Biochemical and Biophysical Research Communications 450 (2014) 219–226

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Molecular phylogeny of C1 inhibitor depicts two immunoglobulin-like domains fusion in fishes and ray-finned fishes specific intron insertion after separation from zebrafish

Abhishek Kumar [a,*], Anita Bhandari [b], Sandeep J. Sarde [a], Chandan Goswami [c]

[a] Department of Genetics & Molecular Biology in Botany, Institute of Botany, Christian-Albrechts-University at Kiel, Kiel, Germany
[b] Molecular Physiology, Zoological Institute, Christian-Albrechts-University at Kiel, Kiel, Germany
[c] National Institute of Science Education and Research, Bhubaneswar, Orissa, India

ARTICLE INFO

Article history:
Received 21 April 2014
Available online 27 May 2014

Keywords:
C1 inhibitor
Serpin G1
Group V4
Synteny
Phylogenetic analysis
Intron gain

ABSTRACT

C1 inhibitor (C1IN) is a multi-facet serine protease inhibitor in the plasma cascades, inhibiting several proteases, notably, regulates both complement and contact system activation. Despite huge advancements in the understanding of C1IN based on biochemical properties and its roles in the plasma cascades, the phylogenetic history of C1IN remains uncharacterized. To date, there is no comprehensive study illustrating the phylogenetic history of C1IN. Herein, we explored phylogenetic history of C1IN gene in vertebrates. Fishes have C1IN with two immunoglobulin like domains attached in the N-terminal region. The RCL regions of CIIN from fishes and tetrapod genomes have variations at the positions P2 and P1′. Gene structures of C1IN gene from selected ray-finned fishes varied in the Ig domain region with creation of novel intron splitting exon Im2 into Im2a and Im2b. This intron is limited to ray-finned fishes with genome size reduced below 1 Gb. Hence, we suggest that genome compaction and associated double-strand break repairs are behind this intron gain. This study reveals the evolutionary history of C1IN and confirmed that this gene remains the same locus for ~450 MY in 52 vertebrates analysed, but it is not found in frogs and lampreys.

1. Introduction

C1 inhibitor (C1IN) is a multi-functional serine protease inhibitor (serpin), which operates by inactivating various serine proteases in different plasmatic cascades including the complement (classical pathway: C1r and C1s; lectin pathway MASP1 and MASP2), contact (factor XII and kallikrein), coagulation (factor XI and thrombin) and fibrinolytic (tPA and plasmin) systems [1,2]. The protein structure of serpin domain consists of three β-sheets, sA-sC and 8–9 α-helices, hA-hI [3]. The trademark of serpin biology is the exposed flexible loop (~17–20 residues) known as reactive center loop (RCL), which serves as a bait mimicking a protease substrate that is cleaved between the active sites P1 and P1′ [3]. Serpins introduce molecular diversities by adding of additional sequences at the terminal end of this core domain structures and also via changing residues in their RCL regions. C1IN belongs to clade G within the serpin superfamily, hence also called as serpinG1 [3]. This gene belongs to vertebrate group V4 in the group-wise vertebrate serpin classification system [4,5]. The hallmark of C1 inhibitor is its deficiency, which causes a rare genetic disorder, hereditary angioedema (HAE), swelling caused by leakage of fluid from blood vessels into connective tissue [2]. In the year 1957, Ratnoff and Lepow discovered CIIN as inhibitor of C1 protein in the plasma [6]. Over last five decades, C1IN has been extensively characterized by biochemical and biophysical methods to demonstrate its roles in the plasmatic cascades. However, its molecular phylogenetic analysis has not been focused, primarily due to the difficulties associated with the reconstruction of phylogenetic relationships among different animals and presence of several paralogs across animal genomes [4]. Hence, an investigation on molecular phylogenetic aspects of C1IN is required. Herein, we elucidated the detailed molecular phylogeny of C1IN genes by combining protein sequence, gene structures and genomic organization from 52 vertebrates.

2. Materials and methods

2.1. Sequence collection

We collected genomic DNA and protein sequences from different vertebrate genomes via Ensembl (release 75, Feb 2014) [7] using BLAST suite for C1IN are listed in Table S1.

* Corresponding author.
E-mail address: akumar@bot.uni-kiel.de (A. Kumar).

http://dx.doi.org/10.1016/j.bbrc.2014.05.097
0006-291X/

2.2. Gene structure prediction and mapping intron positions

To ensure accuracy, gene structure prediction within the Ensembl [7] was combined with predictions of AUGUSTUS gene prediction tool [8]. Mature human α_1-antitrypsin was used as standard sequence for intron position mapping and numbering of intron positions, followed by suffixes a–c for their location as reported previously [4].

2.3. Chromosomal mapping

Chromosomal locus for C1IN gene was scanned for each species and synteny maps were constructed from fishes to mammalian genomes using Ensembl genome browser [7].

2.4. Sequence & structural analysis

Protein alignment of C1IN was created using the MUSCLE [9] and visualized in GENEDOC [10] as shown in Fig. 1S. Sequence logos of conserved RCL in C1IN protein were constructed by Weblogo 3.3 [11].

2.5. Phylogenetic analyses

Two phylogenetic trees were constructed aided by Bayesian (2 runs, until average standard deviation of split frequencies was lower than 0.0098, 25% burn-in-period) and Maximum likelihood methods (1000 bootstraps) using Mrbayes 3.2.1 [12] and MEGA5 [13] as following vertebrate serpins (259 serpins) and C1IN proteins (47 sequences), respectively. These two trees are based on WAG [5 categories (+*G*, parameter = 4.6121)] and JTT [5 categories (+G, parameter = 1.3836)] protein substitution models, respectively, as their best models were computed in MEGA5 [13].

2.6. Protein modelling

Structural model of C1IN protein from platyfish (XmaC1IN) was created using the I-TASSER [14]. This model was visualized within YASARA [15].

