J Mol Evol (2015) 80:278–291
DOI 10.1007/s00239-015-9682-z

ORIGINAL ARTICLE

A Proposal for the Evolution of *Cathepsin* and *Silicatein* in Sponges

Ana Riesgo[1,2] · Manuel Maldonado[3] · Susanna López-Legentil[4] · Gonzalo Giribet[5]

Received: 9 February 2015 / Accepted: 6 May 2015 / Published online: 19 May 2015

Abstract Cathepsins are enzymes capable of degrading proteins intracellularly. They occur ubiquitously in opisthokonts, but their potential to provide insight across the evolutionary transition from protists to metazoans remains poorly investigated. Here, we explore the evolution of *cathepsins* using comparative analyses of transcriptomic datasets, focusing on both, protists (closely related to metazoans), and early divergent animals (i.e., sponges). We retrieved DNA sequences of nine *cathepsin* types (*B*, *C*, *D*, *F*, *H*, *L*, *O*, *Z*, and *silicatein*) in the surveyed taxa. In choanoflagellates, only five types (*B*, *C*, *L*, *O*, *Z*) were identified, all of them being also found in sponges, indicating that while all *cathepsins* present in protists were conserved across metazoan lineages, *cathepsins F* and *H* (and probably *D*) are metazoan acquisitions. The phylogeny of cysteine protease *cathepsins* (excluding *cathepsin D*) revealed two major lineages: lineage B (*cathepsins B* and *C*) and lineage L (*cathepsins F*, *H*, *L*, *O*, *Z*). In the latter lineage, a mutation at the active site of *cathepsin L* gave rise to *silicatein*, an enzyme exclusively known to date from siliceous sponges and involved in the production of their silica spicules. However, we found that several sponges with siliceous spicules did not express *silicatein* genes and that, in contrast, several aspiculate sponges did contain *silicatein* genes. Our results suggest that the ability to silicify may have evolved independently within sponges, some of them losing this capacity secondarily. We also show that most phylogenies based on *cathepsin* and *silicatein* genes (except for that of *cathepsin O*) failed to recover the major lineages of sponges.

Keywords Porifera · Molecular evolution · Phylogeny · Next-generation sequencing · Cysteine proteases

Electronic supplementary material The online version of this article (doi:10.1007/s00239-015-9682-z) contains supplementary material, which is available to authorized users.

✉ Ana Riesgo
anariesgogil@gmail.com

[1] Department of Animal Biology, University of Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Spain

[2] Present Address: Department of Life Sciences, Natural History Museum of London, Cromwell Road, London SW7 5BD, UK

[3] Department of Marine Ecology, Center for Advanced Studies of Blanes (CEAB-CSIC), Accés a la Cala St. Francesc 14, 17300 Blanes, Girona, Spain

[4] Department of Biology & Marine Biology, and Center for Marine Science, University of North Carolina Wilmington, 5600 Marvin K. Moss Lane, Wilmington, NC 28409, USA

[5] Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

Introduction

Cysteine protease enzymes, generically known as cathepsins, play an essential role in intracellular degradation of proteins in many different organisms, ranging from viruses to animals (Berti and Storer 1995). Cathepsins can be lysosomal (B, C, F, H, K, L, O, S, W, and Z) and be activated by the acid pH of lysosomes (Turk et al. 2001) or non-lysosomal (M, J, Q, R, S, and K). So far, lysosomal cathepsins have been largely studied (Bühling et al. 2002) because of their implication in human disease (e.g., Sloane et al. 1981; Sloane and Honn 1984; Lutgens et al. 2007; Rodriguez-Franco et al. 2012), remaining largely uninvestigated in other phyla. In mammals, the function of

cathepsin B and L is often related to both extracellular degradation of collagen type I in bone and processing of hormones and antigenic proteins (Inui et al. 1997; Mort and Buttle 1997; Söderström 1999; Horn et al. 2005). In other eukaryotes, such as arthropods, cathepsin L is necessary for successful molting (Liu et al. 2006), and also for degrading insect cuticle by entomopathogenic fungi (Samuels and Paterson 1995). Cathepsins L and B have also been reported from a variety of sponges (Krasko et al. 1997; Müller et al. 2003, 2007a), but their roles remain poorly investigated. Interestingly, a modification of cathepsin L seems to have led to a new functional enzyme group exclusive to sponges and known as silicateins (Shimizu et al. 1998).

In silicateins, the ancestral proteolytic activity of cathepsins has shifted into an anabolic activity, that is, the ability to polycondensate enzymatically dissolved silicon to produce biogenic silica, which is the essential skeletal material characterizing many sponge groups (Shimizu et al. 1998; Cha et al. 1999). It was initially thought that this functional modification mainly derived from a change in the original catalytic triad of cathepsin L (i.e., cysteine, histidine, and asparagine), with the cysteine being replaced by a serine. Nevertheless, there is an emerging view that additional amino acid changes must have also occurred to explain the functional change observed (see "Discussion" section).

Three paralogous genes for *silicatein* (*silicatein alpha, beta*, and *gamma*; Wiens et al. 2006) have been identified in sponges (Wang et al. 2012a) but not in the other various groups of livings organisms that are able to produce skeletons of biogenic silica (e.g., diatoms, radiolarians, choanoflagellates, silicoflagellates, chrysophyceans, thecamoebae, etc.; Gröger et al. 2008). However, not all sponges appear to have silicatein, being reliably identified to date in two of the three siliceous classes (Hexactinellida and Demospongiae; Fig. 1a) of the four recognized in the phylum (Shimizu et al. 1998; Cha et al. 1999; Krasko et al. 2000, Funayama et al. 2005; Müller et al. 2007b, 2008a, b; Kozhemyako et al. 2010; Kalyuzhnaya et al. 2011; Veremeichik et al. 2011 Wang et al. 2012b). It is not surprising that the members of the class Calcarea, characterized by producing calcareous spicules, lack silicatein. However, it is puzzling that silicatein has not yet been found in members of the class Homoscleromorpha (Fig. 1a), which are also characterized by producing silica skeletons. From ultrastructural studies of spicule sections and silicifying cells, it was suggested that the homoscleromorphs might have acquired the ability to produce silica independently from the two other siliceous lineages (Maldonado and Riesgo 2007). Homoscleromorphs possess spicules in which the concentric silica layers are separated by unusually thick organic rings in a pattern clearly different from that of demosponges and it is the only sponge clade to date in which pinacocytes are able to produce silica spicules (Maldonado and Riesgo 2007). Within the siliceous sponge lineages that have been confirmed to encode for the enzyme silicatein (i.e., classes Demospongiae and Hexactinellida), the evolutionary pathway of the silicatein is not straightforward either. For instance, several demosponge subclades that are known to lack siliceous skeletons (e.g., Dictyoceratida, Dendroceratida, and part of Verongimorpha) have members that express the gene for silicatein (Kozhemyako et al. 2010).

The molecular evolution of cathepsins B, H, and L has been explored in a handful of animals, predominantly vertebrates (Uinuk-ool et al. 2003). However, few invertebrate taxa were represented in that study, hampering a reliable inference of the relationships among the specific cathepsins from different animal phyla and, reciprocally,

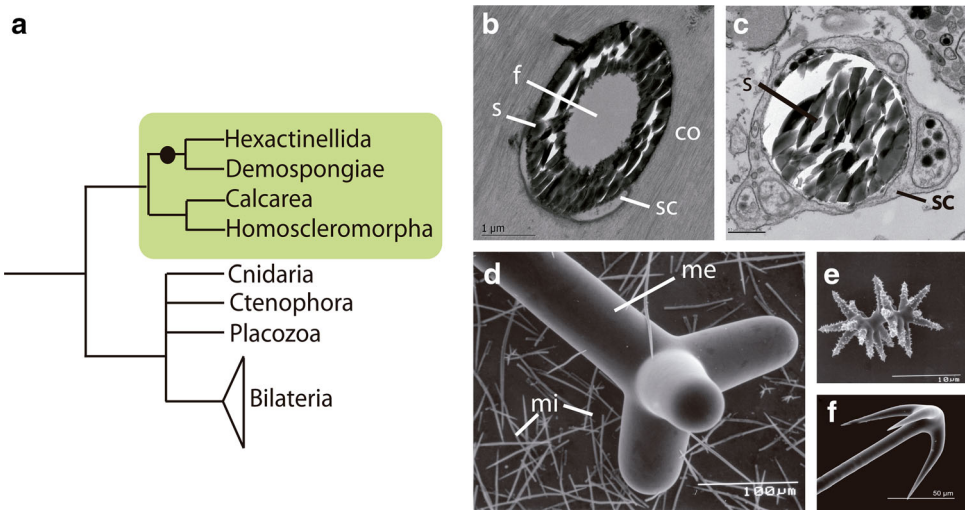

Fig. 1 **a** Phylogenetic scenario of sponge evolution modified from Nosenko et al. (2013) and appearance of *silicatein* (*black dot*). **b**, **c** Transmission electron micrographs of intracellular formation of sponge spicules showing a cross-section of sclerocytes (sc) forming spicules (s) around the axial filament (f). Note the collagen matrix (co) surrounding the sclerocytes in **b**. **d**–**f** Scanning electron micrographs of megasclere (me) spicules (**d**) and microsclere (mi) spicules (**d**–**f**)

making it difficult to assess the suitability of cathepsins for inferring invertebrate evolution. The relationships of sponge silicateins and cathepsins with the cathepsins of other animals remain virtually unaddressed. Deciphering those patterns may also help to improve the current understanding of the relationships within Porifera, given that the standard phylogenetic markers (i.e., *cytochrome c oxidase subunit I*, *28S* and *18S rRNA*) are inconclusive (Cárdenas et al. 2012; Wörheide et al. 2012; Nosenko et al. 2013). Therefore, this study aims to infer the evolution of cathepsins at the leap from protists to animals and further explore the phylogeny of the cathepsine-derived *silicatein*, a unique genetic machinery responsible for the biosilicifying activity in the phylum Porifera.

