

Research article

Open Access

Analysis of vertebrate genomes suggests a new model for clade B serpin evolution

Dion Kaiserman and Phillip I Bird*

Address: Department of Biochemistry & Molecular Biology, Monash University, Clayton, Victoria, Australia

Email: Dion Kaiserman - dion.kaiserman@med.monash.edu.au; Phillip I Bird* - phil.bird@med.monash.edu.au

* Corresponding author

Published: 23 November 2005

Received: 16 September 2005

BMC Genomics 2005, 6:167 doi:10.1186/1471-2164-6-167

Accepted: 23 November 2005

This article is available from: <http://www.biomedcentral.com/1471-2164/6/167>

© 2005 Kaiserman and Bird; licensee BioMed Central Ltd.

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

Abstract

Background: The human genome contains 13 clade B serpin genes at two loci, 6p25 and 18q21. The three genes at 6p25 all conform to a 7-exon gene structure with conserved intron positioning and phasing, however, at 18q21 there are two 7-exon genes and eight genes with an additional exon yielding an 8-exon structure. Currently, it is not known how these two loci evolved, nor which gene structure arose first – did the 8-exon genes gain an exon, or did the 7-exon genes lose one? Here we use the genomes of diverse vertebrate species to plot the emergence of clade B serpin genes and to identify the point at which the two genomic structures arose.

Results: Analysis of the chicken genome indicated the presence of a single clade B serpin gene locus, containing orthologues of both human loci and both genomic structures. The frog genome and the genomes of three fish species presented progressively simpler loci, although only the 7-exon structure could be identified. The *Serp12* gene contains seven exons in the frog genome, but eight exons in chickens and humans, indicating that the additional exon evolved in this gene.

Conclusion: We propose a new model for clade B serpin evolution from a single 7-exon gene (either *Serp1* or *Serp6*). An additional exon was gained in the *Serp12* gene between the tetrapoda and amniota radiations to produce the 8-exon structure. Both structures were then duplicated at a single locus until a chromosomal breakage occurred at some point along the mammalian lineage resulting in the two modern loci.

Background

The serpins are a superfamily of proteins sharing a conserved tertiary structure [1] that has evolved primarily to control proteolytic activity, although some have evolved non-inhibitory functions [2-6]. To date, over 500 serpin sequences have been identified in the genomes of species from almost all phyla including viruses, bacteria, metazoans and plants [1,7].

The serpin mechanism of inhibition is based on the metastable nature of the native serpin fold [8]. The reac-

tive centre loop (RCL) is exposed at the top of the serpin where it acts as a pseudosubstrate for the target proteinase [9]. Following cleavage of the RCL, a conformational rearrangement, termed the stressed to relaxed transition, occurs and the RCL inserts into a β -sheet forming an extra β -strand [10]. This rearrangement is responsible for the distortion and irreversible inactivation of the target proteinase [11].

Phylogenetic analysis has divided serpins into 16 clades and a number of orphan sequences [1]. The clade B ser-

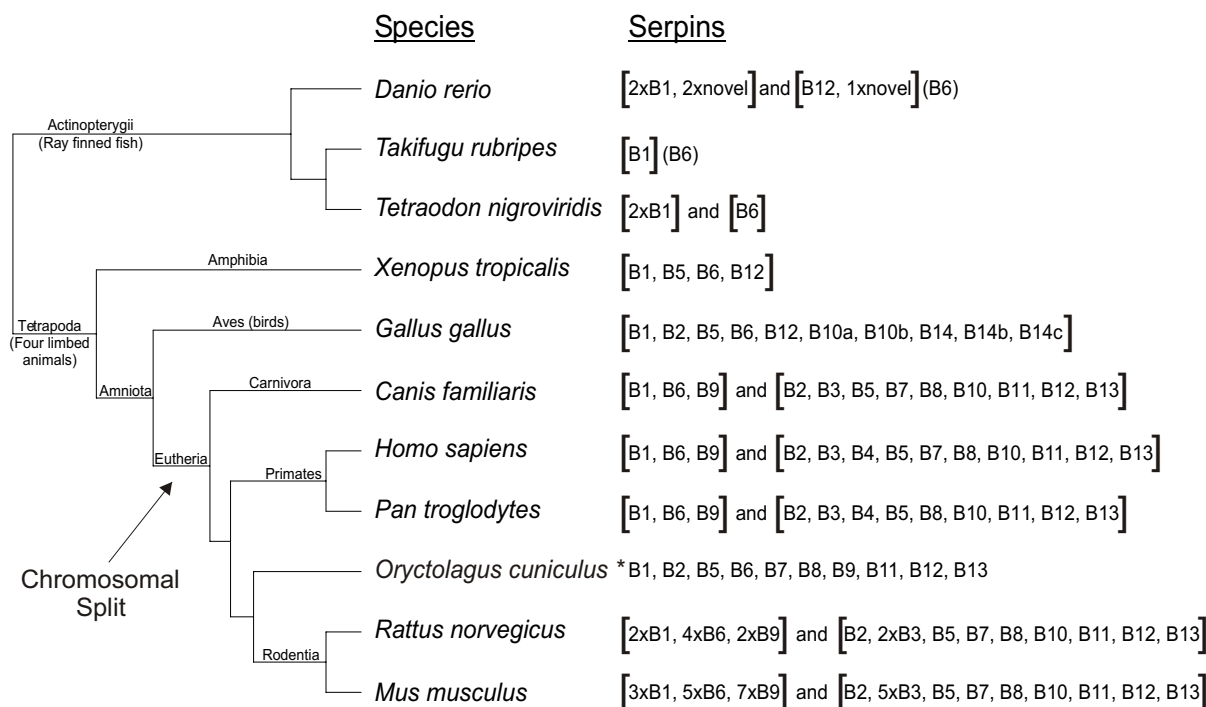


Figure 1
Distribution of clade B serpin genes in vertebrate genomes. A dendrogram showing relationships between vertebrate species with clade B serpin gene loci. The genes identified within each genome are indicated on the right without the *SERPIN* root designation. Multiple loci are delineated by square brackets, while genes identified in EST databases, but not in the genome, are in round brackets. * due to the incomplete nature of the sequencing data, the structure of the rabbit (*Oryctolagus cuniculus*) loci could not be established. Branch lengths are not to scale.

