REVIEW   Open Access

The evolution of metazoan α-carbonic anhydrases and their roles in calcium carbonate biomineralization

Nathalie Le Roy[1], Daniel J Jackson[2*], Benjamin Marie[3], Paula Ramos-Silva[4,5] and Frédéric Marin[5]

Abstract

The carbonic anhydrase (CA; EC 4.2.1.1) superfamily is a class of ubiquitous metallo-enzymes that catalyse the reversible hydration of carbon dioxide. The α-CA family, present in all metazoan clades, is a key enzyme involved in a wide range of physiological functions including pH regulation, respiration, photosynthesis, and biocalcification. This paper reviews the evolution of the α-CA family, with an emphasis on metazoan α-CA members involved in biocalcification. Phylogenetic analyses reveal a complex evolutionary history of α-CAs, and suggest α-CA was independently co-opted into a variety of skeleton forming roles (*e.g.* as a provider of HCO_3^- ions, a structural protein, a nucleation activator, *etc.*) in multiple metazoan lineages. This evolutionary history is most likely the result of multiple gene duplications coupled with the insertion of repetitive or non-repetitive low-complexity domains (RLCDs/LCDs). These domains, of largely unknown function, appear to be lineage-specific, and provide further support for the hypothesis of independent recruitment of α-CAs to diverse metazoan biocalcification processes. An analysis of α-CA sequences associated with biocalcification processes indicates that the domains involved in the activity and conformation of the active site are extremely conserved among metazoans.

Keywords: α-Carbonic anhydrase, Metazoa, Biocalcification, Biomineralization, Molecular evolution, Low complexity domains (LCDs), Repetitive low complexity domains (RLCDs)

Introduction

Biocalcification refers to the physiological and molecular process by which living systems produce mineralized structures based on calcium salts, mainly calcium carbonate. From an evolutionary perspective, biocalcified structures represent a true innovation in the evolutionary history of metazoans. Biocalcitic structures fulfil a wide range of functions [1] such as: support and protection (*e.g.* sponge spicules, coral and bryozoan exoskeletons, mollusc and brachiopod shells, mineralized carapaces of crustaceans, calcified tubes of annelids, and the test, spines and teeth of echinoderms); organs for spatial equilibration (*e.g.* statoliths in jellyfishes, otoliths in fishes); calcium storage structures (*e.g.* gastroliths in crustaceans); organs for reproduction (*e.g.* gastropod love darts); desiccation prevention (*e.g.* avian eggshells); detoxification (*e.g.* calcium

granules in molluscs). Despite the diversity and abundance of metazoan biocalcified structures, and the relevance of their evolutionary origins to our understanding of how metazoan life diversified, the molecular processes that underlie their formation remain generally poorly understood. Nevertheless, many recent molecular studies focused on a broad phylogenetic range of biomineralising animals have revealed that carbonic anhydrase (CA) is a key conserved enzyme involved in the process of biocalcification [2-22].

The carbonic anhydrase super-family

Carbonic anhydrases constitute a group of metallo-enzymes of the carbon-oxygen lyase sub-class and of the hydro-lyase infra-class. They catalyse the reversible hydration of CO_2 to form one bicarbonate ion and one proton: $H_2O + CO_2 \Leftrightarrow H_2CO_3 \Leftrightarrow HCO_3^- + H^+$. In most cases, the active site of these enzymes is composed of one Zinc(II) ion coordinated by three histidine residues. CAs occur in all three domains of life (Eubacteria,

* Correspondence: djackso@uni-goettingen.de
[2]Courant Research Centre Geobiology, Georg-August-University of Göttingen, Goldschmidtstr 3, 37077 Göttingen, Germany
Full list of author information is available at the end of the article

Archaea and Eukaryota), and play roles in diverse metabolic pathways. The CA superfamily is subdivided into five families: α-CA, β-CA, γ-CA, δ-CA and ζ-CA [23,24]. Interestingly, large-scale phylogenetic analyses reveal no homology between these five families [25], and they are therefore thought to be the result of convergent evolution, *i.e.* there was no common ancestor from which these five families evolved. This viewpoint is supported by the fact that CA conformations differ between families: while monomers and dimers are the active forms of α-CAs [26], γ-CAs are active in homotrimeric form [27], and β-CAs can be active in dimeric, tetrameric, hexameric or octameric forms. The amino acids involved in the coordination of the zinc ion are three histidine residues, or two cysteines and one histidine residues in the case of β-CAs (*e.g.* dicotyledonous plant *Pisum sativum*; [28]).

Despite their similar enzymatic properties and distinct evolutionary histories, the presence of a given CA family in a certain lineage does not exclude the presence of another. For example, α-CAs are present in bacteria, algae, in the cytoplasm of green plants and fungi, and in metazoans [29-31], while the β-CAs are present in bacteria, fungi, algae and in the chloroplasts of monocotyledones and dicotyledones [32,33]; γ-CAs are found in archaea and bacteria [34-36], suggesting that they are the oldest class, while the δ-CAs and the ζ-CAs are present in some marine diatoms [37-39]. All CA families are characterized by the presence of a zinc ion, the exception being the ζ family, for which cadmium replaces zinc, for example in marine diatoms such as *Thalassiosira weissflogii* [37,40]. Because α-CAs have most often been associated with biocalcification roles in various metazoans, here we will focus our attention on this family.

The α-carbonic anhydrases family
In 1933 Meldrum and Roughton [41], and simultaneously Stadie and O'Brien [42], discovered the first α-CA in vertebrate erythrocytes *via* its enzymatic activity. Today, with the advances in molecular techniques and the development of high-throughput sequencing methods, hundreds of α-CAs in many isoforms have been described for vertebrate and invertebrate metazoans.

Different α-CA isozymes are produced by distinct tissues. In mammals, they have diverse subcellular localizations: cytosolic (isozymes CAI, II, III, VII and XIII), membrane-bound (isozyme CAIV – GPI-link), transmembrane (isozymes CAIX, XII and XIV), mitochondrial (isozymes CAVA and VB) or extracellular (isozyme CAVI). They are involved in many biological functions including: acid–base balance, CO_2 transport, urea cycling, gluconeogenesis, fatty acids/amino acids synthesis and calcification/decalcification [43]. It is very likely that the primitive function of the α-CA was the intracellular regulation of acid–base balance, and/or the metabolism of CO_2 [44]. However, during evolution, the function of the enzyme was co-opted for other physiological processes, such as biomineralization. In some metazoan groups, α-CAs have been described as a calcium-dissolving enzyme, as in the bones of vertebrates and the spicules of calcareous sponges such as *Sycon raphanus* [45]. Recently, eight CAs in murine, produced by ameloblasts, were identified to be involved in biomineralization during teeth development [46]. In corals, α-CA isoforms interact in the carbon cycle of the endosymbiotic zooxanthellae playing an indirect role in skeletal formation [19]. In the calcifying species listed in Table 1, we have summarized the presence of α-CAs potentially involved in the deposition of $CaCO_3$ biominerals. For some species, such as brachiopods, data are currently missing; for others, such as sponges, crustaceans or cnidarians, they are widely incomplete [7,9,14,47]. In molluscs, sequence data is relatively abundant, but are focused on a limited number of species, such as the pearl oyster (*Pinctada* sp.) or the giant owl limpet (*Lottia gigantea*) [11,21]. Within the deuterostomes, most of the available α-CA data in echinoderms are related to the genome sequencing of the sea urchin *Strongylocentrotus purpuratus* [48]. The urochordates and vertebrates also possess $CaCO_3$ structures (*e.g.* spicules, otoliths, eggshell). Indeed, α-CAs were identified in the calcifying tissues that produce those structures [3,4,6,49].

The α-CA family contains numerous isozymes, corresponding to the expression of a set of paralogous genes that are the product of gene duplication and speciation events that occurred throughout the Phanerozoic [65-67]. All genes encoding α-CAs are likely derived from a single ancestral gene, and are therefore homologous [35].

