

A wide-range phylogenetic analysis of Zic proteins: Implications for correlations between protein structure conservation and body plan complexity[☆]

Jun Aruga^{a,*}, Akiko Kamiya^a, Hirokazu Takahashi^a, Takahiko J. Fujimi^a, Yuri Shimizu^a,
Keiko Ohkawa^a, Shigenobu Yazawa^{b,c}, Yoshihiko Umesono^{b,c}, Hideki Noguchi^d,
Takashi Shimizu^e, Naruya Saitou^f, Katsuhiko Mikoshiba^g, Yoshiyuki Sakaki^d,
Kiyokazu Agata^{b,c}, Atsushi Toyoda^d

^a Laboratory for Comparative Neurogenesis, RIKEN Brain Science Institute, Wako 351-0198, Japan

^b RIKEN Center for Developmental Biology, Kobe 650-0047, Japan

^c Laboratory for Molecular Developmental Biology, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan

^d Sequence Technology Team, RIKEN Genomic Science Center, Yokohama 230-0045, Japan

^e Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

^f Division of Population Genetics, National Institute of Genetics, Mishima 4511-8540, Japan

^g Laboratory for Developmental Neurobiology, RIKEN Brain Science Institute, Wako 351-0198, Japan

Received 4 October 2005; accepted 16 February 2006

Available online 29 March 2006

Abstract

We compared Zic homologues from a wide range of animals. Striking conservation was found in the zinc finger domains, in which an exon–intron boundary has been kept in all bilaterians but not cnidarians, suggesting that all of the bilaterian Zic genes are derived from a single gene in a bilaterian ancestor. There were additional conserved amino acid sequences, ZOC and ZF-NC. Combined analysis of the zinc finger, ZOC, and ZF-NC revealed the presence of two classes of Zic, based on the degree of protein structure conservation. The “conserved” class includes Zic proteins from the Arthropoda, Mollusca, Annelida, Echinodermata, and Chordata (vertebrates and cephalochordates), whereas the “diverged” class contains those from the Platyhelminthes, Cnidaria, Nematoda, and Chordata (urochordates). The result indicates that the ancestral bilaterian Zic protein had already acquired an entire set of conserved domains, but that this was lost and diverged in the platyhelminthes, nematodes, and urochordates.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Comparative genomics; Evolution; Body plan; Exon–intron boundary; Transcription factor; Protein structure; Conserved domain; Zic; Zinc finger

Introduction

Recent studies of Zic family genes have revealed their roles in the development of various animals. In vertebrates, they are involved in neural and neural crest development, skeletal patterning, and left–right axis establishment (reviewed in

[1,2]). In urochordates, they participate in the determination of cell fate toward neural, notochord, and muscle cells [3–5]. In the Protostomia, a fly Zic homologue, Opa, regulates segmentation and midgut morphogenesis [6,7], whereas the nematode Zic homologue, Ref-2, has a role in vulval development [8]. Although the Zic proteins commonly have roles in cell fate decision in the early embryonic stages, there seems to be significant phylogenetic variability in these roles. For example, their role in neural development has been demonstrated in the Deuterostomia, but not in the Protostomia. However, in the Cnidaria, hydra Zic (HyZic) is expressed in a subset of neural cell precursors, suggesting the involvement of these proteins in neural development, as is the case in

Abbreviations: ZF, zinc finger; ZOC, Zic-Opa conserved domain; ZF-NC, zinc finger–N-flanking conserved region; AA, amino acid residue(s); ORF, open reading frame.

[☆] Sequence data from this article have been deposited with the EMBL/GenBank/DBJ Data Libraries under Accession Nos. AB231864–AB231884.

* Corresponding author. Fax: +81 48 467 9792.

E-mail address: jaruga@brain.riken.jp (J. Aruga).

vertebrates [9]. It is possible that structural change in the Zic genes contributed to the establishment of a body plan unique to each animal.

Although the phylogenetic roles of Zic genes are interesting, our understanding may not be sufficient. This seems to be partly due to the fact that previous studies have dealt with small numbers of representative model animals. These circumstances led us to investigate the structure of Zic genes in a broad range of animals, particularly in animal phyla from which Zic homologues have not been reported. Here, we report the genomic structures and predicted amino acid sequences of newly identified Zic genes in the Cnidaria, Platyhelminthes, Annelida, Mollusca, Arthropoda, and Echinodermata, and we compare them with the sequence data currently available in public databases. The results reveal an overall picture of the Zic genes' evolutionary process, suggesting the involvement of Zic protein structure diversification in body plan simplification.

Results

Zic gene homologues are widely distributed in animal genomes

We first searched Zic gene homologues in each kingdom of living organisms by performing a homology search against current databases. We found no obvious Zic

homologues in bacteria, algae, plants, fungi, and protists, although there are a number of ZF domain-encoding genes in these organisms. The highest level of similarity, with the exception of the metazoans, was found in some fungal proteins (data not shown). However, a homology search using these fungal sequences revealed that they were more closely related to GLIS, another subfamily of ZF proteins in the metazoans, raising the possibility that the fungal ZF protein was derived from a common ancestor with GLI/GLIS/ZIC ZF superfamily proteins [1]. On the basis of these results, we assumed that the presence of Zic homologues was limited to metazoans.

To identify Zic gene homologues in metazoans, we established PCR conditions under which we could amplify a region in the ZF domain of Zic-related genes by using nested degenerate primers. PCR allowed us to clone the Zic homologues from a broad range of animals, including cnidarians, platyhelminthes, annelids, molluscs, nematodes, arthropods, echinoderms, and chordates (Table 1). In addition to our cloning and sequence determination, the rapidly increasing whole genome sequence project benefited searching and comparing Zic homologues. However, our attempt to clone a sponge homologue of Zic was not successful (J.A., Naoyuki Iwabe, K.A., Takashi Miyata, unpublished observation). As a whole, these results strongly suggest that Zic genes are retained very widely in the Eumetazoa.