pins are primarily intracellular proteinase inhibitors [12]. In humans this clade consists of 13 members and includes regulators of inflammation, apoptosis and angiogenesis (reviewed in [13]). The genes are arranged at two genomic loci at 6p25 (3 genes: *SERPINB1*, *SERPINB6* and *SERPINB9*) and 18q21 (10 genes: *SERPINB2*, *B3*, *B4*, *B5*, *B7*, *B8*, *B10*, *B11*, *B12* and *B13*), and all clade B serpins described to date conform to two related gene structures of seven or eight exons. Both the phasing and positioning of the intron/exon splice sites are identical for six introns, leading to conserved exon lengths. The additional intron splice site of the 8-exon structure is also conserved, although the extra exon encodes a highly divergent sequence of variable length, termed the CD loop.

Although clade B serpins have been identified in mammals (humans, mice and rats) as well as birds, it is not known how they evolved, nor has the primordial structure or gene been identified. Two competing models exist to explain the presence of two clade B loci in humans. The first proposes duplication of the entire 6p25 locus followed by a number of single gene duplications at 18q21

[14]. The second proposes duplication of a single gene from 18q21 to 6p25, followed by successive duplications at both loci to derive the modern complement [15]. Two key differences are apparent between these models. The former assumes that the 7-exon structure and 6p25 locus are ancestral, whereas the latter assumes the opposite.

Here, we use the completed chicken genome to show that the two mammalian loci are the result of an ancient chromosomal split, as orthologues of genes from both human loci are linked in the chicken. Further analysis of emerging genome sequences suggests that clade B serpin genes arose from a primordial *Serpinb1* or *b6* gene with seven exons, the additional exon being gained in *Serpinb12* after the divergence of amphibians and amniotes.

Results

Clade B serpin genes in the chicken (*Gallus gallus*)

To determine which of the two clade B serpin loci is ancestral, the chicken genome was investigated, as it represents a distantly related species. This analysis was also recently performed independently by another group, yielding sim-