To date, very few molecular phylogenetic studies of α-CAs have been conducted in relation to the biocalcification processes [7,9,10,18,19]. These few previous studies were performed with CA sequences of interest from poriferan, cnidarian and molluscan representatives in addition to various invertebrate/vertebrate datasets (Figure 1). The first phylogenetic study of α-CAs [35] revealed a clear dichotomy between membrane-associated/secreted and cytosolic/mitochondrial α-CAs, and this pattern is generally found in subsequent studies (Figure 1) [7,9,10,18]. In all of these studies poriferan α-CAs form a monophyletic clade, and occupy an early branching position (Figure 1). In other respects α-CA tree topology can differ significantly, for instance the position of vertebrate GPI-linked α-CAs as a sister group to the poriferan α-CAs [68], or the early-branching position of the molluscan nacreins [10]. One significant difficulty that all phylogenetic analyses of the α-CAs face, is the lack of a full genomic complement of α-CAs from a diverse taxonomic range of metazoan representatives. Due to the ever-growing availability of whole genomes and transcriptomes from non-model
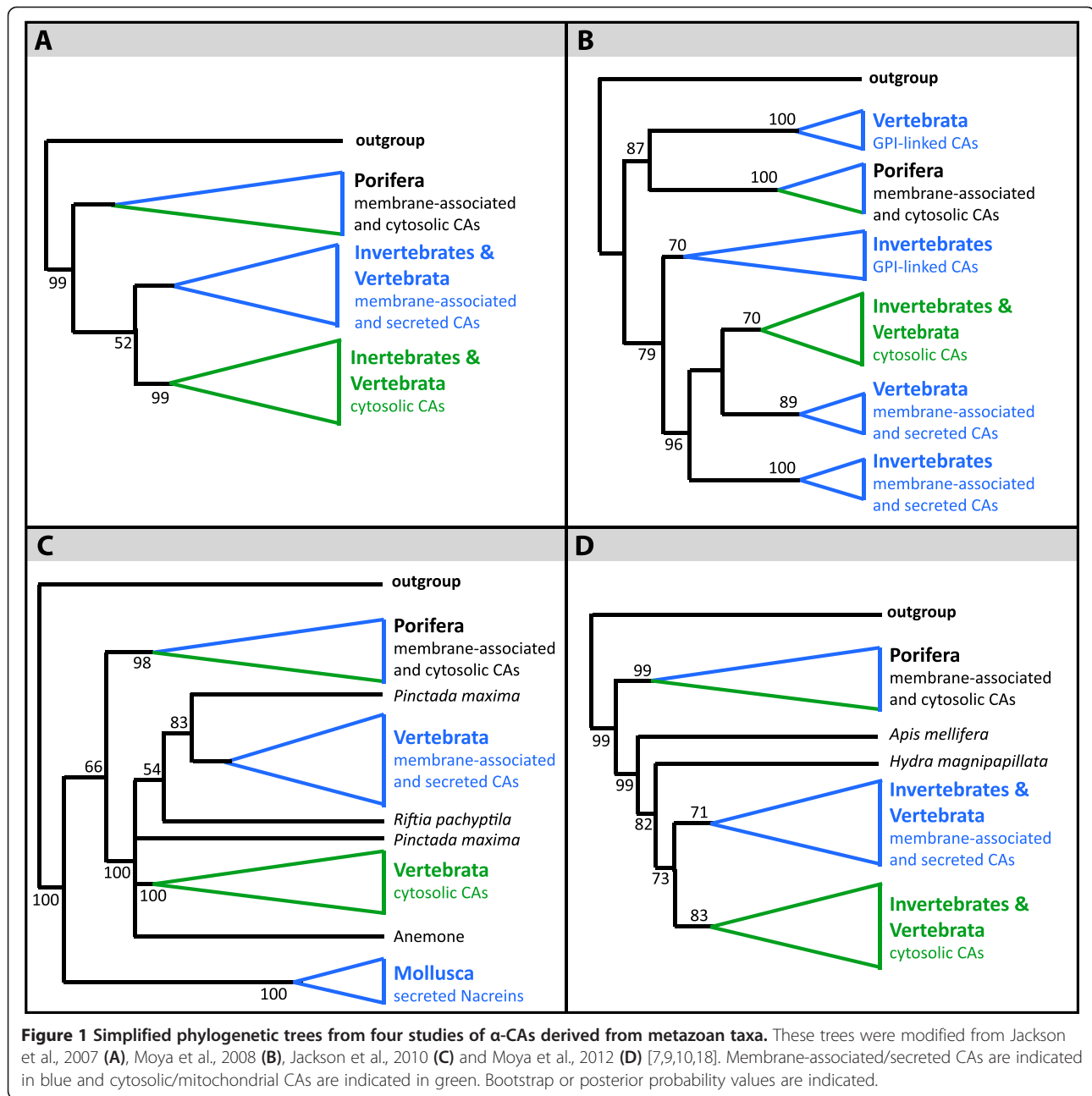
Table 1 Biomineralization types and the presence of α-CA in various metazoan phyla

	Envir.	CaCO$_3$ structure	Predominant form of mineral	Organic phase	Calcifying cells/ tissue	α-CAs	Technique	Tissue localization	Ref.
PORIFERA (appearance in Ediacaran)									
Hexactinellida	DS	Spicules	Silica, aragonite	nd	Sclerocytes	nd	-	-	
Demospongia	M, FW			"Spongine", silicatein, silintaphin-1, chitin		✓	RT-PCR & EST lib.	Adult whole organism, embryos	[7]
Calcarea	M		Calcite, aragonite	nd		✓	EST lib.	Adult whole organism	[45]
CNIDARIA (appearance in Ediacaran)									
Octocorallia	M	Sclerites, spicules, skeleton	High-Mg calcite	Glycoproteins, lipids, polysaccharides…	Calicoblastic cells	✓	biochemistry, RT-PCR, EST lib.	Spicules, tentacules, polyps, larva, embryos	[50]
Hexacorallia		Hard skeleton	Aragonite	Glycoproteins, lipids, polysaccharides…		✓	EST lib., RT-PCR, genome	Calicoblastic cells	[9,14,51-53]
Hydrozoa	M, FW	Granules, skeleton	Aragonite	nd		✓	EST lib., genome	Polyps	[54]
BRACHIOPODA (appearance in Cambrian)									
Craniiformea	M	Shell	Low-Mg calcite	nd	Mantle cells	nd	-	-	
Linguliformea			Ca$_3$(PO$_4$)$_2$	nd		nd	-	-	
Rynchonelliformea			Low-Mg calcite	Glycoproteins + polysaccharides…		nd	-	-	
MOLLUSCA (appearance in Cambrian)									
Solenogastra & Caudofoveata	M	Spicules	Calcite, aragonite	nd	Mantle cells	nd	-	-	
Polyplachophora						nd	-	-	
Bivalvia	M, FW	Shell & granules		Periostracum/OM		✓	RT-PCR, EST lib.	Mantle, gills, adductor muscles, larva	[2,11,13,15]
Gastropoda	M, FW, T					✓	RT-PCR, EST lib., genomes	Mantle, nervous system, whole body	[16,21]
Cephalopoda	M	Internal shell		Nautilin 63		✓	-	-	[55,56]
ANNELIDA (appearance in Cambrian)									
Pogonophora	M	Tube & soft granules	Calcite, aragonite	Mucopoly-saccharides	Epidermal glands	✓	RT-PCR	Branchial plume	[57]
ARTHROPODA (appearance in Cambrian)									
Hexapoda	A, T	Exoskeleton	Calcite	Chitin, proteinaceous organic matrix	Cuticle epidermis	✓	RT-PCR, EST lib., genomes	Whole organism	[54]
Pancrustacea	A, M, FW, T	Exoskeleton, gastrolithes, sternal plates, caecal concretions	Calcite, vaterite, amorphous CaCO$_3$		Cuticle epidermis, stomacal disc	✓	RT-PCR, EST lib.	Larvae, oocytes, gills, epidotes, brain, antennal glands…	[47,58-60]

Table 1 Biomineralization types and the presence of α-CA in various metazoan phyla *(Continued)*

ECHINODERMATA (appearance in Cambrian)									
Echinoidea	M	Test, teeth, spicules, spines	High-Mg calcite	OM	Syncitium/ "fenestrated" mineral	✓	EST lib., genomes	Pluteus ams, calcifying organic matrix	[48,61]
Crinoidea		Articulated ossicules	nd	nd	nd	nd	-	-	
Ophiuroidea			nd	nd	nd	nd	-	-	
Asteroidea		Spicules	nd	nd	nd	✓	Biochemistry	-	[62]
Holothuridea		Tiny spicules or ossicules	nd	nd	nd	nd	-	-	
UROCHORDATA (appearance in Cambrian)									
Ascidiacea	M	Spicules	Calcite	OM	Sclerocytes	✓	EST lib.	nd	[54]
Appendicularia		nd	nd	nd	nd	✓	Genome	nd	[63]
VERTEBRATA (appearance in Silurian)									
Actinopterygii	M, FW	Otoliths	Aragonite, vaterite	Endolymphatic constituants	Sacculus	✓	RT-PCR	Retina, kidney, erythrocytes, gills…	[49,64]
Aves	T	Eggshell	Calcite	Uterine fluid	Distal oviduct	✓	Genome	Testis, retina…	[54]

For each taxon we collected several lines of information: the environment in which each taxon can be found (A: amphibious; DS: deep-sea; FW: freshwater; M: marine; T: terrestrial); the general calcium carbonate structure; the predominant form of the mineral; the general nature of the organic phase within the mineral (OM: organic matrix); the source of the calcifying tissue/cells; whether carbonic anhydrases have been specifically detected within these cells; the method of identification (EST lib.: EST library); tissues where CAs were detected; references to studies from which this information was collated. nd: not determined.

Figure 1 Simplified phylogenetic trees from four studies of α-CAs derived from metazoan taxa. These trees were modified from Jackson et al., 2007 **(A)**, Moya et al., 2008 **(B)**, Jackson et al., 2010 **(C)** and Moya et al., 2012 **(D)** [7,9,10,18]. Membrane-associated/secreted CAs are indicated in blue and cytosolic/mitochondrial CAs are indicated in green. Bootstrap or posterior probability values are indicated.

organisms, this issue is being addressed and will allow more complete analyses to be performed.

Evolution of metazoan α-CAs

We explored the evolutionary relationships of metazoan α-CAs using two datasets compiled from publicly available genomic, EST and RNASeq databases (*i.e.* JGI, SpBase and NCBI [48,54,69]) derived from several metazoan species. In compiling these datasets we endeavoured to employ only those species with complete, or likely close to complete, genomic complements of α-CA. We explicitly excluded α-CA sequences from incomplete genomes and

