

The Evolution of Extracellular Fibrillins and Their Functional Domains

Adam Piha-Gossack¹, Wayne Sossin², Dieter P. Reinhardt^{1,3*}

1 Department of Anatomy and Cell Biology, McGill University, Montreal, Quebec, Canada, **2** Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada, **3** Faculty of Dentistry, McGill University, Montreal, Quebec, Canada

Abstract

Fibrillins constitute the major backbone of multifunctional microfibrils in elastic and non-elastic extracellular matrices, and are known to interact with several binding partners including tropoelastin and integrins. Here, we study the evolution of fibrillin proteins. Following sequence collection from 39 organisms representative of the major evolutionary groups, molecular evolutionary genetics and phylogeny inference software were used to generate a series of evolutionary trees using distance-based and maximum likelihood methods. The resulting trees support the concept of gene duplication as a means of generating the three vertebrate fibrillins. Beginning with a single fibrillin sequence found in invertebrates and jawless fish, a gene duplication event, which coincides with the appearance of elastin, led to the creation of two genes. One of the genes significantly evolved to become the gene for present-day fibrillin-1, while the other underwent evolutionary changes, including a second duplication, to produce present-day fibrillin-2 and fibrillin-3. Detailed analysis of several sequences and domains within the fibrillins reveals distinct similarities and differences across various species. The RGD integrin-binding site in TB4 of all fibrillins is conserved in cephalochordates and vertebrates, while the integrin-binding site within cbEGF18 of fibrillin-3 is a recent evolutionary change. The proline-rich domain in fibrillin-1, glycine-rich domain in fibrillin-2 and proline-/glycine-rich domain in fibrillin-3 are found in all analyzed tetrapod species, whereas it is completely replaced with an EGF-like domain in cnidarians, arthropods, molluscs and urochordates. All collected sequences contain the first 9-cysteine hybrid domain, and the second 8-cysteine hybrid domain with exception of arthropods containing an atypical 10-cysteine hybrid domain 2. Furin cleavage sites within the N- and C-terminal unique domains were found for all analyzed fibrillin sequences, indicating an essential role for processing of the fibrillin pro-proteins. The four cysteines in the unique N-terminus and the two cysteines in the unique C-terminus are also highly conserved.

Citation: Piha-Gossack A, Sossin W, Reinhardt DP (2012) The Evolution of Extracellular Fibrillins and Their Functional Domains. *PLoS ONE* 7(3): e33560. doi:10.1371/journal.pone.0033560

Editor: John Parkinson, Hospital for Sick Children, Canada

Received: November 11, 2011; **Accepted:** February 14, 2012; **Published:** March 16, 2012

Copyright: © 2012 Piha-Gossack et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Canadian Institutes of Health Research (MOP-106494), the Natural Sciences and Engineering Research Council of Canada (RGPIN 375738-09), the Canada Foundation for Innovation, the Canada Research Chairs, and the Fonds de la recherche en santé du Québec. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dieter.reinhardt@mcgill.ca

Introduction

Fibrillins constitute a family of large extracellular proteins that form the core of highly ordered extended and ubiquitously distributed aggregates, termed microfibrils [1]. Evolutionarily, fibrillins and microfibrils are widely distributed from cnidarians to mammals [2–9]. Microfibrils confer structural integrity to individual organ systems and regulate the bioavailability of growth factors of the TGF- β superfamily, including TGF- β and bone morphogenetic proteins [10,11]. In elastin-expressing vertebrate organisms, microfibrils are believed to provide a scaffold for elastogenesis in blood vessels, lung, skin and other elastic tissues [12]. A number of hereditary connective tissue disorders in humans, including Marfan syndrome, dominant Weill-Marchesani syndrome, geleophysic and acromicric dysplasia, stiff skin syndrome and others are associated with mutations in fibrillin-1, whereas mutations in fibrillin-2 lead to congenital contractural arachnodactyly [13–16].

Most of the important structural and functional properties in fibrillins were identified and described with fibrillin nucleotide and protein sequences from mammalian organisms, in particular

humans. Thus, the following introductory description of fibrillin properties refers primarily to information available on human fibrillins. The fibrillin family consists of three homologous isoforms, fibrillin-1, -2 and -3. Fibrillins are composed of a typical sequence of individual domains containing between 40–80 amino acid residues (Fig. 1). This characteristic domain signature is 100% conserved between all three fibrillins of mammalian organisms with the exception of an alternatively spliced domain at the N-terminus of fibrillin-3 [17]. Almost all of the fibrillin domains are characterized by a typical number of cysteine residues ranging between 6–9 cysteine residues per domain. Generally, fibrillins are the extracellular proteins with the highest content of cysteine residues (12–13%) [18].

The most prominent domain in fibrillins is the epidermal growth factor-like (EGF) domain which is present 46–47 times in fibrillins. This domain contains six characteristic cysteine residues which form three stabilizing disulfide bonds in a 1–3, 2–4, 5–6 arrangement [19,20]. The majority of the EGF-like domains contain a consensus sequence ((D/N)X(D/N)(E/Q)X_m(D/N*)X_n(Y/F)) for calcium-binding (cb) at their N-terminus where *m* and *n* are variable numbers of amino acid residues, and the

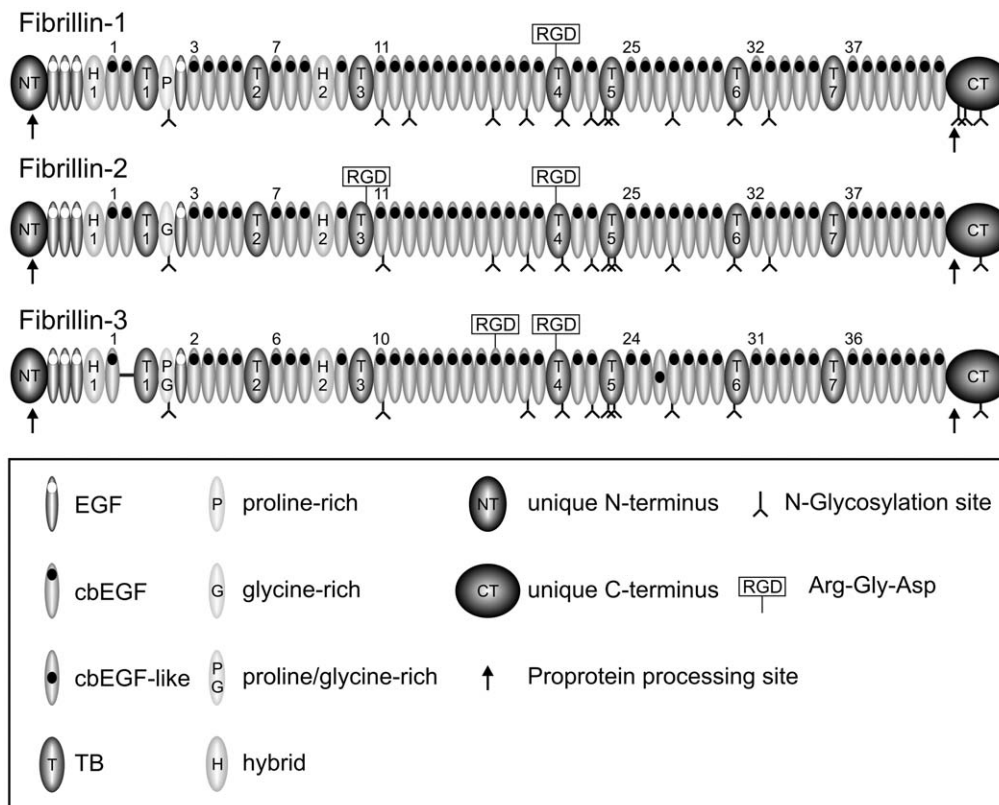


Figure 1. Overview of the domain structure of the human fibrillin family. Numbers above cbEGF domains and within TB and hybrid domains indicate the relative number of the respective domain within the human fibrillin molecule. RGD sites, proprotein processing sites in the unique N- and C-terminal domains (arrows) and predicted N-glycosylation sites (inverted Y symbols) are indicated.
doi:10.1371/journal.pone.0033560.g001

asterisk indicates a potential hydroxylation site [21]. Homologous EGF and cbEGF domains are found in numerous extracellular matrix, cell surface and blood proteins throughout all metazoan organisms [19].