Table 1
Animals used in this study

Phylum	Animal species	Abbr.	Intron	Genome	Genome clone (accession no.)	cDNA accession no. [reference]
Cnidaria	<i>Scolionema suvaense</i>	<i>Ssu</i>	D	Fosmid	Sco26, Sco36 (AB231882)	AB231883
	<i>Hydra vulgaris</i>	<i>Hv</i>	0	Database		[9]
	<i>Nematostella vectensis</i>	<i>Nv</i>	0	BAC	CH314-49A19 (AB231867) CH314-55K22 (AB231868)	
Platyhelminthes	<i>Dugesia japonica</i>	<i>Dj</i>				AB231880, AB231881
	<i>Schmidtea mediterranea</i>	<i>Sme</i>		Database		
	<i>Schistosoma mansoni</i>	<i>Sma</i>	A	BAC	CH103-42N16 (AB231864)	
Annelida	<i>Tubifex tubifex</i>	<i>Tt</i>	A	Fosmid	Tub8, Tub12 (AB231869)	AB231870
Mollusca	<i>Loligo bleekeri</i>	<i>Lb</i>				AB231874
	<i>Octopus ocellatus</i>	<i>Oo</i>				AB231875
	<i>Corbicula</i> sp.	<i>Cj</i>	A	Fosmid	Cor8, Cor10 (AB231873)	
	<i>Spisula solidissima</i>	<i>Sso</i>	A	BAC	CH312-9E16 (AB231865)	
Nematoda	<i>Caenorhabditis elegans</i>	<i>Ce</i>	ACBE	Database		[8]
Arthropoda	<i>Drosophila melanogaster</i>	<i>Dm</i>	AB	Database		[6,7]
	<i>Anopheles gambiae</i>	<i>Ag</i>		Database		XM_321856
	<i>Pandinus imperator</i>	<i>Pi</i>	A	PCR	(AB231876)	AB231877
	<i>Artemia franciscana</i>	<i>Af</i>	AB	Fosmid, PCR	Art1 (AB231878)	AB231879
	<i>Strongylocentrotus purpuratus</i>	<i>Sp</i>	A	Database		
Chordata	<i>Asterina pectinifera</i>	<i>Ap</i>	A	Fosmid	Ast1 (AB231871)	AB231872
	<i>Ciona intestinalis</i>	<i>Ci</i>	AC	Database		[31]
	<i>Ciona savignyi</i>	<i>Cs</i>		Database		[5]
	<i>Halocynthia roretzi</i>	<i>Hr</i>		Database		[3,4]
	<i>Branchiostoma floridae</i>	<i>Bf</i>	A	BAC	CH302-61N11 (AB231866)	[19]
	<i>Xenopus</i> sp.	<i>Xl</i>	A, AB	Database		Ref. in [1]
	<i>Mus musculus</i>	<i>Mm</i>	A, AB	Database		Ref. in [1]
	<i>Homo sapiens</i>	<i>Hs</i>	A, AB	Database		Ref. in [1]

Database means that sequence information was obtained from public databases. Intron indicates the types of intron. Intron 0 indicates no introns in the zinc finger domain.

The copy numbers of Zic in each species varied from 1 to 5. Two Zic genes were identified in two planarians [*Dj* and *Sme* (ZicA and ZicB)] and three urochordates [*Ci*, *Cs*, and *Hr* (ZicL/ZicN and macho1)]. Five Zic genes (ZicA, ZicB, ZicC, ZicD, and ZicE) were present in a sea anemone (*Nv*). Mammals (*Hs*, *Mm*) possessed five related genes. Five Zic genes also were identified in other vertebrate species such as *Xl* and *Gallus gallus* (T.J.F., Y.S., A.T., J.A., unpublished observation). Only a single Zic species was identified in each of the remaining animals surveyed in this study.

The ZF domain amino acid sequence is highly conserved

We next deduced the amino acid (AA) sequence of each Zic protein by sequencing cDNA and/or genomic clones and then subjected the putative protein sequences to comparative analysis (Fig. 1). The highest degree of conservation was found in the ZF domain. The similarity of the AA sequences of any two species was more than 75%, and there were 60 absolutely conserved AA (yellow letters in Fig. 1) within the region ranging from ZF2 to ZF5 (104 AA). The conservation in this region was markedly higher than that in the other regions. There were several clusters of absolutely conserved AA sequences in the ZF2–ZF5 region, and in many positions AA changes within equivalent AA groups were observed.

Some new findings were revealed when we examined the variation within the highly conserved ZF1–ZF5 region. First, vertebrate Zic1–3 had a unique structure in ZF5 whereby the number of AA between the first and the second cysteine residues was 2 instead of 4. Second, the number of AA between the two cysteines of the ZF1 C2H2 motif was highly variable, ranging from 6 (nematoda) to 38 (mosquito). Third, ZF1 was not as strongly conserved as ZF2–ZF5.

Comparison among entire amino acid sequences reveals the presence of evolutionarily conserved domains other than the ZF domain

Comparison of AA sequences in the entire ORFs revealed that there were two additional evolutionarily conserved sequences (ZF-NC and ZOC). ZF-NC could be seen in the N-terminal flanking region of the ZF domain (Fig. 1). The consensus sequence of ZF-NC can be summarized as GAF(F/L)RYMRQP-(0–7 AA)-IKQE. Sequences perfectly matching the ZF-NC consensus could be seen in molluscs, arthropods, annelids, and some chordates (vertebrates and cephalochordates), but not in cnidarians, platyhelminthes, nematodes, or urochordates. However, ZF-NC-lacking species contained minimal conservations that could have been residues of the ZF-NC sequence. ZOC (Zic–Opa conserved domain, Fig. 2A) is the N-terminally located domain, which was found as a sequence conserved between mouse Zic1–3 and fly Odd-paired [10,11]. The consensus sequence can be summarized as (S/T)RDFLxxxR. This sequence motif was conserved in

arthropods, echinoderms, molluscs, annelids, and some chordates (vertebrates and cephalochordates), but not in cnidarians, platyhelminthes, nematodes, or urochordates (Fig. 2B).

Molecular phylogenetic tree analysis reveals two Zic protein clusters

To understand their relationships in light of these similarities, we constructed a molecular phylogenetic tree on the basis of the amino acid sequences of those ZF domains with 20 ZF-NC-containing amino acids at the N-terminal and 10 amino acids in the C-terminal flanking sequences (Fig. 3). We found that Zic proteins were clustered into two classes: “conserved” and “diverged.” The conserved class included the Vertebrata, Cephalochordata, Mollusca, Arthropoda, Echinodermata, and Annelida. The diverged class, characterized by longer branches in the tree, included the Cnidaria, Nematoda, Platyhelminthes, and Urochordata. Interestingly, the ZOC domain was confined to the conserved class, not the diverged class (Figs. 2B and 3).