Methods

Identification of Cathepsin and Silicatein Genes in Sponges: Transcriptomic and Genomic Datasets

Taxon Sampling

For the generation of cDNA libraries from RNA, we designed a taxon sampling comprising species with silica spicules from the 3 "siliceous sponge" lineages: (1) Hexactinellida: *Aphrocallistes vastus* (Vancouver Island, Canada); (2) Demospongiae: *Chondrilla nucula* (Bocas del Toro, Panama), *Petrosia ficiformis* (Catalonia, Spain), *Spongilla lacustris* (Vancouver Island, Canada), *Pseudospongosorites suberitoides* (Florida, USA), and (3) Homoscleromorpha: *Corticium candelabrum* (Catalonia, Spain). For a comparison, we also included a species, *Ircinia fasciculata* (Catalonia, Spain), lacking silica spicules despite belonging to the "siliceous" Demospongiae lineage. A representative of the Calcarea sponge lineage (see Fig. 1a), bearing calcareous rather than siliceous spicules, was also considered: *Sycon coactum* (Vancouver Island, Canada).

For genomic DNA, we collected tissue samples of ca. 2 cm^2 for the siliceous demosponge *Petrosia ficiformis*, as well as several demosponges lacking siliceous spicules: *Spongia lamella*, *Ircinia fasciculata*, *Ircinia oros*, *Aplysina aerophoba*, and *Chondrosia reniformis* from the Catalan coast of Spain. We rinsed tissue pieces at least three times in ethanol and preserved them in 100 % ethanol at −20 °C.

RNA Extraction and cDNA Library Preparation

In order to avoid contamination from sponge-associated microfauna, tissues were carefully but rapidly cleansed under a stereomicroscope to avoid RNA degradation in an RNAse-free, cooled environment using iced dishes. Samples were maintained alive within 1–5 h after collection, depending on the species, prior to preservation. Excised tissue fragments ranged approximately from 0.25 to 0.5 cm^3 in volume. Between 20 and 80 mg of each sponge was used, except for *Petrosia ficiformis* and *Aphrocallistes vastus*, for which 200 mg were analyzed due to their high spicule/cell ratio (see Riesgo et al. 2012a). Samples were either flash-frozen in liquid nitrogen and immediately stored at −80 °C (*Aphrocallistes vastus, Petrosia ficiformis*, *Ircinia fasciculata*, *Pseudospongosorites suberitoides*, *Spongilla lacustris*, and *Sycon coactum*), or immersed in at least 10 volumes of RNA*later*® at 4 °C for 1 h, placed overnight at −20 °C, and subsequently stored in the same buffer at −80 °C until processed (*Corticium candelabrum* and *Chondrilla nucula*).

Two RNA extraction methods were used, depending on the tissue type: (1) total RNA extraction with standard Trizol protocols, followed by mRNA purification using the Dynabeads® mRNA Purification Kit for mRNA Purification from Total RNA preps (Invitrogen, USA) for *Corticium candelabrum*, and (2) direct mRNA extraction for all other species, using the Dynabeads® mRNA DIRECT™ Kit (Invitrogen, USA). Detailed protocols are available elsewhere (Riesgo et al. 2012a). Quantity and quality (purity and integrity) of the resulting mRNA was assessed by three different methods (Agilent Bioanalyzer, Nanodrop, and a Qubit® Fluorometer), as further detailed in Riesgo et al. (2012a). No differences in results should be attributed to the extraction method (see Riesgo et al. 2014).

Next-Generation Sequencing and Assembly

Next-generation sequencing was carried out using the platforms Illumina GAII and HiSeq (Illumina, Inc., San Diego, California, USA) at the FAS Center for Systems Biology, Harvard University. The protocol for cDNA synthesis and library construction is detailed elsewhere (Riesgo et al. 2014). Removal of low-quality raw reads was done using a threshold for Phred scores of 30 with TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and visualized in FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Then, de novo assembly was conducted with Trinity (Grabherr et al. 2011). Quality and completeness of the transcriptomic datasets are described elsewhere (Riesgo et al. 2012b, 2014).

Sequence Annotation

For each species, contigs shorter than 300 bp were removed as most did not return BLAST hits in previous studies (Riesgo et al. 2012b, 2014). The remaining contigs were mapped against a selection of the non-redundant (*nr*) NCBI database (proteins of Metazoa) using the blastx

program of the BLAST suite (hereafter called automated searches). All searches were conducted with blastall (Altschul et al. 1997) using an e-value cut-off of 1e−5.

In order to verify the affiliation of the *cathepsin* and *silicatein* sequences retrieved from the automated blastx searches, we conducted a local blast using tblastn engines incorporated in CLC Genomics workbench, using the contig lists as databases and 1–3 amino acid sequences of closely related metazoans as queries. The open reading frames for each nucleotide sequence were obtained using ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/) and the subsequent amino acid sequences were re-blasted using the blastp suite available at NCBI.

Sequences that showed maximum similarity (over 70 %) against bacteria, protozoans, or fungi were discarded. Likewise, sequences shorter than 100 amino acids were not considered. The translated nucleotide sequences were implemented in SMART (http://smart.embl-heidelberg.de/) to establish the location of domains, and signal peptides were determined using HMMer searches on the SMART database and also PFAM searches (Letunic et al. 2012). The 3D reconstructions were done with the software Phyre2 (Kelley and Sternberg 2009).

We also assessed the presence/absence of the same set of cysteine proteases in four different unicellular eukaryote species (Ichthyosporea: *Sphaeroforma arctica*; Filasterea: *Capsaspora owczarzaki*; Choanoflagellida: *Monosiga brevicollis*, and *Salpingoeca rosetta*) to determine whether some may have been acquired in sponges or whether they were already present in these unicellular organisms. We selected these four species of protists because they are known as the closest relatives to Metazoa and their complete genomes were already sequenced and available (http://www.broadinstitute.org/annotation/genome/multicellularity_project). To determine protein identity, we used the same gene targets and searched in the genomes with the blastp engine implemented in the Broad institute website using default settings.

DNA Extraction and Amplification of Genes Encoding for Silicatein

DNA was extracted using the Qiagen DNeasy Blood and Tissue kit (Qiagen, Valencia, CA, USA), following the manufacturer's instructions. Six sets of degenerated and specific primers (Supplementary File 1) were designed based on the *silicatein* sequences obtained from NCBI. When using degenerated primer sets, two gradient-nested PCR were carried out, the second one using 3 µL of the PCR product obtained during the first PCR and diluted in 47 µL of molecular-grade dH_2O. Both PCRs involved the following program: 2 min at 94 °C, 35 cycles of 94 °C for 30 s, 30 s at 45–50 °C, 2 min at 72 °C, and a final extension step of 7 min at 72 °C. When using specific primers, the PCR program consisted of 2 min at 94 °C, 35 cycles of 94 °C for 30 s, 30 s at 50 °C, 2 min at 72 °C, and a last extension step of 7 min at 72 °C. PCR products were sequenced at Macrogen, Inc. (Seoul, South Korea).

In addition, libraries of genomic DNA from *Petrosia ficiformis*, *Ircinia fasciculata,* and *I. oros* were obtained using 2 µg of high-quality DNA (A260/280 > 1.8) and the GS FLX Titanium Rapid Library Preparation Kit and GS FLX Titanium Rapid Library MID Adaptors Kit. The libraries were run using the Roche Life Sciences 454 GS FLX + System at the Genomics Unit of the Scientific and Technologic Center of the University of Barcelona. BLAST searches of the resulting read files (220,876 for *I. fasciculata*; 126,213 for *I. oros*; and 394,841 for *P. ficiformis*) were executed with CLC Genomics Workbench 5.1 using the tblastn suite.

Phylogenetic Analysis of Cathepsins and Silicateins

Translated amino acid sequences of *cathepsin* and *silicatein* genes were aligned using MUSCLE ver. 3.6 (Edgar 2004) and used for the subsequent phylogenetic analyses. ProtTest ver. 2.4 (Abascal et al. 2005) was used to select the best-fit model of amino acid substitution (LG + Γ + I) under the Akaike Information Criterion (AIC) (Posada and Buckley 2004). In addition, sequences of cathepsin O proteins from sponges were aligned with cathepsin O sequences downloaded from NCBI and also extracted from the transcriptomes of *Chiton olivaceus* (SRX205322), *Cephalothrix hongkongiensis* (SRX205320), *Sipunculus nudus* (SRR619011), *Hormogaster samnitica* (SRX205145), and *Alipes grandidieri* (SRX205685). Alignments of cathepsin O proteins were performed using the same procedures as above.

