

The Origin and Evolution of Vertebrate Glycine Transporters

Max Shpak · Luciana Girotto Gentil ·
Manuel Miranda

Received: 6 November 2013 / Accepted: 20 February 2014 / Published online: 12 March 2014
© Springer Science+Business Media New York 2014

Abstract In the vertebrate central nervous system, glycinergic neurotransmission is regulated by the action of the glycine transporters 1 and 2 (GlyT1 and GlyT2)—members of the solute carrier family 6 (SLC6). Several invertebrate deuterostomes have two paralogous glycine carrier genes, with one gene in the pair having greater sequence identity and higher alignment scores with respect to GlyT1 and the other paralog showing greater similarity to GlyT2. In phylogenetic trees, GlyT2-like sequences from invertebrate deuterostomes form a monophyletic subclade with vertebrate GlyT2, while invertebrate GlyT1-like proteins constitute an outgroup to both the GlyT2-like proteins and to vertebrate GlyT1 sequences. These results are consistent with the hypothesis that vertebrate GlyT1 and GlyT2 are, respectively, derived from GlyT1- and GlyT2-like genes in

invertebrate deuterostomes. This implies that the gene duplication which gave rise to these paralogs occurred prior to the origin of vertebrates. GlyT2 subsequently diverged significantly from its invertebrate orthologs (i.e., through the acquisition of a unique *N*-terminus) as a consequence of functional specialization, being expressed principally in the lower CNS; while GlyT1 has activity in both the lower CNS and several regions of the forebrain.

Keywords Glycine transporters · Paralogy · Gene duplication

Introduction

The amino acid glycine is a significant inhibitory neurotransmitter, principally expressed in the spinal cord and brain stem—where it regulates a wide range of motor and sensory functions (Coleman et al. 2011). Glycine is cleared from the synaptic cleft, and inhibitory neurotransmission is terminated by the uptake of glycine into the presynaptic neuron through the action of two glycine transporters—GlyT1 and GlyT2. These proteins are encoded by two homologous genes (Liu et al. 1993; Morrow et al. 1998) which are located on human chromosomes 1 and 11, respectively.

The GlyTs are Na⁺,Cl⁻-dependent transporters belonging to the solute carrier family 6 (SLC6), with GlyT1 and GlyT2 identified as SLC6 members 9 and 5, respectively (Kristensen et al. 2001). Members of this gene family have a characteristic topology, including a large extracellular loop between domains 3 and 4 and intracellular *N*- and *C*-terminal tails (Gether et al. 2006). SLC6 proteins are widely distributed among eukaryotes, functioning as transporters of various signaling molecules in fungi and unicellular

Electronic supplementary material The online version of this article (doi:10.1007/s00239-014-9615-2) contains supplementary material, which is available to authorized users.

M. Shpak (✉)
NeuroTexas Institute, St. David's Healthcare, Austin, TX 78705,
USA
e-mail: shpak.max@gmail.com

M. Shpak
Center for Systems and Synthetic Biology, University of Texas,
Austin, TX 78712, USA

M. Shpak
Fresh Pond Research Institute, Cambridge, MA 02140, USA

L. G. Gentil · M. Miranda
Department of Biological Sciences, University of Texas at El
Paso, El Paso, TX 79968, USA
e-mail: lggentil@hotmail.com

M. Miranda
e-mail: mmiranda3@utep.edu

eukaryotes, and as metazoan neurotransmitters (Hoglund et al. 2005).

Human GlyT1 and GlyT2 share ~50 % amino acid sequence identity; their main structural differences are highlighted by the length of the *N*-terminus, with GlyT2 characterized by having an extended *N*-terminus (~200 amino acids in mammals) of unknown function (Aragon and Lopez-Corcuera 2003; Eulenberg et al. 2005).

In mammals, GlyT2 is solely expressed in the lower brain (brainstem and cerebellum) and in the spinal cord. GlyT1 is more widely expressed in the lower CNS, the diencephalon, retina, and the neocortex (Zafra et al. 1995; Geerlings et al. 2002). This functional differentiation suggests that glycine transporter evolution was driven by gene duplication and functional divergence (Ohno 1970). This study elucidates the evolutionary history of vertebrate glycine transporters, specifically their phylogenetic relationship to one another and to related genes in invertebrate deuterostomes.

