sp.. Therefore, *Ute* is clearly not monophyletic (*Ute* aff. *syconoides* belongs to clade Grantiidae I), and also the genus *Leucandra* is paraphyletic, as *Leucandra nicolae* and *Leucandra aspera* are found neither in close relationship to each other nor to *Leucandra* sp.[2].

Within the remaining taxa in LEUC II, Jenkinidae, Amphoriscidae and additional taxa of Sycettidae and Grantiidae appearing in this clade are not monophyletic. A clade of (((*Sycon raphanus* +*Leucandra aspera*)+*Leucilla* sp.)+*Anamixilla toressi*) is the sister clade to (Lelapiidae+ Grantiidae II). Basal diverging from the clade consisting of the former taxa is the clade ((*Sycon* cf. *carteri* +*Leucascandra caveolata*) (*Paraleucilla magna* +*Leucandra nicolae*)).

With the presented relationships, the following classically recognized taxa of Calcaronea are not monophyletic (citations are given if the result was included in former studies):

Orders (with more than one sampled family):

Leucosolenida (Manuel et al., 2003; Manuel et al., 2004; Dohrmann et al., 2006)

Families (with more than one sampled genus):

Amphoriscidae

Baeridae (Dohrmann et al., 2006)

Grantiidae (Manuel et al., 2003; Manuel et al., 2004; Dohrmann et al., 2006)

Heteropiidae. (Dohrmann et al., 2006)

Jenkinidae (Dohrmann et al., 2006)

Genera (with more than one sampled species):

Sycon (Manuel et al., 2003; Manuel et al., 2004; Dohrmann et al., 2006)

Leucandra (Dohrmann et al., 2006)

Ute

[2]Note that the specimen of *Leucandra* sp. previously was referred to as "*Aphroceras* sp." (Dohrmann et al., 2006), but reexamination of the sections led us to the conclusion that it belongs to genus *Leucandra* (Suppl. data DESC).

Discussion

Comparison of RNA models

A comprehensive comparison of doublet models was provided by Savill et al. (2001) through the analysis of only the paired sites of rRNA of a five taxon dataset. Our analyses differs significantly in several respects. First, our dataset is much larger (105 taxa). Second, we estimated the phylogeny while it was fixed in the previous study. Third, we analysed the data in a Bayesian framework, allowing us to compare all models with Bayes factors. Fourth, the unpaired nucleotides were not omitted from our dataset and we conducted partitioned analyses. Because in our partitioned analyses we did not fix the model parameters of the REV model for partition loop, these parameters are part of the tested models. However, we found that any fixation would have been completely arbitrary as it would introduce bias in the inference of the topology itself, clearly the optimization of the parameters of the loop model changes with the overall topology of the tree. It should be kept in mind when we refer to a model for the paired partition in the following discussion, we also refer to the REV model of the loop sides with the specific parameters in the corresponding analysis.

All doublet models outperform the 4 state standard REV model for paired sites with our dataset, highlighting the importance to use RNA specific substitution models instead of standard 4-state models (Hudelot et al., 2003; Telford et al., 2005; Erpenbeck et al., 2007a). According to our Bayes factor comparison, we find that 6-state RNA models are to be preferred to 7-state models which themselves perform better than 16-state models. While 16-state models account for all 16 possible pairings of doublets, seven state models merge so-called mismatches (doublets other than canonical Watson-Crick pairing and GU and UG pairs) into a single character. In 6-state models, mismatch pairs are ignored completely. Therefore, some data is ignored in 6-state models, and it seems surprising that they still perform better than seven state and 16-state RNA models. However, when we have a closer look at the doublets that are present in our dataset (Appendix 3, Fig. A3.3), it becomes clear that most of the doublets are standard canonical Watson-Crick pairs. Considerable less doublets are GU and UG base pairs, and only a very small fraction of doublets falls into the mismatch category. From our results it seems that ignoring these mismatches is a better strategy than to invest too much computational power into estimating rate and frequency parameters for each mismatch doublet in 16-state models or even for the pooled mismatch doublets in 7-state models.

Despite the differences between our analyses and the previous one (Savill et al., 2001), our results support the finding of the previous study that the most general doublet model is preferred over more restricted ones. Within each family of doublet models we also could verify most of the other observations made by Savill et al. (2001) e.g. they concluded that double substitutions in doublet models should not be set to zero. Indeed, we also find that the models where some or all double substitutions are set to zero are the least preferred in their family: RNA6E and RNA6D among 6-state models and RNA7C and RNA7E in 7-state models. Also the restriction of reversal base-pair symmetry for frequencies (model RNA6C, RNA6D, RNA7B, RNA7F) according to our test has to be

rejected in favor of the corresponding more general models without base pair symmetry, consistent with the previous work (Savill et al., 2001). Additionally, however, we were able to compare 6-state, 7-state and 16-state models with each other using Bayes factor comparisons, and found that 6-state models performed best. To test whether this result is dependent on the dataset, we reanalyzed an independent SSU rDNA dataset (Voigt et al., 2008) with models RNA6A, RNA7A and RNA16A. A Bayes factors comparison confirms that also with this dataset model RNA6A fitted the data best, while the 7-state model was preferred to the 16-state model.

Model-dependent results

The analysis under different models show that the PP for clades can vary and give a misleading support for certain clades. Here, we mostly concentrated on the deep relationships between phyla and poriferan classes. The relationships or the support for certain clades vary very much depending on the model. Analyses with RNA16E and RNA16D result in a topology where the relationships are resolved with a very high PP and suggest paraphyletic Porifera. Noteworthily, according to our Bayes factor comparison, any other doublet model outperforms these two models, and with the better fitting models some of the clades receive considerably lower support or even a lead to a different topology. By applying a suboptimal doublet model, a bias is introduced that leads to the overestimation of PP, thus suggesting overconfidence in the topology. Therefore, model choice is an important issue with doublet models. Unfortunately, no method or software does exists yet to *a priori* propose a doublet model before analyses, unlike e.g. MODELTEST (Posada and Crandall, 1998) for many of the standard 4x4 models. Therefore, Bayes factor comparisons are the only method presently available to choose the best fitting model, but since it requires complete Bayesian analyses, it is time-consuming and computationally intensive. However, with our dataset, all 6-state models outperform 7-state and 16-state models. Considering similar results with our reanalyses of the SSU rDNA dataset we speculate that this could generally be the case with rRNA data, because the mismatches that are ignored by 6-state models make up only a small fraction of all the doublets in rRNA. Therefore, when a test of models is computationally too demanding, we propose to analyze similar rDNA data with the RNA6A model for the stem partition.

Tree topologies

Relationships of phyla and poriferan classes

Several relationships in the topology in Fig. 3.5 are only supported with low PP (note, however, the differences in models, see above). E.g., we did not find support to resolve the monophyly of Porifera, as suggested by a recent phylogenomic study (Philippe et al., 2009). We assume that the phylogenetic signal in our data is not sufficient to resolve these deep relationships. Several factors may contribute to this lack of resolution, but probably the most important is the suboptimal taxon sampling. Calcarea are over-represented in our dataset to enable unraveling deeper relationships, and especially in Demospongiae we only cover a small fraction of the diversity, because no additional full length LSU rRNA was available. More taxa of Homoscleromorpha should also be included.

However, (Placozoa+Cnidaria) and (Demospongiae+Hexactinellida) are robust clades in our analyses regardless of the applied model. The close relationship of Placozoa and Cnidaria is in conflict with two recently published studies, in which Placozoans are the sister group to the remaining non-Bilateria based on concatenated mitochondrial (mt) genes or concatenated mt and nuclear genes (Schierwater et al., 2009; Dellaporta et al., 2006). However, especially with the mt data, it is obvious that applying more sophisticated models of amino acid evolution, such as the CAT model (Lartillot and Philippe, 2004), decreases the support for basal diverging Placozoa (compare e.g. Fig. 4 in Kayal and Lavrov, 2008). This suggests that analyzing the mt data under suboptimal models in the mentioned studies (Schierwater et al., 2009; Dellaporta et al., 2006) introduced a strong bias to the tree topology (together with methodological problems such as too short mcmc runs). In contrast, our findings are consistent with tree topologies that were recovered by analyses of nuclear genes of fully sequenced genomes (Srivastava et al., 2008, although here no ctenophores were included) and SSU rRNA data (Collins, 1998).

A sister group relationship between Hexactinellida and Demospongiae, as recovered in our analyses, was also found in a former molecular study of SSU DNA (Collins, 1998) and combined SSU and LSU analyses (Medina et al., 2001). Recent studies on hexactinellid phylogeny, in contrast to our results, recover paraphyletic demosponges (Dohrmann et al., 2008; Dohrmann et al., 2009). Unfortunately, we could only include taxa from one of their two recovered demosponge clades. However, the monophyly of demosponges is highly supported by data from 17 demosponge mt genomes (Wang and Lavrov, 2008). One reason for the unexpected result of Demospongiae in the study from Dohrmann et al. (2008) may be the bias in the taxon sampling that included a large number of hexactinellid sequences and only few Demosponge taxa.

One of the most important findings in the comparison of the doublet models is that the application of suboptimal models can result in topologies that significantly differ from the ones recovered with the best fitting model, with high support values for nodes that are not present or supported by analyses with better fitting models. In our case such suboptimal models resulted in trees with high support for sponge paraphyly.

Evolution of Calcarea

With our extended taxon and character set we gained further insights into the evolution of Calcarea and could compare the results with previously proposed hypotheses about their evolution. While our results again provide support for the division into the two subclasses Calcinea and Calcaronea, the difficulties with the classical taxonomy of Calcarean orders, families and genera is getting more and more evident with the inclusion of more taxa.

Evolution of Calcinea

Borojevic and co-workers (Borojevic et al., 1990; Borojevic et al., 2000), proposed a scenario for the evolution of Calcinea and Calcaronea, which is mostly based upon concepts of a gradual evolution in Calcinea and Calcaronea from a simple olynthus grade of organization to more complex forms in several independent lines (Borojevic et al., 1990; Borojevic et al., 2000; Manuel, 2006). In

Calcinea, five evolutionary paths have been proposed by the authors and visualized by Manuel (2006): lineage one and two (Soleneicidae and Levinellidae, respectively) are represented by *Soleneiscus* and *Levinella* in our dataset. But as the relationships between these genera have low support values, our data is not sufficient to decide for or against the scenario. However, together with *Ascandra*, these genera (if they prove to be monophyletic) diverge early in the evolution of Calcinea. The third lineage is Leucaltidae. According to the scenario, *Ascandra*-like sponges developed from an olynthus-like ancestor by an increase of the sponge body diameter and the formation of internal folds of the choanoderm. Subsequently, a primitive syconoid aquiferous system and an atrial skeleton were gained, followed by the formation of a secondary choanoskeleton. Thereby, the atrial and choanoskeleton are interpreted as secondary gains, while the cortex of *Leucaltis* and *Leucettusa* corresponds to the skeleton of the olynthus-like ancestors. In a further evolutionary line that lead to the leuconoid Leucettidae, the sponge body was interpreted to have evolved by formation of a cormus of branching and anastomosing tubes (Borojevic et al., 1990; Manuel, 2006). The cortex in this lineage was a new formation, covering the cormus and by subsequent development of inhalant and exhalant systems, accompanied with a compacting of the anastomosing choanoskeleton evolved into the solid sponge body of leuconoid Leucettidae.

The hypothesis of the independent evolution in Leucaltidae has to be rejected due to the polyphyly of this family. In our dataset, three of four genera of Leucaltidae are included: *Ascandra* HAECKEL,1872, *Leucaltis* HAECKEL,1872, and *Leucettusa* HAECKEL,1872, but they are not closely related. In our tree *Ascandra* is closely related to asconoid sponges like S*oleneiscus*, *Levinella* and *Clathrina* species. Although the deep folds in the choanosome of *Acascandra* have been regarded to as analogous to the syconoid grade of organization in Calcaronea (Borojevic et al., 2002b), we find it clearly representing a special modification of typical asconoid aquiferous system, since *Acascandra* is homocoel and lacks an atrial skeleton. Therefore, the position among other asconoid specimens in our tree is reasonable. The position of *Leucaltis* and *Leucettusa* in clade CLAT I contradicts the idea that the atrial skeleton and the choanoskeleton in these genera are secondary formations (Borojevic et al., 1990; Manuel, 2006).

Similarly to Leucaltidae, Murrayonida are also not monophyletic, as previously found (Dohrmann et al., 2006), and do therefore not represent an independent evolutionary lineage that independently evolved a cortex, leuconoid aquiferous system and a reinforced skeleton (Borojevic et al., 1990).

In summary, we observe asconoid taxa basally diverging in Calcinea (*Soleneiscus*, *Levinella* and *Ascandra*). In *Calthrina* sp. and clade CLAT II, that subsequently branch off, we find asconoid taxa forming cormi of branching and anastomosing tubes (Fig. 3.7, A, C). The remaining Calcinea, (including the genera *Murrayona*, *Lelapiella*, *Ascaltis*, *Leucascus*, *Leucaltis*, *Leucettusa*, *Leucetta* and *Pericharax*) form a monophyletic clade, and are characterized by the possession of a cortex (Fig. 3.7, B). According to our character mapping, this clade had a leuconoid ancestor, the syconoid organization was secondarily gained in *Leucaltis*, and a secondary asconoid aquiferous system is present in *Ascaltis*.

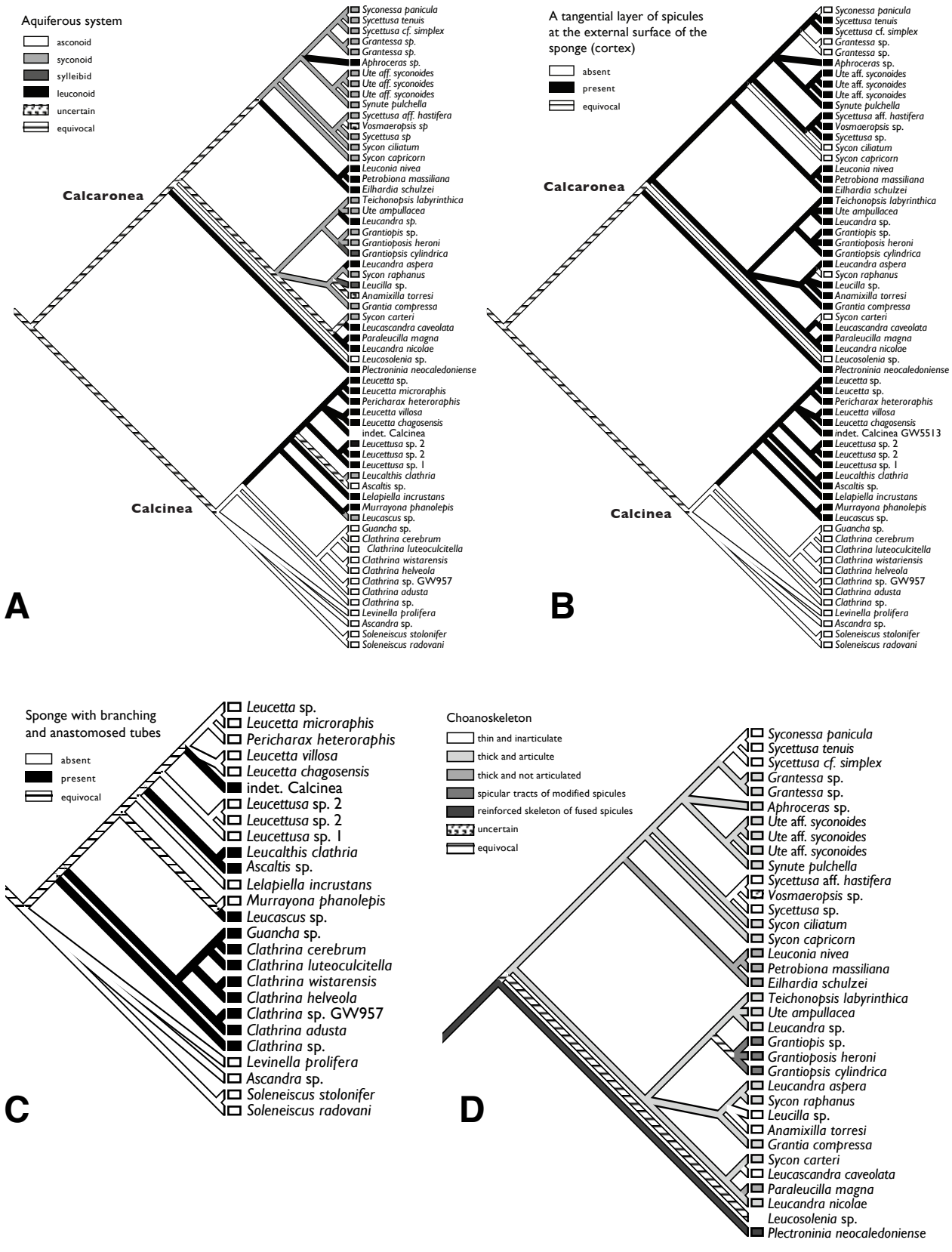


Figure 3.7: Evolution of morphological characters. Tree topology as in Fig. 3.5 with the poorly supported node collapsed. **A:** Organization of the aquiferous system; **B:** Evolution of a cortex; **C:** Evolution of a branching and anastomosed growth form in Calcinea; **D:** Evolution of the choanoskeleton in Calcaronea.

Evolution of Calcaronea

Additionally to conflicts with classical taxonomy in Calcaronea that have been reported before (Manuel et al., 2003; Manuel et al., 2004; Dohrmann et al., 2006), our topology suggest the polyphyly of the family Amphoriscidae and the genus *Ute*, and our phylogenetic hypotheses of the relationships of Calcaronea has an increased resolution compared to the previous studies.

The three new representative specimens of the genus *Ute (Ute* aff. *syconoides)* are found within the highly supported clade LEUC I. They form a clade with *Synute*, a sponge with a distinct organization, consisting of (in our case several, see Appendix 3, Suppl. information 3.1) fused syconoid units, that are covered by a common cortex containing giant longitudinal diactines. Dendy (1892) already noted the close resemblance of *Synute pulchella* and *Ute argentea* regarding the organization of the canal system and of the skeleton. *Ute* (aff.) *syconoides* is even closer to *Synute pulchella* by possessing an articulated skeleton supporting the radial tubes. Therefore, the monophyly of the two 'species' agrees with expectations from morphological observations. However, *Ute ampullacae*, which shows a similar skeletal arrangement, is not related to *Synute* or *Ute* aff. *syconoides* and is not included in LEUC I. Additionally to *Ute* aff. *syconoides* and *Synute*, LEUC I comprises *Aphroceras* (also Grantiidae), two clades of Heteropiidae, *Sycon cilatum* and *Sycon capricorn*. Heteropiidae is characterized by the a sub-cortical layer of pseudosagittal spicules (Borojevic et al., 2000). Borojevic et al. (2000) pointed out that such spicules sporadically occur in other Calcareonea, especially in species with a strong cortex and e.g. in *Sycon ensiferum*, but that they are not organized as subcortical layers. Considering our phylogeny, the parasagittal spicules could however have evolved in LEUC I (maybe after *Sycon capricorn* split off), and than been lost in Grantiidae I, *Aphroceras*, and *Sycon ciliatum*.

The close relationship of *Ute* aff. *syconoides*, *Synute* and *Aphroceras* to the clade Heteropiidae I is unexpected. These three genera of Grantiidae are characterized by giant longitudinal diactines, that support the cortex (Fig. 3.6). Such spicules are also present in the genus *Heteropia* of the family Heteropiidae. Borojevic et al. (2000) mention, that several genera of Grantiide and Heteropiidae have an analogous skeletal complexity, but because of the lack of a subcortical layer of pseudosagittal spicules they are not more closely related to each other (Borojevic et al., 2000). Such analogous genera are *Ute* (Grantiidae)- *Heteropia* (Heteropiidae) and *Amphiute* (Grantiidae)- *Paraheteropia* (Heteropiidae). However, at least with *Ute* aff. *syconoides*, *Synute* and *Aphroceras* the giant longitudinal diactines may prove to be a better indicator of closer relationships with the heteropiid species, than the presence or absence of subcortical pseudosagittal spicules. The inclusion of *Amphiute*, *Heteropia* and *Paraheteropia* in a molecular phylogeny is therefore most desirable.

In our sampling, Heteropiidae I contains only specimens from the coasts of Australia and the Indian Ocean, while the specimens of Heteropiidae II were collected in the Mediterranean and the Red Sea. This observation may be coincidental, but illustrates the need for worldwide sampling, also for other taxa. In our analysis, most specimens originate from Australian museums' collec-

tions, resulting in an overrepresentation of taxa from this region. The paraphyly of Heteropiidae would not have been discovered, had we not included samples from other oceans.

In our analyses we can show that the family Amphoriscidae is paraphyletic, because *Leucilla* sp. is not closely related to *Paraleucilla magna* (Fig. 3.5). Instead, *Leucilla* sp. is more closely related to *Anamixilla* (Jenkinidae). Interestingly, both these species have an inarticulated choanoskeleton. This questions the taxonomical value of the diagnostics of Amphoriscidae, the presence of a subcortical layer that is supported by the apical rays of giant tetractines.

The paraphyly of Jenkinidae has been reported previously (Manuel et al., 2003; Manuel et al., 2004; Dohrmann et al., 2006). *Leucascandra caveolata* according to our analyses forms a clade with *Sycon* cf. *carteri* (Sycettidae). This result is quite unexpected, as several key features are different in both species (Fig. 3.8): *Leucascandra*'s skeleton consists of a cortex of triradiates, the inarticulated choanoskeleton of subatrial triactines that supports an irregular alveolar leuconoid choanosome (Fig. 3.8 B, D). In contrast, *Sycon* cf. *carteri* is showing a typical organization of its genus, i.e. lacking a cortex, the choanoskeleton forms very short radial tubes, each containing a choanocyte chamber of the syconoid aquiferous system (Fig. 3.8, A, C). However, both species also share some characteristics. *Sycon* (cf.) *carteri* is built from tubes 'united in a copiously branching, bushy mass' (Dendy, 1893 ,p. 79, see also Appendix 3, Fig A3.1.1,A). Likewise, specimens of

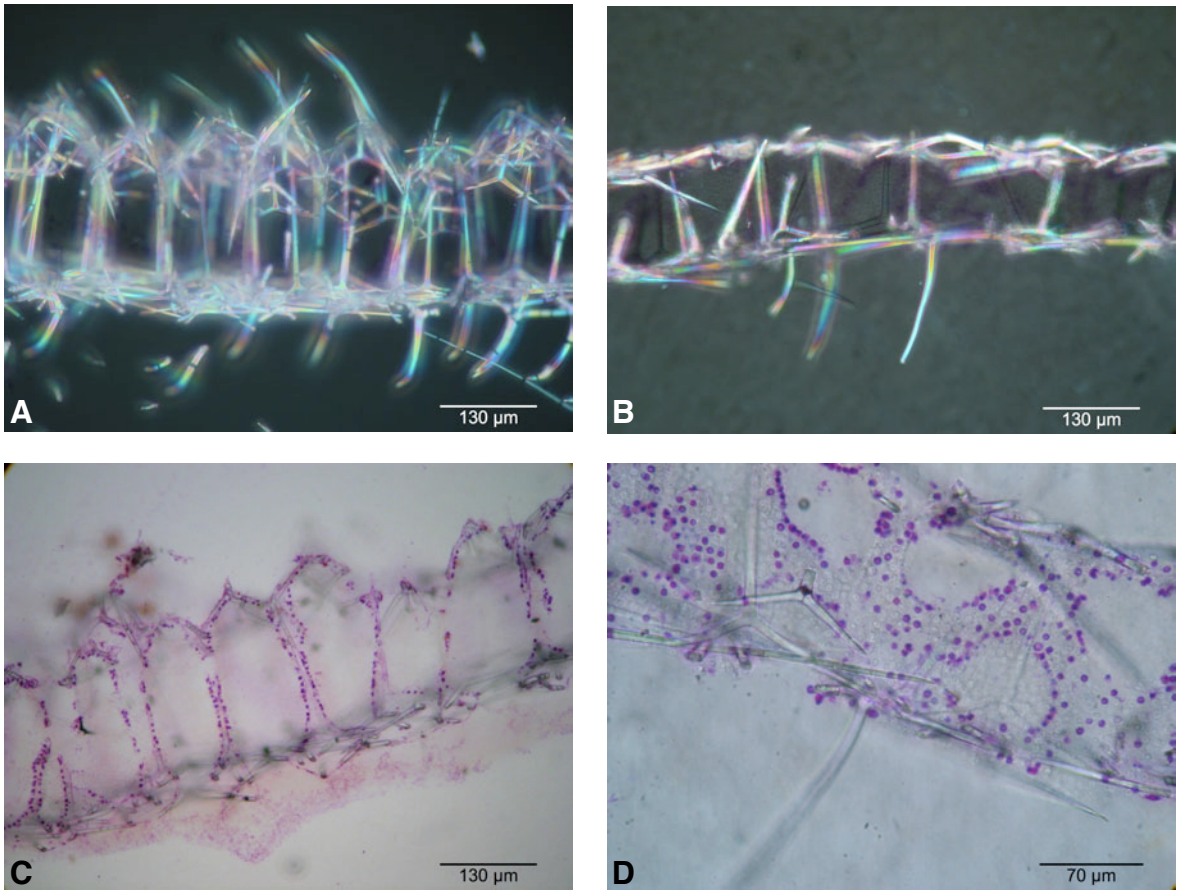


Figure 3.8: Comparison *Sycon* cf. *carteri* (**A,C**) and *Leucascandra caveolata* (QM G316146) (**B,D**). **A, B:** Skeletal arrangement; the atrial skeleton at the lower side, the distal cones or the cortical skeleton respectively on top, **C,D**: syconoid and alveolar leuconoid aquiferous system.

Leucascandra caveolata are formed by 'copiously branched and anastomosed tubes' (Borojevic and Klautau, 2000, p.199). Additionally, the spiculation of both species is quite similar aside from spicule size and the occurrence of diactines in the distal cones of *S. carteri*. Both atrial skeletons include tri- and tetractines. Subatrial triactines have longer unpaired actines that reach the lower end of the distal cones in *S. carteri* or the cortical skeleton in *L. caveolata*. The triactines in *S. carteri* are limited to the upper half of the radial tubes, therefore the choanocyte chambers in their atrial half are only supported by the unpaired actine of the subatrial triactine, giving the choanoskeleton an almost inarticulate appearance (Fig. 3.8, A). The triactines from the upper half of the radial tubes resemble the cortical skeleton of *L. caveolata.* Additionally, *Leucascandra caveolata* is named for its 'honeycomb appearance of the internal surface of the sponge' (Borojevic and Klautau, 2000). Such a regular arrangement is typical for atrial skeletons of syconoid calcareous sponges with radial tubes, and may represent a link between the different organizations of the aquiferous system in both species, with the leuconoid state in *Leucascandra* originating from a syconoid sponge with short radial tubes like observed in *S. carteri*. Taking these observations into account, a hypothetical evolution from a *Sycon*-like organization as in *S. carteri* to an inarticulate, leuconoid organization seems imaginable by the flattening of the distal cones so that the triactines form a cortical layer.

Because at least some Jenkinidae and Amphoriscidae might have taxonomic affinities (i.e. *Anamixilla torresi* and *Leucilla* sp., Fig. 3.7, D), the inarticulated choanoskeleton in these closely related species could be a homologous feature. However, the inarticulated choanoskeleton also appears in different, not closely related clades in the genera *Sycettusa*, *Syconessa*, and *Leucascandra* (Fig. 3.7, D), thus presenting convergent or pleisiomorphic states. This is conceivable, because in species with an articulated skeleton, the thin walls of young individuals show an inarticulate organization (Borojevic et al., 2000). In this light, a multiple loss of an articulated choanoskeleton as implied by our phylogeny (Fig. 3.7, D) could be a plausible evolutionary scenario.

At this stage, the evolution of body plans in Calcaronea is more confusing than in Calcinea. After the divergence of the lineages that lead to *Plectroninia* and *Leucosolenia*, respectively, the ancestor of the remaining sampled Calcaronea apparently already showed a cortex and an articulated choanoskeleton (Fig. 3.7, B & D). According to the mapping of morphological characters to our tree topology the cortex was lost several times (in *Syconessa*, questionable in *Grantessa*, and in the polyphyletic genera *Sycettusa* and *Sycon)*. The evolution of the aquiferous system appears confusing, sponges with leuconoid or sylleibid aquiferous systems apparently evolved several times independently from syconoid ancestors. However, with the available data it is impossible to determine when or how often the syconoid aquiferous system evolved. Through the comparison of *Sycon* cf. *carteri* and *Leucascandra caveolata*, the transition of a syconoid bodyplan with radial tubes to the inarticulated, leuconoid sponge with a cortex is conceivable. However, the loss of a cortex and return to an organization of coalescent radial tubes from sylleibid or leuconoid taxa is hard to imagine. Nonetheless, such transitions may prove to be possible, when the genetic toolkit that is responsible for the formation of axes, patterning and spicules during development will be understood.

Conclusion

In our phylogenetic analyses we have contributed to the understanding of the evolution of Calcarea. Our results confirm previous findings showing that the taxonomy of Calcarea is highly artificial and is in need of thorough revision. We furthermore discovered formerly unexpected relationships and evolutionary pathways in Calcinea and Calcaronea. The fact that most orders, families and several genera are paraphyletic assemblages suggests that classical revisions of such taxa (e.g. of the polyphyletic genus *Sycettusa*) will almost certainly exclude 'unexpected relatives' and therefore will remain artificial and will not result in systematics that reflects the relationships of Calcarea. A basic framework is needed to understand the evolution of characters in this sponge class. DNA phylogenies can provide such a framework and are a source of new, formerly unexpected evolutionary hypotheses. However, a considerable larger taxon sampling is necessary to adequately represent the diversity of Calcarea.

Authors' contributions

O.V. designed the study, determined specimens and generated and analyzed sequence data and wrote the manuscript. E.W. contributed by sequencing complete LSU for specimen that were included in the study from Dohrmann et al. (2006). G.W. conceived the study, provided intellectual input, and critically reviewed the manuscript.

Acknowledgements

We thank John Hooper and Monika Schlacher-Hoenlinger from the Queensland Museum in Brisbane, Jane Fromont from the Western Australian Museum in Perth, Thierry Laperousaz and Shirley Sorokin from the South Australian Museum in Adelaide and Rob van Soest from the Zoologisch Museum, Universiteit van Amsterdam, Amsterdam, Netherlands, for letting us subsample specimens of Calcarea included in this study. We thank Catherine Vogler for many helpful comments on the manuscript and language corrections. This study benefitted from funding by the German Science Foundation (DFG, Project Wo896/3). We would like to thank the Great Barrier Reef Marine Park Authority for permitting the fieldwork (Permit No.: G06/16547.1).

Supplementary information

Appendix: 3.1: Specimen descriptions.

Suppl. Table 3.1: LSU primer sequences

Suppl. Table 3.2: Bayes factors from model comparisons

Suppl. Fig 3.2: Strict consensus tree from the 17 analyses with doublet models

Suppl. Fig. 3.3.: Doublet composition of each sequence

Chapter 4

Mitochondrial diversity of early branching Metazoa is revealed by the complete mt genome of a haplosclerid demosponge

D. Erpenbeck[1,2,3], O. Voigt[3], M. Adamski[1], M. Adamska[1], J. N. A. Hooper[2], G. Wörheide[3] and B. M. Degnan[1*]

[1] School of Integrative Biology, The University of Queensland, St Lucia, Queensland, Australia

[2] Biodiversity Program, Queensland Museum, South Brisbane, Queensland, Australia

[3] Dept. of Geobiology, Geoscience Centre Göttingen, Göttingen, Germany

(Research was conducted at all three institutions)

* corresponding author. Address: School of Integrative Biology, The University of Queensland, QLD 4072 St Lucia, Australia. email: b.degnan@uq.edu.au

This version of the article was published in Molecular Biology and Evolution Biology, [24(1):19–22. 2007, (doi:10.1093/molbev/msl154)] and as Advance Access on October 19, 2006.

Abstract

The first mitochondrial genomes of demosponges have recently been sequenced and appear to be markedly different from published eumetazoan mitochondrial genomes. Here we show that the mitochondrial genome of the haplosclerid demosponge *Amphimedon queenslandica* has features that it shares with both demosponges and eumetazoans. While the *Amphimedon* mitochondrial genome has typical demosponge features, including size, long non-coding regions and bacterial-like rRNA genes, it lacks *atp9*, which is found in the other demosponges sequenced to date. We found strong evidence of a recent transposon mediated transfer of this gene to the nuclear genome. In addition *A. queenslandica* bears an incomplete tRNA set, unusual amino amino acid deletion patterns, and a putative control region. Furthermore, the mitochondrial rRNA genes of *Amphimedon queenslandica* evolve at significantly higher rates than observed in other demosponges, likewise to previously observed rates among the nuclear rRNA genes in other haplosclerid demosponges.

The acceptance of relative uniformity of metazoan mitochondrial (mt) genomes (Lang et al., 1999) has significantly been weakened by the recently published mitochondrial genomes of Porifera (*Geodia neptuni*, *Tethya actinia* and *Axinella corrugata*, all class Demospongiae, Lavrov and Lang, 2005; Lavrov et al., 2005) and Placozoa (*Trichoplax adhaerens*, Dellaporta et al., 2006). The mitochondrial genomes of these early branching animals exceed the typical length of metazoan mitochondrial genomes, which is approximately 16 kb, they possess long non-coding stretches of

DNA, have no identifiable control region, and bear additional open reading frames (ORFs) - *atp9* in demosponges and 5 putative ORFs in *Trichoplax*.

The three sponge mt genomes sequenced by Lavrov and co-workers (Lavrov and Lang, 2005; Lavrov et al., 2005) are from different orders (Table 4.1) and show relatively uniform features. All possess the metazoan standard set of 13 protein genes plus *atp9*, two rRNA genes, and 24-25 tRNA genes for a complete set of amino acids (see Table 4.1 for differences). The arrangement of the protein and rRNA genes is identical (with the only exception of *nad6* in *T. actinia*).

Despite this congruence, it remains unknown if these features are shared between all poriferans. Here, we show that mitochondrial genome evolution of basal diverging Metazoa is far more complex than previously appreciated and in higher (Eu)metazoa, by presenting the complete sequence of the demosponge *Amphimedon queenslandica*[4] (Hooper and van Soest, 2006), which is the target species for the Sponge Genome Project (http://www.jgi.doe.gov/sequencing/why/CSP2005/reniera.htm). *A. queenslandica* is a species of the order Haplosclerida, which is a pivotal order and probably the most successful demosponge taxon with the highest biodiversity in terms of species and habitat (van Soest and Hooper, 2002). Therefore *A. queenslandica* appears an ideal object for studies on demosponge molecular evolution. (See Appendix 4, Table A4.1 for Methods).

