

Research article

Open Access

Molecular evolution of Cide family proteins: Novel domain formation in early vertebrates and the subsequent divergence

Congyang Wu^{†1}, Yinxin Zhang^{†1}, Zhirong Sun² and Peng Li^{*1}

Address: ¹Protein science laboratory of Ministry of Education, Department of Biological Sciences and Biotechnology, Tsinghua University, Beijing 100084, China and ²MOE Key Laboratory of Bioinformatics, State Key Laboratory of Biomembrane and Membrane Biotechnology, Department of Biological Sciences and Biotechnology, Tsinghua University, Beijing 100084, China

Email: Congyang Wu - cy-wu04@mails.tsinghua.edu.cn; Yinxin Zhang - zhangyinxin04@mails.tsinghua.edu.cn; Zhirong Sun - sunzhr@mail.tsinghua.edu.cn; Peng Li* - li-peng@mail.tsinghua.edu.cn

* Corresponding author †Equal contributors

Published: 23 May 2008

Received: 27 October 2007

BMC Evolutionary Biology 2008, 8:159 doi:10.1186/1471-2148-8-159

Accepted: 23 May 2008

This article is available from: <http://www.biomedcentral.com/1471-2148/8/159>

© 2008 Wu et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Cide family proteins including Cidea, Cideb and Cidec/Fsp27, contain an N-terminal CIDE-N domain that shares sequence similarity to the N-terminal CAD domain (NCD) of DNA fragmentation factors Dffa/Dff45/ICAD and Dffb/Dff40/CAD, and a unique C-terminal CIDE-C domain. We have previously shown that Cide proteins are newly emerged regulators closely associated with the development of metabolic diseases such as obesity, diabetes and liver steatosis. They modulate many metabolic processes such as lipolysis, thermogenesis and TAG storage in brown adipose tissue (BAT) and white adipose tissue (WAT), as well as fatty acid oxidation and lipogenesis in the liver.

Results: To understand the evolutionary process of Cide proteins and provide insight into the role of Cide proteins as potential metabolic regulators in various species, we searched various databases and performed comparative genomic analysis to study the sequence conservation, genomic structure, and phylogenetic tree of the CIDE-N and CIDE-C domains of Cide proteins. As a result, we identified signature sequences for the N-terminal region of Dffa, Dffb and Cide proteins and CIDE-C domain of Cide proteins, and observed that sequences homologous to CIDE-N domain displays a wide phylogenetic distribution in species ranging from lower organisms such as hydra (*Hydra vulgaris*) and sea anemone (*Nematostella vectensis*) to mammals, whereas the CIDE-C domain exists only in vertebrates. Further analysis of their genomic structures showed that although evolution of the ancestral CIDE-N domain had undergone different intron insertions to various positions in the domain among invertebrates, the genomic structure of Cide family in vertebrates is stable with conserved intron phase.

Conclusion: Based on our analysis, we speculate that in early vertebrates CIDE-N domain was evolved from the duplication of NCD of Dffa. The CIDE-N domain somehow acquired the CIDE-C domain that was formed around the same time, subsequently generating the Cide protein. Subsequent duplication and evolution have led to the formation of different Cide family proteins that play unique roles in the control of metabolic pathways in different tissues.

Background

Cide family proteins including *Cidea*, *Cideb* and *Cidec*/Fsp27 [1-3] were originally identified by their sequence homology to the N-terminal CAD domain (NCD) [4] of DNA fragmentation factors Dffa and Dffb [5-16]. Whereas NCD specifically refers to the N-terminal domain of Dff factors, CIDE-N denotes the N-terminal sequence shared by Cide proteins in this article. In addition to the CIDE-N domain, Cide proteins also contain a unique conserved C-terminal domain (CIDE-C domain). Despite some variation between NCD and CIDE-N, they all contain a potential yin and yang surface that could mediate weak protein-protein interaction. Recently, they were also found to be structurally homologous to the ubiquitin (UB) α/β roll superfold [17,18], but bear no high similarity to other existing proteins [17,19].

While *Cidea* is expressed at high levels in BAT [20], *Cideb* is more abundantly expressed in the liver, with moderate levels in kidney, small intestine and colon [21]. When over-expressed in heterologous cells such as 293T and COS-7 cells, *Cideb* can form homo- or hetero-dimers with other CIDE family members and induce caspase-independent cell death [22]. Furthermore, CIDE-C domain of *Cideb*, is responsible for *Cideb*-induced cell death and dimerization [22]. *Cidea* protein has also been found to regulate apoptosis induced by TGF- β [23]. However, how Cide proteins induce apoptosis remains unclear. No caspase cleavage site or nuclease specific domain present in Dff factors was identified in Cide proteins. To study the physiological role of Cide proteins, we previously generated *Cidea* null mice, and found that *Cidea*-null mice are lean and resistant to diet-induced obesity and diabetes [20]. *Cidea* controls energy homeostasis in BAT by regulating lipolysis and thermogenesis. A recent study showed that *Cidea* was implicated in human obesity by regulating human adipocyte lipolysis [24] and a V115F polymorphism in human was found to be associated with obesity in certain populations [25]. *Cidea* was the most highly up-regulated gene in the liver of high calorie diet (HC)-fed mice and second most down-regulated gene in the liver of HC plus resveratrol (HCR) aging-improved mice [26]. Similar to *Cidea*, *Cideb* also plays important roles in metabolism. We recently reported that *Cideb* regulates diet-induced obesity, liver steatosis, and insulin sensitivity by controlling lipogenesis and fatty acid oxidation in the liver [21]. In addition, Fsp 27/*Cidec* was found to be associated with lipid droplets and promote triglyceride storage in differentiated 3T3-L1 cells [27,28]. All these studies suggest that Cide family proteins play important roles in modulating energy homeostasis, aging and the development of metabolic diseases such as obesity and diabetes [29-31].

While it is evident that Cide proteins regulate energy homeostasis in mammals, it is unclear about the origin and evolution of Cide family proteins. To provide further insights into the structure and function of Cide proteins, we have employed various databases and bioinformatic tools to study how Cide family proteins have been evolved. A recent analysis of the evolutionary process of Dff family proteins has identified orthologs of Dffa/b in lower organisms such as sea anemone, suggesting that the DNA fragmentation pathway in apoptosis is conserved throughout evolution [32]. Here we defined signature sequences for the N-terminal region of Dffa, Dffb and Cide proteins and CIDE-C domain of Cide proteins and analyzed the evolutionary history of CIDE-N and CIDE-C domains of Cide family proteins. No ortholog of CIDE proteins was identified in invertebrates or other lower organisms. However, a homologous sequence of CIDE-N domain of Cide proteins was identified in hydra, in addition to sea anemone as previously reported [32]. We found that the signature sequences for CIDE-N domain of Cide proteins are similar to those of NCD of Dffa in hydra and sea anemone. More importantly, we found that CIDE-C domain exists only in vertebrates with occasional possible omission from certain ancient fish species. By analyzing the genic structures and intron phases of Cide and Dff family, we found that although the evolution of the ancestral CIDE-N domain includes different intron insertions, the genomic structure of *Cide* family in vertebrates is stable, including 5 conserved exons separated by 4 introns with the sequential phases 2-0-0-2. Based on our observation, we postulate that the origin of Cide proteins may be the result of recombination of sequences encoding CIDE-N and CIDE-C domains in early vertebrates, and subsequent duplication and evolution have led to the formation of different Cide family proteins.

Results

Sequence comparison and species/tissue distribution of CIDE-N and CIDE-C domains

Through sequence alignment of the N-terminal region of Cide and Dff family proteins in human and mouse (Fig 1a), we observed highly conserved 37 amino acid residues around the EDGT protein signature site in the CIDE-N domain and NCD. NMR structural analysis in human *Cideb* suggests that the EDGT signature is located on an important loop of the CIDE-N domain interaction interface zone 1. Within this conserved domain, we observed penta amino acid residues RPXRV unique for CIDE-N domain of Cide family proteins, a VDDXXYF signature for Dffa and a LPXXGSR signature for Dffb (Fig 1a). These specific sequences will be used to distinguish Cide proteins from Dff in our following study. In addition, through the sequence alignment of CIDE-C domain of Cide family proteins in human and mouse (Fig 1b), we identified a highly conserved XARXTFDXYXXN-

Table 1: Hmmer search in Nr.db from NCBI using the CIDE-N motif pfam02017

Organism/Protein Found	Cidea	Cideb	Cidec	Dffa Related	Dffb Related	Drep1* Related	Drep2* Related	Drep3* Related	Drep4* Related
<i>Homo sapiens</i> (human)									
<i>Pan troglodytes</i>									
<i>Pongo pygmaeus</i>									
<i>Macaca mulatta</i>									
<i>Mus musculus</i> (mouse)									
<i>Rattus norvegicus</i>									
<i>Equus caballus</i>									
<i>Bos Taurus</i>									
<i>Sus scrofa</i>									
<i>Canis lupus familiaris</i>									
<i>Monodelphis domestica</i>									
<i>Ornithorhynchus anatinus</i>									
<i>Gallus gallus</i> (chicken)									
<i>Xenopus laevis</i>									
<i>Xenopus tropicalis</i> (<i>X.tropicalis</i>)									
<i>Danio rerio</i> (zebrafish)									
<i>Tetraodon nigroviridis</i> (tetraodon)									
<i>Aedes aegypti</i>									
<i>Anopheles gambiae</i> str. PEST									
<i>Culex pipiens quinquefasciatus</i>									
<i>Drosophila melanogaster</i> (fruit fly)									
<i>Drosophila pseudoobscura</i>									
<i>Bombyx mori</i>									
<i>Apis mellifera</i>									
<i>Tribolium castaneum</i>									

Total Cide Protein Found 41

Total Dff Related Protein Found 48

* Drep1 is Dffa in fruit fly, and Drep4 is Dffb in fruit fly.

identified any protein or open reading frame that contains only the CIDE-C domain and does not contain the CIDE-N domain. The Nr.db and EST search also suggests that CIDE-C domain has appeared in a later stage of evolution and it may have a specific function relating to vertebrates.