Exon–intron organization of Zic genes reveals the existence of a single ancestral bilateralian Zic gene

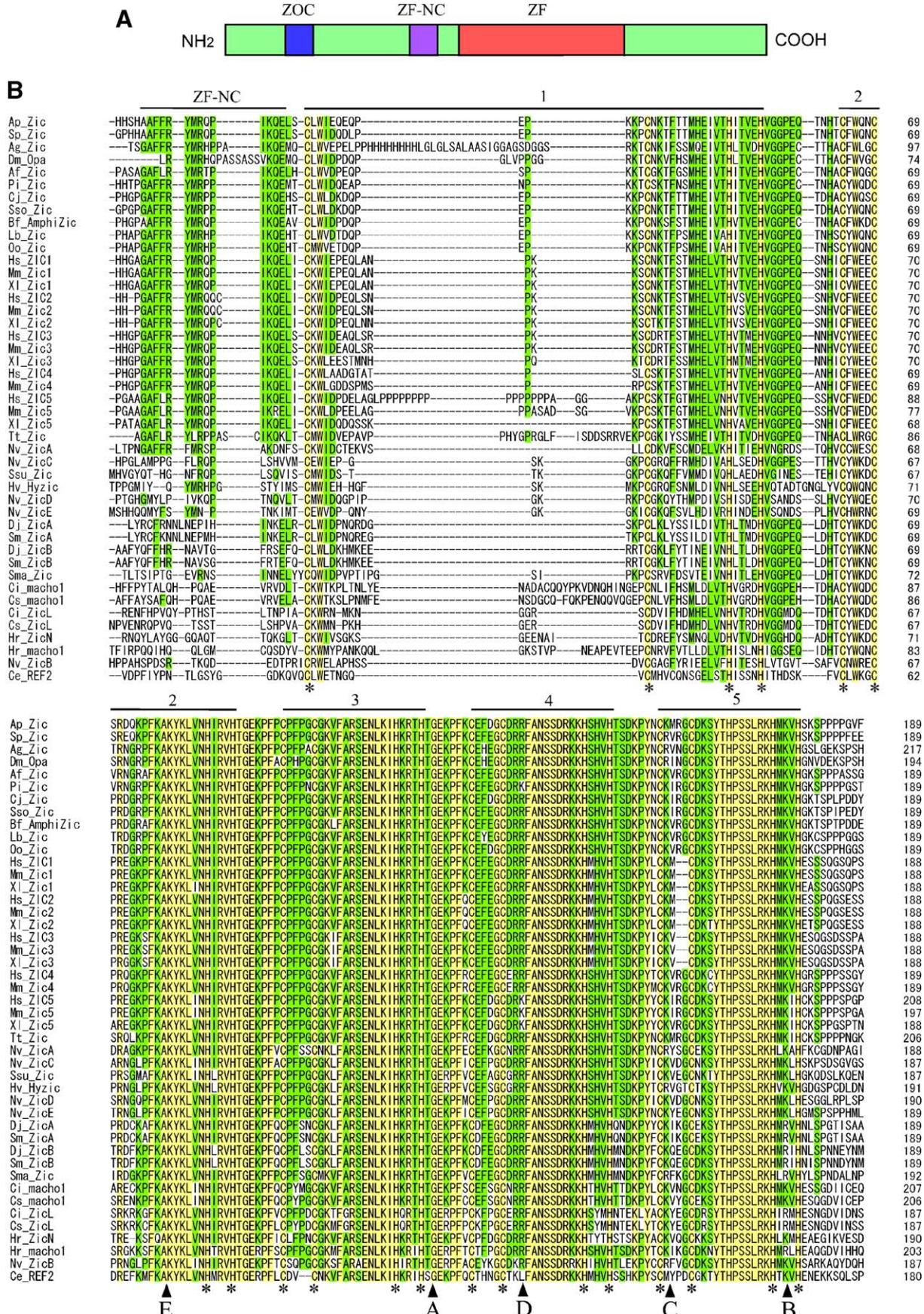
In our examination of genomic organization we focused on the exon–intron boundaries in the ZF domain (Figs. 1 and 4). As indicated in Fig. 1, exon–intron borders were recognized in five locations (A–E). Among them, an exon–intron border located between ZF3 and ZF4, which we refer to hereafter as A-intron (border), was extraordinarily conserved and was distributed in all of the Zic homologues except seven cnidarian Zics, suggesting that the A-intron is a bilateralian-specific structure of Zic genes.

In contrast to the A-intron, the remaining exon–intron borders, designated here as B-, C-, D-, or E-intron, were each found in a small number of animals. B-intron was found in the vertebrate Zic1–3 and certain groups (Hexapoda and Crustacea) of the Arthropoda. C-intron was seen in the two types of ascidian Zic and nematode Ref-2. Nematode Ref-2 uniquely contained another intron (E-intron) in addition to the A-, B-, and C-introns in its ZF domain. In cnidarians such as the jellyfish (*Ssu*), introns (D-introns) occurred in positions of the Zic genes different from those mentioned above. In the hydra (*Hv*) and the sea anemone (*Nv*), however, there were no insertions of introns within the ZF domain [0 (zero)-intron] at all.

A tandem array of multiple Zic genes was revealed in *Nv* and *Hs*. In the case of the cnidarian *Nv*, the three genes ZicC, ZicA, and ZicB were tandemly arrayed over an approximately 80-kb region (Fig. 4). A tandem array of Zic genes was also seen in vertebrate Zic1/Zic4 pairs as well as Zic2/Zic5 pairs; these genes were arrayed in a head-to-head orientation (Fig. 4).

Discussion

Recent molecular phylogenetic analysis indicates that there are three main taxa in bilateralian animals: Deuterostomia,



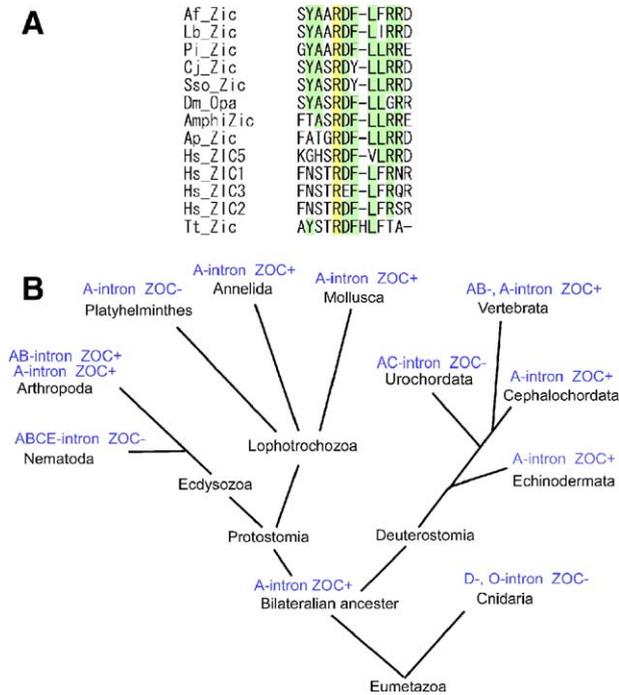


Fig. 2. (A) Amino acid alignment of the ZOC domain and (B) its distribution in a phylogenetic tree. Phylogenetic trees were adapted and modified from Tessmar-Rable and Arendt [37]. Types of exon–intron borders are also indicated.

