

# Bayesian phylogeny analysis of vertebrate serpins illustrates evolutionary conservation of the intron and indels based six groups classification system from lampreys for ~500 MY

Abhishek Kumar

Department of Genetics & Molecular Biology in Botany, Institute of Botany,  
Christian-Albrechts-University at Kiel, Kiel, Germany  
Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ),  
Heidelberg, Germany

## ABSTRACT

The serpin superfamily is characterized by proteins that fold into a conserved tertiary structure and exploits a sophisticated and irreversible suicide-mechanism of inhibition. Vertebrate serpins are classified into six groups (V1–V6), based on three independent biological features—genomic organization, diagnostic amino acid sites and rare indels. However, this classification system was based on the limited number of mammalian genomes available. In this study, several non-mammalian genomes are used to validate this classification system using the powerful Bayesian phylogenetic method. This method supports the intron and indel based vertebrate classification and proves that serpins have been maintained from lampreys to humans for about 500 MY. Lampreys have fewer than 10 serpins, which expand into 36 serpins in humans. The two expanding groups V1 and V2 have SERPINB1/SERPINB6 and SERPINA8/SERPIND1 as the ancestral serpins, respectively. Large clusters of serpins are formed by local duplications of these serpins in tetrapod genomes. Interestingly, the ancestral HCII/SERPIND1 locus (nested within PIK4CA) possesses group V4 serpin (A2APL1, homolog of  $\alpha_2$ -AP/SERPINF2) of lampreys; hence, pointing to the fact that group V4 might have originated from group V2. Additionally in this study, details of the phylogenetic history and genomic characteristics of vertebrate serpins are revisited.

Submitted 22 March 2015

Accepted 26 May 2015

Published 16 June 2015

Corresponding author

Abhishek Kumar, abhishek.abhishekkumar@gmail.com

Academic editor

María Ángeles Esteban

Additional Information and  
Declarations can be found on  
page 17

DOI 10.7717/peerj.1026

© Copyright  
2015 Kumar

Distributed under  
Creative Commons CC-BY 4.0

**OPEN ACCESS**

**Subjects** Aquaculture, Fisheries and Fish Science, Biochemistry, Evolutionary Studies, Genetics, Genomics

**Keywords** Vertebrates, Serpins, Bayesian phylogeny, Gene structure, Gene duplication, Intron-exon

## INTRODUCTION

Serine proteinase inhibitors (serpins) are one of the major regulators of cellular proteolysis. The superfamily of serpins is involved in an array of fundamental biological processes such as blood coagulation, cell differentiation, cell migration, complement activation, embryo implantation, fibrinolysis, angiogenesis, and inflammation, and tumor suppression

(Silverman *et al.*, 2001). Serpins usually have a single domain (Pfam ID PF00079 or Interpro ID IPR023795) with a conserved core of ~350–400 residues. They often possess N- or C-terminal extensions and an overall molecular mass of ~40–60 kDa. N- and/or O-glycosylations are frequently observed in extracellular serpins (Gettins, 2002; Gettins, Patston & Olson, 1996). The conserved three-dimensional structure of serpins is composed of three  $\beta$ -sheets ( $\beta A$ – $\beta C$ ) and 8–9  $\alpha$ -helices ( $\alpha A$ – $\alpha I$ ). The hallmark of the serpin inhibitory mechanism is a large-scale conformational change involving the reactive center loop (RCL). The RCL is an exposed flexible loop of about 17–20 residues, which interacts with a target protease (Silverman *et al.*, 2001). RCL acts as a bait imitating the protease substrate, and is cleaved between the positions P1 and P1' (Silverman *et al.*, 2001).

In metazoans, serpins have undergone divergent evolution over a period of about 650–700 million years (Kumar & Ragg, 2008). A number of phylogenetic studies have been undertaken using sequence analysis of the serpins. Early investigations suggested the establishment of this multigene family through inter- and intra-chromosomal gene duplications. Several gene clusters have arisen, encoding functionally diverse serpin proteins. In metazoans, serpins display highly variant exon-intron patterns are strongly conserved within some taxa. Gene architecture and other rare genetic characters singularize a robust basis for classifying vertebrate serpins. Based on number, positions, and phases of introns, serpins have been classified into six groups (V1–V6). Vertebrate serpin genes with equivalent gene structures often tend to be organized in clusters (Benarafa & Remold-O'Donnell, 2005). However, close physical linkage cannot always be established. Interestingly, none of the 24-intron positions that have been mapped to the core domain of vertebrate serpins is shared by all of the six gene groups. Nevertheless, characteristic amino acid indels provide further cues to unravel the phylogenetic relationship (Ragg *et al.*, 2001). Previous analyses were performed using limited vertebrate serpins, and mainly focused on human and mouse data. Currently, several non-mammalian vertebrate genomes are known. Hence, combining these genomes for validation of intron-encoded vertebrate serpin classification is possible. In addition, different phylogenetic methods have been applied to vertebrate serpin classification such as maximum-likelihood (ML) and Neighbor-joining (NJ) (Atchley *et al.*, 2001). In the last decade, Bayesian Markov chain Monte Carlo (MCMC) has been enthusiastically corroborated as the state-of-the-art method for phylogenetic reconstruction and was largely driven by the rapid and widespread adoption of MrBayes suite (Ronquist & Huelsenbeck, 2003). This method was recently tested for urochordate serpin classification (Kumar & Bhandari, 2014). Heretofore, there is no report on the use of this phylogenetic method for large-scale analysis of vertebrate serpins. Herein, Bayesian method was employed along with several non-mammalian genomes for reconstructing vertebrate serpin classification system. This study reveals that Bayesian phylogenetic method supports the intron-coded vertebrate serpin classification system.

Moreover, this classification system is conserved from lampreys to human with a few new introns being created in groups V2, V4 and V6 in the serpin core domains of selected ray-finned fishes. Furthermore, different properties of these six serpin groups were also summarized.

**Table 1** Vertebrate genomes used during this study.

Genome	Major database used	References
<i>Homo sapiens</i>	<a href="http://www.ncbi.nlm.nih.gov/genome/guide/human/">http://www.ncbi.nlm.nih.gov/genome/guide/human/</a>	(Venter et al., 2001)
<i>Mus musculus</i>	<a href="http://www.ncbi.nlm.nih.gov/genome/guide/mouse/">http://www.ncbi.nlm.nih.gov/genome/guide/mouse/</a>	(Mouse Genome Sequencing Consortium et al., 2002)
<i>Rattus norvegicus</i>	<a href="http://www.ncbi.nlm.nih.gov/genome/guide/rat/">http://www.ncbi.nlm.nih.gov/genome/guide/rat/</a>	(Gibbs et al., 2004)
<i>Gallus gallus</i>	<a href="http://www.ncbi.nlm.nih.gov/genome/guide/chicken/">http://www.ncbi.nlm.nih.gov/genome/guide/chicken/</a>	(Hillier et al., 2004)
<i>Xenopus tropicalis</i>	<a href="http://genome.jgi-psf.org/Xentr4/Xentr4.home.html">http://genome.jgi-psf.org/Xentr4/Xentr4.home.html</a>	(Hellsten et al., 2010)
<i>Fugu rubripes</i>	<a href="http://genome.jgi-psf.org/Takru4/Takru4.home.html">http://genome.jgi-psf.org/Takru4/Takru4.home.html</a>	(Aparicio et al., 2002)
<i>Tetraodon nigroviridis</i>	<a href="http://www.genoscope.cns.fr/externe/tetranew/">http://www.genoscope.cns.fr/externe/tetranew/</a>	(Jaillon et al., 2004)
<i>Danio rerio</i>	<a href="http://www.ensembl.org/Danio_rerio/index.html">http://www.ensembl.org/Danio_rerio/index.html</a>	(Birney et al., 2006)
<i>Petromyzon marinus</i>	<a href="http://www.ensembl.org/Petromyzon_marinus/Info/Index">http://www.ensembl.org/Petromyzon_marinus/Info/Index</a>	(Smith et al., 2013)

## MATERIALS AND METHODS

### Collection of serpins from selected genomes

Serpins sequences of selected vertebrates (Table 1) were obtained from the Ensembl database (release 72, June 2013) using the BLAST suite (Altschul et al., 1997). Details of identified serpins and comprehensive alignments are provided (Kumar, 2010).

### Gene structure prediction of serpins

To ensure accuracy, gene structure prediction within the Ensembl (Flicek et al., 2013) was taken and combined with predictions of AUGUSTUS gene prediction tool (Stanke & Morgenstern, 2005). Mature human  $\alpha_1$ -antitrypsin was used as the standard sequence for intron position mapping and numbering of intron positions, followed by suffixes a–c for their locations as reported previously (Kumar & Ragg, 2008).

### Protein sequence alignment

Protein alignments of different vertebrate serpins were created by MUSCLE tool (Edgar, 2004) using default parameters.

### Selection of substitution model

Upon evaluation of different amino acid substitution models for this alignment dataset using MEGA5 software suite (Tamura et al., 2011), it turned out that the WAG + G model was the best fit (Table S1).

### Bayesian phylogenetic analysis

To infer the evolutionary history of serpins, the Bayesian phylogenetic tree was constructed using the MrBayes 3.2 suite (Ronquist & Huelsenbeck, 2003) with the following parameters: 5 generations, until average standard deviation of split frequencies was lower than 0.0098, 25% burn-in-period, WAG + G matrix-based model. Nve\_Spn1 from starlet sea anemone (*Nematostella vectensis*) was used as the outgroup.

## Estimation of genome size

Selected vertebrate genome sizes were calculated using the animal genome size database (Gregory, 2014).