Table 4.1:Comparison of the Demosponge mt Genome Features

	Geodia neptuni	***Tethya actinia***	***Axinella corrugata***	***Amphimedon queenslandica***
	(Astrophorida)	(Hadromerida)	(Halichondrida)	(Haplosclerida)
	AY320032	AY320033	AY791693	DQ915601
size	18020bp	19564bp	25610bp	19960bp
gene overlaps	4	4	0	2
noncoding:	21%	76%	231%	121%
- longest seq	59 bp	319 bp	931 bp	1044 bp
- repeats	-	-	-	6x12 bp
protein genes:	14	14	14	13
tRNA genes:	24	24	25	17
- doub. /ident.[a]	I, L, R, S / -	I, L, R, S / M	I, L, R, S / -	R, S / M
- missing[b]	-	-	-	D, H, I, L, T, V
rRNA cluster:	rns-trnG-trnV-rnl	rns-trnG-trnV-rnl	rns-trnG-trnV-rnl	rns-trnG-trnF-rnl
- position[c]	nad5, cox2	nad5, cox2	nad5,(nad6)-cox2	nad3, nad4l
- *rns* term. BL[d]	67	86	46	193
- *rnl* term. BL[d]	69	79	62	168

[a]Refers to double amino acid tRNA genes with / without identical anticodon

[b]Refers to missing amino acid tRNA genes (referring to the anticodon)

[c]Indicates the protein genes flanking the rRNA-cluster

[d]Refers to the terminal branch lengths (BL) as calculated by p-distances from a conservative alignment

[4] Appears in databases and earlier reports under its working-name "*Reniera* sp."

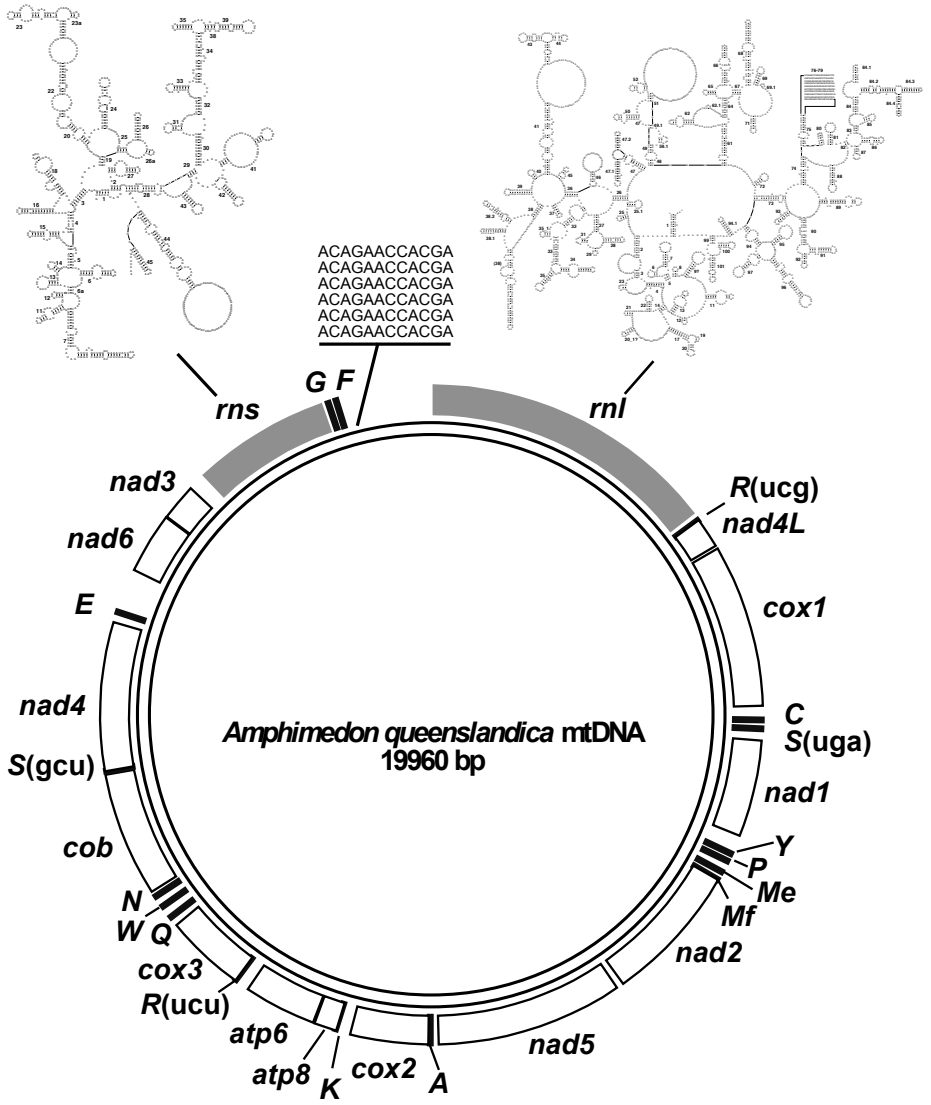


Figure. 4.1: Genomic map of the *A. queenslandica* mitochondrium. All protein genes have the same transcriptional orientation and no introns. Start codons are ATG with the exception of *cox3* (GTG) and *nad6* (TTG). All codons are present in the mitochondrial ORFs, although some are rare; e.g., CGT (1x), CGC (2x), CTC (3x). Above the rRNA genes, schematic drawings of the predicted secondary structures are given (see also Appendix 4, Figs. A4.1 & A4.2).

The *Amphimedon queenslandica* mitochondrial genome (Fig. 4.1) is a circular molecule of 19960 bp, which is a typical length for demosponges (Table 4.1). There are only two gene pairs that overlap and most genes are irregularly interspersed by noncoding regions, which comprise about 12% of the genome (Fig. 4.1). The longest noncoding region is 1044 bp and is the longest yet found in Porifera. It possesses the first mitochondrial repeat sequence detected in sponge mitochondria, which, in combination with its position at the mt-rRNA cluster, resembles higher metazoan control region features. Therefore, it provides strong first evidence for the presence of a mitochondrial control region in poriferan mitochondria. No other genes have been confirmed existing amongst the ORF of intragenic spacers as has been observed in anthozoan cnidarians (e.g., mutS; Pont-Kingdon et al., 1995).

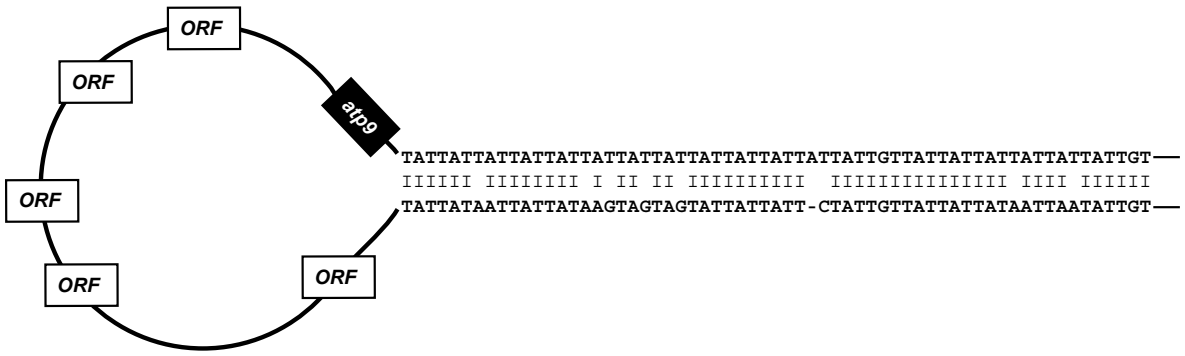


Figure 4.2: Alignment of the inverted terminal repeats (ITR). For clarity reasons is the upper sequence displayed as complement.

A. queenslandica possesses 32 mitochondrial genes, which is the smallest gene number observed in a demosponge mitochondrial genome to date. The *A. queenslandica* mtDNA codes for 13 proteins and lacks the *atp9* gene, as observed in many eumetazoan mitochondrial genomes; *atp9* is present in other demosponge mitochondrial genomes. *Cob* and *trnS* genes, which flank *atp9* in other demosponges, overlap in *A. queenslandica* and we found no evidence for an *atp9* pseudogene, which indicaties an excision event of *atp9* in the *A. queenslandica* lineage. We located *atp9* in the *A. queenslandica* nuclear genome: It encodes a protein that is five amino acids shorter than the mitochondrial counterparts and is translatable into the same functional protein with both the universal and the poriferan mitochondrial code. Interestingly, the nuclear *atp9* has flanking inverted terminal repeats (ITR) regions of 63 bp of which 87% are matching (see Fig. 4.2) providing evidence for the transposon-mediated transport of *atp9* from the mitochondrion to the nucleus. The high similarity of the ITR regions suggest a rather recent transposition event. The ITRs are located 72 bp and 2.5 kb respectively from *atp9* and enclose ORFs with yet unidentified function. Nevertheless, the fixation status of the *atp9* transposition in the genus has to be observed in closely related species (there are > 40 *Amphimedon* species *cf.* World Porifera Database; www.vliz.be/vmdcdata/porifera/). The transposition of *atp9* to a different locus prevents its usage for demosponge mitochondrial phylogenetics (see also Delsuc et al., 2005, on phylogenomics).

Other peculiar protein features of *A. queenslandica* comprise deletions in *cox1, nad1,2,5,6* and *atp6,* which are not observed in other early branching taxa (including the other demosponges).

A. queenslandica is the most early branching Metazoa found without a complete set of tRNAs in its mt genome. Six amino acids are not coded by tRNAs (D, H, I, L, T, V, Table 4. 1), a feature that is only known from Cnidaria but has not previously been detected in other diploblast mitochondrial genomes. Remarkable is the lack of *trnL*, which is usually present in two copies in metazoan mt genomes, and $trnI_{CAU}$, which is an ancestral feature lost in Eumetazoa (Lavrov and Lang, 2005). An *A. queenslandica* $tRNA^{Ile}$ is unlikely to be formed out of posttranscriptional edition of one of the two *trnM*s as both are clearly functionally distinguished into initiator (*trnMf*) and elongator (rnMe, Drabkin et al., 1998). The *trnM* tandem arrangement is a peculiar feature and, considering their high sequence difference (51%), unlikely the result of a single tandem duplication.

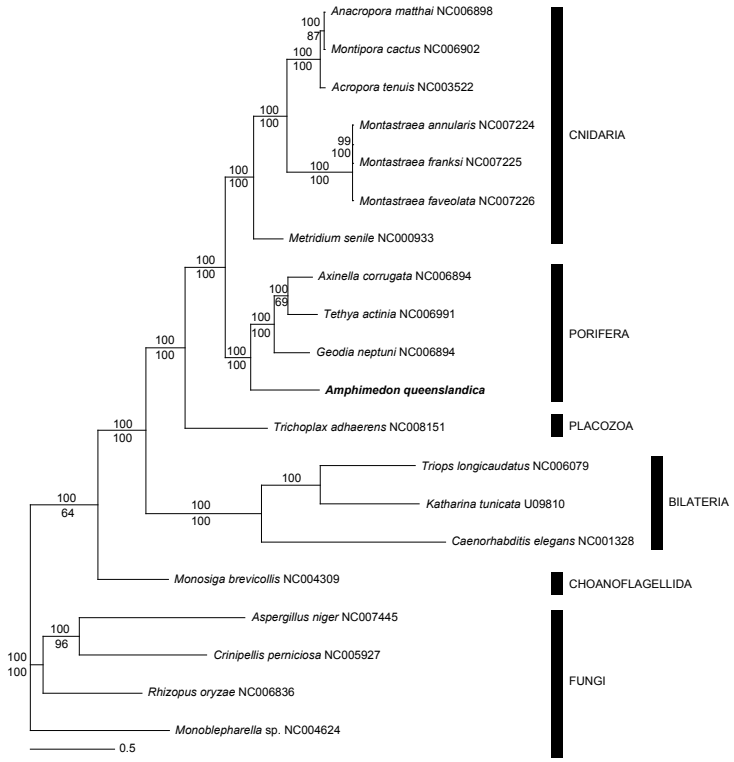


Figure 4.3:
Phylogenetic reconstruction (BI) from the mitochondrial protein sequences of representative taxa. Species names are followed by the Genbank accession. Numbers above the branches are Bayesian Inference posterior probabilities calculated under the Cprev+G+I+F model. Numbers below the branches are Maximum Likelihood bootstrap probabilities (100 repeats).

The rRNA cluster is uniquely translocated among the demosponges within the genome (Table 4.1) and diverges from the 'demosponge + choanoflagellate + "many bilaterians"- motif' *rns-(trnG/trnV/trnG-trnV)-rnl* (Lavrov et al., 2005). Its rRNA secondary structures are bacteria-like as in all demosponges, but possess several unique structural features, including extra and missing helices (secondary structures are presented in Appendix 4, Figs. A4.1, A4.2). The *nuclear* rRNA of Haplosclerida has previously been shown to evolve in different patterns and significantly higher rates than in other demosponge orders (Erpenbeck et al., 2004). Interestingly, similar tests in our present study reveal higher evolutionary rates likewise for the mitochondrial rRNA genes and the mt-protein coding genes ($p<0.05$, Table 4.1 and Fig. 4.3) of *A. queenslandica* opposed to the non-haplosclerid sponges. These higher rates in both Haplosclerida genomes may cause their higher adaptation potential and biological success.

The availability of mitochondrial DNA in combination with nuclear genomic traces, which revealed transposon mediated export of genetic material from the mitochondrium in this study, makes *A. queenslandica* an ideal candidate for mitochondrial evolution studies. The mitochondrial genome of *A. queenslandica* shares features with both known poriferan and eumetazoan (excluding *T. adhaerens*) genomes, consisting of ancestral and derived features. However, it is clear that this demosponge is not a 'missing link' between both groups (see molecular analyses in Fig. 4.3 and morphology, Hooper and van Soest, 2006). Instead, it appears that *A. queenslandica* displays convergent features of lower metazoan mitochondrial evolution, which may be a reflection of its higher rates of evolution that have been likewise observed in other nuclear genomes of this order (Erpenbeck et al., 2004). Consequently, elevated evolutionary rates in combination with given genetic precursors result in the evolution of the *A. queenslandica* mitochondrial genome convergent

to its eumetazoan counterparts. It shows how gain and loss of genes, partial genes and (control) regions occurred multiple times, presumably even throughout smaller lineages in lower Metazoa, and that one can expect far more divergence at the root of the Metazoa before obtaining a realistic picture of metazoan mitochondrial evolution.

Acknowledgements:

We thank two anonymous reviewers for their valuable comments and acknowledge financial support from the European Union (Marie-Curie outgoing fellowship MOIF-CT-2004) to D.E., the UQ Postdoctoral Fellowship Scheme and ARC Discovery Grant to M.A., the German Research Foundation (DFG; Projects Wo896/3,5,6) and the European Union through the Marie-Curie project HOTSPOTS (MEST-CT-2005-020561) to G.W., and grants from the US Department of Energy Joint Genome Institute through the Community Sequencing Program and the Australian Research Council to B. M. D

Supplementary Material

Appendix 4:

Table A4.1: Methods.
Figure A4.1: Mitochondrial (mt) 12 S rRNA secondary structures of *Amphimedon. queenslandica*.
Figure A4.2: Mitochondrial (mt) 16 S rRNA secondary structures of *Amphimedon queenslandica*

Chapter 5

A fragmented metazoan organellar genome: the two mitochondrial chromosomes of *Hydra magnipapillata*

Oliver Voigt, Dirk Erpenbeck, Gert Wörheide[§]

Adress: Courant Research Center Geobiology, Georg-August-Universität Göttingen, Goldschmidtstr. 3, 37077 Göttingen, Germany
Email addresses: Oliver Voigt: ovoigt@gwdg.de; Dirk Erpenbeck: derpenb@gwdg.de; Gert Wörheide§: gert.woerheide@geo.uni-goettingen.de

[§]Corresponding author

This version of the article was published in BMC Genomics on 26 July 2008 (*BMC Genomics* 2008, **9**:350 doi:10.1186/1471-2164-9-350) and is available from: http://www.biomedcentral.com/1471-2164/9/350

Abstract

Background

Animal mitochondrial (mt) genomes are characteristically circular molecules of ~16-20 kb. Medusozoa (Cnidaria excluding Anthozoa) are exceptional in that their mt genomes are linear and sometimes subdivided into two to presumably four different molecules. In the genus *Hydra*, the mt genome comprises one or two mt chromosomes. Here, we present the whole mt genome sequence from the hydrozoan *Hydra magnipapillata*, comprising the first sequence of a fragmented metazoan mt genome encoded on two linear mt chromosomes (mt1 and mt2).

Results

The *H. magnipapillata* mt chromosomes contain the typical metazoan set of 13 genes for respiratory proteins, the two rRNA genes and two tRNA genes. All genes are unidirectionally oriented on mt1 and mt2, and several genes overlap. The gene arrangement suggests that the two mt chromosomes originated from one linear molecule that separated between *nd5* and *rns*. Strong correlations between the AT content of rRNA genes (*rns* and *rnl*) and the AT content of protein-coding genes among 24 cnidarian genomes imply that base composition is mainly determined by mt genome-wide constraints. We show that identical inverted terminal repeats (ITR) occur on both chromosomes; these ITR contain a partial copy or part of the 3' end of *cox1* (54 bp). Additionally, both mt chromosomes possess identical oriented sequences (IOS) at the 5' and 3' ends (5' and 3' IOS) adjacent to the ITR. The 5' IOS contains *trnM* and non-coding sequences (119 bp), whereas the 3' IOS comprises a larger part (mt2) with a larger partial copy of *cox1* (243 bp).

Conclusions

ITR are also documented in the two other available medusozoan mt genomes (*Aurelia aurita* and *Hydra oligactis*). In *H. magnipapillata*, the arrangement of ITR and 5' IOS and 3' IOS suggest that these regions are crucial for mt DNA replication and/or transcription initiation. An analogous organization occurs in a highly fragmented ichthyosporean mt genome. With our data, we can reject a model of mt replication that has previously been proposed for *Hydra*. This raises new questions regarding replication mechanisms probably employed by all medusozoans, and also has general implications for the expected organization of fragmented linear mt chromosomes of other taxa.

Background

Mitochondria were most likely acquired by the common ancestor of eukaryotes (Lang et al., 1999; Gray et al., 1999; Burger et al., 2003b). Presumably, these organelles originated from incorporated α-proteobacteria and still carry their own, reduced genome (Gray et al., 1999). Mitochondrial (mt) genomes show very diverse organizations and are of a very broad range of sizes (Burger et al., 2003b; Nosek and Tomaska, 2003; Gray et al., 2004). In comparison to many protists and plants, metazoans possess an even more reduced set of mt genes and fewer non-coding regions (Boore, 1999). Typical metazoan mt genomes are circular DNA molecules of 16-20 kb (Boore, 1999; Lavrov, 2007). Remarkable exceptions are the linear mt genomes of medusozoan cnidarians (Cnidaria excluding Anthozoa). Linear mt genomes have not been found in other metazoan taxa, but from various other eukaryotes (e.g., Pritchard and Cummings, 1981; Dinouel et al., 1993; Vahrenholz et al., 1993; Fan and Lee, 2002; Burger et al., 2003a). The linear structure of cnidarian mt genomes is known from the work of Warrior (1987), who separated the DNA of isolated mitochondria via electrophoresis, as well as from Bridge et al. (1992) and Ender and Schierwater (2003), who applied *rnl*-probes to electrophoretically separated DNA extracts. Most of the medusozoan mt genomes from these studies were encoded on one ~16 kb molecule, which has been verified by the first two sequences of such linear metazoan genomes (*Aurelia aurita*, Scyphozoa: Shao et al., 2006, and *Hydra oligactis,* Hydrozoa: Kayal and Lavrov, 2008). However, in some *Hydra* species, and apparently in some Cubozoa, the mt DNA is divided onto at least two different molecules (Warrior, 1987; Bridge et al., 1992; Pont-Kingdon et al., 2000; Ender and Schierwater, 2003). In the genus *Hydra*, mt genomes are organized on one ~16 kb molecule or two ~8 kb molecules (Warrior, 1987; Bridge et al., 1992), making this genus an excellent candidate in which to examine changes due to fragmentation of mt genomes from one to two chromosomes. The *Hydra oligactis* mt genome, a 16.3 kb linear DNA molecule, was published recently (Kayal and Lavrov, 2008). Pont-Kingdon et al. (2000) sequenced a terminal section (3,232 bp) of one of the two mt chromosomes from *Hydra vulgaris* (as *Hydra attenuata*), and previous hybridization experiments have shown that in this species all four termini possess a 150-200 bp identical sequence with unknown orientation to one another (Warrior, 1987). By providing the complete mt genome sequence of *Hydra magnipapillata*, encoded on two mt chromosomes, we now present in detail the organization of such a fragmented linear mt genome from early diverging Metazoa.

Methods

We assembled the two mt chromosomes by using publicly available sequences from the *Hydra magnipapillata* whole-genome shotgun sequencing project by conducting BLAST searches (Altschul et al., 1997) of several mt protein-coding genes against the traces of *H. magnipapillata* (available via GenBank http://www.ncbi.nlm.nih.gov/). Hits were used to initiate local genome assembly in a bioinformatical pipeline (applying the cap3 assembler, Huang and Madan, 1999) to obtain the two mt chromosome sequences. The chromosomes will be referred to as mt1 (containing the *rnl* gene; available at [EMBL: BN001179]) and mt2 (containing the *rns* gene; available at [EMBL: BN001180]). mt1 and mt2 of *Hydra* share almost identical but inverted sequences of 191-196 bp at each 5' and 3' end (inverted terminal repeats (ITR), Figs. 5.1A, 5.1B, see below). Warrior (1987) previously reported that about 200 bp were identical (with unknown orientation to one another) at the ends of the mt chromosomes of *Hydra vulgaris*, and the 5' end of one chromosome (corresponding to mt1) of this species was sequenced by Pont-Kingdon et al. (2000). By comparing our sequences with the experimentally verified 5' end of mt1 from *H. vulgaris* (Pont-Kingdon et al., 2000), we were able to infer the ends of the *H. magnipapillata* mt chromosomes and found that these predicted ends coincide with an abrupt decrease in coverage in our assemblies (Appendix 5.1). This in combination with the reported sizes of one (*H. magnipapillata;* Pont-Kingdon et al., 2000) or both (*H. magnipapillata* and *H. vulgaris*; Warrior, 1987) mt chromosomes from the same or a closely related *Hydra* species suggests that the excess sequences are artifacts, and consequently they were omitted.

To exclude the possibility that other observations in our assemblies originated from methodological artifacts, we conducted additional experimental procedures and tests as follows:

1. PCR experiments

PCR experiments were done with a closely related *Hydra* species. We obtained specimens of *Hydra* sp. from the Schulbiologie-Zentrum Hannover. DNA from one polyp was prepared with the Chelex method (protocol as described in Voigt et al., 2004), 1 µl of the undiluted supernatant or 1 µl of a 1:10 dilution was used as template. Phylogenetic analysis with partial *cox1* data verified that our *Hydra* sp. specimen is very closely related to *Hydra magnipapillata* (see Results).

Primers (Appendix 5, Table A5.1) were designed to confirm our bioinformatically derived observations via PCR. The fragments shown in Fig. 5.1 were amplified and sequenced (some in two overlapping parts, see Appendix 5, Table A5.1, for details). Sequences have been submitted to GenBank [GenBank: EU683621-EU683624].

2. Additional local genome assembly experiments

To exclude the possibility that we had amplified a nuclear mt pseudogene (NUMT) of the *nd5* and partial *cox1* fragment, we started different assemblies with the pipeline originating from blast hits of a 200 bp fragment (100 bp both down- and up-stream of the connection of *nd5* and *cox1* in the assembly), as well as two assemblies starting from the last 100 bp 3' of *nd5*, and the first 100 bp of

the incomplete copy of *cox1*. In the first two cases, only one assembly was received, each time consistent with our former assemblies. By starting with 100 bp of the partial *cox1*, we recovered different assemblies, which were compatible with one chromosomal assembly or the other. In no case did we observe any inconsistencies or any presence of nuclear genes, which would have indicated that the *nd5-cox1* arrangement is part of a NUMT.

Phylogenetic analyses

The *H. magnipapillata* and *Hydra*. sp *cox1* sequences from our assembly and sequencing were manually aligned with additional sequences from other *Hydra* and outgroup species available from GenBank in the SeaView editor (Galtier et al., 1996). The final dataset contained 560 characters. A maximum likelihood analysis was carried out in PAUP* 4.0b10 (Swofford, 2003) under a model of nucleotide evolution suggested by the hierarchical likelihood ratio test in Modeltest 3.7 (Posada and Crandall, 1998); for the bootstrap analysis (1,000 replicates) we applied the same model. The dataset was also analyzed with MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) (six substitution rates with a proportion of invariant sites, two runs with eight chains each for 2 million generations with a sample frequency of 100 generations, and a burn-in of 50,000 generations). Parameter stabilization of the chains in MrBayes was monitored with Tracer 1.4 (http://tree.bio.ed.ac.uk/software/tracer/), and convergence of chains was examined with the diagnostics provided by the AWTY server (Nylander et al., 2008).

Base compositions of cnidarian mt genomes

Sequences of 23 additional cnidarian genomes were downloaded from GenBank (see Appendix 5, Table 5.3 for taxa and accession numbers). The sequences of the 13 respiratory chain protein genes shared between all 24 mt genomes and the sequences of *rns* and *rnl* were extracted from the GenBank format using the Artemis software v.9 (Rutherford et al., 2000). In some mt genomes the rRNA genes were not entirely annotated in their full length. We therefore considered the non-coding regions around the apparently too small genes as rRNAs. The corresponding taxa and positions in each sequence were: *Nematostella* sp. [GenBank: NC_008164] (*rns*: 5054..6171; *rnl*: 10342..12484); *Mussa angulosa* [GenBank: NC_008163] (*rns*: 6901..8038; *rnl*: 15327..17170); *Astrangia* sp. [GenBank: NC_008161] (*rns*: 6899..7797; *rnl*: 12982..14681). The AT contents of rRNAs and protein codon positions of the mt genomes are shown in Appendix 5, Table 5.3.

Results

Genes, base composition and codon usage

The two chromosomes of the *H. magnipapillata* mt genome each carry one rRNA gene (mt1: *rnl*; mt2: *rns*). Each of the assembled contigs of mt1 and mt2 is represented by > 7,000 single sequence reads from the trace archive (http://www.biomedcentral.com/content/supplementary/1471-2164-9-350-s1.txt). Our consensus sequence for mt1 is 8,194 bp long. This matches the length reported by Bridge et al. (1992) for this *H. magnipapillata* mt chromosome. The sequence of mt2 is shorter (7,686 bp).

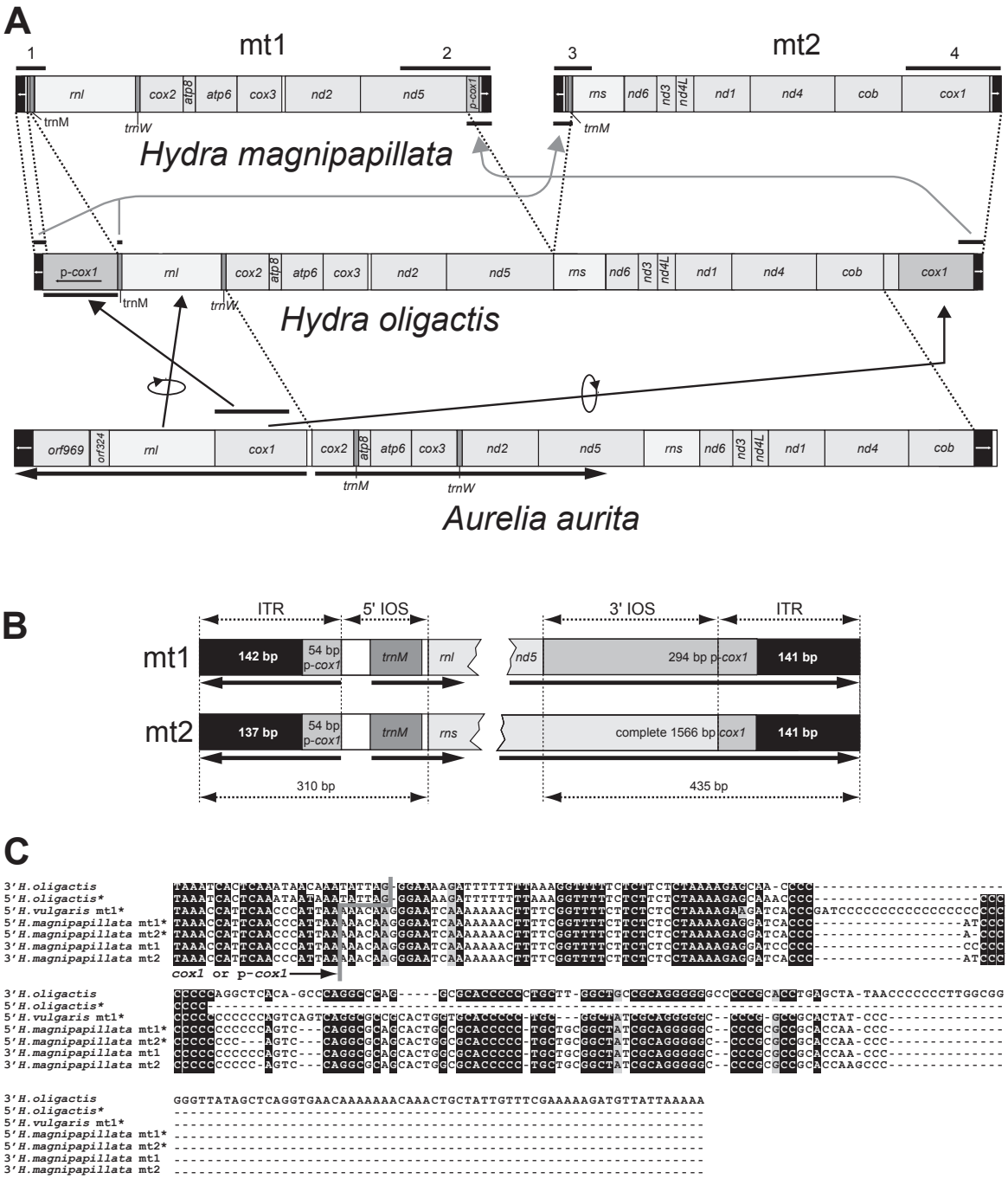


Figure 5.1: Organization of the *H. magnipapillata* mt chromosomes (mt1 and mt2). **A:** In comparison to the linear mt genome of *H. oligactis* (Hydrozoa) and *Aurelia aurita* (Scyphozoa), drawn to scale. Arrows indicate orientation of genes in *Aurelia*. Numbered black bars above *H. magnipapillata* mt chromosomes correspond to the PCR fragments amplified from *Hydra* sp. (Additional file 3). Arrows in grey indicate the proposed duplications of terminal sequences in the mt chromosome separation process. **B:** Organization at the 5' and 3' ends of mt1 and mt2 in *H. magnipapillata*. Arrows in the inverted terminal repeats (ITR) are drawn according to the orientation of the cox1 fragment. **C.** Alignment of the ends of the ITR from *H. oligactis*, *H. vulgaris* (mt1) and *H. magnipapillata* (mt1 and mt2). * = sequence displayed as reverse complement.

The *H. magnipapillata* mt genome includes 13 protein-coding genes of the respiratory chain usually found in other Metazoa. mt1 contains 6 protein-coding genes, *rnl* and two tRNA genes; mt2 contains 7 protein-coding genes, *rns* and one tRNA gene (Fig. 5.1A). All genes are unidirectionally encoded on each of the two molecules and densely arranged along the chromosomes. As in *H. oligactis*, the longest non-coding intergenic region is 52 bp between *cox3* and *nd2* (Kayal and Lavrov, 2008). Otherwise, subsequent genes are separated by 0-5 bp or overlap for up to 10 bp (in *nd6-nd3* and *nd1-nd4*).

Like many other Cnidaria (Beagley et al., 1998; Medina et al., 2006; Shao et al., 2006; Brugler and France, 2007; Kayal and Lavrov, 2008), the *H. magnipapillata* mt genome possesses only the two tRNA genes for methionine (*trnM*; CAU) and tryptophan (*trnW*; UCA). *trnW* is only found on mt1, whereas identical copies of *trnM* are present on both chromosomes (Fig. 5.1B).

Six amino acid codons are not used in the 13 protein-coding genes (Table 5.1), and all genes are terminated by TAA. Apparently synonymous codons that posses an A or T, instead of a G or C, at the third codon position are preferred in *H. magnipapillata*. To test whether this observation is caused by mechanisms that affect base composition in the whole mt genome, we analyzed codon usage in the 13 respiratory protein-coding genes in 24 mt genomes of Cnidaria. We plotted the AT content at each of the three codon positions against the AT contents of the rRNA genes for every genome, as rRNA coding genes represent a different part of the mt genomes in terms of functional constraints compared to protein-coding genes. Remarkably, *H. magnipapilllata* showed the highest values for AT content at the third codon positions (89.8%) and in the rRNA genes (78.1%; Fig. 5.2, black filled symbols). Moreover, a high AT content in rRNA genes generally correlates with the usage of A and T at third codon positions in all Cnidaria (significant at p= 0.001), suggesting that codon usage might be the result of a general selection for base composition on the mt genome caused by interaction of mutational, repair, replication and translational mechanisms (Perna and Kocher, 1995). The AT content at the first and second codon positions also correlates with that of

Table 5.1: Codon frequency among the 3,987 codons of the 13 protein-coding genes in *H. magnipapillata*.