Tandem arrays of EGF and cbEGF domains in fibrillins separate two other types of domains, the transforming growth factor (TGF)- β binding protein domains (TB) and the hybrid domains. These two domains are unique to fibrillins and to the latent TGF- β binding proteins (LTBPs) [22]. The TB domains occur seven times in fibrillins and are characterized by eight cysteine residues that form four stabilizing disulfide bonds organized in a 1–3, 2–6, 4–7, 5–8 pattern [23–25]. Three of these cysteine residues (Cys3, Cys4, Cys5) are present in a characteristic triple Cys-Cys-Cys motif. Sequence and structural analysis demonstrated that the two hybrid domains present in fibrillins have phylogenetically evolved by fusion of the N-terminus of a TB domain with the C-terminus of a cbEGF domain [7,26,27]. Hybrid domains are characterized by a typical Cys-Cys repeat and a 1–3, 2–5, 4–6, 7–8 disulfide-bond pattern [7,27]. The first hybrid domain in mammalian fibrillins contains an extra cysteine inserted between Cys2 and Cys3. This unpaired cysteine residue were shown to be surface-exposed, suggesting a role in intermolecular disulfide-bond formation [28].

Fibrillins can be structurally distinguished by a characteristic domain without cysteine residues immediately following the first TB domain. This unique domain is rich in proline residues in fibrillin-1, rich in glycine residues in fibrillin-2 and rich in both proline and glycine residues in fibrillin-3 [7–9]. For fibrillin-1, the proline-rich region has been suggested as a hinge region

facilitating a folding mechanism of the fibrillin molecule in microfibrils [29]. The proline-rich region in fibrillin-1 and possibly the glycine-rich region in fibrillin-2 are involved in the interaction with tropoelastin [30,31]. The fibrillin N-terminal domain contains four cysteine residues and the C-terminal domain contains 2 cysteine residues. Both terminal domains also contain a tribasic consensus sequence (Arg-Xaa-(Lys/Arg)-Arg) that are recognition signals for proprotein convertases of the Furin/PACE type [32,33]. Both the N- and the C-terminal domains are proteolytically processed at these sites [28,34–37]. It has been shown for fibrillin-1 that only the processed form becomes incorporated into the extracellular matrix, suggesting that profibrillin-1 processing within its terminal domains plays a regulatory role for its assembly into microfibrils [38]. All three fibrillins in mammalian organisms contain one (fibrillin-1) or two (fibrillin-2 and -3) Arg-Gly-Asp (RGD) sites. Common to all three mammalian fibrillins is a RGD site present in the TB4 domain at the tip of an extensible loop [25]. This site mediates the interactions with $\alpha\beta3$, $\alpha5\beta1$ and $\alpha\beta6$ integrins [25,39–42]. Human fibrillin-2 and fibrillin-3 contain a second RGD site in TB3 and cbEGF18, respectively. It is not known whether these RGD sites represent functional integrin-binding epitopes. Another differential feature between the three fibrillins is the number and position of predicted N-linked glycosylation sites (see Fig. 1).

Recently, a global evolutionary analysis of TB domain-containing proteins highlighted that the characteristic fibrillin domain organization emerged over 600 million years ago prior to the divergence of cnidarians and bilaterians and before LTBPs emerged [43]. This study further described the unpaired cysteine

residue in the first hybrid domain, as well as the CXXC motif in the C-terminus as absolutely conserved between species, while the RGD site in TB4 is not conserved outside vertebrates.

In the present study we have retrieved and extensively curated 78 fibrillin gene and protein sequences from 39 organisms ranging from cnidarians to mammals. We have developed evolutionary trees suggesting that a single ancestral fibrillin gene still present in invertebrates and jawless fish (agnathans) underwent duplication. One of the resulting genes evolved to become fibrillin-1, while the other underwent evolutionary changes, including a second duplication, to produce fibrillin-2 and fibrillin-3. We have further analyzed the evolution of critical functional motifs and domains in fibrillins.

Methods

Sequence Collection

Fibrillin nucleotide and protein sequences for 39 organisms were sequentially retrieved from publically available databases including the DOE Joint Genome Institute (<http://www.jgi.doe.gov/>), Ensembl (<http://uswest.ensembl.org/>), the National Center for Biotechnology Information nucleotide database (<http://www.ncbi.nlm.nih.gov/nucleotide/>), the UCSC Genome Bioinformatics (<http://genome.ucsc.edu>) and the EMBL European Bioinformatics Institute (<http://www.ebi.ac.uk>). A complete list of retrieved sequences and the corresponding databases is shown in Table S1. Keyword searches with the terms “fibrillin,” “fbn,” “fbn-like,” and “similar to fibrillin” were used to find annotated fibrillin sequences. In order to identify non-annotated fibrillin sequences, BLASTP, TBLASTN and BLASTN searches were conducted using sequences of characteristic fibrillin domains as queries, including the TB and the hybrid domain [44]. FASTA comparisons with known fibrillin sequences were used on incomplete or uncertain non-annotated sequences in order to distinguish between fibrillins and LTBP s [45,46]. Retrieved sequences were stored in a GCG software package version 11.1 database (Accelrys). Following a complete sequence alignment with ClustalX 2.0 [47], minor sequence adjustments were made to remove missense insertions, correct missense regions and fill in gap regions with the appropriate sequence. Missense insertions were only removed if they were located between intron/exon boundaries. Adjustments were made based on evidence from the genomic scaffolds or contigs used to derive the original sequences.

Manual Sequence Reconstruction

Many of the protein database entries are predicted by automated computation from genomic nucleotide sequences. As a consequence, certain sequences were either incomplete or appeared to contain errors such as insertions and deletions due to misinterpretations of intron/exon boundaries and open reading frames. To compare these critical regions with other sequences, corresponding raw genomic chromosomes, scaffolds or contigs were downloaded from databases and translated into all six reading frames. Upon determining the appropriate direction, the three corresponding reading frames were simultaneously viewed and a known fibrillin sequence was used as a template to highlight and extract individual domains from the raw sequence. With this method, we successfully reconstructed six complete fibrillin sequences, as well as filled in partial or missing regions for various other sequences. For some sequences, precise intron/exon boundaries were difficult to determine and as a consequence, the sequences around certain intron/exon boundaries are uncertain. However, by analyzing these reconstructed sequences with ClustalX, we confirmed the validity of the sequences based on

homology. A complete list of manually reconstructed and modified sequences is provided as part of Table S1.

Phylogenetic Analysis

Given that signal peptides are often species-specific [48], they were identified using the SignalP 3.0 server and then removed from the sequences (<http://www.cbs.dtu.dk/services/SignalP/>) [49]. In cases with low signal peptide cleavage site probability, sequence homology was included as a further criterion. A total of 78 protein sequences from 39 organisms were used in the analysis. A multiple sequence file was generated using GCG's ClustalW+ program. A complete alignment was then conducted on the resulting file using ClustalX 2.0 with default settings. Gap-only columns were subsequently removed. Using the ProtTest-HPC 3.0 objective model testing program with Akaike Information Criterion (AIC), the Jones-Taylor-Thornton (JTT) model with invariant sites and a gamma distribution was determined to be the most appropriate for the given alignment ($\Delta AIC = 0.00$, $AIC_w = 1.00$) [50,51]. Maximum likelihood phylogenetic reconstructions were generated with this substitution model using Molecular Evolutionary Genetics Analysis (MEGA) 5.05 software [52]. The Nearest-Neighbor-Interchange (NNI) search method was employed. Reliability of the trees was evaluated using 1,000 bootstrap repetitions. Similar results (not shown) were obtained using the JTT substitution model with the Randomized Accelerated Maximum Likelihood (RAxML) program (version 7.2.8) on the CIPRES Portal (http://www.phylo.org/sub_sections/portal) [53,54]. Distance-based phylogenetic trees were constructed using a series of four sequentially executed programs from the PHYLIP version 3.69 software package [55]. First, “Seqboot” was used to generate 1,000 data sets from the alignment file in the PHYLIP format. Next, “Protdist” was used to generate a distance matrix using the JTT model. Using the program “Neighbor,” we then constructed phylogenetic trees from each data set using the Neighbor-Joining method [56]. Finally, the program “Consense” was used to generate the consensus tree.