## RESULTS AND DISCUSSION

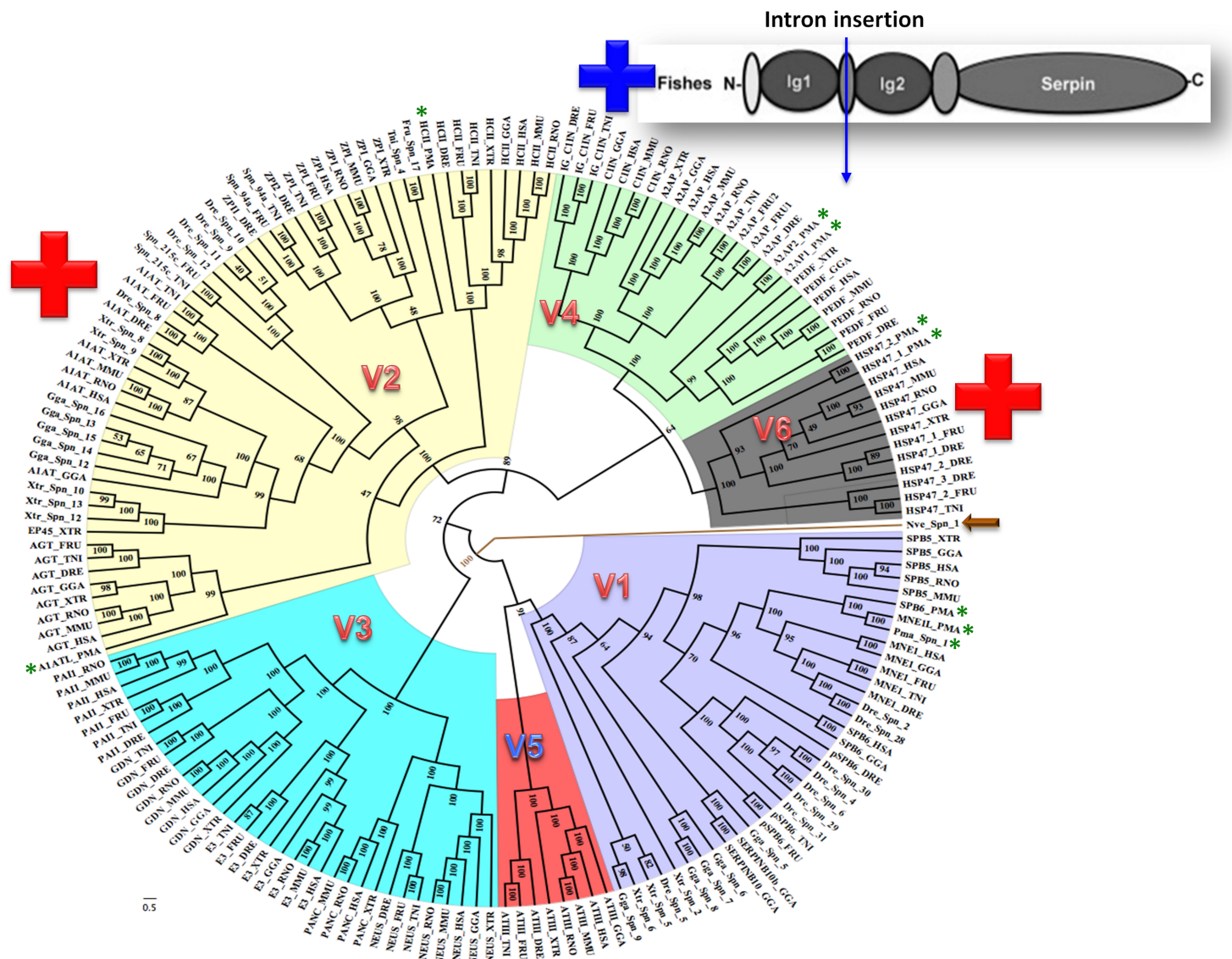
### Bayesian phylogeny classifies serpins into six groups V1–V6

Bayesian phylogenetic analysis reveals six groups of vertebrate serpins as depicted in different colors (Fig. 1). Posterior probability values are marked, with the lowest being 47, because several paralogs of group V2 serpins are known in tetrapod genomes (Fig. 1). Sea anemone serpin (Nve\_Spn1) is the out-group for this phylogenetic analysis (marked by a brown arrow). This clustering matches with intron-indel based vertebrate serpins classification system of six groups (V1–V6) as illustrated in Fig. 2. Lampreys have only eight serpins as evident from BLAST analysis against sea lamprey (*Petromyzon marinus*) genome and cDNA of European river lamprey, *Lampetra fluviatilis* in the Genbank (Table 2). These serpins of lampreys are only distributed into four groups (marked by green stars in Fig. 2). I will describe and discuss each of these groups in next sections.

### Group V1 has several clade B members expanded from fishes to human

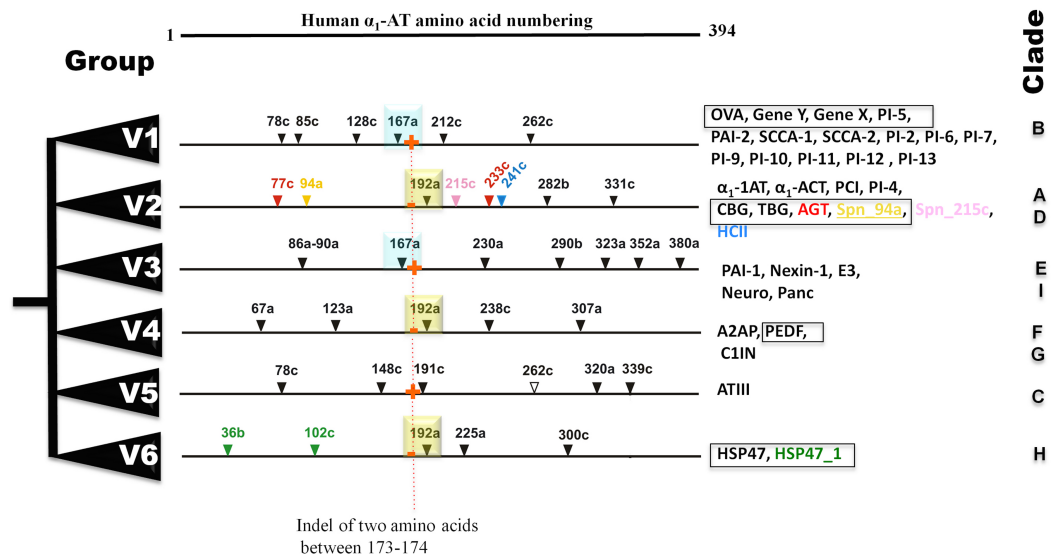
Group V1 serpins has been defined by a gene structure depicting five introns at positions—78c, 128c, 167a, 212c, and 262c, in their coding region (Fig. 2). An additional intron at the position 85c is found in some group V1 members and the presence and absence of this intron, defines sub-groups V1a and V1b, respectively. Group V1 is multi-membered, consisting of ovalbumin-like serpins that are involved in different physiological roles and often called as ov-serpins (Benarafa & Remold-O'Donnell, 2005). These serpins belong to clade B under clade-based classification system of serpins (Silverman et al., 2001). They are usually inhibitors of serine or cysteine proteases (cross-class inhibition), but some of them are non-inhibitory members (e.g., maspin/SERPINB5). They are mostly intracellular since they lack N-terminal signal peptide with few exceptions (Benarafa & Remold-O'Donnell, 2005; Izuhara et al., 2008; Kaiserman & Bird, 2005). They are also deprived of the C-terminal extensions (Benarafa & Remold-O'Donnell, 2005; Izuhara et al., 2008; Kaiserman & Bird, 2005). These serpins are localized in two clusters in the human genome. Human chromosome 6p25 region harbors three genes—SERPINB1, SERPINB6, and SERPINB9 as the first cluster while, the remaining genes namely, SERPINB2, SERPINB3, SERPINB4, SERPINB5, SERPINB7, SERPINB8, SERPINB10, SERPINB11, SERPINB12, and SERPINB13 are located in the 18q21 region (Benarafa & Remold-O'Donnell, 2005; Izuhara et al., 2008; Kaiserman & Bird, 2005).

This cluster originated by duplication of SERPINB1/MNEI1-like gene (Benarafa & Remold-O'Donnell, 2005). In contrast, the chicken has only one cluster on the chromosome 2q. Therefore, it is corroborated that there was a split after mammal/bird divergence at about 310 MY (Benarafa & Remold-O'Donnell, 2005; Izuhara et al., 2008; Kaiserman & Bird, 2005). Similar genomic organization of group V1 was also found in frogs and fishes.



**Figure 1** Bayesian phylogenetic history of vertebrate serpins reveals that exon-intron and rare indel based classification system is retained over period of 500 MY with conserved patterns from early diverging lampreys. Novel introns are inserted in groups V2 and V6 serpins in core domains are marked by red + while introns inserted in additional Ig domains of fish-specific C1 inhibitors is shown with a blue +. A sea anemone serpin (Nve\_Spn1) is the out-group for this phylogenetic analysis as marked by an arrow. Lamprey serpins are marked by green stars. HSP47 has two isoforms in lamprey named as HSP47\_1.PMA and HSP47\_2.PMA. DRE, Danio rerio; H SA, Homo sapiens; GGA, Gallus gallus; MMU, Mus musculus; PMA, *Petromyzon marinus*; RNO, *Rattus norvegicus*; TRU, *Takifugu rubripes*; TNI, *Tetraodon nigroviridis*; XTR, *Xenopus tropicalis*. *p*-paralog of a gene.

In addition to this syntenic organization, fishes possess some paralogous clusters of group V1 serpins while an additional cluster containing two serpins adjacent to the conserved orthologous cluster is found in frogs. The serpins SERPINB1/SERPINB6 of group V1 are probably the ancestors of all group V1 serpins, as these genes are found in lampreys (Table 2) and other fishes and are also conserved across other vertebrate taxa (Kumar, 2010). The group V1 serpins may be classified into sub-groups V1a and V1b, since these



**Figure 2** Summary of six groups (V1–V6) classification system of vertebrate serpins, based on introns and rare indels. Conserved intron positions are shown in cyan and yellow boxes for positions 167a and 192a, respectively. Fish-specific introns inserted in selected serpins are illustrated in different colors. Non-inhibitory serpins are shown in square boxes. Presence and absence of sequence indel of two amino acid between positions 173–174 are marked in by red + and – signs. OVA, Ovalbumin; Gene Y, Chicken gene Y protein; Gene X, Chicken gene X protein; PAI, Plasminogen activator inhibitor; SCCA, Squamous cell carcinoma antigen;  $\alpha_1$ -AT,  $\alpha_1$ -antitrypsin;  $\alpha_1$ -ACT,  $\alpha_1$ -antichymotrypsin; CBG, -Corticosteroid-binding globulin; TBG, Thyroxine-binding globulin; HCII, Heparin cofactor II; PCI, Protein C inhibitor; AGT, Angiotensinogen; E3, SerpinE3; Neuro, Neuroserpin; Panc, Pancreatin; A2AP,  $\alpha_2$ -Antiplasmin; PEDF, Pigment epithelium derived factor; C1IN, C1-Inhibitor; ATIII, Antithrombin III; HSP47, Heat shock protein 47kDa.

differ by one intron. It has been argued that a serpin gene of group V1b (7 exons) is the ancestor of group V1a (8 exons) that has emerged in birds after the divergence of frogs (Benarafa & Remold-O'Donnell, 2005; Izuhara et al., 2008; Kaiserman & Bird, 2005). The first argument coincides with the current data and corroborates that groups V1a serpins are derived from 7-exon genes (such as MNEI/SPB6). However, the argument that 8-exon genes first arose in chickens is not in agreement with the current data, since Xtr-Spn-5 in *X. tropicalis* and pSPB6 in *T. nigroviridis*, are group V1a members with the 8 exons/7 introns architecture (Kumar, 2010). Therefore, it is proposed that group V1b is ancestral to all group V1 serpins and group V1a is suggested to have arisen independently several times in different vertebrates from fishes to mammals. The ancestor of group V1 serpins appears to have been generated during the emergence of vertebrates. The oldest group V1 serpins are SPB1/SPB6 orthologs, which are present in lamprey (Table 2). A recent study claimed an ancestor of serpinB6 to be present in urochordates (Benarafa & Remold-O'Donnell, 2005). However, another study depicted six different groups of urochordate serpins, based on intron-encoded classification system, which markedly differs from vertebrates six groups (Kumar & Bhandari, 2014).