	Codon	n		Codon	n		Codon	n		Codon	n
Phe	TTT	406	**Ser**	TCT	150	**Tyr**	TAT	169	**Cys**	TGT	39
	TTC	32		TCC	15		TAC	14		TGC	2
Leu	TTA	456		TCA	95	**TER**	TAA	13	**Trp**	TGA	74
	TTG	33		TCG	2		TAG	0		TGG	3
Leu	CTT	48	**Pro**	CCT	61	**His**	CAT	72	**Arg**	CGT	6
	CTC	4		CCC	9		CAC	8		CGC	0
	CTA	55		CCA	53	**Gln**	CAA	56		CGA	0
	CTG	3		CCG	2		CAG	3		CGG	0
Ile	ATT	304	**Thr**	ACT	92	**Asn**	AAT	217	**Ser**	AGT	79
	ATC	36		ACC	11		AAC	43		AGC	16
	ATA	298		ACA	51	**Lys**	AAA	119	**Arg**	AGA	51
Met	ATG	86		ACG	0		AAG	11		AGG	0
Val	GTT	77	**Ala**	GCT	90	**Asp**	GAT	67	**Gly**	GGT	65
	GTC	7		GCC	7		GAC	14		GGC	9
	GTA	84		GCA	42	**Glu**	GAA	82		GGA	111
	GTG	6		GCG	0		GAG	3		GGG	26

the rRNA genes (significant at p= 0.001), but here AT content rise at a lower rate than the increasing AT content of the rRNAs (regression line slopes: first codon position: 0.46; second: 0.33; third: 1.18). This is likely the result of selection on certain amino acids. Cnidarians posses a lower AT content at the first codon position than at the second (Fig. 5.2), with *H. magnipapillata* and *H. oligactis* being the only exceptions (73.1% vs. 70.9% for *H. magnipapillata*, filled symbols in Fig. 5.2).

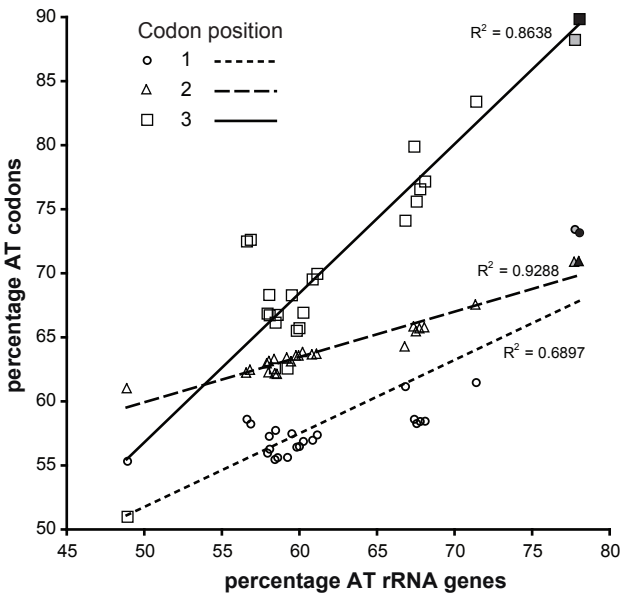


Figure 5.2: Base composition in cnidarian mt genomes. Correlations of AT content (%) of mt rRNAs and the AT content (%) in the codon positions 1, 2 and 3 calculated from 13 protein-coding genes of 23 cnidarian mt genomes (Appendix, Suppl-Table 5.x). Black filled symbols = *H. magnipapillata*; grey filled symbols = *H. oligactis*.

Gene arrangement and inverted terminal repeats

Compared to the gene arrangement of *A. aurita* and *H. oligactis*, only a few changes can be observed in *H. magnipapillata*. Neglecting the positions of tRNAs, two blocks (*cox2*, *atp8*, *atp6*, *cox3*, *nd2*, *nd5* and*, rns*; *nd6*, *nd3*, *nd4L*, *nd1*, *nd4*, *cob*) of genes are identical across the three genomes, occurring on mt1 or mt2, respectively, in *H. magnipapillata* (Fig. 5.1A). The mt genomes of *H. oligactis* and of *H. magnipapillata* are entirely alignable and display a sequence divergence of 12.3% (excluding the terminal chromosome structures; see below).

As mentioned before, we found 191-196 bp of ITR at both ends of mt1 and mt2. In the linear mt genomes of *H. oligactis* and *A. aurita*, ITR were also present but were longer (*H. oligactis*: 1,488 bp; *A. aurita*: 471 bp; Shao et al., 2006; Kayal and Lavrov, 2008) assuming symmetry for unsequenced ends). Unlike ITR in *Aurelia* (Shao et al., 2006), ITR in *H. magnipapillata* have a higher GC content than the rest of the molecule (52.2% GC in ITR vs. 25.2% GC in 5' IOS [see below], 27.6% GC in 3' IOS [see below] and a mean of 22.5% GC for all remaining regions). We found that a smaller part of 3' *cox1* (54 bp) is included in all ITR of *H. magnipapillata*. Probably because the 3' end of *cox1* is not very conserved, Pont-Kingdon et al. (2000) missed this feature in their mt1 fragment of *H. vulgaris*. The ITR regions of *H. oligactis* contain a larger *cox1* fragment (one non-functional copy at the 5' end, functional *cox1* at 3' end, Fig. 5.1A). The remaining sequenced 3' region of ITR in *H. oligactis* is very similar to those found in *H. magnipapillata* and *H. vulgaris* (Fig. 5.1C), but longer. Between *H. magnipapillata* and *H. vulgaris*, the major difference is that a stretch of Gs (31 in *H. vulgaris*) is significantly shorter in *H. magnipapillata* (11-16 at the homologous region).

In *H. magnipapillata* mt1 and mt2, we found additional identical sequences at the 5' and 3' ends following (at the 5’ ends) and preceding (at the 3’ ends) the ITR. We refer to those regions as identically oriented sequences (5' and 3' IOS, Fig. 5.1B). After the ITR, the 5' IOS of both molecules contain identical copies of non-coding DNA and *trnM*. At the 3' IOS we found a larger partial copy of the 5' region of *cox1* on mt1. As a consequence of this arrangement, mt1 and mt2 share 310 bp (ITR+5' IOS) at the 5' end and 436 bp (3' IOS+ITR) at the 3' end, giving both molecules a specific orientation.

Using PCR experiments with the closely related *Hydra* sp., we verified the following arrangements initially observed in the *H. magnipapillata* sequences (compare Fig. 5.1A): (i) the presence and orientation of the ITR at all four chromosome ends could be shown, as well as the presence of partial *cox1* sequences in the ITR; (ii) identical regions are shared at the 5' and 3' end, respectively, between mt1 and mt2 adjacent to the ITR; and (iii) within the latter regions, the 5' motif contains *trnM*, which therefore appears in two copies in the genome, and a larger sequence of *cox1* forms the shared 3' motif of mt1 and mt2.

Phylogenetic analysis

The tree topology derived from our phylogenetic analysis of *cox1* shows the close relationship of *Hydra* sp. and *H. magnipapillata* (Fig. 5.3B), thus ensuring that we used an appropriate taxon to test our results. *H. vulgaris* (Two sequences from GenBank) is paraphyletic, which reflects the difficult taxonomy of the genus (Hemmrich et al., 2007). The presented phylogeny, in combination with the mt genome organization, supports the view that the ancestral state of mt genome organization in the genus *Hydra* was a single linear mt chromosome.

Discussion

Linear mt genomes and fragmentation of mt chromosomes in Cnidaria

Linearity of mt genomes seems to have evolved once after the divergence of Medusozoa from Anthozoa. Fig. 5.3A summarizes the results of different studies (Warrior, 1987; Ender and Schierwater, 2003; Shao et al., 2006; Kayal and Lavrov, 2008), mapped on a cnidarian phylogeny (Collins et al., 2006). A fragmentation of mt genomes has been reported from several *Hydra* species (Hydrozoa) (Warrior, 1987; Bridge et al., 1992) and Cubozoa (Ender and Schierwater, 2003). Uncertainties remain for Cubozoa: Bridge et al. (1992) studied the same cubozoan species *Carybdea marsupialis* as Ender and Schierwater (2003), but reported a single ~16 kb linear mt genome, while in the more recent work, a ~4 kb fragment was shown to carry the *rnl* gene. Because Ender and Schierwater (2003) were able to repeat the experiments with different DNA isolates of *C. marsupialis* and obtained concordant results from an additional cubozoan species (*Tripedalia cystophora*), an experimental error seems unlikely. However, their conclusion of four equally-sized mt chromosomes in Cubozoa is not directly supported by their identification of a 4 kb chromosome carrying *rnl*. Alternatively, one could assume the presence of a single ~12 kb mt counterpart, as indicated in Fig. 5.3A. Such an arrangement is possible, e.g., if *rnl* and *cox1*, the two genes that are encoded

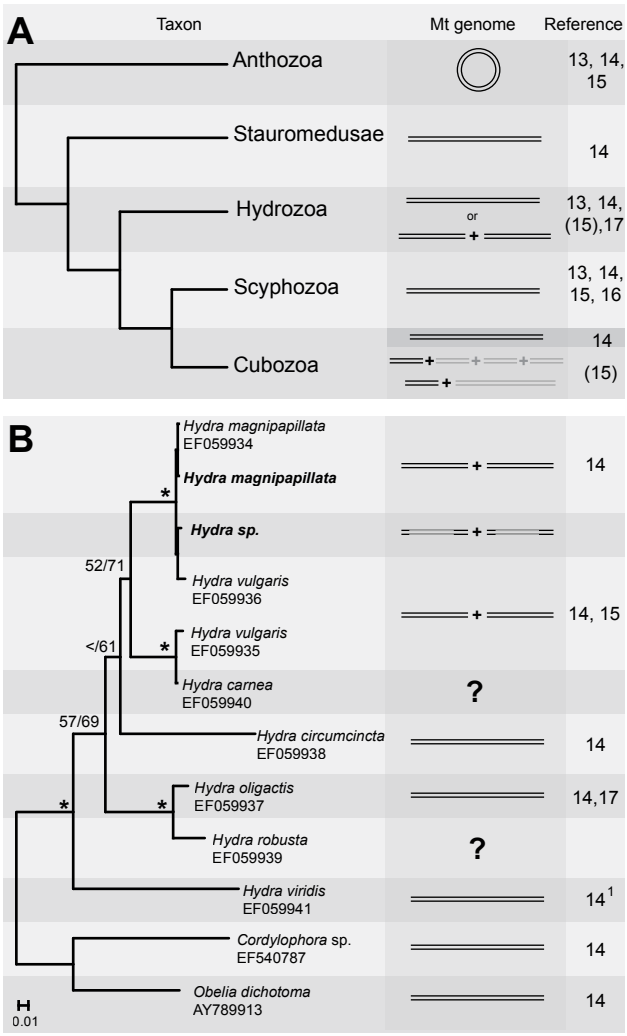


Figure 5.3: Evolution of cnidarian mt organization.
A: Summary of relationships of higher cnidarian taxa according to nuclear Small and Large Subunit rRNA data (Collins et al., 2006), and the organization of mt genomes., Where only the rnl-carrying chromosome was examined, reference are **13:** Warrior (1987); **14:** Bridge et al. (1992); 1**5:** Ender and Schierwater (2003); **16:** Shao et al. (2006); **17:** Kayal et al. (2008).
B: Summary of relationships within the genus *Hydra* based upon our ML and Bayesian analyses of partial cox1 data, rooted with other hydrozoan sequences from GenBank (accession numbers are given after each species name). Support values >50 are shown above branches (ML bootstraps/Bayesian posterior probability, * = 100 in both analyses). Sequences from this study are bold. Expected mt genome organization is shown in grey. [1]syn.: *Hydra viridis*; *H. viridissima*

in different orientation to the other mt genes in *A. aurita* (Shao et al., 2006), were encoded in one chromosome in Cubozoa, and the remaining genes on a second chromosome.

However, given the available data it seems reasonable to assume that fragmented linear genomes occur in both Cubozoa and Hydrozoa (in some members of the genus *Hydra*). This suggests from an evolutionary perspective that the mt genome in the common ancestor of Medusozoa was linear and then independently split into different chromosomes in *Hydra* (Fig. 5.3B), and in at least some Cubozoa (compare Fig. 5.3A).

A possible mechanism for the origin of linear chromosomes from a circular molecule is the integration of one or more resolution elements (Nosek and Tomaska, 2003). The circular DNA molecule would be split into one or more linear molecules with identical ends. In Medusozoa, the processes of linearization and the split of one linear into two linear chromosomes were obviously different processes as shown in the phylogenetic trees (Fig. 5.3). The linearization, possibly occurring in the last common ancestor of medusozoans, seems to have preceded the splitting of the chromosomes by a long time. If the ancestral linear mt chromosome of Medusozoa originated by introduction of a resolution element, one probably would not expect to observe its original motifs, which would occur as identical repeats at the two ends of the linear molecule (Nosek and

Tomaska, 2003). Indeed, the ends of linear medusozoan mt chromosomes have inverted terminal motifs (the ITR), instead of direct repeats. The splitting of ancestral linear mt chromosomes as in *H. magnipapillata* (and possibly Cubozoa) happened much later in evolutionary history, contradicting the view that the two or more linear mt chromosomes in Medusozoa directly originated from one circular DNA molecule.

Fragmented mt genomes are present in various eukaryotic taxa, e.g., in dinoflagellates (Slamovits et al., 2007; Jackson et al., 2007), Ichthyosporea (Burger et al., 2003a) and Fungi (Burger and Lang, 2003). In Metazoa, fragmented mt genomes are known from the genera *Globodera* (Nematoda; Armstrong et al., 2000; Gibson et al., 2007b; Gibson et al., 2007a), *Dicyema* (Mesozoa; Watanabe et al., 1999) and the rotifer *Brachionus plicatilis* (Suga et al., 2008), but unlike in *H. magnipapillata*, in these taxa the genomes are encoded on several small circular molecules. The mt chromosomal organization observed in *H. magnipapillata* supports the hypothesis of an ancestral, linear chromosome in *Hydra* (Fig. 5.3B), as represented by the mt genome of *H. oligactis* (Kayal and Lavrov, 2008), which has been split in two between *nd5* and *rns*.

Function of ITR and IOS

Warrior (1987) already suggested the presence of identical terminal sequences on both chromosomes of *H. vulgaris*. We now show that these ends are arranged as ITR on mt1 and mt2, as in other medusozoans (Shao et al., 2006; Kayal and Lavrov, 2008). In *H. oligactis*, which in the phylogenetic tree branches off before *Hydra* species carrying two mt DNA molecules (Fig. 5.3B), the single linear mt chromosome has ITR containing a large copy of the 5' end of *cox1*. Only the ITR at the 3’ end has been completely sequenced (Kayal and Lavrov, 2008). Based on our findings in *H. magnipapillata,* we predict that the unsequenced 5' end is almost identical to the 3' motif (Fig. 5.1), and we expect that about 150 bp remain unsequenced on the 5' end (in contrast to the 65 bp that have been proposed Kayal and Lavrov, 2008). In *Hydra*, partial copies of *cox1* play a crucial role as part in ITR regions at the chromosome ends (Fig. 5.1, Kayal and Lavrov, 2008). The ITR in *H. magnipapillata* contains only a short sequence of the 3' end of *cox1* (54 bp, compared to the 1284 bp in *H. oligactis*), suggesting that large parts of the *cox1* copies were lost. A simultaneous duplication of 5' ITR (containing the already shortened partial *cox1* copy) and the 5' IOS motif seems likely to have occurred in the process of chromosome splitting. In this case, the longer *cox1* copy (containing additional 240 bp of *cox1*) is a duplication of the functional *cox1* of the original 5’ end of a single mt chromosome (Fig. 5.1A).

ITR of linear mt molecules are present in other taxa besides medusozoans, e.g., in yeasts (e.g., Dinouel et al., 1993) and in the green algae *Chlamydomonas reinhardtii* (Vahrenholz et al., 1993). Furthermore, in the green algae *Polytomella parva,* identical ITR are present at all ends of the two linear mt chromosomes (Fan and Lee, 2002), similar to what we observe in *H. magnipapillata*. We report 5' and 3' IOS as an additional shared feature of the two mt chromosomes. Interestingly, such an arrangement of ITR and 5' and 3' IOS is also seen in another, highly fragmented eukaryotic mt genome. In the ichthyosporean *Amoebidium parasiticum,* mt genes are distributed over several

hundred different chromosomes, each of which also possesses ITR and 5' and 3' IOS (Burger et al., 2003a).

Pont-Kingdon et al. (2000) speculated that there may be a role for transcription initiation at the 240 bp 5' of *trnM*, which they found in their *H. vulgaris* (as *H. attenuata*) partial mt1 sequence. Considering that transcription initiation within the ITR would result in energetically expensive nonsense transcripts (since all genes are encoded on only one strand), transcription is more likely to start in the adjacent, non-coding regions of the 5' IOS. In *H. magnipapillata* and *H. vulgaris* this region within the 5' IOS is 40 bp long and lies between the *cox1* copy and *trnM* (Fig. 5.1B). In *H. oligactis*, the non-coding region between the ITR and *trnM* is only 6 bp. However, a striking sequence similarity can be observed near *trnM* between *H. oligactis* and *H. vulgaris* (with the same sequence in this region as *H. magnipapillata,* Kayal and Lavrov, 2008). There is a 14-bp motif (TTATTTRRTCTTCT) that is shared between the species and differs by the last 3 bp from the 3' ITR+3bp counterpart in *H. oligactis*. This motif might be involved in transcription initiation. If so, the difference in the very last 3 bp between the 5' end and its counterpart on the reverse strand in the ITR of *H. oligactis* prevents a functional transcription signal on the non-coding strand in this species. A crucial function for transcription initiation would explain selective pressure for maintaining the 5' IOS of both molecules after the ITR in *H. magnipapillata*. All mt chromosomes from *Amoebidium parasiticum* that contain coding genes are transcribed from 5' IOS to 3' IOS (Burger et al., 2003a), as in *H. magnipapillata*. This observation led Burger et al. (2003a) to the conclusion that the IOS in *Amoebidium* are responsible for transcription initiation (5' IOS) and termination (3' IOS). While in *H. magnipapillata* we expect the same function for 5' IOS, the role of the additional partial *cox1* copy within the 3' IOS of mt1 and mt2, if any, remains unknown; considering that the end of *cox1* is part of the ITR, transcription can only be terminated in ITR and not in the 3' IOS. The sequence homologies of ITR and IOS within or between mt1 and mt2 are probably not the result of a relatively recent origin from ancestral sequences, as a first duplication of partial *cox1* is already observed in *H. oligactis* and therefore predates the separation process. The substitutions between ITR (partial *cox1*) of the two species are found in all ITR copies in each mt genome. Considering this and the fact that similar arrangements are found in other eukaryotes (Pritchard and Cummings, 1981; Dinouel et al., 1993; Vahrenholz et al., 1993; Fan and Lee, 2002; Burger et al., 2003a; Shao et al., 2006), it seems more likely that concerted evolution maintains the almost identical sequences, probably caused or influenced by the yet unknown mt genome replication mechanism. Terminal sequences of linear DNA molecules play a crucial role in mt replication (Nosek and Tomaska, 2003). The main problem in the process of linear chromosome replication is the maintenance of the 5' ends. In nuclear (nc) chromosomes this is normally achieved by telomerase, an enzyme that adds short sequence motifs in tandem repeats at the end of each molecule to compensate for loss at the 5' end that occurs in each replication cycle (Nosek et al., 2006). Consequently, in *Hydra* as in most Metazoa, the motif $(TTAGGG)_n$ is found at the end of nc chromosomes (Traut et al., 2007). The termini of *Hydra* mt chromosomes are much more complex, and their maintenance during replication is not yet understood but is most likely telomerase independent (Warrior, 1987). Warrior (1987) suggested an mt replication mechanism for the two *H.*

vulgaris mt chromosomes similar to that of T7 bacteriophages. His conclusion was based upon observations from hybridization experiments, which showed the presence of identical terminal sequences that he assumed to have the same orientation at the 5' and 3' ends. According to this model, intermediate concatamers are formed, and via ligation, site-specific nicking and elongation, the 5' ends are finally filled. Based on our data we can reject this model, because we showed that the terminal sequences in *H. magnipapillata* are inverted (ITR) and therefore do not allow the necessary concatamerization in the proposed way. Similarly, ephemeral circularization as in the phage lambda is not possible when terminal sequences are not direct repeats, but inverted.

Different replication mechanisms for linear chromosomes have been reported or proposed (for review, see Nosek et al., 1998). Solutions for maintaining the terminal structure in sequences with ITR include (a) covalently bound proteins, which also could serve as primers for a 'racket frame' replication (e.g., in linear mt chromosomes of plants and fungi or adenoviruses Nosek et al., 1998; Sakaguchi, 1990), (b) 5' and 3' ends of chromosomes that are connected by a hairpin-loop (e.g., in some yeasts; Dinouel et al., 1993; Nosek et al., 1998) and (c) single-stranded 3' overhangs (e.g., in *Chlamydomonas reinhardtii*; Vahrenholz et al., 1993). Warrior (1987) showed that the first two possibilities are not realized in *H. vulgaris*. The mt-replication mechanism in *Chlamydomonas* requires an internal repeat of the single-stranded 3' overhangs (Vahrenholz et al., 1993). We cannot entirely rule out the existence of short single-stranded 3' overhangs, as it is possible they might have been missed due to our methods. The outermost sequence at least cannot be part of a repeat motif, as our PCR amplification of *Hydra* sp. did not yield fragments of different sizes. Furthermore, neither in mt1 of *H. vulgaris* nor in the mt genome of *H. oligactis* were additional sequences or repeats found (Warrior, 1987; Kayal and Lavrov, 2008). Therefore, although a similarity to the mt replication of *C. reinhardtii* cannot be excluded, we find that mechanisms for linear mt chromosome replication are too diverse and that too many details are still unknown (Nosek et al., 1998; Nosek and Tomaska, 2003) for us to draw further conclusions about the replication process in *H. magnipapillata* from the presence of the ITR alone. Considering that ITR are a shared feature among all three available sequences of medusozoan mt genomes [*H. oligactis*, *A. aurita* (Shao et al., 2006; Kayal and Lavrov, 2008) and *H. magnipapillata*], it is very likely that the mechanisms for mt replication are similar in all medusozoans and are still the same after the fragmentation of the mt genome. Keeping in mind that similar arrangements of ITR and IOS are found in *Amoebidium parasiticum*, the mt replication mechanisms of *H. magnipapillata* and Medusozoa are probably not unique among eukaryotes.

Conclusions

The *H. magnipapillata* mt genome represents the first complete sequence of a linear metazoan mt genome that consists of two separate molecules. The gene arrangements and our phylogenetic analysis suggest that mt1 and mt2 originated from an ancestral linear mt genome, as found in *H. oligactis*, that at some point divided in two between *nd5* and *rns* (Figs. 5.1 and 5.3). Of most interest is the organization at the ends of the two mt chromosomes (Fig. 5.1). We show that *H. magnipapillata* has ITR, which include a part of *cox1* (at the 3' ITR of mt2) or partial copies of the 5' end

of *cox1* (all other ITR). We conclude that mechanisms for mt replication in *Hydra* species are different from the previously proposed ones and probably are shared among all medusozoans. In addition to ITR, both mt chromosomes of *H. magnipapillata* have identical motifs on their 5' and 3' ends, called respectively the 5' IOS and 3' IOS (Fig. 5.1). The 5' IOS includes *trnM* and a non-coding region, including a motif that may play a role in transcription initiation. The 5' IOS is probably the result of a duplication during the separation process of a single ancestral mt chromosome. The organization of the ITR and 5' and 3' IOS is not unique among eukaryotes with fragmented linear mt genomes. ITR most likely play a role in mt replication, while the duplication of the 5' end of an single ancestral linear and unidirectionally-encoded mt chromosome (with the presence of 5' IOS) and its concerted evolution ensure that transcription of all mt genes is maintained after fragmentation of linear mt chromosomes. A similar arrangement of ITR and IOS regions can therefore be expected in the apparently fragmented mt genomes of Cubozoa and other eukaryotes with two or more linear mt DNA molecules.

Authors' contributions

O.V. contributed to the conception and design of the study, acquired, analyzed and interpreted the data, and wrote the manuscript. D.E. and G.W. contributed to the interpretation of data and critically reviewed the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

This work was financially supported by the German Research Foundation (DFG, Projects Wo896/3-1, 3 & Wo896/6-1) and is a contribution from the DFG Priority Programme SPP1174 "Deep Metazoan Phylogeny". D.E. acknowledges financial support from the European Union under a Marie Curie outgoing fellowship (MOIF-CT-2004 Contract No. 2882). All authors acknowledge the J. Craig Venter Institute for providing *H.magnipapillata* WGS traces via NCBI's trace archive, and the National Human Genome Research Institute for funding the *H. magnipapillata* genome project. We thank Bernie Degnan and Ben Woodcroft (School of Integrative Biology, The University of Queensland, Brisbane) for providing access to their 'Reefedge' pipeline for local genome assembly and the Schulbiologiezentrum Hannover for providing *Hydra* specimens for our PCR experiments. O.V. thanks Catherine Vogler for help with R and many useful discussions and remarks on the manuscript.

Supplementary information

Trace IDs of sequences included in the *H. magnipapillata* assemblies can be accessed under: .
http://www.biomedcentral.com/content/supplementary/1471-2164-9-350-s1.txt

Appendix 5:

Figure A 5.1: Coverage of mt1 and mt2 assemblies.

Table A5.1: Primer sequences.

Table A5.2: Taxa, GenBank accession numbers and AT contents of cnidarian mt genomes

Summary of results and conclusions

Secondary structure of hyper-variable insertions in ribosomal RNA genes

Through the analysis of ribosomal RNA (rRNA) gene data (rDNA) in this work, several open questions concerning the phylogeny of taxonomically difficult taxa in the phylum Porifera were clarified. In marine Haplosclerida, we presented alternative support for a tree topology that contradicts the current taxonomical scheme by analyzing the secondary structure of hyper-variable insertions into the SSU rRNA gene (**chapter 1**). Thereby we demonstrated that such insertions – that are usually excluded from phylogenetic analyses due to their high variability – contain an important phylogenetic signal and therefore should not be neglected. This also appears to be the case for Hexactinellida, where the SSU rRNA contains similar insertions, and hence similar analyses could be performed. But unfortunately, in many metazoan taxa, the insertions are absent in the SSU rRNA. However, the method could probably be extended to further taxa by taking into account the LSU rRNA gene, where longer insertions with taxon-specific secondary structures occur (e.g., Schnare et al., 1996).

Doublet model in analyses of rRNA data

The results of **chapters 1-3** highlight the importance of analyzing rRNA data with specific doublet models to take the coevolution of paired nucleotides into account. Although rRNA data is frequently used in phylogenetic studies, the majority of analyses are performed under standard models of nucleotide substitution. Even when doublet models are used, the model used is often chosen more or less arbitrarily, as a thorough comparison of the performance of all models is hindered by high computational demands. However, I showed in **chapter 3** that suboptimal model choice could lead to over-estimation of support for wrong or poorly supported clades in phylogenetic inferences. Therefore, special care should be taken when choosing doublet models. My analyses enabled comparing seventeen 6-state, 7-state and 16-state models. Such an exhaustive comparison was not available to date. The results suggested that the most general 6-state doublet model 6A performs best in our analyses of SSU and LSU rRNA data. The better performance of 6A compared to analyses using the most general 7-state and 16-state models (7A and 16A) could also be verified by reanalyzing the independent dataset of **chapter 1**. Therefore, when a comparison of doublet models is not possible, applying the 6A model seems to be the most reasonable choice, when the data is similar to the SSU and LSU rDNA data presented here. For SSU rRNA of sponges, the application of such models is facilitated by the secondary structures and Perl-scripts I presented in a database as supplement for **chapter 1**. (http://www.palaeontologie.geo.lmu.de/molpal/RRNA).

Implications for the taxonomy of Calcarea

By performing phylogenetic analyses of the poriferan class Calcarea (**chapters 2 and 3**), we found tremendous discrepancies in the system of Calcarea as it was understood until today, questioning

the validity of many of the recognized orders, families and genera. The previous subdivision of Calcarea in the subclasses Calcinea and Calcaronea (Bidder, 1898; Manuel et al., 2003; Manuel et al., 2004; Manuel, 2006) could be confirmed with the presented studies. In contrast, most of the taxa below that level, where more than one representative species was included in the analyses, proved to be poly- or paraphyletic, with only the exceptions of the order Baerida (Calcaronea, but here several families were not included) and possibly the family Leucettidae (Calcinea, although one included sample without certain identification might not belong to this family).

Therefore, the taxonomy of Calcarea is even more challenging than previously appreciated, and former hypotheses on the evolution of certain morphological features in Calcarea now seem obsolete. At the same time, we can at least to some degree bring forward new hypotheses. The results of **chapters 2 and 3** imply that the system of Calcarea as a whole is in need of thorough revision, and because of the many unexpected relationships, only a combined morphological and molecular approach can succeed in unraveling the relationships within this class of sponges. But despite recent efforts that include analyses of DNA data (e.g., Valderrama et al., 2009), taxonomic revisions are often still restricted to certain taxa, (e.g., the revision of the genus *Clathrina*, Klautau and Valentine, 2003). However, with the presented results that most taxa (genera, families and orders) are not monophyletic, such studies could be misleading: taxa that might have to be included to resolve the relationships of the group of interest will be neglected, because the taxonomy does not reflect the phylogeny. Unfortunately, the evolution of morphological traits is still too little understood from our analyses to provide an alternative classification scheme supported by morphological synapomorphies. Thus, a natural system is not yet at hand. But it raises the question of how such a system could be established in the future. I propose an approach in which every considered species should be included in a molecular phylogeny, by sequencing a given DNA fragment. Such methods are used in the 'barcoding of life' initiative (http://www.dnabarcodes.org), which aims to identify species by their unique DNA-codes (Hebert et al., 2003). For Calcarea, the goal of a sequencing campaign should not primarily be the identification species, but providing a phylogenetic backbone for all further studies, especially to understand the complex evolution of morphological traits. Given a broad enough taxon sampling, such studies will most likely allow identifying previously overlooked synapomorphies (or combination of new morphological characters) for monophyletic groups that in turn can serve as diagnostic characters for a revised taxonomy of the group. But what gene should be used for such an approach? The mt cytochrome oxidase subunit I gene (*cox1*) is normally applied in barcoding approaches of diverse metazoan taxa, including Demosponges and Hexactinellida (Wörheide and Erpenbeck, 2007). From my experience, however, PCR amplification of this gene from Calcarea proved extremely difficult, and I had similar difficulties in amplifying other mitochondrial (mt) genes. The few amplicons available from two species of Leucettidae (data not shown in this work) suggest a very high variation in this gene and other mt genes in Calcarea, which is unusual in the sense that in other basal diverging metazoan taxa including demosponges the evolutionary rate is lower than in Bilateria (e.g., Lavrov et al., 2005). Due to the difficulties of obtaining the sequences and the lack of comparative calcarean *cox1* sequences in public databases, I suggest using the (according to the analysis in **chapter 3**)

most variable part of the LSU rRNA in Calcarea (C-extension helices, see Wuyts et al, 2004) to estimate affinities of specimens and to obtain a phylogenetic backbone. With the LSU rRNA sequences presented in **chapter 2 and 3** a considerable comparative dataset is already available.

Additional morphological characters should however also be considered. So far, mostly macroscopic features were considered –particularly the organization of aquiferous system, the spicule form or the spicule arrangement in the skeleton. Cytological characters on the other hand, are a potential additional source of information. Most cytological works on Calcarea focus on a certain species (see e.g., (Eerkes-Medrano and Leys, 2006; Leys and Di, 2006)). Cytological disparities were however reported between closely related species in a few cases. E.g., Johnson (1978) described differences in size and number granules of pinacocytes and porocytes between two *Clathrina* species, and also mentioned dissimilarities in the surface fine structure of the spicules, which however were to variable within each species to be quantified. Similarly, Wörheide and Hooper (1999) reported differences between several *Clathrina* species in the density of choanocytes, their arrangement in the choanoderm, in presence/absence of granular cells, and in the form and arrangement of porocytes. However, according to them, a more thorough, ultrastructural approach than the one used in their paper is to adequately describe and identify cell types.

Both these studies suggest, that cytological and spicule fine structure should not be neglected, especially considering the high amount of homoplasy in macroscopic morphological features of Calcarea.

The combination of molecular and morphological approaches will allow defining (or redefining) calcarean taxa based upon their phylogenetic relationships and identify the abundant morphological homoplasies (Manuel et al., 2003; Manuel, 2006) that have hampered phylogenetic classification in the past.

Evolution of mitochondrial genomes

The two studies presented in **chapters 4 and 5** contributed to increasing our understanding of different aspects of the complex evolution of mt genome evolution:

1. Gene transfer from the mt genome to the nuclear genome, and evidence for the mechanisms responsible (**chapter 4**)

2. The organization of fragmented mt genomes in Metazoa, (**chapter 5**), which, contrary to former hypotheses (Warrior, 1987), apparently do not require special forms of mt genome replication

The mt genome sequence of *Amphimedon queenslandica* (Demospongiae, Haplosclerida) proved that the diversity of mt organization in poriferan mt genomes is higher than previously expected (**chapter 4**). The gene *atp9,* which in Metazoa only occurs in the mt genomes of Porifera (Lavrov et al., 2005), is missing in *Amphimedon*. Although additional mt genomes of Porifera have been sequenced since the presented study (e.g., Haen et al., 2007; Rosengarten et al., 2008; Wang and Lavrov, 2008; Lukić-Bilela et al., 2008), the lack of *atp9* remains a unique feature in the mt genome

of *Amphimedon queenslandica*, suggesting that the loss of genes can occur independently in Metazoa. Interestingly, the a*tp9* gene could be located in *Amphimedon queenslandica*'s nuclear genome, where the gene is flanked by inverted terminal repeats typical for transposons. This study thus provided evidence for a possible transposon-mediated transposition of the *atp9* gene from the mt genome to the nuclear genome.

The arrangement of a specially modified mt genome was explored by determining the sequence of the two linear mt chromosomes of *Hydra magnipapillata* (**chapter 5**). Despite the fragmentation of the two chromosomes, the organization of coding genes was not fundamentally different to unfragmented linear mt genomes in *Hydra*. In contrast to former assumptions however, the two linear molecules of the fragmented mt genome possess identical inverted terminal repeats (ITRs), which are also known from other, unfragmented linear mt genomes. Because the terminal ends of linear mt chromosomes play a crucial role in mt replication, the result implies that the replication of fragmented linear mt genomes has not to be different from the one found in other medusozoans (which, however, is also not understood yet). Additionally, the regions adjacent to the ITRs are identical between the 5' and 3' ends of both mt chromosomes. Also, because of the resemblance in mt genome organization at the ends of each of the two mt chromosome between *Hydra magnipapillata* and other fragmented and linear mt genomes of unicellular eukaryotes, we can expect comparable adaptations in other linear mt genomes of Cnidaria and other eukaryote taxa.