2.7. Calculations of genome sizes

Selected vertebrate genome sizes were calculated using the animal genome size database [16] and the Ensembl genome database (release 75, Feb 2014) [7].

3. Results

3.1. Two immunoglobulin domains are fused in the N-terminal end of C1 inhibitor gene in selected fishes

Human C1IN is a 500 amino acid spanning protein with first ~120 amino acids are N-terminal extension including signal peptide (Fig. 1A). This domain organization (with N-terminal extension in the core serpin domain) is maintained in reptiles, birds and mammals. However, fishes have the core serpin domain but they possess two immunoglobulin-like domains (marked red in Fig. 1A) separated by non-immunoglobulin region (grey color). HMMER (http://hmmer.janelia.org/) based Pfam domain scanning reveals that platyfish C1IN protein (Ensembl protein Id – ENSXMAP00000009959) has two immunoglobulin-like domains (Pfam Id – PF13895.1) in the regions 24–107 and 110–197 amino acids and serpin domain (Pfam Id – PF00079.15) follows in the region of 241–597. Generally, these Ig domains are about 80 residues long, which are composed of two-layers of sandwich between 7 and 9 β-strands arranged in two β-sheets with a Greek key topology [17,18] and involved in either protein–protein or protein ligand interactions [19]. To visualize we created homology model of platyfish C1IN protein using I-TASSER tool [14]. Fig. 1B illustrates this model structure, which clearly shows a serpin domain with intact RCL region (marked blue in Fig. 1B) with positions of P1–P1′ marked.

C1IN is conserved in all analyzed vertebrates from cave fish to human with variable sequence identities and similarities (Table 1). Human C1IN shares 46/60, 31/47, 22/41 and 21/38 percentage identities/similarities with opossum, turkey, tilapia and cave fish, respectively. There are 24, 84 and 84 amino acids in C1IN alignment, which are conserved 100%, 70–99% and 50–69%, respectively (Table 2). Additionally, there are 51 amino acids conserved in the core serpin domain in more than 70% of serpins [20] and from these 51 residues, 48 were maintained in vertebrate C1IN (Table 2).

C1IN with two Ig domains has been reported in other ray-finned fishes such as the Japanese flounder, *Paralichthys olivaceus* [21], rainbow rout, *Oncorhynchus mykiss* [22] and rock beam, *Oplegnathus fasciatus* [23]. Hence, the C1IN with Ig domains are common features of fishes. Fish specific alignment of C1IN is shown in Fig. S1 and Ig and serpin domains are marked by red and blue lines. Ig domains have invariant cysteine residues marked by red color where conserved disulfide bridges are marked by blue colors as C1 and C2 pairs. RCL regions of C1IN protiens from tetrapods and fishes show some variations but are inhibitory in nature (Figs. 1C and 1S). Major differences are at P2 positions as alanine and serine are present in tetrapod and fish C1IN respectively. P1′ position is variable in tetrapods where as fishes have a serine in higher frequency. We computed fish specific sequence identities and similarities as shown in Table 3, which hints that Ig domain regions are more diverged than serpin domain, for example platyfish show percentage sequence identities and similarities with rainbow rout as 46/64 and 57/76 in the full length and in the serpin domain, respectively.

All in all, fishes have specialized C1 inhibitor with fusion of two Ig domains in N-terminal part.

3.2. Spliceosomal intron gain in the Ig domain region in fish-specific C1IN gene

Human C1IN gene has five introns at positions 67a, 123a, 192a1, 238c and 307a (according to mature human α_1-antitrypsin numbering) creating six exons (termed as Sp1–Sp6) in the core serpin and one extra exon in the N-terminal extension (designed as Sp0) (Fig. 2). Turkey has same gene structure without the exon Sp0, but turtles have this exon Sp0 in the N-terminal extension in the serpin. This reflects that N-terminal extension is common in tetrapod genomes, but it is extended in species-specific manner. Notably, the fishes have two immunoglobulin-like (Ig) domains (as depicted in Fig. 1), this portion leads into three exons, Im1-Im3 in *Danio rerio* and in cave fish (not shown). However, a novel intron is inserted the exon Im2 (red color), splitting this exon into two exons Im2a and Im3b (green color) in selected ray-finned fishes such as *Takifugu*, *Tetraodon*, rock beam (deduced from a recent study [24]), stickleback (as shown in Fig. 2), platyfish and tilapia (not shown). The novel intron is localized in between two Ig domains (Fig. 1S). The novel intron is ranged from 76 bp (in *Tetraodon*) to 123 bp (in stickleback), while created exons Im2a and Im2b are ranged as 109–112 bp and 162–168 bp, respectively; while single exon Im2 is 274 bp in *D. rerio*. Additionally, exons Im1 and Im3 are constant as 48 bp and 291–339 bp, respectively. Size of intron between the exons Im1-Im2 is largely variable and ranged from 69 bp (in *Takifugu*) to 921 bp (in *D. rerio*). Whereas intron size of the intron between the exons Im2-Im3 is ranged from 113 bp (in *Takifugu*) and 2011 bp (in *D. rerio*). In fishes, the