Phylogenetic trees for fibrillin proteins were initially generated in the following three principal ways. Separate distance-based phylogenetic reconstructions were computed for each of the seven TB domains; sequences with gaps or apparent missense regions in these corresponding domains were not included in the analysis. This allowed us to determine the evolutionary history of each individual TB domain and to include the largest number (59–69) of fibrillin sequences in the analysis. An additional distance-based phylogenetic tree using as input the seven concatenated TB domains was also generated in order to calculate an “average” tree. However, since not all fibrillin sequences are completely available, the total number of fibrillin sequences in this analysis was reduced to 48. A final distance-based phylogenetic tree was generated for the entire sequence using 37 fibrillins; partial and highly-gapped sequences were not included in the analysis. The second TB domain from human LTBP-2 was used as the outgroup for the seven individual fibrillin TB domain analyses. With regards to the tree generated from the seven concatenated TB domains, the same outgroup was used, however it was duplicated seven times to generate a comparable 7-TB-domain structure. Alternatively, the phylogenetic analyses of the seven concatenated fibrillin TB domains were performed with an outgroup consisting of the concatenated individual three TB domains found in human LTBP-2 (not a repeated single TB domain). This method produced very similar results (not shown). The full human LTBP-2 sequence was used as the outgroup for the tree generated from the complete fibrillin sequences. Phylogenetic reconstructions were repeated using the maximum likelihood method for both the

concatenated TB domains sequence and the complete fibrillin sequence. Non-fibrillin outgroup sequences were omitted. Final consensus trees were re-rooted around the base of the single fibrillin clade. Similar results were obtained from an additional maximum likelihood phylogenetic reconstruction of the concatenated TB domains using the corresponding 7-TB-domain human LTBP-2 sequence as an outgroup.

Analyses of Short Sequence Motifs

The FindPatterns program from the GCG software package was used to search for matching and homologous short sequence motifs, including integrin-binding sites and furin-cleavage sites. The pattern RGD was used to identify RGD-dependent integrin-binding sites. Furin-cleavage sites were located using the pattern RX(K,R)R, interpreted as arginine followed by any amino acid, followed by either lysine or arginine, followed by arginine.

Results and Discussion

We have retrieved 78 fibrillin sequences from a total of 39 organisms representing several taxonomic groups, including chordates, arthropods, annelids, echinoderms, molluscs and cnidarians from publically available databases (Table S1). After careful inspection and correction of sequence errors as explained in detail in Material and Methods, we performed phylogenetic analyses and studied the evolution of known key functional sequences. Only high quality sequences were included in the individual analyses, whereas regions of low quality sequences were not included. Results and discussions that include domain numbers use the numbering system of human fibrillin-1 as a reference (Fig. 1).

Phylogenetic Analysis

All fibrillins from all species are characterized by a typical fibrillin domain signature containing seven isolated TB and one or two hybrid domains interspersed by characteristic numbers of cbEGF domains, as originally identified in human fibrillins (Fig. 1). As expected, all mammals, except rodents, armadillo and platypus, have three functional fibrillin genes. It was previously described that the rodent fibrillin-3 gene is inactivated, potentially due to chromosome fission events [17]. A fibrillin-3 database entry was found for guinea pig, however several non-homologous insertions were found throughout the sequence. With the present data, we are unable to determine if these insertions are in fact present *in vivo*, thus yielding a non-functional transcript consistent with other rodents, or if these regions merely reflect sequencing errors. The lack of fibrillin-3 in armadillo and platypus may be due to either incomplete sequencing of these genomes or species-specific gene losses. Three fibrillin genes were also identified in chicken, zebra finch, lizard, frog and zebrafish (see Text S1 for more information on zebrafish fibrillins). Other ray-finned fish (actinopterygian) species, including the Japanese pufferfish, the spotted green pufferfish and the medaka fish, contain two fibrillin genes. A single fibrillin was found in species representative of several phyla and subphyla, including cnidarians, molluscs, annelids, arthropods, echinoderms, urochordates, cephalochordates and vertebrates. The single fibrillin found in vertebrates is restricted to agnathans. An overview of the fibrillin genes and proteins identified and analyzed is presented in Table S1.

Phylogenetic trees were generated with the entire fibrillin sequences, with the seven concatenated TB domains from each fibrillin and with the individual TB domains. Generally, all generated phylogenetic trees support the existence of an ancestral fibrillin gene, consistently placed outside of the fibrillin-1, -2, and -

3 clades, and corresponding with species containing a single fibrillin gene (Fig. 2). While a single fibrillin sequence was found in several insects, including the honey bee, ant, beetle, louse and aphid, none was found in the fruit fly (*Drosophila melanogaster*). Furthermore, contrary to existing database entries annotated as fibrillin homologs in the nematode *Caenorhabditis elegans* (NCBI: NM_066269.2; Ensembl: Y64G10A.7), the absence of the typical fibrillin domain signature, including TB and hybrid domains, does not support the existence of fibrillin homologs in *C. elegans*, nor in several other nematode species. The same rationale applies to database entries annotated as putative fibrillin sequences in *Schistosoma mansoni*, a trematode (EnsemblMetazoa: Smp_167190, Smp_067800). These findings are in agreement with the previous report by Robertson *et al.* and suggest that the fibrillin gene was either lost or severely disrupted during the evolution of these organisms [43].

All distance-based and maximum likelihood phylogenetic reconstructions support the concept of two gene duplications as a means of generating the three present-day fibrillin sequences from a single ancestral fibrillin (Fig. 2, Table 1). The ray-finned fish species surveyed in this analysis, including the zebrafish, Japanese pufferfish, medaka fish, and the spotted green pufferfish, are the oldest taxonomic groups found to contain more than one fibrillin gene. The presence of only a single fibrillin gene in invertebrates, including the sea squirt (*Ciona intestinalis*), amphioxus (*Branchiostoma floridae*), and the agnathan sea lamprey (*Petromyzon marinus*) strongly suggests that the first fibrillin gene duplication occurred prior to the divergence of ray-finned fish (Fig. 2, Table S1). This gene duplication resulted in a fibrillin-1 gene that significantly evolved further to become the gene expressing present-day fibrillin-1 (Fig. 2). The second gene, termed fibrillin-2/3, underwent further evolutionary changes, including a second duplication, to produce the genes for present-day fibrillin-2 and fibrillin-3 (Fig. 2). This second duplication appears to coincide with the divergence of actinopterygians and tetrapods, and is supported by the presence of three fibrillin genes in the amphibians, as well as in all analyzed reptile, bird and mammalian species. In considering the distance-based phylogenetic trees, the bootstrap values for the initial gene duplication leading to fibrillin-1 and fibrillin-2/3 (100% for concatenated TB domains and 99.9% for the complete protein) and for the second gene duplication to form separate fibrillin-2 and fibrillin-3 genes (100% for concatenated TB domains and 96.7% for the complete protein) strongly support this conclusion (Table 1). Bootstrap values obtained from the analyses with the individual TB domains are typically more variable, as expected for shorter sequences. For the first gene duplication, high bootstrap values are obtained for TB1, TB2, TB3 and TB6 (76.5–94.3%), while TB4, TB5 and TB7 showed lower bootstrap values (45.4–53.2%). Generally, these values provide additional support for the initial gene duplication. Bootstrap values from individual TB domains for the second split are generally relatively low (21–47.5%) for all TB domains except for TB2 (66.5%). However, this gene duplication hypothesis is strongly supported by the maximum likelihood phylogenetic reconstructions for the concatenated TB domains (100% for the first duplication, 87% for the second duplication) and the complete fibrillin sequences (100% for the first duplication, 100% for the second duplication). In summary, while the distance-based phylogenetic reconstructions from each individual TB domain do not always provide high bootstrap values in support of the duplication events, the analysis of the trees generated from the concatenated TB domains and the complete fibrillin sequences using both distance-based and maximum likelihood methods provides sufficient evidence to support the series of two gene