Methods

Human GlyT1 and GlyT2 proteins were used as query sequences to conduct BLAST (Altschul et al. 1990) searches in GenBank, using default gap penalties and BLOSUM62 scoring matrices. Each search was restricted by taxon.

The nematode *Caenorhabditis elegans* was chosen as a representative protostome. The gene in the *C. elegans* genome with the highest BLAST alignment scores to GlyT1-2 is a GABA transporter. Therefore, the *C. elegans* GABA transporter served as both a protostome outgroup and as a non-glycine transporter outgroup within the SLC6 gene family. *Strongylocentrotus purpuratus* was selected as a representative non-chordate deuterostome (echinoderm), while *Branchiostoma floridae* and *Ciona intestinalis* were chosen as model invertebrate chordates. Representing non-mammalian vertebrates, the genomes of *Danio rerio* and *Xenopus laevis* were queried.

The chromosome positions of the GlyT1-2 homologs were used to identify overlapping or closely linked genes (within ~100 kb) annotated in the Genomic Context and Gene Mapping information in GenBank, in order to establish whether synteny among genes linked to GlyT1-2 was conserved across distantly related deuterostome taxa. These data were available for vertebrates, for *Ciona* and (partially) for *Strongylocentrotus* genomes.

Sequences with the highest alignment scores and the lowest *e*-values against GlyT1 and GlyT2, respectively, were aligned using ClustalX (Thompson et al. 1997) with BLOSUM62 scoring and default gap penalties. The multiple sequence alignments were imported into the protein

maximum parsimony and maximum likelihood modules in PHYLIP (Felsenstein 1989) to construct phylogenetic trees, the latter assuming a PAM model of amino acid substitution. A bootstrap analysis over 100 replicates was carried out for both the parsimony and likelihood trees to determine the level of support for each subclade in the consensus trees.

Results

Database searches

A complete list of sequences used in this analysis, along with their GenBank accession numbers, is provided in the supplementary file S1.

BLAST searches using human GlyT1 and GlyT2 recovered significant alignments in the *C. elegans* GABA transporter. Queried searches of the invertebrate deuterostome genomes revealed several SLC6 sequences homologous to both GlyT1 and GlyT2 with *e*-values at or effectively 0. One of each paralog pair in these deuterostomes gave higher alignment scores against human GlyT1 (typically annotated in GenBank as “GlyT1-like” or “SLC6 member 9-like”), while the second gave a higher score in alignments against human GlyT2 (i.e., “GlyT2-like” or as “SLC6 member 5-like”). Note that in the case of the *Ciona* glycine transporters, both paralogs gave higher alignment scores and lower *e*-values against GlyT2 and are annotated as such in GenBank. However, the alignment score of 672 for *Ciona*B (GlyT2-like) against a GlyT2 query was much higher than that of 547 for the *Ciona*A sequence.

Summary statistics of the BLAST alignments of GlyT1-2 and related invertebrate homologs against human sequence queries are shown in Table 1, including percent identity, alignment scores, and *e*-values. Essentially the same BLAST scores, *e*-values, and percent identity (albeit with coverage between 99 and 100 %) were obtained by performing reverse BLAST of the invertebrate sequences as queries against the human genome. By comparison, the alignment score, percent coverage, *e*-value, and percent identity for human GlyT2 in a BLAST search (with BLOSUM62) against its human paralog GlyT1 are 501, 79 %, 10^{-167} , and 44 %, respectively; consistent with a deeper divergence between GlyT1 and GlyT2 paralogs than among orthologs across any two species.