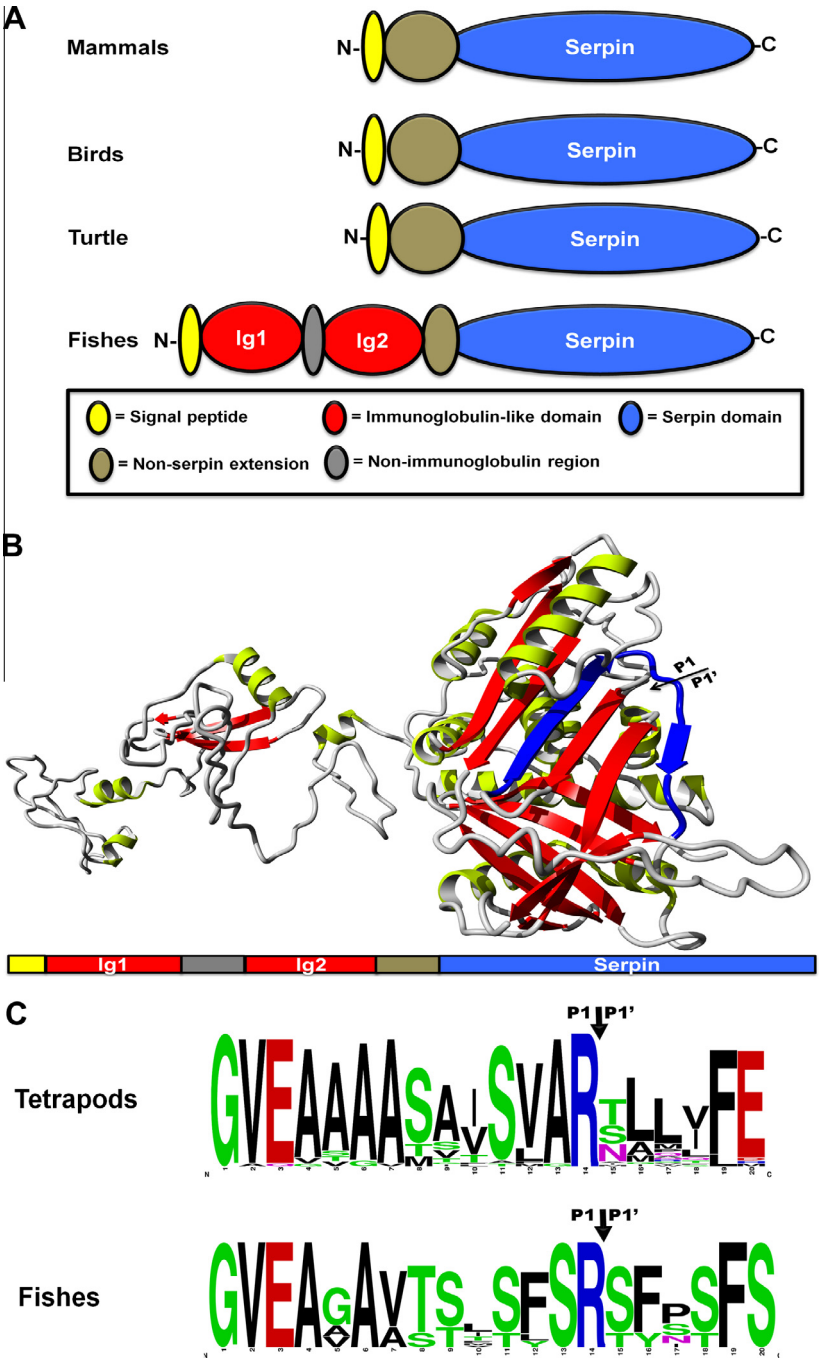

Fig. 1. Fishes have two immunoglobulin domains in C1IN, but not in any tetrapod. (A) Protein domain architecture of vertebrate C1IN illustrates fishes have additional two Ig domains in addition to core serpin domain in C1IN. (B) Protein model of C1IN protein from platyfish (*Xiphophorus maculatus*) depicts two Ig domains in the N-terminal end. (C) Comparisons of RCL regions of tetrapod and fishes.

intron connecting exons Im3 to the exon Sp1 is ranged from 80 bp (in stickleback) and 660 (in *Takifugu*). Sizes of the exons Sp1 to Sp6 are ranged as 207–499 bp, 135–144 bp, 204 bp (fixed size), 140–143 bp, 217–223 bp and 248–254 bp, respectively. Sizes of the five introns at positions 67a, 123a, 192a1, 238c and 307a in the core serpin are ranged as 83–1784 bp, 71–3840 bp, 72–346 bp, 154–5169 bp and 70–2391 bp, respectively. This hints that intron of CIIN gene varies from few base pairs to some thousands base pairs and sizes are normally smaller for fishes with same genomes such as Fugu and *Tetraodon*.

Overall, selected ray-finned fishes have CI inhibitors with two Ig domains as the N-terminal extension. One intron is inserted in the exon Im2, creating two exons (designed as Im2a and Im2b) after separation from *D. rerio*.

3.3. Synteny analyses reveal the chromosomal locus of C1IN is conserved from ~450 MY

In the human chromosome 11, C1IN gene is flanked by hexad of genes (Fig. 3) – APLNR (apelin receptor), SSRP1 (structure specific recognition protein 1) P2RX3 (purinergic receptor P2X, ligand-gated ion channel, 3), RTN4RL2 (reticulon 4 receptor-like 2), SLC43A1 (solute carrier family 43A, member 1) and TIMM10 (translocase of inner mitochondrial membrane 10 yeast-homolog)

Table 1
Percentage identities and similarities values of C1IN for selected vertebrates.

	1	2	3	4	5	6	7	8	9	10	11	12	13	
1. Human		46	29	31	23	23	22	22	24	22	16	24	21	Sequence identities
2. Opossum	60		29	31	27	22	22	22	22	22	14	20	20	
3. Chicken	45	46		70	22	20	19	20	22	22	12	21	19	
4. Turkey	47	49	74		22	21	19	21	22	22	13	21	20	
5. Cod	39	43	38	38		35	35	37	34	40	15	30	30	
6. Fugu	41	38	36	35	50		47	50	66	57	20	33	33	
7. Platyfish	41	39	36	35	51	68		54	48	56	17	35	34	
8. Tilapia	41	38	37	37	54	67	70		50	60	20	35	33	
9. Tetraodon	41	38	37	36	50	79	69	71		55	20	36	32	
10. Stickleback	40	37	37	37	55	74	76	75	72		19	36	35	
11. Spottedgar	28	25	22	23	25	33	31	34	34	34		19	19	
12. Zebrafish	40	36	36	37	42	54	53	54	53	53	32		41	
13. Cavefish	38	35	33	34	41	53	53	51	52	52	32	60		
	Sequence similarities													

Table 2
Summary of amino acid conservations in secondary structural elements of C1IN protein in vertebrates.