To further investigate the origin and function of different Cide proteins, we checked the tissue distribution of ESTs encoding Cide proteins in vertebrates (Table 2). In the mouse, *Cidea* is predominantly expressed in BAT, with small amounts of mRNA detected in heart, brain, skeletal muscle, lymph node, thymus, appendix and bone marrow [2,20]. The expression of *Cidec/Fsp27* is more widespread, with high levels in WAT and moderate levels in BAT and

skeletal muscle[3,20]. *Cideb* is more abundantly expressed in the liver, with moderate levels in kidney, small intestine and colon [21]. Results from the analysis of the distribution of Cide ESTs are in good agreement with the above observations, revealing that a large number of EST for *Cideb* found in the liver, and of *Cidec* in WAT. We also found Cide proteins are expressed at varying levels in many different tissues in the lower vertebrates. *Cidea* is expressed in the eye of zebrafish (*Danio rerio*), testis of Atlantic salmon (*Salmo salar*), ovary and brain of *X. tropicalis* and chicken (*Gallus gallus*), caecal tonsil, intestinal lymphocyte and liver of chicken; *Cideb* is expressed in the liver of zebrafish, medaka (*Oryzias latipes*) and *X. tropicalis*, gut/intestine of zebrafish, Atlantic salmon and *X. tropica-*

Table 2: EST distribution chart for organisms found with CIDE related domains

Organisms Found with Cide Related Domains	Species	Search Criterior		Tissue Distribution#		
		CIDE-N 37aa	CIDE-C 35aa	Cidea	Cideb	Cidec/Fsp27
<i>Homo sapiens</i> (human)	Mammals	281	126	BAT+, WAT, skin, heart, sympathetic trunk	liver, small intestine+, kidney, colon, lung, brain, thalamus, cervix, thymus, T-lymphocytes, lymph, placenta, stomach, prostate, testis, eye	WAT, white matter, brain, lung, skin, eye, breast, ovary, uterine, pancreas, colon, testis, nasopharynx
<i>Macaca fascicularis</i>		3	1			
<i>Macaca nemestrina</i>		1				
<i>Macaca mulatta</i>		1	1			
<i>Pongo pygmaeus</i>		3				
<i>Papio anubis</i>		2				
<i>Spermophilus lateralis</i>			1			
<i>Mus musculus</i> (mouse)		185	130	BAT+, mammary gland, salivary gland, aorta, thyroids, eye, head, lung, kidney	liver, kidney, colon, small intestine+, bowel, placenta	WAT+, mammary gland, salivary gland, thyroid, colon, lung, kidney, placenta
<i>Rattus norvegicus</i>		13	15			
<i>Rattus sp.</i>		1	1			
<i>Sus scrofa</i>		89	74			
<i>Bos taurus</i>		98	90			
<i>Ovis aries</i>		13	5			
<i>Capra hircus</i>		1				
<i>Canis lupus familiaris</i>		9	10			
<i>Macropus eugenii</i>		2	2			
<i>Trichosurus vulpecula</i>		11	9			
<i>Gallus gallus</i> (chicken)	Birds	40	36	ovary, brain, caecal tonsil, intestinal lymphocyte, liver		liver, small intestine, fat body, intestinal lymphocyte, hearts
<i>Taeniopygia guttata</i>		4	4			
<i>Meleagris gallopavo</i>		1				
<i>Xenopus tropicalis</i> (<i>X.tropicalis</i>)	Amphibians	149	146	ovary, brain	liver, gut/intestine, fat body, oviduct	Small intestine, lung brain
<i>Xenopus laevis</i>		88	77			
<i>Ambystoma mexicanum</i>		1	1			
<i>Danio rerio</i> (zebrafish)	Bone Fishes	56	16	eyes	liver, gut/intestine	liver
<i>Dicentrarchus labrax</i>		1	1			
<i>Cyprinus carpio</i>		1	1			
<i>Fundulus heteroclitus</i>		4	3			
<i>Gasterosteus aculeatus</i>		13	3			
<i>Gadus morhua</i>		4	1			
<i>Misgurnus anguillicaudatus</i>		1				
<i>Ictalurus furcatus</i>		1	1			
<i>Oncorhynchus mykiss</i>		30	30			
<i>Oncorhynchus nerka</i>		1	1			
<i>Oryzias latipes</i> (medaka)		49	44		liver	ovary
<i>Pimephales promelas</i>		3				
<i>Platichthys flesus</i>		1	1			
<i>Plecoglossus altivelis altivelis</i>		1	1			

Table 2: EST distribution chart for organisms found with CIDE related domains (Continued)

<i>Salmo salar</i> (Atlantic salmon)		12	15	testis	gut/intestine, head kidney, thymus, pyloric caecum	thyroid, thymus, spleen, pyloric caecum
<i>Sparus aurata</i>		2	2			
<i>Takifugu rubripes</i> (fugu)		2	1			
<i>Thunnus thynnus</i>		2	1			
<i>Zosterisessor ophiocephalus</i>		1	1			
<i>Leucoraja erinacea</i> (little skate)	Sharks & Rays	8	2			Liver
<i>Squalus acanthias</i> (spiny dogfish)		3	3			
<i>Branchiostoma floridae</i> (amphioxus)	Lancelets	3				
<i>Drosophila melanogaster</i> (fruit fly)	Insects	8				
<i>Drosophila pseudoobscura</i>		2				
<i>Drosophila ananassae</i>		1				
<i>Drosophila erecta</i>		1				
<i>Aedes aegypti</i>		3				
<i>Anopheles gambiae</i>		1				
<i>Lutzomyia longipalpis</i>		1				
<i>Phlebotomus papatasi</i>		1				
<i>Bombyx mori</i>		3				
<i>Apis mellifera</i>		1				
<i>Acyrtosiphon pisum</i>		3				
<i>Aphis gossypii</i>		1				
<i>Heliconius erato/himera</i> mixed		1				
<i>Heliconius melpomene</i>		1				
<i>Tribolium castaneum</i>		3				
<i>Diabrotica virgifera virgifera</i>		1				
<i>Rhipicephalus appendiculatus</i>	Arachnids	1				
<i>Ixodes scapularis</i>		2				
<i>Daphnia pulex</i>	Crustaceans	1				
<i>Homarus americanus</i>		1				
<i>Lottia gigantea</i>	Molluscs	4				
<i>Aplysia californica</i>		2				
<i>Crassostrea gigas</i>		2				
<i>Mytilus californianus</i>		2				
<i>Capitella</i> sp. I ECS-2004	Annelids	3				
<i>Nematostella vectensis</i> (sea anemone)	Cnidarians	5				
<i>Hydra vulgaris</i> (hydra)		1				
Total EST Hits		1251	857			

The order of tissues is based on the copy number of cDNAs encoding corresponding protein. The cDNAs from mixed tissues are ignored in counting.

+ The high expression of Cide proteins in these tissues is verified by experiments, but not observed in the current EST database.

lis, in the thymus, head kidney, and pyloric caecum of Atlantic salmon, in fat body and oviduct of *X. tropicalis*; while *Cidec/Fsp27* is expressed in the liver of little skate (*Leucoraja erinacea*), zebrafish and chicken, in the small intestine of *X. tropicalis* and chicken, in the ovary of medaka, in thyroid, thymus, spleen and pyloric caecum of Atlantic salmon, in the brain and lung of *X. tropicalis*, as well as in the fat body, intestinal lymphocyte and hearts of chicken.