**Table 2** Summary of serpins in two lampreys namely, sea lamprey (*Petromyzon marinus*) and European river lamprey (*Lampetra fluviatilis*).

Name given	Ensembl Accession id	Serpin name	Group	Clade	RCL with P1–P1'	Ortholog in <i>L. fluviatilis</i> Genbank ID
Pma-Spn-1	<a href="#">ENSPMAG00000001963</a>	SERPINB1-like	V1	B	GTEAAAATAAIVMMR–CARMG	
MNE1L_PMA/Pma-Spn-2	<a href="#">ENSPMAG00000009027</a>	MNEI1-like/SERPINB1-like	V1	B	GTEAAAATAVTKLR–CAMPT	
SPB6_PMA (Pma-Spn-3)	<a href="#">ENSPMAG00000009040</a>	SPB6/SERPINB6-like	V1	B	GTEAAAATAISVLM–CAMPT	
A1ATL_PMA (Pma-Spn-4)	<a href="#">ENSPMAG00000006108</a>	A1AT-like, angiotensinogen, SERPINA8	V2	A	GTEAKAETVVGIMPI–SMPPT	<a href="#">CAV16871.1/CAV29466.1</a>
A1ATL_PMA (Pma-Spn-5)	<a href="#">ENSPMAG00000008131</a>	Heparin cofactor II/SERPIND1	V2	D	GSEAAAVTTVGFPL–TSHNR	<a href="#">CAX18777.1/AIA57696.1</a>
A2APL1_PMA (Pma-Spn-6)	<a href="#">ENSPMAG00000008124</a>	Alpha-2-antiplasmin-like 1	V4	F	GVKATAATGIMISLM–SVQHS	<a href="#">CAX18777.1/CAX18778.1</a>
A2APL2_PMA (Pma-Spn-7)	<a href="#">ENSPMAG00000002992</a>	Alpha-2-antiplasmin-like 2, A2APL2_PMA	V4	F	GAEAAAVTGVFLSRT–NPIYP	<a href="#">AIE16052.1/AIE16053.1</a>
HSP47_PMA (Pma-Spn-8)	<a href="#">ENSPMAG00000007485</a>	HSP47/SERPINH1	V6	H	GEEYDMSVHGHPDM–RNPHL	

## Group V2 possesses $\alpha_1$ -antitrypsin-like serpins, angiotensinogen (clade A) and heparin cofactor II (clade D)

Group V2 serpins are characterized by three introns at homologous positions—192a, 282b, and 331c (Fig. 2) in their coding regions, and most of the members have an intron mapping to the untranslated regions. This group is multi-membered, composed of  $\alpha_1$ -antitrypsin-like serpins that are involved in different physiological roles, including inhibitors, like  $\alpha_1$ -antitrypsin or antichymotrypsin as well as non-inhibitory members, like angiotensinogen (AGT/SERPINA8 [Kumar, Sarde & Bhandari, 2014](#)).

Gene structures of heparin cofactor II (HCII/SERPIND1) are variable in fishes with a novel intron gain at the position 241c, but this gene is nested in the large intron of phosphatidylinositol 4-Kinase (PIK4CA) gene for  $\sim 500$  MYA ([Kumar et al., 2014a](#)). Human HCII/SERPIND1 has 985 germline variants, identified from 1,092 human genomes. This includes 37 statistically deleterious missense variants ([Kumar et al., 2014a](#)).

Recently, it was reported that the gene structures of AGT from selected ray-finned fishes varied in exons I and II, with insertions of two novel introns in the core domain for ray-finned fishes at positions 77c and 233c, respectively ([Kumar, Sarde & Bhandari, 2014](#)). It was also reported that the AGT loci is conserved from lampreys to human and was estimated to be older than 500 MY ([Kumar, Sarde & Bhandari, 2014](#)). Interestingly, the RCL of AGT protein is inhibitory in lampreys and evolved to become non-inhibitory in human over a period of 500 MY ([Kumar, Sarde & Bhandari, 2014](#)). Additionally, 690 AGT variants were detected by analyzing 1,092 human genomes with the top three variation classes belonging to single nucleotide polymorphisms (SNPs, 89.7%), somatic SNVs (5.2%) and deletion (2.9%) ([Kumar, Sarde & Bhandari, 2014](#)). Furthermore, 121 missense variants of AGT (including 32 statistically deleterious variants) were deciphered ([Kumar, Sarde & Bhandari, 2014](#)).

From fishes to humans, group V2 comprises of multiple paralogs of  $\alpha_1$ -antitrypsin-like genes. Genuine orthologs of AGT/SERPINA8 and HCII SERPIND1 were identified from lampreys to humans, using synteny and signature sequences. Concerning the other genes of group V2; since in most tetrapod genomes,  $\alpha_1$ -antitrypsin-like gene clusters are derived from recent duplication events, which results in proteins with high sequence similarities, one-to-one orthology allocation proved to be difficult. This poses notorious challenges in detection of orthologs within this cluster and often leads into problems in generating phylogenetic trees (Fig. 1).

In the cluster of  $\alpha_1$ -antitrypsin-like genes, the protein Z-dependent protease inhibitor (ZPI/SERPINA10) is localized at the end of the cluster with other conserved marker genes ([Kumar, 2010](#)). This assisted in the detection of ZPI/SERPINA10 orthologs using synteny analysis. Recently published report on the serpins of channel catfish, *Ictalurus punctatus* ([Li et al., 2015](#)) also supported this finding.

Additionally, two group V2 serpins are found only in ray-finned fishes. The first serpin was detected with a novel intron at position 94a and hence it is named as the Spn\_94a gene fishes, which have conserved in the same genomic organization in these fishes. This corroborates that fish specific ortholog Spn\_94a shows sequence similarity with



ZPI/SERPINA10 gene and hence it is paralog of ZPI/SERPINA10. Similarly, *Fugu* and *T. nigroviridis* possess the second fish-specific group V2 gene with an additional intron at the position 215c (Spn\_215c), which indicates that they are orthologs (Kumar, 2010). The origin of these genes, however, is unclear.

No orthologs of the hormone binding serpins (corticosteroid-binding globulin [CBG/SERPINA6] and thyroxine-binding globulin [TBG/SERPINA6]) were detected in non-mammalian vertebrates. In short, the conserved set of group V2 comprises only orthologs of AGT/SERPINA8 (Kumar, Sarde & Bhandari, 2014) and HCII/SERPIND1 (Kumar et al., 2014a). In contrast, some fish-specific group V2 genes and the  $\alpha_1$ -antitrypsin-like genes are differentially expanded in vertebrates, particularly in mammalian lineages, such as rodents (Forsyth, Horvath & Coughlin, 2003) and cattle (Pelissier et al., 2008). The expansion of group V2 members should be further explored by analyzing marsupials and Platypus, which branched out early in mammalian evolution. The presence of group V2 members in the lamprey genome suggests that this group originated during emergence of vertebrates (Table 2). Further investigation of group V2 members in the hagfish genome and several lamprey genomes will shed more light on this issue.

### **Group V3 is composed of 5 members, which belongs to two clades E and I**

Group V3 serpins have seven introns at the positions - 86a/88a/90a, 167a, 230a, 290b, 323a, 352a and 380a in their coding regions (Fig. 2). The exact location of the first intron is uncertain in different group V3 serpins, due to alignment ambiguities. Group V3 has five inhibitory serpins, which are involved in different physiological processes namely, plasminogen activator inhibitor 1 (PAI1/SERPINE1), glia derived nexin (GDN/SERPINE2/), SERPINE3, neuroserpin/SERPINI1, and pancpin/SERPINI2. PAI1/SERPINE1 is conserved in vertebrates, depicting 38–80% sequence identity and 59–95% sequence similarity at the amino acid level with human PAI1/SERPINE1. The inhibitory RCL region is conserved and consists of R–M at the P1–P1'. GDN/SERPINE2 is also highly conserved in vertebrates and it shows 51–84% sequence identity and 70–93% sequence similarity with human GDN/SERPINE2. The helix-D region is highly conserved among GDN/SERPINE2 orthologs of different vertebrates and an N-glycosylation site (positions 163–165) is also conserved. The inhibitory RCL region is also strongly conserved. SERPINE3 is maintained in vertebrates, show 27–64% sequence identity, and 37–74% sequence similarity on the amino acid level with human serpinE3. The inhibitory RCL of SERPINE3 is conserved with a cluster of hydrophobic amino acids preceding the presumptive P1 position.

The neuroserpin/SERPINI1 is highly conserved in vertebrates, and the protein shows 47–81% sequence identity and 65–95% sequence similarity with the human ortholog. The inhibitory RCL region of neuroserpin/SERPINI1 always contains an arginine at the position P1. An N-glycosylation signal at residues 163–165, and a C-terminal extension that has have been shown to direct neuroserpin to the regulated secretory pathway (Ishigami et al., 2007) are strongly conserved.

Discriminatory data at the genomic, gene and protein levels offered comprehensive insight into the phylogenetic history of neuroserpin/SERPINI1 (Kumar & Ragg, 2008).