Conclusion

With the presented work, I contributed to the evaluation of different methodologies in the field of molecular evolution. Although nowadays more and more studies use phylogenomic approaches to resolve the deeper nodes of the Metazoan tree and the relationships of phyla in basal diverging Metazoa (Delsuc et al., 2005; Dunn et al., 2008; Philippe et al., 2009), the analyses of rRNA genes for the studies of larger phylogenies at a shallower taxonomic level still proved a valuable tool that can be used to solve pending questions in the classification of taxonomically challenging taxa. The full potential of rRNA gene analyses had not been exhaustively exploited before, because the advantages of doublet models were long neglected in many phylogenetic analyses. Additionally, the secondary structure of hyper-variable insertions in rRNAs can serve as a source of phylogenetic information. The largest benefit from rRNA data comes from the amount of available data, which probably provides the best coverage in sense of taxonomic sampling. Future phylogenetic studies should therefore combine the advantages of these markers, and it should become good practice to analyze them with doublet models to maximize the amount of phylogenetic information.

Similarly, studying individual mt genomes in depth helps us understand molecular evolutionary processes acting on (organellar) genomes. In the presented studies, publicly available data from genome projects – complemented with some PCR experiments – were used to assemble complete mt genomes. Although to my knowledge such methods have not previously been used, it proved very suitable to infer the sequence of (almost) complete mt genomes. As more and more genome traces are available from an increasing number of genome projects, the approach can easily be

applied to other taxa. The ongoing studies of the processes of molecular evolution on gene and mt genome level will help us gain a clearer picture of the evolution and relationships at the very base of the metazoan tree of life.

Bibliography

Adams, C. L., McInerney, J. and Kelly, M. (1999). Indications of relationships between poriferan classes using full-length 18S rRNA gene sequences. *Memoirs of the Queensland Museum* **44:** 33-44.

Akaike, H. (1974). A new look at the statistical model identifications. *IEEE transactions on automatic contro*l **AC-19**: 716-723.

Alfaro, M. E., Zoller, S. and Lutzoni, F. (2003). Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution* **20:** 255-266.

Altekar, G., Dwarkadas, S., Huelsenbeck, J. P. and Ronquist, F. (2004). Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20(3) :** 407-415.

Altschul, S., Gisch, W., Miller, W., Myers, E. W. and Lipman, D. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215(3)** : 403-410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25(17)** : 3389-3402.

Armstrong, M. R., Blok, V. C. and Phillips, M. S. (2000). A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. *Genetics* **154(1)** : 181-192.

Ax, P. (1995). Das System der Metazoa I. Ein Lehrbuch der phylogenetischen Systematik. pp. 226. Gustav Fischer Verlag, Stuttgart.

Ban, N., Nissen, P., Hansen, J., Moore, P. B. and Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289(5481) :** 905-920.

Beagley, C. T., Okimoto, R. and Wolstenholme, D. R. (1998). The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): Introns, a paucity of tRNA genes, and a near-standard genetic code. *Genetics* **148(3) :** 1091-1108.

Ben Ali, A., Wuyts, J., De Wachter, R., Meyer, A. and Van de Peer, Y. (1999). Construction of a variability map for eukaryotic large subunit ribosomal RNA. *Nucleic Acids Research* **27(14) :** 2825-2831.

Bidder, G. P. (1898). The skeleton and classification of calcareous sponges. *Proceedings of the Royal Society, London* **6(403) :** 61-76.

Böger, H. (1988). Versuch über das phylogenetische System der Porifera. *Meyniana* **40:** 143-154.

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research* **27(8) :** 1767-1780.

Borchiellini, C., Chombard, C., Manuel, M., Alivon, E., Vacelet, J. and Boury-Esnault, N. (2004). Molecular phylogeny of Demospongiae: implications for classification and scenarios of character evolution. *Molecular Phylogenetics and Evolution* **32(3) :** 823-837.

Borchiellini, C., Manuel, M., Alivon, E., Boury-Esnault, N., Vacelet, J. and Parco, Y., Le (2001). Sponge paraphyly and the origin of Metazoa. *Journal of Evolutionary Biology* **14(1)** : 171-179.

Borojevic, R., Boury-Esnault, N., Manuel, M. and Vacelet, J. (2002a). Order Baerida Borojevic, Boury-Esnault & Vacelet, 2000. In: *Systema Porifera. A Guide to the Classification of Sponges.* (eds. J. N. A. Hooper and R. W. M. Van Soest), pp. 1193-1199. Plenum, New York.

Borojevic, R., Boury-Esnault, N., Manuel, M. and Vacelet, J. (2002b). Order Clathrinida Hartman, 1958. In: *Systema Porifera. A Guide to the Classification of Sponges.* (eds. J. N. A. Hooper and R. W. M. Van Soest), pp. 1141-1152. Plenum, New York.

Borojevic, R., Boury-Esnault, N., Manuel, M. and Vacelet, J. (2002c). Order Leucosolenida Hartman, 1958. In: *Systema Porifera. A Guide to the Classification of Sponges.* (eds. J. N. A. Hooper and R. W. M. Van Soest), pp. 1157-1184. Plenum, New York.

Borojevic, R. (1979). Evolution des spongiaires Calcarea. In: *Biologie des Spongiaires* (eds. C. Lévi and N. Boury-Esnault), pp. 527-530. Editions du C.N.R.S, Paris.

Borojevic, R., Boury-Esnault, N. and Vacelet, J. (1990). A revision of the supraspecific classification of the subclass Calcinea (Porifera, class Calcarea). *Bulletin du Museum National d'Histoire Naturelle, Section A, Zoologie Biologie et Ecologie Animales* **12(2) :** 243-276.

Borojevic, R., Boury-Esnault, N. and Vacelet, J. (2000). A revision of the supraspecific classification of the subclass Calcaronea (Porifera, class Calcarea). *Zoosystema* **22(2) :** 203-263.

Borojevic, R. and Klautau, M. (2000). Calcareous sponges from New Caledonia. *Zoosystema* **22(2) :** 187-201.

Boury-Esnault, N., Ereskovsky, A., Bezac, C. and Tokina, D. (2003). Larval development in the Homoscleromorpha (Porifera, Demospongiae). *Invertebrate Biology* **122(3) :** 187-202.

Boyle, J. S. and Lew, A. M. (1995). An inexpensive alterntive to glassmilk for DNA purification. *Trends in Genetics* **11:** 8.

Bridge, D., Cunningham, C. W., Schierwater, B., DeSalle, R. and Buss, L. W. (1992). Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. *Proceedings of the National Academy of Sciences of the United States of America* **89(18) :** 8750-8753.

Brugler, M. and France, S. (2008). The mitochondrial genome of a deep-sea bamboo coral (Cnidaria, Anthozoa, Octocorallia, Isididae): Genome structure and putative origins of replication are not conserved among octocorals. *Journal of Molecular Evolution* **67(2) :** 125-136.

Brugler, M. R. and France, S. C. (2007). The complete mitochondrial genome of the black coral *Chrysopathes formosa* (Cnidaria:Anthozoa:Antipatharia) supports classification of antipatharians within the subclass Hexacorallia. *Molecular Phylogenetics and Evolution* **42(3) :** 776-788.

Burger, G., Forget, L., Zhu, Y., Gray, M. W. and Lang, B. F. (2003a). Unique mitochondrial genome architecture in unicellular relatives of animals. *Proceedings of the National Academy of Sciences of the United States of America* **100(3) :** 892-897.

Burger, G., Gray, M. W. and Lang, B. F. (2003b). Mitochondrial genomes: anything goes. *Trends in Genetics* **19(12) :** 709-716.

Burger, G. and Lang, B. F. (2003). Parallels in genome evolution in mitochondria and bacterial symbionts. *IUBMB Life* **55(4-5) :** 205-212.

Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., Pande, N., Shang, Z., Yu, N. and Gutell, R. R. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**: 2.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17(4) :** 540-552.

Cavalier-Smith, T. and Chao, E. E. (2003). Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. *Journal of Molecular Evolution* **56(5) :** 540-563.

Cavalier-Smith, T., Chao, E. E., BouryEsnault, N. and Vacelet, J. (1996). Sponge phylogeny, animal monophyly, and the origin of the nervous system: 18S rRNA evidence. *Canadian Journal of Zoology* **74(11) :** 2031-2045.

Clemons, W. M., May, J. L., Wimberly, B. T., McCutcheon, J. P., Capel, M. S. and Ramakrishnan, V. (1999). Structure of a bacterial 30S ribosomal subunit at 5.5 A resolution. *Nature* **400(6747) :** 833-840.

Collins, A. G. (1998). Evaluating multiple alternative hypotheses for the origin of Bilateria: an analysis of 18S rRNA molecular evidence. *Proceedings of the National Academy of Sciences of the United States of America* **95(26) :** 15458-15463.

Collins, A. G., Schuchert, P., Marques, A. C., Jankowski, T., Medina, M. and Schierwater, B. (2006). Medusozoan phylogeny and character evolution clarified by new large and small subunit rDNA data and an assessment of the utility of phylogenetic mixture models. *Systematic Biology* **55(1) :** 97-115.

Crease, T. J. and Taylor, D. J. (1998). The origin and evolution of variable-region helices in V4 and V7 of the small-subunit ribosomal RNA of branchiopod crustaceans. *Molecular Biology and Evolution* **15(11) :** 1430-1446.

Daly, M., Brugler, M. R., Cartwright, P., Collins, A. G., Dawson, M., Fautin, D., France, S. C., McFadden, C., Opresko, D., Rodriguez, E., Romano, S. L. and Stake, J. (2007). The phylum Cnidaria: A review of phylogenetic patterns and diversity 300 years after Linnaeus. *Zootaxa* **1668**: 127-182.

De Rijk, P., Wuyts, J. and De Wachter, R. (2003). RnaViz 2: an improved representation of RNA secondary structure. *Bioinformatics* **19(2) :** 299-300.

Dellaporta, S. L., Xu, A., Sagasser, S., Jakob, W., Moreno, M. A., Buss, L. W. and Schierwater, B. (2006). Mitochondrial genome of *Trichoplax adhaerens* supports Placozoa as the basal lower metazoan phylum. *Proceedings of the National Academy of Sciences of the United States of America* **103(23) :** 8751-8756.

Delsuc, F., Brinkmann, H. and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics* **6:** 361-375.

Dendy, A. (1892). Preliminary account of *Synute pulchella*, a new genus and species of calcareous sponges. *Proceedings of the Royal Society of Victoria* **4(1) :** 1-6.

Dendy, A. (1893). Synopsis of the Ausralian Calcarea Heterocoela; with a proposed classification of the group and descriptions of some new genera and species. *Proceedings of the Royal Society of Victoria* **5:** 69-116.

Dinouel, N., Drissi, R., Miyakawa, I., Sor, F., Rousset, S. and Fukuhara, H. (1993). Linear mitochondrial DNAs of yeasts: closed-loop structure of the termini and possible linear-circular conversion mechanisms. *Molecular and Cellular Biology* **13(4) :** 2315-2323.

Dixon, M. T. and Hillis, D. M. (1993). Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. *Molecular Biology and Evolution* **10(1) :** 256-267.

Dohrmann, M., Voigt, O., Erpenbeck, D. and Wörheide, G. (2006). Non-monophyly of most supraspecific taxa of calcareous sponges (Porifera, Calcarea) revealed by increased taxon sampling and partitioned Bayesian analysis of ribosomal DNA. *Molecular Phylogenetics and Evolution* **40(3) :** 830-843.

Dohrmann, M., Collins, A. G. and Wörheide, G. (2009). New insights into the phylogeny of glass sponges (Porifera, Hexactinellida): monophyly of Lyssacinosida and Euplectellinae, and the phylogenetic position of Euretidae. *Molecular Phylogenetics and Evolution* **52(1) :** 257-262.

Dohrmann, M., Janussen, D., Reitner, J., Collins, A. G. and Wörheide, G. (2008). Phylogeny and evolution of glass sponges (Porifera, Hexactinellida). *Systematic Biology* **57(3) :** 388-405.

Drabkin, H. J., Estrella, M. and Rajbhandary, U. L. (1998). Initiator-elongator discrimination in vertebrate tRNAs for protein synthesis. *Molecular and Cellular Biology* **18(3) :** 1459-1466.

Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., Seaver, E., Rouse, G. W., Obst, M., Edgecombe, G. D., Sørensen, M. V., Haddock, S. H., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R. M. et al. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452(7188)** : 745-749.

Elder, J. F. and Turner, B. J. (1995). Concerted Evolution of Repetitive DNA sequences in Eukaryotes. *The Quarterly Review of Biology* **70(3) :** 297-320.

Ender, A. and Schierwater, B. (2003). Placozoa are not derived cnidarians: evidence from molecular morphology. *Molecular Biology and Evolution* **20(1) :** 130-134.

Erixon, P., Svennblad, B., Britton, T. and Oxelman, B. (2003). Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic Biology* **52(5) :** 665-673.

Eerkes-Medrano, D. I. and Leys, S. P. (2006). Ultrastructure and embryonic development of a syconoid calcareous sponge. *Invertebrate Biology* **125(3) :** 177-194.

Erpenbeck, D., Nichols, S. A., Voigt, O., Dohrmann, M., Degnan, B. M., Hooper, J. N. and Wörheide, G. (2007a). Phylogenetic analyses under secondary structure-specific substitution models outperform traditional approaches: Case studies with diploblast LSU. *Journal of Molecular Evolution* **64(5) :** 543-557.

Erpenbeck, D., Voigt, O., Adamski, M., Adamska, M., Hooper, J. N., Wörheide, G. and Degnan, B. M. (2007b). Mitochondrial diversity of early-branching metazoa is revealed by the complete mt genome of a haplosclerid demosponge. *Molecular Biology and Evolution* **24(1)(1) :** 19-22.

Erpenbeck, D., Breeuwer, J. A., Parra-Velandia, F. J. and van Soest, R. W. (2006). Speculation with spiculation?-Three independent gene fragments and biochemical characters versus morphology in demosponge higher classification. *Molecular Phylogenetics and Evolution* **38(2) :** 293-305.

Erpenbeck, D., McCormack, G. P., Breeuwer, J. A. J. and van Soest, R. W. M. (2004). Order level differences in the structure of partial LSU across demosponges (Porifera): new insights into an old taxon. *Molecular Phylogenetics and Evolution* **32(1) :** 388-395.

Erpenbeck, D. and Wörheide, G. (2007). On the molecular phylogeny of sponges (Porifera). *Zootaxa* **1668:** 107-126.

Fan, J. and Lee, R. W. (2002). Mitochondrial genome of the colorless green alga *Polytomella parva*: Two linear DNA molecules with homologous inverted repeat termini. *Molecular Biology and Evolution* **19(7) :** 999-1007.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39(4) :** 783-791.

Galtier, N. (2004). Sampling properties of the bootstrap support in molecular phylogeny: Influence of non-independence among sites. *Systematic Biology* **53(1) :** 38-46.

Galtier, N., Gouy, M. and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Computer applications in the biosciences : CABIOS* **12(6) :** 543-548.

Gibson, T., Blok, V. C. and Dowton, M. (2007a). Sequence and characterization of six mitochondrial subgenomes from *Globodera rostochiensis*: multipartite structure is conserved among close nematode relatives. *Journal of Molecular Evolution* **65(3) :** 308-315.

Gibson, T., Blok, V. C., Phillips, M. S., Hong, G., Kumarasinghe, D., Riley, I. T. and Dowton, M. (2007b). The mitochondrial subgenomes of the nematode *Globodera pallida* are mosaics: evidence of recombination in an animal mitochondrial genome. *Journal of Molecular Evolution* **64(4)** : 463-471.

Gillespie, J. J., McKenna, C. H., Yoder, M. J., Gutell, R. R., Johnston, J. S., Kathirithamby, J. and Cognato, A. I. (2005a). Assessing the odd secondary structural properties of nuclear small subunit ribosomal RNA sequences (18S) of the twisted-wing parasites (Insecta: Strepsiptera). *Insect molecular biology* **14(6) :** 625-643.

Gillespie, J. J., Yoder, M. J. and Wharton, R. A. (2005b). Predicted secondary structure for 28S and 18S rRNA from Ichneumonoidea (Insecta: Hymenoptera: Apocrita): impact on sequence alignment and phylogeny estimation. *Journal of Molecular Evolution* **61(1) :** 114-137.

Gray, M. W., Burger, G. and Lang, B. F. (1999). Mitochondrial evolution. *Science* **283** : 1476-1481.

Gray, M. W., Lang, B. F. and Burger, G. (2004). Mitochondria of protists. *Annual Review of Genetics* **38** : 477-524.

Green, R. and Noller, H. F. (1997). Ribosomes and translation. *Annual Review of Biochemistry* **66:** 679-716.

Grell, K. G. (1972). Eibildung und Furchung von Trichoplax adhaerens F.E. Schulze (Placozoa). Zeitschrift für Morphologie und Ökologie der Tiere **73:** 297-314.

Grell, K. G. and Ruthmann, A. (1991). Placozoa. In: *Microscopic Anatomy of Invertebrates* (eds. F. W. Harrison and E. E. Ruppert), pp. 13-27. Wiley-Liss, Inc, New York.

Gu, X., Fu, Y. X. and Li, W. H. (1995). Maximum-likelihood-estimation of the heterogeneity of substitution rate among nucleotide sites. *Molecular Biology and Evolution* **12(4)** : 546-557.

Gutell, R. R. (1993). Collection of small subunit (16S- and 16S-like) ribosomal RNA structures. *Nucleic Acids Research* **21(13) :** 3051-3054.

Haeckel, E. (1872a). Die Kalkschwämme, Band 1: Biologie der Kalkschwämme (Calcispongien oder Grantien). pp. 484. Georg Reimer, Berlin.

Haeckel, E. (1872b). Die Kalkschwämme, Band 2: System der Kalkschwämme (Calcispongien oder Grantien). pp. 416. Georg Reimer, Berlin.

Haeckel, E. (1872c). Die Kalkschwämme, Band 3: Atlas der Kalkschwämme (Calcispongien oder Grantien). Georg Reimer, Berlin.

Haen, K. M., Lang, B. F., Pomponi, S. A. and Lavrov, D. V. (2007). Glass sponges and bilaterian animals share derived mitochondrial genomic features: a common ancestry or parallel evolution? *Molecular Biology and Evolution* **24(7) :** 1518-1527.

Halanych, K. M. (2004). The new view of animal phylogeny. Annual Review of Ecology, Evolution, and Systematics **35:** 229-256.

Hancock, J. M., Tautz, D. and Dover, G. A. (1988). Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. *Molecular Biology and Evolution* **5(4) :** 393-414.

Hancock, J. M. and Vogler, A. P. (2000). How slippage-derived sequences are incorporated into rRNA variable-region secondary structure: implications for phylogeny reconstruction. *Molecular Phylogenetics and Evolution* **14(3) :** 366-374.

Hartman, W. D. (1958). A re-examination of Bidder's classification of the Calcarea. *Systematic Zoology* **7:** 97-110.

Hebert, P. D. N., Cywinska, A., Ball, S. L. and DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B* **270(1512)** : 313-321.

Hemmrich, G., Anokhin, B., Zacharias, H. and Bosch, T. C. (2007). Molecular phylogenetics in *Hydra,* a classical model in evolutionary developmental biology. *Molecular Phylogenetics and Evolution* **44(1) :** 281-290.

Henze, K. and Martin, W. (2001). How do mitochondrial genes get into the nucleus? *Trends in Genetics* **17(7) :** 383-387.

Hickson, R. E., Simon, C. and Perrey, S. W. (2000). The performance of several multiple-sequence alignment programs in relation to secondary-structure features for an rRNA sequence. *Molecular Biology and Evolution* **17(4) :** 530-539.

Higgs, P. G. (2000). RNA secondary structure: physical and computational aspects. *Quarterly Reviews of Biophysics* **33(3) :** 199-253.

Hillis, D. and Bull, J. J. (1993). An empirical-test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42(2) :** 182-192.

Hillis, D. M. and Dixon, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly Review of Biology* **66(4) :** 411-453.

Hooper, J. and van Soest, R. W. M. (2002). Class Demospongiae Sollas,1885. In: Systema Porifera: A Guide to the Classification of Sponges (eds. J. Hooper and R. W. M. van Soest), pp. 15-18.

Hooper, J. N., van Soest, R. W. M. and Debrenne, F. (2002). Phylum Porifera Grant,1836. In: Systema Porifera: A Guide to the Classification of Sponges (eds. J. N. Hooper and R. W. M. van Soest), pp. 9-13.

Hooper, J. N. A. and van Soest, R. W. M. (2006). A new species of *Amphimedon* (Porifera, Demospongiae, Haplosclerida, Niphatidae) from the Capricorn-Bunker Group of Islands, Great Barrier Reef, Australia: target species for the 'sponge genome project'. *Zootaxa* **1314:** 31-39.

Hooper, J. N. A. and van Soest, R. W. M. (2002). Systema Porifera: a guide to the classification of sponges. (eds. J. N. A. Hooper and R. W. M. van Soest), Kluwer Academic/ Plenum Publishers, New York.

Huang, X. and Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research* **9(9) :** 868-877.

Hudelot, C., Gowri-Shankar, V., Jow, H., Rattray, M. and Higgs, P. G. (2003). RNA-based phylogenetic methods: application to mammalian mitochondrial RNA sequences. *Molecular Phylogenetics and Evolution* **28(2) :** 241-252.

Huelsenbeck, J. and Rannala, B. (2004). Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology* **53(6) :** 904-913.

Huelsenbeck, J. and Ronquist, F. (2005). Bayesian analysis of molecular evolution using MrBayes. In: *Statistical methods in molecular evolution* (ed. R. Nielsen), pp. 183-323. Springer, New York.

Huelsenbeck, J. P., Larget, B., Miller, R. E. and Ronquist, F. (2002). Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* **51(5) :** 673-688.

Huelsenbeck, J. P. and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17(8) :** 754-755.

Jackson, C. J., Norman, J. E., Schnare, M. N., Gray, M. W., Keeling, P. J. and Waller, R. F. (2007). Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. *BMC Biology* **5(1) :** 41.

James-Clark, H. (1866). Conclusive prrofs on the animality of the ciliate sponges, and their affinities with the Infusoria Flagellata. *American Journal of Science, Series 2* **42:** 320-325.

Johnson, M. (1978). Significance of life-history studies of calcareous sponges for species determination. *Bulletin of Marine Science* **28(3)** : 570-574.

Jow, H., Hudelot, C., Rattray, M. and Higgs, P. G. (2002). Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. *Molecular Biology and Evolution* **19(9) :** 1591-1601.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. Journal of the American Statistical Association **90(430) :** 773-795.

Kayal, E. and Lavrov, D. V. (2008). The mitochondrial genome of *Hydra oligactis* (Cnidaria, Hydrozoa) sheds new light on animal mtDNA evolution and cnidarian phylogeny. *Gene* **410(1) :** 177-186.

Keane, T. M., Naughton, T. J., Travers, S. A., McInerney, J. O. and McCormack, G. P. (2005). DPRml: distributed phylogeny reconstruction by maximum likelihood. *Bioinformatics* **21(7)** : 969-974.

Kjer, K. M. (1995). Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. *Molecular Phylogenetics and Evolution* **4(3) :** 314-330.

K**lautau, M. and Valentine, C.** (2003). Revision of the genus *Clathrina* (Porifera, Calcarea). *Zoological Journal of the Linnean Society* **139(1) :** 1-62.

Kruse, M., Leys, S. P., Muller, I. M. and Müller, W. E. G. (1998). Phylogenetic position of the hexactinellida within the phylum porifera based on the amino acid sequence of the protein kinase C from *Rhabdocalyptus dawsoni*. *Journal of Molecular Evolution* **46(6) :** 721-728.

Lafay, B., Boury-Esnault, N., Vacelet, J. and Christen, R. (1992). An analysis of partial 28S ribosomal RNA sequences suggests early radiations of sponges. *BioSystems* **28(1-3) :** 139-151.

Lafontaine, D. L. and Tollervey, D. (2001). The function and synthesis of ribosomes. *Nature Reviews. Molecular Cell Biology* **2(7) :** 514.

Lang, B. F., O'Kelly, C. J., Nerad, T., Gray, M. W. and Burger, G. (2002). The closest unicellular relatives of animals. *Current Biology* **12(20)** : 1773-1778.

Lang, B. F., Gray, M. W. and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annual Review of Genetics* **33:** 351-397.

Larget, B. and Simon, D. L. (1999). Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution* **16(6) :** 750-759.

Lartillot, N. and Philippe, H. (2004). A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. *Molecular Biology and Evolution* **21(6) :** 1095-1109.

Lavrov, D. V., Forget, L., Kelly, M. and Lang, B. F. (2005). Mitochondrial genomes of two demosponges provide insights into an early stage of animal evolution. *Molecular Biology and Evolution* **22(5) :** 1231-1239.

Lavrov, D. V. and Lang, B. F. (2005). Transfer RNA gene recruitment in mitochondrial DNA. *Trends in Genetics* **21(3) :** 129-133.

Lavrov, D. V. (2007). Key transitions in animal evolution: a mitochondrial DNA perspective. *Integrative and Comparative Biology* **47(5) :** 734-743.

Lee, J. C. and Gutell, R. R. (2004). Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs. *Journal of Molecular Biology* **344(5) :** 1225-1249.

Lemmon, A. R. and Moriarty, E. C. (2004). The importance of proper model assumption in Bayesian phylogenetics. *Systematic Biology* **53(2) :** 265-277.

Leys, S.P. and Eerkes-Medrano, D. I. (2006). Feeding in a calcareous sponge: particle uptake by pseudopodia. *The Biological Bulletin* **211(2) :** 157-171.

Liao, D. (1999). Concerted evolution: molecular mechanism and biological implications. *American Journal of Human Genetics* **64(1) :** 24-30.

Lukić-Bilela, L., Brandt, D., Pojskić, N., Wiens, M., Gamulin, V. and Müller, W. E. (2008). Mitochondrial genome of *Suberites domuncula*: Palindromes and inverted repeats are abundant in non-coding regions. *Gene* **412(1-2)** : 1-11.

Maddison, W. P. and Maddison, D. R. (2002). MacClade. Sinauer Associates, Sunderland, MA.

Manuel, M., Borchiellini, C., Alivon, E. and Boury-Esnault, N. (2004). Molecular phylogeny of calcareous sponges using 18S rRNA and 28S rRNA sequences. *Bollettino dei musei e degli istituti biologici dell'Università di Genova* **68:** 449-461.

Manuel, M., Borchiellini, C., Alivon, E., Le Parco, Y., Vacelet, J. and Boury-Esnault, N. (2003). Phylogeny and evolution of calcareous sponges: Monophyly of Calcinea and Calcaronea, high level of morphological homoplasy, and the primitive nature of axial symmetry. *Systematic Biology* **52(3) :** 311-333.

Manuel, M., Borojevic, R., Boury-Esnault, N. and Vacelet, J. (2002). Class Calcarea Bowerbank,1864. In: *Systema Porifera:A Guide to the Classification of Sponges* (eds. N. A. Hooper and R. W. M. van Soest), pp. 1103-1110. Kluwer Academic/Plenum Publishers, New York.

Manuel, M. (2006). Phylogeny and evolution of calcareous sponges. *Canadian Journal of Zoology* **84:** 225-241.

Manuel, M., Borojevic, R., Boury-Esnault, N. and Vacelet, J. (2002). Class Calcarea Bowerbank 1864. In: *Systema Porifera: A Guide to the Classification of Sponges* (eds. J. N. A. Hooper and R. W. M. van Soest), pp. 1103-1110. Kluwer Academic/ Plenum Publishers, New York.

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M. and Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America* **101(19) :** 7287-7292.

McCormack, G., P., Erpenbeck, D. and van, S.,R. W. M. (2002). Major discrepancy between phylogenetic hypotheses based on molecular and morphological criteria within the Order Haplosclerida (Phylum Porifera: Class Demospongiae). J*ournal of zoological systematics and evolutionary research* **40(4) :** 237-240.

Medina, M., Collins, A. G., Silberman, J. D. and Sogin, M. L. (2001). Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proceedings of the National Academy of Sciences* **98(17) :** 9707-9712.

Medina, M., Collins, A. G., Takaoka, T. L., Kuehl, J. V. and Boore, J. L. (2006). Naked corals: skeleton loss in Scleractinia. *Proceedings of the National Academy of Sciences of the United States of America* **103(24) :** 9096-9100.

Michot, B., Qu, L. H. and Bachellerie, J. P. (1990). Evolution of large-subunit rRNA structure. The diversification of divergent D3 domain among major phylogenetic groups. *European Journal of Biochemistry* **188(2) :** 219-229.

Minchin, E. A. (1896). Suggestions for a Natural Classification of the Asconidae. *Annals and Magazine of Natural History (6)* **18(107) :** 349-362.

Minelli, A. (2009). Perspectives in animal phylogeny and evolution . pp. 345. Oxford University Press, Oxford.

Muse, S. V. (1995). Evolutionary analyses of DNA sequences subject to constraints on secondary structure. *Genetics* **139:** 1429-1439.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B* **56:** 3-48.

Nichols, S. A. (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(1) :** 81-96.

Nielsen, C., Scharff, N. and EibyeJacobsen, D. (1996). Cladistic analyses of the animal kingdom. *Biological Journal of the Linnean Society* **57(4) :** 385-410.

Nielsen, C. (2001). Animal evolution. Interrelationships of the living phyla. pp. 578. Oxford University Press, New York.

Nissen, P., Hansen, J., Ban, N., Moore, P. B. and Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* **289(5481) :** 920-930.

Noller, H. F. (2005). RNA structure: Reading the ribosome. *Science* **309(5740) :** 1508-1514.

Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., Dohme, F., Herr, W., Stahl, D. A., Gupta, R. and Waese, C. R. (1981). Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Research* **9(22)** : 6167-6189.

Nosek, J., Kosa, P. and Tomaska, L. (2006). On the origin of telomeres: a glimpse at the pre-telomerase world. *BioEssays* **28(2) :** 182-190.

Nosek, J. and Tomaska, L. (2003). Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. *Current Genetics* **44(2) :** 73-84.

Nosek, J., Tomaska, L. U., Fukuhara, H., Suyama, Y. and Kovac, L. (1998). Linear mitochondrial genomes: 30 years down the line. *Trends in Genetics* **14(5)** : 184-188.

Nylander, J. A., Ronquist, F., Huelsenbeck, J. P. and Nieves-Aldrey, J. L. (2004). Bayesian phylogenetic analysis of combined data. *Systematic Biology* **53(1**) : 47-67.

Nylander, J. A., Wilgenbusch, J. C., Warren, D. L. and Swofford, D. L. (2008). AWTY (Are We There Yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* **24(4) :** 581-583.

Pearse, V. B., Pearse, J. S., Buchsbaum, M. and Buchsbaum, R. (1987). Living invertebrates. pp. 848. Boxwood Press, Pacific Grove.

Pearse, V. B. and Voigt, O. (2007). Field biology of placozoans (*Trichoplax*): distribution, diversity, biotic interactions. *Integrative and Comparative Biology* **47(5) :** 677-692.

Perna, N. T. and Kocher, T. D. (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution* **41(3) :** 353-358.

Peterson, K. J. and Butterfield, N. J. (2005). Origin of the Eumetazoa: testing ecological predictions of molecular clocks against the Proterozoic fossil record. *Proceedings of the National Academy of Sciences of the United States of America* **102(27) :** 9547-9552.

Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Quéinnec, E., Da Silva, C., Wincker, P., Le Guyade, H., Leys, S., Jackson, D. J. et al. (2009). Phylogenomics revives traditional views on deep animal relationships. *Current Biology* **19(8)** : 706-712.

Pickett, J. (2002). Fossil Calcarea. In: *Systema Porifera: a guide to the classification of sponges.* (eds. J. Hooper and R. W. M. van Soest), pp. 1117-1119. Kluwer Academic/ Plenum Publishers, New York.

Poléjaeff, N. (1883). Report on the Calcarea dredged by H.M.S. 'Challenger', during the years 1873-1876. *Report on the scientific results of the voyage of H.M.S. 'Challenger', 1873-1876. Zoology* **8(2) :** 1-76.

Pons, J. and Vogler, A. P. (2005). Complex pattern of coalescence and fast evolution of a mitochondrial rRNA pseudogene in a recent radiation of tiger beetles. *Molecular Biology and Evolution* **22(4) :** 991-1000.

Pont-Kingdon, G., Okada, N. A., Macfarlane, J. L., Beagley, C. T., Watkins-Sims, C. D., Cavalier-Smith, T., Clark-Walker, G. D. and Wolstenholme, D. R. (1998). Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. *Journal of Molecular Evolution* **46(4)** : 419-431.

Pont-Kingdon, G., Vassort, C. G., Warrior, R., Okimoto, R., Beagley, C. T. and Wolstenholme, D. R. (2000). Mitochondrial DNA of *Hydra attenuata* (Cnidaria): a sequence that includes an end of one linear molecule and the genes for l-rRNA, tRNA(f-Met), tRNA(Trp), COII, and ATPase8. *Journal of Molecular Evolution* **51(4) :** 404-415.

P**ont-Kingdon, G. A., Okada, N. A., Macfarlane, J. L., Beagley, C. T., Wolstenholme, D. R., Cavalier-Smith, T. and Clark-Walker, G. D**. (1995). A coral mitochondrial *mutS* gene. *Nature* **375** : 109-111.

Posada, D. and Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14(9) :** 817-818.

Pritchard, A. E. and Cummings, D. J. (1981). Replication of linear mitochondrial DNA from Paramecium: sequence and structure of the initiation-end crosslink. *Proceedings of the National Academy of Sciences of the United States of America* **78(12) :** 7341-7345.

Raleigh, J., Redmond, N. E., Delahan, E., Torpey, S., van Soest, R. W. M., Kelly, M. and McCormack, G. P. (2007). Mitochondrial Cytochrome oxidase 1 phylogeny supports alternative taxonomic scheme for the marine Haplosclerida. *Journal of the Marine Biological Association of the UK* **87(6) :** 1577-1584.

Redmond, N. E., van Soest, R. W., Kelly, M, Raleigh, J., Travers, S. A. and GP, M. (2007). Reassessment of the classification of the Order Haplosclerida (Class Demospongiae, Phylum Porifera) using 18S rRNA gene sequence data. *Molecular Phylogenetics and Evolution* **43(1) :** 344-352.

Reiswig, H. M. (2002). Class Hexactinellida Schmidt,1870. In: Systema Porifera: A Guide to the Classification of Sponges (eds. J. N. Hooper and R. W. M. van Soest), pp. 1201-1202.

Reitner, J. (1992). "Coralline Spongien". Der Versuch einer phylogenetisch taxonomischen Analyse. *Berliner Geowissenschaftliche Abhandlungen, Reihe (E), Paläobiologie* **1:** 1-352.

Rokas, A. and Carroll, S. B. (2006). Bushes in the tree of life. *PLoS Biology* **4(11) :** e352.