Phylogenetic inference

The consensus trees for both bootstrapped maximum likelihood (Fig. 1a) and maximum parsimony (Fig. 1b) phylogenies support the monophyly of deuterostome

Table 1 BLAST output using human query sequences

Query sequence	Human GlyT1				Human GlyT2			
	Score	Cover	<i>e</i> -value	Identity	Score	Cover	<i>e</i> -value	Identity
<i>Ciona</i> A	532	81 %	2e−180	47 %	547	74 %	0.0	49 %
<i>Ciona</i> B	587	81 %	0.0	50 %	672	74 %	0.0	57 %
<i>Branchiostoma</i> GlyT1-like	600	83 %	0.0	51 %	583	75 %	0.0	46 %
<i>Branchiostoma</i> GlyT2-like	629	83 %	0.0	51 %	715	79 %	0.0	56 %
<i>Strongylocentrotus</i> GlyT1-like	588	82 %	0.0	48 %	573	74 %	0.0	48 %
<i>Strongylocentrotus</i> GlyT2-like	578	82 %	0.0	46 %	610	74 %	0.0	51 %
<i>C. elegans</i> GABA transporter	432	78 %	1e−141	39 %	474	71 %	5e−157	43 %

Comparison of percentage identity, alignment scores, and *e*-values of human GlyT1 and GlyT2 query sequence against various homologs in a BLAST search, using the BLOSUM 62 scoring matrix. “Cover” and “Identity” measure the percentage of sequence overlap and the percentage of identical residues in the pairwise alignments, respectively

glycine transporters within the SCL6 gene family, as well as the monophyly of vertebrate GlyT2 + invertebrate GlyT2-like sequences. The branch lengths in Fig. 1a were computed under the assumption that the topology of the consensus tree represents an estimate of the actual phylogeny, as there is no standard method of computing the consensus branch lengths across the different topologies produced for each replicate.

Vertebrate GlyT1’s form a monophyletic clade, while the invertebrate GlyT1-like sequences are outgroups to both vertebrate GlyT1 and GlyT2/GlyT2-like subclades. The positions of specific invertebrate orthologs differ between the two trees, with either *Ciona* (parsimony) or *Strongylocentrotus* (likelihood) appearing as the immediate outgroup to the {GlyT2 + vertebrate GlyT1} clade.

Comparative genomics

For the invertebrate deuterostomes, chromosome location data were available for GlyT1 homologs in *Ciona*. Both GlyT1- and GlyT2-like sequences are located on chromosome 7 of the *Ciona* genome. Chromosome coordinates for GlyT1-2 homologs in representative taxa are provided in Table 2.

The table also lists genes that overlap with or are closely linked to (i.e., within ~100 kb) the glycine transporters. Note that synteny among the genes is conserved for several GlyT1- and GlyT2-linked genes between the *Danio* and human, but not the *Ciona* genomes. BLAST searches located *Ciona* homologs of CCDC24, RNF220, and ATP6VOB on chromosome 8 rather than 7 (with CCDC24 and RNF220 within 100 kb of one another); B4GALT2 and DMAP homologs map to *Ciona* chromosome 9. For human genes linked to GlyT2, *Ciona*’s homolog of NELL1 is located in chromosome 1; while its PRMT3 gene is located on chromosome 7 (at position 2095147–2087480, it is over

1 megabase from the GlyT2 homolog and >400 kb from the GlyT1-like gene). We remark that among the genes linked to the GlyT1-like gene in *Ciona* is an SLC6A7-like gene, homologous to the SLC6A7 proline transporter active in vertebrate brains.