Maximum likelihood analyses of entire dataset and also that containing only cathepsin O sequences were conducted in RAxML ver. 7.2.7 (Stamatakis 2006). For the likelihood searches, a unique LG model of amino acid substitution with corrections for a discrete gamma distribution and invariant sites (LG + Γ + I) was specified, according to the AIC analysis, and 500 independent searches were conducted. Nodal support was estimated via 500 replicates of the rapid bootstrap algorithm of Stamatakis et al. (2008) using the PROTGAMMAI model. Bootstrap resampling frequencies were thereafter mapped onto the optimal tree resulting from the independent searches.

Bayesian inference analyses of the entire dataset were conducted in MrBayes ver. 3.1.2 (Huelsenbeck and Ronquist 2005) with a unique WAG model of amino acid substitution and corrections for a discrete gamma distribution and a proportion of invariant sites (WAG + Γ + I; Yang 1996). WAG + G + I was the second best-fit model

suggested by ProtTest but the LG model was not implemented in MrBayes ver. 3.1.2. Two runs, each with three hot and one cold chains, were conducted in MrBayes for 20 million generations, sampling every 3000th generation and using random starting trees. The analysis was run twice and 25 % of the runs were discarded as burn-in after checking for stationarity with Tracer v.1.4 (http://beast.bio.ed.ac.uk/Tracer). The remaining trees were combined to find the maximum a posteriori probability estimate of the phylogeny. The topology of the resulting trees was rooted according to results obtained by several studies that provided structural and phylogenetic evidence for the existence of two distinct subfamilies, the L-like and the B-like subfamilies, distinguished by the presence and absence of the ERFNIN motif, respectively (Karrer et al. 1993; Berti and Storer 1995; Uinuk-ool et al. 2003).

Distance values among Bayesian inference and maximum likelihood trees based on Robinson-Foulds metric (Steel and Penny 1993) were calculated with the *phangorn* package in R version 3.0.3 (R Development Core Team 2011). RF values were rescaled using the formula RF/N-2, where N is the number of terminals used (188), in order to provide a rough estimate of the number of terminals whose position changed from tree to tree.

Sequences derived from the present study were deposited in GenBank with accession numbers shown in Supplementary File 2. Accession numbers of *cathepsin* sequences from other metazoans and protists retrieved from GenBank to conduct phylogenetic analysis are also shown in Supplementary File 2.

Identity and Similarity Analyses of Cathepsin and Silicatein Proteins

An alignment of cathepsin and silicatein proteins of sponges ($n = 74$) was implemented in MatGat (Campanella et al. 2012) to obtain pairwise levels of similarity and identity using the Myers and Miller global alignment algorithm and the BLOSUM62 matrix, allowing gaps in 12 sites. Identity is defined as the degree of identical correspondence between two sequences when no gaps between the sequences are allowed, assuming similarity of function when values are over 25 %. In turn, similarity refers to the degree of resemblance between sequences, which is dependent on the identity. Within and between groups (silicatein and cathepsin L), mean distances were computed with MEGA 5.2 (Tamura et al. 2011) using the bootstrap variance estimation method with 100 replications and the two only available models in MEGA, Poisson and JTT, with uniform rates among sites and pairwise deletion.

Results

Cathepsin and Silicatein Genes: Presence/Absence and Structure

Non-lysosomal *cathepsins M*, *J*, *Q*, *R*, *S*, and *K* were not found in our transcriptomic datasets. Lysosomal *cathepsins B, C, Z, O*, and *L* appeared in protist lineages (Figs. 2, 3, 4), but while *cathepsins B*, *L*, and *Z* were obtained from all four protist species, *cathepsin O* was only observed in the choanoflagellates *Monosiga brevicollis* and *Salpingoeca rosetta*, and *cathepsin C* in *Sphaeroforma arctica* and *Capsaspora owczarzaki* (Fig. 2). Among the recovered *cathepsins*, none was exclusive to protists. In contrast, *cathepsins H* and *F* appeared to be exclusive metazoan acquisitions (Fig. 2), with *F* present in all four major sponge lineages (i.e., Demospongiae, Hexactinellida, Homoscleromorpha, Calcarea), but *H* missing in our representative of the class Calcarea (Table 1; Fig. 2).

In the selected sponge species, we retrieved a total of eight cysteine proteases (*cathepsins B*, *C*, *H*, *L*, *F*, *O*, *Z*, and *silicatein*), although with inconsistent patterns of presence/absence and duplication patterns across the phylum (Table 1). In addition, 31 sequences tentatively assigned to an *aspartyl protease* (i.e., *cathepsin D*) were found (Table 1), constituting another plausible metazoan acquisition. In some cases, the different gene copies might correspond to distinct alleles rather than paralogs, but without a reference genome to map, the position of the genes in the same location (alleles) or different locations (paralogs) remains impossible to assess it at this point. Accordingly, even though our approach might not recover the paralogy with 100 % certainty, we were not able to provide more specific assignations, and assumed that since sequences were very divergent among them (see Supplementary File 3), in most cases, we might have recovered paralogs rather than highly divergent alleles. The *cathepsin* sequences identified using automated searches with blastx involved many fragments of unassembled genes and also sequences derived from sponge bacterial and eukaryotic symbionts (not shown) that were later discarded. In many instances, both automated and local searches retrieved the same number of sequences, but often the automated retrieval outperformed the local searches (Table 1). Although our transcriptomic datasets showed high parameter values for quality descriptors (see Supplementary File 4), absences of genes could not be interpreted with reliability as true absences.

Out of the five demosponges whose transcriptome was completed in this study, the *silicatein* gene was found expressed in three (Table 1). Surprisingly, *silicatein* was not retrieved from a fourth siliceous demosponge, *Chondrilla*

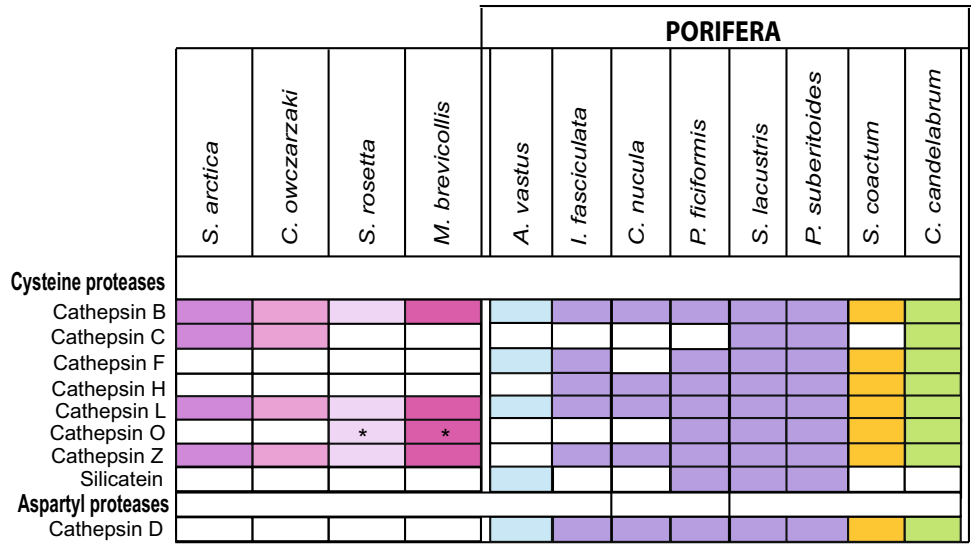

Fig. 2 Diagram comparing the occurrence of *cathepsin* genes in four protist lineages (data obtained from genomes), and sponges (data obtained from ncbi databases and our transcriptomic datasets). *Asterisks* indicate very divergent sequences in protists that were not further used for phylogenetic reconstructions of Fig. 4a