Structural components	Id-100 (red)	Id-70–99 (green)	Id-50–69 (yellow)	Status of 51 conserved amino acids proposed by Irving et al. (2000) [20]. Bold – Missing Italics – Conserved in few species				
Signal peptide	0	3	6					
N-terminal end after signal peptide	0	3	7					
hA	2	11	12	F33				
s6B	0	1	2	N49	S53			
hB	3	5	2	P54	*S56*	L61	G67	
hC	0	2	1	T72	L80			
hD	0	1	1					
s2A	0	5	4					
hE	2	3	2	F130				
s1A	1	0	0					
hF	4	2	4	**F147**	*I157*	N158	V161	T165
Loop between hF/s3A	1	5	6	*I169*	*T180*			
s3A	1	5	0	L184	N186	F190	**K191**	*G192*
hF1	0	0	1					
s4C	1	0	0	F198	*T203*	F208		
s3C	1	3	4	V218	*M220*	*M221*		
s1B	1	0	1					
s2B	1	1	2	**Y244**				
s3B	3	2	2	L254	P255			
hG	1	1	2					
hH	0	0	1					
s2C	2	1	4					
s6A	0	3	2	P289	*K290*			
hI	0	1	2	*L299*	*L303*	G307		
hI1	0	2	0					
Loop between hI/s5A	0	4	4	*F312*	A316	*L327*		
s5A	0	3	1	H334	E342			
s4A (RCL)	0	7	5	G344	A347			
s1C	0	1	1					
s4B	0	3	3	P369	F370			
s5B	0	5	1	L383	F384	G386		
C terminus	0	1	1	P391				

Id = Identity.

Table 3
Percentage identities and similarities values of C1IN for selected fishes, Values are for full-length C1 inhibitor outside bracket and only serpin domain in the bracket.

	1	2	3	4	5	6	7	8	9	10	11	
1. Cavefish		33 (42)	36 (45)	33 (43)	35 (46)	35 (44)	34 (42)	32 (42)	33 (42)	35 (44)	41 (50)	Sequence identities
2. Fugu	52 (64)		59 (66)	51 (62)	46 (57)	60 (69)	47 (58)	66 (74)	50 (60)	57 (67)	34 (44)	
3. Larimichthys	53 (63)	75 (82)		59 (68)	52 (61)	72 (77)	56 (66)	56 (63)	59 (68)	65 (72)	38 (47)	
4. Maylandia	51 (64)	68 (77)	73 (82)		50 (61)	64 (75)	54 (67)	51 (61)	92 (94)	60 (71)	35 (45)	
5. Oncorhynchus	52 (66)	65 (73)	67 (77)	66 (77)		53 (62)	46 (57)	46 (57)	49 (60)	53 (63)	38 (51)	
6. Oplegnathus	52 (63)	77 (84)	83 (87)	77 (85)	70 (79)		60 (70)	61 (71)	64 (74)	74 (81)	37 (47)	
7. Platyfish	53 (63)	68 (78)	73 (80)	69 (80)	64 (76)	78 (85)		48 (59)	54 (67)	56 (66)	35 (46)	
8. Tetraodon	52 (64)	79 (86)	74 (80)	71 (80)	64 (75)	78 (86)	70 (79)		50 (61)	55 (65)	37 (47)	
9. Tilapia	51 (64)	67 (76)	74 (83)	96 (97)	66 (77)	77 (86)	70 (80)	70 (80)		61 (70)	35 (45)	
10. Stickleback	52 (64)	74 (82)	81 (87)	75 (84)	69 (79)	85 (91)	76 (84)	72 (80)	75 (85)		37 (46)	
11. Zebrafish	60 (68)	55 (67)	56 (66)	55 (65)	56 (70)	55 (66)	55 (66)	54 (66)	55 (65)	54 (66)		
	Sequence similarities											