incomplete and/or highly divergent α-CA sequences that were insufficiently conserved for alignment. We performed two phylogenetic analyses using Maximum likelihood [70,71] and Bayesian methods [72,73] (Additional file 1) that generated largely congruent tree topologies. The first analysis was conducted using 138 α-CA sequences (6 α-CAs from non-metazoans as outgroups, and 132 α-CAs from metazoans, Figure 2, Additional files 2 and 3: Figure S1 and S2), and the second, from 98 α-CA sequences (3 α-CAs from protists and 94 α-CAs from metazoans with a focus on molluscan α-CAs likely to be involved in biocalcification, Figure 3, Additional file 4: Figure S3).
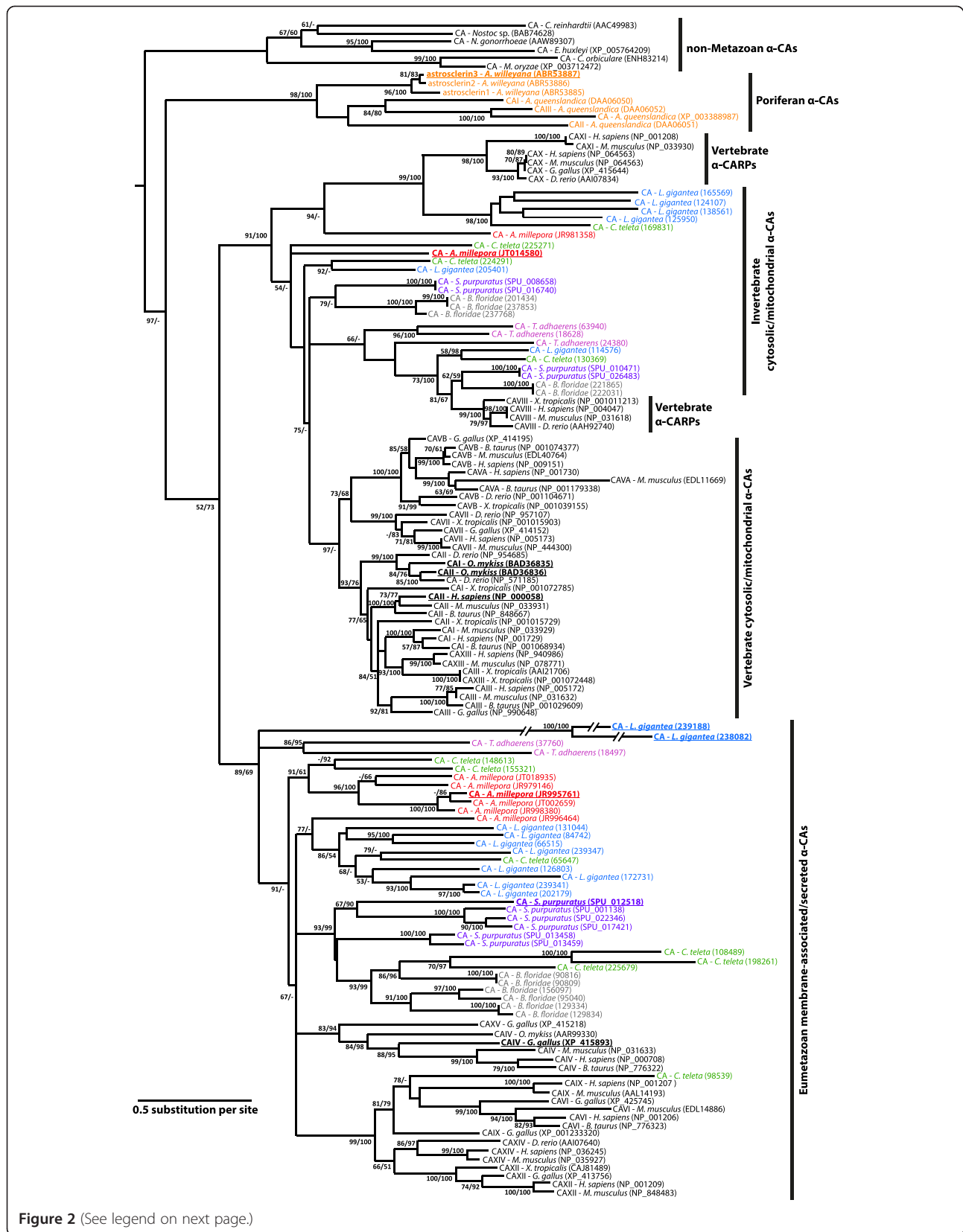
Figure 2 (See legend on next page.)

(See figure on previous page.)

Figure 2 Phylogenetic reconstruction of metazoan α-CAs relationships. The phylogeny was reconstructed using PhyML and Bayesian inference methods, the topology presented is that resulting from the PhyML method (see Additional file 2: Figure S1 for the Bayesian topology). Node support values indicate aLRT values, and Bayesian posterior probabilities expressed as a percentage respectively. Only values above 50 are indicated. Sequence data was derived from 4 vertebrate species (*Homo sapiens, Mus musculus, Bos taurus, Gallus gallus, Oncorhynchus mykiss* and *Xenopus tropicalis*, in black)*;* 1 cephalochordate (*Branchiostoma floridae*; in grey)*;* 1 sea urchin (*Strongylocentrotus purpuratus*, in purple)*;* 1 annelid (*Capitella teleta*, in green)*;* 1 mollusc (*Lottia gigantea*, in blue)*;* 1 coral (*Acropora millepora*, in red)*;* 1 placozoan (*Trichoplax adhaerens*, in pink)*;* 2 sponges (*Amphimedon queenslandica* and *Astrosclera willeyana*, in orange)*;* 1 fungi (*Colletotrichum orbiculare*, in black)*;* 2 chromalveolates (*Emiliania huxleyi* and *Vaucheria litorea*, in black)*;* 1 chlorophyte (*Chlamydomonas reinhardtii*, in black)*;* and 2 bacteria (*Nostoc* sp. *and Neisseria gonorrhoeae*, in black). The α-CAs involved in biocalcification are in bold/underlined. CARPs: Carbonic Anhydrase Related-Proteins.

In these trees, as previously mentioned, α-CA isozymes can be located in one of four broad positions: cytosolic, membrane-associated (*i.e.* transmembrane, membrane-bound, GPI-linked), secreted and mitochondrial. The location of a given α-CA seems to be married to its function; for example the majority of molluscan α-CAs known to be involved in shell formation are located extracellularly [2,15,21], while CAV members are thought to be involved in insulin regulation in the pancreas [74]. The fact that the ancient (>550 MYA) relationships of α-CA family members are largely reflected in where they are located highlights the important role that gene duplication played in
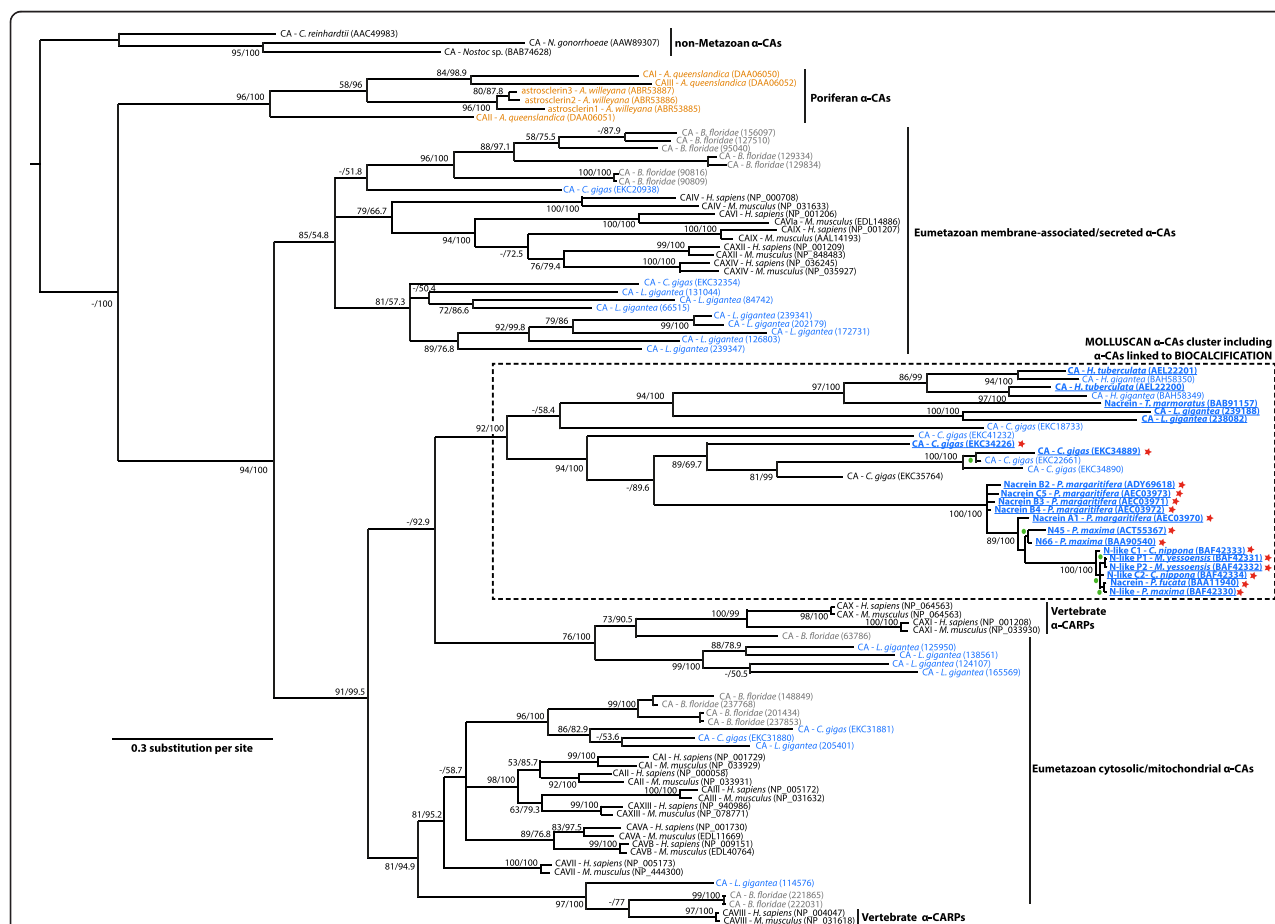
Figure 3 Phylogenetic reconstruction of metazoan α-CAs with a focus on molluscan sequences. The phylogeny was reconstructed using PhyML and Bayesian inference methods, the topology presented is that resulting from the Bayesian method (see Additional file 4: Figure S3 for the PhyML topology). Node support values indicate aLRT values, and Bayesian posterior probabilities expressed as a percentage respectively. Only values above 50 are indicated. Sequence data was derived from 2 vertebrate species (*Homo sapiens* and *Mus musculus*, in black)*;* 1 cephalochordate (*Branchiostoma floridae*, in grey)*;* 6 bivalve molluscs (*Crassostrea gigas, Crassostrea nippona, Patinopecten yessoensis, Pinctada fucata, Pinctada margaritifera* and *Pinctada maxima*, in blue)*;* 4 gastropod molluscs (*Haliotis gigantea, Haliotis tuberculata, Lottia gigantea, Turbo marmoratus*, in blue)*;* 2 sponges (*Amphimedon queenslandica* and *Astrosclera willeyana*, in orange)*;* and 3 unicellular organisms (*Nostoc* sp., *Chlamydomonas reinhardtii* and *Neisseria gonorrhoeae*, in black). The α-CAs involved in biocalcification are in bold/underlined. Nacreins and nacrein-likes are indicated by a red star. Well supported aLRT/bootstrap values (>90) are indicated by a green point due to lack of space. CARPs: Carbonic Anhydrase Related-Proteins.

allowing this class of enzymes to diversify and acquire novel functions.

