BMC Evolutionary Biology

BioMed Central

Research article

Open Access

Functional evolution of ADAMTS genes: Evidence from analyses of phylogeny and gene organization

Ainsley C Nicholson†[1], Shehre-Banoo Malik†[2], John M Logsdon Jr[2] and Erwin G Van Meir*[1]

Address: [1]Laboratory of Molecular Neuro-Oncology, Neurosurgery Department and Winship Cancer Institute, 1365-C Clifton Road, Room C5078, Emory University, Atlanta GA 30322 USA and [2]Roy J. Carver Center for Comparative Genomics, Department of Biological Sciences, 300 Old Biology Building, University of Iowa, Iowa City IA 52242-1324 USA

Email: Ainsley C Nicholson - agn0@cdc.gov; Shehre-Banoo Malik - banoo-malik@uiowa.edu; John M Logsdon - john-logsdon@uiowa.edu; Erwin G Van Meir* - evanmei@emory.edu

* Corresponding author †Equal contributors

Published: 04 February 2005

BMC Evolutionary Biology 2005, **5**:11 doi:10.1186/1471-2148-5-11

Received: 08 September 2004
Accepted: 04 February 2005

This article is available from: http://www.biomedcentral.com/1471-2148/5/11

© 2005 Nicholson et al; licensee BioMed Central Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The ADAMTS (A Disintegrin-like and Metalloprotease with Thrombospondin motifs) proteins are a family of metalloproteases with sequence similarity to the ADAM proteases, that contain the thrombospondin type 1 sequence repeat motifs (TSRs) common to extracellular matrix proteins. ADAMTS proteins have recently gained attention with the discovery of their role in a variety of diseases, including tissue and blood disorders, cancer, osteoarthritis, Alzheimer's and the genetic syndromes Weill-Marchesani syndrome (ADAMTS10), thrombotic thrombocytopenic purpura (ADAMTS13), and Ehlers-Danlos syndrome type VIIC (ADAMTS2) in humans and belted white-spotting mutation in mice (*ADAMTS20*).

Results: Phylogenetic analysis and comparison of the exon/intron organization of vertebrate (*Homo*, *Mus*, *Fugu*), chordate (*Ciona*) and invertebrate (*Drosophila* and *Caenorhabditis*) ADAMTS homologs has elucidated the evolutionary relationships of this important gene family, which comprises 19 members in humans.

Conclusions: The evolutionary history of *ADAMTS* genes in vertebrate genomes has been marked by rampant gene duplication, including a retrotransposition that gave rise to a distinct *ADAMTS* subfamily (*ADAMTS1*, *-4*, *-5*, *-8*, *-15*) that may have distinct aggrecanase and angiogenesis functions.

Background

ADAMTS (A Disintegrin-like and Metalloprotease with Thrombospondin motifs) proteins have homology with the metalloprotease region of the ADAM proteases, but also have at least one of the Thrombospondin type 1 Sequence Repeat (TSR) motifs that are common in extracellular matrix proteins. Since the discovery of a gene encoding ADAMTS1 in 1997 [1], a total of 19 similar genes have been found in the human genome [2], numbered *ADAMTS1-20*; there is no *ADAMTS11* because early reports of an *ADAMTS11* [3] were later found to describe *ADAMTS5*. Many of these genes have been implicated in a variety of diseases, including connective tissue disorders [4], cancer [5-7], osteoarthritis [3,8], and possibly Alzheimer's disease [6,9]. Recently, an autosomal recessive form of Weill-Marchesani syndrome (WMS) has been attributed to null mutations of the *ADAMTS10* gene [10]. The symptoms characteristic of this syndrome (short

stature, brachydactyly, joint stiffness, and anomalies of the eye lenses), together with its expression patterns, suggest a role for the gene encoded by this protein in normal growth and in skin, eye, and heart development.

ADAMTS proteins are characterized by a pro-domain, a metalloprotease domain, the so-called disintegrin-like and spacer domains, and a tail of TSR repeats. The pro-domain of ADAMTS1 and -4 is cleaved at the RX(K/R)R furin cleavage site [11] in the Golgi [12,13], releasing an active protein [14]. There are clearly conserved furin cleavage sites for most human ADAMTS proteins (positions 578–581 of the alignment) [Additional File 2]. While this site was less well conserved in ADAMTS10 and ADAMTS12, the pro-domain of ADAMTS12 was also shown to be removed by a furin-mediated process [7]. On the basis of this combined evidence, it is commonly believed that furin cleavage of the pro-domain might occur for all ADAMTS proteins.

The metalloprotease domain of ADAMTS proteins is shared with the related ADAM proteins, and the catalytic Zn2+-binding motif HEXGHXXXXXHD [15] is well conserved, shown at amino acid positions 761–772 [Additional File 2]. While the metalloprotease domain of ADAM proteins is followed by a disintegrin domain which binds integrins at a conserved X(D/E)ECD site [16,17], the corresponding amino acids in the disintegrin-like domain of ADAMTS proteins are not well conserved. We also note that the so-called spacer domain following this disintegrin-like domain (amino acids 1060–1400) [Additional File 2] in fact has many highly conserved residues, despite its comparatively reduced overall conservation of amino acid sequence.

There are four matrix metalloprotease (MMP) cleavage sites in the spacer domain of ADAMTS1 [14,18], including the highly conserved IPAGA site at amino acid positions 1229–1233 [Additional File 2] (*L. Iruela-Arispe, personal communication*). Further proteolytic processing within this domain has been demonstrated for ADAMTS1, -2, -5, and -12 [3,6,7,19]. For ADAMTS1, this second proteolytic step is mediated by several MMPs, and results in removal of the C-terminal TSRs that interact with the extracellular matrix (ECM). This leads to release of the protein from the endothelial cell membrane, reducing its ability to inhibit endothelial cell proliferation and probably reducing its anti-angiogenic potential [14]. Release of ECM-bound proteins *via* proteolytic removal of their TSR domains may be a common theme, as we see similar proteolytic removal of the C-terminal TSRs of the unrelated neuronal guidance protein F-spondin by plasmin, releasing it from ECM binding [20]. While the exact mechanism of the proteolytic processing of ADAMTS proteins remains somewhat controversial, there is an intriguing possibility that regulation of the ratio of ECM-bound *vs*. free ADAMTS protein could be mediated by MMPs. The region containing these sites is conserved to varying degrees in the newly discovered ADAMTS proteins, suggesting variable (perhaps tissue-specific) MMP processing of these proteins. ADAMTS4, which lacks a TSR tail, may not have an ECM-bound form.

The region between the metalloprotease domain and the TSR repeat tail was demonstrated to be necessary for *gon-1*, a *Caenorhabditis ADAMTS* homolog, to mediate cell migration during gonadogenesis [21]. A variant of this region that lacks the conserved amino acids upstream of the classic TSR but maintains the highly conserved spacer residues is found in papilin, where it has been implicated in influencing cell rearrangements during organogenesis [22] and in the *Manduca sexta* lacunin protein which plays a role in basal lamina remodeling during morphogenesis [23]. It will be interesting to investigate whether there is to be a common theme of organogenesis function among proteins that contain this configuration of domains.

There is evidence that several mammalian ADAMTS proteins are expressed organogenesis. For example, mutations in the mouse *ADAMTS20* gene have been found to cause the belted white-spotting mutation, resulting from a defect in melanocyte development or migration during embryogenesis [29], the ADAMTS1 protein is necessary for mouse gonadogenesis [30], ADAMTS12 is specifically expressed in fetal lung [7], and several of the newly described ADAMTS proteins [28] are expressed solely or primarily in fetal tissue.

ADAMTS proteins contain a single "classic" TSR motif (WXXWXXW) in the disintegrin-like domain, and a variable number of variant TSRs within the C-terminal tail of the protein, which contain 4 amino acids between tryptophan residues (W4XW) rather than two. TSRs can be divided in several structural groups, based on the presence and spacing of cysteines [24]. The precise function of each type of TSR has not yet been determined, although it is known that the sequence CSVTCG in one of the thrombospondin-1 TSR's mediates endothelial cell apoptosis through binding to CD36 [25,26]. About 70 proteins in the human genome contain TSRs [27] and many of them are matrix binding proteins.