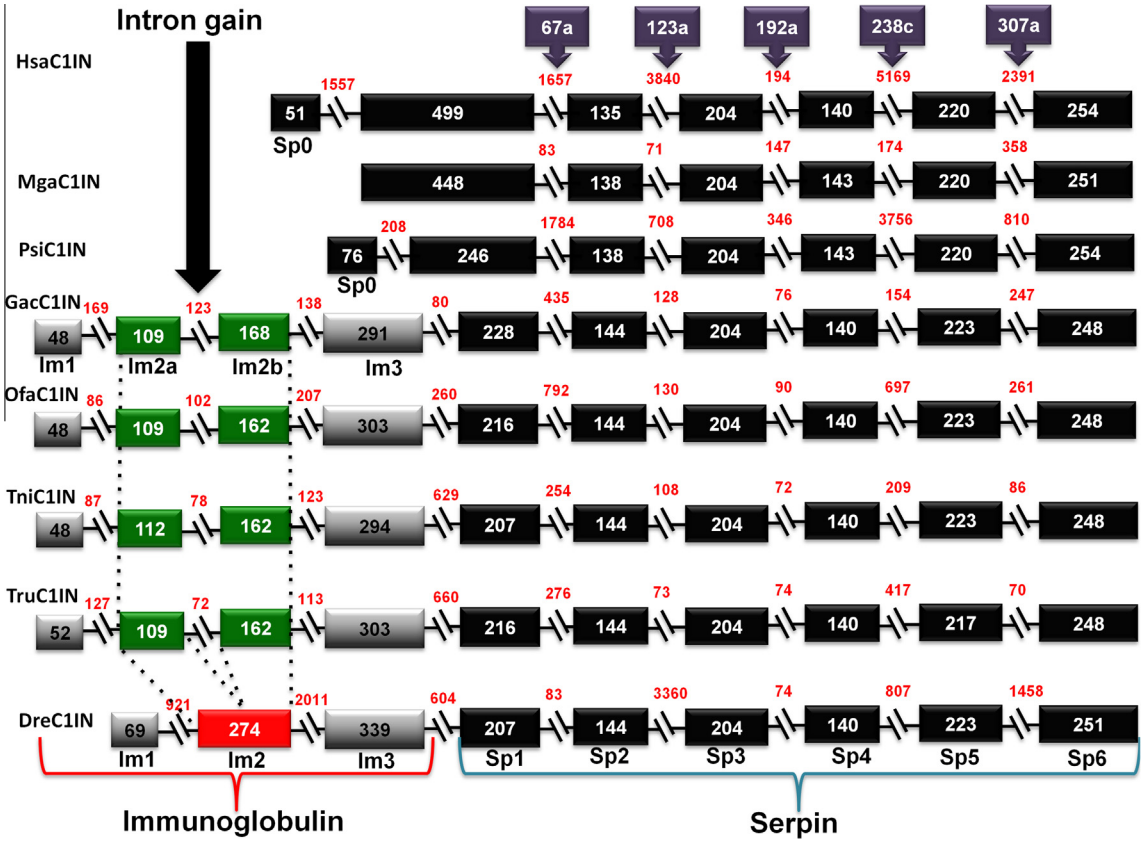

Fig. 2. Gene structure patterns of vertebrate C1IN gene demonstrate that Ig domain part has a gain of intron in selected ray-finned fishes after separation from *D. rerio*. Exon Im2 (red color) of *D. rerio* splitted into Im2a and Im2b (green boxes) in other ray-finned fishes. Exons in immunoglobulin and serpin domains are named with prefixes Im and Sp, respectively. There is no change of gene structure in the serpin domain throughout vertebrate evolution, only N-terminal extensions in human and turtle have extra exons named as Sp0. Dre – Zebrafish (*Danio rerio*); Hsa – Human (Homo sapiens); Gac – Stickleback (*Gasterosteus aculeatus*); Mga – Turkey (*Meleagris gallopavo*); Ofa – Rock bream (*Oplegnathus fasciatus*); Psi – Turtle (*Pelodiscus sinensis*); Tni – *Tetraodon nigroviridis*; Tru - Fugu (*Takifugu rubripes*). Gene structure of C1IN of rock beam, *Oplegnathus fasciatus* is deduced from a recent study [23]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

on the one side. The other side has a triad of genes – ZDHHC5 (zinc finger, DHHC-type containing 5), BTBD18 (BTB/POZ domain containing 18) and OR6Q1 (olfactory receptor, family 6, subfamily Q, member 1) (Fig. 3). This genomic organization is conserved across all mammals such as mouse (chromosome 2), rat (chromosome 3), horse (chromosome 12) and opossum (chromosome 5). This syntenic arrangement is conserved in birds such as chicken (chromosome 2) flycatcher (scaffold JH603356.1) zebra finch (chromosome 15) and turkey (chromosome 5). Two reptile species also possesses this genomic fragment, such as lizard (chromosome 1) and turtle (scaffold JH211036.1). Fishes have different sets of flanking markers such as diad of genes on both side as GPR83 (G protein-coupled receptor 83) and TREM134 (transmembrane protein 134) on one side where as the other side contains the second diad consist of DOC2D (double C2-like domains, delta) and CABP4 (calcium binding protein 4). This locus is maintained in fishes namely Atlantic cod (genescaffold 3668), medaka (chromosome 1), platyfish (scaffold JH556863.1), stickleback (group IX), *Takifugu* (scaffold 281), Tetraodon (chromosome Un_ran), and tilapia (Scaffold GL831282.1) with novel intron inserted. Coelacanth (scaffold JH205199.1), cave fish (scaffold KB871929.1), spotted gar (chromosome LG9) and zebrafish (chromosome 8) also shared this syntenic arrangement but do not have novel introns. By reverse tracing these four marker genes (flanking C1IN in fishes), we found that human (chromosome 11) and horse (chromosome 12) have two of these markers TREM134 and CABP4 in same orientation but distanced by 10 Mb. Similarly we found that birds and reptiles also maintained these two marker genes on same genomic region. However, these two markers are conserved on another chromosome in mouse (chromosome 19), Rat (chromosome 1), horse (chromosome 12) and opossum (chromosome 8). This suggested that C1 inhibitor is conserved on same locus from fishes to human but there is shuffling of C1 inhibitor gene from this locus for ~10 Mb in human. Interestingly, frogs have no C1 inhibitor gene as not detected in genomes of *Xenopus laevis* and *X. tropicalis*. However, we found that TIMM10 and ZDHHC5 are localized on two different scaffolds (GL172925.1 and GL173350.1, respectively) on detected flanking genes (Fig. 3). Additionally, lampreys also have no C1 inhibitors but they possess other group V4 members, such as α2-antiplasmin-like (A2APL) serpins [24]. These data suggest that this serpin originated at the early stages of vertebrate evolution, ca 450 MYA.

3.4. Phylogenetic analysis of C1IN

Serpins are divided into six groups (namely V1–V6) as depicted in Fig. 4A. Group V1 is clade B members whereas group V2 possess several members of α_1-antitrypsin-like serpins. Other groups have limited numbers of serpins with group V5–V6 being single serpin groups. C1IN gene is a member of group V4, along with pigment epithelium derived factor (PEDF), α_2-antiplasmin as demonstrated by evolutionary history of vertebrate serpins (Fig. 4A). To evaluate location of novel intron insertions, species-wide phylogenetic tree was constructed (Fig. 4B), which reveals that after separation of cave fish and zebrafish from other-ray-finned fishes, this novel intron is embedded in the canonical introns. Fig. 4B also depicts

