The complex evolutionary history of sulfoxide synthase in ovothiol biosynthesis

Marco Gerdol[1*], Marco Sollitto[1], Alberto Pallavicini[1,2], Immacolata Castellano[2*]

[1] Department of Life Sciences, University of Trieste, Italy

[2] Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Naples, Italy.

[*]To whom correspondence should be addressed:

Immacolata Castellano

e-mail: immacolata.castellano@szn.it

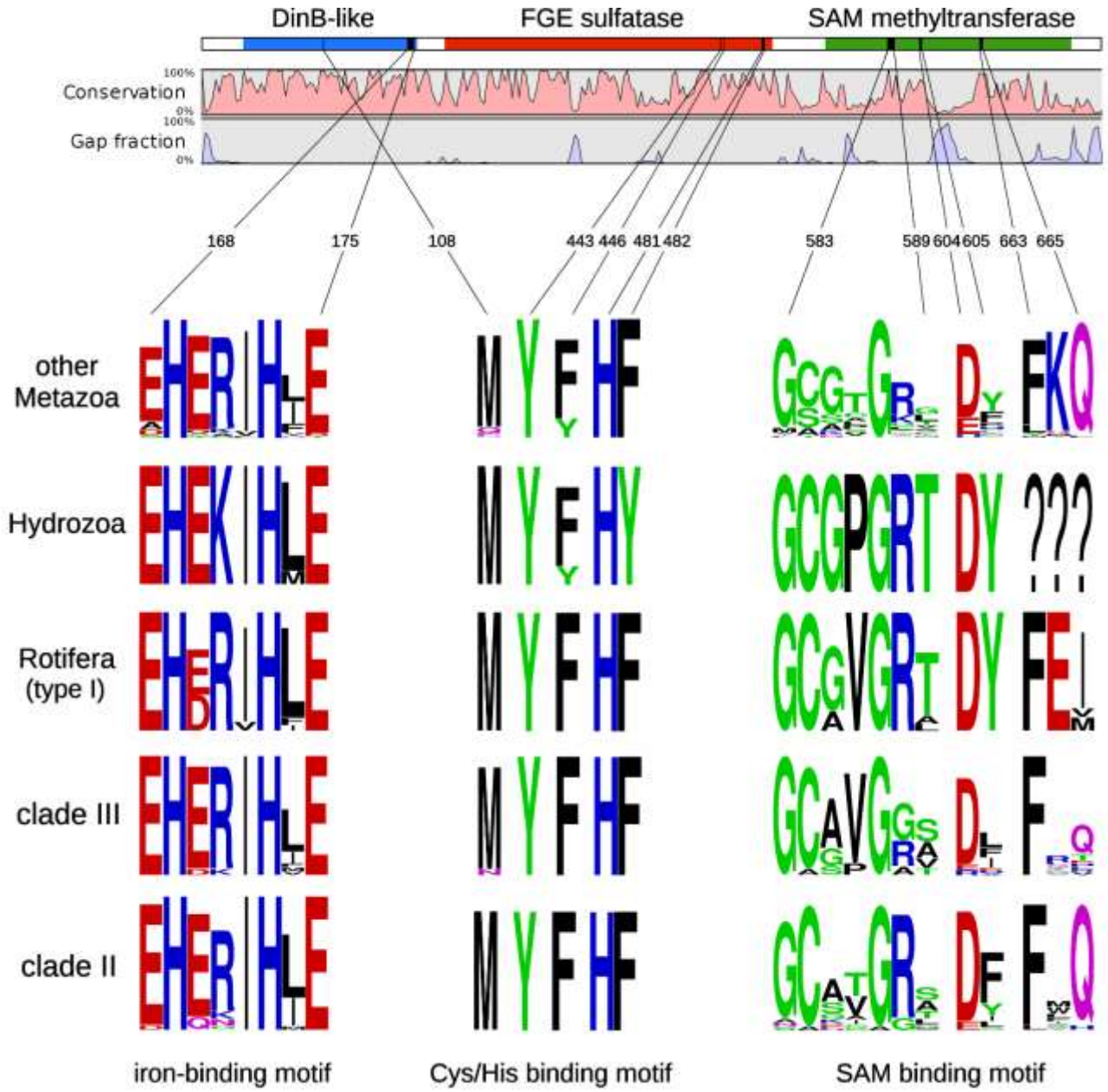
Marco Gerdol

e-mail: mgerdol@units.it

Supplementary Data Note 1. Comparison of OvoA protein sequences from metazoans

The multi-exonic architecture of metazoan OvoA genes covers the canonical protein structure composed of the N-terminal DinB-like domain, the central FGE-sulfatase, and the C-terminal SAM-transferase domain (Supplementary Figure 1), which are also found in the vast majority of non-metazoan sequences pertaining to clade II and clade III (for reference, see Figure 1 in the main text). The motif responsible for iron binding, i.e. HX3HXE, in the DinB-like domain was conserved, with slight variations, in all OvoA sequences from all organisms. The