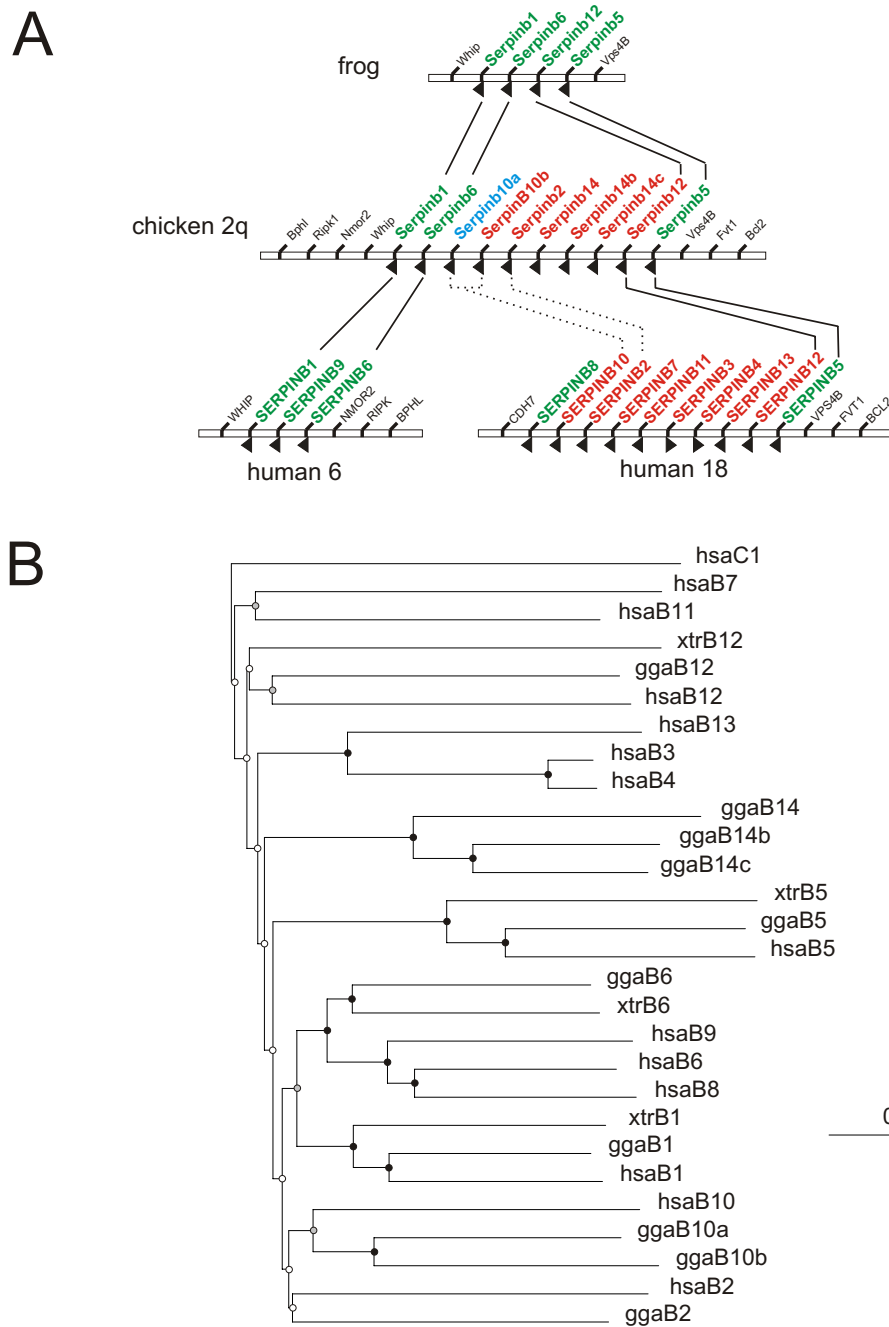


Figure 2

Comparison of frog, chicken and human clade B serpin loci. The structure of the serpin locus in frog, chicken and human is shown with 7-exon serpin genes in green, 8-exon serpin genes in red and non-serpin genes in black. Gene structure could not be predicted for *Serpina10a* (blue). Arrows indicate the direction of transcription. Orthologous genes with strong RCL conservation are linked with solid lines, dotted lines denote inter-species orthologues with weak RCL homology. (B) The amino acid sequences of human, chicken and frog clade B serpins were aligned with human antithrombin (*SERPINC1*) and a Neighbour-Joining tree constructed (gapped positions removed, 1000 bootstraps) to show evolutionary relationships. Scale bar indicates the number of substitutions at each site. Node colour indicates the bootstrap value (black >75%, grey 50–75%, white <50%). Genes are named without the *SERPIN* root. hsa = *Homo sapiens*; gga = *Gallus gallus*; xtr = *Xenopus tropicalis*.

Table 1: Accession numbers of sequences used in this study

Gene (Species)	Accession Number(s)
<i>Serp1b1</i> (<i>Oncorhynchus mykiss</i>)	[Genbank: CX143380]
<i>Serp1b6</i> (<i>Oncorhynchus mykiss</i>)	[Genbank: AY606039]
<i>Serp1b1</i> (<i>Salmo salar</i>)	[Genbank: CA059107 ; Genbank: CA051086]
<i>Serp1b6</i> (<i>Salmo salar</i>)	[Genbank: CK877893]
<i>Serp1b1</i> (<i>Oryzias latipes</i>)	[Genbank: BJ721045]
<i>Serp1b6</i> (<i>Oryzias latipes</i>)	[Genbank: BJ530676]
<i>Serp1b6</i> (<i>Ictalurus punctatus</i>)	[Genbank: CK420790]
<i>Serp1b1</i> (<i>Haplochromis</i>)	[Genbank: BJ692672]
<i>Serp1b1</i> (<i>Gasterosteus aculeatus</i>)	[Genbank: CD504030]
<i>Serp1b1</i> (<i>Danio rerio</i>)	[Genbank: AL912530]
<i>Serp1b6</i> (<i>Danio rerio</i>)	[Genbank: CD015363]
<i>Serp1b1</i> (<i>Xenopus tropicalis</i>)	[Genbank: CX372321 ; Genbank: CF593092 ; Genbank: CX383158 ; Genbank: DR857859]
<i>Serp1b5</i> (<i>Xenopus tropicalis</i>)	[Genbank: CX847450 ; Genbank: CF346211]
<i>Serp1b6</i> (<i>Xenopus tropicalis</i>)	[Genbank: BX709151 ; Genbank: DR860385 ; Genbank: DN093291]
<i>Serp1b12</i> (<i>Xenopus tropicalis</i>)	[Genbank: DR880497]
<i>Serp1b1</i> (<i>Gallus gallus</i>)	[Genbank: CD216625 ; Genbank: CF255528 ; Genbank: CF257148]
<i>Serp1b2</i> (<i>Gallus gallus</i>)	[Genbank: BU409199 ; Genbank: BU205882 ; Genbank: BU410099]
<i>Serp1b5</i> (<i>Gallus gallus</i>)	[Genbank: CD739917]
<i>Serp1b6</i> (<i>Gallus gallus</i>)	[Genbank: CF250725 ; Genbank: BU116697 ; BU133732 ; Genbank: AJ734077]
<i>Serp1b12</i> (<i>Gallus gallus</i>)	[Genbank: BU442268]

ilar results [16]. As illustrated in Figure 2A, the chicken genome contains a single clade B serpin locus on chromosome 2, composed of 10 genes, nine of which are supported by EST data. Both gene structures are also evident within the locus. Single orthologues exist in the chicken for the 7-exon human genes SERPINB1 (MNEI), SERPINB5 (maspin) and SERPINB6 (PI6), and also for the 8-exon genes SERPINB2 (PAI-2) and SERPINB12 (yukopin). Human SERPINB10 (bomapin), an 8-exon gene, has 2 chicken homologues – MENT (which fits the 8-exon arrangement) and a MENT-like gene for which the structure cannot be completely determined. The three remaining chicken genes Serpinb14 (ovalbumin), Serpinb14b (geneY) and Serpinb14c (geneX) are highly related [17] and have no human orthologues. This suggests that the common ancestor of chickens and mammals had a complement of six clade B serpin genes: SERPINB1, B2, B5, B6, B10 and B12. Furthermore, the organization of the chicken clade B serpin locus matches that of humans, contradicting both previously proposed models of clade B serpin gene evolution. Rather than two independently evolving loci, it suggests that both the 7- and 8-exon gene structures evolved at a single locus in the common ancestor of birds and mammals that then split to yield the 2 loci apparent in humans and rodents.

Evolution of clade B serpin genes in vertebrates

Since both the 7 and 8-exon gene structures were present in the chicken genome, we investigated more distantly related genomes to identify which structure arose first. Each of the 13 human protein sequences were used to probe the current builds of vertebrate genomes to identify

clade B serpin genes. This identified genes with strong RCL similarity to human clade B serpins in the genomes of fish, amphibians and various mammals. By plotting the presence of genes on a dendrogram of species a pattern of gene evolution can be proposed (Figure 1).