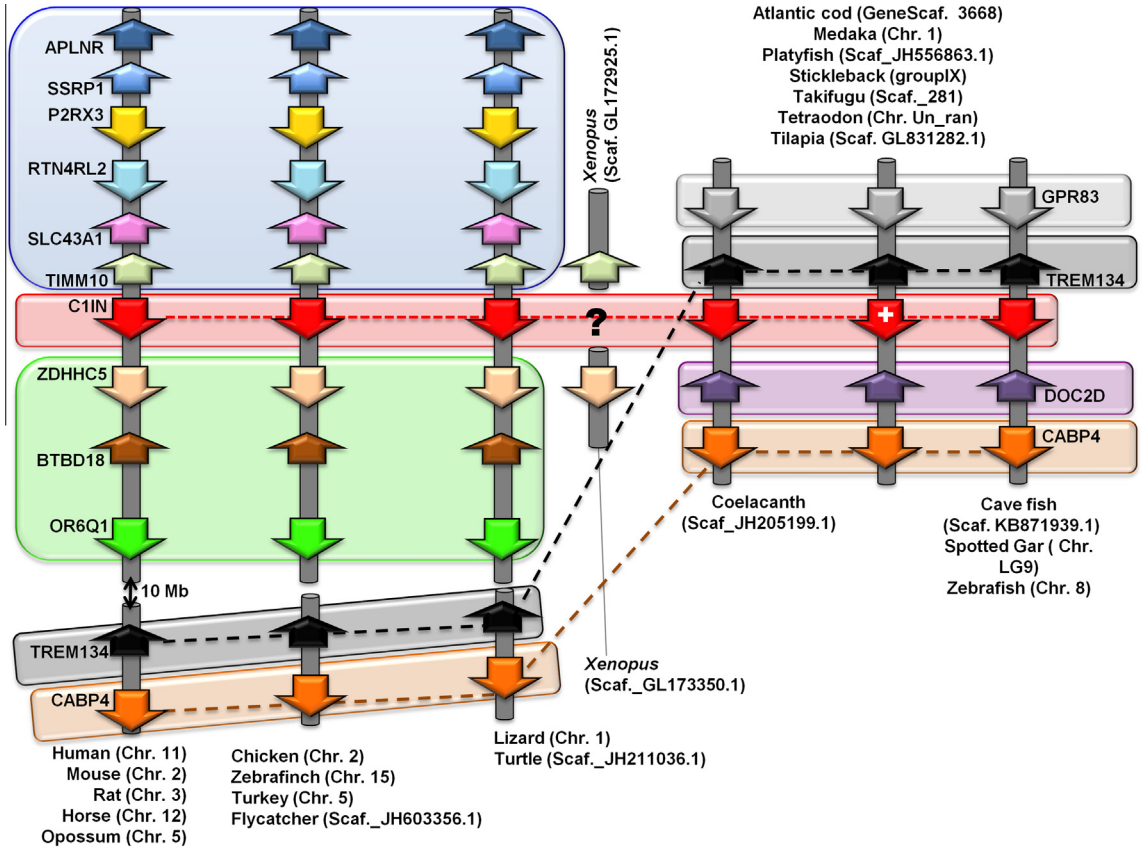

Fig. 3. Syntenic organizations of genes flanking C1IN (serpinG1) reveal changes in fishes with respect to tetrapod. Frogs have no C1IN, however two flanking genes are detected in the scaffold in the *Xenopus* genome.

that domain fusion events of Ig and serpins only occurred in ray-finned fishes.

4. Discussion

C1IN is a critical regulator of the activation of various plasmatic cascades associated with immunity and inflammation. However, despite great efforts on discovering the molecular mechanisms and clinical relevance of C1IN gene and protein functions, questions about molecular phylogeny of remain unanswered. The current study provides an updated repository of the C1IN gene from 55 vertebrate species (Table 1S) and summarizes major concepts considering three critical factors: i.e. sequence, structure and phylogeny of C1IN across vertebrate genomes.

C1IN gene is characterized by the presence of six exons namely Sp1–Sp6 in vertebrates in the core serpin domain. However, fishes have a N-terminal fusion of two Ig domains, which has three exons Im1–Im3 but the exon Im2 is divided into pieces (Im2a and Im2b) by the invasion of novel intron (Fig. 2). Eukaryotic genes are expressed as pre-mRNAs that are converted to mRNA by splicing mechanisms, which removes introns and exons, creating expressing segment of the genes [25]. Spliceosomal introns and its splicing machinery are hallmarks of eukaryotic genomes. However, the mystery about their creation remains puzzling [26]. There are total 24 conserved introns in vertebrate serpins encompassing group V1–V6 [5] with six additional introns that were gained in selected ray finned fishes among serpin genes [6]. But this current study is the first example of conserved introns in non-serpin domain, as no other reports exist by today. Ray-finned fishes with novel introns are examples of unique genomes with reducing genetic contents as shown in Table 4. Several types of gene rearrangements (such as inversions, translocations, duplications and transpositions) are typical features of genome evolution and these lead into gene fusion and gains of introns such as what observed in this study. So far, there are seven different mechanisms, which have been proposed for intron gain/invasions [27,28]: (a) intron transposition with partial recombination, (b) transposon insertion, (c) tandem genomic duplication using duplicated splice sites, (d) double-strand break repair (DSBR), (e) group II intron insertion, (f) intron transfer (g) intronization. Genome compaction and associated double-strand break repair (DSBR) were attributed with several examples of intron creations in selected ray-finned fishes whose genome underwent compaction events in the serpin superfamily [6] and in the GPCR superfamily [29]. These repair processes involved in successful genome compaction best-explained gains of introns in ray-finned fishes.

The C1IN gene is conserved on the similar locus from ray-finned fishes to human, except for frog and lampreys, dating ~450 MY (Fig. 3). C1IN is a group V4 member (Fig. 4A). Variations are observed in ray-finned fishes with only few marker genes are maintained (Fig. 3). These observations also suggest that this single serpin gene is originated by chromosomal duplications events at the beginning of vertebrate evolution. Some of these serpins remained single gene in the chromosomal fragments such as angiotensinogen [30], antithrombin III [31] and heparin cofactor II [32], whereas other serpins underwent rapid tandem duplications that originated several serpin paralogs on the same locus, such as members present in clade A and clade B.