key residues supposed to be involved in the binding of the substrates histidine and cysteine [ref. 5 in the text] in OvoA (Met108, Tyr443, Phe446, His481 and Phe482) were extremely well conserved in most of OvoA sequences, regardless of their taxonomic origin. The only major difference observed was the replacement of Phe482 with a tyrosine residue in the horizontally transferred sequences of Hydrozoa.

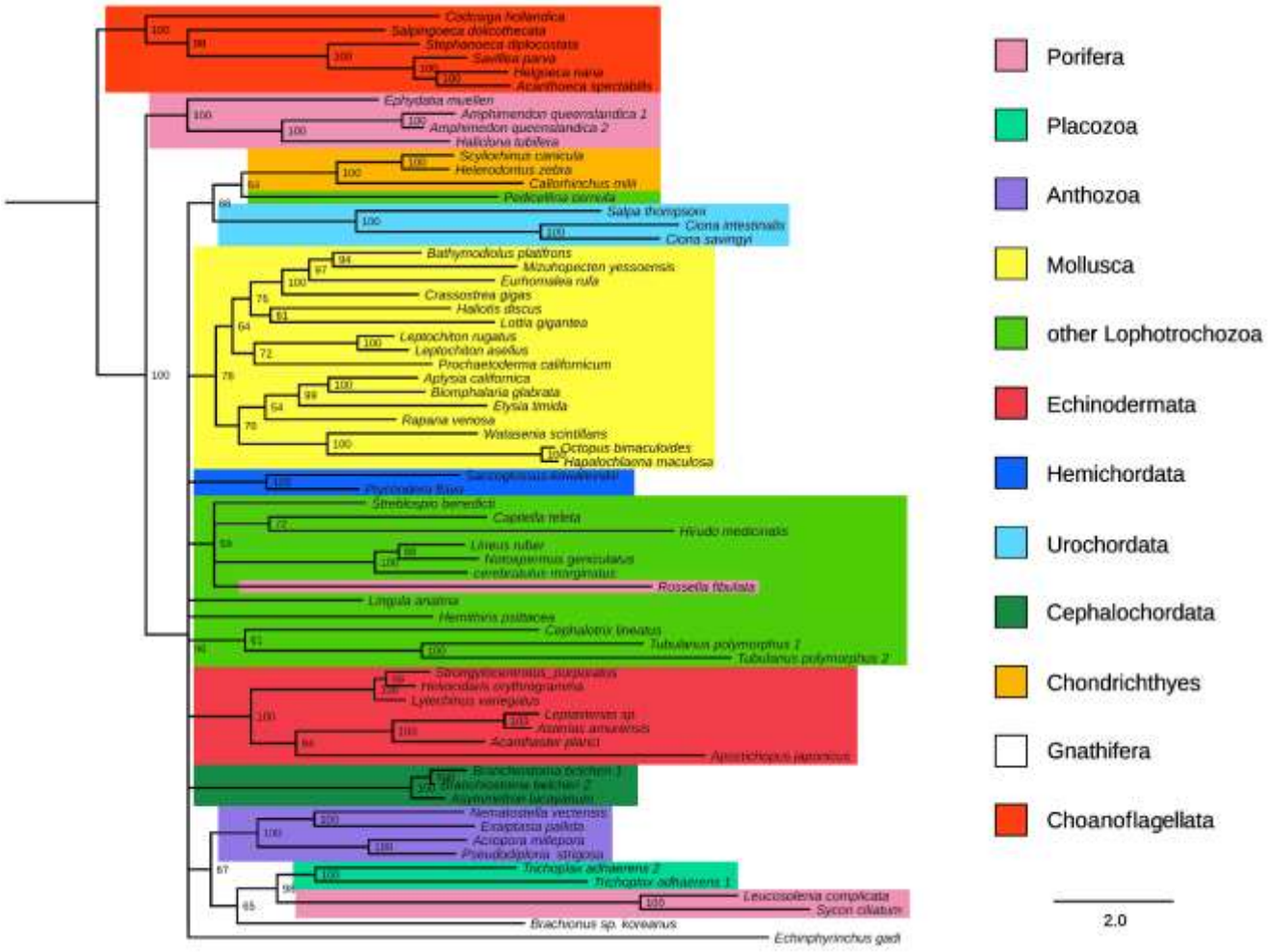
On the other hand, the residues involved in the binding of the S-adenosyl-methionine substrate in the SAM-transferase domain displayed a lower degree of conservation in metazoans, even though some key residues were invariable (i.e. Gly587 and Gln665). The SAM-binding motifs were relatively well conserved also in rotifers and, partly, of hydrozoans, with the exception of the third sub-motif (with an FKQ consensus in other metazoans). Indeed, this region was highly divergent in *Hydra*, also partly due to the significantly longer size of this region in these sequences compared to other OvoA sequences (see main text), and the homologous residues could not be identified.