As described above, the common ancestor of mammals and birds contained a single clade B serpin gene locus of six genes (*Serp1b1*, *b2*, *b5*, *b6*, *b10* and *b12*). All of the mammalian genomes examined so far (human, chimpanzee, dog, rat and mouse) contain two loci, suggesting that the chromosomal split occurred early in mammalian evolution (Figure 1). The locus syntenic with human 6p25 contains *SERPINB1*, *B6* and *B9* in humans, chimpanzees, and dogs and an expanded repertoire of 15 and 8 in mice and rats, respectively [18,19], indicating further expansion within the rodent lineage. The incomplete rabbit genome yielded hits for 10 of the human clade B serpins (*Serp1b3*, *b4* and *b10* were not identified), with no evidence of the expansion observed in rats or mice. Furthermore, the *Serp1b9* RCL sequence of rabbits is predicted to match that of humans whereas the rodent *Serp1b9* RCL is mutated.

The locus syntenic to human 18q21 is identical between mammals with the exception of the *SERPINB3* and *SERPINB4* genes (SCCA-1 and -2), which are unique to each species (data not shown and [20]). This suggests that 12 genes were present in the mammalian common ancestor – *SERPINB1*, *B6* and *B9* at one locus, and *SERPINB2*, *B5*, *B7*, *B8*, *B10*, *B11*, *B12*, *B13* and the primordial *SERPINB3/B4* at the other. Since then, the ancestral

Table 2: Intron/exon phasing in vertebrate clade B serpin genes

Species	Gene ¹	A ²	B	C ³	D	E	F	G
<i>D. rerio</i>	B1	UTR	0	NP	0	I	0	0
<i>T. nigroviridis</i> [†]	B1	UTR	0	NP	0	I	0	0
<i>T. rubripes</i> [†]								
<i>X. tropicalis</i>	B1	UTR	0	NP	0	I	0	0
	B5	UTR	0	NP	0	I	0	0
	B6	UTR	0	NP	0	I	0	0
	B12	UTR	0	NP	0	I	0	0
<i>G. gallus</i>	B1	UTR	0	NP	0	I	0	0
	B2	UTR	0	0	0	I	0	0
	B5	UTR	0	NP	0	I	0	0
	B6	UTR	0	NP	0	I	0	0
	B10a	UTR	0	*	0	I	0	0
	B10b	UTR	0	0	0	I	0	0
	B12	UTR	0	0	0	I	0	0
	B14	UTR	0	0	0	I	0	0
	B14b	UTR	0	0	0	I	0	0
	B14c	UTR	0	0	0	I	0	0
<i>H. Sapiens</i>	B1	UTR	0	NP	0	I	0	0
	B2	UTR	0	0	0	I	0	0
	B3/4	UTR	0	0	0	I	0	0
	B5	UTR	0	NP	0	I	0	0
	B6	UTR	0	NP	0	I	0	0
	B7	UTR	0	0	0	I	0	0
	B8	UTR	0	NP	0	I	0	0
	B9	UTR	0	NP	0	I	0	0
	B10	UTR	0	0	0	I	0	0
	B11	UTR	0	0	0	I	0	0
	B12	UTR	0	0	0	I	0	0
	B13	UTR	0	0	0	I	0	0

¹ Gene names are given without the Serpin root.

² UTR = intron occurs in the 5' untranslated region and thus has no phase

³ NP = Intron not present, * = Intron cannot be reliably predicted and no EST exists.

[†] Intron phasing predicted from consensus sequence

SERPINB3/B4 has evolved independently in each mammalian species.

Further back in evolution, chickens and mammals shared a common ancestor with amphibians approximately 300 million years ago, and the genome of *Xenopus tropicalis* (the western clawed frog) provides an insight into this more primitive gene cluster (Figure 1). The *X. tropicalis* genome contains a single clade B serpin locus of 4 genes (*Serp1b1*, *Serp1b6*, *Serp1b5* and *Serp1b12*), bounded by orthologues of *WHIP* and *VPS4B*. The placement of both serpin and non-serpin genes is conserved between the frog and chicken genomes, strengthening our hypothesis that a single clade B serpin gene cluster has split at some point in mammalian evolution (Figure 2A). Again, the phylogenetic tree confirms that clade B serpin genes cluster on the basis of conserved function, rather than by species (Figure 2B).

450 million years ago, amphibians, birds and mammals shared a common ancestor with fish [21]. Three fish

genomes were available for analysis: *Danio rerio* (zebrafish), *Tetraodon nigroviridis* (green pufferfish) and *Takifugu rubripes* (japanese pufferfish). The *T. rubripes* and *T. nigroviridis* genomes each contain one or two copies of *Serp1b1*, respectively, with no other serpin genes for over 100 kb upstream or downstream, while the *D. rerio* locus contains two copies of *Serp1b1*, linked to two other genes with RCL sequences that do not match any human serpins. However, ESTs derived from both *Serp1b1* and *Serp1b6* can be identified from all three species, indicating the presence of another clade B serpin locus. Indeed, a second locus is identifiable in the *D. rerio* and *T. nigroviridis* genomes. The *T. nigroviridis* locus contains only *Serp1b6*, although the contig is very short and other genes may exist. The *D. rerio* region contains a number of clade B serpin gene fragments as well as a gene with RCL homology to human *SERP1B12*, although no orthologue of *SERP1B6* can be identified. However, the BAC clone finishes in intronic sequence between the final two exons of a serpin gene, and therefore this fragment may be the 5' region of *D. rerio Serp1b6*.