Identification of the ancestral CIDE-N domain in hydra

Hydra and sea anemone belong to the phylum cnidaria which is one of the earliest animal phyla [33,34]. We found 1 cDNA (GenBank: [DY447116](#)) in hydra and 5 cDNAs (GenBank: [DV089654](#), GenBank: [DV085979](#), GenBank: [FC181163](#), GenBank: [FC273871](#) and GenBank: [FC274613](#)) in sea anemone which encode proteins homologous to the most conserved 37 amino acids of CIDE-N domain from mouse *Cideb*. Further analysis revealed that the cDNA in hydra encode *Dffa*, while the other 5 cDNAs in sea anemone all encode *Dffa*, as observed by Eckhart et al [32].

Sequence alignment of N-terminal region for *Cide* and *Dff* family proteins in hydra, sea anemone and human (Fig 2) showed a remarkable similarity between *Cide* proteins and hydra *Dffa*. Using pair-wise comparison, we found the NCD of hydra *Dffa* shares approximately 42.3 percent sequence similarity to the NCD of human *Dffa*, and 42.9 percent sequence similarity to the CIDE-N domain of human *Cideb*. *Dffa* in hydra and sea anemone has signatures of both *Cide* family proteins and *Dffa* (RPXRV and VDDXXYF), but *Dffb* in sea anemone only contain signature sequences for *Dffb* (LPXXGSR). These data, together with the above sequence comparison and species distribution data, suggest that CIDE-N domain of *Cide* proteins is derived from the NCD of *Dffa*, but not *Dffb*, in lower organism like hydra and sea anemone.

Thus we define herein the NCDs of *Dffa* in hydra and sea anemone as the ancestral CIDE-N domain.

Comparative genomic analysis of genic structures and intron phases of Cide and Dff gene family

By searching the genomic data base of various species, we observed that the gene structure of all *Cide* family proteins in vertebrates consists of 5 exons and 4 introns with the sequential phases 2-0-0-2, while vertebrate *Dffa* gene consists of 6 exons and 5 introns with the sequential phases 1-1-0-1-0. Vertebrate *Dffb* consists of 7 exons and 6 introns with the sequential phases 0-1-1-0-0-2. The length of exons of *Cide* gene family is also conserved in vertebrates. By matching their exons to the corresponding protein sequences of *Cide* and *Dff* family proteins, we found that CIDE-N domain is encoded by exon 2 and exon 3 of *Cide* genes whereas the conserved CIDE-C domain is encoded by exon 4 and exon 5 of *Cide* genes. NCDs of *Dffa* and *Dffb* are encoded mainly by its exon 1 and exon 2, respectively (Fig 3a,d).

Unlike the conserved genic structures and intron phases of *Dff* gene family in 10 vertebrates, the *Dff* gene family in 3 representative invertebrates has different genic structures and intron phases. The ancestral CIDE-N domain of *Dffa* in sea anemone was split by one phase 0 intron in fruit fly (*Drosophila melanogaster*), but by phase 1 intron in a different position in amphioxus (*Branchiostoma floridae*) (Fig 3b,c). Based on the genomic structure and analysis from non-redundant proteins databases and EST database, we conclude that the *Cide* gene family exists only in vertebrates, while the *dff* gene family exists in both vertebrates and invertebrates.

The absence of some Cide family proteins in several vertebrate species

From the result of tblastn search in available EST data bases, we found the most ancient CIDE-N domain exists in hydra and sea anemone, and the most ancient CIDE-C

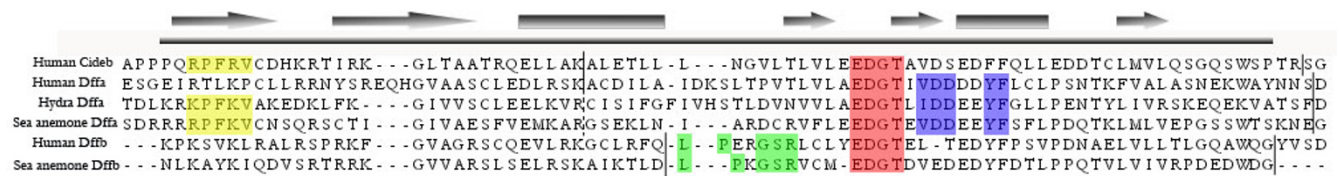


Figure 2
Sequence alignment of N-terminals for Cide and Dff family proteins in hydra, sea anemone, and human. Shown here are six N-terminals of *Cide* and *Dff* proteins, with the sequence from human *Cideb* representing the CIDE-N domain and the rest representing the NCD domains. The most conserved EDGT signature is highlighted in red. The RPXRV signature of *Cide* family proteins is highlighted in yellow, the VDDXXYF signature of *Dffa* in purple, and the LPXXGSR signature of *Dffb* in green. The exon boundaries are marked by black vertical lines. The potential exon boundaries are marked by black dotted lines. The secondary structure of the human *Cideb*'s CIDE-N domain is presented on the top of the alignment [17]; cylinders and arrows represent α helices and β strands, respectively.