Supplementary Figure 1: Schematic representation of an OvoA protein sequence, with residue conservation and gap fraction obtained from the multiple sequence alignment of all metazoan sequences used for the generation of the phylogenetic tree presented in Figure 4. Numbers are relative to the reference OvoA sequence of *S. purpuratus*. The conservation of the three key iron-binding, cysteine/histidine binding and SAM binding motifs in metazoan OvoA sequences, in the horizontally transferred genes from Hydrozoa and Rotifera, and in the sequences from clade II and clade III (see Figure 1A in the main text) is reported as a sequence logo.

Supplementary Data Note 2. Phylogenetic analysis of OvoA protein sequences from metazoans

Although Choanoflagellata were placed in a basal position, as expected, the tree topology obtained displayed significant incongruences compared with currently accepted species phylogeny. In detail, with the exception of Demospongiae (placed with high posterior probability support, = 100%, in a basal clade), all the other OvoA sequences were part of a single very large polytomy. While the sequences of some phyla clustered together with high support (e.g. Mollusca -78% posterior probability; Echinodermata -100% posterior probability; Hemichordata -100% posterior probability; Cephalochordata -100% posterior probability), others were either split among different small subclades (e.g. Porifera, with the aforementioned Demospongiae placed in a basal position, and Calcarea/Hexactinellida scattered in different positions), or grouped together in clades with moderate posterior probability support but not in line with the know relationship among species.



Supplementary Figure 2: Bayesian phylogeny of metazoan OvoA sequences, inferred based on a LG+I+G model of molecular evolution and four independent MCMC chains run for 3,000,000 generations. Posterior probabilities support values are shown close to each node. Numbers near the species name indicate multiple sequences identified in the same species. Choanoflagellate OvoA sequences pertaining to clade I (see Figure 1A) were used for rooting purposes. The scale bar represents the number of substitutions per site. The multiple sequence alignment, Bayesian and ML tree files are available in supplementary file 2.