Ecdysozoa, and Lophotrochozoa (reviewed in [12,13]). The Deuterostomia include echinoderms and the chordates, the latter of which have three groups: vertebrates, cephalochordates (lancelets), and urochordates (ascidians). Ecdysozoa and Lophotrochozoa are two sister groups in the Protostomia. Ecdysozoa include arthropods and nematodes. The Lophotrochozoa contain annelids, molluscs, and platyhelminthes. Phylogenetic trees based on molecular phylogeny are not always consistent with classical phylogenetic classifications based on morphological criteria, such as the presence or absence of a coelom. In this regard, pseudocoelomic animals (nematode) and acoelomic animals (platyhelminthes) have been placed near the roots of bilateralian phylogenetic trees. Accumulating evidence raises the possibility that nematodes and a majority of the platyhelminthes are derived from a presumptive complex bilateralian ancestor that already had a coelom with a mesodermal lining. This bilateralian ancestor possessed the various genes required for a complex body plan, such as numerous transcription factors including homeobox genes and genes required for intercellular signaling (TGF- β , Hedgehog, Notch, and EGF) (reviewed in [13]). We should discuss the evolutionary processes of Zic genes and

their significance in animal evolution in the light of the above-mentioned background.

Evolutionary processes of Zic genes

From our current results, we can envisage the evolutionary process of Zic genes as follows (Fig. 5A). The earliest Zic genes that appeared in ancestral Eumetazoa or metazoans, probably in the pre-Cambrian era, may have shared common structural features, namely the ZF and ZF-NC domains. Exon–intron boundaries are not conserved between Eumetazoa and cnidarians. Therefore, we postulate that the earliest Zic genes may not have had any introns in their ZF regions. The identity of the direct ancestor of the earliest Zic gene is not certain, but the presence of Gli/Nkl/Zic superfamily-related genes in fungi suggests that the Zic genes have diverged from a common ancestor of the Gli/Nkl/Zic superfamily genes.

In a bilateralian ancestor, a single Zic gene would have acquired an A-intron in the ZF3 region. All of the bilateralian Zic genes are descendents of this bilateralian common ancestral Zic (urbilateralian Zic), as supported by the presence of an A-intron in all the bilateralian Zic genes that we examined. The urbilateralian Zic may have already possessed the ZOC domain, as is supported by the widespread occurrence of ZOC in bilateralian progenies. Therefore, the full set of conserved domains (ZOC, ZF-NC, and ZF) had probably already appeared in the urbilateralian Zic.

Additional variation may have been given to urbilateralian Zic later in the evolutionary process. Intron sequences were inserted into Zic genes in vertebrates, urochordates, arthropods, and nematodes; the positions of the introns diverged but seem to have been conserved in each phylum. The urochordate (ascidian) has a unique C-intron in both types of *Ci-Zic* genes, macho-1 and ZicL, suggesting that a gene duplication event occurred after insertion of the C-intron in the urochordate clade. In nematodes, five additional introns, three of which were located within the ZF domain, seem to have been inserted.

In the case of vertebrate Zic1–3 and insect Zic homologues, there is an apparently common exon–intron boundary (referred to as B-intron). The intron sequence is located in the carboxy-terminally located histidine residue in the five tandemly repeated C2H2 motifs. Two hypotheses could explain the occurrence of B-intron in the evolutionary process. One is the insertion of B-intron before the divergence of the Deuterostomia and Protostomia. The other assumes the independent occurrence of intron insertion in vertebrates and arthropods. We think that the latter is more likely because there are no B-intron-containing Zic

Fig. 1. (A) Domain structure of typical Zic protein. (B) Amino acid alignment of the ZF domain and ZF-NC domain. Letters with yellow background indicate completely conserved AA residues, and those with green indicate AA residues conserved more than 49% among the listed sequences. Bars with ZF-NC or numbers (1–5) indicate the extent of the ZF-NC domain and C2H2 ZF motifs 1–5, respectively. Asterisks indicate the cysteine and histidine residues that form C2H2 ZF motifs. A–E with arrowheads indicate the positions of the evolutionarily conserved exon–intron boundaries in the ZF. The phases of the exon–intron boundaries in the translation frames are A, 1; B, 0; C, 0; D, 2; E, 0. Abbreviations of the animal species are given in Table 1.

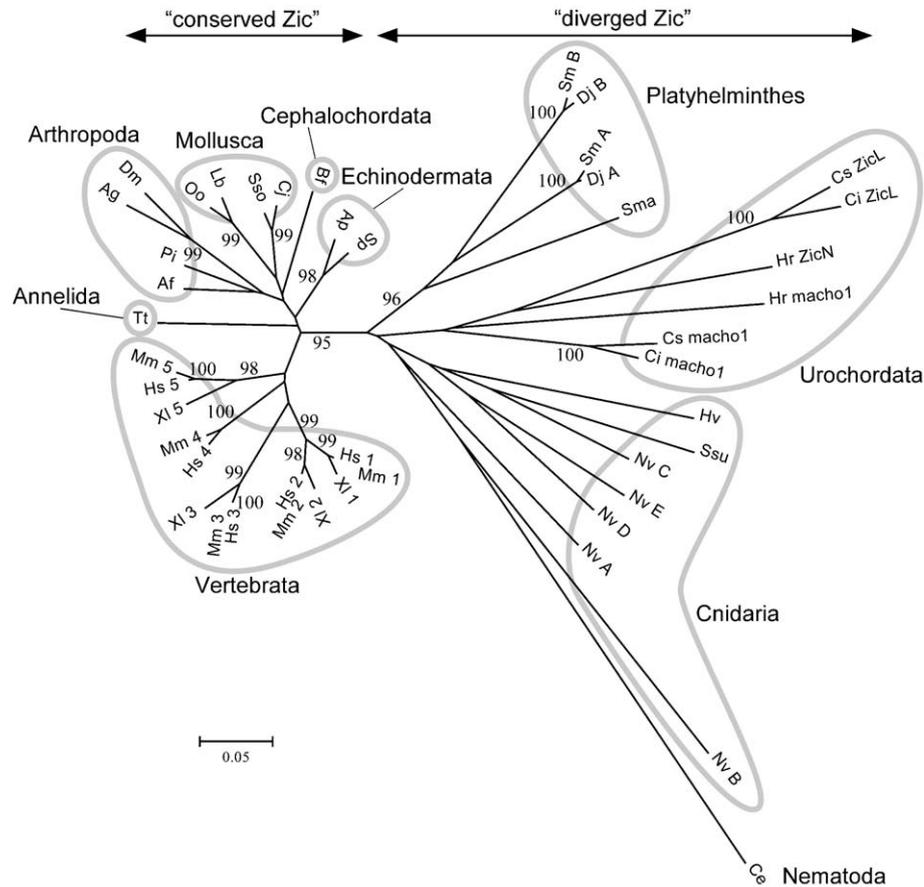


Fig. 3. Molecular phylogenetic tree using ZF and ZF-NC domains. Neighbor-joining tree was drawn by MEGA3.1 [32] after a multiple alignment by CLUSTALW [38]. Amino acid sequences containing ZF domains with 20 AA at the N-terminal and 10 AA at the C-terminal flanking sequences were used for the analysis. Animals in the same phyla or subphyla are encircled by the gray lines. The numbers near the internal branches indicate the statistical significance calculated by bootstrap test. Values of more than 94% are indicated in the tree.

genes in the other Protostomia and Deuterostomia species. In addition, it seems plausible that B-intron-containing vertebrate Zic1, Zic2, and Zic3 homologues have been generated after unique gene duplication events in this animal taxon (see below).