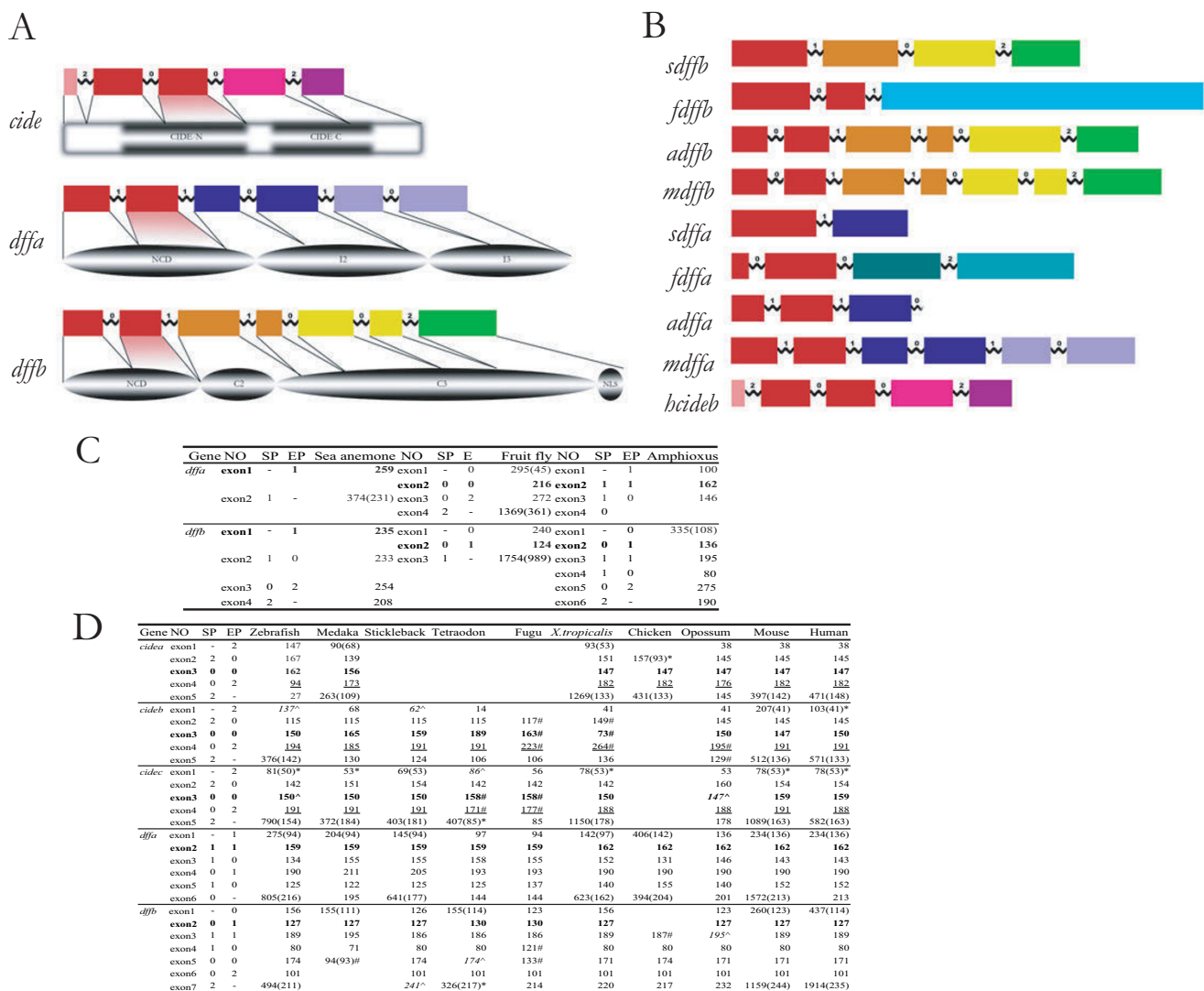


Figure 3
Genic structures and intron phases of Cide and Dff gene family. (A) The gene structures of human *Cideb*, mouse *Dffa* and mouse *Dffb* are shown above their corresponding proteins. Exons and introns are shown in boxes and wavy lines, respectively [4, 17]. Exons are drawn to scale. Black lines link a particular exon to its translated region of the protein. The matching areas flanked by black lines between the most conserved proteins sequences and their corresponding exons are marked in red. (B) Genic structures and intron phases of *Cide* and *Dff* gene family from several representative model organisms are drawn to scale. The regions from the various species in the same color share sequence homology. (C, D) The genic and intron phases of *Cide* and *Dff* gene family are divergent in invertebrates but conserved in vertebrates. Exon sizes in different organisms are shown in evolutionary order. The bordering intron phases of the exons are shown in the left. The number in '()' of exon 1 of each gene indicates its nucleotide position downstream of ATG (inclusive of the three nucleotides). The number in '()' of each last exon indicates its nucleotide position upstream of its stop codon. '#' indicate before the first exon (or after the last exon) noncoding exon regions. '^' refers to exons with deviational start phase and the end phase from the others in the same exon group. '^' indicates discard of some flanking exons. Italicized numbers are adjusted exon sizes. Exons in bold encode the most conserved protein regions. Abbreviations for species: s, sea anemone (*Nematostella vectensis*); f, fruit fly (*Drosophila melanogaster*); a, amphioxus (*Branchiostoma floridae*); m, mouse (*Mus musculus*); h, human (*Homo sapiens*). Other model organisms are zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), tetraodon (*Tetraodon nigroviridis*), fugu (*Takifugu rubripes*), *X.tropicalis* (*Xenopus tropicalis*), chicken (*Gallus gallus*), opossum (*Monodelphis domestica*). SP, start phase, EP, end phase.

domain exists in spiny dogfish (*Squalus acanthias*) and little skate (*Leucoraja erinacea*). No sequence homologous to CIDE-N domain was identified before phylum cnidaria such as yeast (*S. cerevisiae*). In addition, no sequence homologous to CIDE-C was found before the phylum vertebrata.

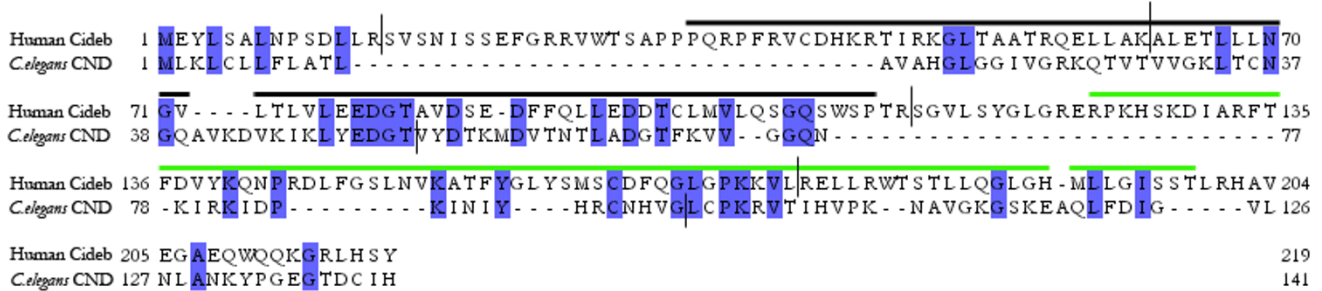
Interestingly, although the whole genome of nematodes (*C. elegans*) was sequenced and analyzed extensively, no genomic sequence encoding proteins that share sequence similarity to Dffa/b or Cide proteins were identified. We only found 1 protein (GenBank: [Y51A2D.10](#)) in *C. elegans* with limited homology to the conserved 37 amino acids including the signature EDGT motif of CIDE-N domain from human Cideb (Fig 4), with no homology to any other region. Furthermore, the exon boundaries between this protein and human Cideb are not conserved. Therefore, this protein is unlikely to be the ortholog of Dff or Cide in *C. elegans*.

In the *Petromyzon marinus*-3.0 Contigs database (sea lamprey (*Petromyzon marinus*) genome data base last modified on Apr 16, 2007) we found an ortholog of Dffb, but not Dffa or Cide proteins. As sea lamprey is regarded as the most primitive vertebrate, it would be interesting to determine whether any ortholog of Cide proteins exists in this organism. In the large EST database for little skate and spiny dogfish, we failed to identify any Cidea-like protein. Thus it is highly possible that ortholog of Cidea may not exist in little skate and spiny dogfish. Furthermore, after searching the whole genome sequences of fugu (*Takifugu rubripes*) and tetraodon (*Tetraodon nigroviridis*) [35,36], no

ortholog for Cidea protein was identified in these species. No Cidea-like protein was identified in stickleback (*Gasterosteus aculeatus*) either. In addition, no ortholog of Cideb was identified in the currently available chicken (*Gallus gallus*) genome data base [37] or EST databases.

The phylogenetic tree of CIDE-N and CIDE-C domains

In this study, three methods including neighbor-joining (NJ), maximum likelihood (ML), and unweighted pair group method with arithmetic mean (UPGMA) were used to construct the phylogenetic trees. These three methods often gave the same trees, except for some minor details. From the phylogeny of 31 selected Cide and Dff family proteins from various model organisms using the CIDE-N domain and NCD, respectively (Fig 5a), we found that Cide family proteins form an independent sub-clade from the Dffa proteins and the vertebrate CIDE-N domains have close relationship with the NCD of amphioxus Dffa. These results confirm that CIDE-N domain is derived from NCD of Dffa in early vertebrates. Through the two phylogenetic trees of 17 selected Cide family proteins in vertebrates using the CIDE-N and CIDE-C domains, respectively (Fig 5b,c), we found that all the Cide family proteins can be divided into 3 subfamilies, Cidea, Cideb and Fsp27/Cidec. The CIDE-N domain NJ phylogeny is rooted by the NCD of amphioxus Dffa (Fig 5a,c). In addition, the CIDE-N domain ML phylogeny and the CIDE-C domain phylogeny generated by NJ and UPGMA analysis are rooted at midpoint (Fig 5b). These data suggest that Cideb is the most ancient member in Cide family, and its duplication resulted in the formation of Cidec and Cidea. However, the CIDE-N domain UPGMA phylogeny rooted



19.6% Positivity
9.6% Identity

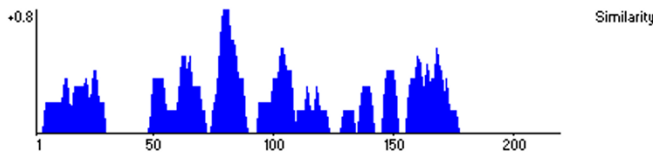


Figure 4
The full sequence alignment of human Cideb and the putative CIDE-N domain-containing protein (CND) in *C. elegans*. The CIDE-N domain is marked by black bold line and the CIDE-C domain by green bold line. The exon boundaries are marked by solid vertical lines.