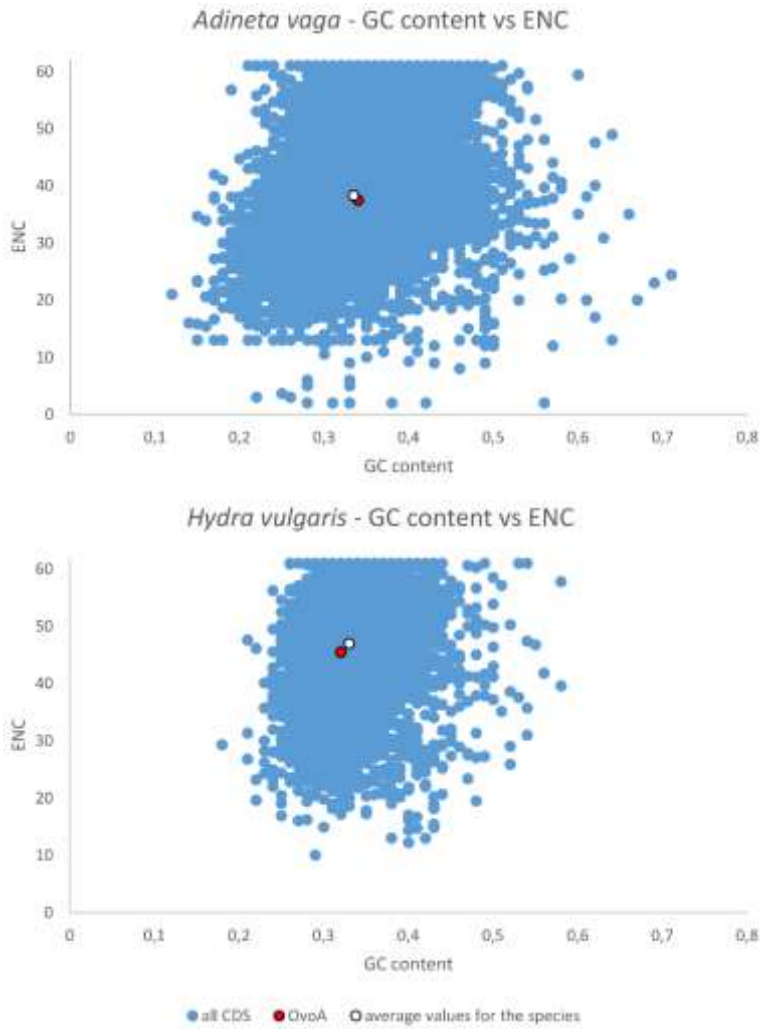
Supplementary Data Note 3: number of OvoA genes in metazoan genomes

Most metazoan genomes were found to harbor a single functional OvoA, with a few exceptions. For example, the genomes of the sponge *Amphimedon queenslandica* and the placozoan *Trichoplax adherens* contain two tandemly replicated OvoA, both actively transcribed [5,6]. Some deuterostomes possess two or more OvoA gene copies, even though the actual status of genome assemblies did not enable to discriminate between paralogous gene copies and uncollapsed allelic variants. The hemichordate *Saccoglossus kowalevskii*, for example, possess two similar gene copies on the same genomic scaffold, but transcriptomic evidence from the congeneric species *Saccoglossus mereschkowskii* suggests that only a single locus is expressed, and that the second gene might be a pseudogenic paralog. *Ptychodera flava*, another hemichordate species, also possess two highly similar paralogous genes, but one of them displays exonic indels that would impair its functionality. Among Urochordata, *Ciona savignyi*, whose genome is highly heterozygous [7], possess three highly similar OvoA genes. Gene duplications also occurred in cephalochordates, with two tandemly repeated OvoA, sharing 81% identity in *Branchiostoma belcheri*. Rotifers display a variable number of gene copies, depending on the species. *Rotaria macrura*, for example, possess two OvoA genes, located on the same genomic scaffold, which encode highly similar proteins (91% identity); a third pseudogene, containing multiple nonsense mutations, is placed on a different scaffold.

Supplementary Data Note 4: hydrozoan and rotifer genes are not the product of contaminant exogenous genomic DNA sequence

The possibility that the OvoA genes identified in Hydrozoa and in Rotifera (Bdelloidea) could be originated by a contamination of the genetic material used to generate genome sequencing libraries with genomic DNA of exogenous origin was tested by plotting two important metrics related with nucleotide composition (GC content) and codon usage, whose combination usually reveals lineage-specific signatures.