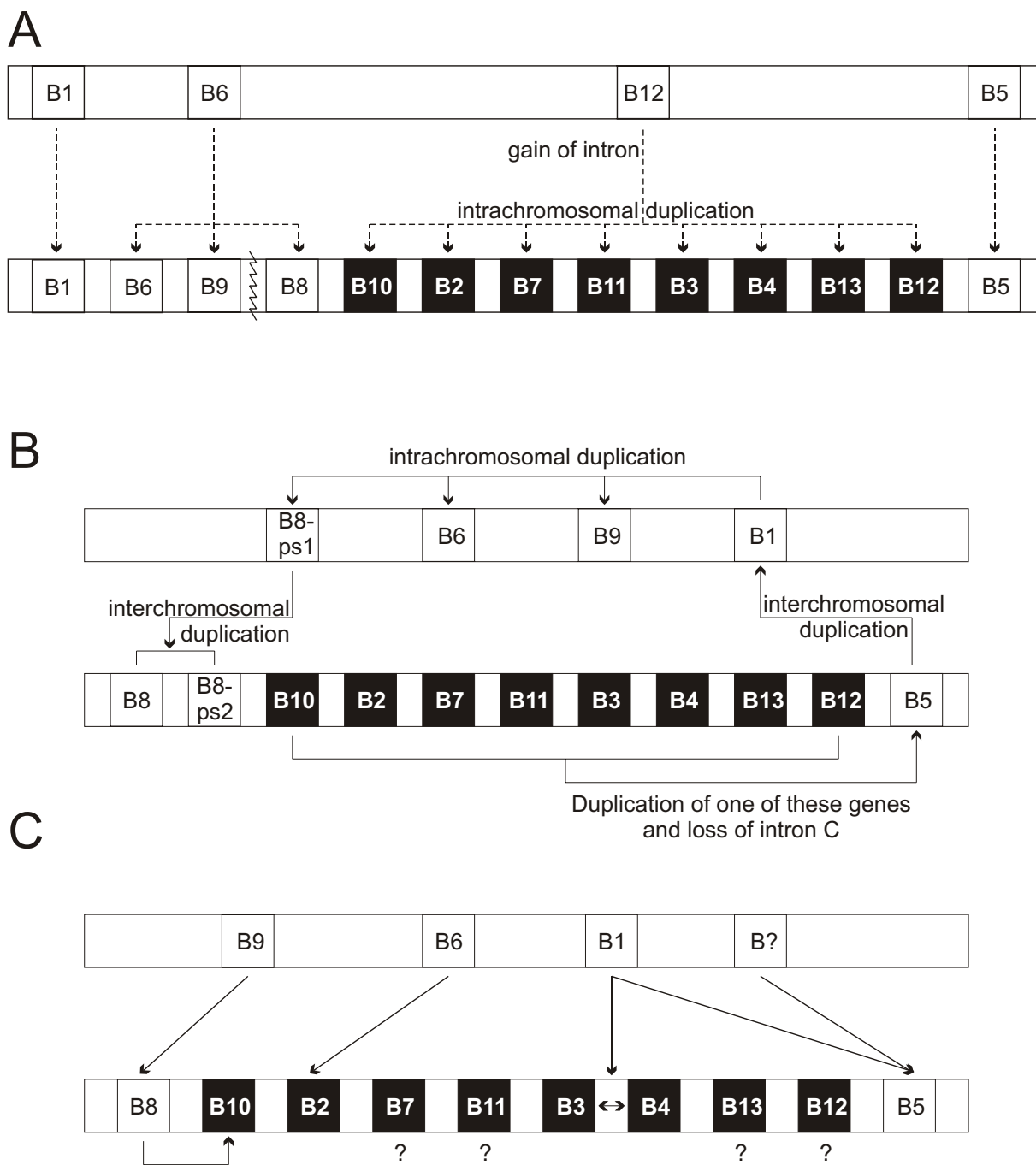


Figure 3
A new mechanism of clade B serpin gene expansion. (A) A new model for evolution of two clade B serpin loci in mammals, from a single locus of 4 genes. A series of intrachromosomal duplications results in 13 genes, followed by chromosomal breakage to yield two loci. Two previously proposed mechanisms based on duplication of an ancestral locus as described by (B) Bartuski *et al.* [14]; and (C) Scott *et al.* [15]. Arrows represent gene duplication events. Black boxes indicate genes displaying the 8-exon structure, white boxes are 7-exon genes. Genes are named without the *SERPIN* root. Diagram not to scale.

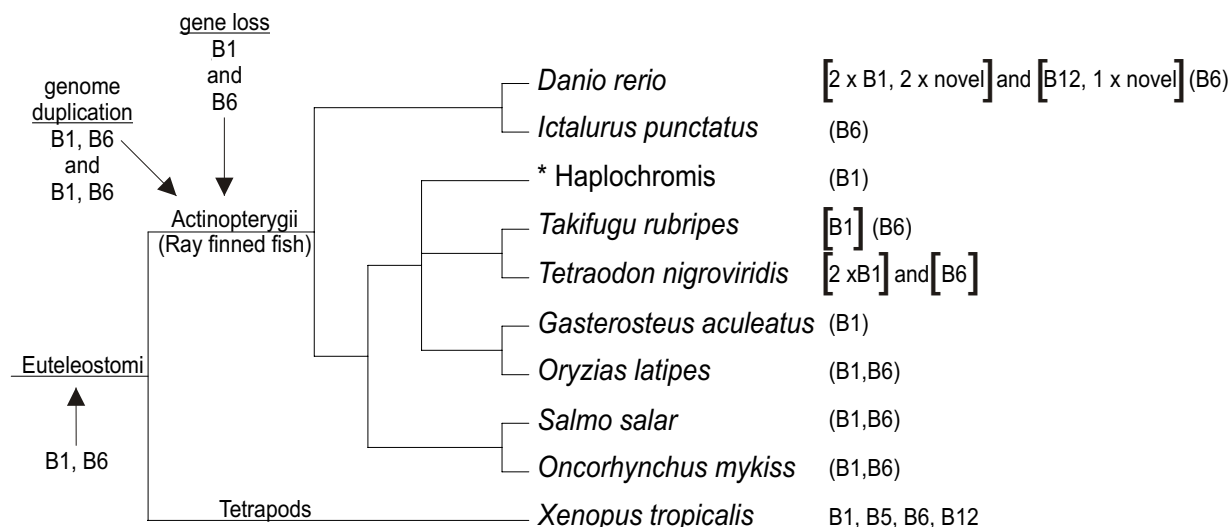


Figure 4
Distribution of clade B serpin ESTs in fish. A dendrogram of fish species is shown with clade B serpins ESTs shown on the right in brackets. The tetrapoda lineage is also indicated, with the simplest locus (*Xenopus tropicalis*) shown for comparison. Genes are named without the *SERPIN* root. The relative timing of genome duplication and subsequent gene loss within ray finned fish are shown as well as the composition of the loci after each event. Asterisk indicates an EST for which only the genus name was given.