ADAMTS1 and -8 inhibit angiogenesis [31], and gene expression profiling suggests that ADAMTS4 also has a role in angiogenesis [32]. Several of these proteins (ADAMTS1, -4, and -5) have also been shown to cleave aggrecan, the proteoglycan that makes cartilage elastic and compressible [19,33,34], and ADAMTS4 was recently shown to cleave cartilage oligomeric matrix protein (TSP5) [35]. The ADAMTS2, -3, and -14 proteins appear

functionally related. ADAMTS3 is a procollagen II N-propeptidase, and ADAMTS14 appears also to be an aminoprocollagen peptidase [36]. ADAMTS2 is an aminoprocollagen peptidase of procollagen I and II, and deficiency of this protein causes Ehlers-Danlos syndrome type VIIC [4]. ADAMTS13 is a von Willebrand factor-cleaving protease. Mutations in the *ADAMTS13* gene result in inappropriate platelet activation, leading to the blood disorder thrombotic thrombocytopenic purpura (TTP) [37-40].

Recently, an intriguing link has been discovered between ADAMTS metalloproteases. The proinflammatory cytokines IL17 [41], IL1β [42] and TGFβ [43] induce expression of *ADAMTS4*. TGFβ also induces *ADAMTS2* [44] and TNFα was found to up-regulate *ADAMTS1*, *ADAMTS6*, and *ADAMTS9* in ocular cells [45], suggesting a role for these proteases in inflammatory eye disease. Similarly, TNFα produced a marked induction of *ADAMTS1* in endothelial cells [46]. As several ADAMTS proteins, ADAMTS4 in particular, are implicated in rheumatoid arthritis, and TNFα inhibitors have been recently been used with great success in its treatment [47], we speculate that part of the effect of the TNFα inhibitors is an indirect downregulation of the ADAMTS proteins that break down connective tissues. As TNFα inhibitors are not without inherent risks [48,49], transcriptional inhibition of specific *ADAMTS* genes may ultimately provide similar benefits with fewer risks.

To better understand the multiple functions of the ADAMTS proteins, we carried out the most detailed and comprehensive analysis to date of the phylogenetic relatedness and intron/exon organization of all human *ADAMTS* genes, including their comparison with invertebrate and chordate *ADAMTS* homologs. Prior analyses included fewer species and did not address the sequence of genomic events that resulted in the current *ADAMTS* genomic structure [2,28,50]. Our analysis reveals distinct sub-families with unique functions and reveals a history of gene duplications, retrotransposition, and the loss and gain of introns during animal evolution. For example, *ADAMTS1*, *-4*, *-5*, *-8*, and *-15* genes all derive from a retrotransposition event that occurred prior to the divergence of vertebrates and the urochordate *Ciona intestinalis*, and from subsequent gene duplications that occurred prior to the divergence of mammals and the pufferfish, *Fugu rubripes*. This ADAMTS protein subfamily encompasses proteins that share aggrecanase and angiogenesis-related activities.

Results & Discussion

Gene discovery

To perform a phylogenetic analysis of the ADAMTS gene family we first examined several genome databases to have a comprehensive survey of all ADAMTS genes in humans and other species. When we started this analysis in 2002 only 12 ADAMTS proteins were known. We predicted coding sequences of the human *ADAMTS* genes based on several sequence databases (see Material and Methods) and found them to be identical to those subsequently published by Cal, *et al* [28] based on cloned cDNAs. This work confirmed that a total of 19 ADAMTS genes exist in the human genome, with various isoforms. Since there were several cases in which alternative splicing or different exon predictions resulted in ADAMTS proteins of varying lengths, the most complete (longest) translations were considered in our analyses, and their Genbank accession numbers are indicated [Additional File 2]. *ADAMTS9* has three known splice variants, of which the long variant that we used for analysis was NM_182920. *ADAMTS20* has two known splice variants, of which we used NM_025003. *ADAMTS18* has two known variants, of which we used NM_199355, minus the final exon. *ADAMTS13* has four known variants, of which we used NM_139025.

ADAMTS10 and *ADAMTS6* each have a single known coding sequence, and we found evidence of others. The variant of *ADAMTS6* that is published (NM_014273) is a short form, and contains a non-consensus exon immediately following the metalloprotease catalytic site, while the variant of *ADAMTS10* that is published (NM_030957) has a consensus exon at this location, and is in the long form. We predict a long form of *ADAMTS6*, as well as a consensus exon for *ADAMTS6* and a non-consensus exon for *ADAMTS10*. We used the consensus exons of both genes in our analyses.

Phylogenetic analyses

We compared the 19 known human ADAMTS protein sequences with ADAMTS homologs from invertebrates (*Drosophila* and *Caenorhabditis*) from which entire genome sequences were recently determined, to elucidate their evolutionary relationships (Figure 1A and 1B) [Additional File 1]. This revealed a series of gene duplications among human *ADAMTS* genes, of uncertain affinity to these invertebrate relatives. With the goal to further elucidate this gene duplication history, the human and invertebrate *ADAMTS* orthologs were compared with *ADAMTS* orthologs from *Mus*, *Fugu* and *Ciona*, which diverged between humans and invertebrates, and with an additional invertebrate, the honeybee, *Apis mellifera* (Figure 2) [Additional File 2].

To perform this analysis we inferred the coding sequence of sixteen ADAMTS proteins in the *Fugu rubripes* genome [52], and drew from Genbank nine representative mouse orthologs, three orthologs from the *Drosophila* genome and four from *Caenorhabditis* (see Figures 1A and 2). Other *Mus* (and *Rattus*) *ADAMTS* orthologs that were