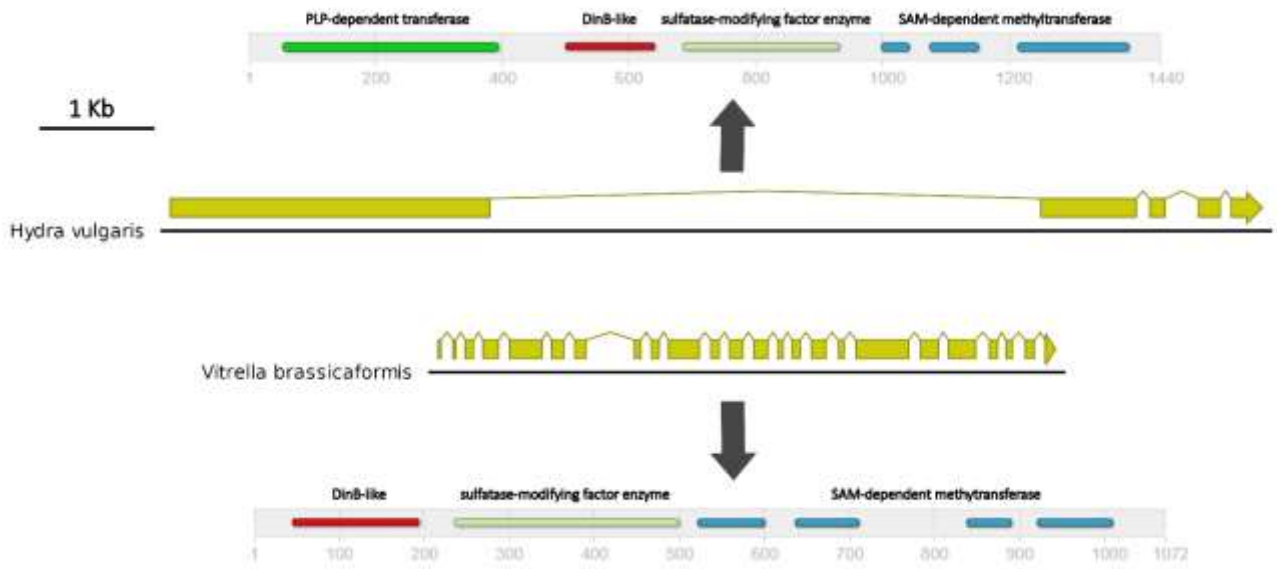
In both cases, the OvoA gene displayed metrics very close to the average values observed in the genomes, indicating that the hypothesis of an exogenous contamination was extremely unlikely. At the same time, one might argue that horizontally-transferred genes do originally carry a nucleotide and codon composition similar to those of the donor species. While this is, in principle, correct, several studies have indicated that genes subject to HGT, once fixed in the new genome, tend to quickly modify their GC content and codon composition to match those of the new organisms due to selective forces [1,2]. Since we infer that these two HGT events are ancient, we might expect such selective forces to have reshaped the nucleotide composition of the two OvoA genes to match the codon bias of the species, to optimize translational speed and accuracy.



Supplementary Figure 3: Scatter plots showing between GC content and effective number of codons (ENC) of all the coding sequences (CDS) extracted from the genomes of *Adineta vaga* and *Hydra vulgaris*, compared with OvoA. Average values are shown for reference.

Supplementary Data Note 5: organization of the hydrozoan OvoA gene

Although the chromerid *Vitrella brassicaformis* OvoA gene has been identified by phylogenetic analyses as the closest known relative to the horizontally-transferred OvoA gene of *Hydra vulgaris* (see Figure 1B in the main text), the comparative analysis of the genome of the two species [3,4] revealed a completely different genomic architecture. The additional N-terminal PLP-dependent transferase domain of *Hydra* was found to be entirely encoded by the very large exon 1, which did also contain the DinB-like and FGE sulfatase domains. Hence, the alignable portion of the two proteins was split among just 5 exons in *H. vulgaris*, and 24 exons in *V. brassicaformis* (Supplementary Figure 3). The lack of similarity between intron number and position of the two species suggest a very ancient origin for the hypothesized HGT event, which would have permitted a massive re-arrangement of the locus, which included a fusion with a second OvoB-like gene. This observation is consistent with the data presented in Figure 2, as the hydrozoan OvoA HGT event has been inferred to have taken place before the split between Trachylinae and Hydroidoline, but after the separation between Hydrozoa and all the other major cnidarian lineages.