In conclusion, C1IN gene is revisited from sequence-structural, phylogenetic and variants perspective in the post-genomics era. There is fusion of two Ig domains to serpin core domain of C1

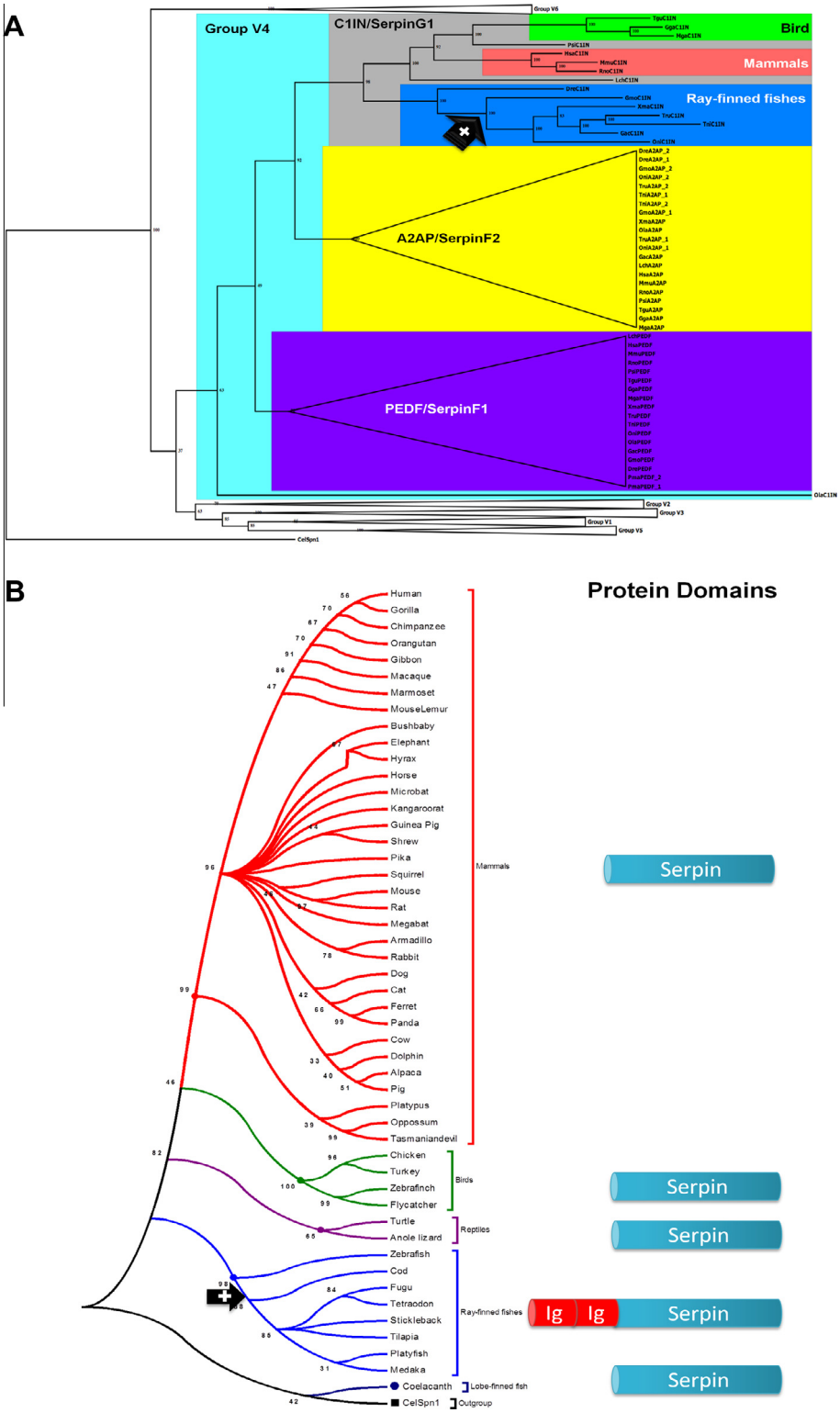

Fig. 4. Evolutionary history of vertebrate serpins and C1 inhibitors. (A) Bayesian phylogenetic history of vertebrate serpins illustrates vertebrate group-wise (V1-V6) distribution with three members in group V4 as C1IN (serpinG1), α_2-antiplasmin (serpinF1) and PEDF (serpinF1). (B) Maximum-Likelihood-based phylogenetic tree depicts species-wise distribution of C1IN and domain architecture. Location of intron gain is shown as black arrow with "+" sign. CelSpn1 from *Caenorhabditis elegans* (Genbank accession id – NP_503315) served as outgroup.

Table 4
Ig domain based intron insertion has occurred in ray-finned fishes with genome size below 1 Gb, marked in bold letters.

Organisms	Genome size
Human (*Homo sapiens*)	3.50
Mouse (*Mus musculus*)	3.25
African elephant (*Loxodonta africana*)	4.4
Platypus (*Ornithorhynchus anatinus*)	3.06
Chicken (*Gallus gallus*)	1.25
Flycatcher (*Myiarchus crinitus*)	1.32
Turkey (*Meleagris gallopavo*)	1.4
Zebra Finch (*Taeniopygia guttata*)	1.25
Frog (*Xenopus tropicalis*)	3.69
Girdled lizard (*Cordylus tropidosternum*)	3.93
Asian leaf turtle (*Cyclemys dentata*)	3.05
Blood python (*Python curtus*)	1.83
Fugu (*Takifugu rubripes*)	**0.40**
Tetraodon (*Tetraodon nigroviridis*)	**0.35**
Stickleback (*Gasterosteus aculeatus*)	**0.70**
Medaka (*Oryzias latipes*)	**0.75**
Atlantic cod (*Gadus morhua*)	**0.93**
Rock beam (*Oplegnathus fasciatus*)	**0.93**
Platyfish (*Xiphophorus maculatus*)	**0.92**
Tilapia (*Oreochromis niloticus*)	**0.95**
Coelacanth (*Latimeria chalumnae*)	2.8
Cave fish (*Astyanax mexicanus*)	1.0
Spotted Gar (*Lepisosteus oculatus*)	1.4
Zebrafish (*Drosophila*)	1.78
Sea lamprey (*Petromyzon marinus*)	2.44
European river lamprey (*Lampetra fluviatilis*)	1.45

inhibitor in fishes. Gene structures of C1IN gene are variable in genome compacted fishes with intron gained in the Ig domains but no change of gene structure is observed in the serpin core *per se*.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2014.05.097.