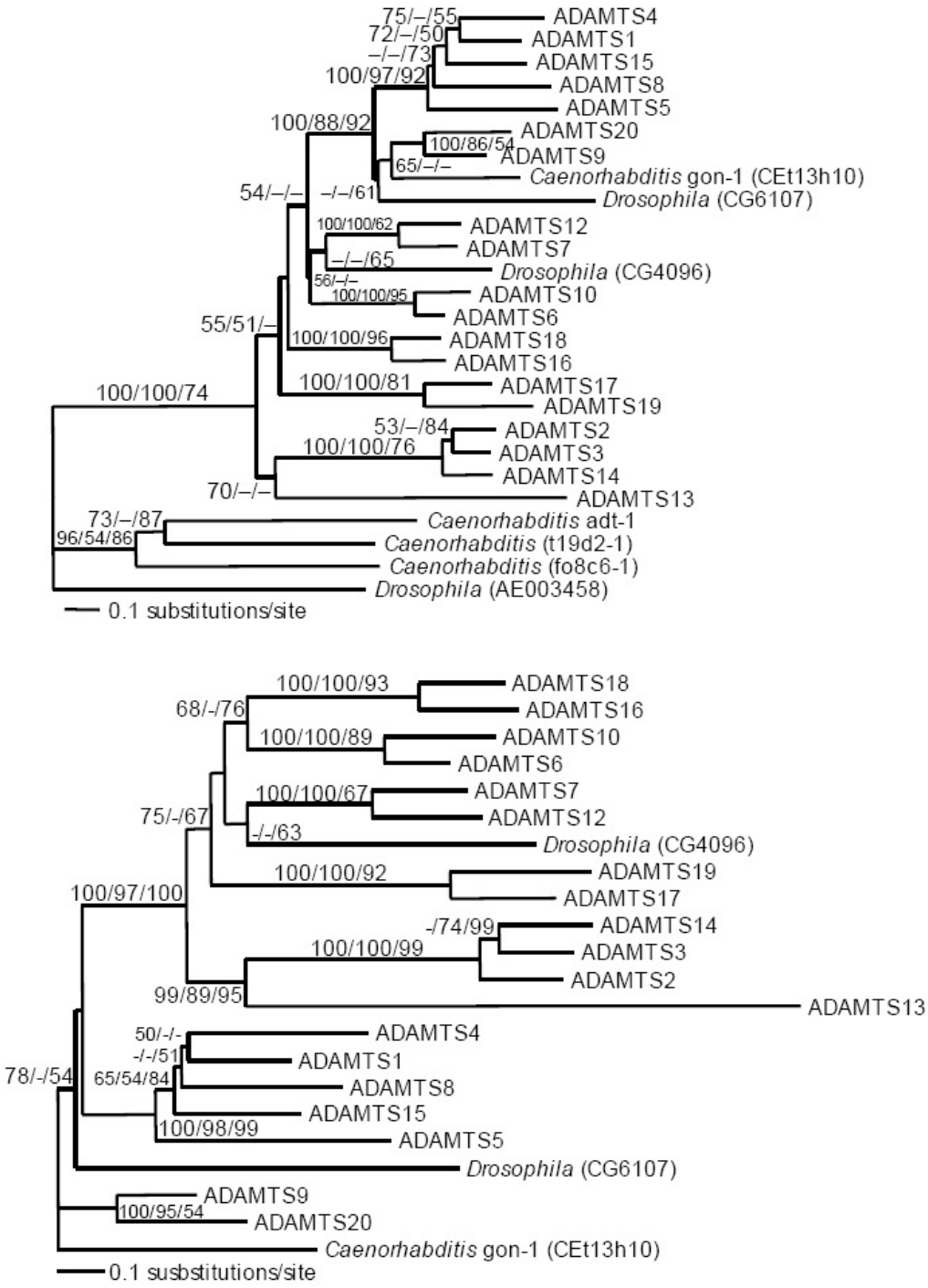

Figure 1

Phylogenetic analysis of human ADAMTS proteins and invertebrate homologs. Unambiguously aligned amino acids were analyzed by distance (Protdist+NJ), maximum parsimony (MP) and maximum likelihood (ML) methods. The trees shown are the ML distance topologies. Numbers at the nodes represent the percent of bootstrap replicates of distance (NJ) and parsimony (MP), and the percent of quartet puzzling steps (QP) in support of each group. **(A) Phylogenetic tree of human and distantly related invertebrate ADAMTS homologs** inferred from a 359-amino acid alignment, with α = 1.42 and proportion of invariable sites (pI) = 0.09. **(B) Phylogeny of human and invertebrate ADAMTS homologs with long branches removed,** inferred from 543 aligned amino acids, with α = 1.48 and proportion of invariable sites (pI) = 0.10. For reference, Genbank GI numbers for the sequences are provided [Additional File 2].

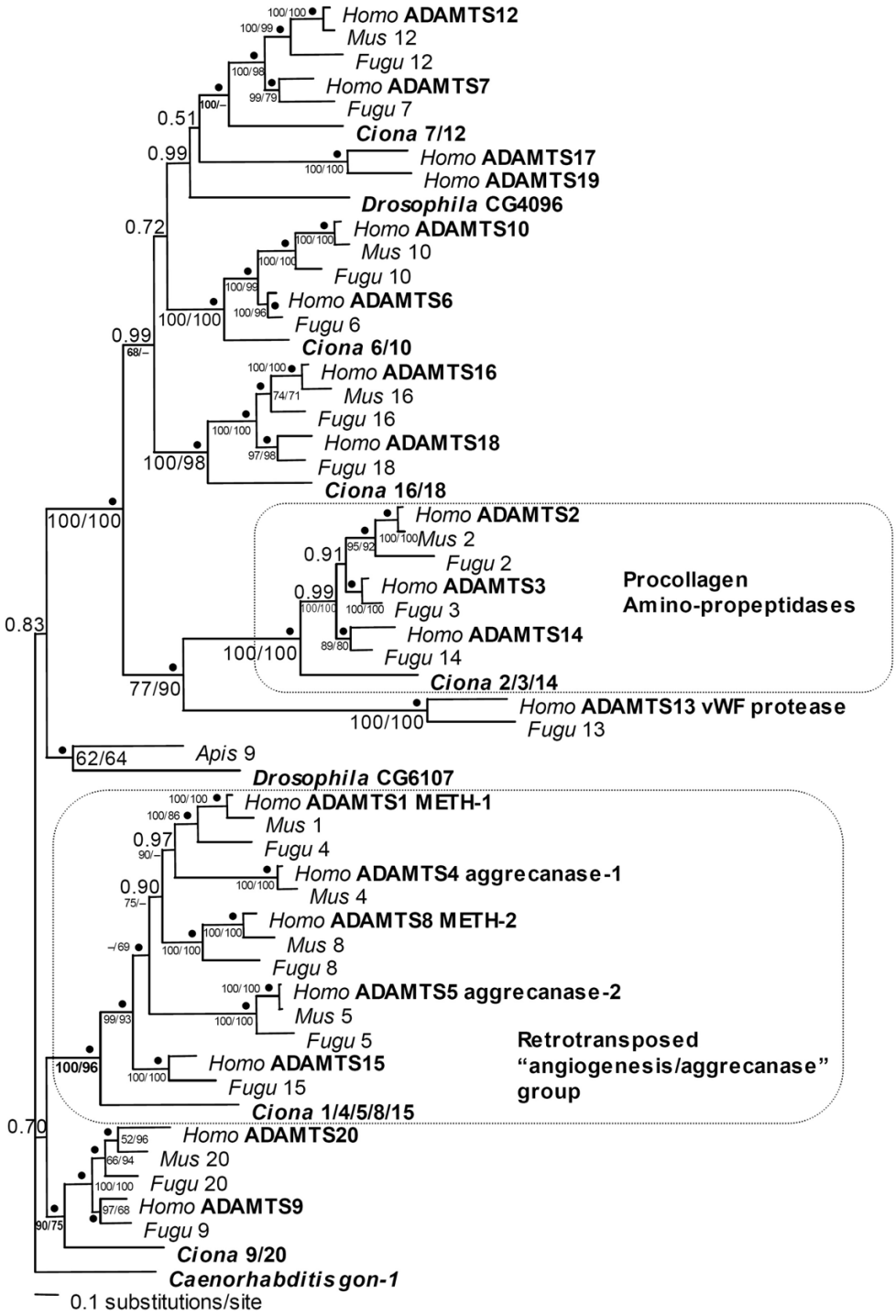

Figure 2

Phylogenetic analysis of animal ADAMTS homologs. This is the consensus maximum likelihood tree topology determined from 900 trees with the highest posterior probabilities inferred by Bayesian analysis of protein sequences. 571 aligned amino-acid sites were analyzed, with mean α = 1.59 (1.38 < α < 1.83), pI = 0.10 (0.07 < pI < 0.13) and lnL = -37875.26. Numbers at nodes represent Bayesian posterior probabilities for that relationship, with the best-supported posterior probabilities (1.00) represented by bullets (•). The percent of 1000 bootstrap replicates in support of the nodes, as found by distance and parsimony analyses, are also reported. Accession numbers and scaffold numbers for sequences are provided [Additional File 2].

unannotated at the time of our initial work have since been identified by others [51,53], and some are annotated in the MEROPS database[82]. However, they were not included for our final analyses presented here since they are so similar to the human sequences (69–99% identical, [53]) that they offered no help in elucidating the gene duplication history. The invertebrate genomes were surveyed extensively for additional *ADAMTS* genes, and those most closely related to the vertebrate *ADAMTS* orthologs were retained in the analyses presented herein. The most divergent *Drosophila* and *Caenorhabditis* ADAMTS homologs represented in Figure 1A were removed from further analyses in attempt to avoid systematic bias known as "long branch attraction" where divergent but putatively unrelated sequences group together because of their divergence rather than due to shared characters [54]. All ADAMTS proteins introduced here contain the same basic domain structure as previously described ADAMTS proteins. A complete alignment of all human and invertebrate ADAMTS protein sequences, and representative *Ciona*, *Fugu* and *Mus* orthologs, annotated with intron positions and phases, is available [Additional File 2].