Supplementary Figure 4: comparison between the gene structure and protein domain architecture of *Hydra vulgaris* OvoA and its closest relative identified by phylogenetic analyses, the chromerid *Vitrella brassicaformis*.

Supplementary Data Note 6: parallelism between rotiferan OvoA-like genes and EgtD

The OvoA-like sequences found in bdelloid rotifers are highly divergent from canonical OvoA genes and possess a different C-terminal domain, i.e. a SAM-dependent histidine-specific methyltransferase (PF10017) (Supplementary Figure 4). The origins of these genes, likely to derive from HGT due to their intron-less nature, are uncertain. Indeed, the encoded proteins only displayed a rather low degree of similarity with bacterial EgtB sequences, with a generalized lack of significant matches in the C-terminal domain region, which is characterized by the presence of a domain similar to the domain present in bacterial EgtD proteins. Curiously, the genome of the cyanobacterium *Moorea producens* encodes a canonical OvoA protein and a second monofunctional sulfoxide synthase, which lacks the C-terminal methyltransferase domain. This is a typical feature of most cyanobacterial OvoA-like proteins, which are co-encoded with an EgtD homolog. The presence of an Asn-Trp motif in cyanobacterial OvoA-like proteins is associated with the production of ergothioneine instead of ovothiol [8], implying that the acquisition of an EgtD gene by an ovothiol-producing organism (e.g. an ancient cyanobacterium) has been likely sufficient to enable the production of ergothioneine. Hence, bdelloid rotifers might represent an interesting target for future investigations aimed at the identification of alternative metazoan thiol-biosynthetic pathways, due to the contemporary presence of canonical OvoA and atypical OvoA-like proteins with an EgtD-like domain.

Supplementary figure 5: schematic domain organization of rotiferan Ovoa-like proteins, as exemplified by the case of *Rotaria macrura*.

References

1. Koski LB, Morton RA, Golding GB. 2001 Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol. Biol. Evol.* **18**, 404–412. (doi:10.1093/oxfordjournals.molbev.a003816)

2. Friedman R, Ely B. 2012 Codon usage methods for horizontal gene transfer detection generate an abundance of false positive and false negative results. *Curr. Microbiol.* **65**, 639–642. (doi:10.1007/s00284-012-0205-5)

3. Chapman JA *et al.* 2010 The dynamic genome of Hydra. *Nature* **464**, 592–596. (doi:10.1038/nature08830)

4. Oborník M *et al.* 2012 Morphology, ultrastructure and life cycle of Vitrella brassicaformis n. sp., n. gen., a novel chromerid from the Great Barrier Reef. *Protist* **163**, 306–323. (doi:10.1016/j.protis.2011.09.001)

5. Kamm K, Osigus H-J, Stadler PF, DeSalle R, Schierwater B. 2018 Trichoplax genomes reveal profound admixture and suggest stable wild populations without bisexual reproduction. *Sci. Rep.* **8**, 11168. (doi:10.1038/s41598-018-29400-y)

6. Fernandez-Valverde SL, Calcino AD, Degnan BM. 2015 Deep developmental transcriptome sequencing uncovers numerous new genes and enhances gene annotation in the sponge Amphimedon queenslandica. *BMC Genomics* **16**, 387. (doi:10.1186/s12864-015-1588-z)

7. Kim JH, Waterman MS, Li LM. 2007 Diploid genome reconstruction of Ciona intestinalis and comparative analysis with Ciona savignyi. *Genome Res.* **17**, 1101–1110. (doi:10.1101/gr.5894107)

8. Liao C, Seebeck FP. 2017 Convergent Evolution of Ergothioneine Biosynthesis in Cyanobacteria. *Chembiochem Eur. J. Chem. Biol.* **18**, 2115–2118. (doi:10.1002/cbic.201700354)