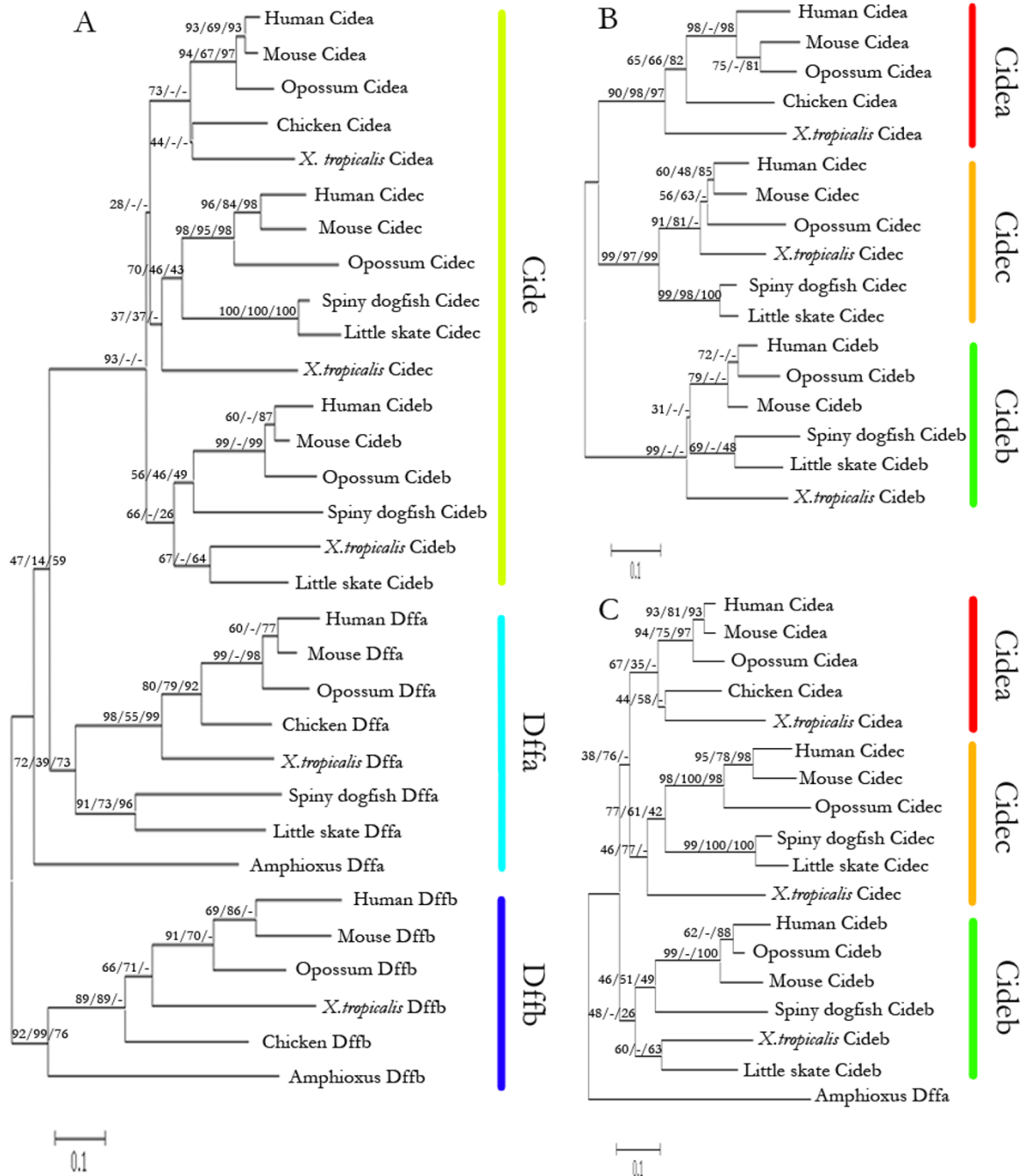


Figure 5

Neighbor-joining phylogenetic trees of selected Cide or Dff family proteins from model organisms. (A) Shown here is the NJ phylogeny of 31 representative Cide and Dff family proteins from various model organisms. The protein sequences conserved by CIDE and Dff family proteins were used and the tree was drawn by using MEGA 4.0. (B) The NJ phylogeny of 17 selected Cide family proteins in vertebrates based on the CIDE-C domain rooted by the NCD of amphioxus Dffa. (C) The NJ phylogeny of 17 selected Cide family proteins in vertebrates based on the CIDE-N domain rooted at midpoint. Bootstrap values for NJ, ML and UPGMA analyses (first, second and third values, respectively) are presented for each clade. The scale bar indicates the number of amino acid substitutions per site. Model organisms used are amphioxus (*Branchiostoma floridae*), little skate (*Leucoraja erinacea*), spiny dogfish (*Squalus acanthias*), *X.tropicalis*, chicken (*Gallus gallus*), opossum (*Monodelphis domestica*), mouse (*Mus musculus*), human (*Homo sapiens*).

by the NCD of amphioxus *Dffa* indicates *Cidec* as the most ancestral *Cide* protein. The CIDE-C domain ML phylogeny rooted at midpoint indicates *Cidea* as the most ancestral *Cide* protein. To sum up, these results point to a likelihood that *Cideb* is the most ancestral *Cide* protein.

Discussion

Here we have defined signature sequences for the N-terminal region of *Dffa*, *Dffb* and *Cide* proteins, CIDE-C domain of *Cide* proteins, and analyzed the evolutionary history of CIDE-N and CIDE-C domains of *Cide* proteins using various databases and bioinformatic tools. We identified the ancestral CIDE-N domain in hydra and found CIDE-C domain exists only in vertebrates. Furthermore, genomic structures and intron phases of *Cide* family proteins are conserved in vertebrates.

The formation of ancestral CIDE-N in early metazoan

Based on currently available data, we have found the ancestral CIDE-N domain in *H. vulgaris*, in addition to *N. vectensis* as previously reported [32]. In tropical demosponge *Reniera*, more primitive to the above two cnidarians, we did not find any sequence homologous to CIDE-N domain, with the caution that its genome sequencing is still ongoing [38]. Similarly, we did not find any CIDE-N homology in the current database of EST from sponges [39]. However, other important apoptosis genes, such as the proapoptotic molecule DD2 [40], the cell survival proteins, Bcl-2-related molecules [40], and caspases [41] have been found in the sponge. More importantly, the caspase-3 dependent DNA fragmentation was observed in sponge [41]. These data indicate that the DNA fragmentation pathway in apoptosis is conserved in sponge. Thus an ancestral CIDE-N domain, the NCD of *Dffa*, should also exist in sponge. The exact conclusion awaits further sequencing information from sponges.

The formation of CIDE-N and CIDE-C in early vertebrates

The mechanisms for the origin of new genes, such as exon shuffling, gene duplication and retroposition, have been thoroughly explored and extensively discussed [42]. It is well established that recombination of sequences encoding protein domains play a major part in protein evolution. However, there is less evidence to suggest how the novel protein domain, themselves, arise [43]. Here, we found homology sequences of CIDE-C exist only in vertebrates and all of them are linked to CIDE-N to form *Cide*. Based on protein sequence alignments (Fig 1, 2) and the phylogenetic tree (Fig 5a), CIDE-N domain of *Cide* family proteins must have been derived from the NCD of *Dffa*. In addition, from the exon boundaries derived from the protein sequence alignments results (Fig 1, 2), we could tell that exon 3 of *Cide* gene is derived from the exon 2 of *Dffa* gene. By comparison of the genic structures and intron phases between *Dffa* and *Cide* gene family, we found that

the formation of CIDE-N domain had undergone three steps of intron changes from NCD of *Dffa* (See Fig 3a for relative intron positions). First, the phase-1 intron1 of *Dffa* was changed to the phase-0 intron 2 of *Cide*. Second, the phase-1 intron 2 of *Dffa* was changed to the phase-0 intron 3 of *Cide*. Third, the phase-2 intron1 of *Cide* was formed. These three intron changes underlie the evolution of the NCD of *Dffa* into CIDE-N, and provide insight into the molecular mechanisms regarding the origin of *Cide* family. Completion of the whole genome sequencing of sea lamprey, regarded as the most primitive vertebrate thus far, will help ultimately resolve the formation of CIDE-N and CIDE-C domain.

The divergence of *Cide* family proteins in advanced vertebrates

The 2R hypothesis (two rounds of whole genome duplication in early vertebrate evolution) suggests that one round whole genome duplication happened at the root of the vertebrate lineage, followed by another round in Agnatha and Gnathostomata [44]. We found the evolution path of *Cide* proteins is in good agreement with the hypothesis. We propose that the appearance of ancient *Cide* family protein occurred at the root of the vertebrate lineage along with the first round of whole genome duplication. We suspect that the most ancient *Cide* family protein was most similar to *Cideb*, which then gave rise to the ancient *Cidea/c* around the Agnatha and Gnathostomata period which was probably accompanied with another round of whole genome duplication. *Cidec* and *Cidea* were derived from ancient *Cidea/c* at the emerging of the actinopterygian fishes. In support of the above conclusions are the following observations: no *Cide* protein was found in amphioxus; only *Cideb* and *Cidec* are found in little skate and spiny dogfish; all of *Cidea*, *Cideb* and *Cidec* are found in zebrafish and *X. tropicalis*. According to the phylogenies of CIDE-N and CIDE-C domains (Fig 5) and the evolution of other gene families such as the Hox gene family in early vertebrates [45,46], we speculate that there should be one *Cide* protein in sea lamprey with strong resemblance to *Cideb*. The lack of evidence for the presence of any *Cide* protein in sea lamprey must be due to the incomplete genome database available.

We have also compared the tissue distributions of EST clones and experimental data (Table 2), and found that there are some differences in the expression patterns of *Cide* proteins between mammals and lower vertebrates, which surprisingly revealed an expression overlap between *Cideb* and *Cidec* in lower vertebrates but not in mammals. The coexpression of *Cideb* and *Cidec* in the liver, guts and WAT of lower vertebrates is in accordance with our above-mentioned evolutionary model for the early divergence of *Cideb* and *Cidec*. *Cidea* seems to be