Intron position and phase

We compared the positions of introns and their phases between *Homo*, *Fugu*, *Ciona*, *Drosophila* and *Caenorhabditis* homologs of ADAMTS genes, in attempt to corroborate and further elucidate their evolutionary relationships, as shown in Figure 3. The term intron phase refers to the position of the splice site with respect to the codon, where phase 0 describes a splice site 5' of the codon, phase 1 describes a splice site between the first and second base of a codon, and phase 2 describes a splice site between the second and third base of a codon. Introns at the same position that have the same phase in homologous genes are considered to be shared characters that were conserved during evolution. The lack of an intron at a conserved position may either suggest that the gene lost an intron at that position during its evolution, or that it never had that intron, and the intron conserved in the other homologs was gained after those homologs diverged from the intron-lacking homolog. In combination with the phylogenetic analysis of the ADAMTS protein sequences, a parsimonious interpretation of the data summarized in Figure 3 that invokes the fewest changes should help to distinguish between older and more recent gene duplication events in this gene family. Three of our most striking observations of the intron distribution are that (i) some intron positions are shared between worm, fly, chordate and vertebrates, (ii) recently duplicated genes share similar patterns of introns, and that (iii) the complete absence of ancient introns and the presence of introns at new positions in *ADAMTS1*, *4*, *5*, *8* and *15* of vertebrates and *Ciona* reveals that this subgroup of genes evolved by retrotransposition early during chordate evolution, was repopulated by new introns (in some cases, separately in vertebrates and *Ciona*), and subsequently underwent gene duplication during the evolution of vertebrates.

The phylogenetic analysis of animal ADAMTS homologs reveals that proteins that are known to have similar biological activities are closely related, and that they have evolved by a series of gene duplication events (Figure 2). Since the functions of only some ADAMTS proteins have been empirically tested, estimates of evolutionary relatedness amongst the entire family may imply closer functional relatedness, and thus guide the future design of more specific empirical tests of protein functions. An interesting property of the vertebrate ADAMTS proteins are the extensive sequence similarity between many pairs of sequences, as indicated in Figures 1, 2, 3, 4 [and Additional file 2]. Although in many cases little is known about the functions of these proteins, we can speculate that the two proteins in each pair may share similar biological activities due to their shared primary sequence. It is also possible that these ADAMTS proteins act as heterodimers, in a manner similar to the ADAM proteins fertilin α and β [55].

As shown in Figures 1 and 2 and summarized in Figure 4, animal *ADAMTS* homologs have undergone several gene duplications. Assuming that the Ecdysozoa hypothesis is true and arthropods and nematodes are united as a group [56-62], our results indicate that a single *ADAMTS* gene duplication preceded the divergences of Ecdysozoa and chordates. At least 3–4 *ADAMTS* gene duplications occurred in chordates prior to the divergence of *Ciona* and vertebrates, followed by additional gene duplications in vertebrates prior to the divergence of *Fugu* and mammals (Figures 2 and 4).

Ciona intestinalis, the urochordate sea squirt, was found to have at least six *ADAMTS* genes (Figure 2), which correspond to six of the seven major groups of vertebrate ADAMTS homologs. *Ciona* ADAMTS6 is the sister of the group comprised of vertebrate ADAMTS6 and -10, indicating that ADAMTS6 and -10 evolved by gene duplication early during vertebrate evolution, preceding the divergence of pufferfish and mammals, but after their divergence from urochordates. Similarly, *Ciona* ADAMTS16 is a sister to the group comprised of vertebrate ADAMTS16 and -18, *Ciona* ADAMTS7 is a sister to the group comprised of vertebrate ADAMTS7 and -12, *Ciona* ADAMTS3 is a sister to the group comprised of vertebrate ADAMTS2, -3 and -14, *Ciona* ADAMTS9 is a sister to the group comprised of vertebrate ADAMTS9 and -20 and *Ciona* ADAMTS15 is the sister to the group comprised of vertebrate ADAMTS1, -4, -5, -8 and -15. This reveals that both gene duplications early in chordate evolution as well as

Distribution and phases of intron positions across *ADAMTS* genes in *Homo, Fugu, Ciona, Drosophila* and *Caenorhabditis*. Intron positions are numbered in the order in which they appear from 5' > 3' ends of the *ADAMTS* coding region. Introns when present at these positions are marked as cells in the table, with the numbers 0, 1 or 2 marking the phase of the intron at that position. The absence of an intron is indicated by "–" and missing information indicated by "?".

| | | | # | | | | | | | * | | | | | | |

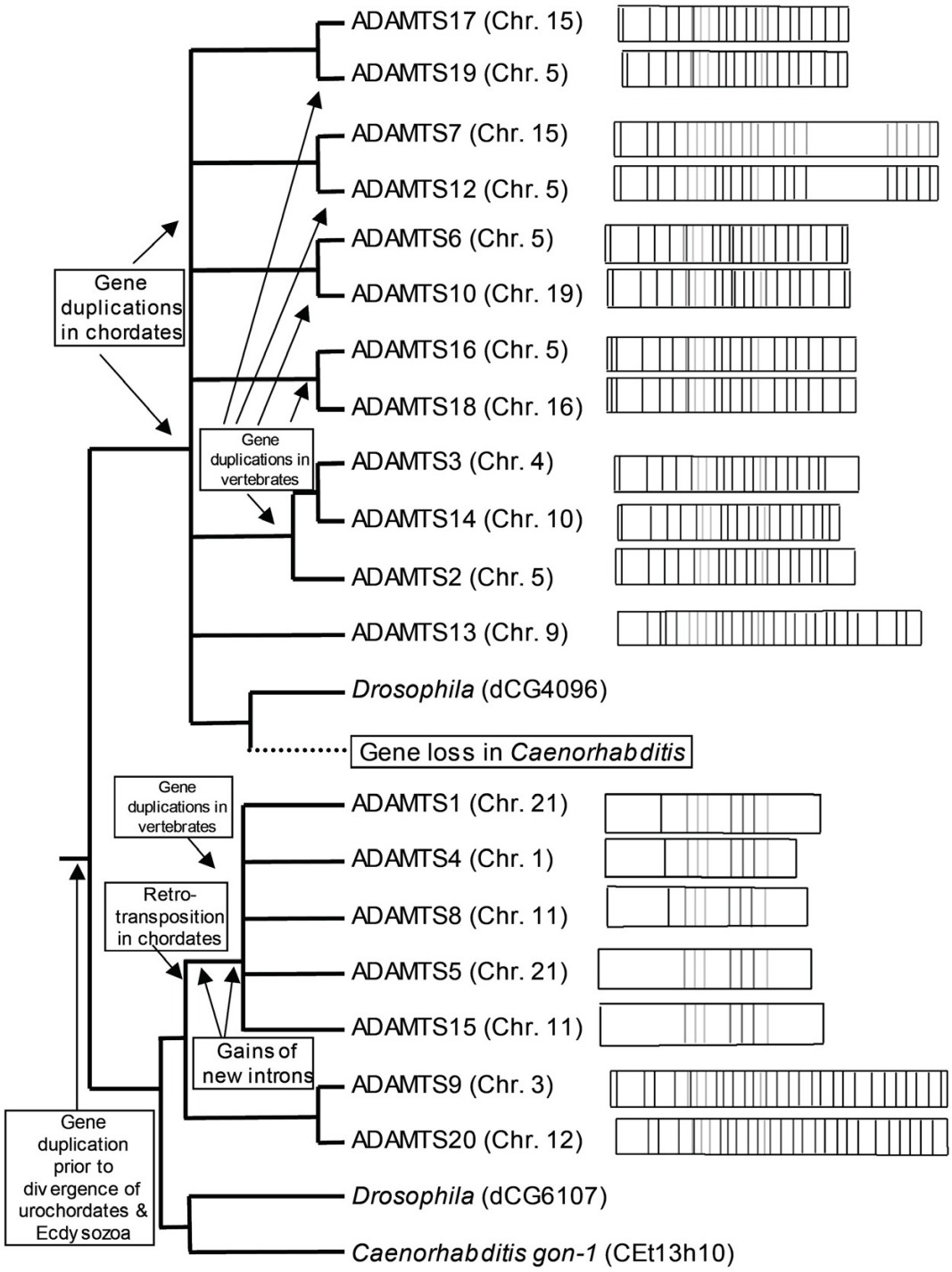

Figure 4

Proposed scenario for the evolutionary history of ADAMTS proteins. During chordate evolution a series of gene duplications resulted in six ADAMTS proteins present in the *Ciona* genome, while an early retrotransposition event gave rise to the "angiogenesis clade" of ADAMTS proteins. This proliferation of ADAMTS proteins did not occur in invertebrates, and there is evidence of the loss of one ADAMTS ortholog from *Caenorhabditis*. More recent duplications that occurred early during vertebrate evolution resulted in the paired sets of ADAMTS proteins present in the human genome. The chromosomal locations of the human ADAMTS genes are indicated in parentheses and the exon structure of each human gene is diagrammed to the right of its position in the schematic phylogenetic tree, and shown in more detail in the Additional Files.