Rokas, A., Kruger, D. and Carroll, S. B. (2005). Animal evolution and the molecular signature of radiations compressed in time. *Science* **310(5756) :** 1933-1938.

Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19(12) :** 1572-1574.

Rosengarten, R. D., Sperling, E. A., Moreno, M. A., Leys, S. P. and Dellaporta, S. (2008). The mitochondrial genome of the hexactinellid sponge *Aphrocallistes vastus*: evidence for programmed translational frameshifting. *BMC Genomics* **9(1)** : 33.

Rousset, F., Pélandakis, M. and Solignac, M. (1991). Evolution of compensatory substitutions through G.U intermediate state in Drosophila rRNA. *Proceedings of the National Academy of Sciences of the United States of America* **88(22) :** 10032-10036.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. and Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16(10)** : 944-945.

Sakaguchi, K. (1990). Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of adeno-type viruses. *Microbiological Reviews* **54(1) :** 66-74.

Sánchez, J. A., Aguilar, C., Dorado, D. and Manrique, N. (2007). Phenotypic plasticity and morphological integration in a marine modular invertebrate. *BMC Evolutionary Biology* **7:** 122.

Savill, N. J., Hoyle, D. C. and Higgs, P. G. (2001). RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. *Genetics* **157(1) :** 399-411.

Schierwater, B., Eitel, M., Jakob, W., Osigus, H. J., Hadrys, H., Dellaporta, S. L., Kolokotronis, S. O. and Desalle, R. (2009). Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLoS Biology* **7(1)** : e20.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. and Yonath, A. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell* **102(5)** : 615-623.

Schnare, M. N., Damberger, S. H., Gray, M. W. and Gutell, R. R. (1996). Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23 S-like) ribosomal RNA. *Journal of Molecular Biology* **256(4) :** 701-719.

Schöniger, M. and von Haeseler, A. (1994). A stochastic model for the evolution of autocorrelated DNA sequences. *Molecular Phylogenetics and Evolution* **3(3) :** 240-247.

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of statistics* **6** : 461-464.

Sempere, L. F., Cole, C. N., McPeek, M. A. and Peterson, K. J. (2006). The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **306(6) :** 575-588.

Sempere, L. F., Martinez, P., Cole, C., Baguñà, J. and Peterson, K. J. (2007). Phylogenetic distribution of microRNAs supports the basal position of acoel flatworms and the polyphyly of Platyhelminthes. *Evolution & Development* **9(5) :** 409-415.

Shao, Z., Graf, S., Chaga, O. Y. and Lavrov, D. V. (2006). Mitochondrial genome of the moon jelly *Aurelia aurita* (Cnidaria, Scyphozoa): A linear DNA molecule encoding a putative DNA-dependent DNA polymerase. *Gene* **381:** 92-101.

Siddall, M. E., Martin, D. S., Bridge, D., Desser, S. S. and Cone, D. K. (1995). The demise of a phylum of protists: phylogeny of Myxozoa and other parasitic cnidaria. *Journal of Parasitology* **81(6) :** 961-967.

Signorovitch, A. Y., Dellaporta, S. L. and Buss, L. W. (2005). Molecular signatures for sex in the Placozoa. *Proceedings of the National Academy of Sciences of the United States of America* **102(43) :** 15518-15522.

Sinniger, F., Chevaldonné, P. and Pawlowski, J. (2007). Mitochondrial genome of Savalia savaglia (Cnidaria, Hexacorallia) and early metazoan phylogeny. *Journal of Molecular Evolution* **64(2) :** 196-203.

Slamovits, C. H., Saldarriaga, J. F., Larocque, A. and Keeling, P. J. (2007). The highly reduced and fragmented mitochondrial genome of the early-branching dinoflagellate *Oxyrrhis marina* shares characteristics with both apicomplexan and dinoflagellate mitochondrial genomes. *Journal of Molecular Biology* **372(2) :** 356-368.

Smit, S., Smit, S., Widmann, J., Widmann, J., Knight, R. and Knight, R. (2007). Evolutionary rates vary among rRNA structural elements. *Nucleic Acids Research* **35(10**) : 3339.

Sperling, E., Pisani, D. and Peterson, K. (2007). Poriferan paraphyly and its implications for Precambrian palaeobiology. In: *The Rise and Fall of the Ediacaran Biota* (eds. P. Vickers-Rich and P. Komarower), pp. 355-368. Geological Society, London, Special Publications

S**rivastava, M., Begovic, E., Chapman, J., Putnam, N. H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M. L., Signorovitch, A. Y., Moreno, M. A., Kamm, K., Grimwood, J., Schmutz, J. et al**. (2008). The *Trichoplax* genome and the nature of placozoans. *Nature* **454**: 955-960.

Storch, V. and Welsch, U. (1997). Systematische Zoologie. pp. 804. Gustav Fischer Verlag, Stuttgart.

Suga, K., Mark Welch, D. B., Tanaka, Y., Sakakura, Y. and Hagiwara, A. (2008). Two circular chromosomes of unequal copy number make up the mitochondrial genome of the rotifer *Brachionus plicatilis*. *Molecular Biology and Evolution* **25(6) :** 1129-1137.

Suzuki, Y., Glazko, G. V. and Nei, M. (2002). Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences of the United States of America* **99(25) :** 16138-16143.

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.

Swofford, D. L., Olsen, G. J., Waddell, P. J. and Hillis, D. M. (1996). Phylogenetic inference. In: *Molecular Systematics* (eds. D. M. Hillis, C. Moritz and B. K. Mable), pp. 407-514. Sinauer Associates, Sunderland, MA.

Tamura, K. and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochodrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10** : 512-526.

Tavaré, S. (1986). Some probabilistic and statistical problems on the analysis of DNA sequences. *Lec. Math. Life Sci.* **17:** 57-86.

Telford, M. J., Wise, M. J. and Gowri-Shankar, V. (2005). Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the Bilateria. *Molecular Biology and Evolution* **22(4) :** 1129-1136.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25(24) :** 4876-4882.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22(22) :** 4673-4680.

Tillier, E. R. M. and Collins, R. A. (1995). Neighbor Joining and Maximum-Likelihood with RNA sequences - addressing the interdependence of sites. *Biochemical Systematics and Ecology* **12(1) :** 7-15.

Tillier, E. R. M. and Collins, R. A. (1998). High apparent rate of simultaneous compensatory base-pair substitutions in ribosomal RNA. *Genetics* **148(4) :** 1993-2002.

Timmis, J. N., Ayliffe, M. A., Huang, C. Y. and Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* **5(2) :** 123-135.

Traut, W., Szczepanowski, M., Vítková, M., Opitz, C., Marec, F. and Zrzavý, J. (2007). The telomere repeat motif of basal Metazoa. *Chromosome research* **15(3)** : 371-382.

Vacelet, J. (1991). Recent Calcarea with a reinforced skeleton (Pharetronids). In: *Fossil and recent sponges* (eds. J. Reitner and H. Kneupp), pp. 252-268. Springer-Verlag, Heidelberg.

Vacelet, J., Borojevic, R., Boury-Esnault, N. and Manuel, M. (2002a). Order Lithonida Vacelet, 1981, recent. In: *Systema Porifera: A Guide to the Classification of Sponges* (eds. J. N. A. Hooper and R. W. M. Van Soest), pp. 1185-1192. Plenum, New York.

Vacelet, J., Borojevic, R., Boury-Esnault, N. and Manuel, M. (2002b). Order Murrayonida Vacelet, 1981. In: *Systema Porifera: A Guide to the Classification of Sponges* (eds. J. N. A. Hooper and R. W. M. Van Soest), pp. 1153-1156. Plenum, New York.

Vahrenholz, C., Riemen, G., Pratje, E., Dujon, B. and Michaelis, G. (1993). Mitochondrial DNA of *Chlamydomonas reinhardtii*: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication. *Current Genetics* **24(3) :** 241-247.

Valderrama, D., Rossi, A., Solé-Cava, A., Rapp, H. T. and Klautau, M. Revalidation of *Leucetta floridana* (Haeckel, 1872) (Porifera, Calcarea): a widespread species in the tropical western Atlantic. *Zoological Journal of the Linnean Society,* **in press**.

van Soest, R. W. M. (1996). Porifera. In: *Spezielle Zoologie. Teil 1: Einzeller und Wirbellose Tiere* (eds. W. Westheide and R. Rieger), pp. 98-119. Gustav Fischer Verlag, Stuttgart.

van Soest, R. W. M. and Hooper, J. N. A. (2002). Order Haplosclerida. In: *Systema Porifera. A guide to the classification of sponges.* (eds. J. N. A. Hooper and R. W. M. van Soest), pp. 831-832. Kluwer Academic/ Plenum Publishers, New York, Boston, Dordrecht, London, Moscow.

van Soest, R. W. M., Boury-Esnault, N., Hooper, J. N. A., Rützler, K., de Voogd, N. J., Alvarez, B., Hajdu, E., Pisera, A. B., Vacelet, J., Manconi, R., Schoenberg, C., Janussen, D., Tabachnick, K. R. and Klautau, M. (2009). The World Porifera Database.: http://www.marinespecies.org/porifera/

Voigt, O., Collins, A. G., Pearse, V. B., Pearse, J. S., Ender, A., Hadrys, H. and Schierwater, B. (2004). Placozoa -- no longer a phylum of one. *Current Biology* **14(22)** : R944-5.

Voigt, O., Erpenbeck, D. and Wörheide, G. (2008). Molecular evolution of rDNA in early diverging Metazoa: First comparative analysis and phylogenetic application of complete SSU rRNA secondary structures in Porifera. *BMC Evolutionary Biology* **8(69)**

Wainright, P. O., Hinkle, G., Sogin, M. L. and Stickel, S. K. (1993). Monophyletic origins of the metazoa: an evolutionary link with fungi. *Science* **260(5106) :** 340-342.

Wang, X. and Lavrov, D. V. (2007). Mitochondrial genome of the homoscleromorph *Oscarella carmela* (Porifera, Demospongiae) reveals unexpected complexity in the common ancestor of sponges and other animals. *Molecular Biology and Evolution* **24(2) :** 363-373.

Wang, X. and Lavrov, D. V. (2008). Seventeen new complete mtDNA sequences reveal extensive mitochondrial genome evolution within the Demospongiae. *PLoS ONE* **3(7) :** e2723.

Warrior, R. (1987). The mitochondrial DNA of *Hydra attenuata* consists of two linear molecules. **Ph.D. thesis:** Yale University, New Haven, USA.

Watanabe, K. I., Bessho, Y., Kawasaki, M. and Hori, H. (1999). Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal *Dicyema*. *Journal of Molecular Biology* **286(3) :** 645-650.

Westheide, W. and Rieger, R. (1996). Spezielle Zoologie. Teil 1: Einzeller und Wirbellose Tiere. (eds. W. Westheide and R. Rieger), pp. 909. Gustav Fischer Verlag, Stuttgart.

Wheeler, B. M., Heimberg, A. M., Moy, V. N., Sperling, E. A., Holstein, T. W., Heber, S. and Peterson, K. J. (2009). The deep evolution of metazoan microRNAs. *Evolution & Development* **11(1)** : 50-68.

Wheeler, W. C. and Honeycutt, R. L. (1988). Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. *Molecular Biology and Evolution* **5(1) :** 90-96.

Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T. and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* **407**: 327-339.

Woese, C. R., Gutell, R. R., Gupta, R. and Noller, H. F. (1983). Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiological Reviews* **47(4) :** 621-669.

Woese, C. R., Magrum, L. J., Gupta, R., Siegel, R. B., Stahl, D. A., Kop, J., Crawford, N., Brosius, J., Gutell, R. R., Hogan, J. J. and Noller, H. F. (1980). Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Research* **8(10)** : 2275-2293.

Wörheide, G. and Erpenbeck, D. (2007). DNA taxonomy of sponges—progress and perspectives. *Journal of the Marine Biological Association of the UK* **87(06) :** 1629–1633.

Wörheide, G., Nichols, S. A. and Goldberg, J. (2004). Intragenomic variation of the rDNA internal transcribed spacers in sponges (Phylum Porifera): implications for phylogenetic studies. *Molecular Phylogenetics and Evolution* **33(3) :** 816-830.

Wörheide, G. and Hooper, J. N. A. (1999). Calcarea from the Great Barrier Reef. 1: Cryptic Calcinea from Heron Island and Wistari Reef (Capricorn-Bunker Group). *Memoirs of the Queensland Museum* **43(2) :** 859-891.

Wuyts, J., De Rijk, P., Van de Peer, Y., Pison, G., Rousseeuw, P. and De Wachter, R. (2000). Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA. *Nucleic Acids Research* **28(23) :** 4698-4708.

Wuyts, J., Perrière, G. and Van De Peer, Y. (2004). The European ribosomal RNA database. *Nucleic Acids Research* **32(Database issue) :** D101-3.

Wuyts, J., Van de Peer, Y. and De Wachter, R. (2001). Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. *Nucleic Acids Research* **29(24) :** 5017-5028.

Wuyts, J., Van de Peer, Y., Winkelmans, T. and De Wachter, R. (2002). The European database on small subunit ribosomal RNA. *Nucleic Acids Research* **30(1) :** 183-185.

Zrzavy, J., Mihulka, S., Kepka, P., Bezdek, A. and Tietz, D. (1998). Phylogeny of the Metazoa based on morphological and 18S ribosomal DNA evidence. *Cladistics* **14:** 249-285.

Appendix 1

Table A1.1: Taxa and their GenBank Accession

Order / Family / Species	Accession
Spongiidae	
Hippospongia communis	AF246616
Spongia officinalis	AY348888
Thorectidae	
Smenospongia aurea	AY591806
Hadromerida	
Suberitidae	
Suberites domuncula	AJ620112
Suberites ficus	AF100947, AJ627184
Chondrillidae	
Chondrosia reniformis	AY348876
Clionaidae	
Spheciospongia vesparium	AY734440
Tethyidae	
Tethya actinia	AY878079
Halichondrida	
Axinellidae	
Axinella corrugata	AY737637
Axinella damicornis	AY348887
Axinella polypoides	U43190
Dragmacidon lunaecharta	AY734442
Dragmacidon reticulata	AJ705046
Ptilocaulis gracilis	AY737638
Dictyonellidae	
Dictyonella incisa	AY348880
Scopalina ruetzleri	AJ621546
Halichondriidae	
Halichondria melanodocia	AY737639
Spongosorites genitrix	AY348885
Haplosclerida	
Callyspongiidae	
Callyspongia sp. 1	DQ927310
Callyspongia sp. 2	DQ927314
Siphonochalina sp.	DQ927311
Chalinidae	
Chalinula hooperi	DQ927319
Haliclona amphioxa	AJ703887
Haliclona cinerea	DQ927306
Haliclona fascigera	DQ927315
Haliclona mediterranea	AY348879
Haliclona oculata	AY734450, DQ927307
Haliclona sp. 1	AJ703889
Haliclona sp. 2	AY734444
Haliclona sp. 3	DQ927309
Lubomirskiidae	
Baikalospongia bacillifera	DQ176775
Baikalospongia intermedia	AY769090
Lubomirskia baicalensis	DQ176776
Metaniidae	
Corvomeyenia sp.	DQ176774
Niphatidae	
Amphimedon queenslandica	see text.
Aka mucosum	DQ927322
Cribochalina vasculum	DQ927308
Dasychalina fragilis	DQ927316
Niphates sp.	DQ927312
Petrosiidae	
Acanthostrongylophora ingens	DQ927318
Petrosia sp. 1	DQ927321
Petrosia sp. 2	DQ927320
Xestospongia muta	AY621510
Phloeodictyidae	
Calyx podatypa	AY734447
Calyx sp.	DQ927313
Oceanapia sp.	DQ927317
Spongillidae	
Ephydatia cooperensis	AF140354
Ephydatia fluviatilis	AY578146, AJ705048
Ephydatia muelleri	AF121110
Eunapius fragilis	AF121111
Nudospongilla sp.	DQ927323
Spongilla lacustris	AF121112, AJ703890
Trochospongilla horrida	AY609320
Homosclerophorida	
Plakinidae	
Oscarella tuberculata	AY348883
Plakortis simplex	AY348884
Plakortis sp.	AF100948
Lithistida	
Corallistidae	
Corallistes sp.	AY737636
Poecilosclerida	
Microcionidae	
Microciona prolifera	L10825, AJ705047
Raspailiidae	
Eurypon cf. *clavatum*	AJ621547
Mycalidae	
Mycale fibrexilis	AF100946, AJ627185
Mycale sp.	AY737643
Crellidae	
Crella elegans	AY348882
Hymedesmiidae	
Phorbas tenacior	AY348881
Iotrochotidae	
Iotrochota birotulata	AY737641
Tedaniidae	
Tedania ignis	AY737642, AJ704975

Order **Family** Species	Accession
Spirophorida	
Tetillidae	
Cinachyrella apion	AJ627186
Cinachyrella sp. 1	AY734437
Cinachyrella sp. 2	AY734438
Cinachyrella sp. 3	AY734439
Tetilla japonica	D15067
Verongida	
Aplysinidae	
Aiolochroia crassa	AJ621544, AY591805
Aplysina aerophoba	AY591799
Aplysina archeri	AY591801
Aplysina cavernicola	AY348875, AY591800
Aplysina fistularis	AJ621545
Aplysina lacunosa 'hard'	AY591802
Aplysina lacunosa 'soft'	AY591803
Verongula gigantea	AY591804
Ianthellidae	
Hexadella pruvoti	AY348877
Class Hexactinellida	
Amphidiscosida	
Pheronematidae	
Semperella schulzei	AM886410
Hexactinosida	
Aphrocallistidae	
Aphrocallistes vastus	AM886406
Farreidae	
Farrea occa	AF159623
Lyssacinosida	
Leucopsacidae	
Oopsacas minuta	AF207844
Rossellidae	
Acanthascus dawsoni	AF100949

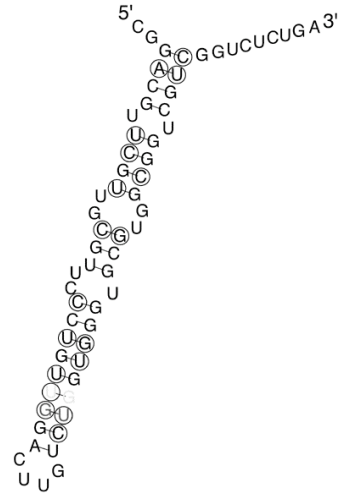
Figure A1.1: : Trace IDs (TI)

from GenBank trace archive used to assemble the *Amphimedon queenslandica* SSU rDNA sequence.

858502160, 858507879, 858263784, 858630376, 858630760, 858634403, 858660883, 858661925, 858662146, 913727848, 913739760, 913740143, 913766300, 913759001, 913759741, 913762063, 913807366, 913800228, 922086177, 913804596, 858272580, 913819001, 913819182, 922090130, 922092432, 858269503, 922087734, 931281954, 922156723, 922157099, 922167280, 922165839, 922162740, 858261628, 922184662, 929287845, 922185952, 922224450, 929291625, 922225360, 922226129, 922220894, 922220222, 922221890, 922219704, 922219800, 922220847, 922220175, 922241123, 930304867, 930305250, 930302865, 929295597, 922249739, 922255990, 922257730, 922254758, 922260071, 922263990, 931287113, 922276733, 922279487, 922286149, 922286303, 922290133, 858123121, 858294358, 929303285, 929303668, 930342673, 930336675, 930336992, 930338891, 930359346, 930353363, 930353459, 930355384, 858302190, 858302573, 930362441, 930365060, 858303862, 858305259, 930376656, 930422817, 930420218, 930429587, 930435315, 931289356, 858317066, 929382778, 929401682, 930469211, 929412935, 858322313, 929425919, 929434563, 929427218, 929442355, 930524117, 930526173, 930523053, 930546576, 929447200, 858327335, 858324348, 858327427, 858324875, 858133606, 858381241, 858382464, 858130301, 858394896, 858134625, 858398326, 858410779, 858131782, 858412106, 858412299, 858419861, 858420391, 858136213, 858419025, 858479948, 858489652, 858490035, 858135544, 858481504, 858486660, 858487178, 858511235, 859670721, 859674658, 859674782, 859673724, 859678208, 859676044, 859703678, 859706579, 859695509, 859700517, 859699887, 859926459, 859927090, 859928836, 859273544, 913830314, 859279581, 859951203, 859955609, 859270625, 922307864, 913849928, 929457535, 929457632, 913850586, 913850790, 913855731, 913903633, 913904160, 922380548, 922393569, 922396932, 922399445, 922402220, 922412890, 922416806, 922421393, 922433517, 859298800, 922438131, 922446932, 922459704, 922450254, 922455818, 859287564, 922462849, 930549436, 922466959, 922472814, 922471416, 930550142, 859286380, 859288617, 922486783, 922487316, 922492482, 922496341, 930551915, 930553976, 930562094, 930558411, 930568097, 930578832, 930571516, 929484319, 929489099, 929491310, 931302728, 931304081, 931304177, 859300224, 930585350, 930592658, 930587151, 930591671, 929503682, 929504882, 933297708, 933298660, 933299040, 933296732, 929519673, 929523708, 931320371, 933303539, 929525539, 931402650, 930617133, 859214500, 859215770, 930654948, 859305811, 930650062, 859322241, 933330978, 859306990, 933335950, 933333574, 931408213, 931409692, 931405510, 933324908, 930677640, 930676250, 930676638, 933328440, 933356711, 931423345, 931426074, 933372225, 931445924, 933372616, 933372634, 933398189, 933391805, 933388729, 933393045, 933393653, 931454200, 933395600, 859315149, 933399994, 859348938, 859360080, 859385623, 859371069, 859382935, 859467116, 859467582, 859490269, 859497089, 859210463, 859620521, 859621469, 859225285, 859622162, 859626487, 866270729, 866270803, 870357343, 881574018, 866275387, 881576580, 890849065, 890844357, 890866217, 890872162, 890868516, 870352092, 870350279, 870355281

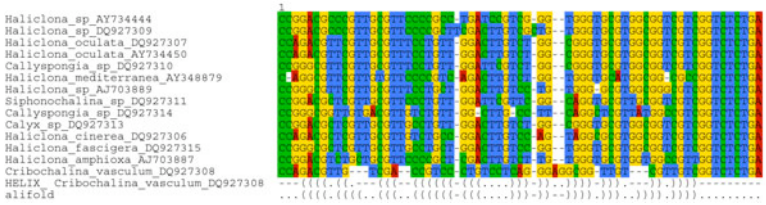
Figure A1.2: Compensatory base changes and alignments

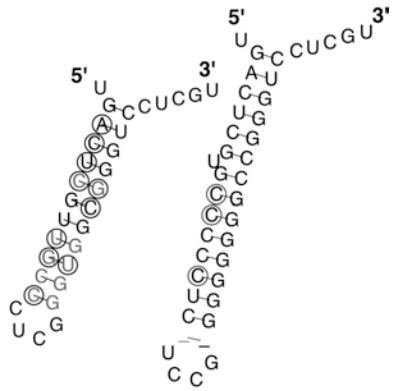
of predicted secondary structures of clade III marine Haplosclerida (Demospongiae) and Hexasterophora (Hexactinellida). In the presented structure plots, positions with base changes supporting the structure are encircled. Positions with mismatches in some sequences are in grey. In the tables, pairs at positions with compensatory base changes are in dark green, semi-compensatory change in light green. Mismatches are shaded in orange. Numbers of occurrence are given after each pair. Only sequences, which allowed an alignment, were included.



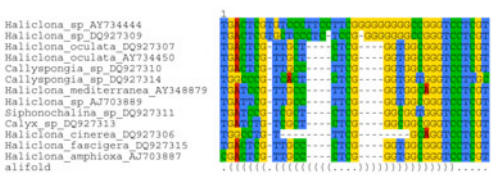
5'	3'							
4	60	G:C 13	G:U 1					
5	59	A:U 9	G:U 4	G:C 1				
6	58	C:G 14						
7	57	G:C 13	G:U 1					
9	55	U:G 11	C:G 3					
10	54	C:G 11	U:G 2	G:C 1				
11	53	G:C 11	G:U 2	-:- 1				
12	52	U:G 11	U:A 1	C:G 1				
15	49	C:G 13	U:A 1					
16	48	G:C 13	G:U 1					
17	47	G:U 12	A:U 1	U:U 1				
20	45	C:G 12	U:G 2					
21	44	C:G 14						
22	43	U:G 6	C:G 4	U:A 3	G:C 1			
23	42	G:U 7	G:C 6	U:G 1				
24	39	C:G 7	U:G 6	U:U 1				
25	38	U:G 7	C:G 3	U:U 3	C:A 1			
27	36	G:C 5	G:U 4	C:G 2	U:G 1	A:U 1	C:U 1	
28	35	G:C 12	G:U 1	U:A 1				
29	34	A:U 12	G:C 1	-:- 1				

Marine Haplosclerida, Helix E23_1a.1: 14 taxa, *Cribochalina* was aligned according to minimum free energy folding predictions (pos.1 is included in E23_1b and not shown in structure)





5'	3'						
2	36	G:C 13					
3	35	A:U 11	G:U 2				
4	34	C:G 9	U:G 4				
5	33	U:G 8	C:G 5				
6	32	C:G 10	U:A 1	U:G 1	C:A 1		
7	31	G:C 11	G:U 2				
9	30	U:G 11	G:C 2				
10	29	U:G 8	C:G 4	-:- 1			
11	28	G:U 7	G:C 2	A:U 1	C:G 1	U:G 1	-:- 1
12	27	C:G 12	-:- 1				
13	26	C:G 7	U:G 5	-:- 1			
14	25	-:- 11	C:G 1	U:G 1			
15	24	-:- 11	U:G 2				
16	23	-:- 11	C:G 2				
17	22	-:- 12	C:G 1				

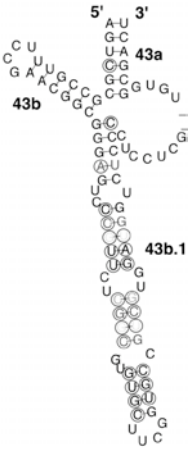


Marine Haplosclerida, Helix E23_14a.1

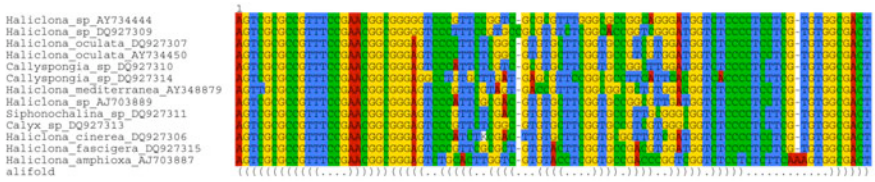
11 taxa (right structure) +2 taxa (left structure). Alignment and compensatory base changes are combined.

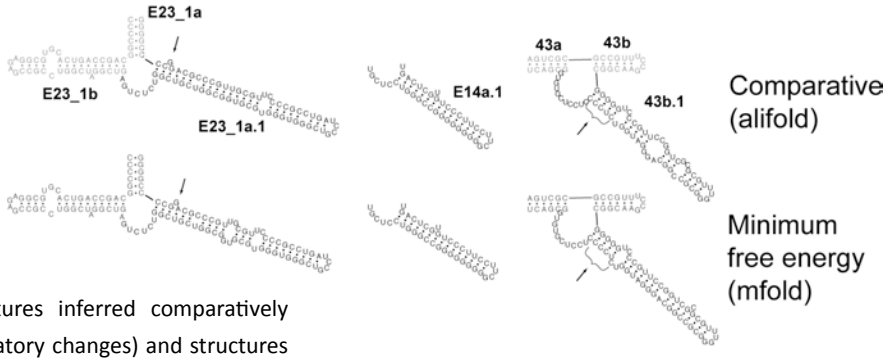
Marine Haplosclerida, Helix 43b.1

13 taxa, parts of conserved helix 43 (a+b) included in structure and alignment. Compensatory base changes are only shown for insertion.



5'	3'									
23	73	G:C 12	G:U 1							
24	72	G:C 13								
25	71	G:C 13								
26	70	A:U 10	G:U 2	A:A 1						
27	69	G:C 13								
30	67	C:G 12	U:G 1							
31	66	C:G 11	U:G 1	G:G 1						
32	65	G:C 4	G:U 2	A:U 3	U:U 1	C:C 1	C:U 2			
33	64	U:A 10	U:G 2	A:U 1						
34	63	U:G 10	C:G 2	G:C 1						
37	60	C:G 7	G:C 2	U:G 1	U:A 1	G:A 1	C:U 1			
38	59	G:C 9	G:U 2	A:U 1	U:U 1					
39	58	G:U 3	A:U 3	U:G 2	G:C 1	U:A 1	A:G 1	C:A 1	G:G 1	
40	57	C:G 10	U:G 2	U:U 1						
44	55	G:C 12	C:G 1							
45	54	U:G 9	C:G 3	U:A 1						
46	53	G:U 7	G:C 4	A:U 2						
47	52	C:G 9	U:G 4							

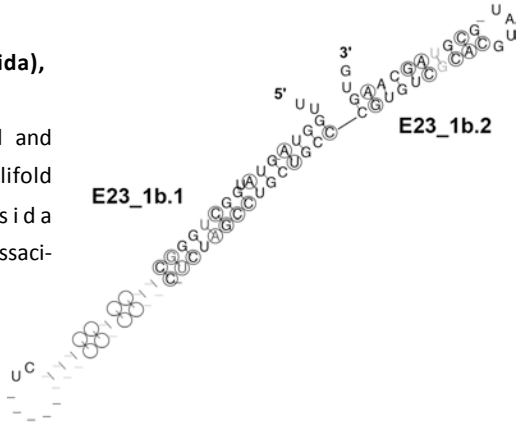




Comparison of structures inferred comparatively (considering compensatory changes) and structures inferred only with minimum free folding predictions for *Haliclona* sp. (AY734444). Arrows show differences.

Hexasterophora (Hexactinellida), Helix E23_1b.1 and E23_1b.2

Initial alignments were refined and combined after separated alifold analyses of Hexactinosida (*Aphrocallistes +Farrea*) and Lyssacinosida (*Acanthascus+Oopsacas*).



5'	3'				
3	62	G:C 3	G:U 1		Helix E23_1b.1
4	61	G:C 4			
5	60	U:G 4			
6	59	A:U 2	G:C 1	G:U 1	
8	58	G:C 4			
9	57	U:G 4			
10	56	A:U 3	G:U 1		
12	55	G:C 2	G:U 2		
13	54	G:C 2	G:U 2		
14	53	C:G 2	U:A 1	U:G 1	
15	52	U:G 2	U:A 1	-:A 1	
16	51	G:U 4			
17	50	G:C 3	G:U 1		
18	49	G:U 2	U:G 2	U:U 1	
19	47	G:C 2	C:G 2		
20	46	U:G 2	-:- 2		
21	45	G:U 2	-:- 2		
22	44	G:C 1	U:A 1	-:- 2	
23	43	G:C 1	U:G 1	-:- 2	
24	42	U:G 2	-:- 2		
25	41	U:A 1	G:U 1	-:- 2	
26	40	C:G 1	G:U 1	-:- 2	
27	39	C:G 1	G:G 1	-:- 2	
28	38	C:G 2	-:- 2		
29	37	U:G 2	-:- 2		
63	87	C:G 4			Helix E23_1b.2
64	86	U:A 2	G:U 2		
65	85	U:A 4			
66	84	G:C 4			
67	83	U:A 1	U:G 3		
68	82	C:G 2	U:A 2		
69	81	G:U 2	R:U 1	G:Y 1	
70	80	C:G 2	U:G 2		
71	79	G:C 2	A:U 2		
72	78	C:G 2	U:G 2		

Table A1.2: SSU rRNA base composition and fragment

of examined rRNA fragment (see Chapter 1, Methods. Fragment corresponds to *Amphimedon . queenslandica* SSU rRNA position 48-1896 (Chapter 1, Fig 1.1).