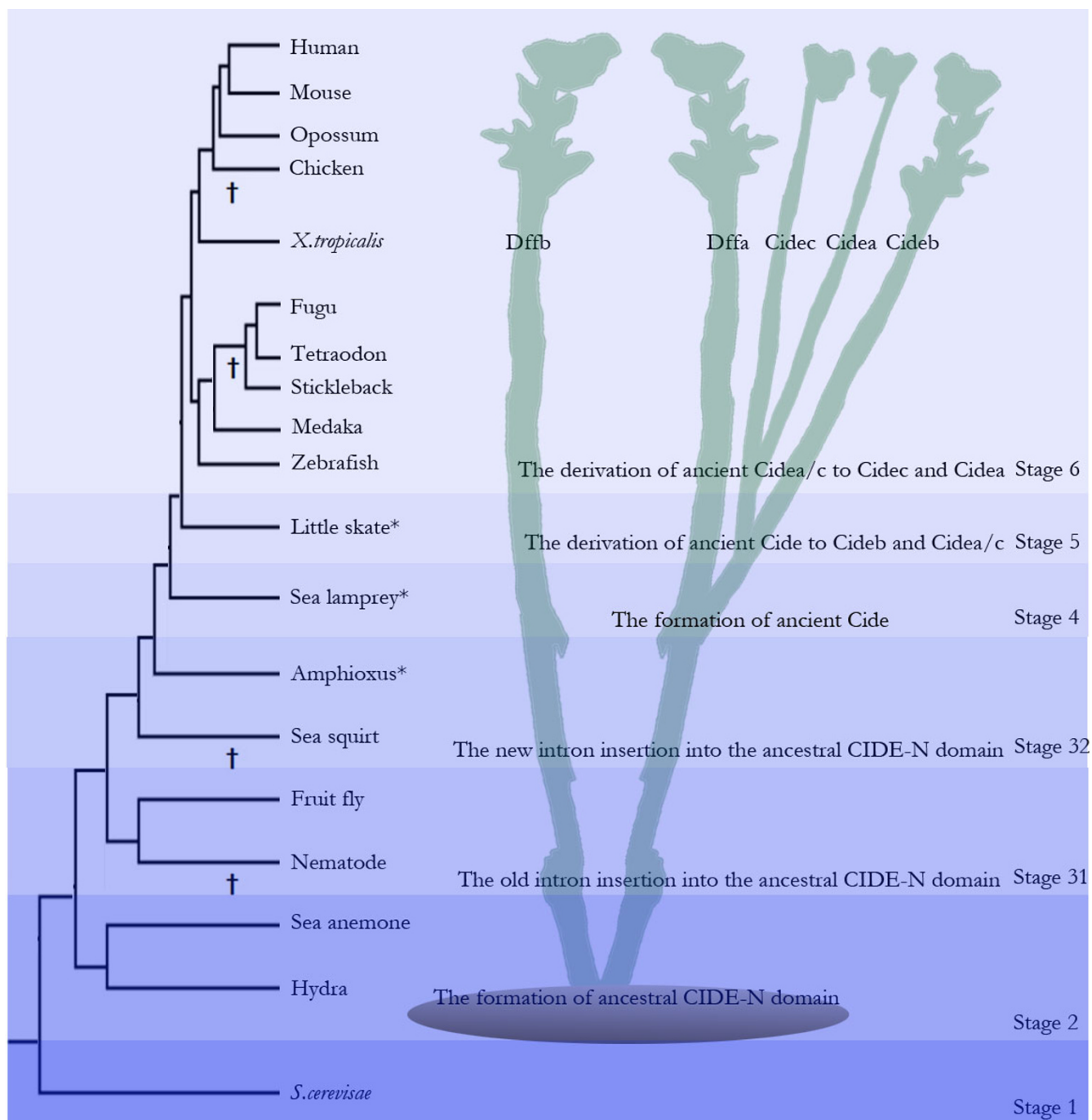


Figure 6
The Evolutionary history of Cide family proteins based on a phylogenetic tree of animals. The evolutionary history of Cide family proteins is divided into several stages based on the phylogenetic tree of animals [63–66]. †* indicate three key model organisms whose whole genomes have not been fully sequenced thus far. Absence of Cide or Dff family proteins from each species is indicated by a cross next to the respective branch of the evolutionary tree. Model organisms used are *S. cerevisiae*, hydra (*Hydra vulgaris*), sea anemone (*Nematostella vectensis*), nematode (*Caenorhabditis elegans*), fruit fly (*Drosophila melanogaster*), sea squirt (*Diazona violacea*), amphioxus (*Branchiostoma floridae*), sea lamprey (*Petromyzon marinus*), little skate (*Leucoraja erinacea*), zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), tetraodon (*Tetraodon nigroviridis*), fugu (*Takifugu rubripes*), *X. tropicalis*, chicken (*Gallus gallus*), opossum (*Monodelphis domestica*), mouse (*Mus musculus*), human (*Homo sapiens*).

highly expressed in tissues unique to mammals, including BAT and mammary gland.

Although Cide family proteins were originally identified to induce cell death [2], many studies have also found they could play an important part in modulating energy homeostasis, aging and the development of metabolic diseases such as obesity and diabetes [20,21,24-26]. Considering the urgent need in evolution for vertebrates to modulate energy homeostasis and the emerging of warm blood animals, Cide family proteins may originally to function as "thrifty" genes and gradually evolve to be important regulator of metabolic pathways in mammals [47]. Combining the vertebrate origin and the control of metabolic pathways, Cide family proteins could be ideal targets for therapeutic intervention of metabolic diseases such as obesity and diabetes.

Model for the evolutionary history of Cide family proteins

Based on our results, we mapped the presence or absence of the Cide and Dff family proteins to the phylogenetic tree of the animals, and summarized the evolutionary history of CIDE-N and CIDE-C domains into six stages (Fig 6). Around the transition of unicellular protozoan to multicellular metazoan, or the evolution of Bilateria from diploblasts (possibly the results of Cambrian explosion), one ancient or ancestral NCD for Dffa and Dffb was formed, encoded by an ancient exon bordering a phase 1 intron. Subsequent duplication led to the separation of Dffa and Dffb in cnidarians, and only the NCD of Dffa, but not Dffb, comprised the ancestral CIDE-N domain (Stage 2). In arthropods the ancient exon that encoded the ancestral CIDE-N domain was spliced by one phase 0 intron, while in nematodes the whole Dff family proteins, including the ancestral CIDE-N domain, were lost for some unknown reason (Stage 31). Around the same time in cephalochordates another phase 1 intron was inserted into a different position of the ancestral CIDE-N. This new intron insertion of the ancestral CIDE-N was later passed on to vertebrates. Also Dff family proteins might have disappeared from urochordates at this time (Stage 32). In early vertebrates like agnathan fishes, NCD of Dffa/the ancestral CIDE-N domain underwent duplication. One duplicated NCD of Dffa became the CIDE-N domain and merged with the newly formed CIDE-C domain to generate one ancient Cideb-like protein (Stage 4). Subsequent duplication led to the ancient Cidea/c protein which bears strong resemblance to Cidec in chondrichthyan fishes (Stage 5). When actinopterygian fishes occurred, Cidea was formed from the duplication of ancient Cidea/c. Some Cide family proteins might have disappeared in several vertebrate species (Stage 6).

Conclusion

In this article, we searched various databases and performed comparative genomic analysis to study the sequence conservation, genomic structure, and phylogenetic tree of the CIDE-N and CIDE-C domains of Cide proteins. We were able to define signature sequences of CIDE-N domain and CIDE-C domain for Cide proteins, and NCD for Dff proteins, respectively. Our study identified the ancestral CIDE-N domain in cnidarians, and found the CIDE-C domain exists only in vertebrates. Further analysis of genomic structure such as exon length and intron phase patterns showed although evolution of the ancestral CIDE-N domain had undergone different intron insertions to various positions in the domain among invertebrates, the genomic structure of Cide family in vertebrates is stable with conserved intron phase. We propose that NCD of Dffa was duplicated in early vertebrates, and one of the duplicated copies became CIDE-N domain that merged with the newly formed CIDE-C domain, generating an ancient Cide family protein. Subsequent duplication and evolution led to the formation of different Cide family proteins that exert their specific roles in the control of metabolic pathways in different tissues.

Methods

We retrieved human and mouse Cide and Dff family protein sequences using NCBI Entrez [48], their accession numbers are as follows: human Cidea (GenBank: [AAQ65241](#)) 219aa; human Cideb (GenBank: [AAH35970](#)) 219aa; human Cidec (GenBank: [AAH16851](#)) 238aa; human Dffa (GenBank: [AAH07721](#)) 331aa; human Dffb (GenBank: [AAC39709](#)) 338aa; mouse Cidea (GenBank: [AAH96649](#)) 217aa; mouse Cideb (GenBank: [AAH12664](#)) 219aa; mouse cidec/Fsp27 (Swiss-Prot: [P56198](#)) 239aa; mouse Dffa (GenBank: [AAH58213](#)) 331aa; mouse Dffb (GenBank: [AAH53052](#)) 343aa.

Using the MAFFT algorithm [49] implanted in Jalview [50], we performed alignments of the N-terminal regions of human and mouse Cide and Dff family proteins, and the CIDE-C domain of Cide proteins.

Hmmer search of Nr.db from NCBI

Using the well-defined CIDE-N motif pfam02017, we searched potential Cide and Dff family proteins in the downloaded NCBI non-redundant protein database [48] by Hidden Markov model search program HMMER [51], and found 287 proteins with satisfying E cutoff (<10). Then we performed multiple sequence alignment analysis to identify the resultant proteins through Jalview [50]. The nonredundant entries were summarized in Table 1.

tblastn search in EST database from NCBI

Using the two most conserved regions of mouse Cidea (a region in the CIDE-N domain of 37 amino acids:

TLVLEEDGTVDTEEFFQTLRDNTHFMILEKGQKWTP, and the other in the CIDE-C domain of the 35 amino acids: IARVTFDLYRLNPKDFLG CLNVKATMYEMYS-VSYD) revealed by the multiple sequence alignment of human and mouse Cide proteins, we conducted two tblastn searches with the EST database from NCBI [52]. In order to analyze and compare the sequences we identified, we translated all of the cDNA sequences into protein sequences using the Translate Tool from the ExPaSy server [53]. Further sequence composition analysis and alignments were performed using Jalview [50].