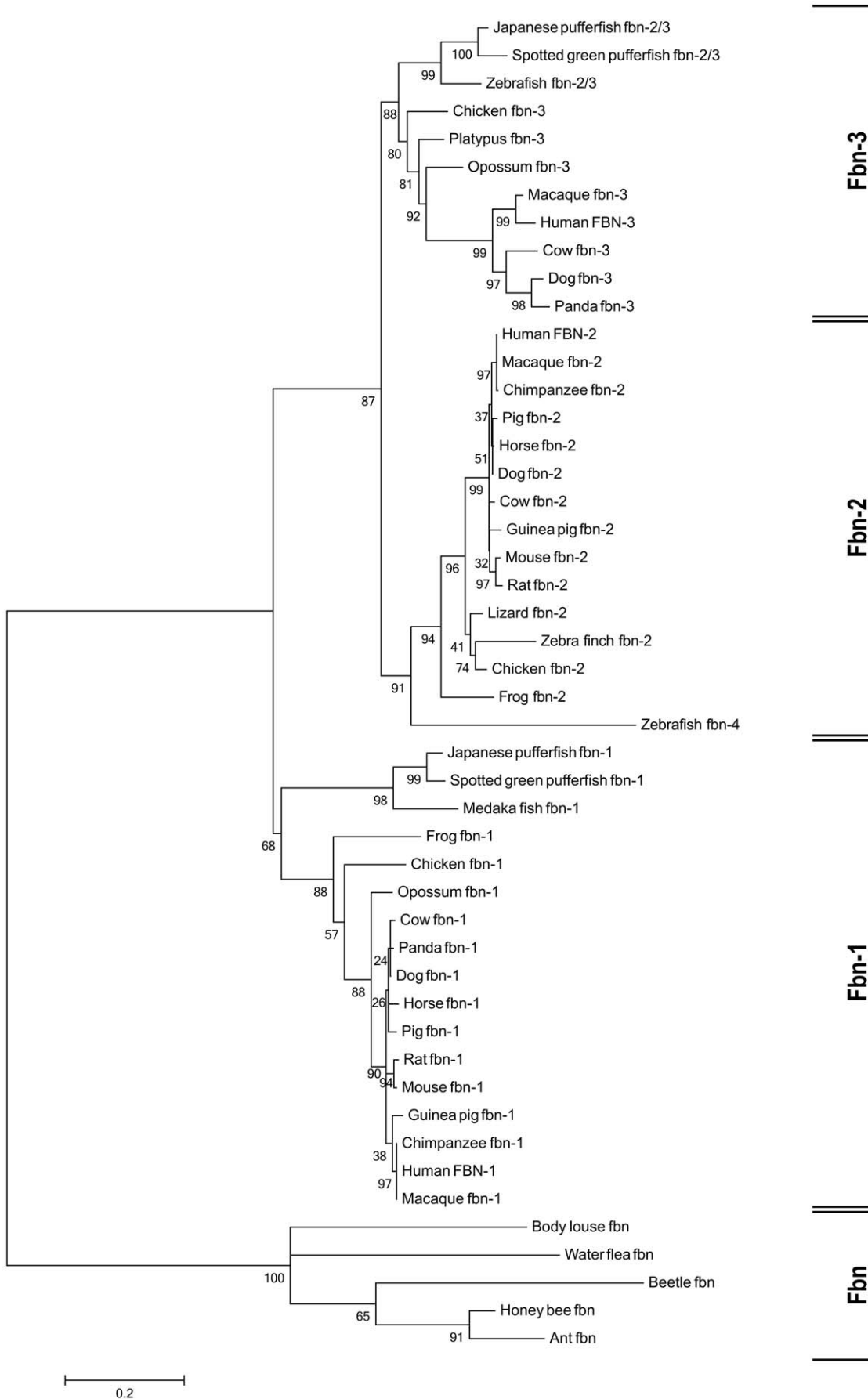


Figure 2. Maximum likelihood phylogenetic tree generated from seven concatenated TB domains of each fibrillin. The fibrillin protein family is grouped into four separate clades (from bottom to top): single fibrillin (Fbn), fibrillin-1 (Fbn-1), fibrillin-2 (Fbn-2), and fibrillin-3 (Fbn-3). Bootstrap values shown on each node represent the percentage of trees (out of 1,000) yielding the same two-set partition of sequences. The Jones-Taylor-Thornton model with invariant sites and a gamma distribution with four discrete categories was used. The tree was re-rooted around the base of the single fibrillin clade. The scale bar indicates the estimated average number of expected substitutions per site. Analyses of full length fibrillins or of full length fibrillins lacking the unique region resulted in similar phylogenetic trees. doi:10.1371/journal.pone.0033560.g002

duplications to yield three present-day fibrillins from one ancestral fibrillin.

Analysis of Functional Regions

RGD Integrin-Binding Sites. A characteristic feature that distinguishes amongst the fibrillin isoforms in human and other mammals is the differential presence and location of cell surface integrin-binding sites, characterized by the amino acid sequence RGD (Fig. 1). All three forms of human fibrillin contain a RGD site within TB4, while fibrillin-2 and fibrillin-3 have an additional RGD site located in TB3 and cbEGF18, respectively [57]. Analysis of the evolutionary conservation of these sites confirmed that all vertebrate fibrillin-1, -2, and -3 sequences contain a RGD site within TB4, located between Cys1 and Cys2 of this TB domain (Fig. 3B) [43]. The single fibrillin found in both sea lamprey (jawless fish) and amphioxus (cephalochordate) also has a RGD site in this domain. The presence of a RGD site in TB4 of a cephalochordate contradicts recent findings and emphasize that this site is not strictly limited to vertebrates [43]. Furthermore, sea lamprey contains a RGD site in TB3, while in amphioxus, the binding site is incomplete, although sequence quality in this region is lower than in other regions (Fig. 3A). The fibrillin sequence found in the urochordate *Ciona intestinalis* does not have a RGD site in either of these domains. The single fibrillin found in other invertebrates, including cnidarians, arthropods, molluscs and echinoderms, does not have a RGD site in TB3 or TB4. Thus, the introduction of RGD integrin-binding sites in both TB3 and TB4 occurred during the evolution of chordates, prior to the emergence of sea lamprey and possibly even before that of amphioxus.

With respect to fibrillin-3, only human and chimpanzee fibrillin-3 have RGD sites in cbEGF18, while KGD is found in this domain in the Rhesus macaque, a primate, and NAD, NAE, NAN, or RAD are found in other fibrillin-3 sequences. This finding strongly suggests that the presence of an integrin-binding site in fibrillin-3 within cbEGF18 is a very recent evolutionary change. It is currently not known whether the RGD site in cbEGF18 of human and chimpanzee fibrillin-3 is functional with respect to cell binding.

The integrin-binding site analysis further revealed additional characteristics of the fibrillin family as it evolved over time.

Integrin-binding sites were found in TB3 of fibrillin-1 and fibrillin-2/3 sequences, at the homologous fibrillin-2 RGD site, for all surveyed ray-finned fish species (Fig. 3A). The fibrillin-3 sequences for the chicken and the frog also contain a RGD site in TB3. The platypus (monotreme) and opossum (marsupial) fibrillin-1 sequences contain the residues RGE, suggesting that the loss of the integrin-binding site within TB3 of fibrillin-1 occurred after the divergence of ray-finned fish but prior to the evolution of mammals. Within monotreme and marsupial fibrillin-3 sequences, the glycine residue is absent from the integrin-binding site, a feature that is maintained in more recently diverged groups such as placental mammals (eutherians). Therefore, we conclude that prior to the initial gene duplication, the ancestral fibrillin protein contained integrin-binding sites in TB3 and TB4, as supported by the presence of these sites in ray-finned fish fibrillin-1 and fibrillin-2/3. The fibrillin-1 RGD site in TB3 was lost prior to the evolution of amphibians, while the RGD site in TB3 of fibrillin-2/3 remained even after the second gene duplication, where it is conserved in all fibrillin-2 sequences, as well as in fibrillin-3 sequences of amphibians and birds. As the fibrillin-3 sequence continued to change, the evolution of mammals brought with it the loss of the RGD site in TB3, and only the most recent fibrillin-3 sequences, including that of humans and chimpanzees, experienced the addition of an integrin-binding site within cbEGF18. Therefore, we propose that the integrin-binding sites of fibrillin-2, found within TB3 and TB4, are the most conserved and representative of the RGD sites in the single fibrillin protein that exists in cephalochordates and agnathans. Thus, the presence of integrin-binding sites is highly conserved in chordates, although gene duplication allowed for the loss of the binding sites in TB3 in fibrillin-1 and -3. In addition to these RGD sites, which provide insight into the evolutionary relationship of fibrillins, several other RGD sites were noted in various fibrillin domains of organisms containing only one fibrillin gene. For completeness, these sites are summarized in Table S2.