Syntenic analysis proved to be very instrumental in this respect, demonstrating that rare genomic characters can provide very useful information for decoding of links between protein families with intricate evolutionary history. The strongly conserved syntenic association of PDCD10 and neuroserpin/SERPINI1 orthologs during diversification of deuterostomes is unraveled (Kumar & Ragg, 2008). These head-to-head oriented genes (Neuroserpin/SERPINI1 and PDCD10) have common bi-directional and asymmetrical promoter region inserted within the ~0.9 kb intergenic region (Chen et al., 2007). Requirements of common regulatory units could have driven the preservation of this linkage. In the era of next-generation genome sequencing, the rapidly accumulating genome sequences will certainly continue to provide further discriminatory markers, such as codon usage dichotomy (Krem & Di Cera, 2003), in order to facilitate robust classification of other metazoan serpins.

Some serpins have signals for sub-cellular localization at the C-terminal end and they have been involved in the secretory pathway. It is not just limited to vertebrate serpins and several examples exist in invertebrates such as in some urochordate serpins (Kumar & Bhandari, 2014). Using these signals, ancestral orthologs of neuroserpin/SERPINI1 were deciphered (Kumar & Ragg, 2008). A C-terminal KDEL-like motif deters the secretion of soluble endoplasmic reticulum (ER)—resident proteins (Lewis, Sweet & Pelham, 1990; Raykhel et al., 2007; Semenza et al., 1990). There are 24 possible variants of ER retention signals listed as a PROSITE motif—[KRHQSA]-[DENQ]-E-L in the PROSITE database (Sigrist et al., 2013). In addition, there are some ER retention signals that do not fit into the classical PROSITE motif (Raykhel et al., 2007). These ER retention signals are present across eukaryotic genomes such as in the BEM46 protein of *Neurospora crassa* (Kumar, Kollath-Leiss & Kempken, 2013). In early diverging deuterostomia, the neuroserpin orthologs in *Strongylocentrotus* (Spu-spn-1), lancelet (Bfl-Spn-1) and sea anemone (Nve-Spn-1) have HEEL, KDEL, and SDEL at their C-terminal ends, respectively. These are variants of the PROSITE motif for ER retention/retrieval signal. In contrast, the C-terminal end of tetrapod neuroserpin is HDFEEL. In HeLa cells that express three different KDEL receptors with overlapping, but differential passenger specificities, the “FEEL” sub-sequence targets attached passenger proteins primarily to the Golgi, though one-fourth of cells depict ER localization (Raykhel et al., 2007). However, in transfected COS cells, intracellular neuroserpin localizes either to the ER or to Golgi (Ishigami et al., 2007). Conversely, in cells with a regulated secretory pathway, neuroserpin/SERPINI1 resides in large dense core vesicles, which is assisted by a C-terminal extension encompassing the last 13 amino acids (ETMNTSGHDFEEL) including the FEEL sequence (Ishigami et al., 2007). Collectively, these data suggest that in the neuroserpin orthologs from deep-branching metazoans, a two amino acid insertion ‘FE’ in combination with additional residues constitutes a modified sorting signal, which attributes a specialized subcellular localization. Surveillance of the secretory pathway routes by serpins is an ancient and conserved trait in eukaryotes as indicated by the putative neuroserpin ortholog present in the sea anemone genome. It will be interesting to investigate experimentally,

whether the C-terminal extensions of neuroserpin orthologs from fishes are functional and mediate differential localization in a fashion similar to mammalian neuroserpin.

Due to variations in their RSL region, ER-localized serpins may work differently in the secretory pathway. Neuroserpin from vertebrates inhibits tissue-type plasminogen activator (tPA) *in vitro*, using the Arg residue at the P1 position in the RSL region (Osterwalder *et al.*, 1998). The cleavage site of Bfl-spn-1 is preceded by the dipeptide motif Lys-Arg (KR), a distinct feature for substrates and inhibitors of proprotein convertases (PCs). Similar features were reported for Bla-Spn-1 from *B. lanceolatum* as well (Bentele *et al.*, 2006). Since the serpins Bfl-spn-1 (*B. floridae*), Spu-spn-1 (sea urchin) and Nve-Spn-1 (sea anemone) also possess the Lys-Arg (KR) dipeptide motif. Thus, similar physiological role of these serpins can be expected. Status of the neuroserpin ortholog in the arthropod lineage is recently becoming clear. Examples of classical ER targeting signal (HDEL) possessing serpins were found in *D. melanogaster* as Spn4, which is a furin inhibitor (Oley, Letzel & Ragg, 2004; Osterwalder *et al.*, 2004; Richer *et al.*, 2004) and its homologous gene in *Anopheles gambiae* as Spn10 (Danielli, Kafatos & Loukeris, 2003). Recently, the crystal structure of fly Spn4 was determined and this serpin exhibits structural properties as of human neuroserpin/SERPINI1 (Ellisdon *et al.*, 2014), which provided first evidences of the orthologous nature of these serpins.

The pancpin/SERPINI2 gene is localized in close proximity to the neuroserpin gene. Pancpin also possesses a C-terminal extension and indels like neuroserpin, suggesting its close relatedness to these proteins. Pancpin/SERPINI2 orthologs are found only in mammals and in *Xenopus*, showing 49–76% sequence identity, and 68–88% sequence similarity at the amino acid level. The C-terminal end is strongly conserved (Kumar, 2010). Absence of pancpin/SERPINI2 in fishes hints that the pancpin gene may have originated by tandem duplication of neuroserpin after separation of tetrapods from the fish lineage.

In the human genome, the other group V3 members such as PAI1/SERPINE1 (chromosome 3), GDN/SERPINE2 (chromosome 2) and SERPINE3 (chromosome 13) are present at various genomic locations. This suggests that they originated at independent loci in the vertebrates.

### **Group V4 has three serpins—two in the clade F and one in the clade G; surprisingly, fishes have C1IN with two immunoglobulin domains**

Group V4 of vertebrate serpins have a gene structure consisting a conserved set of five introns at positions 67a, 123a, 192a, 238c, and 307a in the coding regions (Fig. 2). In mammals, group V4 serpins consists of three genes—pigment epithelium derived factor (PEDF/SERPINF1),  $\alpha_2$ -antiplasmin ( $\alpha_2$ -AP/SERPINF2) and C1 inhibitor (C1IN/SERPING1). Group V4 serpins are involved in very different physiological functions. PEDF is a non-inhibitory serpin that possesses neuroprotective and antiangiogenic functions (Sawant *et al.*, 2004; Steele *et al.*, 1993; Tombran-Tink, 2005).  $\alpha_2$ -antiplasmin is an inhibitor of plasmin and its fibrin bound form is a major regulator of blood clot lysis (Coughlin, 2005). C1 inhibitor (C1IN/SERPING1) is a multi-functional serpin, which operates by inactivating various serine proteases in different plasmatic cascades including

the complement (classical pathway—C1r and C1s; as well as lectin pathways—MASP1 and MASP2), contact (Factor XII and kallikrein), coagulation (Factor XI and thrombin) and fibrinolytic (tPA and plasmin) systems (Davis, Cai & Liu, 2007; Davis, Lu & Mejia, 2010). Fishes have C1IN/SERPING1 with two immunoglobulin-like domains attached at the N-terminal region (Fig. 4). The RCL regions of C1IN/SERPING1 have variations at the positions P2 and P1' from fishes and tetrapods. Gene structures of C1IN/SERPING1 from selected ray-finned fishes varied in the Ig domain region with the insertion of a novel intron splitting exon Im2 into Im2a and Im2b (Kumar et al., 2014b). Kumar & Bhandari (2014) depicted that C1IN/SERPING1 gene has remained on the same locus for ~450 MY in 52 vertebrates, but it is missing in frogs and lampreys (Kumar et al., 2014b).

Protein sequence analyses depicts that orthologs of PEDF/SERPINF1 and  $\alpha_2$ -AP (SERPINF2)-like genes are conserved throughout vertebrates (Kumar, 2010). However, on close scrutiny of group V4 serpins, orthologs of most human group V4 serpins other than A2AP1\_FRU in *Fugu* cannot be found in current genomic sequence versions of fish genomes (Kumar, 2010). Fishes have paralogs, probably due to fish-specific genome duplications and diversifications (Kumar, 2010). In addition, the sea lamprey genome has two members of group V4 were detected resembling  $\alpha_2$ -AP-like genes (A2APL1\_PMA and A2APL2\_PMA) with orthologs in the European sea lamprey (Table 1). This suggests that group V4 serpins existed since the beginning of vertebrates. Recently, it was shown that the A2APL1\_PMA gene is present in the nested state in the largest intron of PIK4CA gene along with HCII/SERPIND1 gene in the reverse orientation (Kumar et al., 2014a). However, only HCII/SERPIND1 gene is found as nested gene in PIK4CA in the ray-finned fishes to humans (Kumar et al., 2014a) and hence, it is postulated that ancestral group V4 gene were originated at adjacent to HCII/SERPIND1, which was mostly likely lost in the other lineages. This also corroborates that the origin of group V4 serpins is associated with group V2, as assumed from the conservation of basal intron at the position 192a (Fig. 2).

### Group V5 comprises only antithrombin III (ATIII) aka SERPINC1

Group V5 consists of a single member—antithrombin III (ATIII/SERPINC1). This gene encompasses seven exons and six introns with conserved intron positions (Fig. 2). In the human genome, the ATIII/SERPINC1 gene is located on chromosome 1q23–q25. ATIII/SERPINC1 is the major thrombin inhibitor in the blood coagulation cascade (Jordan, 1983), requires heparin for activation and has potent anti-angiogenic activity in certain conformations (Gettins, Patston & Olson, 1996).