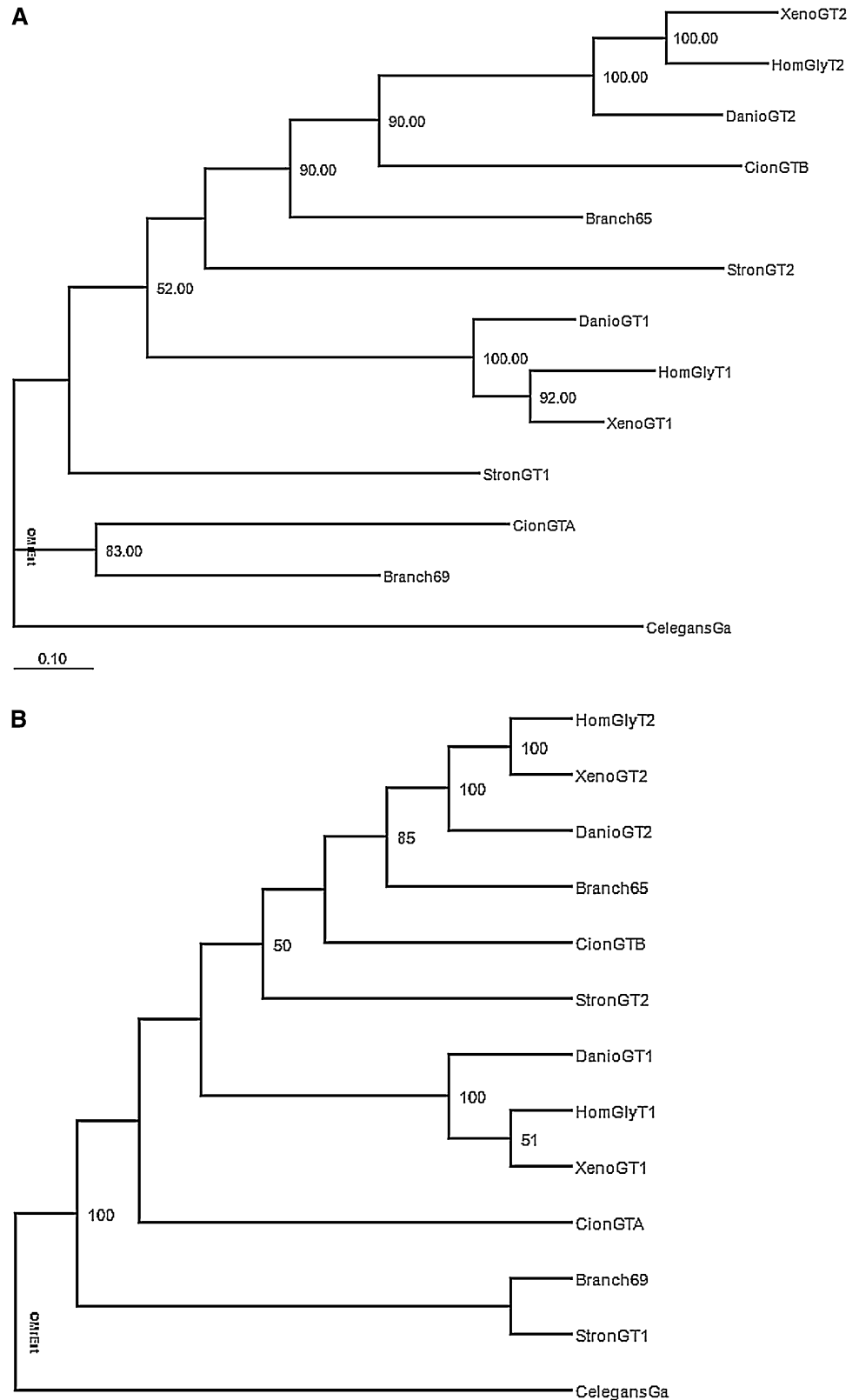
Chromosome mapping data are not available for *Branchiostoma* or *Strongylocentrotus* genomes, nor are the genes flanking GlyT1 and GlyT2 homologs identified and annotated in the former. Table 2 does include information on the identity of genes linked to GlyT1- and GlyT2-like sequences in the partially assembled genome of *Strongylocentrotus* provided in GenBank. None of the genes overlapping with or adjacent to GlyT1- or GlyT2-like genes is homologous to those linked to their respective (presumed) vertebrate orthologs, nor to those in *Ciona*.

Discussion

Phylogenetic analysis indicates that the origin of sequences ancestral to GlyT1 and GlyT2 predates the origin of vertebrates. This is supported by the observation that the genomes of invertebrate deuterostomes have two paralogous glycine transporters, one of which is unambiguously related to vertebrate GlyT2; while the second paralog is an outgroup to both GlyT2 and GlyT1 sequences in vertebrates. Strictly speaking, the phylogenies suggest that the invertebrate GlyT1-like sequences represent an ancestral deuterostome glycine transporter from which both GlyT2 and vertebrate GlyT1 genes were derived, rather than true orthologs to vertebrate GlyT1 as such.