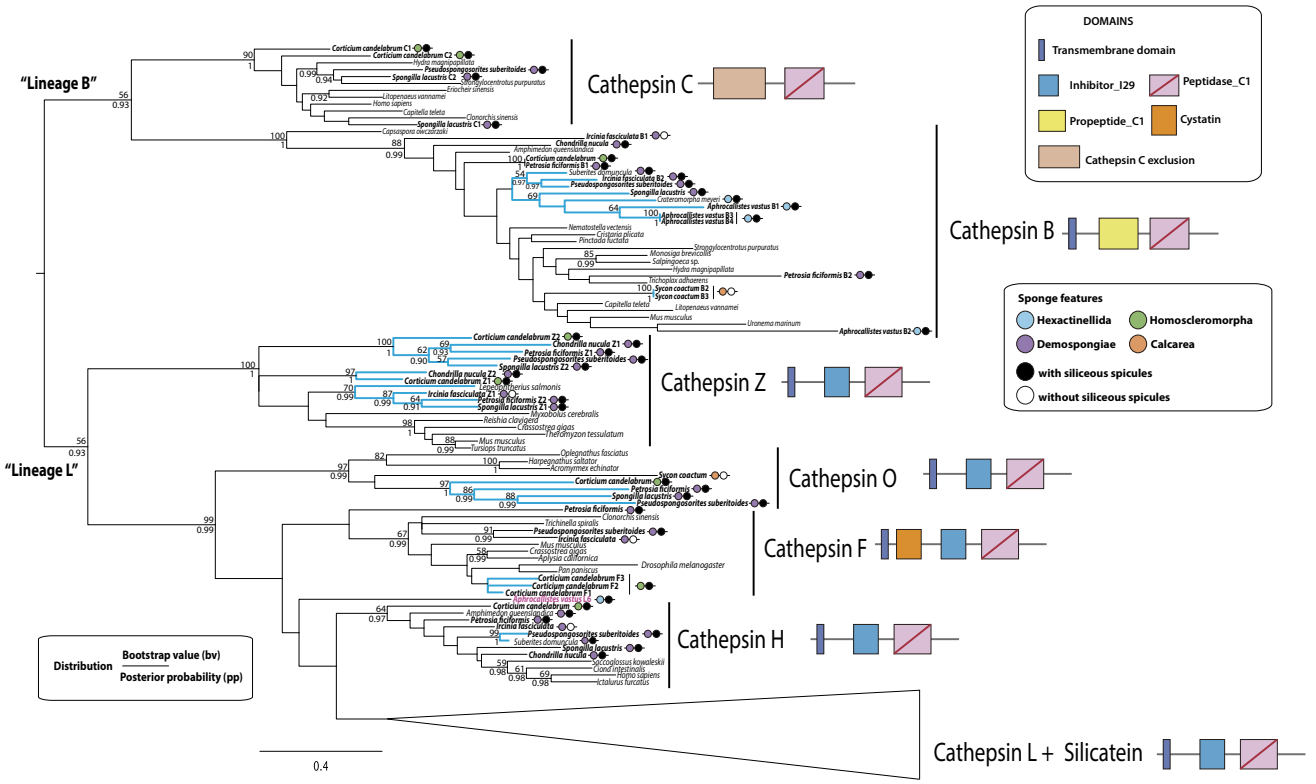

Fig. 3 Phylogenetic reconstruction of the *cathepsins B*, *C*, *Z*, *H*, *O*, *L*, and *silicatein* using maximum likelihood with posterior probabilities shown in clades also supported by Bayesian inference. Only bootstrap and posterior probabilities over 70 % and 0.9, respectively, are shown. Sequences obtained in this study are in *bold letters*. Clades containing only sponge sequences are shown in *blue*. Sequences from *cathepsin L* and *silicatein* are collapsed and shown in detail in Fig. 4. The sequence of *cathepsin L6* from *Aphrocallistes vastus* is shown in *red* to highlight its clustering outside *cathepsin L* (Color figure online)

nucula, despite this species being known to elaborate a dense skeleton of silica spicules. It was also puzzling that *silicatein* expression was found neither in the heavily silicified hexactinellid *Aphrocallistes vastus* nor in the homoscleropmorph *Corticium candelabrum*, each representing in our analyses major lineages of siliceous sponges. To make the evolutionary scenario even more complicated, a demosponge from a subclade that lacks siliceous spicules (*Spongia lamella*) provided a short, highly divergent *silicatein* sequence, retrieved when using

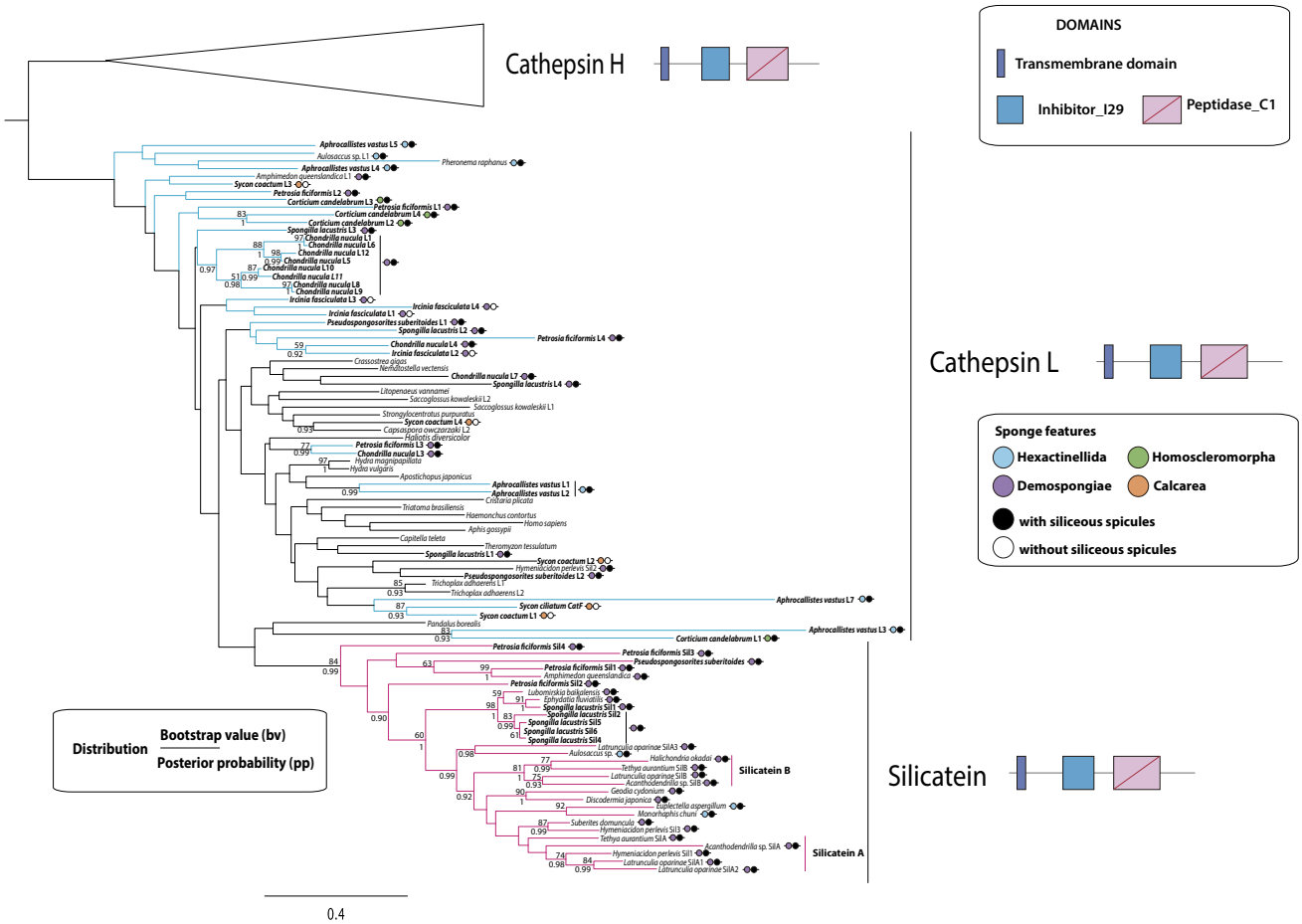

Fig. 4 Phylogenetic reconstruction of the *cathepsins H*, *L*, and *silicatein* using maximum likelihood with posterior probabilities shown in clades also supported by Bayesian inference. Only bootstrap and posterior probabilities over 70 % and 0.9, respectively, are shown. Sequences obtained in this study are in *bold letters*. Clades containing only sponge sequences are shown in *blue*. Sequences from *cathepsin H* are collapsed and shown in detail in Fig. 3 (Color figure online)

specific primers (fragment of 331 bp). The open reading frame of such fragment was too short to be included in our phylogenetic analysis (44 aa), but since the retrieved amino acid sequence was highly similar to two demosponge *silicateins* previously reported from *Petrosia ficiformis* (Q6YD92.1) and *Amphimedon queenslandica* (XP_003383103.1), we conducted an additional exploratory phylogenetic analysis with a reduced dataset containing only *cathepsin L* and *silicatein* amino acid sequences from sponges (Supplementary File 5). This exploratory analysis indicated that the *silicatein* amino acid sequence in *Spongia lamella* was quite divergent, being vaguely related to *silicatein* sequences from haplosclerid demosponges (Supplementary File 5). In the 44 amino acids predicted for the *silicatein-like* gene of *S. lamella*, the asparagine residue of the catalytic triad was replaced by a cysteine (Supplementary File 6). Such a substitution should theoretically be enough to render this "*silicatein*" dysfunctional for silicification.

Using SMART, we found that the eight identified cysteine proteases (*cathepsins* and *silicateins*) derived from the transcriptomic datasets had a very similar secondary structure, all having a catalytically inactive peptidase C1 domain. In *cathepsin B*, three domains were identified: a transmembrane domain, a propeptide C1, and a peptidase C1 domain (Figs. 3, 4). *Cathepsin C* encoded for a cathepsin C exclusion domain before the peptidase C1 domain (Figs. 3, 4). The structure of *cathepsins H*, *L*, *O*, and *Z* had in common a transmembrane domain, a domain inhibitor I29 acting as a propeptide, and a peptidase C1 domain (Figs. 3, 4). *Cathepsin F* had a prosegment with structural similarities with cysteine protease inhibitors of the cystatin superfamily before the inhibitor I29 and then a peptidase C1 domain. Since the primary structure of the cystatin-like inhibitor diverged greatly from the domains present in the rest of the *cathepsins* considered in the analysis, we did not include it in further analyses.