The ATIII/SERPINC1 protein is highly conserved, and the sequence identity and similarity from fishes to mammals falls within the range of 50–87% and 67–97% respectively. ATIII/SERPINC1 has been maintained for over 450 MY on the same genomic loci in vertebrates with a few changes in ray-finned fishes. ATIII/SERPINC1 gene has lost an intron (262c) in tetrapods and in the lobed-finned fish coelacanth, *Latimeria chalumnae*. In addition, it has gained an intron at the position 262c in the ray-finned fishes, a characteristic feature, which is shared by group V1 members as well (Kumar et al., 2013a). ATIII/SERPINC1 comprises of several proteins motifs (Fig. 3), heparin binding

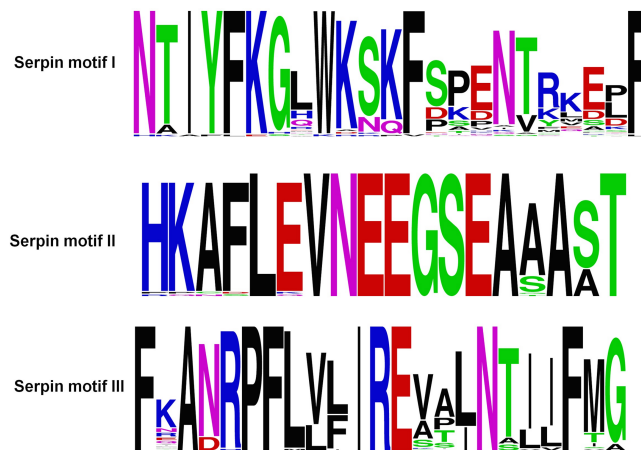


Figure 3 Serpin motifs of ATIII proteins.

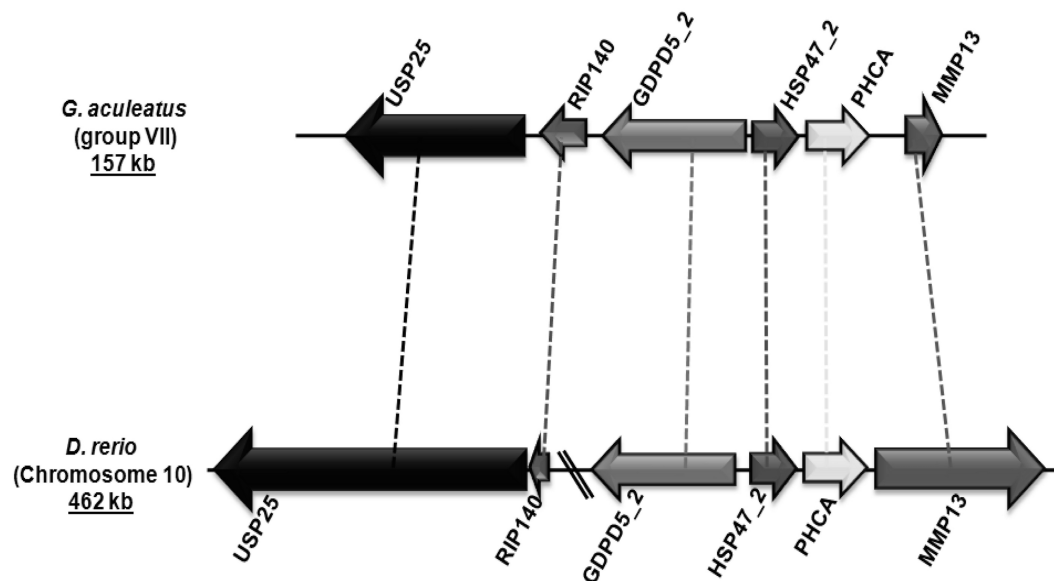


Figure 4 Genomic localization of fish-specific HSP47.2 gene.

basic residues, the hD helix, 3 pairs of Cys-Cys salt bridges, *N*-glycosylation sites, serpin motifs and inhibitory RCL (*Kumar et al., 2013a*). 1997 ATIII/SERPINC1 variants have been identified from 1,092 human genomes. These variants have been categorized into 76.2% SNPs, 11.8% deletions and 8.1% insertions (*Kumar et al., 2013a*).

### Group V6 is composed of HSP47 (SERPINH1) ortholog and fishes have 1–3 paralogs of HSP47

Group V6 is characterized by a gene structure depicting three introns at positions 192a, 225a and 300c in their coding regions (*Fig. 2*). This gene encodes for heat shock protein 47 kDa (HSP47/SERPINH1), which possesses a C-terminal endoplasmic reticulum (ER) retention signal (*Pelham, 1990*). HSP47/SERPINH1 is a non-inhibitory serpin that is

**Table 3** Sequence comparisons of HSP47 homologs in vertebrates. Percentage sequence identity (SI) and percentage sequence similarity (SS) values are shown as compared to HSP47\_HSA and A1AT\_HSA. Synteny based clustering divides group V6 genes into three sets: set I—true mammalian HSP47 orthologs (normal), set II—fish specific paralogs (bold) and set III (italics).

Human Serpins	Values (%)	HSP47_MMU	HSP47_RNO	HSP47_GGA	HSP47_XTR	HSP47_1_FRU	HSP47_2_FRU	HSP47_TNI	HSP47_1_DRE	HSP47_2_DRE	HSP47_3_DRE	HSP47_PMA
HSP47_HSA	SI	96	96	76	70	<b>63</b>	29	22	65	<b>64</b>	29	46
	SS	98	98	88	83	<b>82</b>	46	37	83	<b>82</b>	52	65
A1AT_HSA	SI	23	23	25	24	<b>24</b>	18	14	25	<b>24</b>	17	23
	SS	45	45	45	46	<b>44</b>	35	26	45	<b>46</b>	37	41

found in the ER of collagen producing cells where it is involved in the correct folding of procollagen triplet helices. Furthermore, it assists in the transport of procollagen from the ER to the Golgi complex (Hendershot & Bulleid, 2000; Lamande & Bateman, 1999; Nagata, 1996; Sauk, Nikitakis & Siavash, 2005). Tetrapods have a single copy of HSP47 genes, while fishes have up to three copies. The first HSP47/SERPINH1 is common to all vertebrates (HSP47\_1). The second copy is conserved in few ray-finned fishes, such as *G. aculeatus* and *D. rerio* with conserved syntenic organization (Fig. 4). The third one is only present in some fishes such as *D. rerio* (Table 3). HSP47/SERPINH1-like gene is conserved from lampreys to mammals, and this gene show 22–96% sequence identity and 37–98% sequence similarity with human HSP47/SERPINH1, respectively (Table 3). The HSP47\_TNI protein is highly diverged from standard HSP47/SERPINH1 protein as well as from all other serpin sequences (Table 3).

Orthology of the group V6 gene, lamprey HSP47\_PMA (normal font) cannot be decided on this basis. Group V6 comprises of the HSP47 gene and its paralogs in different vertebrates. Tetrapods have a single copy of the HSP47/SERPINH1 gene, while there are two or three HSP47-like genes in some fishes (Fig. 1).

### Status of thrombin inhibitors

Four human serpins inhibit thrombin, namely ATIII/SERPINC1, HCII/SERPIND1, protein C inhibitor (PCI/SERPINA5) and nexin I/SERPINE2 (Huntington, 2013; Huntington, 2014). These serpins exhibit higher rates of thrombin inhibition after binding to the glycosaminoglycan (GAG); however they have evolved radically different inhibition mechanisms (Huntington, 2013). Apart from these four, recent studies revealed that a fifth serpin, namely AGT/SERPINA8 gene also act as thrombin inhibitor,—at least in lampreys (Kumar, Sarde & Bhandari, 2014; Wang & Ragg, 2011; Wong & Takei, 2011). Lamprey AGT/SERPINA8 gene possesses inhibitory RCL and regulates thrombin along with HCII (Kumar, Sarde & Bhandari, 2014). Lampreys have no ATIII/SERPINC1 gene (Kumar et al., 2013a), but after the emergence of ATIII/SERPINC1 gene in tetrapods, AGT/SERPINA8 gene became non-inhibitory serpin (Kumar, Sarde & Bhandari, 2014). It is surprising that lampreys do not have the major thrombin inhibitor, ATIII/SERPINC1. However, lampreys also lack other immunologically critical genes such as recombination-activation genes which are essential for V(D)J recombination process that yields and assembles the variable regions of immunoglobulin and T-cell receptor genes in developing B- and T-lymphocytes

(Kumar et al., 2015). This suggests that lampreys may not need some of the very essential vertebrate genes and ATIII/SERPINC1 is in this category. This can be explained since AGT/SERPINA8 is the bi-functional serpin in lampreys, which acts as a thrombin inhibitor as well as a blood pressure regulator (Kumar, Sarde & Bhandari, 2014; Wang & Ragg, 2011; Wong & Takei, 2011).

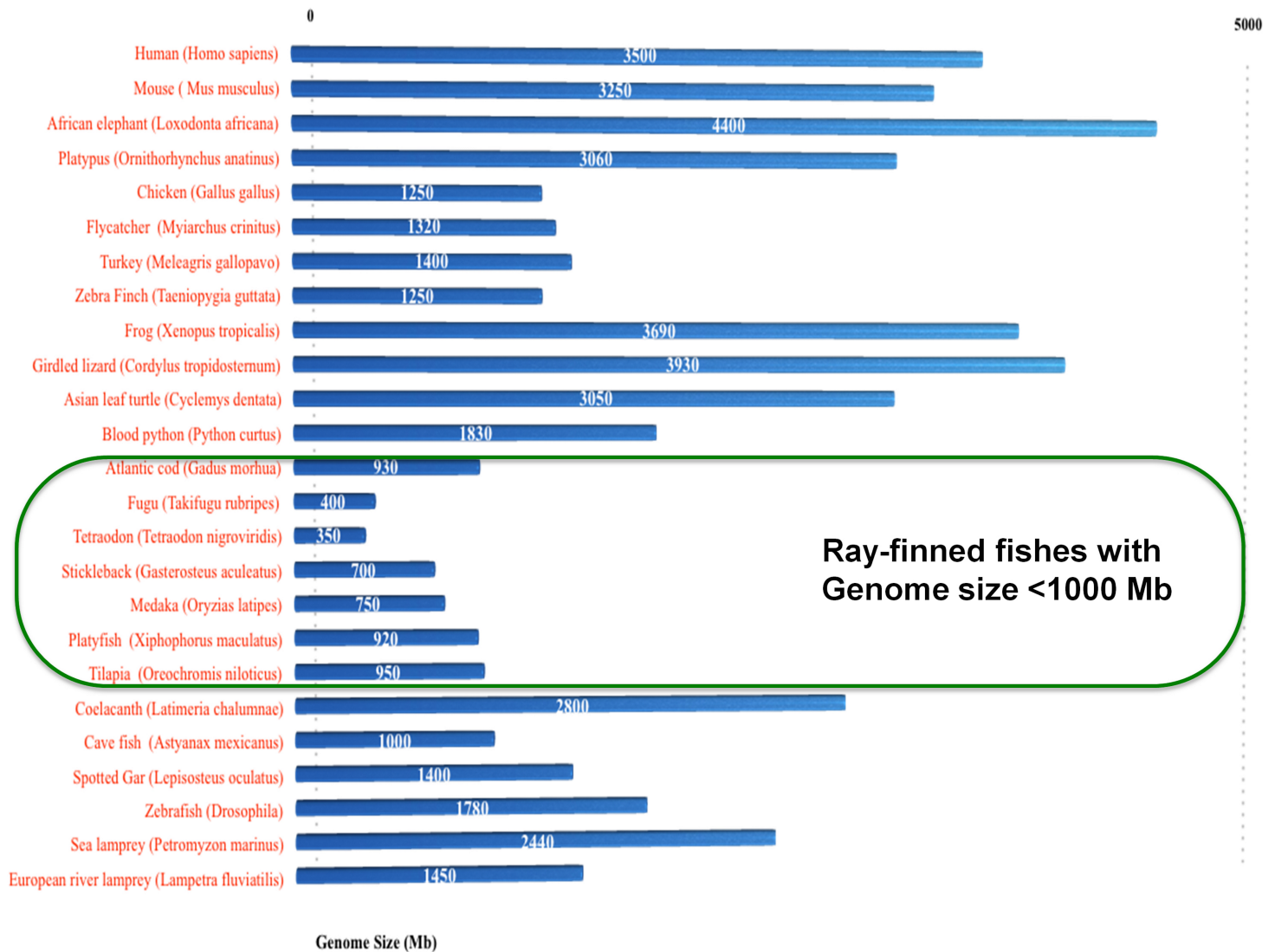
Major thrombin regulating thrombin serpins are ATIII/SERPINC1 (Kumar et al., 2013a) and HCII/SERPIND1 (Kumar et al., 2014a), which facilitates thrombin inhibition in two different locations such as in the vascular space and in the extravascular space, respectively.