Gene structure

The gene structures of the newly identified genes from *Xenopus* and fish were determined by alignment of ESTs to genomic sequence as well as by GENSCAN predictions (Table 2). Human *SERPINB1* is a 7-exon gene, and this is also the case for *Serpinb1* in the three fish genomes investigated, as well as the novel genes identified in *Danio rerio*. Lack of EST coverage resulted in an inability to determine intron phasing for the *T. nigroviridis* and *T. rubripes* genes, however, the positioning of the introns within the predicted mRNA was consistent with the mammalian clade B serpins.

Overlapping ESTs allowed construction of complete mRNA sequences for all 4 of the *Xenopus* genes. *Serpinb1*, *b5* and *b6* conformed to the 7-exon structure, and both positioning and phasing of the intron/exon boundaries was conserved with humans. The other gene to appear in frogs is *Serpinb12*, an 8-exon gene in the chicken and mammalian genomes. However, this gene in frogs contained only 7-exons, lacking the extra exon encoding the CD loop. This gene structure was confirmed by EST DR880497 which spans all 6 intron/exon boundaries. As shown in Figure 2B, *Xenopus Serpinb12* is clearly related to human and chicken *Serpinb12*, suggesting that the 8-exon structure arose some time between the divergence of amphibians and amniotes in the *Serpinb12* gene.

Discussion

The human genome contains two clusters of clade B serpin genes at 6p25 and 18q21. The 13 genes all conform to either a 7- or 8-exon gene structure with conservation of both phasing and positioning of intron/exon boundaries. To address possible mechanisms of clade B gene evolution, the chicken genome was analysed and found to contain a single clade B gene cluster of 10 genes. Another recent investigation of the chicken genome also identified these 10 genes, however, the authors concluded that the newly identified MENT-like chicken gene, *Serpinb10a*, and human *SERPINB10* represented direct orthologues [16]. This was based on the overall protein similarity as well as homology within the CD loop and RCL. While the RCL sequence of chicken *Serpinb10a* and human *SERPINB10* have some similarities, there are a number of charge differences throughout the RCL, and optimal alignment requires insertion of a gap at the P side (N-terminal to the point of proteolytic cleavage), a feature not observed between other chicken-human orthologues. Furthermore, although the presence of a CD loop would be predicted in *Serpinb10a* given its evolutionary ancestry, the presence of an eighth exon requires experimental proof. Without direct sequencing of a transcript, it remains possible that the exon may be missing or contain a premature stop codon or mis-sense mutation resulting in a lack of functional protein expression. The exon identified by Benarafa

and Remold-O'Donnell had a very low statistical score [16], and decreasing the GENSCAN settings to detect this exon leads to the identification of a number of potential exons in other serpin genes that are not transcribed (data not shown).

The identification of orthologues (matching RCLs) of both mammalian clusters (*SERPINB2*, *SERPINB5* and *SERPINB12* from human 18q21 and *SERPINB1* and *SERPINB6* from human 6p25) at a single genomic locus in the chicken raises the intriguing possibility that the serpin clusters arose from an ancient chromosomal breakage, rather than chromosomal duplication events. More distantly related genomes appear to support this conclusion as a single clade B serpin gene cluster exists in the *Xenopus tropicalis* genome, consisting of *Serp1b1*, *b5*, *b6* and *b12*. Therefore, the two loci observed in mammals arose from this single cluster of four 7-exon genes. *Serp1b12* gained an extra exon and was duplicated to give rise to the 8-exon genes, followed by a chromosomal split yielding two loci. This presents a much simpler mechanism for the observation of different genetic structures at a single locus than previously proposed models [14,15] (Figure 3).

Since *Serp1b1* and *Serp1b6* are linked in frogs, chickens and humans, it is likely that they were linked in their common ancestor with fish, however, this was not the case in any of the fish genomes investigated. This may be due to a whole genome duplication that occurred early in the evolution of ray finned fish, followed by rapid loss of duplicate genes [22]. This could give rise to the two serpin loci that have since evolved separately (Figure 4). Therefore, we propose that the single primordial serpin clade B locus (in the common ancestor of fish and tetrapods) contained only *Serp1b1* and *Serp1b6*, as these are the only two genes identifiable in ESTs from multiple species of fish. Following the whole genome duplication, one gene was lost from each locus (resulting in one locus with *Serp1b1* and another with *Serp1b6*) and each of these loci have since evolved uniquely in descendant fish species (Figure 4). The fish family Polypteridae (bichirs) diverged from teleost fish before the whole genome duplication [22], and as such, a genomic analysis within this family would be predicted to show a single 'primordial' clade B serpin gene locus of *Serp1b1* and *Serp1b6*.

The identification of a *Serp1b12* homologue in the *D. rerio* genome suggests that this gene may have evolved significantly earlier than the tetrapod lineage, however, no corroborating ESTs can be found in other fish species. A number of unique clade B serpins have evolved in zebrafish, at both genomic loci, which cannot be found in other fish genomes or EST databases, indicating a more rapid evolution in this fish species. Furthermore, phyloge-

netic analysis indicates that *Danio Serp1b12* does not cluster with *Serp1b12* from other species, but rather with *Danio Serp1b6* (data not shown), and suggests that the *Danio rerio Serp1b12* gene may be a case of convergent evolution.

Six genes (*Serp1b3/b4*, *b7*, *b8*, *b9*, *b11* and *b13*) were gained early along the mammalian branch and the cluster split into two loci, although when these events occurred in relation to each other cannot currently be discerned. The current effort to sequence the platypus (a monotreme) genome, as well as a number of marsupials (the wallaby and opossum), which branched off early in mammalian evolution may shed some light on this.