Table 1 Occurrence of the different *cathepsins* in the sponge transcriptome datasets identified using automated blastx searches (A) and local tblastn searches in CLC Genomics Workbench (L)

Cathepsin	Hexactinellida		Demospongiae										Homoscleromorpha		Calcarea	
	Aphrocallistes		*Chondrilla*		*Ircinia*		*Petrosia*		*Spongilla*		*Pseudospongosorites*		*Corticium*		*Sycon*	
	A	L	A	L	A	L	A	L	A	L	A	L	A	L	A	L
Cysteine proteases																
B	4	4	2	1	6	2	4	2	6	1	1	1	7	1	6	2
C	0	0	0	0	0	0	0	0	1	2	0	1	0	2	0	0
F	1	1	0	0	0	2	1	1	1	1	0	1	3	3	1	2
H	1	0	1	1	1	1	1	1	1	1	1	1	4	1	0	0
L	7	7	11	11	7	4	8	4	18	4	2	2	16	4	12	4
O	0	0	0	0	0	0	0	1	0	1	1	1	0	1	0	1
Z	0	0	4	4	0	1	4	2	0	2	1	1	4	2	2	1
Silicatein	0	0	0	0	0	0	4	4	15	5	3	1	0	0	0	0
Aspartyl proteases																
D	2	2	2	2	2	1	6	4	4	4	1	2	9	9	5	5

Only the sequences that showed their maximum similarity against metazoans or those with open reading frames longer than 100 aa were considered for the subsequent phylogenetic analyses

The primary structure resulting from the *silicatein* sequences, for which almost a full-length transcript was obtained, is shown in Supplementary File 6. The catalytic triad is typically formed by cysteine, histidine, and asparagine in cathepsin L sequences, and by serine, histidine, and asparagine in silicateins. However, in the *silicatein 4* of *Petrosia ficiformis*, the first residue of the catalytic triad (i.e., serine) was replaced by a cysteine (Supplementary File 6), a change that is supposed to turn the enzyme into a non-functional configuration for silicification. The catalytic triad was not observed in the remaining sequences derived from our transcriptomic datasets. It was surprising to find *silicatein* only in Demospongiae. The absence of *silicatein* transcripts in the hexactinelid representative is a major mismatch relative to the results of previous work describing the occurrence of *silicatein* in at least some members of the Hexactinellida (Veremeichik et al. 2011).

Identity and Similarity Analyses of Cathepsin and Silicatein Proteins

Using MEGA, we found that the genetic divergence within the cathepsin L group of sequences was 0.53 ± 0.03 under the Poisson model and 0.67 ± 0.05 under the JTT model. Similarly, within the silicatein group of amino acid sequences, the genetic divergence was 0.53 ± 0.03 under the Poisson model and 0.58 ± 0.05 under JTT model. In turn, genetic divergence between cathepsin L and silicatein groups was 0.76 ± 0.05 under the Poisson model and 0.93 ± 0.07 under the JTT model (Supplementary File 3). MatGat provided pairwise comparisons between sequences for identity and similarity that are tabulated in Supplementary File 3. Similarities within the silicatein proteins ranged between 20 and 97.2 %, within cathepsin L varied from 17.3 to 74 %, and finally between silicatein and cathepsin L protein sequences from 14.1 to 78.4 % (Supplementary File 3). The 3D reconstructions also showed a high similarity in the domain organization between the cathepsin L and silicatein proteins (Fig. 5), and especially between cathepsin L4 in *Corticium candelabrum* and silicateins (Fig. 5).

Phylogeny of Cathepsin Proteins in Sponges

Maximum likelihood (ML) and Bayesian inference (BI) recovered similar tree topologies although with some discrepancies; Figs. 3 and 4 show the topology obtained under ML with the bootstrap proportions and posterior probabilities mapped on clades. Additionally, the topology of the tree obtained with BI analysis is provided in Supplementary File 7. The Robinson-Foulds metric was 87 and 0.46 when rescaled, indicating that the position of roughly 46 % terminals differed between trees. Rooting of the ML and BI trees was done on the split of lineages B and L (Figs. 3, 4), based on the presence of ERFNIN motif in cathepsins Z, F, O, H, L, and silicatein (subfamily or lineage L) and absence in cathepsins B and C (subfamily or lineage B). All cathepsins in lineage L shared a similar secondary structure (one transmembrane domain, the inhibitor I29 domain, and the peptidase C1), while cathepsins

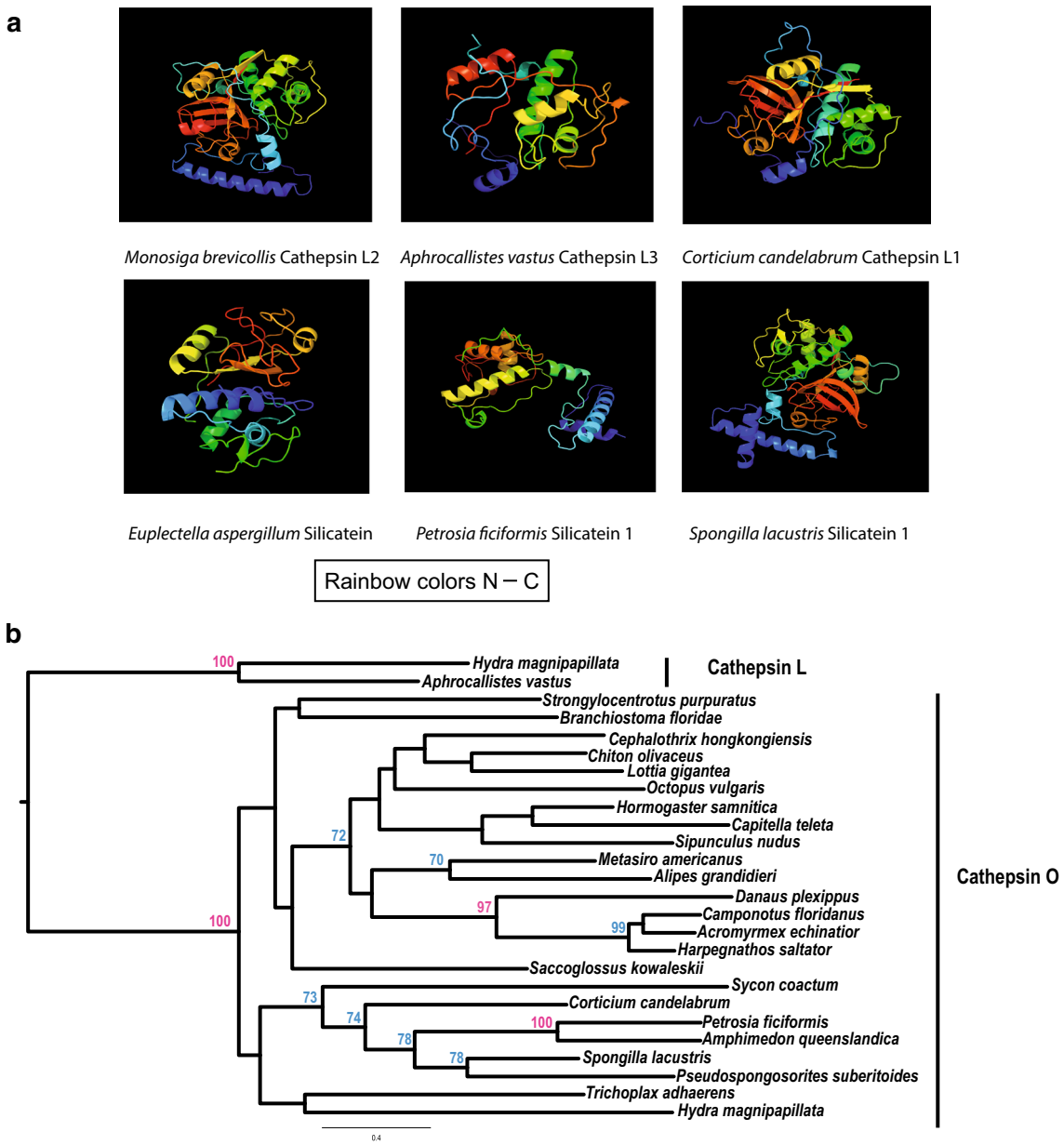

Fig. 5 **a** Three-dimensional reconstructions of *cathepsin L* sequences in the choanoflagellate *Monosiga brevicollis* and the sponges *Aphrocallistes vastus* and *Corticium candelabrum*, and *silicatein* in *Euplectella aspergillum*, *Petrosia ficiformis*, and *Spongilla lacustris*. Similar domains are coded with the same colors. **b** Maximum likelihood reconstruction of relationships within the *cathepsin O* clade. Bootstraps obtained from 500 replicates over 70 are shown on the branches

in lineage B only had the peptidase C1 domain in common (Figs. 3, 4).