A survey of genes linked to GlyT1 and GlyT2 in vertebrates indicates conservation of synteny between mammalian and teleost (*Danio*) sequences, but not with the invertebrate GlyT1 or GlyT2 homologs. This reflects the fact that mammals and ray-finned fish shared a common ancestor in the early Devonian (~400 ma) while

Fig. 1 a Consensus tree of maximum likelihood trees for vertebrate glycine transporters and representative homologous and representative homologous SLC6 family sequences. Bootstrap support values >50 are shown for each node (for 100 replicates), and branch lengths are computed for the consensus tree. Taxa and sequence abbreviations are as follows: Hom (human), Xeno (*Xenopus laevis*), and Danio (*Danio rerio*) with the suffixes GT1 and GT2 referring to GlyT1/2. Strong1/2, CionaA/B, and Branch 69/65 refer to the GlyT1 and GlyT2-like sequences, respectively, from *Strongylocentrotus purpuratus*, *Ciona intestinalis*, and *Branchiostoma floridae*. Branch lengths are on a scale of 0–1.0 (note the *scale bar*, and the length of the invertebrate sequence branches relative to vertebrate terminal branches). Trees and tree annotation were formatted using MrEnt (Zuccon and Zuccon 2006). **b** Consensus tree for 100 bootstrap replicates of the maximum parsimony phylogeny, computed for the same sequences (and with the same notation) as in **a**



vertebrates and invertebrate chordates diverged in the early Cambrian (~540 ma), while echinoderms and chordates shared a common ancestor earlier still (e.g., Valentine

2004). One would not expect to see extensive conservation of gene order or syntenic gene clusters across such distantly related clades.

Table 2 Chromosome coordinates of GlyT1 and GlyT2 homologs in representative vertebrates and invertebrate chordates, together with a list of known genes that overlap with or located within <100 kb from GlyT1 and GlyT2 (those marked ^a slightly exceed 100 kb distance)

Synteny data are not available for *Branchiostoma* genomes near GlyT1 and GlyT2 homologs

	Chromosome	Locations	Linked genes
Human GlyT1	1	44462155–44497171	CCDC24, KLF17, B4GALT2, ATP6VOB, DMAP1, and RNF220 ^a
Human GlyT2	11	20620946–20676610	PRMT3 and NELL1
<i>Danio</i> GlyT1	2	33053161–33134536	CCDC24, KLF17, and RNF220
<i>Danio</i> GlyT2	7	18682162–18700849	PRMT3 ^a
<i>Ciona</i> A	7	1620438–1629689	SLC6A7, TRIM71-like, and Na ⁺ dependent glucose transporter 1C-like
<i>Ciona</i> B	7	927306–934934	Sialin-like
<i>Strongylocentrotus</i> , GlyT1-like	Unknown	Unknown	Retinol dehydrogenase 8-like, histone H4-like, histone H3-like, and methionine adenosyltransferase 2-like
<i>Strongylocentrotus</i> , GlyT2-like	Unknown	Unknown	Protein O-linked mannose- β acetylglucosaminyltransferase and GDP-mannose transporter-like

Among orthologs, the GlyT2-like subclade topology is congruent with what is known of chordate phylogeny, i.e., *Ciona* is an outgroup to the vertebrates and *Branchiostoma* to the other chordate sequences (consistent with Delsuc et al. 2006). With GlyT1-like sequences, however, we have the anomalous position of invertebrate chordates as an outgroup relative to echinoderms and vertebrates in the likelihood tree, and the grouping of *Strongylocentrotus* and *Branchiostoma* as a subclade in the parsimony tree. These nodes are weakly supported (<50 % bootstrap values), presumably as a consequence of long-branch attraction (Felsenstein 1978) of the invertebrate sequences. Specifically, note that in Fig. 1a, the longest branches are associated with the invertebrate sequences. The branch lengths are on a scale of 0–1.0, with the vertebrate terminal sequences between 0.13 and 0.16. In contrast, the GlyT2-like sequences give branch lengths of 0.46, 0.36, and 0.66 for *Ciona*, *Branchiostoma*, and *Strongylocentrotus*, respectively (see supplementary PHYLIP output file S2 documenting the branch lengths in Fig. 1a). The GlyT1-like sequences in these respective taxa are 0.53, 0.37, and 0.52. Regardless, the inconsistency in the topology of ortholog subtrees does not contradict the main thesis of vertebrates inheriting GlyT1/T2-like paralogs from invertebrate ancestors.