The analysis of frog *Serp1b12* clearly shows that it is a 7-exon gene, despite having eight exons in both chickens and mammals. This shows that the 7-exon structure arose first and that the eighth exon was gained in *Serp1b12* some time after amphibians evolved, but before mammals and birds split. The gain of the CD loop has enabled some members of the clade B serpin family to evolve motifs for secondary functions, without affecting their inhibitory role [23]. The AT-hook motif and nuclear localisation signal (NLS) of chicken MENT [3] are found in the CD-loop, as is the NLS of human bomapin [24]. The extended CD-loop of PAI-2 contains motifs necessary for its suppression of tumour necrosis factor α induced death [23], redox sensitivity [25,26] and cross-linking to cell membranes [27,28] as well as a novel motif for binding to retinoblastoma protein [29].

These analyses suggest that either *Serp1b1* or *Serp1b6* is the original clade B serpin gene, and that it was present in the common ancestor of euteleostomi (the bony vertebrates, including both fish and mammals) approximately 450 million years ago. A possible functional orthologue of *Serp1b6* is evident in the urochordate species *Ciona intestinalis* (sea squirt). However, this gene does not contain any introns, although other serpin genes in the *Ciona* genome do have introns. This makes it difficult to suggest the *Ciona Serp1b6* gene as a primordial clade B serpin. Furthermore, the suggestion that clade B serpins evolved from an intronless precursor would also be contrary to a recent study suggesting that they arose from an intron-containing gene [30].

The question of where clade B genes arose may be answered when the genomes of the skate and lamprey are completely sequenced. These species diverged at different times between the sea squirt-human and fish-human common ancestors, and may indicate which gene, *Serp1b1* or *Serp1b6*, evolved first, and whether or not clade B serpins are descendant from an intronless precursor.

Conclusion

By investigating the genomes of various vertebrate species, we have shown that the all clade B serpins are descended from a single gene (either *Serp1b1* or *Serp1b6*) that contained seven exons. The additional exon present in the 8-exon structure appeared in the *Serp1b12* gene at some point after the divergence of amphibians from amniotes, but before the divergence of birds and mammals. Therefore, both the 7-exon and 8-exon genes evolved at a single locus that split soon after birds and mammals diverged to yield the two loci present in the human genome.

Methods

The current builds of vertebrate genomes [22,31-34] were accessed through the NCBI and mined using BLAST [35]. The *Xenopus tropicalis* genome was accessed from the Department of Energy Joint Genome Institute. Rabbit genomic contigs were downloaded from the Broad Institute. GENSCAN [36] was also used to predict genes from genomic sequence. Intron/exon boundaries were predicted with Spidey. Alignments were created using ClustalW [37] and manually adjusted with BioEdit (version 6.0.7). Phylogenetic relationships were determined using the Neighbour-Joining method with 1,000 bootstraps and trees were viewed with TreeView [38]. Accession numbers for the sequences identified in this study are given in Table 1.

Authors' contributions

DK performed the analysis and drafted the manuscript. PIB assisted in design of the study and in drafting of the manuscript.