that of the mouse *ADAMTS1* gene [63], and our analysis shows four *ADAMTS* genes with this characteristic gene structure in the genome of *F. rubripes*. Therefore, retrotransposition of an ancestor of the angiogenesis/aggrecanase subfamily of genes, its acquisition of new introns, and subsequent gene duplications that produced five related genes occurred prior to the divergence of human, mouse and pufferfish lineages. In at least one case (intron 17 in *ADAMTS8*) we see evidence of acquisition of a new intron following the process of duplication, but prior to the divergence of mammals and pufferfish.

ADAMTS20 and *-9* are most closely related and, along with the members of the angiogenesis/aggrecanase clade, are most closely related to invertebrate *gon-1* (*Caenorhabditis*) and CG6107 (*Drosophila*). The finding that *ADAMTS9* and *-20* together as a group have a single sister gene in *Ciona* indicates that they evolved by gene duplication in the vertebrate lineage, after their divergence from urochordates. However, the results of our phylogenetic analyses demonstrate that neither *ADAMTS9* nor *ADAMTS20* can alone be rightfully dubbed as being orthologous to *gon-1*, as has been recently proposed [64,65]. In fact, our analyses suggest that while *gon-1* and CG6107 are likely orthologs, the chordate ortholog of these genes was the common ancestor of *Ciona ADAMTS9* and *-15*, *i.e.* also the common ancestor of the later-duplicated vertebrate ADAMTS genes *ADAMTS9*, *-20*, *-15*, *-5*, *-8*, *-4*, and *-1* (Figures 1 and 2).

Only one invertebrate sequence (*Drosophila* CG4096) was found that grouped with the remainder of the human *ADAMTS* homologs. The placement of this gene with or within a group of these remaining *ADAMTS* genes would suggest the number of gene duplication events in this gene family that occurred prior to the divergence of vertebrates and invertebrates from a common ancestor. The Kishino-Hasegawa test revealed that the likelihood of *Drosophila* CG4096 being most closely related to an ancestor of the group of all remaining mammalian ADAMTS proteins, or of the various groups nested within it, was not significantly different from the likelihood that *Drosophila* CG4096 is most closely related to *ADAMTS 7* and *-12* (data not shown). The most parsimonious explanation of this result is that a single gene duplication of an ancient ADAMTS homolog occurred early during the evolution of animals, prior to the divergence of chordates from invertebrates, followed by lineage-specific gene loss and gene duplications (Figure 4). If this scenario is correct, the *Caenorhabditis* ortholog of *Drosophila* CG4096 has been subsequently lost, but the vertebrate ortholog has been retained and underwent several gene duplications within and among chromosomes in the vertebrate lineage (*ADAMTS2*, *-3*, *-6*, *-7*, *-10*, *12–14*, *16–19*).

The exons comprising the TSR tail each contain a single variant TSR, with the C(S/T)XCG motif 5' and the W4XW motif at the 3' end. This exon structure may facilitate the formation of alternately spliced isoforms, such as we describe here, but it would also lend itself to duplication or loss of individual repeats. However, the number of variant TSRs in the tails of these proteins has been conserved for the gene pairs *ADAMTS17* and *-19*, *ADAMTS6* and *-10*, and *ADAMTS18* and *-16* (Figure 4). While this may suggest a series of relatively recent gene duplications, a more likely explanation is that each TSR has an important and non-redundant role, or that the presence of a specific number of TSRs is critical for each protein's function.

The apparent absence of any ortholog of *ADAMTS17* or *-19* in the pufferfish genome (Figure 2), but their presence in *Mus* and *Rattus* [51] suggests that this gene duplication either occurred in mammals after they diverged from fish, or that *ADAMTS17* or *-19* evolved earlier than the mammal/fish divergence but were lost in *Fugu*.

Gene duplication is a common way by which new genes with similar functions may evolve. In fact, duplication of large segments of chromosomes has been a common occurrence during animal evolution [66,67]. Both the phylogenetic trees and the intron/exon structures of *ADAMTS* genes show a history of such duplications. In addition to the similarity in intron positions of *ADAMTS-1*, *-4*, *-5*, *-8*, and *-15* in both mouse and pufferfish, human *ADAMTS1* and *-5* are located in tandem on chromosome 21, and human *ADAMTS8* and *-15*, on chromosome 11 (Figure 3). The remaining genes have additional introns at conserved locations, and both proteins in each set of pairs have the same intron/exon structure.

Thus, by combining the phylogenetic analysis and intron/exon structure determinations, we are able to propose the following series of events leading to the ADAMTS protein family (Figure 4): (i) The ancestral *ADAMTS* gene duplicated prior to the divergence of the ecdysozoan and chordate lineages, approximately 673 million years ago [68]. (ii) In the following approximately 250 million years prior to divergence of fish and mammals [69], multiple gene duplications occurred. (iii) A retrotransposition of the common ancestor of the *ADAMTS9* and *-20* gene pair resulted in an intronless gene that proceeded to gain multiple introns, giving rise to the angiogenesis/aggrecanase clade. (iv) This gene was involved in a duplicative chromosomal inversion, and later a duplication of the chromosomal segment containing both *ADAMTS* genes. (v) Another intron was gained, in *ADAMTS8*, prior to the divergence of the pufferfish and mammalian lineages. In the other branch of the *ADAMTS* family, we see a remarkable frequency of genes located on chromosome 5, suggesting it as the location of the ancestor of these genes. We

can speculate that a similar scenario of within-chromosome duplication followed by duplication of chromosomal segments took place, although none are in as close physical proximity as the *ADAMTS1*-subfamily genes.

Conclusions

This comprehensive bioinformatic survey of the human genome affirms the widely held belief, derived from experimental work, that the nineteen known human ADAMTS proteins constitute the complete gene family. By examining both the amino acid sequences using rigorous phylogenetic methods and comparing the exon structure of these proteins, we were able to draw conclusions about the evolutionary history of this family of proteins which, in turn, provides a framework for further analysis of the functions of these clinically-relevant genes.

Methods

Gene discovery

Sequences of *ADAMTS* homolog sequences presented here were identified from 2001 to 2004 using genomic sequences cataloged in the NCBI, JGI (*Fugu* and *Ciona* [52,70]) and Celera genome databases using BLASTp, BLASTn, BLASTx, tBLASTn and tBLASTx searches [71]. We searched databases of Expressed Sequence Tags (ESTs) for all human genes and, with the exception of *ADAMTS15*, found ESTs that corresponded to those genes, confirming their transcription. Splice site predictions were made by neural network *via* the Berkeley Drosophila Genome Project and also by eye using Sequencher 4.2 (Genecodes), with reference to protein multiple sequence alignments. We used the same method to identify *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Fugu rubripes* homologs of *ADAMTS* genes, and their genomic structure. Multiple sequence alignments of the inferred amino acid sequences were prepared using Multalin [72] and ClustalX1.81 [73] and manually refined and annotated within BioEdit [74] and MacClade4.06 [75].