We carried out sequence alignment for the N-terminals of Cide and Dff family proteins from three species: hydra, sea anemone, and human; we also conducted a full sequence alignment of human Cideb and a putative CIDE-N domain-containing protein in *C. elegans* using ClustalW [54]. By doing a pair wise comparison for each of the two proteins mentioned above using Vector NTI [55], we were able to determine the sequence homology between these proteins.

Gene structure analysis using the genome database of 17 model organisms

Seventeen representative model organisms, including 11 vertebrates, 5 invertebrates and 1 fungus, were chosen in our gene structure analysis, as their genome sequences are either fully or mostly available. The genome for sea anemone was obtained from [56,57], amphioxus (*Branchiostoma floridae*) from [58], and for sea lamprey (*Petromyzon marinus*) from [59]. The genome databases for the other 14 organisms were obtained from Ensemble [60]. We summarized the nucleotide composition, length of the exons, and intron phase patterns bordering respective exons in tables, and genic structures to scale and exons to its translated protein regions in schematic figures.

Phylogenetic analysis of Cide and Dff family proteins

We retrieved the sequences of 17 Cide family proteins in selected model vertebrates, and 14 Dff family proteins in selected vertebrates and invertebrates from their genome or EST databases (Additional file 1). By the preliminary multiple sequence alignments using the MAFFT algorithm, we isolated the CIDE-N and CIDE-C domain of Cide family proteins, and NCDs of Dff family proteins. After manual alignment improvement by Jalview (Additional file 2), the phylogeny of the 17 selected CIDE-N domains for Cide and the 14 NCD domains for Dff family proteins, the phylogeny of the CIDE-C domains and that of the CIDE-N domains are separately constructed by the neighbor-joining (NJ), maximum likelihood (ML), and unweighted pair group method with arithmetic mean (UPGMA) methods. We constructed NJ and UPGMA trees using MEGA 4.0 [61], and ML trees by using PHYML V2.4.4 [62]. For NJ and UPGMA trees, Poisson correction

for amino acid sequences and 10,000 bootstrap resamplings were used, while the Jones, Taylor, and Thornton (JTT) model for amino acid sequences and 100 bootstrap resamplings were used in ML analysis. Tree files were viewed by using MEGA 4.0 [61]. NJ trees are shown with bootstrap values for NJ, ML and UPGMA analyses (first, second, and third values, respectively). Finally, we mapped the distribution of the Cide and Dff family proteins to the standard phylogenetic tree of the animals and summarized the evolutionary history of CIDE-N and CIDE-C domains into several stages.

Authors' contributions

CW carried out the analysis of Cide and Dff family proteins in human and mouse, tblastn search in EST database from NCBI and helped to draft the manuscript. YZ carried out hmmer search in Nr.db from NCBI, tissue distribution analysis, genomic structure and phylogenetic analysis using the genome database of model organisms and drafted the manuscript. ZS participated in the hmmer search in Nr.db from NCBI. PL participated in experimental design, data coordination, analysis and interpretation. PL was also responsible for the revision, finalization of the manuscript and the decision to submit the manuscript for publication. All authors read and approved the final manuscript.

Additional material

Additional file 1

Sequences of Cide and Dff family proteins used in our analysis. This table summarizes Accession Numbers of the sequences used in our phylogenetic analysis.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-159-S1.pdf>]

Additional file 2

Multiple sequence alignments for each tree building in this study. This figure depicts the multiple sequence alignments for each tree building in Fig 5 (In ALN format).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-159-S2.pdf>]

Acknowledgements

This work was supported by grants (30530350 to PL) from National Natural Science Foundation of China and National Basic Research Program of China (2006CB503909 and 2007CB914404) from the Ministry of Science and Technology of China, and fund for Innovative Research Team from Tsinghua University and Ministry of Education of China. We thank Dr. Sheng-Cai Lin for his critical reading of the manuscript, and members in Peng Li's laboratory for helpful discussion. P.L. is a Cheung Kong Scholar.

References

- Danesch U, Hoeck W, Ringold GM: **Cloning and transcriptional regulation of a novel adipocyte-specific gene, FSP27. CAAT-enhancer-binding protein (C/EBP) and C/EBP-like proteins interact with sequences required for differentiation-dependent expression.** *J Biol Chem* 1992, **267**(10):7185-7193.
- Inohara N, Koseki T, Chen S, Wu X, Nunez G: **CIDE, a novel family of cell death activators with homology to the 45 kDa subunit of the DNA fragmentation factor.** *Embo J* 1998, **17**(9):2526-2533.
- Liang L, Zhao M, Xu Z, Yokoyama KK, Li T: **Molecular cloning and characterization of CIDE-3, a novel member of the cell-death-inducing DNA-fragmentation-factor (DFF45)-like effector family.** *Biochem J* 2003, **370**(Pt 1):195-203.
- Woo EJ, Kim YG, Kim MS, Han WD, Shin S, Robinson H, Park SY, Oh BH: **Structural mechanism for inactivation and activation of CAD/DFF40 in the apoptotic pathway.** *Mol Cell* 2004, **14**(4):531-539.
- McCarty JS, Toh SY, Li P: **Multiple domains of DFF45 bind synergistically to DFF40: roles of caspase cleavage and sequestration of activator domain of DFF40.** *Biochem Biophys Res Commun* 1999, **264**(1):181-185.
- McCarty JS, Toh SY, Li P: **Study of DFF45 in its role of chaperone and inhibitor: two independent inhibitory domains of DFF40 nuclease activity.** *Biochem Biophys Res Commun* 1999, **264**(1):176-180.
- Liu X, Li P, Widlak P, Zou H, Luo X, Garrard WT, Wang X: **The 40-kDa subunit of DNA fragmentation factor induces DNA fragmentation and chromatin condensation during apoptosis.** *Proc Natl Acad Sci USA* 1998, **95**(15):8461-8466.
- Halenbeck R, MacDonald H, Roulston A, Chen TT, Conroy L, Williams LT: **CPAN, a human nuclease regulated by the caspase-sensitive inhibitor DFF45.** *Curr Biol* 1998, **8**(9):537-540.
- Enari M, Sakahira H, Yokoyama H, Okawa K, Iwamatsu A, Nagata S: **A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD.** *Nature* 1998, **391**(6662):43-50.
- Mukae N, Yokoyama H, Yokokura T, Sakoyama Y, Sakahira H, Nagata S: **Identification and developmental expression of inhibitor of caspase-activated DNase (ICAD) in Drosophila melanogaster.** *J Biol Chem* 2000, **275**(28):21402-21408.
- Inohara N, Koseki T, Chen S, Benedict MA, Nunez G: **Identification of regulatory and catalytic domains in the apoptosis nuclease DFF40/CAD.** *J Biol Chem* 1999, **274**(1):270-274.
- Mukae N, Enari M, Sakahira H, Fukuda Y, Inazawa J, Toh H, Nagata S: **Molecular cloning and characterization of human caspase-activated DNase.** *Proc Natl Acad Sci USA* 1998, **95**(16):9123-9128.
- Liu X, Zou H, Slaughter C, Wang X: **DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis.** *Cell* 1997, **89**(2):175-184.
- Sakahira H, Enari M, Nagata S: **Cleavage of CAD inhibitor in CAD activation and DNA degradation during apoptosis.** *Nature* 1998, **391**(6662):96-99.
- Inohara N, Nunez G: **Genes with homology to DFF/CIDEs found in Drosophila melanogaster.** *Cell Death Differ* 1999, **6**(9):823-824.
- Yokoyama H, Mukae N, Sakahira H, Okawa K, Iwamatsu A, Nagata S: **A novel activation mechanism of caspase-activated DNase from Drosophila melanogaster.** *J Biol Chem* 2000, **275**(17):12978-12986.
- Lugovskoy AA, Zhou P, Chou JJ, McCarty JS, Li P, Wagner G: **Solution structure of the CIDE-N domain of CIDE-B and a model for CIDE-N/CIDE-N interactions in the DNA fragmentation pathway of apoptosis.** *Cell* 1999, **99**(7):747-755.
- Uegaki K, Otomo T, Sakahira H, Shimizu M, Yumoto N, Kyogoku Y, Nagata S, Yamazaki T: **Structure of the CAD domain of caspase-activated DNase and interaction with the CAD domain of its inhibitor.** *J Mol Biol* 2000, **297**(5):1121-1128.
- Zhou P, Lugovskoy AA, McCarty JS, Li P, Wagner G: **Solution structure of DFF40 and DFF45 N-terminal domain complex and mutual chaperone activity of DFF40 and DFF45.** *Proc Natl Acad Sci USA* 2001, **98**(11):6051-6055.
- Zhou Z, Yon Toh S, Chen Z, Guo K, Ng CP, Ponniah S, Lin SC, Hong W, Li P: **Cidea-deficient mice have lean phenotype and are resistant to obesity.** *Nat Genet* 2003, **35**(1):49-56.
- Li JZ, Ye J, Xue B, Qi J, Zhang J, Zhou Z, Li Q, Wen Z, Li P: **Cideb regulates diet-induced obesity, liver steatosis and insulin sensitivity by controlling lipogenesis and fatty acid oxidation.** *Diabetes* 2007, **56**(10):2523-2532.
- Chen Z, Guo K, Toh SY, Zhou Z, Li P: **Mitochondria localization and dimerization are required for CIDE-B to induce apoptosis.** *J Biol Chem* 2000, **275**(30):22619-22622.
- Iwahana H, Yakymovych I, Dubrovskaya A, Hellman U, Souchelnytskyi S: **Glycoproteome profiling of transforming growth factor-beta (TGFbeta) signaling: nonglycosylated cell death-inducing DFF-like effector A inhibits TGFbeta1-dependent apoptosis.** *Proteomics* 2006, **6**(23):6168-6180.
- Nordstrom EA, Ryden M, Backlund EC, Dahlman I, Kaaman M, Blomqvist L, Cannon B, Nedergaard J, Arner P: **A human-specific role of cell death-inducing DFFA (DNA fragmentation factor-alpha)-like effector A (CIDEA) in adipocyte lipolysis and obesity.** *Diabetes* 2005, **54**(6):1726-1734.
- Dahlman I, Kaaman M, Jiao H, Kere J, Laakso M, Arner P: **The CIDEA gene VI15F polymorphism is associated with obesity in Swedish subjects.** *Diabetes* 2005, **54**(10):3032-3034.
- Baur JA, Pearson KJ, Price NL, Jamieson HA, Lerin C, Kalra A, Prabhu VV, Allard JS, Lopez-Lluch G, Lewis K, Pistell PJ, Poosala S, Becker KG, Boss O, Gwinn D, Wang M, Ramaswamy S, Fishbein KW, Spencer RG, Lakatta EG, Le Couteur D, Shaw RJ, Navas P, Puigserver P, Ingram DK, de Cabo R, Sinclair DA: **Resveratrol improves health and survival of mice on a high-calorie diet.** *Nature* 2006, **444**(7117):337-342.
- Puri V, Konda S, Ranjit S, Aouadi M, Chawla A, Chouinard M, Chakladar A, Czech MP: **Fat-specific protein 27, a novel lipid droplet protein that enhances triglyceride storage.** *J Biol Chem* 2007, **282**(47):34213-34218.
- Keller P, Petrie JT, De Rose P, Gerin I, Wright WS, Chiang SH, Nielsen AR, Fischer CP, Pedersen BK, Macdougald OA: **Fat-specific protein 27 regulates storage of triacylglycerol.** *J Biol Chem* 2008, **283**(21):14355-14365.
- Kim JY, Liu K, Zhou S, Tillison K, Wu Y, Smas CM: **Assessment of fat-specific protein 27 in the adipocyte lineage suggests a dual role for FSP27 in adipocyte metabolism and cell death.** *Am J Physiol Endocrinol Metab* 2008, **294**(4):E654-E667.
- Li P: **Cidea, brown fat and obesity.** *Mech Ageing Dev* 2004, **125**(4):337-338.
- Lin SC, Li P: **CIDE-A, a novel link between brown adipose tissue and obesity.** *Trends Mol Med* 2004, **10**(9):434-439.
- Eckhart L, Fischer H, Tschachler E: **Phylogenomics of caspase-activated DNA fragmentation factor.** *Biochem Biophys Res Commun* 2007, **356**(1):293-299.
- Darling JA, Reitzel AR, Burton PM, Mazza ME, Ryan JF, Sullivan JC, Finnerty JR: **Rising starlet: the starlet sea anemone, Nematostella vectensis.** *Bioessays* 2005, **27**(2):211-221.
- Antonio C, Marques J, Collins AG: **Cladistic analysis of Medusozoa and cnidarian evolution.** *Invertebrate Biology* 2004, **123**(1):23-42.
- Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoeff F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S: **Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes.** *Science* 2002, **297**(5585):1301-1310.
- Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher C, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crolius H: **Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype.** *Nature* 2004, **431**(7011):946-957.
- Consortium ICGS: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.** *Nature* 2004, **432**(7018):695-716.