In the tree of Figure 2, we observe that the seven poriferan α-CAs share a monophyletic relationship, and that they occupy an early-branching position. This finding is congruent with earlier publications [7,10,18] (see Figure 1 A, C and D), and confirms a recent analysis reported by Moya et al. [9]. While we do not observe a monophyletic clade of secreted and membrane-bound α-CAs, we do recover a highly supported clade of cytosolic and mitochondrial α-CAs. Puzzling is also the fact that the poriferan α-CAs are excluded from the membrane-associated/secreted cluster in the different phylogenetic reconstructions [7,9,10,18]. Among them, six are membrane-associated or secreted while one is cytosolic (CAII *Amphimedon queenslandica*, GenBank:DAA06051) [7]. The presence of membrane-associated/secreted α-CAs both in the poriferan and in eumetazoan α-CAs, suggests that the ancestral metazoan α-CA was membrane-associated or secreted. It must be noted however, that there is a putative cytosolic α-CA in the poriferan cluster, which belongs to the sponge *A. queenslandica* (CAII, GenBank:DAA06051). This may be explained by either the internalization of an ancestral membrane-associated/secreted enzyme, or the early evolution of a membrane-associated/secreted CA in the poriferan lineage. Nevertheless, it is obvious that the dearth of α-CA sequence data from the phylum Porifera is a critical point for any interpretation of the evolution of metazoan α-CA.

The distribution of metazoan α-CAs in relation to biocalcification

Our second phylogenetic analysis (Figure 3) includes 44 α-CAs from molluscan species, and sheds light on the evolution of A-CAs involved in molluscan biocalcification. One key-feature of this analysis is that molluscan mantle-secreted α-CAs and nacrein/nacrein-like sequences clearly group together. In their 2010 paper, Jackson et al. [10] described a basal position of molluscan nacreins relative to cytosolic/mitochondrial and secreted/membrane-associated α-CAs. Our phylogenetic reconstruction uses a more complete α-CA dataset, and generates a nacrein-type α-CA clade in a more derived position (Figure 3). Together with other molluscan α-CAs (*Lottia gigantea*, JGI:239188, 238082, *Crassostrea gigas*, GenBank:EKC18733, EKC41232, EKC34226, EKC34889, EKC22661, EKC34890, *Haliotis tuberculata*, GenBank:AEL22201, AEL22200, *Haliotis gigantea*, GenBank:BAH58350, BAH58349, *Turbo marmoratus*, GenBank:BAB91157), the nacrein/nacrein-like cluster is the sister group of a clade containing vertebrate related-protein α-CAs (CARPs: CAX and CAXI). Nacreins are clustered with other molluscan α-CAs likely to be involved in biocalcification (*e.g. Lottia gigantea*, JGI:239188, 238082).

With regard to its complement of 17 α-CAs, the giant owl limpet, *L. gigantea*, is unusual. Currently, this represents the largest known repertoire of α-CAs for an invertebrate. Sixteen of these sequences were used in our phylogenetic reconstruction, with 8 of these located within the membrane-associated/secreted clade, 6 within the vertebrate cytosolic/mitochondrial α-CAs and CARPs, and 2 (JGI:239188, 238082]) within the molluscan cluster containing nacreins (Figure 3). Two recent proteomic analyses confirmed the presence of these 2 α-CAs (JGI:239188, 238082) in the shell of *L. gigantea* [17,21]. Nonetheless, among the 8 sequences that are distributed in the membrane-associated/secreted cluster, 3 have a signal peptide (JGI:239341, 202179, 172731) while 5 are incomplete sequences with missing N-termini. We also note the presence of a putatively cytosolic α-CA (JGI:239347) in both trees (Figures 2 and 3), within the membrane-associated/secreted α-CAs. Even so, if these sequences are genuine cytosolic α-CAs, this may be explained by the evolutionary internalisation of a formerly secreted α-CA, as may be the case for the poriferan DAA06051 sequence (see above). Alternatively, these proteins may be secreted by non-canonical secretory pathways, or represent incomplete sequencing/assembly/or gene predictions of the N-terminal region of the protein. Moreover, the location of the majority of the invertebrate α-CAs are yet to be experimentally validated, and therefore the groupings that we observe in our phylogenetic analysis must be considered in that light. Despite these uncertainties, the overall pattern of molluscan α-CA evolution appears to be one that took place largely independently of vertebrate α-CA evolution (see dashed square in Figure 3).

Similar to the *L. gigantea*, the 10 α-CAs from the purple sea urchin *Strongylocentrotus purpuratus,* and the 8 from the stony coral *Acropora millepora,* are spread across the membrane-associated/secreted cluster (6 *A. millepora* α-CAs and 6 *S. purpuratus* α-CAs), and in the cytosolic/mitochondrial cluster (2 *A. millepora* α-CAs and 4 *S. purpuratus* α-CAs; Figure 2). One of the *S. purpuratus* α-CAs (JGI:SPU_012518, also called Sp-CAra7LA) has been identified in the extracellular matrix of the adult urchin test [12]. In our phylogenetic tree, Sp-Cara7LA is localized in the same cluster that contains the 2 α-CAs involved in the shell biocalcification of *L. gigantea* (JGI:239188, 238082) [17,69] and the 2 others involved in the skeleton precipitation of *A. millepora* (GenBank:JR995761 and JT014580) [18]. Despite the large phylogenetic distance between these three organisms they share a common α-CA ancestor that could be independently recruited in the purple sea urchin, in the giant owl limpet and in the stony coral, to be included in their specific biocalcification processes.

The dispersion of α-CAs involved in biocalcification in cnidarians, molluscs and echinoderms (Figures 2 and 3) supports the hypothesis of the independent evolution of

calcifying matrices in these different metazoan lineages. This is also true for vertebrate α-CAs taking part in the calcification processes (*e.g.* otolithes, eggshell). Our finding is congruent with the proteomic analyses on skeletal matrices that show that calcifying secretory repertoires are highly divergent in different metazoan phyla [22,75], even within one clade such as molluscs [10,15,21,68,76].

The position of α-CAs involved in (1) the calcification of eggshell of the red junglefowl *Gallus gallus* (GenBank:XP_415893, [77]), of otoliths of the rainbow trout *Oncorhynchus mykiss* (GenBank:BAD36835, BAD36836, [49,64]) and (2) the decalcification of bones of the human *Homo sapiens* (GenBank:NP_000058, [78]), still remains inside vertebrate α-CA clades (Figure 2). Nevertheless, they are not monophyletic. Among these α-CAs playing a role in the biocalcification, CAII of *H. sapiens* and CAI and II of *O. mykiss* are clustered with cytosolic/mitochondrial α-CAs while CAIV of *G. gallus* is included within membrane-associated/secreted α-CAs. The use of different α-CA isoforms in these species could be explained by an independent recruitment of α-CAs for biocalcification functions inside the vertebrate subphylum.

The phylogenetic distribution of α-CAs involved in metazoan biocalcification (Figure 2) reveals a complex evolutionary history that requires more whole-genome data in order to clarify how the different α-CA isoforms evolved. Obviously, broader taxon sampling from groups such as sponges would serve to shed light on the evolutionary history of these biomineral forming mechanisms.

## Conserved domains in α-CAs involved in the metazoan biocalcification processes

The catalytic domain of all α-CAs is highly conserved. In order to identify the most conserved domains in α-CAs involved in biocalcification, we calculated the substitution rate per site using the methodology described in Petit et al. (2006) [79] and Martin et al. (2009) [80]. Briefly, the topology of the tree provided by the Maximum Likelihood method with MEGA 5.1 (using an alignment of 34 sequences for the alignment of α-CAs involved in the calcification process and another alignment of 33 sequences from cytosolic/mitochondrial α-CAs in eumetazoans, Additional file 5: Table S1) was considered as the user tree to run the parsimony program Protpars included in the PHYLIP v3.69 package [81]. A moving average (window size = 5) was calculated for each substitution number per site, and then plotted in parallel with the structural domains of α-CAs (Figure 4). The major reversed peaks correspond to the catalytic sites that contain histidine residues, but also to conserved domains that maintain the conformation of the catalytic site for a correct activity. Seven short conserved α-CA domains (from 4 to 8 residues) spread