Phylogenetic analysis

Initial phylogenetic analyses were conducted including all of the human ADAMTS proteins along with seven invertebrate homologs, three from *Drosophila melanogaster* and four from *Caenorhabditis elegans* (Figure 1A), and then these analyses were repeated with the most divergent invertebrate sequences removed (Figure 1B). The datasets for these analyses were comprised of 359 and 543 unambiguously aligned amino acid sites, respectively. The alignments were analyzed using parsimony and distance methods. Parsimony analyses used a heuristic search with random stepwise addition of data and tree-bisection-reconnection in PAUP*4.0b10 for 1000 bootstrap replicates [76]. Distance matrices were inferred using the Jones, Taylor, Thornton (JTT) substitution model with PROTDIST in PHYLIP 3.6a3 [77,78]. Neighbor-joining trees were constructed with the input order jumbled for 1000 bootstrap replicates using NEIGHBOR, SEQBOOT and CONSENSE in PHYLIP. Using Tree-Puzzle 5.0 [79] we generated maximum-likelihood distance matrices in which among-site substitution rate heterogeneity was corrected using an invariable and eight gamma-distributed substitution rate categories and the JTT model. 10,000 quartet-puzzling steps were also used (Tree-Puzzle) to assess branch support.

To better resolve the evolutionary relationships of the vertebrate *ADAMTS* subfamily members and to provide information on their presence in other chordate lineages, sequences from *Mus*, *Fugu* and *Ciona* were identified and added to the alignment and phylogenetic analysis. An annotated ADAMTS homolog from the honeybee *Apis mellifera* was also included in the analysis as another representative invertebrate. 571 unambiguously aligned amino acid sites of *ADAMTS*-homologous sequences encoded by *Homo*, *Mus*, *Fugu*, *Ciona*, *Apis*, *Drosophila* and *Caenorhabditis* were analyzed (Figure 2). MrBayes3.0 [80] was used to construct a maximum likelihood (ML) phylogenetic tree from this protein alignment. MrBayes was run for 1000000 generations, with four incrementally heated Markov chains, and a sampling frequency of 1000 generations. The temperature setting was increased to 0.5. Among-site substitution rate heterogeneity was corrected using an invariable and eight gamma-distributed substitution rate categories and the WAG model for amino acid substitutions [81], abbreviated herein as WAG+I+8Γ. The consensus ML tree topology, the arithmetic mean log-likelihood (lnL) for this topology, and branch support were estimated from the set of sampled trees with the best posterior probabilities. Means and 95% confidence intervals for the gamma distribution shape parameter α and the proportion of invariable sites (pI) were also estimated.

List of abbreviations

ADAMTS (**A** **D**isintegrin-like and **M**etalloprotease with **T**hrombo**s**pondin motifs)

ADAM (**A** **D**isintegrin-like and **M**etalloprotease)

TSR (**T**hrombospondin type 1 **S**equence **R**epeat)

MMP (**m**atrix **m**etallo**p**rotease)

ECM (**e**xtra**c**ellular **m**atrix)

SAGE (**s**erial **a**nalysis of **g**ene **e**xpression)

TTP (**t**hrombotic **t**hrombocytopenic **purpura**)

IL (**i**nter**l**eukin)

TGF (**t**ransforming **g**rowth **f**actor)

TNF (**t**umor **n**ecrosis **f**actor)

BLAST (**B**asic **l**ocal **a**lignment **s**earch **t**ool)

Authors' contributions

ACN and EGVM initiated the project and all authors were involved in the design phases. ACN, SBM and JML inferred sequences of previously un-annotated ADAMTS genes from Genbank, Celera, and JGI, and determined intron positions. SBM performed the phylogenetic analyses. ACN and SBM drafted the manuscript and JML and EGVM contributed to writing the paper and advised throughout.

Additional material

Additional File 2

Alignment used for phylogenetic analyses of animal ADAMTS homologs. *Unambiguously aligned amino acid sites, indicated by the "mask" line in the alignment, were used for phylogenetic analyses (Figure 2). Accession (GI) numbers for sequences are provided. Intron positions in the corresponding genomic sequence are indicated, with color-coding for intron phases (Figure 3).*

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-5-11-S2.pdf]

Additional File 1

Alignment used for phylogenetic analyses of human ADAMTS proteins and invertebrate homologs. *Unambiguously aligned amino acid sites, indicated by the "mask" line in the alignment, were used for phylogenetic analyses (figure 1). Intron positions in the corresponding genomic sequence are indicated, with color-coding for intron phases (Figures 3 and 4).*

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-5-11-S1.pdf]

Acknowledgements

ACN was supported by NIH training grant T32-NS07480. SBM was supported by an Emory Graduate Fellowship, an Avis Cone Fellowship (Iowa), and JML's grant. JML was supported by a grant from the NSF (MCB-0216702). EGVM was supported by grants from the NIH (CA86335 and CA87830). *Fugu rubripes* and *Ciona intestinalis* data are from the Joint Genome Institute databases.

References

1. Kuno K, Kanada N, Nakashima E, Fujiki F, Ichimura F, Matsushima K: **Molecular cloning of a gene encoding a new type of metalloproteinase- disintegrin family protein with thrombospondin motifs as an inflammation associated gene.** *J Biol Chem* 1997, **272(1):**556-562.
2. Apte SS: **A disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motifs: the ADAMTS family.** *Int J Biochem Cell Biol* 2004, **36(6):**981-985.
3. Abbaszade I, Liu RQ, Yang F, Rosenfeld SA, Ross OH, Link JR, Ellis DM, Tortorella MD, Pratta MA, Hollis JM, Wynn R, Duke JL, George HJ, Hillman MC Jr, Murphy K, Wiswall BH, Copeland RA, Decicco CP, Bruckner R, Nagase H, Itoh Y, Newton RC, Magolda RL, Trzaskos JM, Hollis GF, Arner EC, Burn TC: **Cloning and characterization of ADAMTS11, an aggrecanase from the ADAMTS family.** *J Biol Chem* 1999, **274(33):**23443-23450.
4. Colige A, Sieron AL, Li SW, Schwarze U, Petty E, Wertelecki W, Wilcox W, Krakow D, Cohn DH, Reardon W, Byers PH, Lapiere CM, Prockop DJ, Nusgens BV: **Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene.** *Am J Hum Genet* 1999, **62(2):**308-317.
5. Okada Y: **Tumor cell-matrix interaction: pericellular matrix degradation and metastasis.** *Verh Dtsch Ges Pathol* 2000, **84:**33-42.
6. Clark ME, Kelner GS, Turbeville LA, Boyer A, Arden KC, Maki RA: **ADAMTS9, a novel member of the ADAM-TS/ metallospondin gene family.** *Genomics* 2000, **67(3):**343-350.
7. Cal S, Arguelles JM, Fernandez PL, Lopez-Otin C: **Identification, characterization, and intracellular processing of ADAM-TS12, a novel human disintegrin with a complex structural organization involving multiple thrombospondin-1 repeats.** *J Biol Chem* 2001, **276(21):**17932-17940.
8. Tortorella MD, Burn TC, Pratta MA, Abbaszade I, Hollis JM, Liu R, Rosenfeld SA, Copeland RA, Decicco CP, Wynn R, Rockwell A, Yang F, Duke JL, Solomon K, George H, Bruckner R, Nagase H, Itoh Y, Ellis DM, Ross H, Wiswall BH, Murphy K, Hillman MC Jr, Hollis GF, Arner EC: **Purification and cloning of aggrecanase-1: a member of the ADAMTS family of proteins.** *Science* 1999, **284(5420):**1664-1666.
9. Satoh K, Suzuki N, Yokota H: **ADAMTS-4 (a disintegrin and metalloproteinase with thrombospondin motifs) is transcriptionally induced in beta-amyloid treated rat astrocytes.** *Neurosci Lett* 2000, **289(3):**177-180.
10. Dagoneau N, Benoist-Lasselin C, Huber C, Faivre L, Megarbane A, Alswaid A, Dollfus H, Alembik Y, Munnich A, Legeai-Mallet L, Cormier-Daire V: **ADAMTS10 mutations in autosomal recessive Weill-Marchesani syndrome.** *Am J Hum Genet* 2004, **75(5):**801-806.
11. Molloy SS, Bresnahan PA, Leppla SH, Klimpel KR, Thomas G: **Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen.** *J Biol Chem* 1992, **267(23):**16396-16402.
12. Longpre JM, Leduc R: **Identification of pro-domain determinants involved in ADAMTS-1 biosynthesis.** *J Biol Chem* 2004, **279(32):**33237-45.
13. Wang P, Tortorella M, England K, Malfait AM, Thomas G, Arner EC, Pei D: **Proprotein convertase furin interacts with and cleaves pro-ADAMTS4 (Aggrecanase-1) in the trans-Golgi network.** *J Biol Chem* 2004, **15(279):**15434-15440.
14. Rodriguez-Manzaneque JC, Milchanowski AB, Dufour EK, Leduc R, Iruela-Arispe ML: **Characterization of METH-1/ADAMTS1 processing reveals two distinct active forms.** *J Biol Chem* 2000, **275(43):**33471-33479.
15. Stone AL, Kroeger M, Sang QX: **Structure-function analysis of the ADAM family of disintegrin-like and metalloproteinase-containing proteins (review).** *J Protein Chem* 1999, **18(4):**447-465.
16. Evans JP: **Fertilin beta and other ADAMs as integrin ligands: insights into cell adhesion and fertilization.** *Bioessays* 2001, **23(7):**628-639.
17. Zhu X, Bansal NP, Evans JP: **Identification of key functional amino acids of the mouse fertilin beta (ADAM2) disintegrin loop for cell-cell adhesion during fertilization.** *J Biol Chem* 2000, **275(11):**7677-7683.
18. Turk BE, Huang LL, Piro ET, Cantley LC: **Determination of protease cleavage site motifs using mixture-based oriented peptide libraries.** *Nat Biotechnol* 2001, **19(7):**661-667.
19. Tortorella M, Pratta M, Liu RQ, Abbaszade I, Ross H, Burn T, Arner E: **The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for aggrecan substrate recognition and cleavage.** *J Biol Chem* 2000, **275(33):**25791-25797.
20. Tzarfaty-Majar V, Lopez-Alemany R, Feinstein Y, Gombau L, Goldshmidt O, Soriano E, Munoz-Canoves P, Klar A: **Plasmin-mediated release of the guidance molecule F-spondin from the extracellular matrix.** *J Biol Chem* 2001, **276(30):**28233-28241.