To explain the complex organization of vertebrate Zic genes, we propose the model outlined in Fig. 5B. The ancestral Zic in the ancestral chordates may have been similar to the urbilateralian Zic, which is represented by a cephalochordate (*Bf*) Zic gene. First, gene duplication occurred, resulting in head-to-head array of Zic genes with A-intron. Then, B-intron was inserted in one of the two Zic genes; this was followed by duplication twice (quadruplication) of the duplicated Zic, concomitant with the entire genome duplication events postulated in accordance with the DNA dose per haploid genome and the molecular phylogenetic analysis of many genes [14,15]. Of the resulting eight Zic genes, three (one AB-intron gene and two A-intron genes) may have disappeared in the course of evolution. The vertebrate Zic1, Zic2, and Zic3 have conserved sequences within this subfamily of Zic genes. In particular, the conservation of their carboxy-terminal sequence region, NFNEWYV, is clear. Partly similar sequences were found

in the midst of the region between the ZF domain and the C-terminal ends of vertebrate Zic4 and Zic5 and the cephalochordate Zic gene, suggesting that truncation of the open reading frame occurred in the prototype Zic genes after the initial duplication and before the quadruplication. It is possible that this domain was endowed with a new property that was functionally or structurally important for Zic1, Zic2, and Zic3 proteins.

The presence of multiple types of exon–intron boundaries in the highly conserved ZF domains is the most prominent feature of Zic genes. Both intron loss and intron gain occur in the course of evolution [16,17]. Although the exon–intron structure is highly dynamic [18,19], the introns are useful as phylogenetic characters as well as nucleotide and protein sequences [20–22]. In the case of Zic genes, the distribution of A-intron attracts our attention. A-intron was identified in all of the identified 27 bilateralian Zic genes compared in this study, but was not in the seven cnidarian Zic genes. In addition, the presence of A-intron can be predicted in the draft genome sequences of four insect species (*Aedes aegypti*, *Tribolium castaneum*, *Bombyx mori*, *Apis mellifera*) and in six vertebrate species (*Gallus gallus*, *Xenopus tropicalis*, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Danio rerio*,

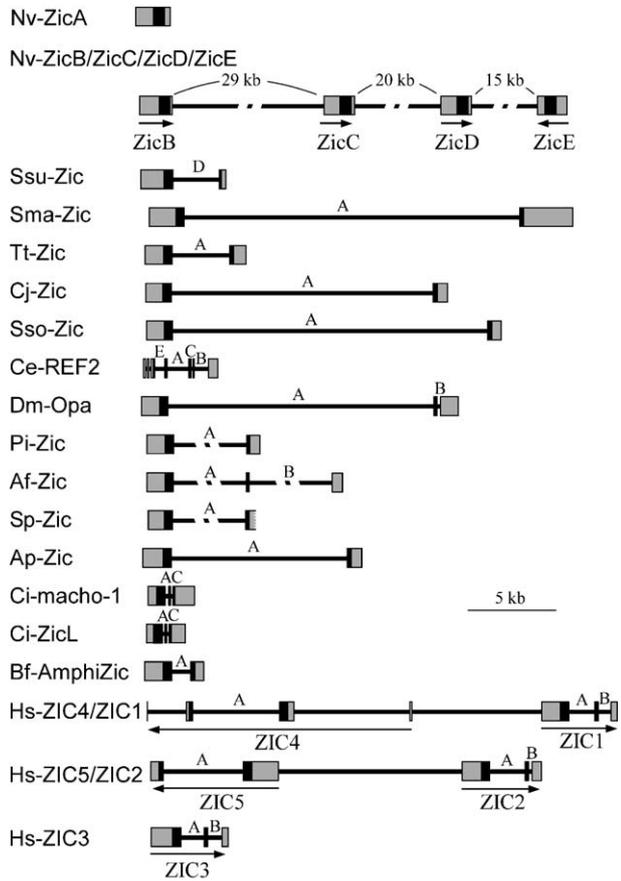


Fig. 4. Schematic drawing of exon–intron structures of Zic genes from various animals. Gray and black boxes indicate non-ZF and ZF parts of the Zic protein coding region, respectively. Horizontal bars indicate introns; capital letters (A–E) above the introns indicate the types of exon–intron borders. General transcription direction is left to right; for Nv_ZicB/ZicC/ZicD/ZicE and Hs ZIC1–5, the direction of the transcription is indicated by arrows. Each arrow also indicates the extent of one Zic protein-coding exon.

Oryzias latipes) (J.A., unpublished observation). The extent of exon–intron structure conservation seems to be higher than that of the genes for animal elongation factor-1 α [18,19] and animal translation initiation factor 2 [23], both of which have been well characterized as to the evolutionary conservation of the exon–intron structure. The location of A-intron in Zic genes may be noted to be highly conserved in bilateralian animals, and its distribution in the animal kingdom should be further investigated. These results also suggest that the extent of phylogenetic conservation of their locations considerably varies among introns. However, this speculation awaits further verification since the number of genes or animal species that have been subjected to the wide-range phylogenetic analysis of exon–intron structure is still limited.