Analysis of the Unique Region. Fibrillin isoforms in humans and other mammals can also be distinguished by a characteristic unique domain located towards the N-terminus of the protein immediately following the TB1 domain. In humans and other mammals, fibrillin-1 is characterized by a proline-rich region, fibrillin-2 by a glycine-rich region and fibrillin-3 by a mixed proline-glycine-rich region. Here, we collectively use the

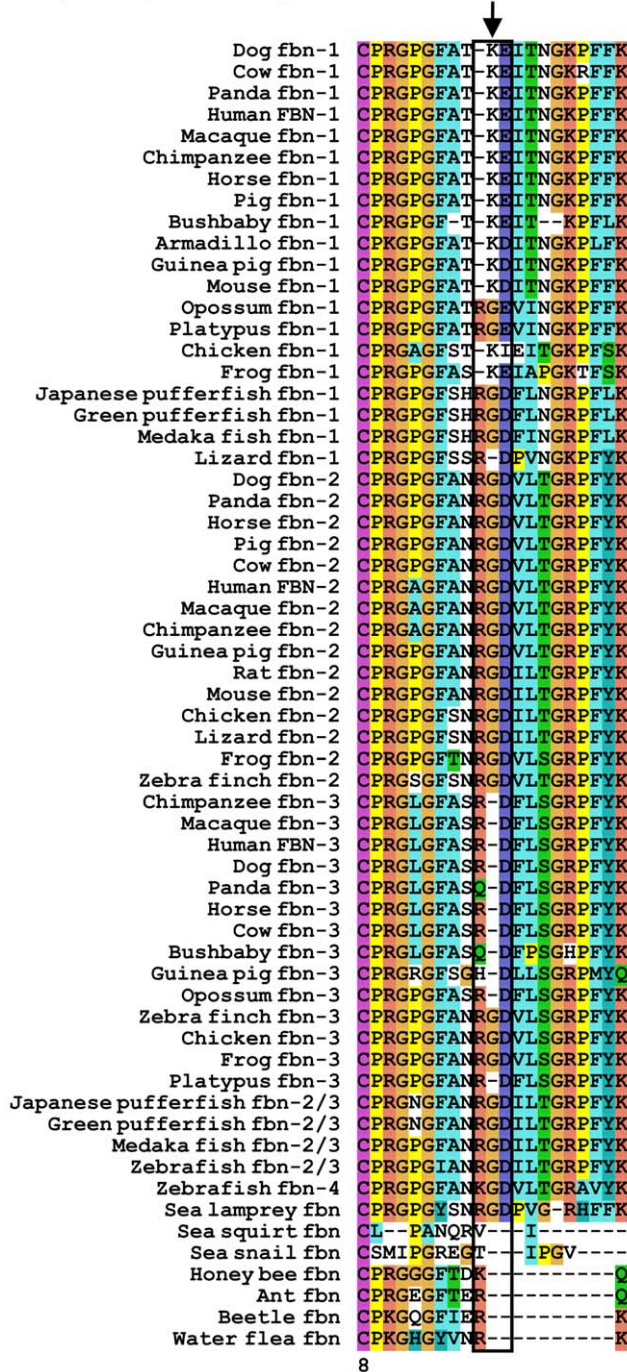
Table 1. Comparison of bootstrap values.

	TB1	TB2	TB3	TB4	TB5	TB6	TB7	Concatenated TB domains (DB/ML)	Complete Protein (DB/ML)
FBN1:FBN2/3 split	94.3	98.0	80.5	46.4	45.4	76.5	53.2	100/100	99.9/100
FBN2:FBN3 split	47.5	66.5	31.4	21.0	26.3	44.4	29.8	100/87	96.7/100
Number of sequences	59	65	62	67	60	65	69	48	37

Bootstrap values, that is, the percentage of trees yielding the same two-set partition of sequences for a given node, are indicated in percentages (two top rows). Bootstrap values were generated from 1,000 data sets. Distance-based phylogenetic trees were generated with individual TB domains (TB1–TB7). Distance based (DB) and maximum likelihood (ML) based methods were generated with all TB domains of a fibrillin protein concatenated together and with the entire fibrillin sequences as indicated. The numbers of fibrillin sequences included in the respective analyses are indicated in the bottom row.

doi:10.1371/journal.pone.0033560.t001

A) TB3, partial sequence



B) TB4, partial sequence

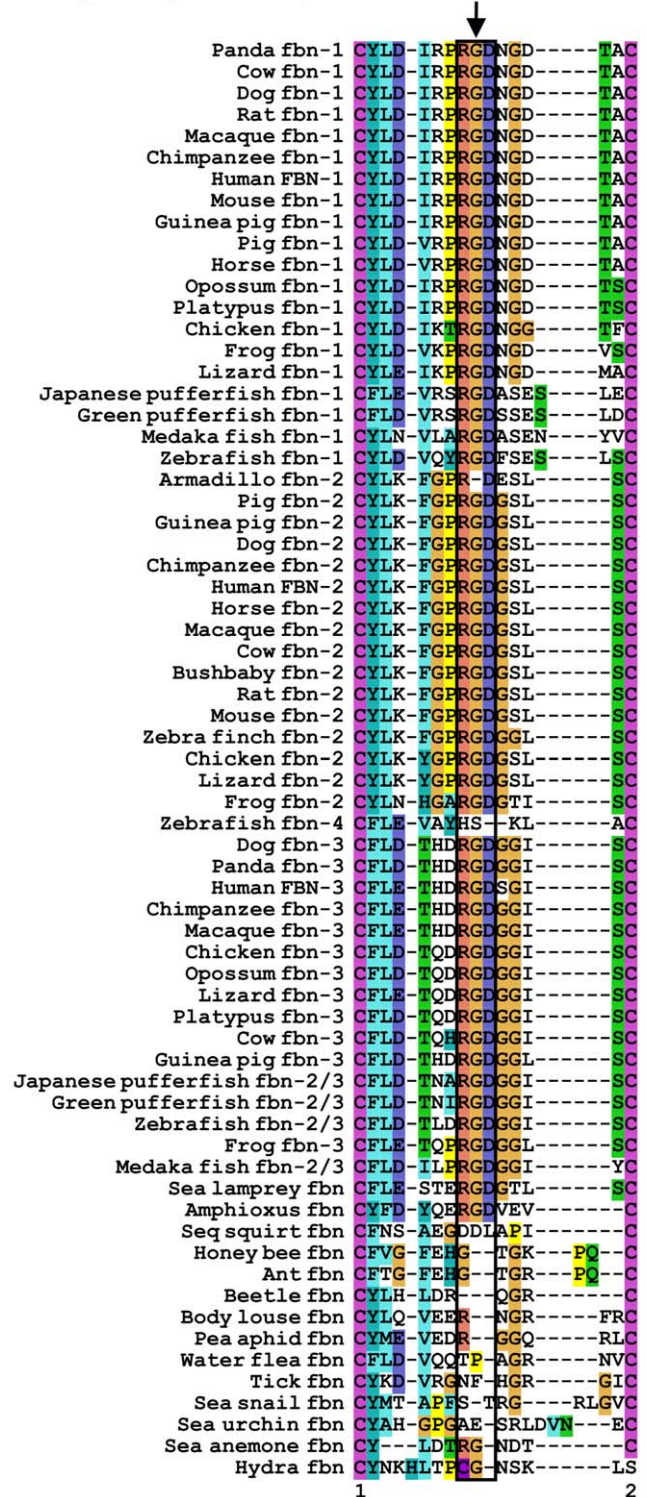


Figure 3. ClustalX alignment of RGD cell interaction sites in A) TB3 and B) TB4. Only the relevant loop containing the RGD sequence is shown. RGD sequences are highlighted by a box and an arrow. The relative numbers of the cysteine residues within the respective TB domain is indicated on the bottom. Note for TB3 the presence of a RGD integrin-binding site in sea lamprey, ray-finned fish fibrillin-1 and fibrillin-2/3, all fibrillin-2 species and several tetrapod fibrillin-3 sequences. For TB4, the RGD integrin-binding site is present in all vertebrate fibrillins, with the exception of armadillo fibrillin-2. Invertebrate fibrillin sequences do not contain an integrin-binding site in TB4.
doi:10.1371/journal.pone.0033560.g003

Table 2. Amino acid composition of the unique region.