Lineage B was clearly divided into two well-supported subgroups (one containing cathepsin B (100 % bootstrap support [BS]; 1.00 posterior probability [pp]), and the other containing cathepsin C (90 % BS, 1.00 pp; Fig. 3). For both subgroups, the phylogenetic relationships of sponge sequences with those of other metazoans were not resolved (Fig. 2). Within lineage L (Fig. 3), cathepsin Z was also well-supported (100 % BS, 1.00 pp) and made the earliest diverging group, followed by a clade supported with 99 % BS and 0.99 pp, respectively that contained cathepsins O, F, and H, and a large clade containing cathepsin L and silicatein (Figs. 3, 4). The internal relationships between cathepsin Z sequences were in most cases robustly supported (>0.95 pp) by Bayesian posterior probabilities (Fig. 3), and included three subclades with internal compositions that do not recover the major accepted sponge lineages (Fig. 3). Sequences of cathepsin O were supported by high values of bootstrap and posterior probabilities (>80 % BS and >0.99 pp, respectively) and the clade was further divided into two subclades (Fig. 3), one containing

only sponge sequences and another containing sequences from other metazoans (arthropods and vertebrates). Further analyses of the clade containing cathepsin O (including more sequences from metazoans retrieved from available transcriptomic datasets) supported monophyly of all sponges except Hexactinellida (which did not contain a cathepsin O sequence) and recovered moderate support (72 % BS) for most bilaterian groups (Fig. 5). Sequences of cathepsin F, excluding that of Petrosia ficiformis, were grouped with 67 % BS from ML and 0.99 pp (Fig. 3 and Supplementary File 7), but within the clade, sponge sequences were mixed with those of other metazoans and not forming a sponge clade (Fig. 3). The group of cathepsin H was supported by 0.97 pp in the BI (Fig. 3); however, bootstrap values from ML analysis only supported a clade formed by two sponge sequences, the hadromerids *Suberites domuncula* and *Pseudospongosorites suberitoides* (Fig. 3).

The relationships among the different sequences of the cathepsin L group remained poorly resolved and incorporated a mix of sequences from sponges and other animals (Fig. 4 and Supplementary File 7). In addition, the putative cathepsin L sequence (cathepsin L6) obtained for *Aphrocallistes vastus* did not fall into the cathepsin L cluster in the ML analysis, branching off earlier than cathepsin H (Fig. 4). In contrast, within the cathepsin L group, all the silicatein sequences formed a supported group (84 % BS; 0.99 pp), except for the silicatein 3 of *Hymeniacidon perlevis* that clustered out of the silicatein group (Fig. 4 and Supplementary File 7). The four major sponge classes were not recovered within the silicatein clade (Fig. 4 and Supplementary File 7). The silicatein sequences of the haplosclerids *Petrosia ficiformis* (4 sequences) and *Amphimedon queenslandica* (1 sequence), and that of the hadromerid *Pseudospongosorites suberitoides* diverged earlier than the rest (Fig. 4), while the remaining silicatein sequences formed a derived subclade supported by a posterior probability of 1 in the BI analysis (Fig. 4 and Supplementary File 7). Within the latter subclade, the silicateins of freshwater sponges (family Spongillidae) made a small cluster, not related to marine haplosclerids (i.e., against the old view postulating that freshwater sponges derived from marine haplosclerids) but sister to a larger, complex cluster (supported only by BI) that contained intermingled sequences from hexactinellids and several non-haplosclerid demosponges (Fig. 4 and Supplementary File 7).

Discussion

Cathepsins

Our study suggests that all the investigated protists possess a small number of *cathepsins*, including *B*, *C*, *L*, and *Z*, which were later expanded in metazoans. In sponges, early branching metazoans, *cathepsins B*, *C*, *F*, *H*, *L*, *O*, and *Z* were already present, also presenting a novel acquisition, *silicatein*, a *cathepsin L*-derived enzyme. Within each *cathepsin* lineage, many different enzymes appeared through duplications during animal diversification, and such enzymes specialized to undertake quite different functions (Uinuk-ool et al. 2003). Interestingly, our analyses indicated that *cathepsin H* evolved within the L lineage, similarly to what was found by Uinuk-ool et al. (2003), although the *cathepsin H* sequences form a well-supported clade within the L group.

In sponges, the levels of gene duplication appeared to be higher for *cathepsins B* and *L* than for the rest of *cathepsins* (and *silicateins*), as we found up to four paralogs of *cathepsin B* in *Aphrocallistes vastus* and eleven of *cathepsin L* in *Chondrilla nucula*. Only *cathepsins F* and *H* appeared to be novel metazoan acquisitions and they occurred in all four major sponge lineages, while *cathepsin H* was not retrieved from our calcareous species. For most of the six major *cathepsin* clades recovered in our analyses, with the exception of *cathepsin O*, taxon relationships were poorly supported, probably suggesting that they are not reliable markers for phylogeny inference. In summary, despite *cathepsins* being ubiquitious enzymes in opisthokont (Fungi, Choanozoa, Metazoa) lineages, their phylogenetic signal was weak, failing to recover major lineages of either metazoans or sponges. However, screening of other protist genomes in the future might shed more light into the origin and acquisition of *cathepsins* in these ancient lineages.

The functional roles of most *cathepsins* in both early metazoans (sponges, ctenophores, cnidarians, and placozoans) and most other non-mammal animals remain largely unknown. The assessment of the potential role of *cathepsins* from lineage B (B and C) in the reproduction and immunity of sponges (as it occurs in other animals, Cha et al. 1999; Niu et al. 2013) appears as a promising field. Currently, it seems that enzymes in lineage L are the ones with major importance for sponge biology given their involvement in spiculogenesis (Kozhemyako et al. 2010), but most of their proteolytic targets and their specific biological functions remain unclear. Sponges *cathepsin L* and *silicatein* have received some attention (Krasko et al. 1997, 2000; Shimizu et al. 1998; Cha et al. 1999; Müller et al. 2003; Schröder et al. 2005, 2006, 2007), and it has been hypothesized that *cathepsin L* might be involved in the elaboration of the collagen templates necessary for spicule formation (Kozhemyako et al. 2010) since it is involved in different aspects of collagen cleavage in other animals (Kakegawa et al. 1993; Felbor et al. 2000; Dalton et al. 2003; Lustigman et al. 2004).

Silicateins

Our results, in agreement with previous studies, corroborate that *cathepsin L* gave rise to *silicatein*, which to

date has only been found in the phylum Porifera. Enzymes encoded by *cathepsin L* possess an active site nucleophilic cysteine residue at position 25 that attacks scissile peptide bonds (Brocklehurst et al. 1987). This cysteine residue is part of a catalytic triad (cysteine, histidine, and asparagine) unique to *cathepsins*. It is thought that the gene *cathepsin L* duplicated and mutated the catalytic cysteine to a serine (Cha et al. 1999), resulting in a new functional enzyme, *silicatein*. The silicatein enzyme is able to polycondensate dissolved silicon into amorphous biogenic silica, the basic material from which siliceous sponge spicules are made (Shimizu et al. 1998; Cha et al. 1999; Schröder et al. 2007; Wang et al. 2010, 2012a, b). Because *silicatein* has been derived from a lysosomal proteolytic enzyme, we propose here that the intracytoplasmic "silica deposition vesicle," in which the silicatein filament is assembled to start silicification within the sponge silicifying cells (sclerocytes), is likely derived from a lysosome.

The functionality of *silicatein* may, however, rely on mechanisms more complex than currently thought. It is worth noting that the *silicatein 4* sequence from the demosponge *Petrosia ficiformis* and the *silicatein* sequence from the glass sponge *Aulosaccus* sp. (Veremeichik et al. 2011) that were previously classified as true *silicateins* do not possess serine, but cysteine in the first residue of the catalytic triad (Supplementary File 6), which would potentially prevent them from being functional *silicateins*. Furthermore, Fairhead et al. (2008) demonstrated that the single mutation of cysteine to serine did not confer significant levels of silica-production activity. In order to obtain high-silicification activity, the residues of either side of the cysteine had to be mutated as well. Additionally, *silicateins* also differ from sponge *capthepsins L* in having (although not in all species) long clusters of serine residues in the former. This serine stretch that comes close to the catalytic histidine has been postulated to participate somehow in the binding of the substrate/product (Müller et al. 2007a, b, 2008a). Yet additional conditions, such as association with galectin, might be required for effective spicule formation (Fairhead et al. 2008). Since we were able to identify four *silicatein* genes in *P. ficiformis*, including the previously known copy (Pozzolini et al. 2004), we hypothesize that different genes may play alternate roles in the process of spicule formation in *Petrosia ficiformis* and could operate alternatively/consecutively during the putative intracellular and extracellular stages of spicule formation.