Implications for body plan evolution

The distribution of the ZOC domain and the molecular phylogenetic tree of the Zic family suggest the presence of two classes of Zic genes (Figs. 2 and 3). One is composed

of the conserved class, comprising the Vertebrata, Cephalochordata, Echinodermata, Annelida, Mollusca, and Arthropoda; the other is the diverged class that includes the Cnidaria, Platyhelminthes, Nematoda, and Urochordata. Widespread occurrence of the conserved class Zic genes in the Ecdysozoa, Lophotrochozoa, and Deuterostomia suggests that the bilateralian ancestor gene belongs to the

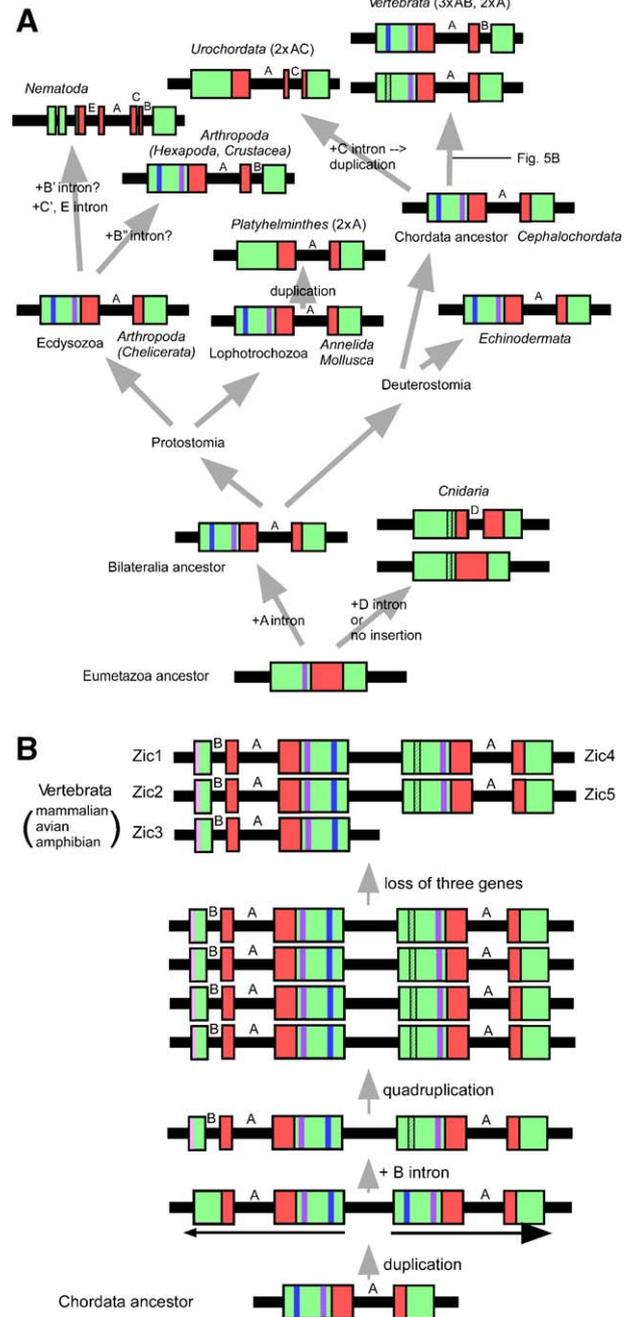


Fig. 5. Evolutionary processes of Zic genes. (A) Hypothetical model that explains the diversity of Zic genes in the Eumetazoa. (B) Establishment of vertebrate Zic1–5 genes. Boxes indicate protein-coding regions. Thick horizontal bars indicate intron and flanking region. Green indicates protein coding regions without clearly conserved domain structures; red indicates zinc finger domains; dark blue indicates ZOC; light blue indicates ZF-NC. Thin arrows in B represent the direction of transcription.

conserved class of Zic. These two classes may reflect the difference in diversification of the ancestor Zic genes. In general terms, the complexity of body structure seems to be well correlated with the extent of the Zic protein structure conservation. Cnidarians are the bilayered animals; Platyhelminthes are acoelomatic animals, and nematodes are pseudocoelomatic. The animals belonging to these three phyla may have a relatively simple body structure in comparison to the other animals examined in this study. The urochordates seem to be unique among the four members of the diverged class. However, their body structure is simpler than that of other chordates in that they do not have either a large CNS or segmented mesodermal structures, both of which are retained in vertebrates and cephalochordates. Therefore, the “complexity” of the body structures seems to be a major determinant of the structural conservation of Zic proteins.

There have been some indications that changes in protein structure between different animal species result in qualitative functional changes that can affect body plan. In both arthropod Hox proteins [24] and proneural proteins [25], structural changes determine their binding protein repertoires. The ZOC domain has been shown to be necessary for the transcriptional activation capacity of the mouse Zic2 protein [10]. I-mfa, an inhibitor of myogenic transcription factor protein, binds the region including ZOC and inhibits Zic2-mediated transcriptional activation by translocating Zic2 proteins from the nucleus to the cytoplasm [10]. It has been proposed that the extent of ZOC domain conservation is related to functional diversity among mouse Zic1–5 proteins [26]. In addition, our recent study showed that an evolutionarily conserved transcription factor could specifically bind the ZOC domain (T. Tohmonda, K. Mizugishi, K.M., J.A., unpublished observation), suggesting that ZOC has multi-functional properties. In contrast to ZOC, there are few known facts that can explain the significance of the ZF-NC domain. Its relevance to ZOC is intriguing, since both domains are often colocalized in the animal phylogenetic tree. It is possible that these two domains are closely located in a mature protein or bind common partner molecules. Structural and functional evaluation of the conserved domains from a phylogenetic point of view is needed to obtain a better understanding.

Materials and methods

Animals

Dugesia japonica was derived from a clonal planarian strain that was originally established by K. Watanabe (Himeji Technical Institute, Hyogo, Japan) [27]. The asexually reproductive jellyfish *Scolionema suvaense* [28], which was originally purchased from a local pet dealer, has been kept in artificial seawater for more than 1 year and fed brine shrimp. The freshwater oligochaete *Tubifex tubifex* was obtained as described [29]. *Loligo bleekeri* was obtained from Aburatsubo Marine Park Aquarium (Kanagawa, Japan). *Octopus ocellatus* was obtained from Ushimado Marine Laboratory, Okayama University. *Corbicula* sp. was purchased from a local fish dealer. A scorpion, *Pandinus imperator*, and a brine shrimp, *Artemia franciscana*, were purchased from local pet dealers. *Asterina pectinifera* was

collected from the seashore in Kanagawa Prefecture, Japan. To avoid contamination with genetic materials from other animal species, embryos in pure water or saline or artificial seawater were collected in the case of *T. tubifex*, *P. imperator*, *L. bleekeri*, *O. ocellatus*, and *As. pectinifera*. 18S ribosomal DNA sequences were determined for *S. suvaense* (Accession No. AB231884), *Ar. franciscana*, *P. imperator*, and *Corbicula* sp. for species confirmation. The *S. suvaense* sequence was most similar to those derived from animals in the phylum Cnidaria, class Hydrozoa, orders Limnomedusae and Trachymedusae, both of which orders are different from that of *Hydra vulgaris* (class Hydrozoa, order Athecata).