21. Blelloch R, Anna-Arriola SS, Gao D, Li Y, Hodgkin J, Kimble J: **The gon-1 gene is required for gonadal morphogenesis in Caenorhabditis elegans.** *Dev Biol* 1999, **216(1):**382-393.
22. Kramerova IA, Kawaguchi N, Fessler LI, Nelson RE, Chen Y, Kramerov AA, Kusche-Gullberg M, Kramer JM, Ackley BD, Sieron AL, Prockop DJ, Fessler JH: **Papilin in development; a pericellular protein with a homology to the ADAMTS metalloproteinases.** *Development* 2000, **127(24):**5475-5485.
23. Nardi JB, Martos R, Walden KK, Lampe DJ, Robertson HM: **Expression of lacunin, a large multidomain extracellular matrix protein, accompanies morphogenesis of epithelial monolayers in Manduca sexta.** *Insect Biochem Mol Biol* 1999, **29(10):**883-897.
24. Huwiler KG, Vestling MM, Annis DS, Mosher DF: **Biophysical characterization, including disulfide bond assignments, of the anti-angiogenic type I domains of human thrombospondin-1.** *Biochemistry* 2002, **48(41):**14329-14339.
25. Reiher FK, Volpert OV, Jimenez B, Crawford SE, Dinney CP, Henkin J, Haviv F, Bouck NP, Campbell SC: **Inhibition of tumor growth by systemic treatment with thrombospondin-1 peptide mimetics.** *Int J Cancer* 2002, **5(98):**682-689.
26. Dawson DW, Pearce SF, Zhong R, Silverstein RL, Frazier WA, Bouck NP: **CD36 mediates the In vitro inhibitory effects of thrombospondin-1 on endothelial cells.** *J Cell Biol* 1997, **3(138):**707-717.
27. de Fraipont F, Nicholson AC, Feige JJ, Van Meir EG: **Thrombospondins and tumor angiogenesis.** *Trends Mol Med* 2001, **9(7):**401-407.
28. Cal S, Obaya AJ, Llamazares M, Garabaya C, Quesada V, Lopez-Otin C: **Cloning, expression analysis, and structural characterization of seven novel human ADAMTSs, a family of metalloproteinases with disintegrin and thrombospondin-1 domains.** *Gene* 2002, **283(1–2):**49-62.
29. Rao C, Foernzler D, Loftus SK, Liu S, McPherson JD, Jungers KA, Apte SS, Pavan WJ, Beier DR: **A defect in a novel ADAMTS family member is the cause of the belted white-spotting mutation.** *Development* 2003, **19(130):**4665-4672.
30. Shindo T, Kurihara H, Kuno K, Yokoyama H, Wada T, Kurihara Y, Imai T, Wang Y, Ogata M, Nishimatsu H, Moriyama N, Oh-hashi Y, Morita H, Ishikawa T, Nagai R, Yazaki Y, Matsushima K: **ADAMTS-1: a metalloproteinase-disintegrin essential for normal growth, fertility, and organ morphology and function.** *J Clin Invest* 2000, **105(10):**1345-1352.
31. Vazquez F, Hastings G, Ortega MA, Lane TF, Oikemus S, Lombardo M, Iruela-Arispe ML: **METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity.** *J Biol Chem* 1999, **274(33):**23349-23357.
32. Kahn J, Mehraban F, Ingle G, Xin X, Bryant JE, Vehar G, Schoenfeld J, Grimaldi CJ, Peale F, Draksharapu A, Lewin DA, Gerritsen ME: **Gene expression profiling in an in vitro model of angiogenesis.** *Am J Pathol* 2000, **156(6):**1887-1900.
33. Vankemmelbeke MN, Holen I, Wilson AG, Ilic MZ, Handley CJ, Kelner GS, Clark M, Liu C, Maki RA, Burnett D, Buttle DJ: **Expression and activity of ADAMTS-5 in synovium.** *Eur J Biochem* 2001, **268(5):**1259-1268.
34. Kuno K, Okada Y, Kawashima H, Nakamura H, Miyasaka M, Ohno H, Matsushima K: **ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.** *FEBS Lett* 2000, **478(3):**241-245.
35. Dickinson SC, Vankemmelbeke MN, Buttle DJ, Rosenberg K, Heinegard D, Hollander AP: **Cleavage of cartilage oligomeric matrix protein (thrombospondin-5) by matrix metalloproteinases and a disintegrin and metalloproteinase with thrombospondin motifs.** *Matrix Biol* 2003, **3(22):**267-278.
36. Colige AC, Vandenberghe I, Thiry M, Lambert CA, Van Beeumen J, Li SW, Prockop DJ, Lapiere CM, Nusgens BV: **Cloning and characterization of ADAMTS-14, a novel ADAMTS displaying high homology with ADAMTS-2 and ADAMTS-3.** *J Biol Chem* 2001, **7:**7.
37. Levy GG, Nichols WC, Lian EC, Foroud T, McClintick JN, McGee BM, Yang AY, Siemieniak DR, Stark KR, Gruppo R, Sarode R, Shurin SB, Chandrasekaran V, Stabler SP, Sabio H, Bouhassira EE, Upshaw JD Jr, Ginsburg D, Tsai HM: **Mutations in a member of the ADAMTS gene family cause thrombotic thrombocytopenic purpura.** *Nature* 2001, **413(6855):**488-494.
38. Brass L: **VWF meets the ADAMTS family.** *Nat Med* 2001, **7(11):**1177-1178.
39. Zheng X, Chung D, Takayama TK, Majerus EM, Sadler JE, Fujikawa K: **Structure of von Willebrand factor-cleaving protease (ADAMTS13), a metalloprotease involved in thrombotic thrombocytopenic purpura.** *J Biol Chem* 2001, **276(44):**41059-41063.
40. Fujikawa K, Suzuki H, McMullen B, Chung D: **Purification of human von Willebrand factor-cleaving protease and its identification as a new member of the metalloproteinase family.** *Blood* 2001, **98(6):**1662-1666.
41. Sylvester J, Liacini A, Li WQ, Zafarullah M: **Interleukin-17 signal transduction pathways implicated in inducing matrix metalloproteinase-3, -13 and aggrecanase-1 genes in articular chondrocytes.** *Cell Signal* 2004, **4(16):**469-476.
42. Tsuzaki M, Guyton G, Garrett W, Archambault JM, Herzog W, Almekinders L, Bynum D, Yang X, Banes AJ: **IL-1 beta induces COX2, MMP-1, -3 and -13, ADAMTS-4, IL-1 beta and IL-6 in human tendon cells.** *J Orthop Res* 2003, **2(21):**256-264.
43. Moulharat N, Lesur C, Thomas M, Rolland-Valognes G, Pastoureau P, Anract P, De Ceuninck F, Sabatini M: **Effects of transforming growth factor-beta on aggrecanase production and proteoglycan degradation by human chondrocytes in vitro.** *Osteoarthritis Cartilage* 2004, **4(12):**296-305.
44. Wang WM, Lee S, Steiglitz BM, Scott IC, Lebares CC, Allen ML, Brenner MC, Takahara K, Greenspan DS: **Transforming growth factor-beta induces secretion of activated ADAMTS-2. A procollagen III N-proteinase.** *J Biol Chem* 2003, **21(278):**19549-19557.
45. Bevitt DJ, Mohamed J, Catterall JB, Li Z, Arris CE, Hiscott P, Sheridan C, Langton KP, Barker MD, Clarke MP, McKie N: **Expression of ADAMTS metalloproteinases in the retinal pigment epithelium derived cell line ARPE-19: transcriptional regulation by TNFalpha.** *Biochim Biophys Acta* 2003, **1–3(1626):**83-91.
46. Norata GD, Bjork H, Hamsten A, Catapano AL, Eriksson P: **High-density lipoprotein subfraction 3 decreases ADAMTS-1 expression induced by lipopolysaccharide and tumor necrosis factor-alpha in human endothelial cells.** *Matrix Biol* 2004, **7(22):**557-560.
47. Cole P, Rabasseda X: **The soluble tumor necrosis factor receptor etanercept: a new strategy for the treatment of autoimmune rheumatic disease.** *Drugs Today* 2004, **4(40):**281-324.
48. Baecklund E, Askling J, Rosenquist R, Ekbom A, Klareskog L: **Rheumatoid arthritis and malignant lymphomas.** *Curr Opin Rheumatol* 2004, **3(16):**254-261.
49. Gomez-Reino JJ, Carmona L, Valverde VR, Mola EM, Montero MD, BIOBADASER Group: **Treatment of rheumatoid arthritis with tumor necrosis factor inhibitors may predispose to significant increase in tuberculosis risk: a multicenter active-surveillance report.** *Arthritis Rheum* 2003, **8(48):**2122-2127.
50. Clark IM, Parker AE: **Metalloproteinases: their role in arthritis and potential as herapeutic targets.** *Expert Opin Ther Targets* 2003, **7(1):**19-34.
51. Puente XS, Lopez-Otin C: **A genomic analysis of rat proteases and protease inhibitors.** *Genome Res* 2004, **14(4):**609-622.
52. 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, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S: **Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes.** *Science* 2002, **297(5585):**1301-1310.
53. Puente XS, Sanchez LM, Overall CM, Lopez-Otin C: **Human and mouse proteases: a comparative genomic approach.** *Nat Rev Genet* 2003, **4(7):**544-558.
54. Felsenstein J: **Cases in which parsimony or compatibility methods will be positively misleading.** *Syst Zool* 1978, **27:**401-410.
55. Cho C, Ge H, Branciforte D, Primakoff P, Myles DG: **Analysis of mouse fertilin in wild-type and fertilin beta(-/-) sperm: evidence for C-terminal modification, alpha/beta dimerization, and lack of essential role of fertilin alpha in sperm-egg fusion.** *Dev Biol* 2000, **222(2):**289-295.
56. Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA: **Evidence for a clade of nematodes, arthropods and other moulting animals.** *Nature* 1997, **387(6632):**489-493.