Although the presence and morphology of silica spicules have been used for about two centuries as a major taxonomic tool to establish the taxonomy within Porifera, the phylogenetic signal of *silicatein*, the main gene known to date involved in spicule formation, remains unclear. If sponges were monophyletic (see recent discussions in Nosenko et al. 2013), *silicatein* should have arisen before the split of the two major siliceous clades, that is, hexactinellids and demosponges (Fig. 1). However, there are several subclades within demosponges unable to produce silica (e.g., orders Dendroceratida, Dictyoceratida, and most members of Verongimorpha), suggesting that the ability to produce silica may have been lost secondarily. Such loss would agree with claims that siliceous sponges suffer a chronic limitation by dissolved silicon in the modern ocean, outcompeted by diatoms (Maldonado et al. 1999, 2010, 2012; Maldonado 2009). On the other hand, we have not identified here any *silicatein* gene in hexactinellids, homoscleromorphs, or in the demosponge *Chondrilla nucula*, which produce siliceous spicules. Although the genome of the aspiculate homoscleromoprh *Oscarella carmela* (unable to make siliceous spicules) revealed that this species does not contain a *silicatein* gene (Hemmrich and Bosch 2008; Nichols et al. 2012), other aspiculate sponges might still possess a copy of the gene. We found a small fragment of *silicatein* in the aspiculate dictyoceratid demosponge *Spongia lamella*, and it showed the substitution of the original asparagine by a cysteine, which could render it not functional for silicification (Supplementary File 6). A partial sequence was also retrieved from the dendroceratid *Acanthodendrilla* sp. by Kozhemyako et al. (2010), which was also too small to contain the complete catalytic triad. Therefore, the presence of *silicatein* gene fragments appears to be a common occurrence in aspiculate demosponge species. These results suggest that a modification of the catalytic triad and possibly of other residues in the *silicatein* gene sequences of these species may have impaired their function and thus their ability to produce biogenic silica. In order words, the ability to silicify would have been secondarily lost in these demosponges.

Because the relationships between the four major sponge lineages remain unsettled, it is difficult to retrace plausible gains and losses of *silicatein* genes within Porifera. If sponges were monolophyletic with its four major lineages grouped in two subclades, i.e., [(Demospongiae, Hexactinellida), (Homoscleromorpha, Calcarea)], as suggested by recent studies (Nosenko et al. 2013; Ryan et al. 2013), our results herein suggest that the functional *silicatein* gene was not present in the common ancestor of all Porifera, but rather would have appeared in the common ancestor of Demospongia and Hexactinellida. As *silicatein* has been found in only some members of Hexactinellida (e.g., Müller et al. 2008a, b; Veremeichik et al. 2011; Schloßmacher 2012), the absence of transcripts in other investigated species (Veremeichik et al. 2011; this study) may also be explained by either unrecognized difficulties to detect expressed transcripts or by expression not being captured in the relatively small body portions and narrow time windows of our sampling approach.

If *silicatein* was acquired in a common ancestor to hexactinellids and demosponges, the spicules observed in some homoscleromorphs may result from an independent acquisition. Indeed, the internal structure of the homoscleromorph spicules has been reported as unconventional relative to that of the remaining studied types, and the spicules, in at least some species, can be elaborated not only by the conventional mesohyl specialized cells (sclerocytes) but also by the pinacocytes (Maldonado and Riesgo 2007). Given the abundance of *cathepsin L* paralogs in homosclerophorids (up to 16 retrieved when using automated BLASTX searches and 4 using local BLAST searches), the absence of *silicatein* could indicate that an alternative silicification system had evolved in this sponge lineage. Alternatively, the absence of *silicatein* could also be the consequence of extensive modifications to an original *cathepsin L* that we were unable to recognize. Indeed, this could be the reason behind the high similarity found between the protein structure of cathepsin L4 in *Corticium candelabrum* and silicateins.

Acknowledgments We are grateful to Sergi Taboada, Joan Mora, and Alicia R. Pérez-Porro who helped with field sampling, and Sally Leys for providing data. This project was funded by a Marie Curie IOF Grant (PIOF-GA-2009-237219) and a Juan de la Cierva fellowship to AR, two grants (BFU2008-00227/BMC and CTM2012-37787) awarded by the Spanish Ministry of Economy and Competitiveness to MM, and the Catalan Government Grant 2014SGR-336 for Consolidated Research Groups. Illumina sequencing was supported by funds from the Museum of Comparative Zoology and by the Center from Systems Biology, Harvard University, to GG.

References

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

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Berti PJ, Storer AC (1995) Alignment/phylogeny of the papain superfamily of cysteine proteases. J Mol Evol 246:273–283

Brocklehurst K, Kowlessur D, O'Driscoll M, Patel G, Quenby S, Salih E, Templeton W, Thomas E, Willenbrock F (1987) Substrate-derived two-protonic-state electrophiles as sensitive kinetic specificity probes for cysteine proteinases. Activation of 2-pyridyl disulphides by hydrogen-bonding. Biochem J 244:173

Bühling F, Fengler A, Brandt W, Welte T, Ansorge S, Nagler DK (2002) Cathepsin K expression in human lung. In: Langner J, Ansorge S (eds) Cellular peptidases in immune functions and diseases 2, vol 477. Springer, New York, pp 41–254

Campanella JJ, Bitincka L, Smalley J (2012) MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4(1):29

Cárdenas P, Pérez T, Boury-Esnault N (2012) Sponge systematics facing new challenges. Adv Mar Biol 61:179–209

Cha JN, Shimizu K, Zhou Y, Christiansen SC, Chmelka BF, Stucky GD, Morse DE (1999) Silicatein filaments and subunits from a marine sponge direct the polymerization of silica and silicones in vitro. PNAS 96:361–365

Dalton JP, Neill SO, Stack C, Collins P, Walshe A, Sekiya M, Doyle S, Mulcahy G, Hoyle D, Khaznadji E (2003) *Fasciola hepatica* cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines. Int J Parasitol 33:1173–1181

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Fairhead M, Kowatz T, McMahon SA, Carter LG, Oke M, Johnson KA, Liu H, Naishmith JH, van der Walle CF (2008) Crystal structure and silica condensing activities of silicatein α-cathepsin L chimeras. Chem Commun 1765–1767

Felbor U, Dreier L, Bryant RA, Ploegh HL, Olsen BR, Mothes W (2000) Secreted cathepsin L generates endostatin from collagen XVIII. EMBO J 19:1187–1194

Funayama N, Nakatsukasa M, Kuraku S, Takechi K, Dohi M, Iwabe N, Miyata T, Agata K (2005) Isolation of *Ef silicatein* and *Ef lectin* as molecular markers of sclerocytes and cells involved in innate immunity in the freshwater sponge *Ephydatia fluviatilis*. Zool Sci 22:1113–1122

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol 29(7):644–652

Gröger C, Sumper M, Brunner E (2008) Silicon uptake and metabolism of the marine diatom *Thalassiosira pseudonana*: solid-state ^{29}Si NMR and fluorescence microscopic studies. J Struct Biol 161:55–63

Hemmrich G, Bosch TCG (2008) Compagen, a comparative genomics platform for early branching metazoan animals reveals early origins of genes regulating stem cell differentiation. BioEssays 20:1010–1018

Horn M, Dolečková-Marešová L, Rulíšek L, Máša M, Vasiljeva O, Turk B, Gan-Erdene T, Baudyš M, Mareš M (2005) Activation processing of cathepsin H impairs recognition by its propeptide. Biol Chem 386:941–947

Huelsenbeck JP, Ronquist F (2005) Statistical methods in molecular evolution. Springer, New York, pp 83–226

Inui T, Ishibashi O, Inaoka T, Origane Y, Kumegawa M, Kokubo T, Yamamura T (1997) Cathepsin K antisense oligodeoxynucleotide inhibits osteoclastic bone resorption. J Biol Chem 272:8109–8112

Kakegawa H, Nikawa T, Tagami K, Kamioka H, Sumitani K, Kawata T, Drobnič-Kosorok M, Lenarčič B, Turk V, Katunuma N (1993) Participation of cathepsin L on bone resorption. FEBS Lett 321:247–250

Kalyuzhnaya O, Krasko A, Grebenyuk V, Itskovich V, Semiturkina N, Solovarov I, Mueller WEG, Belikov S (2011) Freshwater sponge silicateins: comparison of gene sequences and exon-intron structure. Mol Biol 45:567–575

Karrer KM, Peiffer SL, DiTomas ME (1993) Two distinct gene subfamilies within the family of cysteine protease genes. PNAS 90:3063–3067

Kelley LA, Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4(2):363–371

Kozhemyako VB, Veremeichik GN, Shkryl YN, Kovalchuk SN, Krasokhin VB, Rasskazov VA, Zhuravlev YN, Bulgakov VP, Kulchin YN (2010) Silicatein genes in spicule-forming and nonspicule-forming Pacific demosponges. Mar Biotechnol 12:403–409

Krasko A, Gamulin V, Seack J, Steffen R, Schroeder HC, Müller WEG (1997) Cathepsin, a major protease of the marine sponge

Geodia cydonium: purification of the enzyme and molecular cloning of cDNA. Mol Mar Biol Biotechnol 6:296

Krasko A, Lorenz B, Batel R, Schröder HC, Müller IM, Müller WEG (2000) Expression of *silicatein* and *collagen* genes in the marine sponge *Suberites domuncula* is controlled by silicate and myotrophin. Eur J Biochem 267:4878–4887

Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res 40:D302–D305

Liu J, Shi GP, Zhang WQ, Zhang GR, Xu WH (2006) Cathepsin L function in insect moulting: molecular cloning and functional analysis in cotton bollworm, *Helicoverpa armigera*. Insect Mol Biol 15:823–834

Lustigman S, Zhang J, Liu J, Oksov Y, Hashmi S (2004) RNA interference targeting cathepsin L and Z-like cysteine proteases of *Onchocerca volvulus* confirmed their essential function during L3 molting. Mol Biochem Parasitol 138:165–170

Lutgens SPM, Cleutjens KBJM, Daemen MJAP, Heeneman S (2007) Cathepsin cysteine proteases in cardivascular disease. FASEB J 21:3029–3041