along the sequence were identified: QSPI, LHVH, GSEH, EAHL, FVVVGVFL, GSLTTP and ESVLW. These conserved regions represent about 25% of the overall sequences used for the analysis. These domains contain residues essential for the regeneration of the active site [51,82-86], and an amino acid involved in the proton shuffling process during the catalytic activity. Two of the poriferan α-CAs (Astrosclerin 1 and 2: GenBank: ABR53885, ABR53886) are characterized by the replacement of one of the three zinc coordinating histidine residues by a glutamine. In accordance with this Astrosclerin 1 was reported to be catalytically inactive, however recombinant Astrosclerin 2 could not be generated [7]. Such mutations in the active site of α-CA are likely to cause a weakening or elimination of zinc-binding capacity, and therefore of catalytic function [87].
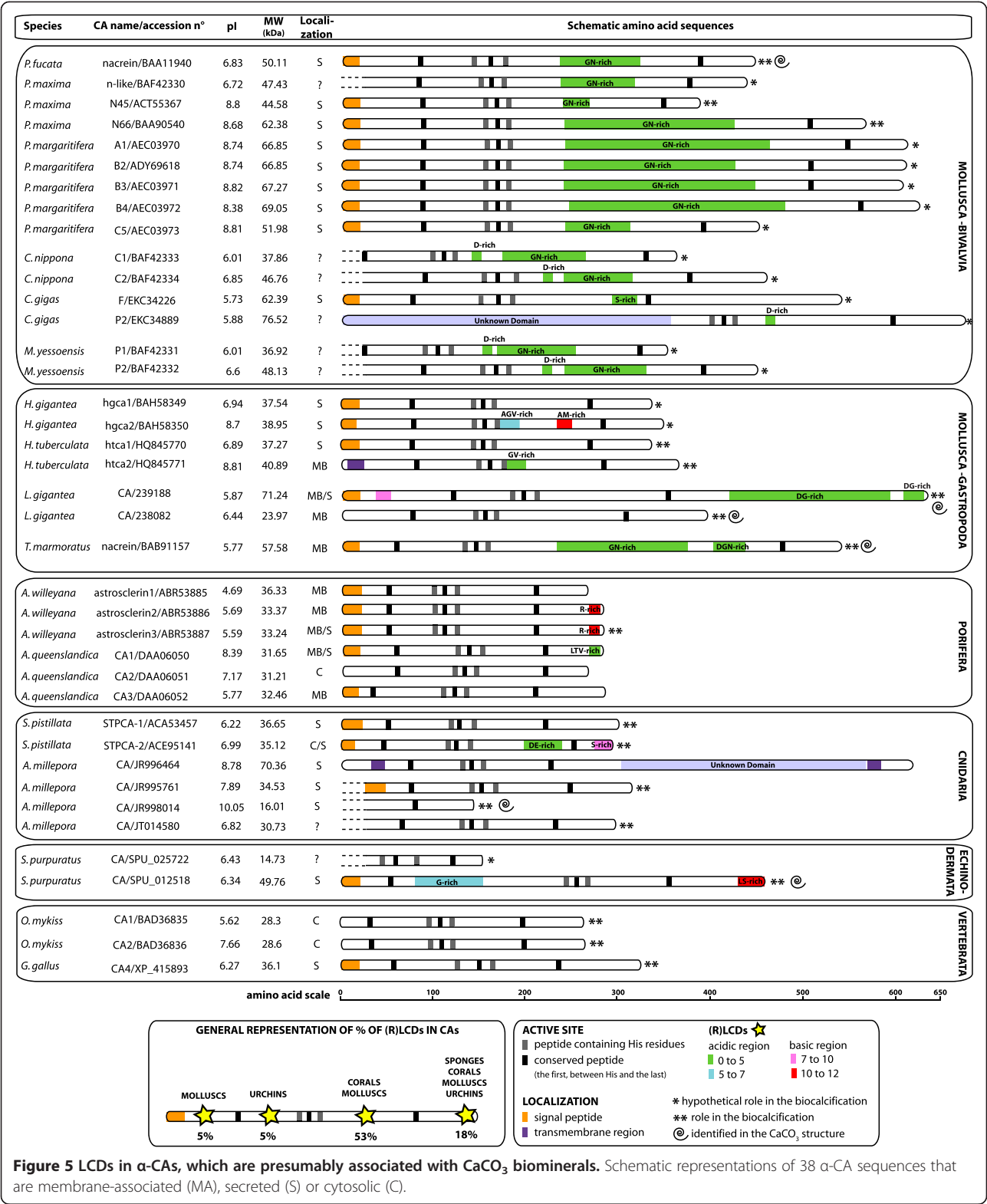
The substitution rate is two times less in the α-CAs involved in the biocalcification (1.6 to 8.3; graph A Figure 4) than in the cytosolic/mitochondrial α-CAs (4 to 16.8; graph B Figure 4). In the cytosolic/mitochondrial α-CAs the substitution rate is higher in regions that are not conserved in comparison with the substitution rate in the first graph (A, Figure 4). In general it seems that α-CAs involved in the biocalcification process are highly conserved when compared to the cytosolic/mitochondrial ones. We observe that the independent recruitment of these α-CAs to the biocalcification process seems to be followed by a higher conservation of amino acid sequences than in α-CAs involved in other functions. This pressure is probably link to the biocalcification function, which needs a high conserved α-CA activity as we can observe in the cluster of molluscan α-CAs linked to the formation of the shell (*e.g.* Figure 3). The catalytic activity could appear at different levels of this process. Indeed, CAs can provide HCO_3^- ions and/or CO_2 depending of the direction of the reaction. The catalytic hydration of CO_2 produces protons in addition to HCO_3^- and these protons have been suggested to promote the uptake of Ca^{2+} during larval shell formation in the freshwater gastropod *Lymnaea stagnalis* [88]. α-CAs may also play structural roles in biomineralization; the enzyme is occluded in the extracellular organic matrix of the biomineral structure as in some coral, molluscan and urchin species (*e.g. A. millepora, U. pictorum, L. gigantea, S. purpuratus*).

Low Complexity Domains (LCDs) in α-CAs related to CaCO₃ biomineralization

In silico sequence analysis (Additional file 1) focused on the identification and characterization of LCD domains present in α-CA sequences. We selected 33 α-CAs among which 33 have been demonstrated to be present in calcifying tissues or in the CaCO₃ biomineral itself (*e.g.* nacreins in the shell of pearl oysters; [2]). Thus, we
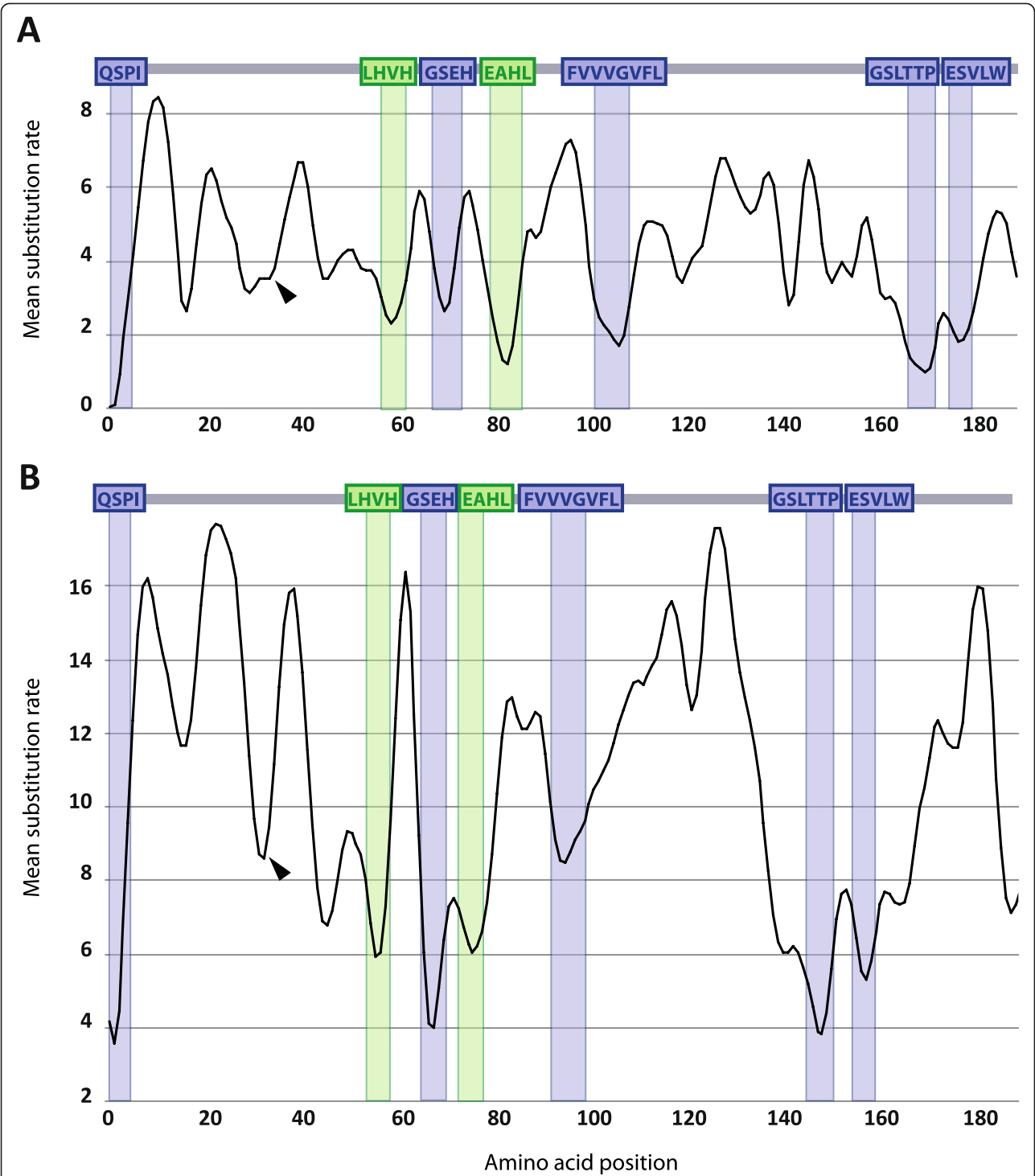
Figure 4 Substitution rates within metazoan α-CAs. The number of substitutions per site between poriferan and vertebrate sequences was calculated for an alignment of amino acids derived from 34 α-CA sequences involved in biocalcification **(A)**, and an alignment of 33 cytosolic/mitochondrial α-CAs **(B)**. Sequences used for the two analyses are listed in the Additional file 5: Table S1. The first position begins with the first α-CA conserved peptide (QSPI). Blue peptides are involved in the CA conformation and green peptides contain histidine residues that coordinate the Zn^{2+} ion in the active site. The black arrowhead indicates the substitution rate of the proton shuffling amino acid.

Figure 5 LCDs in α-CAs, which are presumably associated with CaCO₃ biominerals. Schematic representations of 38 α-CA sequences that are membrane-associated (MA), secreted (S) or cytosolic (C).

assume that these α-CAs are directly involved in the bio-calcification process.

The full-length sequences of these biomineral-associated α-CAs varies from 278 (*A. willeyana*, GenBank:ABR53885) to 678 amino acids (*C. gigas*, GenBank:ECK34889). This disparity in size is primarily due to the presence/absence of LCDs of variable length. Among the selected 39 biomineralization-associated α-CA sequences, 25 of them contain LCDs, which can make up to half the length of the sequence. Figure 5 presents a schematic representation of the primary structure of these 39 CAs.

With three exceptions (*C. gigas*, GenBank:EKC34889; *L. gigantea*, JGI:239188 and *S. purpuratus*, JGI:SPU_012518), all of these LCDs are located in the C-terminal half of the protein. The homology of these domains is difficult to discern as they differ in length, and are composed of a variety of amino acid sequences (Figure 5). Moreover, our phylogenetic analyses indicate that α-CAs containing these LCDs are not necessarily orthologous, but were likely acquired independently from lineage to lineage. The most common LCD is the GXN repeat found in molluscan nacreins [2,89]. Within the bivalves, the homology of such GXN domains is hinted by the degree of nucleic acid similarity (Additional file 6: Table S2). This similarity does not

extend to the GN repeat found in the nacrein of the gastropod *Turbo marmoratus* (Figure 6). Beside issues of LCD homology is the question of origin. In bivalves, the repetition of the GXN motif may be the result of a partial gene duplication followed by a series of unequal recombination events, as it has been suggested for other proteins such as mucins [90-92]. Such events result in the successive, tandem addition of short motifs. Transposable elements may also have generated such RLCDs in molluscan biomineralizing proteins. To test this idea we searched for signatures of transposases and the presence of inverted repeats in RLCD containing α-CA using the RepeatMasker software [93]. We could not find evidence of either in these sequences.