The outgroup position of invertebrate GlyT1-like sequences implies that the ancestral glycine transporter in deuterostomes was similar to GlyT1, and that the duplication giving rise to GlyT2-like genes occurred prior to the divergence of echinoderms and chordates from their common ancestor. This ancestral pair of glycine transporters was inherited by early vertebrates. GlyT1 retained the characteristics of the ancestral GlyT1-like invertebrate sequences, while GlyT2 acquired novel structures (i.e., N-terminus) and function.

The divergence of vertebrate GlyT2 is a consequence of functional divergence between the paralogs. With increased cephalization in vertebrates, GlyT2 became specialized for activity in the brainstem and lower central nervous system; while GlyT1 remained a “generalist” neurotransmitter, expressed in the lower CNS and in the forebrain. As the vertebrate brain became increasingly developed relative to the ganglia of invertebrate chordates, the conservation of the GlyT1 paralog facilitated GlyT2’s specialization. An open area for future investigation is determining the evolutionary history and mechanism of origin of the N-terminus in vertebrate GlyT2 proteins, whose ancestral sequences resembled GlyT2-like proteins in invertebrates which lack this structural characteristic.

Acknowledgments This study was supported by grant NIMH 5SC1MH 086070-04 to MM and also in part by the NIH grant DA029989 to Edward Castaneda.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Aragon C, Lopez-Corcuera B (2003) Structure, function and regulation of glycine neurotransmitters. *Eur J Pharmacol* 479:249–262
- Coleman WL, Fischl MJ, Weimann SR, Burger RM (2011) GABAergic and glycinergic inhibition modulate monaural auditory response properties in the avian superior olivary nucleus. *J Neurophysiol* 105:2405–2420
- Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439:965–968
- Eulenburg V, Arnsen W, Betz H, Gomeza J (2005) Glycine transporters: essential regulators of neurotransmission. *Trends Biochem Sci* 30:325–333
- Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* 27:401–410
- Felsenstein J (1989) PHYLIP—phylogeny inference package (Version 3.2). *Cladistics* 5:164–166

- Geerlings A, Nunez E, Rodenstein L, Lopez-Corcuera B, Aragon C (2002) Glycine transporter isoforms show differential subcellular localization in PC12 cells. *J Neurochem* 82:58–64
- Gether U, Andersen PH, Larsson OM, Schousboe A (2006) Neurotransmitter transporters: molecular function of important drug targets. *Trends Pharmacol Sci* 27:375–383
- Hoglund PJ, Adzic D, Scicluna SJ, Lindblom J, Fredriksson R (2005) The repertoire of solute carriers of family 6: identification of new human and rodent genes. *Biochem Biophys Res Commun* 336:175–189
- Kristensen JA, Jorgensen TN, Eriksen J, Loland CJ, Stromgaard K, Gether U (2001) SLC6 neurotransmitter transporters: structure, function, and regulation. *Pharmacol Rev* 63:585–640
- Liu QR, Lopez-Corcuera B, Mandiyan S, Nelson H, Nelson N (1993) Cloning and expression of a spinal cord- and brain-specific glycine transporter with novel structural features. *J Biol Chem* 268:22802–22808
- Morrow JA, Dunbar DR, Walker GB, Shahid M, Hill DR (1998) Molecular cloning and functional expression of the human glycine transporter GlyT2 and chromosomal localisation of the gene in the human genome. *FEBS Lett* 439:334–340
- Ohno S (1970) *Evolution by gene duplication*. Springer, New York
- Thompson JD et al (1997) The clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Valentine JW (2004) *On the origin of phyla*. University of Chicago Press, Chicago
- Zafra F, Aragon C, Olivares L, Danbold NC, Gimenez C, Storm-Mathisen J (1995) Glycine transporters are differentially expressed among CNS cells. *J Neurosci* 15:3952–3969
- Zuccon A, Zuccon D (2006) MrEnt v1.2. Program distributed by the authors. Department of Vertebrate Zoology & Molecular Systematics Laboratory, Swedish Museum of Natural History, Stockholm