38. **The tropical demosponge Reniera Genome Database** [<http://www.jgi.doe.gov/sequencing/why/CSP2005/reniera.html>]
39. **The SpongeBase Blast Server** [<http://spongebase.uni-mainz.de/cgi-bin/blast/blastserver.cgi>]
40. Wiens M, Krasko A, Muller CI, Muller WE: **Molecular evolution of apoptotic pathways: cloning of key domains from sponges (Bcl-2 homology domains and death domains) and their phylogenetic relationships.** *J Mol Evol* 2000, **50(6)**:520-531.
41. Wiens M, Krasko A, Perovic S, Muller WE: **Caspase-mediated apoptosis in sponges: cloning and function of the phylogenetic oldest apoptotic proteases from Metazoa.** *Biochim Biophys Acta* 2003, **1593(2-3)**:179-189.
42. Long M, Betran E, Thornton K, Wang W: **The origin of new genes: glimpses from the young and old.** *Nat Rev Genet* 2003, **4(11)**:865-875.
43. Schmidt EE, Davies CJ: **The origins of polypeptide domains.** *Bioessays* 2007, **29(3)**:262-270.
44. Makalowski W: **Are we polyploids? A brief history of one hypothesis.** *Genome Res* 2001, **11(5)**:667-670.
45. Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH: **Zebrafish hox clusters and vertebrate genome evolution.** *Science* 1998, **282(5394)**:1711-1714.
46. Meyer A, Malaga-Trillo E: **Vertebrate genomics: More fishy tales about Hox genes.** *Curr Biol* 1999, **9(6)**:R210-213.
47. Lazar MA: **How obesity causes diabetes: not a tall tale.** *Science* 2005, **307(5708)**:373-375.
48. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic Acids Res* 2007:D21-25.
49. Katoh K, Misawa K, Kuma K, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids Res* 2002, **30(14)**:3059-3066.
50. Clamp M, Cuff J, Searle SM, Barton GJ: **The Jalview Java alignment editor.** *Bioinformatics* 2004, **20(3)**:426-427.
51. Eddy SR: **Profile hidden Markov models.** *Bioinformatics* 1998, **14(9)**:755-763.
52. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25(17)**:3389-3402.
53. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A: **ExPASy: The proteomics server for in-depth protein knowledge and analysis.** *Nucleic Acids Res* 2003, **31(13)**:3784-3788.
54. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22(22)**:4673-4680.
55. Lu G, Moriyama EN: **Vector NTI, a balanced all-in-one sequence analysis suite.** *Brief Bioinform* 2004, **5(4)**:378-388.
56. **The Nematostella vectensis Genome Database** [<http://genome.jgi-psf.org/Nemvel1/Nemvel1.home.html>]
57. Sullivan JC, Ryan JF, Watson JA, Webb J, Mullikin JC, Rokhsar D, Finnerty JR: **StellaBase: the Nematostella vectensis Genomics Database.** *Nucleic Acids Res* 2006:D495-499.
58. **The Branchiostoma floridae Genome Database** [http://www.sanger.ac.uk/Projects/B_floridae/]
59. **The Petromyzon marinus Genome Database** [<http://genome.wustl.edu/genome.cgi?GENOME=Petro myzon%20marinus&GROUP=2>]
60. Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Herrero J, Holland R, Howe K, Howe K, Johnson N, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Melsopp C, Megy K, Meidl P, Ouverdin B, Parker A, Prlic A, Rice S, Rios D, Schuster M, Sealy I, Severin J, Slater G, Smedley D, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wood M, Cox T, Curwen V, Durbin R, Fernandez-Suarez XM, Flicek P, Kasprzyk A, Proctor G, Searle S, Smith J, Ureta-Vidal A, Birney E: **Ensembl 2007.** *Nucleic Acids Res* 2007:D610-617.
61. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24(8)**:1596-1599.
62. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52(5)**:696-704.
63. Hedges SB: **The origin and evolution of model organisms.** *Nat Rev Genet* 2002, **3(11)**:838-849.
64. Mulley J, Holland P: **Comparative genomics: small genome, big insights.** *Nature* 2004, **431(7011)**:916-917.
65. Roest Crollius H, Weissenbach J: **Fish genomics and biology.** *Genome Res* 2005, **15(12)**:1675-1682.
66. Philippe H, Lartillot N, Brinkmann H: **Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia.** *Mol Biol Evol* 2005, **22(5)**:1246-1253.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