Maldonado M (2009) Embryonic development of verongid demosponges supports the independent acquisition of spongin skeletons as an alternative to the siliceous skeleton of sponges. Biol J Linnean Soc 97:427–447

Maldonado M, Carmona MC, Uriz MJ, Cruzado (1999) Decline in Mesozoic reef-building sponges explained by silicon limitation. Nature 401(6755):785–788

Maldonado M, Riesgo A (2007) Intra-epithelial spicules in a homosclerophorid sponge. Cell Tissue Res 328:639–650

Maldonado M, Riesgo A, Bucci A, Rützler K (2010) Revisiting silicon budgets at a tropical continental shelf: silica standing stocks surpass those in diatoms. Limnol Oceanogr 55:2001–2010

Maldonado M, Ribes M, van Duyl F (2012) Nutrient fluxes through sponges: biology, budgets, and ecological implications. Adv Mar Biol 62:113–182

Mort JS, Buttle DJ (1997) Cathepsin B. Int J Biochem Cell Biol 29:715–720

Müller WEG, Krasko A, Le Pennec G, Schröder HC (2003) Biochemistry and cell biology of silica formation in sponges. Microsc Res Tech 62:368–377

Müller WEG, Boreiko A, Wang X, Belikov SI, Wiens M, Grebenjuk VA, Schloßmacher U, Schröder HC (2007a) Silicateins, the major biosilica forming enzymes present in demosponges: protein analysis and phylogenetic relationship. Gene 395:62–71

Müller WEG, Schloßmacher U, Eckert C, Krasko A, Boreiko A, Ushijima H, Wolf SE, Tremel W, Müller IM, Schröder HC (2007b) Analysis of the axial filament in spicules of the demosponge *Geodia cydonium*: different silicatein composition in microscleres (asters) and megascleres (oxeas and triaenes). Eur J Cell Biol 86:473–487

Müller WEG, Wang X, Kropf K, Boreiko A, Schloßmacher U, Brandt D, Schröder HC, Wiens M (2008a) Silicatein expression in the hexactinellid *Crateromorpha meyeri*: the lead marker gene restricted to siliceous sponges. Cell Tissue Res 333:339–351

Müller WEG, Boreiko A, Schloßmacher U, Wang X, Eckert C, Kropf K, Li J, Schröder HC (2008b) Identification of a silicatein (-related) protease in the giant spicules of the deep-sea hexactinellid *Monorhaphis chuni*. J Exp Biol 211:300–309

Nichols S, Roberts BW, Richter DJ, Fairclough SR, King N (2012) Origin of metazoan cadherin diversity and the antiquity of the classical cadherin/b-catenin complex. PNAS 109:13046–13051

Niu D, Jin K, Wang L, Sun F, Li J (2013) Identification of cathepsin B in the razor clam *Sinonovacula constricta* and its role in innate immune responses. Dev Comp Immunol 41:94–99

Nosenko T, Schreiber F, Adamska M, Adamski M, Eitel M, Hammel J, Maldonado M, Müller WEG, Nickel M, Schierwater B (2013) Deep metazoan phylogeny: when different genes tell different stories. Mol Phylogenet Evol 67:223–233

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol 53:793–808

Pozzolini M, Sturla L, Cerrano C, Bavestrello G, Camardella L, Parodi AM, Raheli F, Benatti U, Müller WEG, Giovine M (2004) Molecular cloning of silicatein gene from marine sponge *Petrosia ficiformis* (Porifera, Demospongiae) and development of primmorphs as a model for biosilicification studies. Mar Biotechnol 6:594–603

R Development Core Team (2011) R: a language and environment for statistical computing. The R Foundation for Statistical Computing, Vienna, Austria. http://www.Rproject.org/

Riesgo A, Pérez-Porro AR, Carmona S, Leys SP, Giribet G (2012a) Optimization of preservation and storage time of sponge tissues to obtain quality mRNA for next-generation sequencing. Mol Ecol Resour 12:312–322

Riesgo A, Andrade S, Sharma PP, Novo M, Pérez-Porro AR, Vahtera V, González VL, Kawauchi GY, Giribet G (2012b) Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. Front Zool 9:1–24

Riesgo A, Farrar N, Windsor PJ, Giribet G, Leys SP (2014) The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. Mol Biol Evol 31:1102–1120

Rodriguez-Franco EJ, Cantres-Rosario YM, Plaud-Valentin M, Romeu R, Rodríguez Y, Skolasky R, Meléndez V, Cadilla CL, Meléndez LM (2012) Dysregulation of macrophage-secreted cathepsin B contributes to HIV-1-linked neuronal apoptosis. PLoS One 7:e36571

Ryan JF, Pang K, Schnitzler CE, Nguyen AD, Moreland RT, Simmons DK, Koch BJ, Francis WR, Havlak P, NISC Comparative Sequencing Program, Smith SA, Putnam NH, Haddock SHD, Dunn CW, Wolfsberg TG, Mullikin JC, Martindale MQ, Baxevanis AD (2013) The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. Science 342:1242592

Samuels R, Paterson I (1995) Cuticle degrading proteases from insect moulting fluid and culture filtrates of entomopathogenic fungi. Comp Biochem Physiol B 110:661–669

Schloßmacher U (2012) The relevance of the silica metabolizing enzyme silicatein-alpha to biomineralization and the formation of biogenic silica in siliceous sponges. Mainz, Univ., Diss., 2012

Schröder HC, Boreiko O, Krasko A, Reiber A, Schwertner H, Müller WE (2005) Mineralization of SaOS-2 cells on enzymatically (silicatein) modified bioactive osteoblast-stimulating surfaces. J Biomed Mater Res B: Appl Biomater 75(2):387–392

Schröder HC, Boreiko A, Korzhev M, Tahir MN, Tremel W, Eckert C, Ushijima H, Müller IM, Müller WEG (2006) Co-expression and functional interaction of Silicatein with Galectin. Matrix guided formation of siliceous spicules in the marine demosponge *Suberites domuncula*. J Biol Chem 281:12001–12009

Schröder HC, Natalio F, Shukoor I, Tremel W, Schloßmacher U, Wang X, Müller WEG (2007) Apposition of silica lamellae during growth of spicules in the demosponge *Suberites domuncula*: biological/biochemical studies and chemical/biomimetical confirmation. J Struct Biol 159:325–334

Shimizu K, Cha J, Stucky GD, Morse DE (1998) Silicatein α: cathepsin L-like protein in sponge biosilica. PNAS 95:6234–6238

Sloane BF, Honn KV (1984) Cysteine proteinases and metastasis. Cancer Metastasis Rev 3:249–263

Sloane BF, Dunn JR, Honn KV (1981) Lysosomal cathepsin B: correlation with metastatic potential. Science 212:1151–1153

Söderström M, Salminena H, Glumoff V, Kirschkec H, Arob H, Vuorioa E (1999) Cathepsin expression during skeletal development. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1446.1:35–46

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol 57:758–771

Steel MA, Penny D (1993) Distributions of tree comparison metrics—some new results. Syst Biol 42.2:126–141

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Turk V, Turk B, Turk D (2001) Lysosomal cysteine proteases: facts and opportunities. EMBO J 20:4629–4633

Uinuk-ool T, Takezaki N, Kuroda N, Figueroa F, Sato A, Samonte I, Mayer W, Klein J (2003) Phylogeny of antigen-processing enzymes: cathepsins of a cephalochordate, an agnathan and a bony fish. Scand J Immunol 58:436–448

Veremeichik GN, Shkryl YN, Bulgakov VP, Shedko SV, Kozhemyako VB, Kovalchuk SN, Krasokhin VB, Zhuravlev YN, Kulchin YN (2011) Occurrence of a silicatein gene in glass sponges (Hexactinellida: Porifera). Mar Biotechnol 13:810–819

Wang X, Wiens M, Schröder HC, Hu S, Mugnaioli E, Kolb U, Tremel W, Pisignano D, Müller WEG (2010) Morphology of sponge spicules: silicatein a structural protein for bio-silica formation. Adv Eng Mater 12:B422–B437

Wang X, Schloßmacher U, Wiens M, Batel R, Schröder HC, Müller WEG (2012a) Silicateins, silicatein interactors and cellular interplay in sponge skeletogenesis: formation of glass fiber-like spicules. FEBS J 279:1721–1736

Wang X, Schröder HC, Wiens M, Schloßmacher U, Müller WEG (2012b) Biosilica: molecular biology, biochemistry, and function in demosponges as well as its applied aspects for tissue engineering. Adv Mar Biol 62:231–272

Wiens M, Belikov SI, Kaluzhnaya OV, Krasko A, Schröder HC, Perovic-Ottstadt S, Müller WEG (2006) Molecular control of serial module formation along the apical–basal axis in the sponge *Lubomirskia baicalensis*: silicateins, mannose-binding lectin and mago nashi. Dev Genes Evol 216:229–242

Wörheide G, Dohrmann M, Erpenbeck D, Larroux C, Maldonado M, Voigt O, Borchiellini C, Lavrov DV (2012) Deep phylogeny and evolution of sponges (phylum Porifera). Adv Mar Biol 62:1–78

Yang Z (1996) Maximum-likelihood models for combined analyses of multiple sequence data. J Mol Evol 42:587–596