References

- Irving JA, Pike RN, Lesk AM, Whisstock JC: Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Res* 2000, 10:1845-1864.
- Doolittle RF: Angiotensinogen is related to the antitrypsin-antithrombin-ovalbumin family. *Science* 1983, 222:417-419.
- Grigoryev SA, Bednar J, Woodcock CL: **MENT, a heterochromatin protein that mediates higher order chromatin folding, is a new serpin family member.** *J Biol Chem* 1999, 274:5626-5636.
- Zou Z, Anisowicz A, Hendrix MJ, Thor A, Neveu M, Sheng S, Rafidi K, Seftor E, Sager R: **Maspin, a serpin with tumor-suppressing activity in human mammary epithelial cells.** *Science* 1994, 263:526-529.
- Hirayoshi K, Kudo H, Takechi H, Nakai A, Iwamatsu A, Yamada KM, Nagata K: **HSP47: a tissue-specific, transformation-sensitive, collagen-binding heat shock protein of chicken embryo fibroblasts.** *Mol Cell Biol* 1991, 11:4036-4044.
- Hammond GL, Smith CL, Goping IS, Underhill DA, Harley MJ, Reventos J, Musto NA, Gunsalus GL, Bardin CW: **Primary structure of human corticosteroid binding globulin, deduced from hepatic and pulmonary cDNAs, exhibits homology with serine protease inhibitors.** *Proc Natl Acad Sci U S A* 1987, 84:5153-5157.
- Irving JA, Steenbakkers PJ, Lesk AM, Op den Camp HJ, Pike RN, Whisstock JC: **Serpins in prokaryotes.** *Mol Biol Evol* 2002, 19:1881-1890.
- Carrell RW, Owen MC: **Plakalbumin, alpha 1-antitrypsin, antithrombin and the mechanism of inflammatory thrombosis.** *Nature* 1985, 317:730-732.
- Stein PE, Leslie AG, Finch JT, Turnell WG, McLaughlin PJ, Carrell RW: **Crystal structure of ovalbumin as a model for the reactive centre of serpins.** *Nature* 1990, 347:99-102.
- Stratikos E, Gettins PG: **Formation of the covalent serpin-proteinase complex involves translocation of the proteinase by more than 70 Å and full insertion of the reactive center loop into beta-sheet A.** *Proc Natl Acad Sci U S A* 1999, 96:4808-4813.
- Huntington JA, Read RJ, Carrell RW: **Structure of a serpin-protease complex shows inhibition by deformation.** *Nature* 2000, 407:923-926.
- Remold-O'Donnell E: **The ovalbumin family of serpin proteins.** *FEBS Lett* 1993, 315:105-108.
- Silverman GA, Whisstock JC, Askew DJ, Pak SC, Luke CJ, Cataltepe S, Irving JA, Bird PI: **Human clade B serpins (ov-serpins) belong to a cohort of evolutionarily dispersed intracellular proteinase inhibitor clades that protect cells from promiscuous proteolysis.** *Cell Mol Life Sci* 2004, 61:301-325.
- Bartuski AJ, Kamachi Y, Schick C, Overhauser J, Silverman GA: **Cytoplasmic antiproteinase 2 (PI8) and bomapin (PI10) map to the serpin cluster at 18q21.3.** *Genomics* 1997, 43:321-328.
- Scott FL, Eyre HJ, Lioumi M, Ragoussis J, Irving JA, Sutherland GA, Bird PI: **Human ovalbumin serpin evolution: phylogenetic analysis, gene organization, and identification of new PI8-related genes suggest that two interchromosomal and several intrachromosomal duplications generated the gene clusters at 18q21-q23 and 6p25.** *Genomics* 1999, 62:490-499.
- Benarafa C, Remold-O'Donnell E: **The ovalbumin serpins revisited: Perspective from the chicken genome of clade B serpin evolution in vertebrates.** *Proc Natl Acad Sci U S A* 2005, 102:11367-11372.
- Heilig R, Perrin F, Gannon F, Mandel JL, Chambon P: **The ovalbumin gene family: structure of the X gene and evolution of duplicated split genes.** *Cell* 1980, 20:625-637.
- Kaiserman D, Knaggs S, Scarff KL, Gillard A, Mirza G, Cadman M, McKeone R, Denny P, Cooley J, Benarafa C, Remold-O'Donnell E, Ragoussis J, Bird PI: **Comparison of human chromosome 6p25 with mouse chromosome 13 reveals a greatly expanded ov-serpin gene repertoire in the mouse.** *Genomics* 2002, 79:349-362.
- Puente XS, Lopez-Otin C: **A genomic analysis of rat proteases and protease inhibitors.** *Genome Res* 2004, 14:609-622.
- Askew DJ, Askew YS, Kato Y, Turner RF, Dewar K, Lehoczyk J, Silverman GA: **Comparative genomic analysis of the clade B serpin cluster at human chromosome 18q21: amplification within the mouse squamous cell carcinoma antigen gene locus.** *Genomics* 2004, 84:176-184.
- Hedges SB: **The origin and evolution of model organisms.** *Nat Rev Genet* 2002, 3:838-849.
- Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthonard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapelle C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H: **Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype.** *Nature* 2004, 431:946-957.
- Dickinson JL, Norris BJ, Jensen PH, Antal TM: **The C-D interhelical domain of the serpin plasminogen activator inhibitor-type 2 is required for protection from TNF-alpha induced apoptosis.** *Cell Death Differ* 1998, 5:163-171.
- Chuang TL, Schleef RR: **Identification of a nuclear targeting domain in the insertion between helices C and D in protease inhibitor-10.** *J Biol Chem* 1999, 274:11194-11198.
- Wilczynska M, Lobov S, Ohlsson PI, Ny T: **A redox-sensitive loop regulates plasminogen activator inhibitor type 2 (PAI-2) polymerization.** *Embo J* 2003, 22:1753-1761.
- Lobov S, Wilczynska M, Bergstrom F, Johansson LB, Ny T: **Structural bases of the redox-dependent conformational switch in the serpin PAI-2.** *J Mol Biol* 2004, 344:1359-1368.
- Jensen PH, Schuler E, Woodrow G, Richardson M, Goss N, Hojrup P, Petersen TE, Rasmussen LK: **A unique interhelical insertion in**

- plasminogen activator inhibitor-2 contains three glutamines, Gln83, Gln84, Gln86, essential for transglutaminase-mediated cross-linking.** *J Biol Chem* 1994, **269**:15394-15398.
28. Ritchie H, Lawrie LC, Crombie PW, Mosesson MW, Booth NA: **Cross-linking of plasminogen activator inhibitor 2 and alpha 2-antiplasmin to fibrin(ogen).** *J Biol Chem* 2000, **275**:24915-24920.
 29. Darnell GA, Antalis TM, Johnstone RW, Stringer BW, Ogbourne SM, Harrich D, Suhrbier A: **Inhibition of retinoblastoma protein degradation by interaction with the serpin plasminogen activator inhibitor 2 via a novel consensus motif.** *Mol Cell Biol* 2003, **23**:6520-6532.
 30. Ragg H, Lokot T, Kamp PB, Atchley WR, Dress A: **Vertebrate serpins: construction of a conflict-free phylogeny by combining exon-intron and diagnostic site analyses.** *Mol Biol Evol* 2001, **18**:577-584.
 31. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MA, 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 SP, 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 RP, 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 MD, Davies RM, Francis MD, Humphray SJ, Scott CE, Taylor RG, Tickle C, Brown WR, Rogers J, Buerstedde JM, Wilson SA, Stubbs L, Ovcharenko I, Gordon L, Lucas S, Miller MM, Inoko H, Shiina T, Kaufman J, Salomonsen J, Skjoedt K, Wong GK, 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 SM, 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, Betran E, Dupanloup I, Kaessmann H, Hinrichs AS, Bejerano G, Furey TS, Harte RA, Raney B, Siepel A, Kent WJ, Haussler D, Eyas 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: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.** *Nature* 2004, **432**:695-716.
 32. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickinson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordtsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
 33. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoeve F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S: **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*.** *Science* 2002, **297**:1301-1310.
 34. The Chimpanzee Sequencing and Analysis Consortium: **Initial sequence of the chimpanzee genome and comparison with the human genome.** *Nature* 2005, **437**:69-87.
 35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
 36. Burge C, Karlin S: **Prediction of complete gene structures in human genomic DNA.** *J Mol Biol* 1997, **268**:78-94.
 37. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
 38. Page RD: **TreeView: an application to display phylogenetic trees on personal computers.** *Comput Appl Biosci* 1996, **12**:357-358.

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

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

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

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

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