### Special genomic characteristics of serpins

Various types of gene rearrangements characterize a typical evolution of genome. The evolutionary history of serpins is demarked by several such gene rearrangements. Several duplications were the results for expansions of groups V1 (Benarafa & Remold-O'Donnell, 2005) and V2 serpins (Forsyth, Horvath & Coughlin, 2003), from the basal loci of SERPINB1/SERPINB6 and SERPINA8/SERPIND1 in lampreys, respectively. In addition, several serpins in vertebrates are localized as single serpins by chromosomal duplication events such as AGT/SERPINA8 (Kumar, Sarde & Bhandari, 2014), ATIII/SERPINC1 (Kumar et al., 2013a) and HCII/SERPIND1 (Kumar et al., 2014a). HCII/SERPIND1 is conserved from lampreys to humans for ~500 MY, as a nested gene in the largest intron of PIK4CA gene. This is the only serpin that is known so far, to be nested or overlapped (Kumar et al., 2014a), and was initially reported in *Takifugu* (Kumar, 2009a). Nested gene is a gene that is located within a larger gene (Assis et al., 2008; Kumar, 2009b). There are two types of nested genes, either “within intron” genes, which are nested within the intron of the host gene or “non-intronic” genes, nested within the exonic region of the host gene (Kumar, 2009b).

The most interesting part of genomic characters of the serpins is the changes of exon/intron patterns in the vertebrate serpins via either insertion or deletion of spliceosomal introns. Spliceosomal introns and its splicing machinery are the hallmarks of eukaryotes and this adds subtle intricacies to the gene regulation mechanism. However, formation of congruent sequences by mere chance, execution of effective splicing and maintenance of these sequences for several million years due to certain selective forces, remains quite an enigma (Roy & Gilbert, 2006). Intron invasion is assumed to have happened early on in evolution. Nevertheless, there are several examples of late insertion of introns.

In total, 24 conserved introns are reported in vertebrate serpins encompassing group V1–V6 (Kumar & Ragg, 2008), with six additional introns that were gained in selected ray finned fishes among serpin genes (Ragg et al., 2009). Notably, the intron gains in the non-serpin domain of C1IN have also been reported in these selected fishes (Kumar et al., 2014b). Selected ray-finned fishes (namely *Fugu*, medaka, platyfish, *Tetraodon*, tilapia and stickleback) have novel introns and these fishes have genomes ranging in size from 350–950 Mb or below 1,000 Mb (green box in Fig. 5). Introns are either gained or lost through out eukaryotic evolution (Fedorov et al., 2003; Lee & Chang, 2013; Lin et al., 2006; Roy & Irimia, 2009; Verhelst, Van de Peer & Rouzé, 2013; Yenerall, Krupa & Zhou, 2011;



**Figure 5** Spliceosomal introns are inserted only in selected ray-finned fishes with genome size lower than 1,000 Mb.

*Yenerall & Zhou, 2012; Zhu & Niu, 2013*). The only case of intron loss was observed in the ATIII/SERPINC1 gene (*Kumar et al., 2013a*).

The mechanisms proposed by Yenerall's group for spliceosomal introns insertions (*Yenerall, Krupa & Zhou, 2011; Yenerall & Zhou, 2012*) are as follows: (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, and (g) intronization. Double-strand break repair (DSBR) coupled with genome compaction events are the driving forces for several examples of intron insertions in selected ray-finned fishes whose genome underwent compaction events in different group of superfamilies such as in the serpin core domain (*Ragg et al., 2009*), the non-core domain of serpins (*Kumar et al., 2014b*) and in the selected G-protein coupled receptors (*Kumar et al., 2011*).



Intron-exon and higher sequence similarities were maintained in serpins of a particular lineage such as in vertebrates (Fig. 2), urochordates (Kumar & Bhandari, 2014) and within insects (Kumar et al., 2014c). With several 1,000 genomes being currently underway, this issue of gene structure pattern will be revisited within the next decade, when several new animal genomes will be available in the databases.

## CONCLUSIONS

By utilizing Bayesian phylogenetic method, this report corroborated that the six vertebrate serpins groups are conserved from lampreys to humans for circa 500 MY. Moreover, this study provides several vignettes of vertebrate serpins from genomic and phylogenetic perspectives.

## ACKNOWLEDGEMENTS

I thank Chitra Rajakuberan for editing the final version of this manuscript.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

The author received no funding for this article.

### Competing Interests

Abhishek Kumar is an Academic Editor for PeerJ.

### Author Contributions

- Abhishek Kumar conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.1026#supplemental-information>.

## REFERENCES

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:3389–3402 DOI 10.1093/nar/25.17.3389.
- Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Deha P, Christoffels A, Rash S, Hoon S, Smit A, Sollewijn Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJK, 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. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297:1301–1310 DOI 10.1126/science.1072104.

- Assis R, Kondrashov AS, Koonin EV, Kondrashov FA. 2008. Nested genes and increasing organizational complexity of metazoan genomes. *Trends in Genetics* 24:475–478 DOI 10.1016/j.tig.2008.08.003.
- Atchley WR, Lokot T, Wollenberg K, Dress A, Ragg H. 2001. Phylogenetic analyses of amino acid variation in the serpin proteins. *Molecular Biology and Evolution* 18:1502–1511 DOI 10.1093/oxfordjournals.molbev.a003936.
- Benarafa C, Remold-O'Donnell E. 2005. The ovalbumin serpins revisited: perspective from the chicken genome of clade B serpin evolution in vertebrates. *Proceedings of the National Academy of Sciences of the United States of America* 102:11367–11372 DOI 10.1073/pnas.0502934102.
- Bentele C, Kruger O, Todtmann U, Oley M, Ragg H. 2006. A proprotein convertase-inhibiting serpin with an endoplasmic reticulum targeting signal from Branchiostoma lanceolatum, a close relative of vertebrates. *Biochemical Journal* 395:449–456 DOI 10.1042/BJ20051947.
- Birney E, Andrews D, Caccamo M, Chen Y, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez X M, Flicek P, Gräf S, Hammond M, Herrero J, Howe K, Iyer V, Jekosch K, Kähäri A, Kasprzyk A, Keefe D, Kokocinski F, Kulesha E, London D, Longden I, Melsopp C, Meidl P, Overduin B, Parker A, Proctor G, Prlic A, Rae M, Rios D, Redmond S, Schuster M, Sealy I, Searle S, Severin J, Slater G, Smedley D, Smith J, Stabenau A, Stalker J, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodwark C, Hubbard TJP. 2006. Ensembl 2006. *Nucleic Acids Research* 34:D556–D561 DOI 10.1093/nar/gkj133.
- Chen PY, Chang WS, Chou RH, Lai YK, Lin SC, Chi CY, Wu CW. 2007. Two non-homologous brain diseases-related genes, SERPINI1 and PDCD10, are tightly linked by an asymmetric bidirectional promoter in an evolutionarily conserved manner. *BMC Molecular Biology* 8:2 DOI 10.1186/1471-2199-8-2.
- Coughlin PB. 2005. Antiplasmin: the forgotten serpin? *FEBS Journal* 272:4852–4857 DOI 10.1111/j.1742-4658.2005.04881.x.
- Danielli A, Kafatos F, Loukeris T. 2003. Cloning and characterization of four Anopheles gambiae serpin isoforms, differentially induced in the midgut by Plasmodium berghei invasion. *Journal of Biological Chemistry* 278:4184–4193 DOI 10.1074/jbc.M208187200.
- Davis AE, Cai S, Liu D. 2007. C1 inhibitor: biologic activities that are independent of protease inhibition. *Immunobiology* 212:313–323 DOI 10.1016/j.imbio.2006.10.003.
- Davis AE, Lu F, Mejia P. 2010. C1 inhibitor, a multi-functional serine protease inhibitor. *Thrombosis and Haemostasis* 104:886–893 DOI 10.1160/TH10-01-0073.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792–1797 DOI 10.1093/nar/gkh340.
- Ellisdon AM, Zhang Q, Henstridge MA, Johnson TK, Warr CG, Law RH, Whisstock JC. 2014. High resolution structure of cleaved Serpin 42 Da from Drosophila melanogaster. *BMC Structural Biology* 14:14 DOI 10.1186/1472-6807-14-14.
- Fedorov A, Roy S, Fedorova L, Gilbert W. 2003. Mystery of intron gain. *Genome Research* 13:2236–2241 DOI 10.1101/gr.1029803.
- Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, Gil L, Garcia-Giron C, Gordon L, Hourlier T, Hunt S, Juettemann T, Kahari AK, Keenan S, Komorowska M, Kulesha E, Longden I, Maurel T, McLaren WM, Muffato M, Nag R, Overduin B, Pignatelli M, Pritchard B, Pritchard E, Riat HS, Ritchie GRS, Ruffier M, Schuster M, Sheppard D, Sobral D, Taylor K, Thormann A, Trevanion S, White S, Wilder SP, Aken BL, Birney E, Cunningham F, Dunham I,