Sequence (Species as appearing in GenBank + Accession#)	**bp length**	**%GC**	**%A**	**%T**	**%G**	**%C**	**Class or group**
Anamixilla_sp_AF182192	1724	46.40	26.22	27.61	26.57	19.84	Calcarea
Anamixilla_torresi_AF452020	1724	46.17	26.22	27.61	26.45	19.72	Calcarea
Aphroceras_sp_AM180971	1728	45.95	26.39	27.66	26.39	19.56	Calcarea
Clathrina_adusta_AM180962	1721	46.31	26.32	27.37	26.50	19.81	Calcarea
Clathrina_aff_cerebrum_AM180957	1720	45.64	26.51	27.85	26.34	19.30	Calcarea
Clathrina_cerebrum_U42452	1719	45.61	26.47	27.92	26.29	19.31	Calcarea
Clathrina_helveola_AM180958	1720	46.16	26.34	27.50	26.45	19.71	Calcarea
Clathrina_luteoculcitella_AM180959	1720	45.99	26.45	27.56	26.51	19.48	Calcarea
Clathrina_sp_AM180960	1721	46.60	26.21	27.19	26.50	20.10	Calcarea
Clathrina_wistariensis_AM180961	1720	46.16	26.34	27.50	26.45	19.71	Calcarea
Grantiopsis_heroni_AM180978	1728	46.01	26.22	27.78	26.50	19.50	Calcarea
Grantiopsis_sp_AM180977	1726	46.06	26.25	27.81	26.54	19.52	Calcarea
Grantiopsis_sp_AF182193	1726	46.23	26.13	27.75	26.59	19.64	Calcarea
Grantiopsis_sp_AF452019	1726	46.12	26.13	27.75	26.54	19.58	Calcarea
Guancha_sp_AM180963	1720	45.70	26.45	27.85	26.45	19.24	Calcarea
Lelapiella_incrustans_AM180969	1720	46.16	26.22	27.62	26.34	19.83	Calcarea
Leucaltis_clathria_AF452016	1718	45.63	26.31	28.06	26.37	19.27	Calcarea
Leucandra_nicolae_AM180974	1728	45.54	26.50	27.95	26.33	19.21	Calcarea
Leucetta_chagosensis_AF182190	1720	45.87	26.22	27.91	26.45	19.42	Calcarea
Leucetta_microraphis_AM180965	1720	45.87	26.34	27.79	26.34	19.53	Calcarea
Leucetta_sp_AM180964	1720	45.93	26.28	27.79	26.40	19.53	Calcarea
Leucetta_villosa_AM180966	1720	45.93	26.28	27.79	26.40	19.53	Calcarea
Leuconia_nivea_AF182191	1726	46.23	26.59	27.29	26.54	19.70	Calcarea
Leucosolenia_sp_AF100945	1728	45.54	26.50	27.95	26.27	19.27	Calcarea
Leucosolenia_sp_AJ622898	1728	45.54	26.50	27.95	26.27	19.27	Calcarea
Levinella_prolifera_AM180956	1719	45.96	26.18	27.87	26.47	19.49	Calcarea
Murrayona_phanolepis_AM180968	1719	46.07	26.53	27.40	26.24	19.84	Calcarea
Pericharax_heteroraphis_AM180967	1720	45.81	26.34	27.85	26.28	19.53	Calcarea
Petrobiona_massiliana_AF452026	1724	46.23	26.57	27.20	26.62	19.61	Calcarea
Plectroninia_neocaledoniense_AM180979	1731	46.36	26.08	27.61	26.55	19.82	Calcarea
Soleneiscus_radovani_AF452017	1725	46.78	26.03	27.19	26.72	20.06	Calcarea
Soleneiscus_stolonifer_AM180955	1722	46.72	26.19	27.15	26.77	19.95	Calcarea
Sycettusa_sp_AF452025	1725	45.68	26.72	27.59	26.32	19.36	Calcarea
Sycettusa_tenuis_AM180975	1728	45.83	26.62	27.55	26.39	19.44	Calcarea
Sycon_capricorn_AM180970	1728	45.89	26.50	27.60	26.39	19.50	Calcarea
Sycon_ciliatum_AJ627187	1729	45.81	26.72	27.47	26.26	19.55	Calcarea
Sycon_ciliatum_L10827	1724	45.94	26.77	27.52	26.36	19.58	Calcarea
Syconessa_panicula_AM180976	1728	45.95	26.62	27.43	26.39	19.56	Calcarea
Ute_ampullacea_AM180972	1728	45.95	26.27	27.78	26.56	19.39	Calcarea
Vosmaeropsis_sp_AF452018	1727	45.69	26.69	27.62	26.35	19.34	Calcarea
Acanthostrongylophora_ingens_DQ927318	1733	52.25	24.15	23.66	29.75	22.50	marine haplosclerid
Aka_mucosum_DQ927322	1743	51.75	23.47	24.78	29.49	22.26	marine haplosclerid
Amphimedon_queenslandica	1849	56.30	21.20	22.50	31.10	25.20	marine haplosclerid
Callyspongia_sp_DQ927310	1906	52.62	21.93	25.45	29.96	22.67	marine haplosclerid
Callyspongia_sp_DQ927314	1905	51.81	22.52	25.67	29.29	22.52	marine haplosclerid
Calyx_sp_DQ927313	1910	52.88	21.99	25.13	30.05	22.83	marine haplosclerid
Chalinula_hooperi_DQ927319	1734	51.85	24.22	23.93	29.58	22.26	marine haplosclerid

Sequence (Species as appearing in GenBank + Accession#)	**bp length**	**%GC**	**%A**	**%T**	**%G**	**%C**	**Class or group**
Dasychalina_fragilis_DQ927316	1843	54.80	22.03	23.17	31.20	23.60	marine haplosclerid
Haliclona_amphioxa_AJ703887	1907	53.22	21.87	24.91	29.89	23.34	marine haplosclerid
Haliclona_cinerea_DQ927306	1897	52.48	22.56	25.01	29.39	23.09	marine haplosclerid
Haliclona_fascigera_DQ927315	1907	52.60	22.08	25.33	29.89	22.71	marine haplosclerid
Haliclona_oculata_AY734450	1906	52.26	22.19	25.55	29.64	22.61	marine haplosclerid
Haliclona_oculata_DQ927307	1905	52.28	22.15	25.56	29.66	22.62	marine haplosclerid
Haliclona_sp_AJ703889	1906	52.47	22.09	25.45	29.96	22.51	marine haplosclerid
Haliclona_sp_AY734444	1918	53.75	21.85	24.40	30.19	23.57	marine haplosclerid
Niphates_sp_DQ927312	1733	51.76	24.24	24.00	29.49	22.27	marine haplosclerid
Petrosia_sp_DQ927320	1737	52.04	24.24	23.72	29.48	22.57	marine haplosclerid
Petrosia_sp_DQ927321	1734	52.05	24.31	23.88	29.44	22.61	marine haplosclerid
Siphonochalina_sp_DQ927311	1910	52.72	22.04	25.24	29.84	22.88	marine haplosclerid
Xestospongia_muta_AY621510	1907	55.01	21.05	24.04	31.15	23.86	marine haplosclerid
Agelas_clathrodes_AY769087	1732	48.61	25.75	25.64	27.42	21.19	other Demospongiae
Agelas_conifera_AY734443	1733	48.70	25.74	25.56	27.52	21.18	other Demospongiae
Agelas_dispar_AY737640	1731	48.61	25.79	25.65	27.41	21.20	other Demospongiae
Aiolochroia_crassa_AJ621544	1720	47.91	25.58	26.51	27.27	20.64	other Demospongiae
Aiolochroia_crassa_AY591805	1711	48.13	25.75	26.48	27.41	20.72	other Demospongiae
Aplysilla_sulfurea_AF246618	1719	47.70	26.38	26.03	27.08	20.62	other Demospongiae
Aplysina_aerophob_AY591799	1713	48.30	25.72	26.33	27.27	21.03	other Demospongiae
Aplysina_archeri_AY591801	1718	48.34	25.61	26.11	27.36	20.98	other Demospongiae
Aplysina_cavernicola_AY591800	1716	48.31	25.61	26.31	27.21	21.10	other Demospongiae
Aplysina_fistularis_AJ621545	1719	48.40	25.54	26.06	27.23	21.18	other Demospongiae
Aplysina_lacunosa_AY591802	1718	48.37	25.55	26.19	27.27	21.10	other Demospongiae
Aplysina_lacunosa_AY591803	1718	48.54	25.44	26.14	27.36	21.19	other Demospongiae
Axinella_corrugata_AY737637	1732	48.90	25.52	25.58	27.71	21.19	other Demospongiae
Axinella_polypoides_U43190	1736	49.77	24.83	25.40	28.51	21.26	other Demospongiae
Baikalospongia_bacillifera_DQ176775	1728	48.84	25.93	25.23	28.07	20.78	other Demospongiae
Baikalospongia_intermedia_AY769090	1728	48.84	25.93	25.23	28.07	20.78	other Demospongiae
Cinachyrella_apion_AJ627186	1735	51.93	24.38	23.69	29.16	22.77	other Demospongiae
Cinachyrella_sp_AY734437	1735	51.87	24.32	23.80	29.11	22.77	other Demospongiae
Cinachyrella_sp_AY734438	1735	52.05	24.32	23.63	29.16	22.88	other Demospongiae
Cinachyrella_sp_AY734439	1735	51.87	24.50	23.63	29.22	22.65	other Demospongiae
Clypeatula_cooperensis_AF140354	1728	48.90	25.98	25.12	28.13	20.78	other Demospongiae
Corallistes_sp_AY737636	1733	51.13	24.41	24.47	28.97	22.16	other Demospongiae
Corvomeyenia_sp_DQ176774	1728	49.54	25.75	24.71	28.41	21.12	other Demospongiae
Dysidea_avara_AF456326	1719	47.82	25.83	26.59	27.34	20.48	other Demospongiae
Dysidea_sp_AY734449	1724	47.56	25.93	26.51	27.15	20.42	other Demospongiae
Ephydatia_fluviatilis_AJ705048	1728	48.90	25.98	25.12	28.13	20.78	other Demospongiae
Ephydatia_fluviatilis_AY578146	1728	48.90	25.98	25.12	28.13	20.78	other Demospongiae
Ephydatia_muelleri_AF121110	1728	48.96	25.93	25.12	28.07	20.89	other Demospongiae
Eunapius_fragilis_AF121111	1728	48.78	25.93	25.29	28.30	20.49	other Demospongiae
Eurypon_cf_clavatum_AJ621547	1732	48.79	25.58	25.64	27.71	21.07	other Demospongiae
Geodia_neptuni_AY737635	1735	51.12	24.67	24.21	28.82	22.31	other Demospongiae
Halichondria_melanodocia_AY737639	1729	47.08	26.26	26.66	26.72	20.36	other Demospongiae
Hippospongia_communis_AF246616	1723	47.74	26.07	26.25	27.06	20.68	other Demospongiae
Iotrochota_birotulata_AY737641	1733	49.16	25.16	25.68	27.76	21.41	other Demospongiae
Ircinia_felix_AJ703888	1723	47.77	26.18	26.06	27.05	20.72	other Demospongiae
Ircinia_felix_AY734448	1724	47.68	26.10	26.22	26.97	20.71	other Demospongiae
Microciona_prolifera_AJ705047	1731	49.28	25.42	25.30	27.67	21.61	other Demospongiae

Sequence (Species as appearing in GenBank + Accession#)	**bp length**	**%GC**	**%A**	**%T**	**%G**	**%C**	**Class or group**
Microciona_prolifera_L10825	1729	49.33	25.42	25.36	27.79	21.54	other Demospongiae
Mycale_fibrexilis_AF100946	1738	49.08	25.09	25.83	27.79	21.29	other Demospongiae
Mycale_fibrexilis_AJ627185	1739	49.05	25.07	25.88	27.77	21.28	other Demospongiae
Mycale_sp_AY737643	1739	50.43	24.90	24.67	28.35	22.08	other Demospongiae
Nudospongilla_sp_DQ927323	1728	48.84	25.98	25.17	28.24	20.60	other Demospongiae
Plakortis_sp_AF100948	1730	47.86	26.01	26.13	26.88	20.98	other Demospongiae
Pleraplysilla_spinifera_AF246617	1719	47.99	25.92	26.21	27.20	20.80	other Demospongiae
Pseudaxinella_lunaecharta_AY734442	1733	49.91	24.75	25.33	28.33	21.58	other Demospongiae
Pseudaxinella_reticulata_AJ705046	1733	49.91	24.75	25.33	28.33	21.58	other Demospongiae
Ptilocaulis_gracilis_AY737638	1736	50.86	24.60	24.54	28.86	22.00	other Demospongiae
Scopalina_ruetzleri_AJ621546	1724	48.49	26.10	25.41	27.44	21.06	other Demospongiae
Smenospongia_aurea_AY591806	1714	48.66	25.55	26.14	27.48	21.18	other Demospongiae
Spheciospongia_vesparium_AY734440	1729	48.47	25.97	25.56	27.18	21.28	other Demospongiae
Spongilla_lacustris_AF121112	1728	48.84	25.98	25.17	28.24	20.60	other Demospongiae
Spongilla_lacustris_AJ703890	1728	48.78	25.98	25.23	28.24	20.54	other Demospongiae
Suberites_ficus_AF100947	1724	47.21	26.44	26.47	26.68	20.52	other Demospongiae
Suberites_ficus_AJ627184	1727	47.13	26.46	26.40	26.69	20.44	other Demospongiae
Tedania_ignis_AJ704975	1733	49.34	25.10	25.56	27.81	21.52	other Demospongiae
Tedania_ignis_AY737642	1734	49.25	25.32	25.43	27.74	21.51	other Demospongiae
Trochospongilla_horrida_AY609320	1728	49.13	25.81	25.06	28.30	20.83	other Demospongiae
Verongula_gigantea_AY591804	1709	48.85	25.62	26.00	27.58	21.27	other Demospongiae
Aphrocallistes_vastus_AM886406	1889	47.14	24.67	28.24	27.53	19.61	Hexactinellida
Farrea_occa_AF159623	1758	46.36	24.57	29.07	27.42	18.94	Hexactinellida
Oopsacas_minuta_AF207844	1906	47.95	23.77	28.28	28.33	19.62	Hexactinellida
Rhabdocalyptus_dawsoni_AF100949	1904	47.79	23.71	28.65	28.20	19.59	Hexactinellida
Semperella_schulzei_AM886410	1904	50.00	23.84	26.16	28.26	21.74	Hexactinellida

Appendix 2

Table A2:1: Primers used for PCR and sequencing

	Sequence (5'-3')	Used for
28S		
5.8SF*	GGATCACTCGGCTCRTGNRTCGATGAAG	PCR
F673mod*	ACCCGCTGAAYTTAAGCATATHANTMAG	PCR
28S-C2	GAAAAGAACTTTGRARAGAGAGT	Sequencing
NL4F**	GACCCGAAAGATGGTGAACTA	PCR/Sequencing
28S-1350rv	CATCGCCAGTTCTGCTTAC	Sequencing
R1630*	CCYTTCYCCWCTCRGYCTTC	Sequencing
28S-1810fw	CGAAAGGGAATCGGGTTAATATTCC	Sequencing
NL4R**	ACCTTGGAGACCTGATGCG	PCR/Sequencing
28Samp_rev*[1]	ACCTGTCTCACGACGKTCTRAACCCAGCTC	PCR
18S		
18S1***	AACCTGGTTGATCCTGCCA	PCR/Sequencing
18S2***	TGCAGGTTCACCTACAGAA	PCR/Sequencing
18S-AF****	CTGGTTGATCCTGCCAG	Sequencing
18S-BR****	CTGCAGGTTCACCTAC	Sequencing
18S-5F****	GTGCCAGCAGCCGCGG	Sequencing
18S-10F****	GGTGGTGCATGGCCG	Sequencing
18S-10R****	CGGCCATGCACCACC	Sequencing
18S-5R****	GAATTACCGCGGCTGCTG	Sequencing

* **Medina, M.,et al.** (2001). Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proceedings of the National Academy of Sciences* **98(17) :** 9707-9712.)

*[1] reverse complement of primer 28Samp from Medina et al. (2001)

** **Nichols, S. A.** (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(1)** : 81-96.

*** **Manuel, M., et al.** (2003). Phylogeny and evolution of calcareous sponges: Monophyly of Calcinea and Calcaronea, high level of morphological homoplasy, and the primitive nature of axial symmetry. *Systematic Biology* **52(3) :** 311-333

**** **Collins, A.G.** (2002). Phylogeny of Medusozoa and the evolution of cnidarian life cycles. *Journal of evolutionary biology* **15**, 418–432.

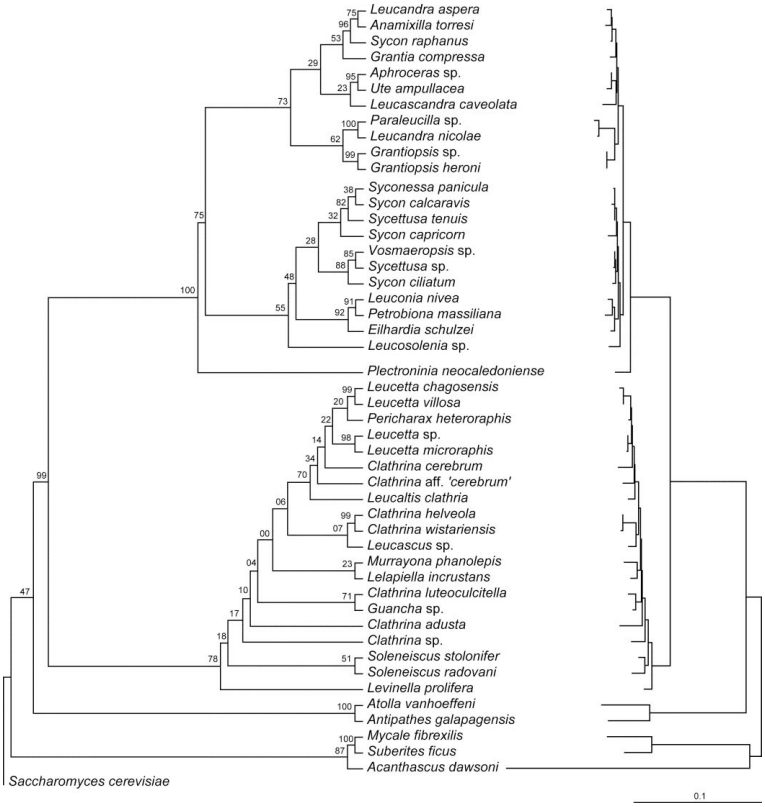


Figure A2.1: Maximum likelihood tree inferred from the 18S rDNA alignment

under the TrN+I+G model. Bootstrap proportions (%) are given above branches. The tree was rooted in TreeView (Page, R.D.M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.) with S. cerivisiae as the outgroup. Phylogram is shown on the right (scale bar, number of expected substitutions per site).

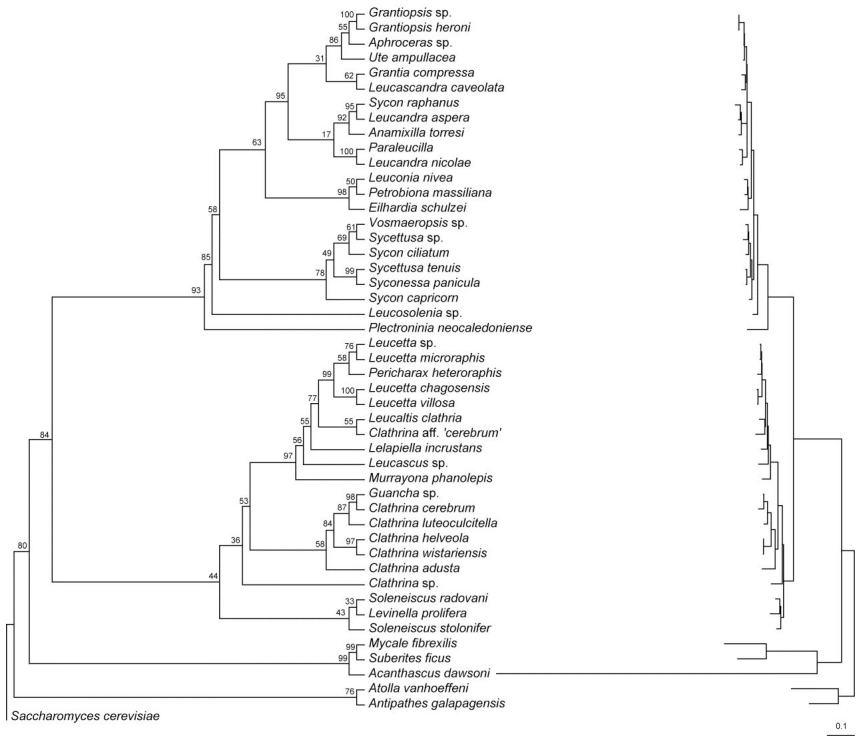


Figure A2.2: Maximum likelihood tree inferred from the 28S rDNA alignment

under the TrN+I+G model. Bootstrap proportions (%) are given above branches. The tree was rooted in TreeView (Page, R.D.M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.) with S. cerivisiae as the outgroup. Phylogram is shown on the right (scale bar, number of expected substitutions per site).

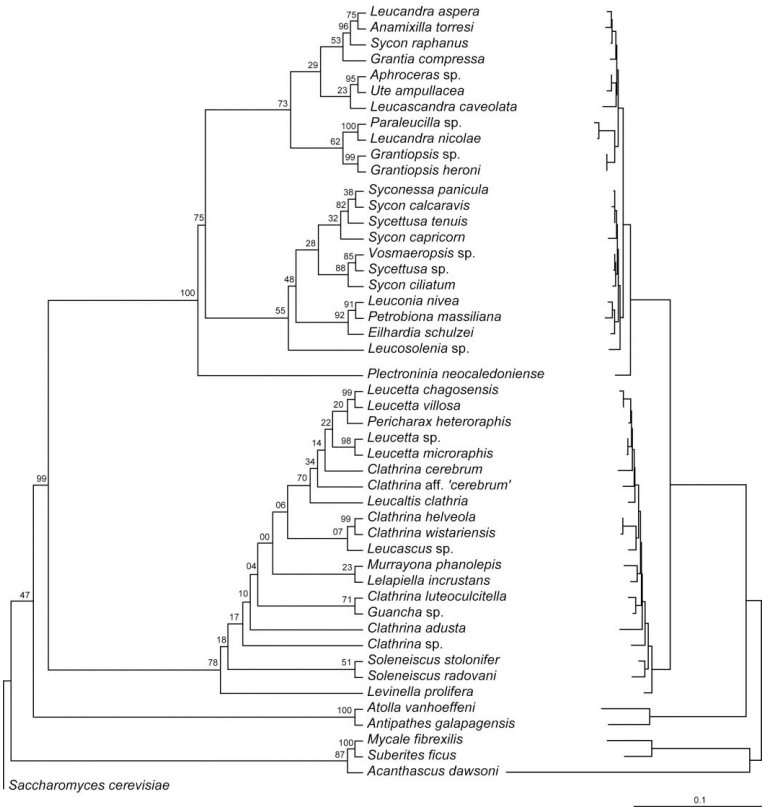


Figure A2.3: Maximum likelihood tree inferred from the combined 18S/28S rDNA alignment

under the TrN+I+G model. Bootstrap proportions (%) are given above branches. The tree was rooted in TreeView (Page, R.D.M., 1996. TREE-VIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.) with S. cerivisiae as the outgroup. Phylogram is shown on the right (scale bar, number of expected substitutions per site).

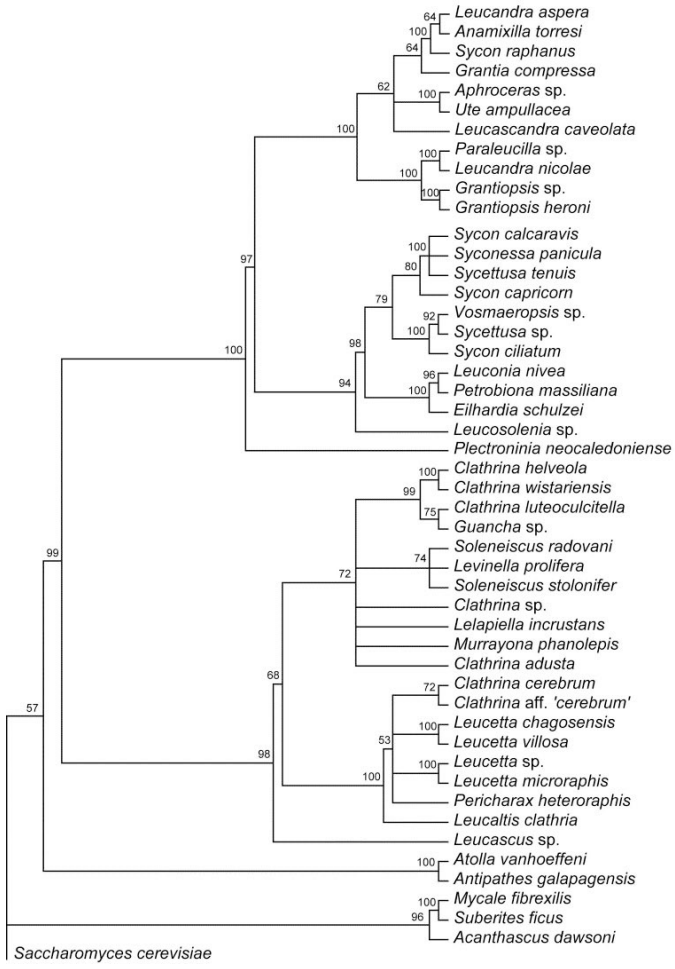


Figure A2.4: Bayesian 50% majority rule consensus tree inferred from the 18S rDNA alignment

under the unpartitioned GTR+I+G model. Bayesian posterior probabilities (%) are given above branches.

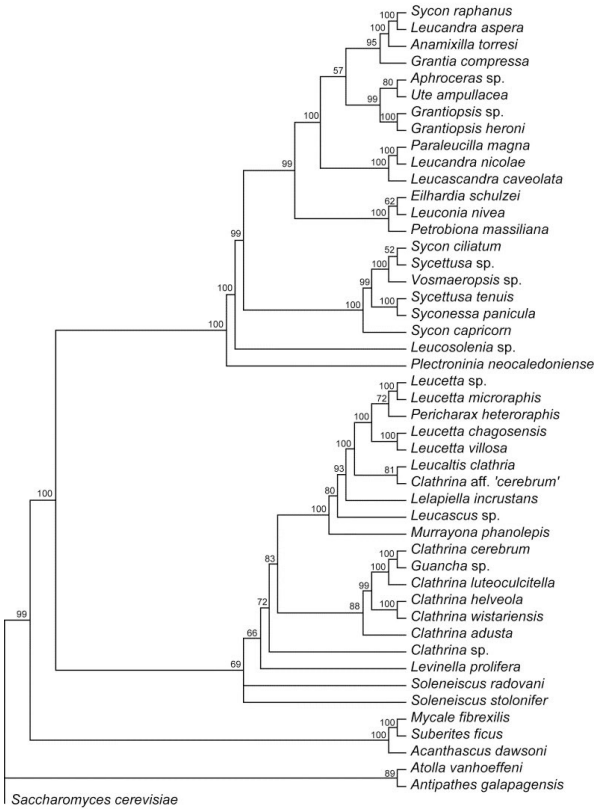


Figure A2.5: Bayesian 50% majority rule consensus tree inferred from the 28S rDNA alignment

under the unpartitioned GTR+I+G model. Bayesian posterior probabilities (%) are given above branches.

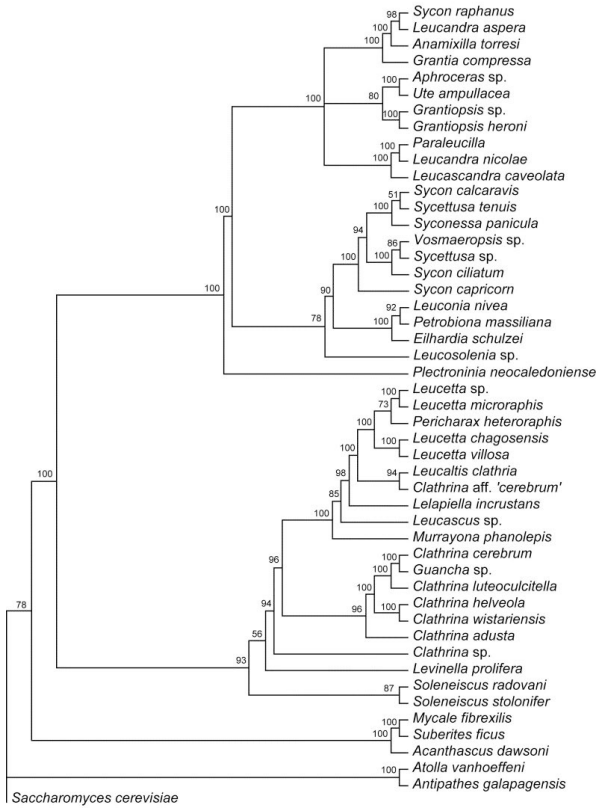


Figure A2.6: Bayesian 50% majority rule consensus tree inferred from the combined 18S/28S rDNA alignment

under the unpartitioned GTR+I+G model. Bayesian posterior probabilities (%) are given above branches.

Appendix 3

Supplementary information 3.1: Specimen descriptions.

In this section we document the calcarean specimen from our study. Determined by OV if not mentioned otherwise. The following abbreviations for collections are used:

GW: Collection Gert Wörheide, LM University Munich, Germany
QM: Queensland Museum, Brisbane, Australia
SAM: South Australian Museum, Adelaide, Australia
WAM: Western Australian Museum, Perth, Australia
ZMA: Zoölogisch Museum, Universiteit van Amsterdam, Amsterdam, Netherlands

Calcaronea

Order: Leucosolenida Hartman, 1958

Family Sycettidae Dendy1892

Sycon Risso

***Sycon (*cf.*) carteri* Dendy 1893**

Specimen: SAM PS 0143

Locality: Australia, Ulladulla

Sponge with many repeatedly branching tubes with a terminal naked osculum (Suppl Fig. 3.1.1, A). The tubes have a syconoid organization (Chapter 3, Fig. 3.8, C), with very short radial tubes that are crowned by few diactines (Chapter 3, Fig. 3.8, A). The triactines of the radial tubes are mostly restricted to the distal cone, leaving the choanocyte chambers mainly supported by the subatrial triactines, whose unpaired actines reaches through the sponge wall until the beginning of the short distal cone. The atrial skeleton is build from large triactines with slender actines and tetractines of the same form but with short apical actines.

Family: Grantiidae Dendy, 1892

Synute Dendy 1892

***Synute pulchella* Dendy 1892**

Specimen: WAM Z1404

Locality: West Australia, Reru Island, Houtman Abrolhos

Dendy described the growth of the *Synute* holotype as follows:

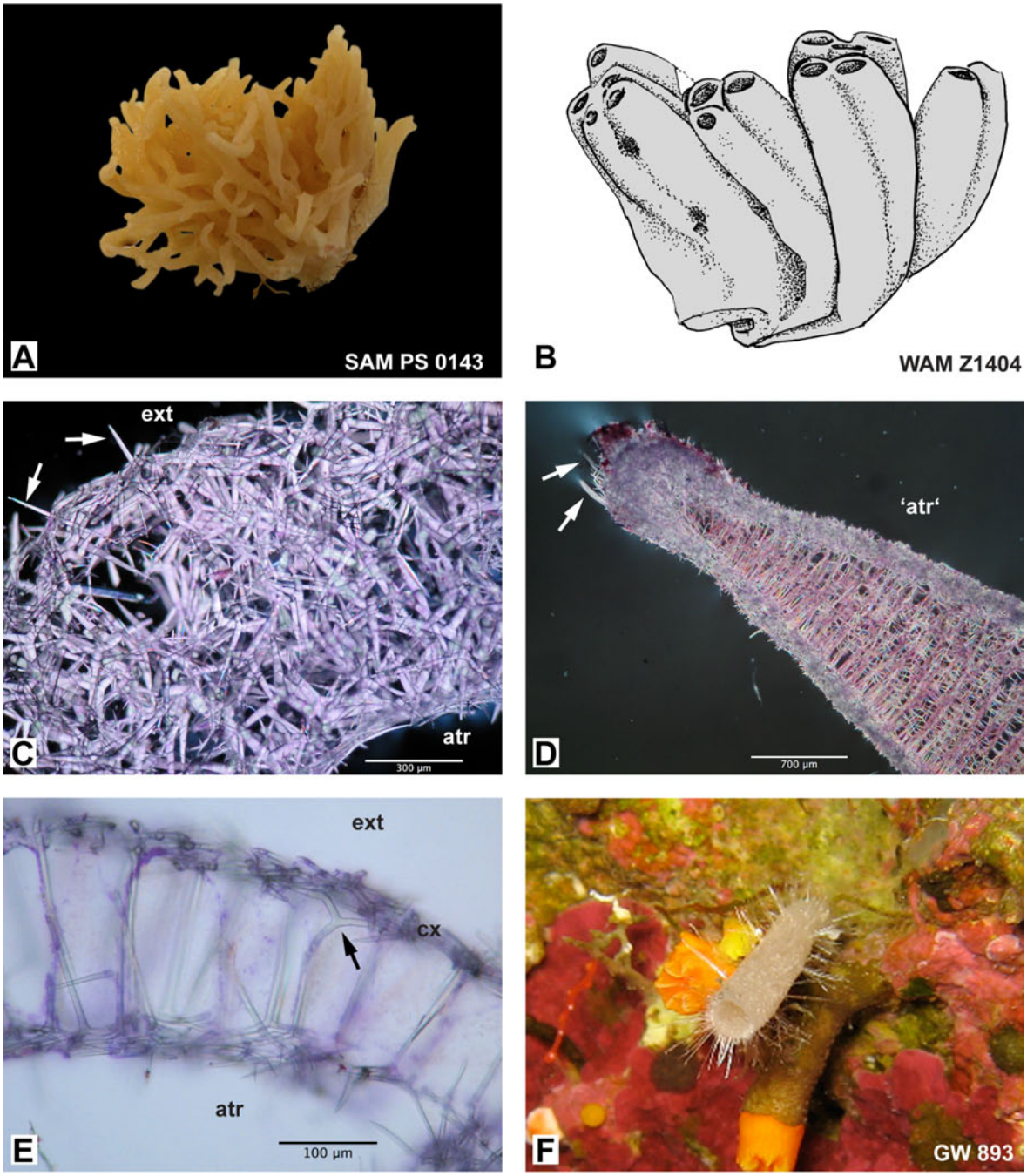


Figure A3.1.1: Specimen of Calcarea.

A: Habitus of *Sycon* cf. *carteri*. The skeletal arrangement and aquiferous system is shown in Chapter 3, Fig 3.8; **B:** Habitus of *Synute pulchella*. For other features see Chapter 3, Fig. 3.6 and Fig. A3.1.2; **C:** Skeletal arrangement of *Leucandra* sp., arrows pint to the large diactines that perpendicular to the surface of the sponge; **D:** Transverse section of *Teichonopsis labyrinthica*. Due to the special growth form of this species, the edge corresponds to the osculum and is supported by large diactines (arrows). The upper side ('atr') corresponds to the gastral surface, the lower side to the external surface. **E:** Transverse section of *Sycettusa* cf. *simplex*, the arrow points to the unpaired angle of an subcortical pseudosagittal spicules.**F:** Habitus of *Sycettusa* aff. *hastifera*; the sponge is about 1.5 cm long and about 3 mm in diameter. Abbreviations: atr=atrium; cx=cortex; ext=exterior of the sponge.