There is strong proteomic evidence that some nacreins are occluded in the calcified molluscan shell, and consequently belong to the so-called "calcifying matrix" [15,94-97]. It is known that the GXN repeat may display different functions related to calcification: (1) it may interact with calcium ions in solution, by behaving as a low affinity, high capacity calcium-binding domain, and could therefore be involved in $CaCO_3$ nucleation [98,99]. This suggests that nacreins work as both as enzymes, converting CO_2 into bicarbonate, and also as $CaCO_3$ nucleating polymers. (2) The GXN repeat domain is also known to

Figure 6 Details of 17 LCDs of α-CAs/nacreins from invertebrate species. Three categories sharing likely homology are indicated (black boxes): C-term basic LCDs in the sponge *A. willeyana* (highlighted in pink), acidic GXN repeated domains in bivalves (highlighted in blue) and acidic LCDs in the two species of gastropods *Haliotis* (highlighted in green).

interact with CaCO$_3$ *in vitro*. Miyamoto et al. [5], conducted *in vitro* calcification assays with wild type and truncated recombinant nacrein proteins. According to that work, the GXN domain inhibits the activity of nacrein and could act as a negative regulator of the CaCO$_3$ precipitation in the shell formation. The reason for such an α-CA inhibitory role is as yet unknown.

Our *in silico* analysis also reveals the occurrence of domains with similar acidic domains to the nacrein GXN repeat in the longer α-CA of *L. gigantea* (JGI:239188) and in the spicule specific α-CA [12] of *S. purpuratus* (JGI: SPU_012518). In *L. gigantea*, an acidic DG-rich peptide is located at the C-terminal end, and another more basic peptide is present just downstream of the signal peptide. In the purple sea urchin, the G-rich domain is less acidic than the nacrein GXN-rich domain. The position of this domain is also uncommon because it is upstream of the active α-CA site, in the N-terminal half of the protein.

Other short LCDs are strikingly different from the nacrein GXN repeat. Some are basic, *i.e.*, positively charged in standard pH conditions and are represented by the C-terminal R-rich, S-rich and LS-rich ends in four α-CAs (GenBank:ABR53886, GenBank:ABR53887, GenBank:ACE95141, JGI:SPU_012518). In *A. willeyana* the basic K-rich domain exhibits the KKRKRR motif containing also several arginine residues similarly to the C-termini of two family of proteins present in the prism matrix of the shell of the pearl oyster *Pinctada fucata*: shematrins and K-rich mantle proteins [100-102]. The presence of such basic domains suggests two possible functions: (1) an interaction with negatively charged bicarbonate ions. If so, the basic domain may concentrate bicarbonate ions at the vicinity of the nucleation site. (2) Electrostatic interaction with acidic (D/E-rich) domains of other proteins of the calcifying matrix. In this case, the basic domains may then anchor the acidic-soluble proteins to the three dimensional framework of the matrix. Here again, these basic domains indicate that the corresponding α-CAs display at least two functions, the enzymatic activity and the interaction with bicarbonate and/or acidic macromolecules.

Although the way by which the "mosaic" α-CAs are evolutionarily constructed is not known, exon shuffling can be envisaged as one possible mechanism [103]. Such a mechanism can generate new functions through the combination of different pre-existing functional domains into one protein. This "genetic tinkering" is well documented for proteins present in vertebrate extracellular matrices [104,105]. It is believed that this mechanism is subject to elevated rates of evolution, in comparison to those observed for intracellular proteins [106].

Conclusion

The α-CA family is a key family for understanding how the molecular mechanisms of metazoan biocalcification strategies evolved. Independent lines of evidence (phylogenetic relationships of biomineral associated α-CAs and patterns of LCD insertions) suggest that the recruitment of α-CAs into biocalcification roles likely occurred independently in different metazoan lineages. A major limitation of this work is the difficulty in reconstructing a robust phylogeny of metazoan α-CAs. The primary causes of this difficulty are likely to include the deep phylogenetic history of the α-CA family, the fact that many independent duplications have generated a variety of isoforms, and that these isoforms have subsequently been recruited to a wide variety of physiological roles. We remain optimistic that both broader and deeper taxon sampling (i.e. complete genome sequences) may bring more resolution to these relationships, and we anticipate that in the coming years such data will make these analyses more accurate. Such complete genome data may also provide additional lines of evidence for assigning CA orthology *via* gene synteny. In addition, physiological, biochemical and functional studies aimed at understanding the role of specific α-CAs in biomineralization processes will provide insight into the evolutionary history of mineral-relevant α-CA isozymes. A rich field of research regarding the origins of the LCDs associated with mineralizing α-CAs also awaits further investigation. Understanding how such LCDs are generated, how and why they come to be associated with mineralizing α-CAs, and what their functions are remain open and intriguing questions.

Additional files

Additional file 1: Description of methods used for Phylogenetic analyses (Figures 2 and 3) and *in silico* analyses (Figure 5).

Additional file 2: Figure S1. Topology of the tree presented in Figure 2 resulting from the Bayesian method. Bayesian posterior probabilities expressed as a percentage above 50 are indicated. For more details of the figure see legend of Figure 2.

Additional file 3: Figure S2. PhyML phylogenetic reconstruction with percentage bootstrap values, using the dataset of Figure 2.

Additional file 4: Figure S3. Topology of the tree presented in Figure 3 resulting from the PhyML method. The aLRT values expressed as a percentage above 50 are indicated. For more details of the figure see legend of Figure 3.

Additional file 5: Table S1. List of α-CA sequences used for the substitution rate analysis (Figure 4). For each α-CA, the name of species and the GenBank or JGI ID are provided.

Additional file 6: Table S2. Percentage of identity between nucleotide sequences encoding GN repeat domains in bivalve nacreins using nBLAST in NCBI.

Abbreviations

CA: Carbonic anhydrase; CaCO$_3$: Calcium carbonate; EST: Expressed sequence tag; GPI: GlycoPhosphatidylInositol; LCD: Low complexity domain; RLCD: Repetitive LCD; CARP: Carbonic anhydrase related-protein.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NLR collected the dataset, carried out alignments and phylogenetic analyses, wrote and drafted the manuscript. DJ participated in the phylogenetic analyses and drafted the manuscript. FM conceived the review, and participated in the design, coordination and drafting of manuscript. All authors read, approved and corrected the final manuscript.

Author details

[1]Centre Scientifique de Monaco, 8 quai Antoine Ier, Monte-Carlo 98000, Monaco. [2]Courant Research Centre Geobiology, Georg-August-University of Göttingen, Goldschmidtstr 3, 37077 Göttingen, Germany. [3]UMR CNRS 7245 MCAM, Muséum National d'Histoire Naturelle, 12 rue Buffon, Paris, France. [4]Instituto Gulbenkian d e Ciência, Rua da Quinta Grande 6, Apartado 14, Oeiras P-2781-901, Portugal. [5]UMR CNRS 6282 Biogéosciences, Université de Bourgogne, 21000 Dijon, France.