PCR cloning of Zic cDNA

The following primers were used for PCR amplification of the ZF domain of the Zic-related genes. The first PCR was carried out using ZF2F_1, 5'-TTYAARGCIAARTAYAARYT-3', and ZF3R, 5'-(T/G)RTGIATYTTIARRTYTC-3', and the second PCR was done using ZF2F_2, 5'-GCIAARTAYAARYTIRTIAYCA-3', and ZF3R. Each PCR consisted of 35 cycles of 94°C for 1 min, 35°C for 1 min, and 72°C for 2 min. The PCR was carried out with ExTaq DNA polymerase (Takara Bio, Shiga, Japan) in the presence of BD TaqStart anti-Taq antibody (BD Biosciences, CA, USA). This PCR was highly specific to clone Zic-related genes. RNAs were isolated by using TRIzol reagent (Invitrogen, CA, USA) in accordance with the manufacturer's recommendation. The homologues were initially identified by nested PCR on cDNA or genomic templates using degenerate primers corresponding to the most ZF2 to ZF3 region. cDNAs corresponding to ZF at their 3' and 5' ends were cloned by using a 3'-Full RACE Core Set and 5'-Full RACE Core Set (Takara Bio). The entire open reading frame region of the cDNAs was again cloned by the primers located outside the target regions. The amino acid sequences were deduced from nucleotide sequencing of multiple PCR fragments.

To determine whether a gene belonged to the Zic gene family in structural terms, we used the following criteria: (1) The ZF domain was composed of five repeats of C2H2 motifs. (2) The last four ZF motifs had an amino acid sequence that had more than 75% similarity to other Zic family proteins. (3) The ZF domain contained the ENLKIHF, FANSSDR, and YTHPSSLRKH sequence motifs, which are conserved in all of the Zic family proteins reported so far. (4) In the first ZF region, the number of amino acid residues between the conserved cysteine residues was greater than that in the other four ZF motifs in each of which four amino acid residues were present.

The newly identified Zic genes were designated as “Zic” preceded by the abbreviation for each animal species. When multiple genes existed in one species, ZicA, ZicB, ZicC, ZicD, and ZicE were used to discriminate each gene. This style did not mandate structural similarity among A, B, C, D, and E from different phyla, in contrast to the Zic1–5 designation in vertebrates. However, in the case of planarians, ZicA and ZicB are structurally related.

Fosmid and BAC cloning of genomic DNA

To isolate high-molecular-weight DNA, library screening was done as described [30]. Fosmid genomic libraries were prepared using CopyControl pCC1FOS vector (Epicentre, WI, USA). The genomic DNA was first sheared by performing an appropriate number of pipettings. After being blunted, the DNA fragments were size-fractionated by field-inversion gel electrophoresis through a Genofield electrophoretic apparatus (ATTO, Tokyo, Japan). The resultant 35- to 45-kb fraction was inserted into the fosmid vector. The BAC filter and clones were purchased from BACPAC Resources, Children's Hospital Oakland Research Institute (OH, USA). Some of the exon–intron boundary genomic sequences in *P. imperator*, *L. bleekeri*, and *As. pectinifera* were determined by using a DNA Walking SpeedUp Kit (Seegene, Seoul, Korea).

DNA sequencing and sequence analysis

Sequencing and data assembly were done as described [31]. Genomic sequences of *Schmidtea mediterranea*, *H. vulgaris*, *Caenorhabditis elegans*, *Strongylocentrotus purpuratus*, and *Ciona intestinalis* were derived from public databases (<http://www.ncbi.nlm.nih.gov/BLAST/tracemb.shtml>). Sequence

analysis was done using DNASISPro (Hitachi Software Engineering, Tokyo, Japan), Sequencher (Gene Codes, Ann Arbor, MI, USA), and Genetyx (Genetyx, Tokyo, Japan) software. Homology search was carried out against a public database (<http://www.ncbi.nlm.nih.gov/BLAST/>) using BLAST and discontinuous Mega BLAST. The molecular phylogenetic tree was drawn using MEGA version 3.1 [32] as follows: The amino acid sequences were first aligned by CLUSTALW [33] under the following parameter settings: Gonnet 250 as protein weight matrix; gap opening and gap extension penalties for pairwise alignments, 35 and 0.75, respectively; gap opening and gap extension penalties for multiple alignments, 15 and 0.3, respectively. The aligned sequences were subjected to neighbor-joining tree construction using Poisson correction distance with bootstrap test (1000 replicates). For the multiple alignment shown in Fig. 1B, CLUSTALW parameters were changed as follows so that cysteine and histidine residues in the C2H2 motifs were aligned correctly: Gonnet 250 as protein weight matrix; gap opening and gap extension penalties for pairwise alignments, 10 and 0.1, respectively; gap opening and gap extension penalties for multiple alignments, 10 and 0.2, respectively.

Exon–intron boundaries of Hs-ZIC2, Mm-Zic1-5, Dm-Opa, and Ce-Ref2 were reported [8,11,34,35,36]. The boundaries of Hs-ZIC1, ZIC3, ZIC4, and ZIC5 were identified in the human genome databases (<http://www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html>). The other exon–intron boundaries were identified by a homology search against the genome sequences using originally determined (Af-Zic, Ap-Zic, Cj-Zic, Pi-Zic, Ssu-Zic, Tt-Zic) or reported (Bf-AmphiZic, Ci-macho-1, Ci-ZicL) cDNA sequences. In the case of Nv-ZicA-E, Sma-Zic, Sp-Zic, and Sso-Zic, the homology search was done using cDNA sequences of Ssu-Zic, Dj-ZicA, Ap-Zic, and Cj-Zic, respectively. All of the exon–intron boundaries followed the GT/AG rule.

Acknowledgments

We thank Ulrich Technau for correspondence concerning Hyzic, Shin-ichi Nemoto for helpful comments on starfish sampling, all the technical staff of Sequencing Technology Team RIKEN GSC, Yayoi Nozaki, and Research Resource Center RIKEN BSI for their technical assistance. This study was done as a collaborative research project of the Strategic Research Program in RIKEN. This work was supported by grants-in-aid from the Japanese Ministry of Education, Science, Sports, and Technology.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ygeno.2006.02.011.