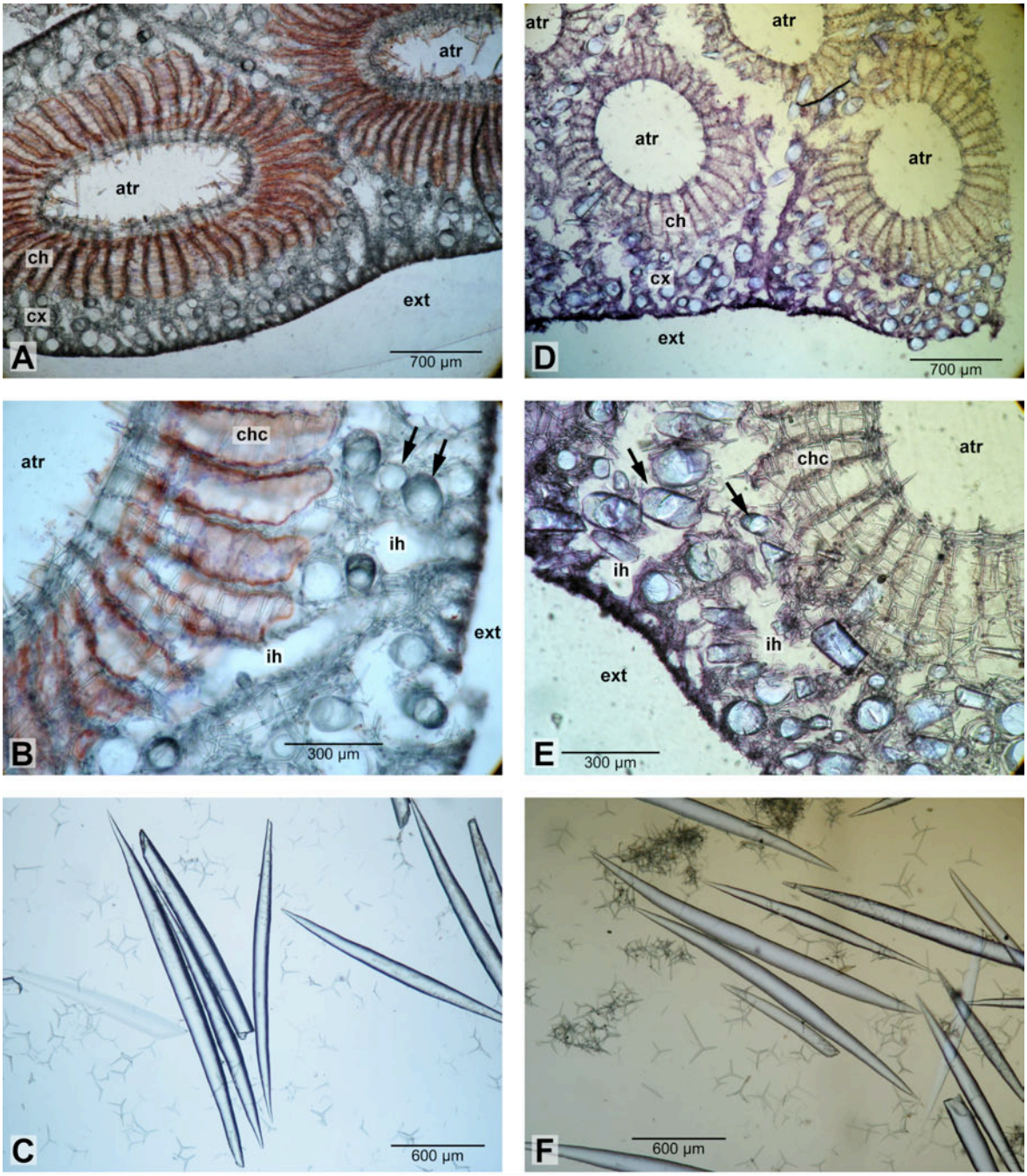


Figure A3.1.2: ***Synute pulchella***

WAM Z1404 (**A,B,C**) in comparison to the holotype (**D,E,F:** Slides from British Museum for Natural History, 25.11.1.1680 sections and spicules from the Dendy collection). **A,B, D, E**: transverse sections, arrows point to giant longitudinal diactines in **B** and **D**. **C, F**:spicule preparations. Abbreviations: Abbreviations: atr=atrium; cx=cortex; ext=exterior of the sponge; ih= inhalant canals.

'If we imagine a colony of the Sycon[3] genus *Ute*, whose component members, growing more or less vertically upwards side by side, have become fused together completely, so that the whole colony forms a single vallate mass in which the individuals can only be recognised externally by their oscula, we have then a tolerably accurate conception of the new genus *Synute*. The fusion of the

[3] Sycon is used in the sense of 'syconoid' here

Sycon individuals of which the colony is composed is complete (extending right up to the oscula) and universal, and by no means partial or accidental, and the entire colony is protected on the outside by a thick common cortex consisting mainly of huge oxeote spicules.' (Dendy, 1892 p. 1).

The observed specimen differs from this description of *Synute pulchella* in that the fusion of syconoid units is not universal and the colony does not form one mass with a meandriniform surface. Instead it is composed of modules of two to four syconoid units, which are fused in their entire length and covered by a common cortex. Nonetheless, the contour of the syconoid units is recognizable by smooth impressions in the cortex (Fig. A3.1.1, B). In the transverse section, the atrial cavities appear elongated in contrast to the almost circular atrial cavities found in the holotype (Fig. A3.1.2, A, D). The exhalant opening of choanocytes chambers of our specimen of *S. pulchella* are not reaching the atrial skeleton, as they do in the holotype. Spiculation and organization otherwise do not seem to significantly differ (Fig. A3.1.2). Without more specimen of this monotypic genus it is not possible to decide whether the mentioned differences are due to plasticity in *Synute pulchella*, and represent different growth form of the same species, or if WAM- Z1404 might belong to a new *Synute* species. For now, we determined this specimen as *Synute pulchella*.

***Ute* Schmidt, 1862**

Ute* aff. *syconoides

Specimens: GW 975, QM G313694, QM G323233

Localities (in the order of specimen): GBR, Australia, Lizard Island; GBR, Yonge Reef; Tasmania, King Island Canyons

Individual tubular sponges. Giant longitudinal diactines (up to four in a row) are present and support the cortex (Chapter 3, Fig. 3.6, A,B). The articulated choanoskeleton is made up by numerous rows of sagittal triactines. The atrial skeleton contains tetractines with the apical ray protruding to the atrium. Subatrial triactines may have a very long unpaired actine (more than four times the size of the paired actines), which is pointing to the outside of the sponge. The aquiferous system is syconoid. Only the individuals of GW 975 and QM G313694 are complete individuals and possess osculae with a fringe consisting of diactines. We found our specimen to closely resemble *Ute syconoides* (described as *Aphroceras syconoides* on p.135 in Carter, 1886; see also Plate 11, Figs 12 and 13 in Dendy, 1893), except that *Ute syconoides* has a naked osculum. Therefore we refer to our specimen as *Ute* aff. *syconoides*.

***Aphroceras* Gray, 1858**

***Aphroceras* sp.**

Specimen: SAM PS 0349

Locality: Tasmanian Peninsula, Waterfall Bay

Sponge formed of cylindrical tubes united at their base. The cortex and the thick artrial skeleton

are supported by giant diactines (Chapter 3, Fig. 3.6. D). The choanosomal skeleton comprises sagittal triactines, the unpaired actine generally pointing to the outside of the sponge and supporting the leuconoid aquiferous system. The atrium is narrow. Large exhalant channels open into the atrium.

Grantiidae with giant longitudinal diactines in the cortical and the artrial skeleton are allocated to the genus *Amphiute,* but the diagnosis of the species includes a syconoid aquiferous system. Therefore we allocated this specimen to genus *Aphroceras*. *Aphroceras* is characterized by giant longitudinal diactines in the cortex and a leuconoid aquiferous system; therefore the presence of the (sub-) atrial longitudinal diactines is not excluded in the genus diagnosis.

***Leucandra* Haeckel, 1872**

***Leucandra* sp.**

Specimen: QM G316285

Locality: Pacific, Coral Sea, Osprey Reef

This specimen was referred to as *Aphroceras* sp. in a previous study (Dohrmann et al., 2006). A closer examination revealed that this specimen belongs to the genus *Leucandra*. A layer of triactines forms the cortex. Larger diactines are present, but do not lay longitudinal to support the cortex as typical for *Aphroceras*. Instead they are radially arranged and protrude the surface of the sponge (Fig. A3.1. 1, C). This is characteristic for Grantiidae of the genus *Leucandra*.

***Teichonopsis* Dendy & Row, 1913**

***Teichonopsis labyrinthica* (Carter 1878)**

Specimen: SAM PS 0228

Locality: Australia, Kangaroo Island

The specimen shows the typical pedunculate calyciform growth form of this species, with a folded wall. The atrium is greatly expanded, comprising the inner side of the cormus (Fig. A3.1.1, D). The syconoid choanosome is supported by an articulated skeleton forming radial tubes with triactine spicules. The edge of the folded wall corresponds to the oscular margin. It contains large diactines with a sharp pointed end pointing inwards and a blunt end pointing to or protruding the edge (Fig. A3.1.1, D, arrows).

Family: Heteropiidae Dendy, 1892

Sycettusa Haeckel, 1872

Sycettusa* cf. *simplex

Specimen: ZMA POR-11566

Locality: Seychelles, near Amirantes

Appears in the Amsterdam collection under its synonym *Grantessa zanzibarensis*. *Grantessa* sensu Borojevic et al. (2000) is characterized by an articulated choanoskeleton, which is missing in this species, see Fig. A3.1.1, E). Determined by R.W.M. van Soest.

Sycettusa* aff. *hastifera

Specimen: GW893

Locality: Red Sea, Gulf of Aqaba

The sponge is a single cylindrical tube of ca. 1,5 cm length and 3 mm in diameter. Large diactines protrude the surface in bundles, some longer than the diameter of the tube (Fig. A3.1.1, F). The bundles of large diactines reach at least half through the sponge wall. The outer tip of the large diactines are lanceolate, and mostly broken off from our specimen. An oscular fringe is also formed by long diactines. The cortex consists of triactines. The choanocyte chambers of the syconoid aquiferous system are supported by the longer of the paired actines of the pseudosagittal subcortical triactines and the unpaired actines of the subatrial triactines (Chapter 3, Fig. 3.2, A). The organization resembles closely to that of *Sycettusa hastifera*. We had the opportunity to compare our specimen to two specimens from the Museum in Amsterdam (ZMA POR-13421 and ZMA POR-13429) that were identified as *Sycettusa hastifera.* Both these specimen were also sampled in the Red Sea, and indeed resemble our specimen. However, these and our specimen differ from the description of *Sycettusa hastifera*. While the examined specimens comprise a single tube, *S. hastifera* is branching and has an overall different appearance (see Row, 1909). Furthermore, the large diactines are much shorter in the original descriptions, and do not project the surface nearly as far as in the examined specimen. Additionally, according to a drawing of the skeletal arrangement, the diactines occur as single spicules and in regular intervals. In our specimens, the spicules are distributed in a more patchy fashion, forming bundles. Therefore, while the form of the diactines suggest an affinity of our species to *S. hastifera*, we find that this and the Museum specimen probably belong to another, closely related species.

***Grantessa* Lendenfeld, 1885**

***Grantessa* sp.**

Specimen: GW 974, GW979

Locality: GBR, Lizard Island

Individual tubes, sometimes two connected at their base (Fig. A3.1.3, A). Each tube narrows toward the osculum. Syconoid, with completely fused radial tubes supported by an articulated skeleton and with short diactine tufts at the distal ends (Chapter 3, Fig. 3.2, B; Fig. A3.1.3, B). Tetractines with a short and bend apical ray are present in the wall of the radial tubes. Pseudosagittal spicules are present at the distal end of each radial tube, with the unpaired actine pointing out

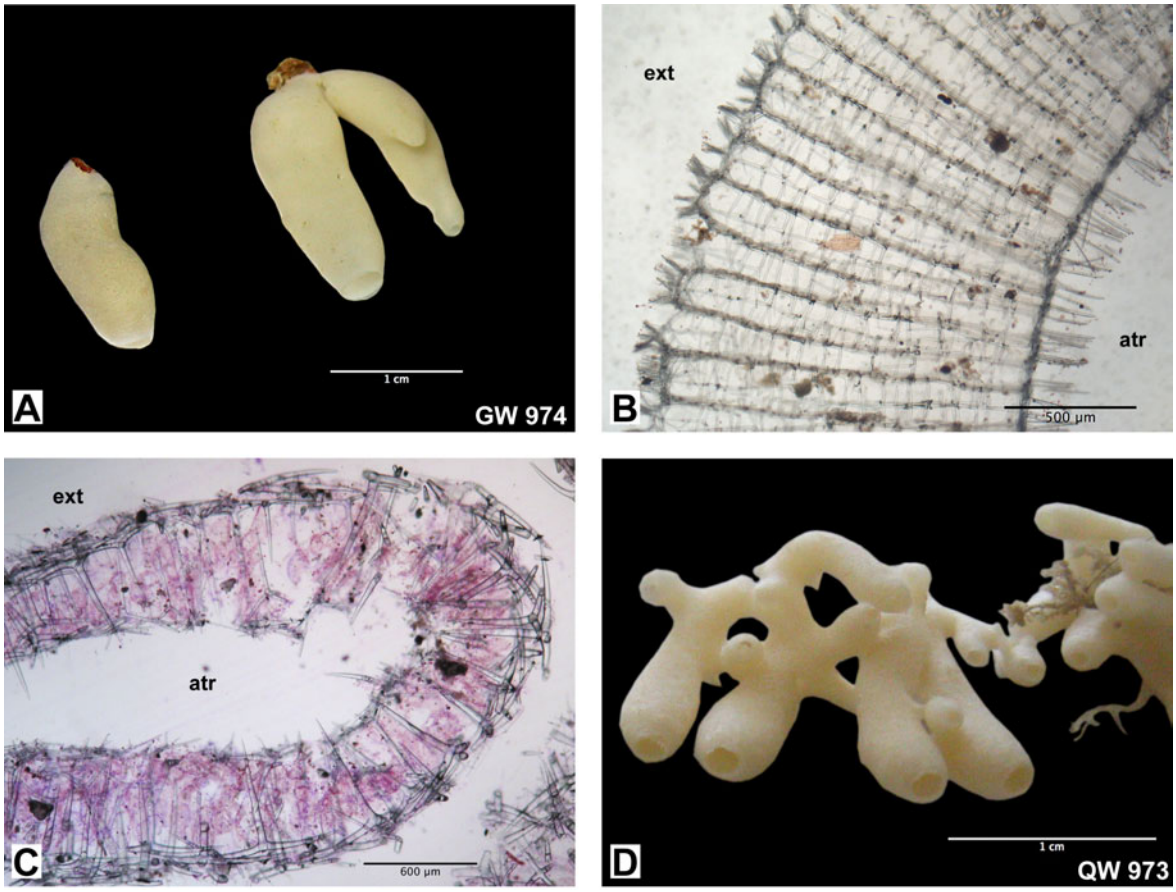


Figure A3.1.3: Specimen of Calacronea.
A: Habitus of *Grantessa* sp.; **B:** transverse section of *Grantessa* sp. (GW 979); **C:** transverse section of *Leucilla* sp. (ZMA-POR 5381); **D:** Habitus of *Grantiopsis* cf. *cylindrica*.

from the radial tube into the neighboring tube. In *Grantessa*, a thin cortex should be present (Borojevic et al., 2000). From our point of view no such thin cortex is present in our specimens. However, even in the drawings of *Grantessa ramosa* by Borojevic et al, (2000, see their Fig. 32) we cannot recognize a clear cortex and found that the displayed arrangement is almost identical to the ones in our specimen (besides the presence of tetractines in the radial tubes, see chapter 3, Fig. 3.2, B). Therefore we assigned our samples to the genus *Grantessa*.

Family: Amphoriscidae Dendy, 1892

***Leucilla* Haeckel, 1872**

Specimen: ZMA POR 5381

Locality: Caribbean, Netherlands Antilles, Curacao

The specimen was formerly determined as *Leucandra barbata* (Grantiidae). We examined the skeletal arrangement and find this specimen to belong to the genus *Leucilla*. The cortex consists of large tetractines, whose apical ray is slightly longer than the basal rays and reaches through the complete choanosome (Fig. A3.1.3, C). The unpaired actine of subatrial sagittal triactines provide

additional support for the choanocyte chambers. Tetractines (much smaller than the cortical ones) are present in the artrial. The organization of the aquiferous system is sylleibid.

Family: Lelapiidae Dendy & Row, 1913

***Grantiopsis* Dendy 1892**

Grantiopsis* cf. *cylindrica

Specimen: GW973

Locality: GBR, Australia, Lizard Island, Bommie Bay Cave

Individual tubes emerging from stolons (Fig. A3.1.3, D). The osculum is smaller than the diameter of the tube. The cortex consists of several layers of large triactines, which gives the tubes a smooth and glistening surface. The unpaired rays of subatrial tetractines are associated with spicular tracts of several modified triactines with two strongly reduced rays (Chapter 3, Fig. 3.1, C). The other unpaired ray of the subatrial tetractines is bend and reaches into the atrium. The atrial skeleton consists of small tetractines with the apical ray sometimes dagger shaped and also reaching into the atrium. The aquiferous system is sylleibid.

Two species have been described in *Grantiopsis*: *G. cylindrica* Dendy 1892 (with two varieties) and *G. heroni* Wörheide and Hooper, 2003. An additional variety of *G. cylindrica –G. cylindrica* var. *fruticosa* Dendy & Frederick 1924– was described from the Abrolhos Island, Western Australia. While *G. heroni* has an osculum of almost the size of the tube, the osculum is smaller in *G. cylindrica* and our specimen. Also, at least the specimens of *G. heron*i from the first description are syconoid (Wörheide and Hooper, 2003). Considering this and the description of Dendy and Frederick (1924), the examined specimen is most similar to *G. cylindrica* var. *fruticosa*, because the typical *G. cylindrica* specimen consisted of single tubes, while *G. cylindrica* var. *fruticosa* is 'colonial, branched and bushy'. Additionally, the cortex in the typical *G. cylindrica* is thicker (about half of the thickness of the tube) than in *G. cylindrica* var. *fruticosa* (one third of the thickness of the wall) (Dendy and Frederick, 1924). In our specimen the cortex is even less developed than described for *G. cylindrica* var. *fruticosa* (ca. one quarter of the thickness of the tube). However, it is not known if this variability between the described species is caused by species boundaries or simply reflect the plasticity within a single *Grantiopsis* species.

Calcinea

Order Clathrinida Hartman,1958

Clathrina Gray 1867

Specimen: GW 957

Locality: GBR, Mac's Reef

The specimen shows the typical branching and anastomsing growth form of the genus (Fig. A3.1.4, A).

Family Leucaltidae Dendy &Row, 1913

***Leucettusa* Haeckel 1872**

***Leucettusa* sp. 1**

Specimen: QM G323232

Locality: Tasmania, Ling Hole

The specimen is a fragment of the base of a tubular sponge (Fig. A3.1.4, C). Several layers of large triactines form the strong cortex (Chapter 3, Fig. 3.1., D). Tetractines are not present in our slides, but examination of a molecular identical specimen (QM-G323268) suggest that such tetractines are present (at least in other parts of the sponge body) and that their apical ray protrude through almost the complete choanosome. The thin atrial skeleton consists of very small tetractines. The choanosome is free of spicules besides scattered tetractines that have are about the size of the spicules of the atrial spicules. The aquiferous system is leuconoid. These features are similar to *L. haeckelina* Poléjaeff, 1883 (compare also plate VII, figs 1-6 in Poléjaeff, 1883). However, a certain determination would require examination of the holotype.

***Leucettusa* sp. 2**

Specimens: QM G323253, QM G323283

Locality: Tasmania

While QM G323253 is only a fragment (Fig. A3.1.4, E), specimen QM G323283 is an individual tubular, clubbed shaped sponge, widening in the distal part and again narrowing toward the small osculum (Fig. A3.1.4, D). The outside has a rough appearance due to the large tangential triactines, which form a several-layered cortex. Large tetractines with their basal rays within the cortex support the choanosome with their apical ray, that almost reaches the atrial skeleton. The atrial skeleton consists of small tetractines. The choanoskeleton only is formed by few scattered small spicules, mostly triactines with one very reduced ray, giving them a v-like shape. The aquiferous

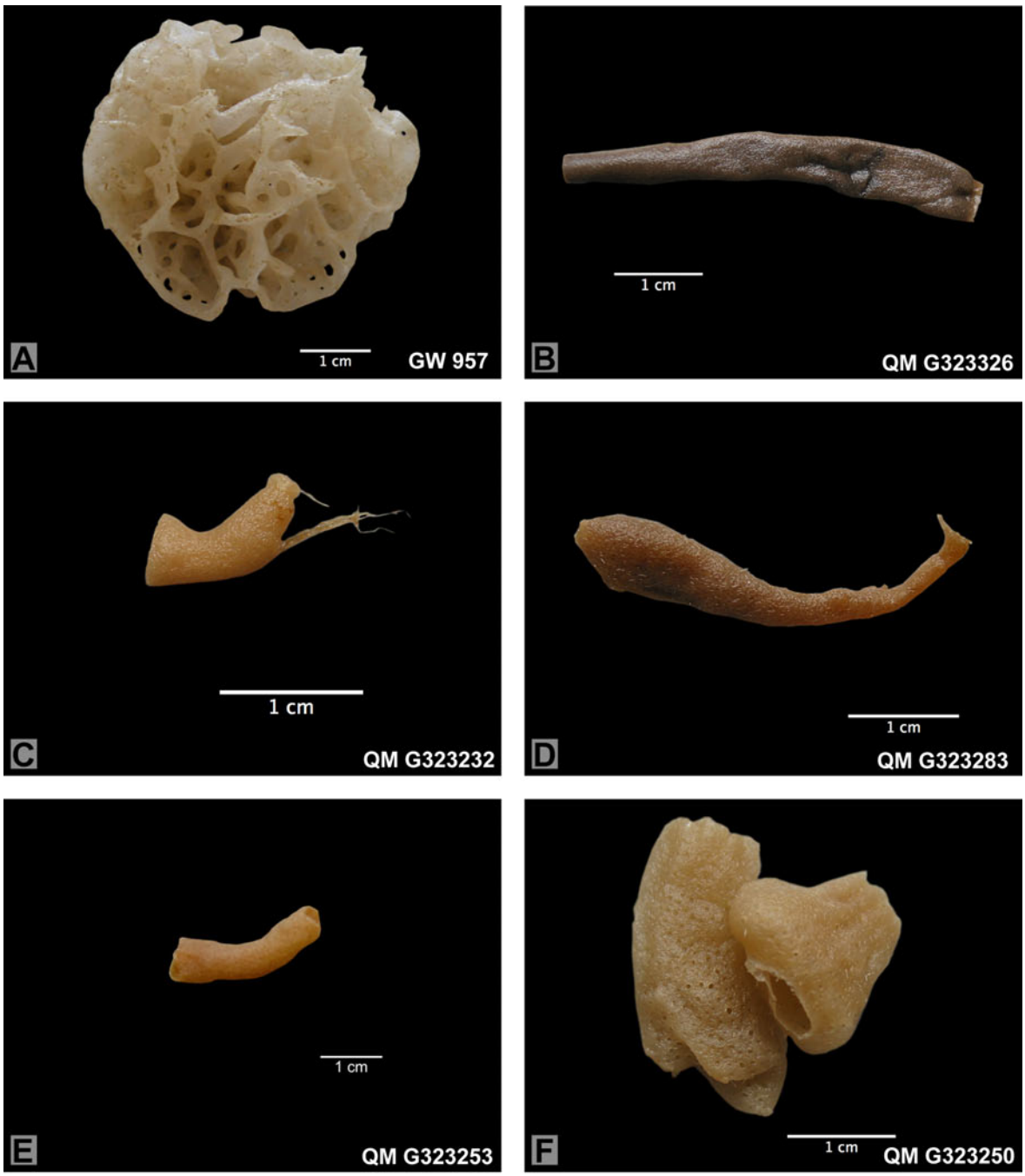


Figure A3.1.4: Specimen of Calcinea, external appearance.
A: *Clathrina* sp.; **B:** *Ascandra* sp.; **C:** *Leucettusa* sp.1; **D&E**: the two specimen of *Leucettusa* sp. 2; **F:** undetermined Calcinea.

system is leuconoid. The specimen has some affinities towards *L. imperfecta* Poléjaeff, 1883 or *L. tubulosa* Dendy 1924.

***Ascandra* Haeckel 1872**

***Ascandra* sp.**

Specimen: QM-G323326

Locality: Tasmania

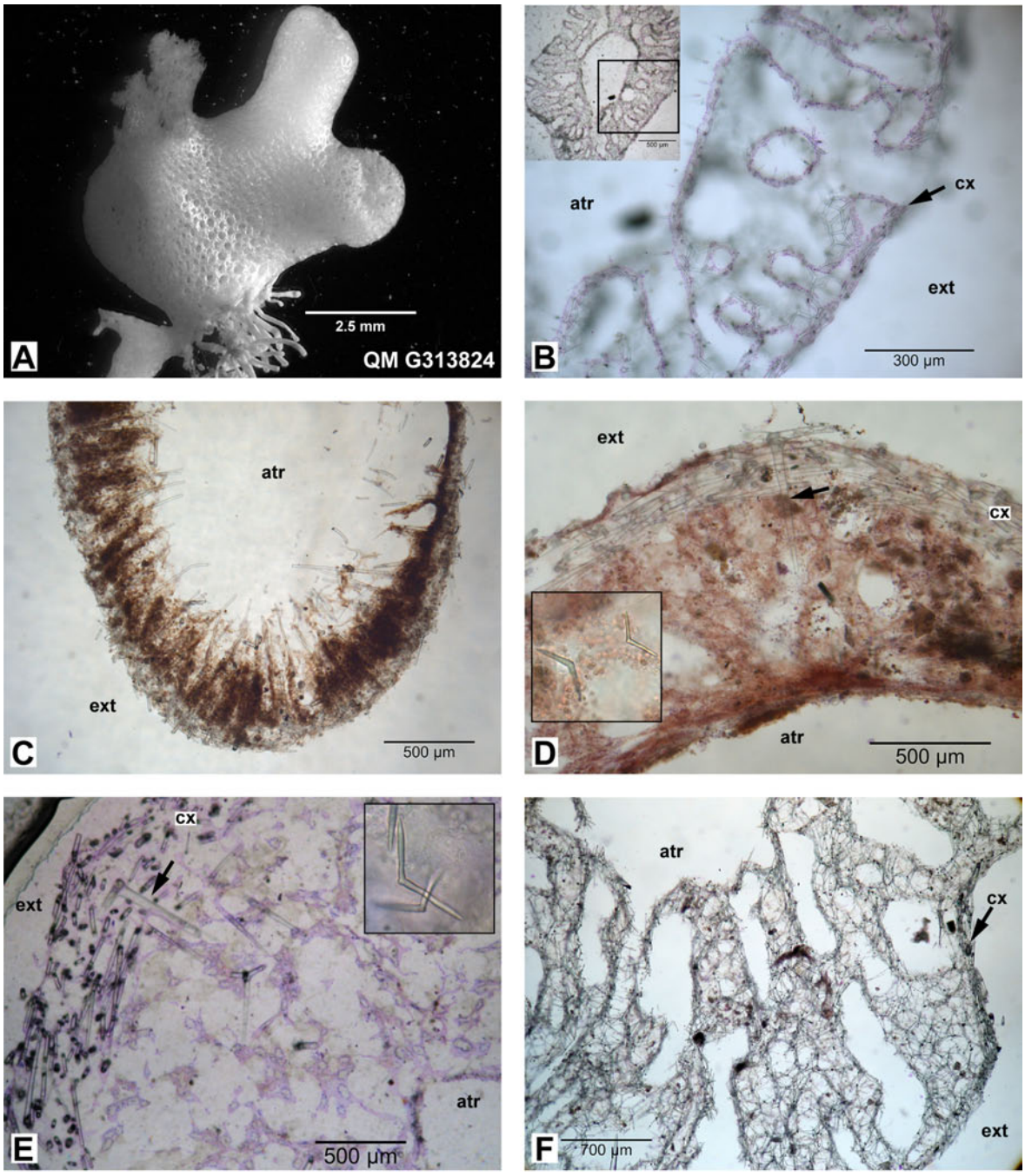


Figure A3.1.5: Specimen of Calcinea

A&B: *Ascaltis* sp. (QM G313824), **A:** habitus; **B:** transverse section, inlet overview over the complete transverse section. Arrow points to the thin cortex. **C:** Transverse section of *Ascandra* sp. . Because of the thickness of this sections, the nature of the folds of the choanoderm only gets clear in the upper right. **D:** *Leucettusa* sp. 2 (QM G323283). Arrow points to the apical ray of a giant tetractine; Inlet: small triactines of the choanoderm with one reduced ray, giving it a 'v'-shaped appearance (spicules are about 40 and 50 µm wide, respectively.) **E:** *Leucettusa* sp. 2 (QM G323253), the arrow points at the apical ray of the large tetractines that support the choanosome. Size of the 'v'-shaped spicules (inlet) about 50 µm. **F:** transverse section through the undetermined Calcinea (QM G323250).

Tubular sponge with a smooth surface (Fig. A3.1.4, B). The prominent spicules are large tetractines, whose apical ray is long, sometimes slightly irregularly bent at towards the end and points into the atrium and provide support for the choanosome (Fig. A3.1.5, C). The sponge has an asco-

noid aquiferous system, but the choanosome forms deep folds along the large apical rays of the spicules.

Family Leucascidae Dendy, 1892

***Ascaltis* Haeckel 1872**

***Ascaltis* sp.**

Specimen: QM G313824

Locality: Australia, Great Barrier Reef, Hook Reef

This specimen was referred to as *Clathrina* aff. *cerebrum* in a previous study (Dohrmann et al., 2006). The basal rays of the tetractines of the choanoskeleton carry spines as in *Clathrina cerebrum.* Some confusion in the former determination may also have arisen from the fact that *Clathrina cerebrum* Haeckel 1872 was originally described as *Ascaltis cerebrum* (Haeckel, 1872). However, after reexamination of the specimen, we find that it represents an *Ascaltis* species. The small sponge branches in short tubes (of which 2 are intact in the specimen), each with a terminal naked osculum (Fig. A3.1.5, A). Inhalant openings are distributed equally over the sponge and delimited by the cortical spicules. The choanoskeleton consists of anastomosed and ramified tubes connected to the central tube (Fig. A3.1.4, B). The basal rays of the tetractines of the choanoskeleton carry spines as in *Clathrina cerebrum.*

Undetermined Calcinea

Specimen: QM G323250

Locality: Tasmania, King Island Canyons

The specimen consists of two parts and has an atrium that opens with a large osculum (Fig. A3.1.4, F). Large triactines form a thin cortex. Big inhalant spaces are visible on the outside of the sponge. The choanoskeleton consists of branching an anastomosing tubes (Fig. A3.1.5, F), the organization is therefore not a solid body as in the closely related Leucettidae species. Unfortunately the organization of the aquiferous system is not recognizable, probably due to suboptimal fixation.

References

Borojevic, R., Boury-Esnault, N. and Vacelet, J. (2000). A revision of the supraspecific classification of the subclass Calcaronea (Porifera, class Calcarea). *Zoosystema* **22(2) :** 203-263.

Carter, H. J. (1886). Descriptions of Sponges from the Neighbourhood of Port Phillip Heads, South Australia, continued. *Annals and Magazine of Natural History* **5(18) :** 34-55,126-149.

Dendy, A. (1892). Preliminary account of *Synute pulchella*, a new genus and species of calcareous sponges. *Proceedings of the Royal Society of Victoria* **4(1) :** 1-6.

Dendy, A. and Frederick, L. M. (1924). On a collection of sponges from the Abrolhos Islands, Western Australia. *Journal of the Linnean Society. Zoology* **35:** 477-519.

Dendy, A. R. T. H. U. R. (1893). Memoirs: Studies on the Comparative Anatomy of Sponges: V. Observations on the Structure and Classification of the Calcarea Heterocala. *The Quarterly Journal of Microscopical Science* **s2-35(138) :** 159-257.

Dohrmann, M., Voigt, O., Erpenbeck, D. and Wörheide, G. (2006). Non-monophyly of most supraspecific taxa of calcareous sponges (Porifera, Calcarea) revealed by increased taxon sampling and partitioned Bayesian analysis of ribosomal DNA. *Molecular Phylogenetics and Evolution* **40(3) :** 830-843.

Haeckel, E. (1872). Die Kalkschwämme, Band 2: System der Kalkschwämme (Calcispongien oder Grantien). pp. 416. Georg Reimer, Berlin.

Poléjaeff, N. (1883). Report on the Calcarea dredged by H.M.S. 'Challenger', during the years 1873-1876. *Report on the scientific results of the voyage of H.M.S. 'Challenger', 1873-1876. Zoology* **8(2) :** 1-76.

Row, R. H. W. (1909). Reports on the marine biology of the Sudanese Red Sea. XIII. Report on the Sponges, collected by Mr. Cyril Crossland in 1904-5. Part I. Calcarea. *Journal of the Linnean Society, Zoology* **31(206) :** 182-214.

Wörheide, G. and Hooper, J. N. A. (2003). New species of Calcaronea (Porifera: Calcarea) from cryptic habitats of the southern Great Barrier Reef (Heron Island and Wistari Reef, Capricorn-Bunker Group, Australia). *Journal of Natural History* **37:** 1-47.

Table A3.1: LSU primer sequences.

Name	Sequence (5'-3')	Reference
F63mod	ACCCGCTGAAYTTAAGCATATHANTMAG	
28S-350rv	CTTTCCCTCACGGTACTTG	
28S-560rv	CTTTCAACGGYTTCACGTGC	
28S-C2-fwd	GAAAAGAACTTTGRARAGAGAGT	Chombard et al 1998 Syst. Biol.
28S-D2-rev	TCCGTGTTTCAAGACGGG	Chombard et al 1998 Syst. Biol.
NL4F	GACCCGAAAGATGGTGAACTA	Nichols pers.com.
NL4R	ACCTTGGAGACCTGATGCG	Nichols pers.com.
28S-1260fw	ATTCTCAAACTTTAAATBGGTAAG	
28S-1340rv	CATCGCCAGTTCTGCTTAC	
28S-1810fw	CGAAAGGGAATCGGGTTAATATTCC	
28S-2490fw	CAACCAAGCGCGGGTAAACG	
28S-2570rv	AATCTCGTTAATCCATTCATGC	
28S-2634fw	TCAAAGTGAAGAAATTCAACCAAGC	
R3264	TTCYGACTTAGAGGCGTTCAG	

Table A3.2: Bayes factors from model comparisons.