Sequence	Most common amino acid (number of sequences/percentage of unique region)	Second-most common amino acid (number of sequences/percentage of unique region)
Fibrillin-1	Proline (18/40%)	Valine (14/14%)
Fibrillin-2/3	Glycine (3/46%)	Asparagine (3/16%)
Fibrillin-2	Glycine (15/38%)	Proline (14/12%)
Pre-eutherian fibrillin-3	Glycine (4/31%)	Proline (4/20%)
Fibrillin-3	Glycine (7/24%)	Proline (4/18%) Leucine & Proline (3/16%)

The most common and second-most common amino acids are shown for each of the 3 modern fibrillins, as well as for ray-finned fish fibrillin-2/3 and pre-eutherian fibrillin-3. The use of the ampersand symbol (&) indicates that both amino acids occur at the same frequency.
doi:10.1371/journal.pone.0033560.t002

diverged prior to placental mammals are characterized by a lower glycine:proline ratio in comparison to fibrillin-2. This ratio decreases even further in eutherian fibrillin-3. While chimpanzee and human fibrillin-3 sequences do maintain this low glycine:proline ratio, they are also characterized by an equal number of leucines and prolines as the second-most common amino acid. Based on these relative amounts of the most common amino acid residues, we speculate that the unique regions in all three present-day fibrillins serve different functions.

Surprisingly, the analysis of the region following TB1 revealed the absence of a characteristic unique region from all organisms containing a single fibrillin sequence. It should be noted, however, that we could not determine with confidence the absence of the unique region in hydra, sea urchin, lamprey and amphioxus due to sequence uncertainties in this region. The sequence of amino acids immediately downstream of TB1 does not contain any prevalent amino acid. Domain analysis reveals that this region consists of a linker followed by an EGF domain. Figure 4 shows the ClustalX alignment for this region in ancestral fibrillins in the context of the vertebrate fibrillin unique regions. This EGF domain contains six characteristically spaced cysteines, as well as all important residues for calcium-binding, except for the first residue of this domain, which differs from the typical calcium-binding consensus sequence in that it is not the required aspartic acid or asparagine residue [20,21]. Since this critical amino acid residue for calcium binding is not present at this position, we annotate this domain as a cbEGF-like domain. Furthermore, given the absence of this residue from each of these sequences, we argue against the possibility of a sequencing error or codon split across an intron/exon boundary. Of further interest, immediately downstream of this cbEGF-like domain, the single fibrillin sequences have a *bona fide* cbEGF domain replacing the (non-calcium-binding) EGF4 domain that typically follows the unique regions in mammalian fibrillins. Thus, the single fibrillin sequences found in protostomes and urochordates contain a cbEGF-like domain, followed by a true cbEGF domain in place of the unique region and the fourth EGF domain, whereas in echinoderms and cephalochordates, this last domain begins to resemble the EGF4 domain found in vertebrate fibrillins.

It was speculated that the proline-rich region in human fibrillin-1 fulfills a flexible hinge-region that may be critical for folding of fibrillin-1 into highly ordered microfibrils [29]. Although no fine structure of microfibrils from invertebrate organisms is available, it is clear that beaded microfibrils, very similar to those found in vertebrates, are also found in invertebrate organisms including, sea cucumber (*Cucumaria frondosa*), lobster (*Homarus americanus*), and jellyfish (*Podocoryne carnea*) [2–4]. The pair of cbEGF-like domain

followed by a cbEGF domain in invertebrate organisms is expected to adopt a relatively stiff rod-like structure, lacking a similar hypothetical flexibility of the unique regions [20,58,59]. However, the presence of a non-calcium binding cbEGF-like domain following TB1 domain suggests that this linker region lacks a typical TB-EGF interdomain interaction and thus still may fulfill a critical flexible role for the folding of fibrillin-1 into microfibrils.

The proline-rich region in human fibrillin-1 may be involved in the interaction with tropoelastin [31]. In addition, a small N-terminal fragment containing the glycine-rich region of human fibrillin-2 was demonstrated to interact with tropoelastin [30]. It is not known whether the unique region in fibrillin-3 plays a similar role. Elastin emerged in evolution with the appearance of a closed circulatory system and is exclusively found in vertebrates except for agnathans, including lamprey [43,60]. The evolution of the unique region in vertebrate fibrillins coincides with or is followed by the appearance of elastin and the first gene duplication of the single fibrillin into fibrillin-1 and fibrillin-2/3. A speculative interpretation of this data is that the unique regions were evolutionarily invented to support the interaction of fibrillins with tropoelastin, and that this interaction may require more than one fibrillin. If this is true, then the fibrillin gene duplication may have been a driving force in the development of closed circulatory systems. Additional support for this hypothesis comes from the genetic deletion of both, fibrillin-1 and -2 in mouse, which had much more severe effects on the structural integrity of elastic fibers in the aortic wall compared to animals with only a single fibrillin deleted [61,62].

Hybrid Domains. The first hybrid domain in human fibrillin-1 and -2 has been demonstrated to contain a free cysteine located on the surface of the domain and was thus suggested to contribute to potential intermolecular disulfide bonds [28]. When primary sequences of hybrid domains in fibrillins were first analyzed, it was noted that they combine sequence homology at their N-termini with TB domains and at their C-termini with EGF domains [7,26]. More recently, a high resolution structure of the human hybrid 2 domain demonstrated that these similarities in the primary sequence are fully reflected in the three-dimensional structure of this domain, proving that the hybrid domain is indeed an evolutionary “hybrid” between TB and cbEGF domains [27].

All high quality fibrillin sequences contain the first 9-cysteine hybrid domain, confirming previous findings [43]. The only exception was the absence of the extra cysteine from dog fibrillin-3. Given that all other collected sequences contain this additional cysteine, we suspect that this is likely due to a sequencing error.

A) N-terminal domain, partial sequence

Dog fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Pig fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Panda fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Horse fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Mouse fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Rat fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Guinea pig fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Opossum fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Lizard fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Zebra finch fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Chicken fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Chimpanzee fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Macaque fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Human FBN-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Cow fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Frog fbn-1 PNVCGSRYNAYCCPGWKTLPGGNQCIVPI
Japanese pufferfish fbn-1 PNVCGSRNNAYCCPGWKTLPGGNQCIVPI
Green pufferfish fbn-1 PNVCGSRNNAYCCPGWKTLPGGNQCIVPI
Medaka fish fbn-1 PNVCGSRNNAYCCPGWKTLPGGNQCIVPI
Dog fbn-1 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Pig fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Cow fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Guinea pig fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Panda fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Chicken fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Zebra finch fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Frog fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Armadillo fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Lizard fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Chimpanzee fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Human FBN-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Macaque fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Horse fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Mouse fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Rat fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Bushbaby fbn-2 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Macaque fbn-3 PNVCGSRFHAYCCPGWRTFPGRSQCVVPI
Chimpanzee fbn-3 PNVCGSRFHAYCCPGWRTFPGRSQCVVPI
Human FBN-3 PNVCGSRFHAYCCPGWRTFPGRSQCVVPI
Dog fbn-3 PNVCGSRFHAYCCPGWRTLPGGNQCIVPI
Panda fbn-3 PNVCGSRFHAYCCPGWRTLPGGNQCIVPI
Pig fbn-3 PNVCGSRFHAYCCPGWRTLPGGNQCIVPI
Cow fbn-3 PNVCGSRFHAYCCPGWRTLPGGNQCIVPI
Opossum fbn-3 PNVCGSRFHSYCCPGWRTLPGGNQCIVPI
Platypus fbn-3 PNVCGSRFHSYCCPGWRTLPGGNQCIVPI
Zebra finch fbn-3 PNVCGSRFHSYCCPGWRTLPGGNQCIVPI
Chicken fbn-3 PNVCGSRFHSYCCPGWRTLPGGNQCIVPI
Frog fbn-3 PNVCGSRFHSYCCPGWRTLPGGNQCIVPI
Green pufferfish fbn-2/3 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Japanese pufferfish fbn-2/3 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Medaka fish fbn-2/3 PNVCGSRFNSYCCPGWKTLAGGNQCIVPI
Zebrafish fbn-2/3 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Zebrafish fbn-4 PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Sea lamprey fbn PNVCGSRFHSYCCPGWKTLPGGNQCIVPI
Sea snail fbn PNVCGSRFRVYCCSGWLRPRGSRHCNIQL
Honey bee fbn PNVCRSRYKNYCCPGWTKKPKETGLCVIPI
Ant fbn PNVCRSRYKNYCCPGWTKKPKETGLCVIPI
Beetle fbn PNIICRSRYSTYCCPGWTHNTLGLCIIPV
Pea aphid fbn PNVCRSFRFNYCCPGWSLKHATGLCIVPV
Sea anemone fbn PHVCGLLNPSCCPGWQRSSGLCIVPV
Sea squirt fbn RHVCHKEKGSTFCCPGWTASTSGKTCISIPK