- Harrow J, Herrero J, Hubbard TJP, Johnson N, Kinsella R, Parker A, Spudich G, Yates A, Zadissa A, Searle SMJ. 2013. Ensembl 2013. *Nucleic Acids Research* 41:D48–D55 DOI 10.1093/nar/gks1236.
- Forsyth S, Horvath A, Coughlin P. 2003. A review and comparison of the murine alpha1-antitrypsin and alpha1-antichymotrypsin multigene clusters with the human clade A serpins. *Genomics* 81:336–345 DOI 10.1016/S0888-7543(02)00041-1.
- Gettins PG. 2002. Serpin structure, mechanism, and function. *Chemical Reviews* 102:4751–4804 DOI 10.1021/cr010170+.
- Gettins PGW, Patston PA, Olson ST. 1996. *Serpins: structure, function and biology*. Austin: R. G. Landes Co.
- Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee H-M, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, de Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Weinstock GM, Havlak PH, Chen R, James Durbin K, Egan A, Ren Y, Song X-Z, Li B, Liu Y, Qin X, Cawley S, Weinstock GM, Worley KC, Cooney AJ, D'Souza LM, Martin K, Qian Wu J, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Smith D, Huang H, Fechtel K, Wang H, Xing H, Weinstock K, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Albà M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hübner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee Y-A, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyras E, Searle SM, Cooper GM, Batzoglou S, Brudno M, Sidow A, Stone EA, Craig Venter J, Payseur BA, Bourque G, López-Otín C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S, Collins F. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428:493–521 DOI 10.1038/nature02426.
- Gregory TR. 2014. Animal genome size database. Available at <http://www.genomesize.com>.
- Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, Ovcharenko I, Putnam NH, Shu S, Taher L, Blitz IL, Blumberg B, Dichmann DS, Dubchak I, Amaya E, Detter JC, Fletcher R, Gerhard DS, Goodstein D, Graves T, Grigoriev IV, Grimwood J,

- Kawashima T, Lindquist E, Lucas SM, Mead PE, Mitros T, Ogino H, Ohta Y, Poliakov AV, Pollet N, Robert J, Salamov A, Sater AK, Schmutz J, Terry A, Vize PD, Warren WC, Wells D, Wills A, Wilson RK, Zimmerman LB, Zorn AM, Grainger R, Grammer T, Khokha MK, Richardson PM, Rokhsar DS. 2010. The genome of the Western clawed frog *Xenopus tropicalis*. *Science* 328:633–636 DOI 10.1126/science.1183670.
- Hendershot LM, Bulleid NJ. 2000. Protein-specific chaperones: the role of hsp47 begins to gel. *Current Biology* 10:R912–915 DOI 10.1016/S0960-9822(00)00850-2.
- Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MAM, Delany ME, Dodgson JB, Chinwalla AT, Cliften PF, Clifton SW, Delehaunty KD, Fronick C, Fulton RS, Graves TA, Kremitzki C, Layman D, Magrini V, McPherson JD, Miner TL, Minx P, Nash WE, Nhan MN, Nelson JO, Oddy LG, Pohl CS, Randall-Maher J, Smith SM, Wallis JW, Yang S-P, Romanov MN, Rondelli CM, Paton B, Smith J, Morrice D, Daniels L, Tempest HG, Robertson L, Masabanda JS, Griffin DK, Vignal A, Fillon V, Jacobsson L, Kerje S, Andersson L, Crooijmans RPM, Aerts J, Van der Poel JJ, Ellegren H, Caldwell RB, Hubbard SJ, Grafham DV, Kierzek AM, McLaren SR, Overton IM, Arakawa H, Beattie KJ, Bezzubov Y, Boardman PE, Bonfield JK, Croning MDR, Davies RM, Francis MD, Humphray SJ, Scott CE, Taylor RG, Tickle C, Brown WRA, Rogers J, Buerstedde J-M, Wilson SA, Stubbs L, Ovcharenko I, Gordon L, Lucas S, Miller MM, Inoko H, Shiina T, Kaufman J, Salomonsen J, Skjoedt K, Wong Gane KAS, Wang J, Liu B, Yu J, Yang H, Nefedov M, Koriabine M, deJong PJ, Goodstadt L, Webber C, Dickens NJ, Letunic I, Suyama M, Torrents D, Von Mering C, Zdobnov EM, Makova K, Nekrutenko A, Elnitski L, Eswara P, King DC, Yang S, Tyekucheva S, Radakrishnan A, Harris RS, Chiaromonte F, Taylor J, He J, Rijnkels M, Griffiths-Jones S, Ureta-Vidal A, Hoffman MM, Severin J, Searle SMJ, Law AS, Speed D, Waddington D, Cheng Z, Tuzun E, Eichler E, Bao Z, Flicek P, Shteynberg DD, Brent MR, Bye JM, Huckle EJ, Chatterji S, Dewey C, Pachter L, Kouranov A, Mourelatos Z, Hatzigeorgiou AG, Paterson AH, Ivarie R, Brandstrom M, Axelsson E, Backstrom N, Berlin S, Webster MT, Pourquie O, Reymond A, Ucla C, Antonarakis SE, Long M, Emerson JJ, Betrán E, Dupanloup I, Kaessmann H, Hinrichs AS, Bejerano G, Furey TS, Harte RA, Raney B, Siepel A, Kent WJ, Haussler D, Eyraes E, Castelo R, Abril JF, Castellano S, Camara F, Parra G, Guigo R, Bourque G, Tesler G, Pevzner PA, Smit A, Fulton LA, Mardis ER, Wilson RK. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716 DOI 10.1038/nature03154.
- Huntington JA. 2013. Thrombin inhibition by the serpins. *Journal of Thrombosis and Haemostasis* 11(Suppl 1):254–264 DOI 10.1111/jth.12252.
- Huntington JA. 2014. Natural inhibitors of thrombin. *Thrombosis and Haemostasis* 111:583–589 DOI 10.1160/TH13-10-0811.
- Ishigami S, Sandkvist M, Tsui F, Moore E, Coleman TA, Lawrence DA. 2007. Identification of a novel targeting sequence for regulated secretion in the serine protease inhibitor neuroserpin. *Biochemical Journal* 402:25–34 DOI 10.1042/BJ20061170.
- Izuhara K, Ohta S, Kanaji S, Shiraishi H, Arima K. 2008. Recent progress in understanding the diversity of the human ov-serpin/clade B serpin family. *Cellular and Molecular Life Science* 65:2541–2553 DOI 10.1007/s00018-008-8049-7.
- Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, 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, Biémont C, Skalli Z, Cattolico L, Poulain J, de

- Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau J-P, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff J-N, Guigó R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quétier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431:946–957 DOI 10.1038/nature03025.
- Jordan RE. 1983. Antithrombin in vertebrate species: conservation of the heparin-dependent anticoagulant mechanism. *Archives of Biochemistry and Biophysics* 227:587–595 DOI 10.1016/0003-9861(83)90488-5.
- Kaiserman D, Bird P. 2005. Analysis of vertebrate genomes suggests a new model for clade B serpin evolution. *BMC Genomics* 6:167 DOI 10.1186/1471-2164-6-167.
- Krem MM, Di Cera E. 2003. Conserved Ser residues, the shutter region, and speciation in serpin evolution. *Journal of Biological Chemistry* 278:37810–37814 DOI 10.1074/jbc.M305088200.
- Kumar A. 2009a. Delving into vertebrate serpins for understanding their evolution. *Nature Precedings* DOI 10.1038/npre.2009.3050.1.
- Kumar A. 2009b. An overview of nested genes in eukaryotic genomes. *Eukaryotic Cell* 8:1321–1329 DOI 10.1128/EC.00143-09.
- Kumar A. 2010. Phylogenomics of vertebrate serpins. Doctoral Thesis, Bielefeld University, Bielefeld. Available at <http://nbn-resolving.de/urn:nbn:de:hbz:361-17480>.
- Kumar A, Bhandari A. 2014. Urochordate serpins are classified into six groups encoded by exon-intron structures, microsynteny and bayesian phylogenetic analyses. *Journal of Genomics* 2:131–140 DOI 10.7150/jgen.9437.
- Kumar A, Bhandari A, Sarde SJ, Goswami C. 2013a. Sequence, phylogenetic and variant analyses of antithrombin III. *Biochemical and Biophysical Research Communications* 440:714–724 DOI 10.1016/j.bbrc.2013.09.134.
- Kumar A, Bhandari A, Sarde SJ, Goswami C. 2014a. Genetic variants and evolutionary analyses of heparin cofactor II. *Immunobiology* 219:713–728 DOI 10.1016/j.imbio.2014.05.003.
- Kumar A, Bhandari A, Sarde SJ, Goswami C. 2014b. Molecular phylogeny of C1 inhibitor depicts two immunoglobulin-like domains fusion in fishes and ray-finned fishes specific intron insertion after separation from zebrafish. *Biochemical and Biophysical Research Communications* 450:219–226 DOI 10.1016/j.bbrc.2014.05.097.
- Kumar A, Bhandari A, Sarde SJ, Muppavarapu S, Tandon R. 2015. Understanding V(D)J recombination initiator RAG1 gene using molecular phylogenetic and genetic variant analyses and upgrading missense and non-coding variants of clinical importance. *Biochemical and Biophysical Research Communications* 462:301–313 DOI 10.1016/j.bbrc.2015.04.125.
- Kumar A, Bhandari A, Sinha R, Goyal P, Grapputo A. 2011. 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:e22046 DOI 10.1371/journal.pone.0022046.
- Kumar A, Congiu L, Lindstrom L, Piironen S, Vidotto M, Grapputo A. 2014c. Sequencing, *De Novo* assembly and annotation of the Colorado potato beetle, *Leptinotarsa decemlineata*, transcriptome. *PLoS ONE* 9:e86012 DOI 10.1371/journal.pone.0086012.
- Kumar A, Kollath-Leiss K, Kempken F. 2013. Characterization of bud emergence 46 (BEM46) protein: sequence, structural, phylogenetic and subcellular localization analyses. *Biochemical and Biophysical Research Communications* 438:526–532 DOI 10.1016/j.bbrc.2013.07.103.
- Kumar A, Ragg H. 2008. Ancestry and evolution of a secretory pathway serpin. *Bmc Evolutionary Biology* 8:250 DOI 10.1186/1471-2148-8-250.