57. Manuel M, Kruse M, Muller WE, Le Parco Y: **The comparison of beta-thymosin homologues among metazoa supports an arthropod-nematode clade.** *J Mol Evol* 2000, **51(4):**378-381.
58. Adoutte A, Balavoine G, Lartillot N, Lespinet O, Prud'homme B, de Rosa R: **The new animal phylogeny: reliability and implications.** *Proc Natl Acad Sci U S A* 2000, **97(9):**4453-4456.
59. Peterson KJ, Eernisse DJ: **Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences.** *Evol Dev* 2001, **3(3):**170-205.
60. Mallatt J, Winchell CJ: **Testing the new animal phylogeny: first use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes.** *Mol Biol Evol* 2002, **19(3):**289-301.
61. Ruiz-Trillo I, Paps J, Loukota M, Ribera C, Jondelius U, Baguna J, Riutort M: **A phylogenetic analysis of myosin heavy chain type II sequences corroborates that Acoela and Nemertodermatida are basal bilaterians.** *Proc Natl Acad Sci* 2002, **99(17):**11246-11251.
62. Copley RR, Aloy P, Russell RB, Telford MJ: **Systematic searches for molecular synapomorphies in model metazoan genomes give some support for Ecdysozoa after accounting for the idiosyncrasies of Caenorhabditis elegans.** *Evol Dev* 2004, **6(3):**164-169.
63. Kuno K, Iizasa H, Ohno S, Matsushima K: **The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.** *Genomics* 1997, **46(3):**466-471.
64. Somerville RP, Longpre JM, Jungers KA, Engle JM, Ross M, Evanko S, Wight TN, Leduc R, Apte SS: **Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS subfamily related to Caenorhabditis elegans GON-1.** *J Biol Chem* 2003, **11(278):**9503-9513.
65. Llamazares M, Cal S, Quesada V, Lopez-Otin C: **Identification and characterization of ADAMTS-20 defines a novel subfamily of metalloproteinases-disintegrins with multiple thrombospondin-1 repeats and a unique GON domain.** *J Biol Chem* 2003, **15(278):**13382-13389.
66. Eichler EE: **Recent duplication, domain accretion and the dynamic mutation of the human genome.** *Trends Genet* 2001, **17(11):**661-669.
67. Wolfe KH: **Yesterday's polyploids and the mystery of diploidization.** *Nat Rev Genet* 2001, **2(5):**333-341.
68. Ayala FJ, Rzhetsky A: **Origin of the metazoan phyla: molecular clocks confirm paleontological estimates.** *Proc Natl Acad Sci U S A* 1998, **95(2):**606-611.
69. Vandepoele K, de Vos W, Taylor JS, Meyer A, Van de Peer Y: **Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates.** *Proc Natl Acad Sci U S A* 2004, **101(6):**1638-1643.
70. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KE, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang HG, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee BI, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS: **The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins.** *Science* 2002, **298(5601):**2157-2167.
71. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25(17):**3389-3402.
72. Corpet F: **Multiple sequence alignment with hierarchical clustering.** *Nucl Acids Res* 1988, **22(16):**10881-10890.
73. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25(24):**4876-4882.
74. Hall T: **BioEdit.** 7.0.0 edition. Carlsbad CA; 2004.
75. Maddison WP, Maddison DR: **MacClade.** 4.06th edition. Sunderland MA: Sinauer Associates; 2003.
76. Swofford DL: **PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods).** 4.0b10 edition. Sunderland, Massachusetts: Sinauer Associates; 2003.
77. Jones DT, Taylor WR, Thornton JM: **The rapid generation of mutation data matrices from protein sequences.** *Comput Appl Biosci* 1992, **8(3):**275-282.
78. Felsenstein J: **PHYLIP, Phylogeny Inference Package.** 3.6a3 edition. Seattle; 2003.
79. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18(3):**502-504.
80. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17(8):**754-755.
81. Whelan S, Goldman N: **A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach.** *Mol Biol Evol* 2001, **18(5):**691-699.
82. **MEROPS – the Peptidase Database** [http://merops.sanger.ac.uk]

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

BioMedcentral