Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Genomics 87 (2006) 783–792
GENOMICS
www.elsevier.com/locate/ygeno

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 bilateralians but not cnidarians, suggesting that all of the bilateralian Zic genes are derived from a single gene in a bilateralian 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 bilateralian Zic protein had already acquired an entire set of conserved domains, but that this was lost and diverged in the platyhelminthes, nematodes, and urochordates.

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/DDBJ Data Libraries under Accession Nos. AB231864–AB231884.

* Corresponding author. Fax: +81 48 467 9792.
E-mail address: jaruga@brain.riken.jp (J. Aruga).

0888-7543/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.02.011

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	*Scolionema suvaense*	*Ssu*	D	Fosmid	Sco26, Sco36 (AB231882)	AB231883
	Hydra vulgaris	*Hv*	0	Database		[9]
	Nematostella vectensis	*Nv*	0	BAC	CH314-49A19 (AB231867) CH314-55K22 (AB231868)	
Platyhelminthes	*Dugesia japonica*	*Dj*				AB231880, AB231881
	Schmidtea mediterranea	*Sme*		Database		
	Schistosoma mansoni	*Sma*	A	BAC	CH103-42N16 (AB231864)	
Annelida	*Tubifex tubifex*	*Tt*	A	Fosmid	Tub8, Tub12 (AB231869)	AB231870
Mollusca	*Loligo bleekeri*	*Lb*				AB231874
	Octopus ocellatus	*Oo*				AB231875
	Corbicula sp.	*Cj*	A	Fosmid	Cor8, Cor10 (AB231873)	
	Spisula solidissima	*Sso*	A	BAC	CH312-9E16 (AB231865)	
Nematoda	*Caenorhabditis elegans*	*Ce*	ACBE	Database		[8]
Arthropoda	*Drosophila melanogaster*	*Dm*	AB	Database		[6,7]
	Anopheles gambiae	*Ag*		Database		XM_321856
	Pandinus imperator	*Pi*	A	PCR	(AB231876)	AB231877
	Artemia franciscana	*Af*	AB	Fosmid, PCR	Art1 (AB231878)	AB231879
Echinodermata	*Strongylocentrotus purpuratus*	*Sp*	A	Database		
	Asterina pectinifera	*Ap*	A	Fosmid	Ast1 (AB231871)	AB231872
Chordata	*Ciona intestinalis*	*Ci*	AC	Database		[31]
	Ciona savignyi	*Cs*		Database		[5]
	Halocynthia roretzi	*Hr*		Database		[3,4]
	Branchiostoma floridae	*Bf*	A	BAC	CH302-61N11 (AB231866)	[19]
	Xenopus sp.	*Xl*	A, AB	Database		Ref. in [1]
	Mus musculus	*Mm*	A, AB	Database		Ref. in [1]
	Homo sapiens	*Hs*	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.