- Kumar A, Sarde SJ, Bhandari A. 2014.** Revising angiotensinogen from phylogenetic and genetic variants perspectives. *Biochemical and Biophysical Research Communications* **446**:504–518 DOI [10.1016/j.bbrc.2014.02.139](https://doi.org/10.1016/j.bbrc.2014.02.139).
- Lamande SR, Bateman JF. 1999.** Procollagen folding and assembly: the role of endoplasmic reticulum enzymes and molecular chaperones. *Seminars in Cell & Developmental Biology* **10**:455–464 DOI [10.1006/scdb.1999.0317](https://doi.org/10.1006/scdb.1999.0317).
- Lee YCG, Chang H-H. 2013.** The evolution and functional significance of nested gene structures in *Drosophila melanogaster*. *Genome Biology and Evolution* **5**:1978–1985 DOI [10.1093/gbe/evt149](https://doi.org/10.1093/gbe/evt149).
- Lewis MJ, Sweet DJ, Pelham HR. 1990.** The ERD2 gene determines the specificity of the luminal ER protein retention system. *Cell* **61**:1359–1363 DOI [10.1016/0092-8674\(90\)90699-F](https://doi.org/10.1016/0092-8674(90)90699-F).
- Li Y, Liu S, Qin Z, Yao J, Jiang C, Song L, Dunham R, Liu Z. 2015.** The serpin superfamily in channel catfish: identification, phylogenetic analysis and expression profiling in mucosal tissues after bacterial infections. *Developmental and Comparative Immunology* **49**:267–277 DOI [10.1016/j.dci.2014.12.006](https://doi.org/10.1016/j.dci.2014.12.006).
- Lin H, Zhu W, Silva JC, Gu X, Buell CR. 2006.** Intron gain and loss in segmentally duplicated genes in rice. *Genome Biology* **7**:R41 DOI [10.1186/gb-2006-7-5-r41](https://doi.org/10.1186/gb-2006-7-5-r41).
- Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyraas E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigó R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korfi I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES. 2002.** Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**(6915):520–562 DOI [10.1038/nature01262](https://doi.org/10.1038/nature01262).
- Nagata K. 1996.** Hsp47: a collagen-specific molecular chaperone. *Trends in Biochemical Sciences* **21**:22–26 DOI [10.1016/S0968-0004\(06\)80023-X](https://doi.org/10.1016/S0968-0004(06)80023-X).

- Oley M, Letzel M, Ragg H. 2004. Inhibition of furin by serpin Spn4A from *Drosophila melanogaster*. *FEBS Letters* 577:165–169 DOI 10.1016/j.febslet.2004.10.003.
- Osterwalder T, Cinelli P, Baici A, Pennella A, Krueger S, Schrimpf S, Meins M, Sonderegger P. 1998. The axonally secreted serine proteinase inhibitor, neuroserpin, inhibits plasminogen activators and plasmin but not thrombin. *Journal of Biological Chemistry* 273:2312–2321 DOI 10.1074/jbc.273.4.2312.
- Osterwalder T, Kuhnen A, Leiserson W, Kim Y, Keshishian H. 2004. *Drosophila* serpin 4 functions as a neuroserpin-like inhibitor of subtilisin-like proprotein convertases. *Journal of Neuroscience* 24:5482–5491 DOI 10.1523/JNEUROSCI.5577-03.2004.
- Pelham HR. 1990. The retention signal for soluble proteins of the endoplasmic reticulum. *Trends in Biochemical Sciences* 15:483–486 DOI 10.1016/0968-0004(90)90303-S.
- Pelissier P, Delourme D, Germot A, Blanchet X, Becila S, Maftah A, Leveziel H, Ouali A, Bremaud L. 2008. An original SERPINA3 gene cluster: elucidation of genomic organization and gene expression in the *Bos taurus* 21q24 region. *BMC Genomics* 9:151 DOI 10.1186/1471-2164-9-151.
- Ragg H, Kumar A, Koster K, Bentele C, Wang Y, Frese MA, Prib N, Kruger O. 2009. Multiple gains of spliceosomal introns in a superfamily of vertebrate protease inhibitor genes. *BMC Evolutionary Biology* 9:208 DOI 10.1186/1471-2148-9-208.
- Ragg H, Lokot T, Kamp PB, Atchley WR, Dress A. 2001. Vertebrate serpins: construction of a conflict-free phylogeny by combining exon-intron and diagnostic site analyses. *Molecular Biology and Evolution* 18:577–584 DOI 10.1093/oxfordjournals.molbev.a003838.
- Raykhel I, Alanen H, Salo K, Jurvansuu J, Nguyen VD, Latva-Ranta M, Ruddock L. 2007. A molecular specificity code for the three mammalian KDEL receptors. *Journal of Cell Biology* 179:1193–1204 DOI 10.1083/jcb.200705180.
- Richer M, Keays C, Waterhouse J, Minhas J, Hashimoto C, Jean F. 2004. The Spn4 gene of *Drosophila* encodes a potent furin-directed secretory pathway serpin. *Proceedings of the National Academy of Sciences of the United States of America* 101:10560–10565 DOI 10.1073/pnas.0401406101.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574 DOI 10.1093/bioinformatics/btg180.
- Roy SW, Gilbert W. 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nature Reviews Genetics* 7:211–221.
- Roy SW, Irimia M. 2009. Mystery of intron gain: new data and new models. *Trends in Genetics* 25:67–73 DOI 10.1016/j.tig.2008.11.004.
- Sauk JJ, Nikitakis N, Siavash H. 2005. Hsp47 a novel collagen binding serpin chaperone, autoantigen and therapeutic target. *Frontiers in Bioscience* 10:107–118 DOI 10.2741/1513.
- Sawant S, Aparicio S, Tink AR, Lara N, Barnstable CJ, Tombran-Tink J. 2004. Regulation of factors controlling angiogenesis in liver development: a role for PEDF in the formation and maintenance of normal vasculature. *Biochemical and Biophysical Research Communications* 325:408–413 DOI 10.1016/j.bbrc.2004.10.041.
- Semenza JC, Hardwick KG, Dean N, Pelham HR. 1990. ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway. *Cell* 61:1349–1357 DOI 10.1016/0092-8674(90)90698-E.
- Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. 2013. New and continuing developments at PROSITE. *Nucleic Acids Research* 41:D344–D347 DOI 10.1093/nar/gks1067.

- Silverman GA, Bird PI, Carrell RW, Church FC, Coughlin PB, Gettins PG, Irving JA, Lomas DA, Luke CJ, Moyer RW, Pemberton PA, Remold-O'Donnell E, Salvesen GS, Travis J, Whisstock JC. 2001. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *Journal of Biological Chemistry* 276:33293–33296 DOI 10.1074/jbc.R100016200.
- Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, Morgan JR, Buxbaum JD, Sachidanandam R, Sims C, Garruss AS, Cook M, Krumlauf R, Wiedemann LM, Sower SA, Decatur WA, Hall JA, Amemiya CT, Saha NR, Buckley KM, Rast JP, Das S, Hirano M, McCurley N, Guo P, Rohner N, Tabin CJ, Piccinelli P, Elgar G, Ruffier M, Aken BL, Searle SMJ, Muffato M, Pignatelli M, Herrero J, Jones M, Brown CT, Chung-Davidson Y-W, Nanlohy KG, Libants SV, Yeh C-Y, McCauley DW, Langeland JA, Pancer Z, Fritzsche B, de Jong PJ, Zhu B, Fulton LL, Theising B, Flicek P, Bronner ME, Warren WC, Clifton SW, Wilson RK, Li W. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nature Genetics* 45:415–421 DOI 10.1038/ng.2568.
- Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* 33:W465–W467 DOI 10.1093/nar/gki458.
- Steele FR, Chader GJ, Johnson LV, Tombran-Tink J. 1993. Pigment epithelium-derived factor: neurotrophic activity and identification as a member of the serine protease inhibitor gene family. *Proceedings of the National Academy of Sciences of the United States of America* 90:1526–1530 DOI 10.1073/pnas.90.4.1526.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28:2731–2739 DOI 10.1093/molbev/msr121.
- Tombran-Tink J. 2005. The neuroprotective and angiogenesis inhibitory serpin, PEDF: new insights into phylogeny, function, and signaling. *Frontiers in Bioscience* 10:2131–2149 DOI 10.2741/1686.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji R-R, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang ZY, Wang A, Wang X, Wang J, Wei M-H, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu SC, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Lai Cheng M, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferreira S, Garg N,



- Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers Y-H, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang Y-H, Coyne M, Dahlke C, Deslattes Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. 2001. The sequence of the human genome. *Science* 291:1304–1351 DOI [10.1126/science.1058040](https://doi.org/10.1126/science.1058040).
- Verhelst B, Van de Peer Y, Rouzé P. 2013. The complex intron landscape and massive intron invasion in a picoeukaryote provides insights into intron evolution. *Genome Biology and Evolution* 5:2393–2401 DOI [10.1093/gbe/evt189](https://doi.org/10.1093/gbe/evt189).
- Wang Y, Ragg H. 2011. An unexpected link between angiotensinogen and thrombin. *FEBS Letters* 585:2395–2399 DOI [10.1016/j.febslet.2011.06.021](https://doi.org/10.1016/j.febslet.2011.06.021).
- Wong MKS, Takei Y. 2011. Characterization of a native angiotensin from an anciently diverged serine protease inhibitor in lamprey. *Journal of Endocrinology* 209:127–137 DOI [10.1530/JOE-10-0422](https://doi.org/10.1530/JOE-10-0422).
- Yenerall P, Krupa B, Zhou L. 2011. Mechanisms of intron gain and loss in *Drosophila*. *BMC Evolutionary Biology* 11:364 DOI [10.1186/1471-2148-11-364](https://doi.org/10.1186/1471-2148-11-364).
- Yenerall P, Zhou L. 2012. Identifying the mechanisms of intron gain: progress and trends. *Biology Direct* 7:29 DOI [10.1186/1745-6150-7-29](https://doi.org/10.1186/1745-6150-7-29).
- Zhu T, Niu DK. 2013. Mechanisms of intron loss and gain in the fission yeast *Schizosaccharomyces*. *PLoS ONE* 8:e61683 DOI [10.1371/journal.pone.0061683](https://doi.org/10.1371/journal.pone.0061683).