References

[1] A.E. Davis, S. Cai, D. Liu, C1 inhibitor: biologic activities that are independent of protease inhibition, Immunobiology 212 (2007) 313–323.
[2] A.E. Davis, F. Lu, P. Mejia, C1 inhibitor, a multi-functional serine protease inhibitor, Thromb. Haemost. 104 (2010) 886–893.
[3] G.A. Silverman, P.I. Bird, R.W. Carrell, et al., The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature, J. Biol. Chem. 276 (2001) 33293–33296.
[4] A. Kumar, H. Ragg, Ancestry and evolution of a secretory pathway serpin, BMC Evol. Biol. 8 (2008) 250.
[5] H. Ragg, A. Kumar, K. Köster, et al., Multiple gains of spliceosomal introns in a superfamily of vertebrate protease inhibitor genes, BMC Evol. Biol. 9 (2009) 208.
[6] O.D. Ratnoff, I.H. Lepow, Some properties of an esterase derived from preparations of the first component of complement, J. Exp. Med. 106 (1957) 327–343.
[7] P. Flicek, I. Ahmed, M.R. Amode, et al., Ensembl 2013, Nucleic Acids Res. 41 (2013) D48–D55.
[8] M. Stanke, B. Morgenstern, AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints, Nucleic Acids Res. 33 (2005) W465–W467.
[9] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, BMC Bioinformatics 5 (2004) 113.
[10] K.B. Nicholas, Nicholas H.B. Jr., D.W.I. and Deerfield, GeneDoc: Analysis and Visualization of Genetic Variation, EMBNEW.NEWS 4 (1997) 14.
[11] G.E. Crooks, G. Hon, J.M. Chandonia, et al., WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190.
[12] F. Ronquist, M. Teslenko, P. van der Mark, et al., MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space, Syst. Biol. 61 (2012) 539–542.
[13] K. Tamura, D. Peterson, N. Peterson, et al., MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol. Biol. Evol. 28 (2011) 2731–2739.
[14] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, Nat. Protoc. 5 (2010) 725–738.
[15] E. Krieger, G. Koraimann, G. Vriend, Increasing the precision of comparative models with YASARA NOVA – a self-parameterizing force field, Proteins 47 (2002) 393–402.
[16] T.R. Gregory, Animal Genome Size Database. http://www.genomesize.com., 2014.
[17] T. Brümmendorf, F.G. Rathjen, Cell adhesion molecules 1: immunoglobulin superfamily, Protein Profile 2 (1995) 963–1108.
[18] P. Bork, L. Holm, C. Sander, The immunoglobulin fold. Structural classification, sequence patterns and common core, J. Mol. Biol. 242 (1994) 309–320.
[19] A.F. Williams, A.N. Barclay, The immunoglobulin superfamily – domains for cell surface recognition, Annu. Rev. Immunol. 6 (1988) 381–405.
[20] J.A. Irving, R.N. Pike, A.M. Lesk, et al., Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function, Genome Res. 10 (2000) 1845–1864.
[21] S. Inoue, B.H. Nam, I. Hirono, et al., A survey of expressed genes in Japanese flounder (*Paralichthys olivaceus*) liver and spleen, Mol. Mar. Biol. Biotechnol. 6 (1997) 376–380.
[22] T. Wang, C.J. Secombes, Complete sequencing and expression of three complement components, C1r, C4 and C1 inhibitor, of the classical activation pathway of the complement system in rainbow trout *Oncorhynchus mykiss*, Immunogenetics 55 (2003) 615–628.
[23] N. Umasuthan, S.D.N.K. Bathige, K.S. Revathy, et al., A C1 inhibitor ortholog from rock bream (*Oplegnathus fasciatus*): molecular perspectives of a central regulator in terms of its genomic arrangement, transcriptional profiles and anti-protease activities of recombinant peptide, Dev. Comp. Immunol. 42 (2014) 197–210.
[24] A. Kumar, Phylogenomics of vertebrate serpins, Technical Faculty, University of Bielefeld, Bielefeld, 2010, pp. 2016.
[25] C.L. Will, R. Lührmann, Spliceosome structure and function, Cold Spring Harb. Perspect. Biol. 3 (2011) 1–23.
[26] S.W. Roy, W. Gilbert, The evolution of spliceosomal introns: patterns, puzzles and progress, Nat. Rev. Genet. 7 (2006) 211–221.
[27] P. Yenerall, B. Krupa, L. Zhou, Mechanisms of intron gain and loss in Drosophila, BMC Evol. Biol. 11 (2011) 364.
[28] P. Yenerall, L. Zhou, Identifying the mechanisms of intron gain: progress and trends, Biol. Direct 7 (2012) 29.
[29] A. Kumar, A. Bhandari, R. Sinha, et al., Spliceosomal intron insertions in genome compacted ray-finned fishes as evident from phylogeny of MC receptors, also supported by a few other GPCRs, PLoS One 6 (2011) e22046.
[30] A. Kumar, S.J. Sarde, A. Bhandari, Revising angiotensinogen from phylogenetic and genetic variants perspectives, Biochem. Biophys. Res. Commun. 446 (2014) 504–518.
[31] A. Kumar, A. Bhandari, S.J. Sarde, et al., Sequence, phylogenetic and variant analyses of antithrombin III, Biochem. Biophys. Res. Commun. 440 (2013) 714–724.
[32] A. Kumar, A. Bhandari, S.J. Sarde, et al., Genetic variants and evolutionary analyses of heparin cofactor II, Immunobiology in press, 2014.