B) C-terminal domain, partial sequence

Chicken fbn-1 EACYECKI
Cow fbn-1 EACYECKI
Pig fbn-1 EACYECKI
Horse fbn-1 EACYECKI
Panda fbn-1 EACYECKI
Guinea pig fbn-1 EACYECKI
Rat fbn-1 EACYECKI
Mouse fbn-1 EACYECKI
Dog fbn-1 EACYECKI
Armadillo fbn-1 EACYECKI
Human FBN-1 EACYECKI
Chimpanzee fbn-1 EACYECKI
Macaque fbn-1 EACYECKI
Opossum fbn-1 EACYECKI
Green pufferfish fbn-1 EACYECKI
Japanese pufferfish fbn-1 EACYECKI
Medaka fish fbn-1 EACYECKI
Lizard fbn-1 EGCYDCKI
Frog fbn-1 EACYDCKI
Dog fbn-2 EACYECKI
Panda fbn-2 EACYECKI
Horse fbn-2 EACYECKI
Cow fbn-2 EACYECKI
Armadillo fbn-2 EACYECKI
Macaque fbn-2 EACYECKI
Chimpanzee fbn-2 EACYECKI
Human FBN-2 EACYECKI
Chicken fbn-2 EACYECKI
Zebra finch fbn-2 EVCYECKI
Frog fbn-2 EGCYEQQL
Lizard fbn-2 EACYECKI
Guinea pig fbn-2 EACYECKI
Rat fbn-2 EACYECKI
Mouse fbn-2 EACYECKI
Bushbaby fbn-2 ETCYECKV
Pig fbn-2 ETCYECKI
Horse fbn-3 EACYECKI
Panda fbn-3 EACYECKI
Pig fbn-3 EACYECKI
Cow fbn-3 EACYECKI
Bushbaby fbn-3 EACYECKI
Dog fbn-3 EACYECKI
Chicken early fbn-3 ETCYECKI
Lizard early fbn-3 EACYECKL
Human fbn-3 EACYECKI
Chimpanzee fbn-3 EACYECKI
Macaque fbn-3 EACYECKI
Zebrafish fbn-2/3 EACYECKI
Medaka fish fbn-2/3 EACYECKI
Japanese pufferfish fbn-2/3 EACYECKI
Green pufferfish fbn-2/3 EACYECKI
Frog fbn-3 ENCYECKI
Platypus fbn-3 DTCYECKI
Opossum fbn-3 DTCYECKI
Amphioxus fbn DTCYENL
Honey bee fbn EGCFSCKV
Water flea fbn EGCFSCKV
Beetle fbn EGCFSCKV
Ant fbn EGCFSCKV
Body louse fbn EGCFSCKV
Pea aphid fbn EGCFSCKI
Tick fbn EGCYSCKL
Sea snail fbn EGCYHCDK
Worm fbn EGCYCDQ
Sea squirt fbn SVCYACMP

Figure 6. ClustalX alignment of the regions containing cysteines in the unique A) N- and B) C-terminal domains. Note the high conservation of the six cysteine residues and the surrounding sequences in both domains.
doi:10.1371/journal.pone.0033560.g006

for the development of elastic fibers and a closed circulatory system. Following the first duplication, one sequence underwent the loss of the TB3 RGD site, as well as several other changes, in order to form fibrillin-1, while the second sequence, coined as fibrillin-2/3, retained the characteristics of the parent sequence

and sustained the introduction of a glycine-rich region. The second duplication event coincides with the evolution of tetrapods, thus allowing ray-finned fish to evolve in parallel and develop the present-day form of fibrillin-2/3 found in these species. Following this duplication, one fibrillin copy maintained

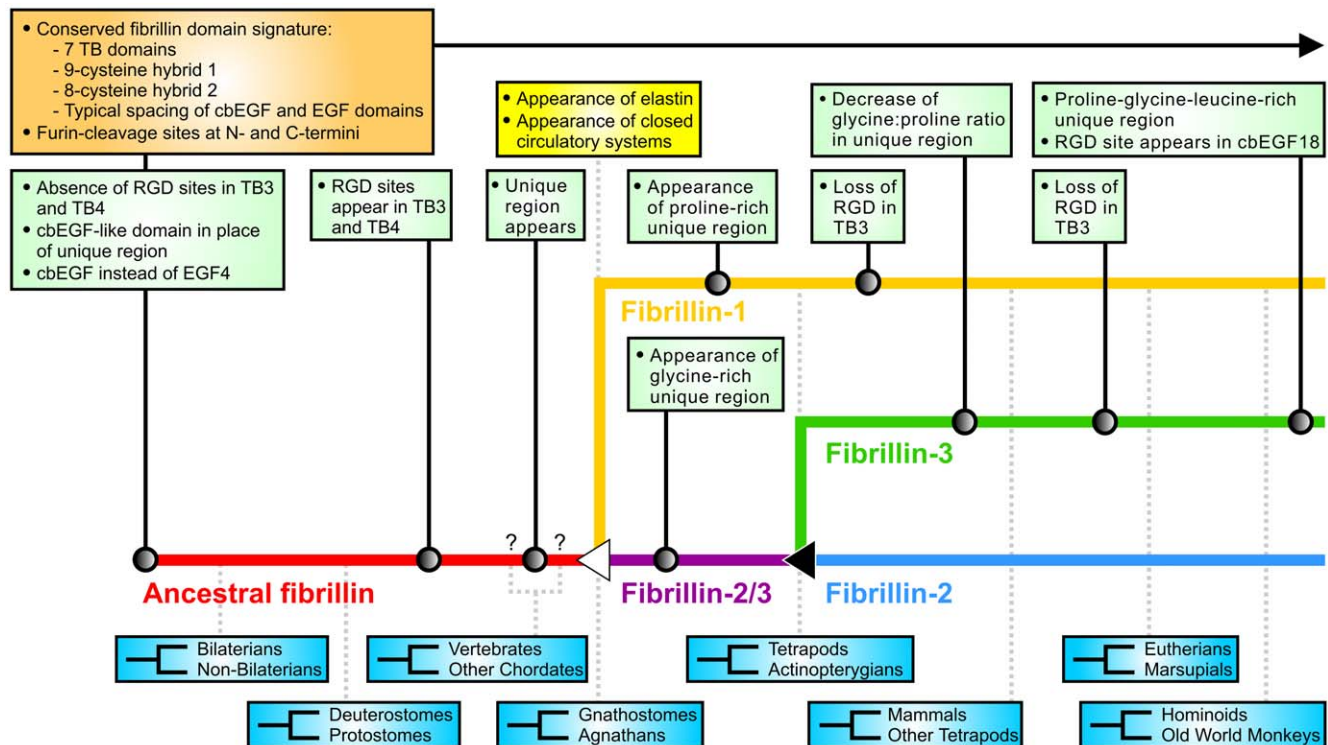


Figure 7. Hypothetical evolution of human fibrillins. An ancestral fibrillin (red) existed prior to the split of bilaterians from non-bilaterians. A gene duplication event (white triangle), presumably at the divergence of gnathostomes (jawed vertebrates) from agnathans (jawless fish), led to the genes for fibrillin-1 (orange) and the ancestral fibrillin-2 and -3 gene, coined as fibrillin-2/3 (purple), and further coincides with the appearance of elastin and the evolution of closed circulatory systems. A second duplication event (black triangle), coinciding with the evolution of tetrapods, led to the genes for fibrillin-2 (blue) and fibrillin-3 (green). The pre-eutherian fibrillin-3 gene underwent significant evolutionary changes to become present-day fibrillin-3. Taxonomic classifications and evolutionary divergence events are found in blue boxes and indicated by fork symbols, with dotted light gray connectors to the individual fibrillins. The upper classification shares a common ancestor with humans. Question marks indicate uncertainties as to whether the unique region appeared before or after the divergence of vertebrates from other chordates. Spacing between taxonomic classifications is not proportional to their evolutionary timeline. Identified conserved characteristic features and evolutionary events found in all analyzed fibrillins are highlighted in an orange-colored box with a black arrow. Evolutionary characteristics and events specific to individual fibrillins are shown in green boxes with solid connectors and circles. Important related evolutionary events are shown in a yellow box. doi:10.1371/journal.pone.0033560.g007

the features of the parent sequence to become fibrillin-2, while the other, underwent significant evolutionary changes to form fibrillin-3. The transition of the initial tetrapod fibrillin-3 sequence to the fibrillin-3 sequence found in humans and other mammals includes the loss of the RGD site in TB3, the novel introduction of an RGD site in cbEGF18, as well as a reduction of the glycine:proline ratio in the unique region. Divergence of the fibrillin-3 gene is observed with the evolution of mammals, evident through the loss of the RGD site in TB3 of monotremes and marsupials. Highly conserved sequences throughout the evolution of fibrillins include the pro-protein furin-type processing sites within the unique N- and C-terminal domains, indicating an evolutionarily conserved microfibril assembly mechanism with regards to pro-protein processing. Other notable highly conserved regions in all analyzed fibrillins are the cysteine patterns in the unique N- and C-terminal domains. The results from this study were used to propose key events for the evolution of the three human fibrillins (Fig. 7).