References

- [1] J. Aruga, The role of Zic genes in neural development, *Mol. Cell Neurosci.* 26 (2004) 205–221.
- [2] G.E. Herman, H.M. El-Hodiri, The role of ZIC3 in vertebrate development, *Cytogenet. Genome Res.* 99 (2002) 229–235.
- [3] S. Wada, H. Saiga, HrzicN, a new Zic family gene of ascidians, plays essential roles in the neural tube and notochord development, *Development* 129 (2002) 5597–5608.
- [4] H. Nishida, K. Sawada, macho-1 encodes a localized mRNA in ascidian eggs that specifies muscle fate during embryogenesis, *Nature* 409 (2001) 724–729.
- [5] K.S. Imai, Y. Satou, N. Satoh, Multiple functions of a Zic-like gene in the differentiation of notochord, central nervous system and muscle in *Ciona savignyi* embryos, *Development* 129 (2002) 2723–2732.
- [6] M.J. Benedyk, J.R. Mullen, S. DiNardo, odd-paired: a zinc finger pair-rule protein required for the timely activation of engrailed and wingless in *Drosophila* embryos, *Genes Dev.* 8 (1994) 105–117.
- [7] D.M. Cimbora, S. Sakonju, *Drosophila* midgut morphogenesis requires the function of the segmentation gene odd-paired, *Dev. Biol.* 169 (1995) 580–595.
- [8] S. Alper, C. Kenyon, The zinc finger protein REF-2 functions with the Hox genes to inhibit cell fusion in the ventral epidermis of *C. elegans*, *Development* 129 (2002) 3335–3348.
- [9] D. Lindgens, T.W. Holstein, U. Technau, Hyzic, the Hydra homolog of the zic/odd-paired gene, is involved in the early specification of the sensory nematocytes, *Development* 131 (2004) 191–201.
- [10] K. Mizugishi, et al., Myogenic repressor I-mfa interferes with function of Zic family proteins, *Biochem. Biophys. Res. Commun.* 320 (2004) 233–240.
- [11] J. Aruga, et al., The mouse zic gene family: homologues of the *Drosophila* pair-rule gene odd-paired, *J. Biol. Chem.* 271 (1996) 1043–1047.
- [12] A. Adoutte, G. Balavoine, N. Lartillot, O. Lespinet, B. Prud'homme, R. de Rosa, The new animal phylogeny: reliability and implications, *Proc. Natl. Acad. Sci. USA* 97 (2000) 4453–4456.
- [13] A.H. Knoll, S.B. Carroll, Early animal evolution: emerging views from comparative biology and geology, *Science* 284 (1999) 2129–2137.
- [14] S. Ohno, *Evolution by Gene Duplication*, Springer-Verlag, Berlin/Heidelberg/New York, 1970.
- [15] P.W. Holland, Gene duplication: past, present and future, *Semin. Cell Dev. Biol.* 10 (1999) 541–547.
- [16] J.M. Logsdon Jr., The recent origins of spliceosomal introns revisited, *Curr. Opin. Genet. Dev.* 8 (1998) 637–648.
- [17] S.W. Roy, W. Gilbert, Rates of intron loss and gain: implications for early eukaryotic evolution, *Proc. Natl. Acad. Sci. USA* 102 (2005) 5773–5778.
- [18] H. Wada, M. Kobayashi, R. Sato, N. Satoh, H. Miyasaka, Y. Shirayama, Dynamic insertion–deletion of introns in deuterostome EF-1alpha genes, *J. Mol. Evol.* 54 (2002) 118–128.
- [19] I.B. Rogozin, A.V. Sverdlov, V.N. Babenko, E.V. Koonin, Analysis of evolution of exon–intron structure of eukaryotic genes, *Brief Bioinform.* 6 (2005) 118–134.
- [20] A. Rokas, J. Kathirithamby, P.W. Holland, Intron insertion as a phylogenetic character: the engrailed homeobox of Strepsiptera does not indicate affinity with Diptera, *Insect Mol. Biol.* 8 (1999) 527–530.
- [21] B. Venkatesh, Y. Ning, S. Brenner, Late changes in spliceosomal introns define clades in vertebrate evolution, *Proc. Natl. Acad. Sci. USA* 96 (1999) 10267–10271.
- [22] S.W. Roy, W. Gilbert, Resolution of a deep animal divergence by the pattern of intron conservation, *Proc. Natl. Acad. Sci. USA* 102 (2005) 4403–4408.
- [23] V. Krauss, M. Pecyna, K. Kurz, H. Sass, Phylogenetic mapping of intron positions: a case study of translation initiation factor eIF2gamma, *Mol. Biol. Evol.* 22 (2005) 74–84.
- [24] U. Lohr, L. Pick, Cofactor-interaction motifs and the cooption of a homeotic Hox protein into the segmentation pathway of *Drosophila melanogaster*, *Curr. Biol.* 15 (2005) 643–649.
- [25] X.J. Quan, et al., Evolution of neural precursor selection: functional divergence of proneural proteins, *Development* 131 (2004) 1679–1689.
- [26] A. Ishiguro, T. Inoue, K. Mikoshiba, J. Aruga, Molecular properties of Zic4 and Zic5 proteins: functional diversity within Zic family, *Biochem. Biophys. Res. Commun.* 324 (2004) 302–307.
- [27] Y. Umesonono, K. Watanabe, K. Agata, A planarian orthopedia homolog is specifically expressed in the branch region of both the mature and regenerating brain, *Dev. Growth Differ.* 39 (1997) 723–727.
- [28] T. Uchida, Y. Sugiura, On the medusa-budding found in limnomedusa, *Scolionema suvaense*, *J. Fac. Sci. Hokkaido Univ. Ser. VI, Zool.* 20 (1976) 600–604.
- [29] T. Shimizu, Development in the freshwater oligochaete Tubifex, in: F. Harrison, R.R. Cowden (Eds.), *Development Biology of Freshwater Invertebrates*, A. R. Liss, New York, 1982, pp. 286–316.
- [30] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1989.
- [31] A. Toyoda, et al., Comparative genomic sequence analysis of the human chromosome 21 Down syndrome critical region, *Genome Res.* 12 (2002) 1323–1332.
- [32] S. Kumar, K. Tamura, M. Nei, MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment, *Brief Bioinform.* 5 (2004) 150–163.

- [33] R. Chenna, et al., Multiple sequence alignment with the Clustal series of programs, *Nucleic Acids Res.* 31 (2003) 3497–3500.
- [34] S.A. Brown, et al., Holoprosencephaly due to mutations in ZIC2, a homologue of *Drosophila* odd-paired, *Nat. Genet.* 20 (1998) 180–183.
- [35] J. Aruga, A. Yozu, Y. Hayashizaki, Y. Okazaki, V.M. Chapman, K. Mikoshiba, Identification and characterization of Zic4, a new member of the mouse Zic gene family, *Gene* 172 (1996) 291–294.
- [36] T. Inoue, M. Hatayama, T. Tohmonda, S. Itoharu, J. Aruga, K. Mikoshiba, Mouse Zic5 deficiency results in neural tube defects and hypoplasia of cephalic neural crest derivatives, *Dev. Biol.* 270 (2004) 146–162.
- [37] K. Tessmar-Raible, D. Arendt, Emerging systems: between vertebrates and arthropods, the Lophotrochozoa, *Curr. Opin. Genet. Dev.* 13 (2003) 331–340.
- [38] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.