			RNA																
model	ln P (model I data)	S.E.	6A	6B	6C	6D	6E	7A	7B	7C	7D	7E	7F	16A	16B	16C	16D	16E	16F	REV
RNA6A	-40836.873	+/- 0.19	-	40.2	50.0	956.3	946.6	8098.8	8107.6	9235.0	8195.4	9404.3	8197.9	14775.0	16649.5	14800.4	15693.9	16269.5	16444.4	21859.2
RNA6B	-40856.951	+/- 0.23	-40.2	-	9.9	916.2	906.4	8058.7	8067.4	9194.8	8155.3	9364.1	8157.7	14734.8	16609.3	14760.3	15653.7	16229.4	16404.2	21819.1
RNA6C	-40861.896	+/- 0.213	-50.0	-9.9	-	906.3	896.5	8048.8	8057.5	9184.9	8145.4	9354.2	8147.8	14724.9	16599.4	14750.4	15643.8	16219.5	16394.3	21809.2
RNA6D	-41315.04	+/- 0.198	-956.3	-916.2	-906.3	-	-9.8	7142.5	7151.2	8278.6	7239.1	8447.9	7241.5	13818.6	15693.1	13844.1	14737.5	15313.2	15488.1	20902.9
RNA6E	-41310.148	+/- 0.213	-946.6	-906.4	-896.5	9.8	-	7152.3	7161.0	8288.4	7248.9	8457.7	7251.3	13828.4	15702.9	13853.9	14747.3	15323.0	15497.8	20912.7
RNA7A	-44886.276	+/- 0.211	-8098.8	-8058.7	-8048.8	-7142.5	-7152.3	-	8.8	1136.2	96.6	1305.4	99.1	6676.2	8550.7	6701.6	7595.1	8170.7	8345.6	13760.4
RNA7B	-44890.653	+/- 0.204	-8107.6	-8067.4	-8057.5	-7151.2	-7161.0	-8.8	-	1127.4	87.9	1296.7	90.3	6667.4	8541.9	6692.9	7586.3	8162.0	8336.8	13751.7
RNA7C	-45454.358	+/- 0.218	-9235.0	-9194.8	-9184.9	-8278.6	-8288.4	-1136.2	-1127.4	-	-1039.6	169.3	-1037.1	5540.0	7414.5	5565.5	6458.9	7034.6	7209.4	12624.2
RNA7D	-44934.579	+/- 0.204	-8195.4	-8155.3	-8145.4	-7239.1	-7248.9	-96.6	-87.9	1039.6	-	1208.8	2.5	6579.6	8454.0	6605.0	7498.5	8074.1	8249.0	13663.8
RNA7E	-45538.998	+/- 0.188	-9404.3	-9364.1	-9354.2	-8447.9	-8457.7	-1305.4	-1296.7	-169.3	-1208.8	-	-1206.4	5370.7	7245.2	5396.2	6289.6	6865.3	7040.1	12455.0
RNA7F	-44935.804	+/- 0.177	-8197.9	-8157.7	-8147.8	-7241.5	-7251.3	-99.1	-90.3	1037.1	-2.5	1206.4	-	6577.1	8451.6	6602.6	7496.0	8071.7	8246.5	13661.4
RNA16A	-48224.363	+/- 0.201	-14775.0	-14734.8	-14724.9	-13818.6	-13828.4	-6676.2	-6667.4	-5540.0	-6579.6	-5370.7	-6577.1	-	1874.5	25.5	918.9	1494.5	1669.4	7084.2
RNA16B	-49161.601	+/- 0.197	-16649.5	-16609.3	-16599.4	-15693.1	-15702.9	-8550.7	-8541.9	-7414.5	-8454.0	-7245.2	-8451.6	-1874.5	-	-1849.0	-955.6	-379.9	-205.1	5209.8
RNA16C	-48237.093	+/- 0.168	-14800.4	-14760.3	-14750.4	-13844.1	-13853.9	-6701.6	-6692.9	-5565.5	-6605.0	-5396.2	-6602.6	-25.5	1849.0	-	893.4	1469.1	1643.9	7058.8
RNA16D	-48683.81	+/- 0.17	-15693.9	-15653.7	-15643.8	-14737.5	-14747.3	-7595.1	-7586.3	-6458.9	-7498.5	-6289.6	-7496.0	-918.9	955.6	-893.4	-	575.7	750.5	6165.3
RNA16E	-48971.636	+/- 0.201	-16269.5	-16229.4	-16219.5	-15313.2	-15323.0	-8170.7	-8162.0	-7034.6	-8074.1	-6865.3	-8071.7	-1494.5	379.9	-1469.1	-575.7	-	174.9	5589.7
RNA16F	-49059.065	+/- 0.171	-16444.4	-16404.2	-16394.3	-15488.1	-15497.8	-8345.6	-8336.8	-7209.4	-8249.0	-7040.1	-8246.5	-1669.4	205.1	-1643.9	-750.5	-174.9	-	5414.8
REV	-51766.479	+/- 0.198	-21859.2	-21819.1	-21809.2	-20902.9	-20912.7	-13760.4	-13751.7	-12624.2	-13663.8	-12455.0	-13661.4	-7084.2	-5209.8	-7058.8	-6165.3	-5589.7	-5414.8	-

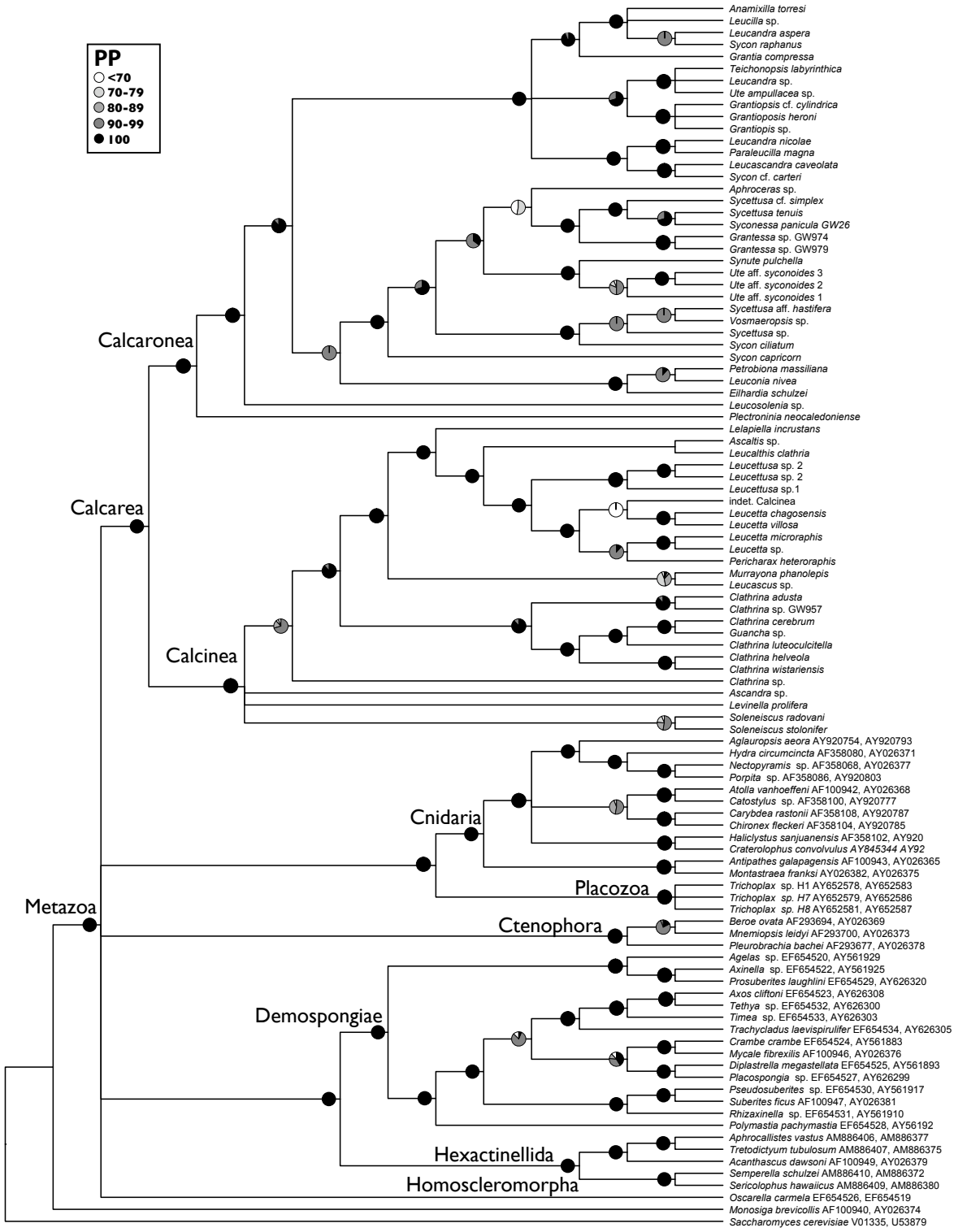


Figure A3.2: Strict consensus tree from the 17 analyses with doublet models

for partition stem. Polytomies reflect differences between the analyses. Pie charts display the range of PP values that were recovered at the shared nodes between the differnt runs. Note that in most cases differences are moderate. Genbank accession numbers are given after the outgroup taxa (SSU, LSU).

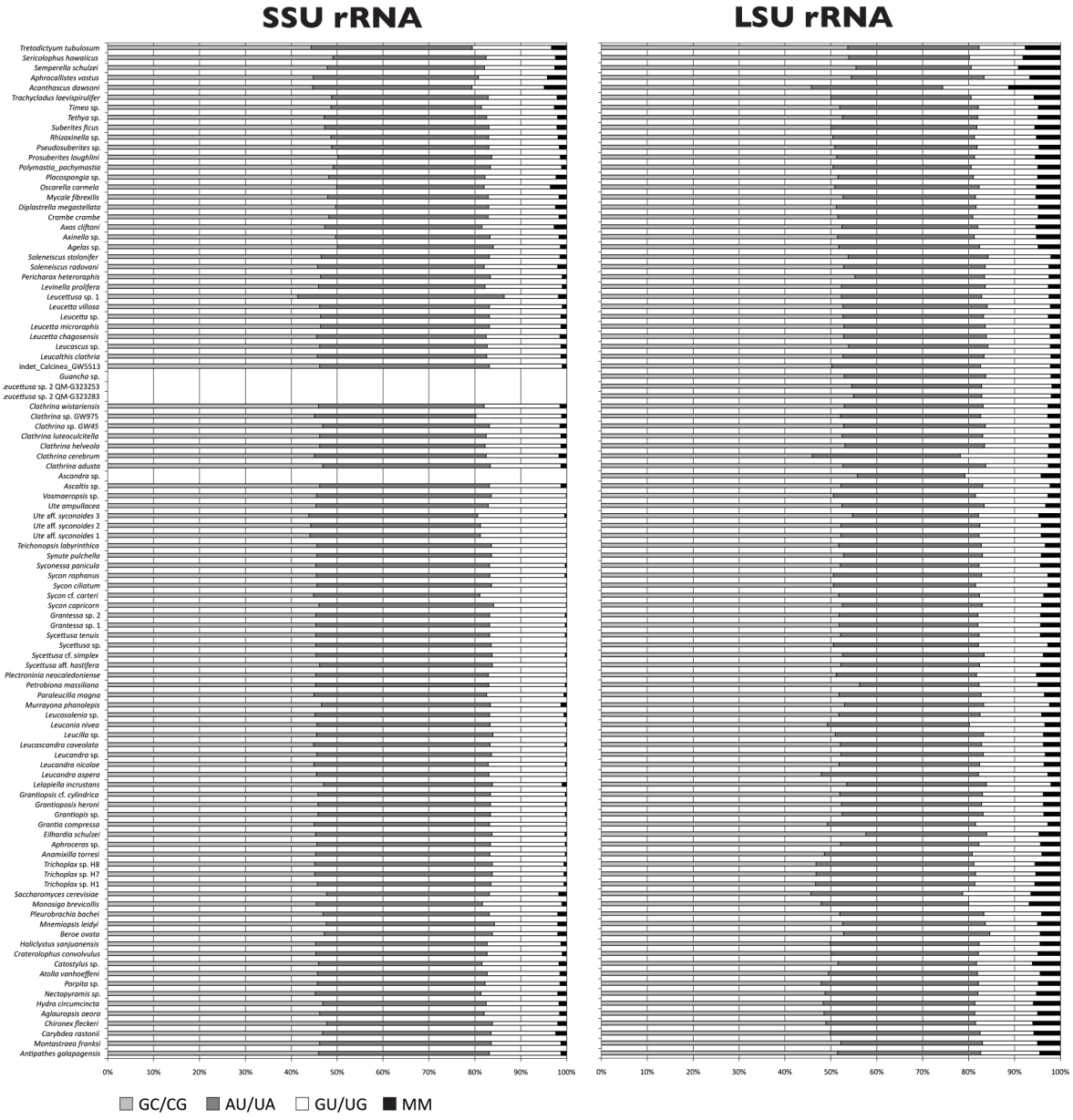


Figure A3.3: Doublet composition of each sequence.

100% refers to all defined doublets in the sequence; missing data or doublets including ambiguities were excluded.

Appendix 4

Table A4.1: Methods.

Action	Program / Primer	Source	Conditions
Origin of traces	Genbank	http://www.ncbi.nlm.nih.gov/	
Database search	BLAST, MEGBLAST	e.g., Altschul et al. 1997	default
Assembly	"The Bommie"	http://reefedge.sols.uq.edu.au/	
Assembly	Codon Code Aligner 1.2	Codon Code Corperation	
Annotation proteins	ORF finder	http://www.ncbi.nlm.nih.gov/gorf/gorf.html	Genetic code: 4: (Mold, Protozoan... mt).
Annotation tRNAs	tRNA-scanSE	Lowe and Eddy 1997	Default search; invertebrate mitochondria.
Visualization	Artemis 8	The Sanger Institute	default
Repeat detection	The Repeat Finder	http://sgdp.iop.kcl.ac.uk/nikammar/repeatfinder.html	default
Alignment	Muscle 3.6	Edgar 2004	default
Alignment processing	Gblocks 0.91b	Castresana 2000	minimum blocksize: 5; allow gap positions: half; *atp8*: *C. elegans* and *T. adhaerens* excluded, later with dashes re-included.
aa-model prediction	ProtTest	Abascal, Zardoya, and Posada 2005	fast algorithm.
Phylogenetic reconst.	PHYML	Guindon and Gascuel 2003	model after ProtTest-results; 100 bootstrap replicates
Phylogenetic reconstruction	MrBayes 3.1.2	Ronquist and Huelsenbeck 2003	4 MCMC chains; stopvalue=0.001; model after ProtTest-results.
Relative rate test	PHYLTEST 2.0	http://iubio.bio.indiana.edu:7780/archive/public/molbio/evolution/phylo/PhylTest/phyltst.exe	p-distance
Sequencing a missing region directly from holotype	FA19484 (fwd): GAATCTGAAATGCAGAGACTAGCGG FB19449 (fwd): GAGCATTTGATTGTTAATGAAGGCG RB20132 (rev): CGCAACCAAGTGTATCAGGGACAACCCC RA20195 (rev): GGGCACTTTGTCTGAATCTGACGGG		All four combinations: annealing 55°C; extension time 1.30 min.

References:

Abascal, F., R. Zardoya, and D. Posada. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104-2105.

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

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540-552.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-1797.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52:696-704.

Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25:955-964.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574

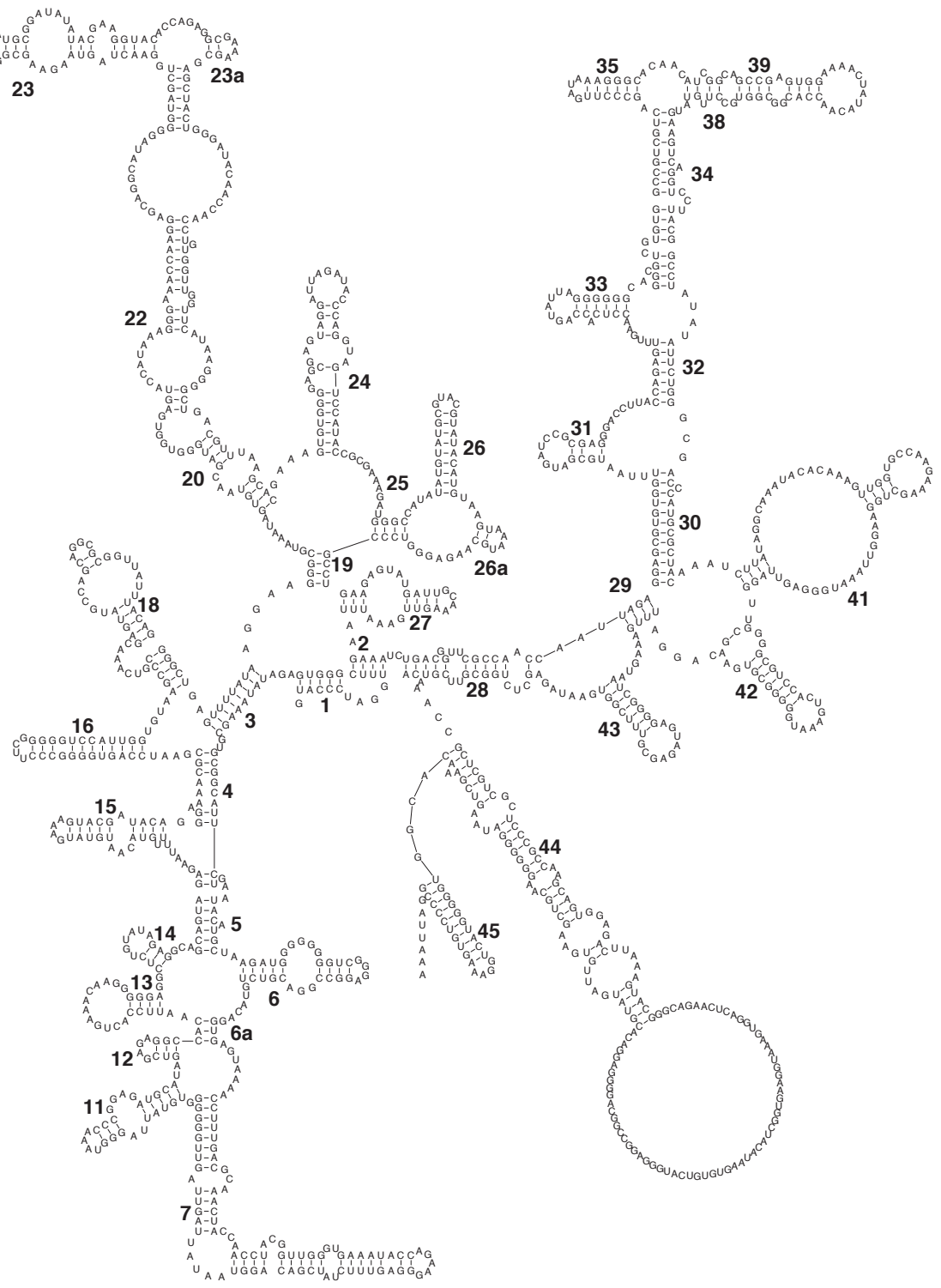


Figure A4.1: Mitochondrial (mt) 12 S rRNA

Predicted secondary structure of Amphimedon. queenslandica.

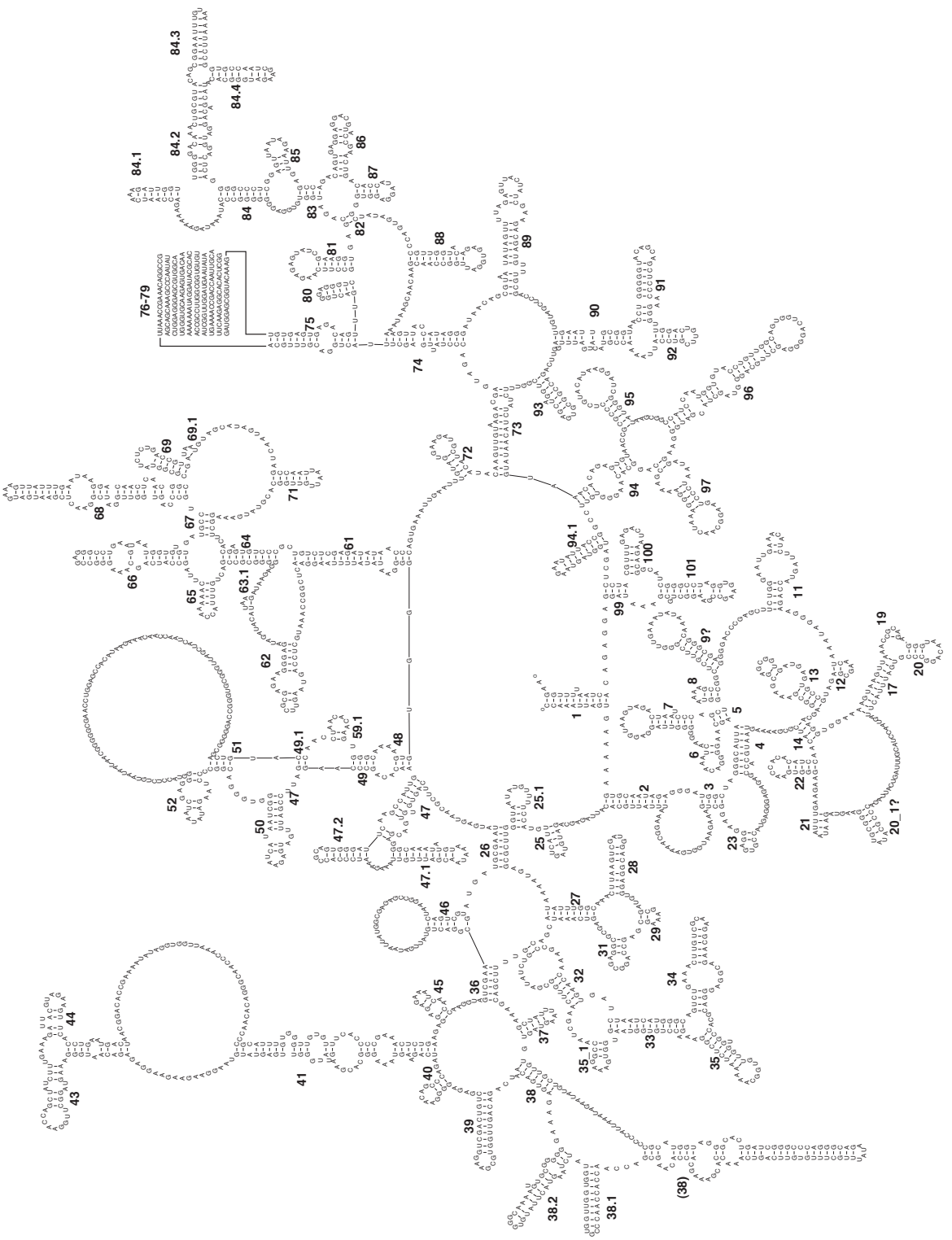
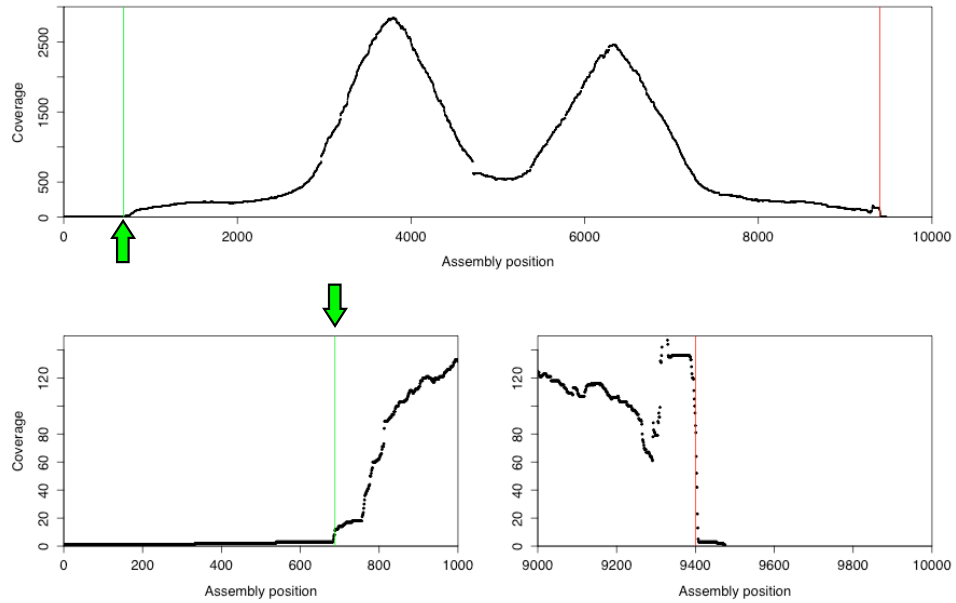


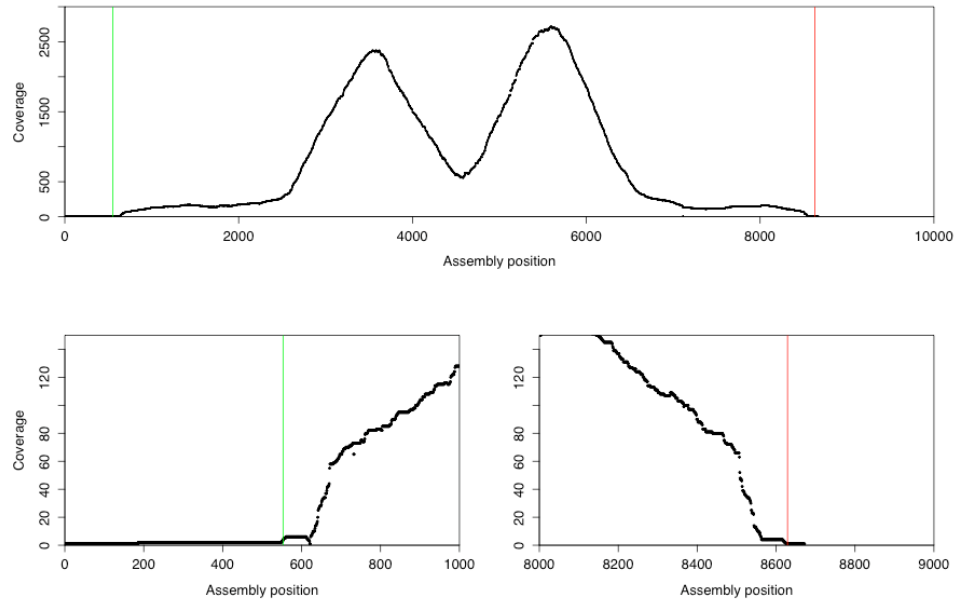
Figure A4.2: Mitochondrial (mt) 16 S rRNA secondary

Predicted structure of *Amphimedon queenslandica*.

Appendix 5



Coverage (number of traces) along the mt1 assembly. Assembly positions refer to the consensus sequence including gaps. Top: complete overview, bottom: excerpts from the assembly ends.



Coverage (number of traces) along the mt2 assembly. Assembly positions refer to the consensus sequence including gaps. Top: complete overview, bottom: excerpts from the assembly ends.

Figure A5.1: Coverage of mt1 and mt2 assemblies.

Presumed ends (vertical green and red lines) were recovered considering the 3' end of mt1 (arrows) from previous studies (Warrior 1998, Pont-Kingdon et al. 2000)

Table A5.1: Primer sequences used in the PCR experiments

[with the following PCR program: 95°C/5min; 37x (95°C/30s; 48°C/30s; 72°C/1-2min; 72°C/3min)] and cycle sequencing reactions.

Primer name	Sequence (5'-3')	PCR fragment	Postion in *H. magnipapillata*
Hydra_mtTIS1	CTGCGATAGCCGCAG	2a,3,4b	mt1: 24; 8157
			mt2: 24; 7648
Hydra_TIS2-seq	GGGTGATCCTCTTTTAGGAG	1,2*,3*,4b*	mt1: 85; 8091
			mt2: 81; 7583
Hydra_l-rRNA-rv	CATGAAAAACCAGCTATCTC	1	mt1: 751
Hydra_nd5_1fw	TGATTACCTGATGCDATGG	2b	mt1: 6621
Hydra_nd5_2fw	TTGAAATGTTATCTTTACAACTT	2a	mt1: 7463
Hydra_cox1_rv	CTTCTAGGCATTCCTGCTAA	2b,4a	mt1: 7787
			mt2: 7279
Hydra_s-rRNA-rv	CGTCTGCTGGCACTTA	3	mt2: 562
C1-L1490 (=cox1_fwd) [1]	GGTCAACAAATCATAAAGATATTGG	4a	mt2: 6005
Hydra_cox1fw2	ACTGTAGGAATGGATGTTGA	4b	mt2: 6862

Table A5.2: Taxa, GenBank accession numbers and AT contents

of protein-coding genes and rRNA genes in 24 cnidarian mt genomes

Species	Accession#	Taxon	% AT in rRNAs	%AT at codon positions 1	2	3
Hydra magnipapillata	BN001179, BN001180	Hydrozoa	78.06	73.16	70.88	89.84
Hydra oligactis	EU237491	Hydrozoa	77.77	73.42	70.83	88.22
Aurelia aurita	NC_008446	Scyphozoa	66.83	61.15	64.22	74.10
Nematostella sp.	NC_008164	Anthozoa (Hexacorallia)	58.60	55.61	62.08	66.75
Metridium senile	NC_000933	Anthozoa (Hexacorallia)	58.06	57.27	62.20	68.31
Chrysopathes formosa	NC_008411	Anthozoa (Hexacorallia)	58.46	57.73	62.12	66.15
Rhodactis sp.	NC_008158	Anthozoa (Hexacorallia)	58.08	56.27	63.08	66.72
Discosoma sp.	NC_008072	Anthozoa (Hexacorallia)	57.95	55.95	62.91	66.85
Ricordea florida	NC_008159	Anthozoa (Hexacorallia)	59.51	57.47	63.06	68.28
Acropora tenuis	NC_003522	Anthozoa (Hexacorallia)	60.26	56.87	63.78	66.92
Anacropora matthai	NC_006898	Anthozoa (Hexacorallia)	60.00	56.48	63.51	65.71
Montipora cactus	NC_006902	Anthozoa (Hexacorallia)	59.83	56.43	63.51	65.51
Pocillopora damicornis	NC_009797	Anthozoa (Hexacorallia)	71.40	61.47	67.49	83.40
Colpophyllia natans	NC_008162	Anthozoa (Hexacorallia)	68.10	58.45	65.72	77.16
Montastraea faveolata	NC_007226	Anthozoa (Hexacorallia)	67.56	58.27	65.40	75.59
Mussa angulosa	NC_008163	Anthozoa (Hexacorallia)	67.79	58.43	65.64	76.56
Astrangia sp.	NC_008161	Anthozoa (Hexacorallia)	67.41	58.60	65.78	79.89
Agaricia humilis	NC_008160	Anthozoa (Hexacorallia)	59.23	55.63	63.35	62.56
Pavona clavus	NC_008165	Anthozoa (Hexacorallia)	58.42	55.46	63.24	62.58
Porites porites	NC_008166	Anthozoa (Hexacorallia)	61.15	57.38	63.64	69.95
Siderastrea radians	NC_008167	Anthozoa (Hexacorallia)	60.85	56.96	63.59	69.52
Savalia savaglia	NC_008827	Anthozoa (Hexacorallia)	48.92	55.32	60.96	50.99
Pseudopterogorgia bipinnata	NC_008157	Anthozoa (Octocorallia)	56.62	58.60	62.17	72.49
Briareum asbestinum	NC_008073	Anthozoa (Octocorallia)	56.86	58.23	62.40	72.61

Acknowledgements

I thank Gert Wörheide for giving me the great opportunity to work on the research of my interest in such an inspiring scientific environment, and his patience when it came to 'deadlines' for finishing manuscripts. I also benefitted much from his encouragement and support for my participation in training courses and conferences, and his willingness to let me do some flat-animal project on the side. I also like to thank Allen Collins for accepting as second referee on this work, and for raising and supporting my interest in research on the 'lower 'Metazoa, and Volker Thiel, for being part of my Ph.D. committee and helpful comments .

Many thanks goes to Cat Vogler, and Dirk Erpenbeck and for the many helpful discussions, ideas and support, it helped enjoying work very much, as did the after-work beers (if only the bars would not close so early... where are all the students, anyway?). Additionally, I like to thank Cat for patiently reviewing my manuscripts and correcting my English language and grammar.

I thank Eilika Wülfing for her skilled and productive assistance in producing LSU sequences.

For kindly letting me subsample specimens from theis collections, I like to thank John Hooper and Monika Schlacher-Hoenlinger from the Queensland Museum in Brisbane, Jane Fromont from the Western Australian Museum in Perth, Thierry Laperousaz and Shirley Sorokin from the South Australian Museum in Adelaide and Rob van Soest from the Zoologisch Museum, Universiteit van Amsterdam, Amsterdam, Netherlands.

The work in the Göttingen lab would have been less fun without all my other colleagues, not mentioned so far: Dan Jackson, Kerstin Pick Klemi Karlinska, Luciana Macis and Martin Dohrman. I enjoyed working with you very much.

Last but not least, I like to thank my parents, my sister, Carsten Feige and Petra Nolting for their support and for always offering a refuge from work when necessary.

Erklärung über eigene Leistungen

Ich versichere, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Die Stellen, die anderen Werken wörtlich oder sinngemäß entnommen sind, sind als solche kenntlich gemacht. Eigene Beiträge im Verhältnis zu denen von Koautoren bei bereits publizierten oder zur Publikation einzureichenden Teilen dieser Arbeit sind wie folgt:

Kaptiel 1: Oliver Voigt konzipierte die Studie in wesentlichen Teilen, erhob, analysierte und interpretierte die Daten und verfasste das Manuskript. Dirk Erpenbeck und Gert Wörheide trugen zur Konzeption der Studie und zur Verbesserung des Manuskripts durch kritische Kommentare bei.

Kapitel 2: Martin Dohrmann erhob einen wesentlichen Anteil der Daten, führte die Sequenzanalysen durch, und verfasste das Manuskript. Gerd Wörheide führte die Feldarbeit durch. Gert Wörheide und Martin Dohrmann bestimmten die Proben und konzipierten die Studie. Oliver Voigt generierte einen erheblichen Teil der Sequenzdaten und stellte die Sekundärstrukturinformation für die 18S Analysen zusammen. Dirk Erpenbeck stellte die Sekundärstrukturinformation für die 28S Analysen zusammen. Dirk Erpenbeck, Gert Wörheide und Oliver Voigt trugen zur Verbesserung des Manuskripts durch kritische Kommentare bei.

Kapitel 3: Oliver Voigt konzipierte die Studie in wesentlichen Teilen, erhob, analysierte und interpretierte die Daten und verfasste das Manuskript. Eilika Wülfing erhob einen nicht unerheblichen Teil der LSU Daten. Gert Wörheide trug zur Konzeption der Studie und zur Verbesserung des Manuskripts durch kritische Kommentare bei.

Kapitel 4: Dirk Erpenbeck führte wesentliche Teile der Datenerhebung und -analysen durch und verfasste das Manuskript. Oliver Voigt trug zur Datenerhebung und -analyse bei. Marcin Adamski, Maja Adamska, John Hooper, und Bernard Degnan halfen bei der Datenerhebung aus den Genomdaten von *Amphimedon*, Gert Wörheide trug wie alle Autoren zur Verbesserung des Manuskripts durch kritische Kommentare bei.

Kapitel 5: Oliver Voigt trug wesentlich zur Konzeption der Studie bei, erhob und interpretierte die Daten und verfasste das Manuskript. Dirk Erpenbeck und Gert Wörheide halfen bei der Interpretation der Daten und trugen zur Verbesserung des Manuskripts durch kritische Kommentare bei.

Ich versichere weiterhin, dass diese Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat.

Göttingen, August 2009