References

1. Lowenstam HA, Weiner S: *On Biomineralization.* New York: Oxford University Press; 1989.
2. Miyamoto H, Miyashita T, Okushima M, Nakano S, Morita T, Matsushiro A: **A carbonic anhydrase from the nacreous layer in oyster pearls.** *Proc Natl Acad Sci U S A* 1996, **93**:9657–9660.
3. Tohse H, Mugiya Y: **Effect of enzyme and anion transport inhibitors on in vitro incorporation of inorganic carbon and calcium into endolymph and otoliths in salmon** *Oncorhynchus masou. Comp Biochem Physiol A* 2001, **128**:177–184.
4. Beier M, Anken RH, Rahmann H: **Susceptibility to abnormal (kinetotic) swimming fish correlates with inner ear carbonic anhydrase-reactivity.** *Neurosci Lett* 2002, **335**:17–20.
5. Miyamoto H, Miyoshi F, Kohno J: **The carbonic anhydrase domain protein nacrein is expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in the mollusc** *Pinctada fucata. Zool Sci (Tokyo)* 2005, **22**:311–315.
6. Beier M, Anken R: **On the role of carbonic anhydrase in the early phase of fish otolith mineralization.** *Adv Space Res* 2006, **38**:1119–1122.
7. Jackson DJ, Macis L, Reitner J, Degnan BM, Wörheide G: **Sponge paleogenomics reveals an ancient role for carbonic anhydrase in skeletogenesis.** *Science* 2007, **216**:1893–1895.
8. Marie B, Luquet G, Bédouet L, Milet C, Guichard N, Medakovic D, Marin F: **Nacre calcification in the freshwater mussel** *Unio pictorum*: **carbonic anhydrase activity and purification of a 95 kDa calcium-binding glycoprotein.** *ChemBioChem* 2008, **9**:2515–2523.
9. Moya A, Tambutté S, Bertucci A, Tambutté E, Lotto L, Vullo D, Supuran CT, Allemand D, Zoccola D: **Carbonic anhydrase in the scleractinian coral** *Stylophora pistillata. J Biol Chem* 2008, **283**:25475–25484.
10. Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Rokhsar DS, Montagnani C, Joubert C, Piquemal D, Degnan BM: **Parallel evolution of nacre building gene sets in molluscs.** *Mol Biol Evol* 2010, **27**:591–608.
11. Joubert C, Piquemal D, Marie B, Manchon L, Pierrat F, Zanella-Cléon I, Cochennec-Laureau N, Gueguen Y, Montagnani C: **Transcriptome and proteome analysis of** *Pinctada margaritifera* **calcifying mantle and shell: focus on biomineralization.** *BMC Genomics* 2010, **11**:613.
12. Mann K, Poustka A, Mann M: **Phosphoproteomes of** *Strongylocentrotus purpuratus* **shell and tooth matrix: identification of a major acidic sea urchin tooth phosphoprotein, phosphodontin.** *Proteome Sci* 2010, **8**:6.
13. Marie B, Zanella-Cléon I, Le Roy N, Becchi M, Luquet G, Marin F: **Proteomic analysis of the acid-soluble nacre matrix of the bivalve** *Unio pictorum*: **detection of novel carbonic anhydrase and putative protease inhibitor proteins.** *ChemBioChem* 2010, **11**:2138–2147.
14. Bertucci A, Tambutté S, Supuran CT, Allemand D, Zoccola D: **A new coral carbonic anhydrase in** *Stylophora pistillata. Mar Biotechnol* 2011, **13**:992–1002.
15. Marie B, Le Roy N, Zanella-Cléon I, Becchi M, Marin F: **Molecular evolution of mollusc shell proteins: insights from proteomic analysis of the edible mussel** *Mytilus. J Mol Evol* 2011, **72**:531–546.
16. Le Roy N, Marie B, Gaume B, Guichard N, Delgado S, Zanella-Cléon I, Becchi M, Auzoux-Bordenave S, Sire JY, Marin F: **Identification of two carbonic anhydrases in the shell-forming mantle of the European abalone** *Haliotis tuberculata* **(Gastropoda, Haliotidae): phylogenetic implications.** *J Exp Zool Part B* 2012, **318**:353–367.
17. Mann K, Edsinger-Gonzales E, Mann M: **In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet** *Lottia gigantea. Proteome Sci* 2012, **10**:28.
18. Moya A, Huisman L, Ball EE, Hayward DC, Grasso LC, Chua CM, Woo HN, Gattuso JP, Forêt S, Miller DJ: **Whole transcriptome analysis of the coral** *Acropora millepora* **reveals complex responses to CO2-driven acidification during the initiation of calcification.** *Mol Ecol* 2012, **21**:2440–2454.
19. Bertucci A, Moya A, Tambutté S, Allemand D, Supuran CT, Zoccola D: **Carbonic anhydrase in anthozoan corals-a review.** *Bioorg Med Chem* 2013, **21**:1437–1450.
20. Drake J, Mass T, Haramaty L, Zelzion E, Bhattacharya D, Falkowski PG: **Proteomic analysis of skeletal organic matrix from the stony coral** *Stylophora pistillata. Proc Natl Acad Sci U S A* 2013, **110**:3788–3793.
21. Marie B, Jackson DJ, Ramos-Silva P, Zanell-Cléon I, Guichard N, Marin F: **The shell-formaing proteome of** *Lottia gigantea* **reveals both deep conservations and lineage-specific novelties.** *FEBS J* 2013, **280**:214–232.
22. Ramos-Silva P, Kaandorp J, Huissman L, Marie B, Zanemmé-Cléon I, Guichard N, Miller DJ, Marin F: **The skeletal proteome of the coral** *Acropora millepora*: **the evolution of calcification by co-option and domain shuffling.** *Mol Biol Evol* 2013, **30**:2099–2112.
23. Supuran CT: **Carbonic anhydrase: novel therapeutic applications for inhibitors and activators.** *Nature Rev Drug Discov* 2008, **7**:168–181.
24. Elleuche S, Pöggeler S: **Evolution of carbonic anhydrase in fungi.** *Curr Genet* 2009, **55**:211–222.
25. Capasso C, Supuran CT: **An overview of the alpha-, beta- and gamma-carbonic anhydrases from bacteria: can bacterial carbonic anhydrases shed new light on evolution of bacteria?** *J Enzyme Inhib Med Chem.* doi:10.3109/14756366.2014.910202.
26. Alterio V, Di Fiore A, D'Ambrosio K, Supuran CT, De Simone G: **Multiple binding modes of inhibitors to carbonic anhydrases: how to design specific drugs targeting 15 different isoforms?** *Chem Rev* 2012, **112**:4421–4468.
27. Ferry JG: **The γ class of carbonic anhydrase.** *Biochim Biophys Acta* 2009, **1804**:374–381.
28. Kimber MS, Pai EF: **The active site architecture of** *Pisum sativum* **β-carbonic anhydrase is a mirror image of that of α-carbonic anhydrases.** *EMBO J* 2000, **19**:1407–1418.
29. Moroney JV, Bartlett SG, Samuelsson G: **Carbonic anhydrase in plants and algae.** *Plant Cell Physiol* 2001, **24**:141–153.
30. Tripp BC, Smith K, Ferry JG: **Carbonic anhydrase: new insights for an ancient enzyme.** *J Biol Chem* 2001, **276**:48615–48618.
31. Bahn YS, Mühlschlegel FA: **CO₂ sensing in fungi and beyond.** *Curr Opin Microbiol* 2006, **9**:572–578.
32. Fawcett TW, Browse JA, Volokita M, Bartlett SG: **Spinach carbonic anhydrase primary structure deduced from the sequence of a cDNA clone.** *J Biol Chem* 1990, **265**:5414–5417.
33. Burnell JN, Suzuki I, Sugiyama T: **Light induction and the effect of nitrogen status upon the activity of carbonic anhydrase in maize leaves.** *Plant Physiol* 1990, **94**:384–387.
34. Alber BE, Ferry JG: **Characterization of heterologously produced carbonic anhydrase from** *Methanosarcina thermophila. J Bacteriol* 1994, **178**:3270–3274.
35. Hewett-Emmett D, Tashian RE: **Functional diversity, conservation, and convergence in the evolution of the α-, β-, and γ-carbonic anhydrase gene families.** *Mol Phylogenet Evol* 1996, **5**:50–77.

36. Smith KS, Ferry JG: A plant-type (β-class) carbonic anhydrase in the thermophilic methanoarchaeon *Methanobacterium thermoautotrophicum*. *J Bacteriol* 2000, **181**:6247–6253.

37. Lane TW, Saito MA, George GN, Pickering IJ, Prince RC, Morel FM: A cadmium enzyme from a marine diatom. *Nature* 2005, **435**:42.

38. Park H, Song B, Morel FMM: Diversity of the cadmium-containing carbonic anhydrase in marine diatoms and natural waters. *Environ Microbiol* 2007, **9**:403–413.

39. Vullo D, Del Prete S, Osman SM, De Luca V, Scozzafava A, AlOthman Z, Supuran CT, Capasso C: Sulfonamide inhibition studies of the d-carbonic anhydrase from the diatom *Thalassiosira weissflogii*. *Bioorg Med Chem Lett* 2014, **24**:275–279.

40. Xu Y, Feng L, Jeffrey PD, Shi YG, Morel FMM: Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. *Nature* 2008, **452**:56–61.

41. Meldrum N, Roughton FJW: Carbonic anhydrase: its preparation and properties. *J Physiol* 1933, **80**:113–142.

42. Stadie W, O'Brien H: The catalysis of the hydration of carbon dioxide and dehydration of carbonic acid by an enzyme isolated from red blood cells. *J Biol Chem* 1933, **103**:521–529.

43. Tashian RE: The carbonic anhydrase: widening perspectives on their evolution, expression and function. *Bioessays* 1989, **10**:186–192.

44. Henry RP: Multiple roles of carbonic anhydrase in cellular transport and metabolism. *Annu Rev Physiol* 1996, **58**:523–538.

45. Müller WEG, Wang X, Grebenjuk VA, Korzhev M, Wiens M, Schlossmacher U, Schröder HC: Common genetic denominators for Ca^{2+} –based skeleton in metazoa: role of osteoclast-stimulating factor and of carbonic anhydrase in a calcareous sponge. *PLoS One* 2012, **7**:14.

46. Reibring C-G, El Shahawy M, Hallberg K, Kannius-Janson M, Nilsson J, Parkkila S, Sly WS, Waheed A, Linde A, Gritli-Linde A: Expression patterns and subcellular localization of carbonic anhydrases are developmentally regulated during tooth formation. *PLoS One* 2014, **9**:e96007.

47. Pongsomboon S, Udomlertpreecha S, Amparyup P, Wuthisuthimethavee S, Tassanakajon A: Gene expression and activity of carbonic anhydrase in salinity stressed *Penaeus monodon*. *Comp Biochem Physiol A* 2009, **152**:225–233.

48. SpBase Strongylocentrotus purpuratus. In [http://www.spbase.org]

49. Tohse H, Murayama E, Ohira T, Takagi Y, Nagasawa H: Localization and diurnal variations of carbonic anhydrase mRNA expression in the inner ear of the rainbow trout *Oncorhynchus mykiss*. *Comp Biochem Physiol B* 2006, **145**:257–264.

50. Rahman M, Isa Y, Takemura A, Uehara T: Analysis of proteinaceous components of the organic matrix of endoskeletal sclerites from the alcyonarian *Lobophytum crassum*. *Calcif Tissue Int* 2006, **78**:178–185.

51. Silverman D, Lindskog S: The catalytic mechanism of carbonic anhydrase implications of a rate-limiting protolysis of water. *Accounts Chem Res* 1988, **21**:30–36.

52. De Boer M, Krupp DA, Weis VM: Two atypical carbonic anhydrases homologs from the planula larva of the scleractinian coral *Fungia scutaria*. *Biol Bull* 2006, **211**:18–30.

53. Grasso L, Maindonald J, Rudd S, Hayward DC, Saint R, Miller DJ, Ball EE: Microarray analysis identifies candidate genes for key roles in coral development. *BMC Genomics* 2008, **9**:540.

54. National Center for Biotechnology Information. In [http://www.ncbi.nlm.nih.gov]

55. Hirohashi N, Alvarez L, Shiba K, Fujiwara E, Iwata Y, Mohri T, Inaba K, Chiba K, Ochi H, Supuran CT, Kotzur N, Kakiuchi Y, Kaupp UB, Baba SA: Sperm from sneaker male squids exhibit chemotactic swarming to CO_2. *Curr Biol* 2013, **23**:775–781.

56. Hu MY, Lee J-R, Lin L-Y, Shih T-H, Stumpp M, Lee M-F, Hwang P-P, Tseng Y-C: Development in a naturally acidified environment: Na^+/H^+-exchanger 3-based proton secretion leads to CO_2 tolerance in cephalopod embryos. *Front Zool* 2013, **10**:51.

57. De Cian M, Andersen AC, Bailly X, Lallier FH: Expression and localization of carbonic anhydrase and ATPases in the symbiotic tubeworm *Riftia pachyptila*. *J Exp Biol* 2003, **206**:399–409.

58. Leu J-H, Chang C-C, Wu J-L, Hsu C-W, Hirono I, Aoki T, Juan H-F, Lo C-F, Kou G-H, Huang H-C: Comparative analysis of differentially expressed genes in normal and white spot syndrome virus infected *Penaeus monodon*. *BMC Genomics* 2007, **8**:120.

59. Towle DW, Smith CM: Gene discovery in *Carcinus maenas* and *Homarus americanus* via expressed sequence tags. *Integr Comp Biol* 2006, **46**(6):912–918.

60. Serrano L, Halanych KM, Henry RP: Salinity-stimulated changes in expression and activity of two carbonic anhydrase isozymes in the blue crab *Callinectes sapidus*. *J Exp Biol* 2007, **210**:2320–2332.

61. Love AC, Andrews ME, Raff RA: Gene expression patterns in a novel animal appendage: the sea urchin pluteus arm. *Evol Dev* 2007, **9**:51–68.

62. Morino Y, Koga H, Tachibana K, Shoguchi E, Kiyomoto M, Wada H: Heterochronic activation of VEGF signaling and the evolution of the skeleton in echinoderm pluteus larvae. *Evol Dev* 2012, **14**(5):428–436.

63. Denoeud F, Henriet S, Mungpakdee S, Aury JM, Da Silva C, Brinkmann H, Mikhaleva J, Olsen LC, Jubin C, Canestro C, Bouquet JM, Danks G, Poulain J, Campsteijn C, Adamski M, Cross I, Yadetie F, Muffato M, Louis A, Butcher S, Tsagkogeorga G, Konrad A, Singh S, Jensen MF, Cong EH, Eikeseth-Otteraa H, Noel B, Anthouard V, Porcel BM, Kachouri-Lafond R, et al: Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* 2010, **330**:1381–1385.

64. Esbaugh AJ, Perry SF, Bayaa M, Georgalis T, Nickerson J, Tufts BL, Gilmour KM: Cytoplasmic carbonic anhydrase isozymes in rainbow trout *Oncorhynchus mykiss*: comparative physiology and molecular evolution. *J Exp Biol* 2005, **208**:1951–1961.

65. Eicher EM, Stern RH, Womack JE, Davisson MT, Roderick TH, Reynolds SC: Evolution of mammalian carbonic anhydrase loci by tandem duplication: close linkage of Car-1 and Car-2 to the centromere region of chromosome of the mouse. *Biochem Genet* 1976, **14**:651–660.

66. Hewett-Emmett D, Hopkins PJ, Tashian RE, Czelusniak J: Origins and molecular evolution of the carbonic anhydrase isozymes. In *Biology and Chemistry of the Carbonic Anhydrase*, Volume 429. Edited by Tashian R, Hewett-Emmett D. New York: Annals of the New York Academy of Sciences; 1984:338–358.

67. Peterson RE, Tu C, Linser PJ: Isolation and characterization of a carbonic anhydrase homologue from the zebrafish (*Danio rerio*). *J Mol Evol* 1997, **44**:432–439.

68. Jackson DJ, McDougall C, Green K, Simpson F, Wörheide G, Degnan BM: A rapidly evolving secretome builds and patterns a sea shell. *BMC Biol* 2006, **4**:40.

69. Doe Joint Genome Istitute. In [http://www.jgi.doe.gov]

70. Méthodes et Algorithmes pour la Bioinformatique LIRMM. In [www.phylogeny.fr]

71. Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003, **52**:696–704.

72. Huelsenbeck JP, Ronquist F: MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 2001, **17**:754–755.

73. Ronquist F, Huelsenbeck JP: MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003, **19**:1572–1574.

74. Parkkila AK, Scarim AL, Parkkila S, Waheed A, Corbett JA, Sly WS: Expression of carbonic anhydrase V in pancreatic beta cells suggests role for mitochondrial carbonic anhydrase in insulin secretio. *J Biol Chem* 1998, **273**:24620–24623.

75. Livingston BT, Killian CE, Wilt F, Cameron A, Landrum MJ, Ermolaeva O, Sapojnikov V, Maglott DR, Buchanan AM, Ettensohn CA: A genorne-wide analysis of biomineralization-related proteins in the sea urchin *Strongylocentrotus purpuratus*. *Dev Biol* 2006, **300**:335–348.

76. Marie B, Le Roy N, Marie A, Dubost L, Milet C, Bédouet L, Becchi M, Zanella-Cléon I, Jackson D, Degnan B, Luquet G, Marin F: Nacre evolution: A proteomic approach. In *MRS Symposium Proceedings; 14–17 April 2009*. Edited by Kisailus DEL, Gupta H, Landis W, Zavattieri P. San Fransisco: Materials Research Society; 2009:69–75.

77. Nys Y, De Laage X: Effects of suppression of eggshell calcification and of 1,25 $(OH)2D3$ on Mg^{2+}, Ca^{2+} and Mg^{2+} HCO_3 ATPase, alkaline-phosphatase, carbonic-anhydrase and CABP levels. 1: the laying hen uterus. *Comp Biochem Physiol A* 1984, **78**:833–838.

78. Lehenkari P, Hentunen TA, Laitala-Leinonen T, Tuukkanen J, Väänänen K: Carbonic anhydrase II plays a major role in osteoclast differentiation and bone resorption by effecting the steady state intracellular pH and Ca^{2+}. *Exp Cell Res* 1998, **242**:128–137.

79. Petit D, Maftah A, Julien R, Petit JM: En bloc duplications, mutation rates, and densities of amino acid changes clarify the evolution of vertebrate alpha-1,3/4-fucosyltransferase. *J Mol Evol* 2006, **63**:353–364.

80. Martin R, Gallet PF, Rocha D, Petit D: **Polymorphism of the prion protein in mammals: a phylogenetic approach.** *Recent Pat DNA Gene Seq* 2009, **3**:63–71.

81. Felsenstein J: *PHYLIP (Phylogeny Inference Package) version 3.6. Distribution by the author.* Seattle: Department of Genome Sciences, University of Washington; 2004.

82. Vallee B, Auld DS: **Active-site zinc ligands and activated H_2O of zinc enzymes.** *Proc Natl Acad Sci U S A* 1990, **87**:220–224.

83. Stams T, Christianson DW: **X-ray crystallographic studies of mammalian carbonic anhydrase isozymes.** *EXS* 2000, **90**:159–174.

84. Briganti F, Mangani S, Orioli P, Scozzafava A, Vernagoline G, Supuran CT: **Carbonic anhydrase activators: X-ray crystallographic and spectroscopic investigations for the interaction of isozymes I and II with histamine.** *Biochemistry* 1997, **36**:10384–10392.

85. Di Fiore A, Monti SM, Hilvo M, Parkkila S, Romano V, Scaloni A, Pedone C, Scozzafava A, Supuran CT, De Simone G: **Crystal structure of human carbonic anhydrase XIII an dits complex with the inhibitor acetazolamide.** *Proteins* 2008, **74**:164–175.

86. Toba S, Colombo G, Merz KM Jr: **Solvent dynamics and mechanism of proton transfer in human carbonic anhydrase II.** *J Am Chem Soc* 1999, **121**:2290–2302.

87. Muhammad E, Leventhal N, Parvari G, Hanukoglu A, Hanukoglu I, Chalifa-Caspi V, Feinstein Y, Weinbrand J, Jacoby H, Manor E, Nagar T, Beck JC, Sheffield VC, Hershkovitz E, Parvari R: **Autosomal recessive hyponatremia due to isolated salt wasting in sweat associated with a mutation in the active site of carbonic anhydrase 12.** *Hum Genet* 2011, **129**:397–405.

88. Ebanks SC, O'Donnell MJ, Grosell M: **Characterization of mechanisms for Ca^{2+} and HCO_3^-/CO_3^{2-} acquisition for shell formation in embryos of the freshwater common pond snail *Lymnea stagnalis*.** *J Exp Biol* 2010, **213**:4092–4098.

89. Miyamoto H, Yano M, Miyashita T: **Similarities in the structure of nacrein, the shell-matrix protein, in a bivalve and a gastropod.** *J Molluscan Stud* 2003, **69**:87–89.

90. Gum JR: **Mucin genes and the proteins they encode-structure, diversity, and regulation.** *Am J Respir Cell Mol Biol* 1992, **7**:557–564.

91. Desseyn J, Aubert JP, Porchet N, Laine A: **Evolution of the large secreted gel-forming mucins.** *Mol Biol Evol* 2000, **17**:1175–1184.

92. Lane T, Hansson GC, Samuelsson T: **Gelforming mucins appeared early in metazoan evolution.** *Proc Natl Acad Sci U S A* 2007, **104**:16209–16214.

93. **Current Version: open-4.0.3 (RMLib: 20130422 & Dfam: 1.2).** In [http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker]

94. Marin F, Luquet G: **Molluscan shell proteins.** *C R Palevol* 2004, **3**:469–492.

95. Gaume B, Fouchereau-Peron M, Badou A, Helléouet M-N, Huchette S, Auzoux-Bordenave S: **Biomineralization markers during early shell formation in the European abalone *Haliotis tuberculata*, Linnaeus.** *Mar Biotechnol* 2011, **158**:341–353.

96. Marin F, Le Roy N, Marie B: **He formation and mineralization of mollusk shell.** *Front Biosci* 2012, **S4**:1099–1125.

97. Marin F, Marie B, Ben Hamada S, Silva P, Le Roy N, Guichard N, Wolf S, Montagnani C, Joubert C, Piquemal D, Saulnier D, Gueguen Y: **"Shellome": proteins involved in mollusc shell biomineralization – diversity, functions.** In *Recent Advances in Pearl Research.* Edited by Watabe S, Maeyama K, Nagasawa H. Tokyo: Terrapub; 2013:149–168.

98. Shiba K, Minamisawa T: **A synthesis approach to understanding repeated peptides conserved in mineralization proteins.** *Biomacromolecules* 2007, **8**:2659–2664.

99. Isowa Y, Sarashina I, Setiamarga DHE, Endo K: **A comparative study of the shell matrix protein aspein in pterioid bivalves.** *J Mol Evol* 2012, **75**:11–18.

100. Yano M, Nagai K, Morimoto K, Miyamoto H: **Shematrin: a family of glycin-rich structural proteins in the shell of the pearl oyster.** *Comp Biochem Physiol B* 2006, **144**:254–262.

101. Zhang C, Xie L, Huang J, Liu X, Zhang R: **A novel matrix protein family participating in the prismatic layer formation of pearl oyster, *Pinctada fucata*.** *Biochem Biophys Res Commun* 2006, **344**:735–740.

102. McDougall C, Aguilera F, Degnan BM: **Rapid evolution of pearl oyster shell matrix proteins with repetitive, low complexity domains.** *J R Soc Interface* 2013, **10**:20130041.

103. Bork P: **Shuffled domains in extracellular proteins.** *FEBS L* 1991, **286**:47–54.

104. Patthy L: **Genome evolution and the evolution of exon-shuffling – a review.** *Gene* 1999, **238**:103–114.

105. Patthy L: **Modular assembly of genes and the evolution of new functions.** *Genetica* 2003, **118**:217–231.

106. Julenius K, Pedersen AG: **Protein evolution is faster outside the cell.** *Mol Biol Evol* 2006, **23**:2039–2048.