References

1. Kiely CM, Sherratt MJ, Marson A, Baldock C (2005) Fibrillin microfibrils. *Adv Protein Chem* 70: 405–436.
2. Reber-Müller S, Spisinger T, Schuchert P, Spring J, Schmid V (1995) An extracellular matrix protein of jellyfish homologous to mammalian fibrillins

Supporting Information

Text S1 Additional information about naming fibrillins, zebra-fish fibrillins, and sea urchin fibrillin. (PDF)

Table S1 Sequences analyzed in this study. (PDF)

Table S2 Additional RGD integrin-binding sites in ancestral fibrillin sequences. (PDF)

Acknowledgments

We are grateful to Glen Corson for insightful discussions.

Author Contributions

Conceived and designed the experiments: APG DPR. Performed the experiments: APG. Analyzed the data: APG WS DPR. Wrote the paper: APG DPR.

- forms different fibrils depending on the life stage of the animal. *Dev Biol* 169: 662–672.
3. Davison IG, Wright GM, DeMont ME (1995) The structure and physical properties of invertebrate and primitive vertebrate arteries. *J Exp Biol* 198: 2185–2196.
 4. Thurmond FA, Trotter JA (1996) Morphology and biomechanics of the microfibrillar network of sea cucumber dermis. *J Exp Biol* 199: 1817–1828.
 5. Lee B, Godfrey M, Vitale E, Hori H, Mattei MG, et al. (1991) Linkage of Marfan syndrome and a phenotypically related disorder to two different fibrillin genes. *Nature* 352: 330–334.
 6. Maslen CL, Corson GM, Maddox BK, Glanville RW, Sakai LY (1991) Partial sequence of a candidate gene for the Marfan syndrome. *Nature* 352: 334–337.
 7. Corson GM, Chalberg SC, Dietz HC, Charbonneau NL, Sakai LY (1993) Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure and alternatively spliced exons at the 5' end. *Genomics* 17: 476–484.
 8. Zhang H, Apfelroth SD, Hu W, Davis EC, Sanguineti C, et al. (1994) Structure and expression of fibrillin-2, a novel microfibrillar component preferentially located in elastic matrices. *J Cell Biol* 124: 855–863.
 9. Nagase T, Nakayama M, Nakajima D, Kikuno R, Ohara O (2001) Prediction of the coding sequences of unidentified human genes. XXII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res* 8: 85–95.
 10. Wagenseil JE, Mecham RP (2007) New insights into elastic fiber assembly. *Birth Defects Res C Embryo Today* 81: 229–240.
 11. Ramirez F, Rifkin DB (2009) Extracellular microfibrils: contextual platforms for TGF-beta and BMP signaling. *Curr Opin Cell Biol* 21: 616–622.
 12. Wagenseil JE, Mecham RP (2009) Vascular extracellular matrix and arterial mechanics. *Physiol Rev* 89: 957–989.
 13. Robinson P, Arteaga-Solis E, Baldock C, Collod-Beroud G, Booms P, et al. (2006) The molecular genetics of Marfan syndrome and related disorders. *J Med Genet* 43: 769–787.
 14. Loeyts BL, Gerber EE, Riegert-Johnson D, Iqbal S, Whiteman P, et al. (2010) Mutations in Fibrillin-1 Cause Congenital Scleroderma: Stiff Skin Syndrome. *Sci Transl Med* 2: 23ra20.
 15. Le Goff C, Mahaut C, Wang LW, Allali S, Abhyankar A, et al. (2011) Mutations in the TGFbeta binding-protein-like domain 5 of FBN1 are responsible for acromioclavicular and geophytic dysplasias. *Am J Hum Genet* 89: 7–14.
 16. Putnam EA, Zhang H, Ramirez F, Milewicz DM (1995) Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder, congenital contractural arachnoidactyly. *Nat Genet* 11: 456–458.
 17. Corson GM, Charbonneau NL, Keene DR, Sakai LY (2004) Differential expression of fibrillin-3 adds to microfibril variety in human and avian, but not rodent, connective tissues. *Genomics* 83: 461–472.
 18. Hubmacher D, Sabatier L, Annis DS, Mosher DF, Reinhardt DP (2011) Homocysteine modifies structural and functional properties of fibronectin and interferes with the fibronectin-fibrillin-1 interaction. *Biochemistry* 50: 5322–5332.
 19. Campbell ID, Bork P (1993) Epidermal growth factor-like modules. *Curr Opin Struct Biol* 3: 385–392.
 20. Downing AK, Knott V, Werner JM, Cardy CM, Campbell ID, et al. (1996) Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell* 85: 597–605.
 21. Handford PA, Mayhew M, Baron M, Winship PR, Campbell ID, et al. (1991) Key residues involved in calcium-binding motifs in EGF-like domains. *Nature* 351: 164–167.
 22. Hubmacher D, Reinhardt DP (2011) Microfibrils and fibrillin. In: Mecham RP, ed. *Biology of Extracellular Matrix*. New York: Springer. pp 233–265.
 23. Yuan X, Downing AK, Knott V, Handford PA (1997) Solution structure of the transforming growth factor beta-binding protein-like module, a domain associated with matrix fibrils. *EMBO J* 16: 6659–6666.
 24. Lack J, O'Leary JM, Knott V, Yuan X, Rifkin DB, et al. (2003) Solution structure of the third TB domain from LTBP1 provides insight into assembly of the large latent complex that sequesters latent TGF-beta. *J Mol Biol* 334: 281–291.
 25. Lee SS, Knott V, Jovanovic J, Harlos K, Grimes JM, et al. (2004) Structure of the integrin binding fragment from fibrillin-1 gives new insights into microfibril organization. *Structure* 12: 717–729.
 26. Pereira L, D'Alessio M, Ramirez F, Lynch JR, Sykes B, et al. (1993) Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan syndrome. *Hum Mol Genet* 2: 961–968.
 27. Jensen SA, Iqbal S, Lowe ED, Redfield C, Handford PA (2009) Structure and interdomain interactions of a hybrid domain: a disulphide-rich module of the fibrillin/LTBP superfamily of matrix proteins. *Structure* 17: 759–768.
 28. Reinhardt DP, Gambee JE, Ono RN, Bächinger HP, Sakai LY (2000) Initial steps in assembly of microfibrils. Formation of disulfide-cross-linked multimers containing fibrillin-1. *J Biol Chem* 275: 2205–2210.
 29. Baldock C, Koster AJ, Ziese U, Rock MJ, Sherratt MJ, et al. (2001) The supramolecular organization of fibrillin-rich microfibrils. *J Cell Biol* 152: 1045–1056.
 30. Trask TM, Crippes Trask B, Ritty TM, Abrams WR, Rosenbloom J, et al. (2000) Interaction of tropoelastin with the amino-terminal domains of fibrillin-1 and fibrillin-2 suggests a role for the fibrillins in elastic fiber assembly. *J Biol Chem* 275: 24400–24406.
 31. Rock MJ, Cain SA, Freeman LJ, Morgan A, Mellody K, et al. (2004) Molecular basis of elastic fiber formation. Critical interactions and tropoelastin-fibrillin-1 cross-link. *J Biol Chem* 279: 23748–23758.
 32. Molloy SS, Bresnahan PA, Lepplä SH, Klimpel KR, Thomas G (1992) 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* 267: 16396–16402.
 33. Taylor NA, van de Ven WJM, Creemers JWM (2003) Curbing activation: proinhibitor convertases in homeostasis and pathology. *FASEB J* 17: 1215–1227.
 34. Reinhardt DP, Keene DR, Corson GM, Pöschl E, Bächinger HP, et al. (1996) Fibrillin 1: organization in microfibrils and structural properties. *J Mol Biol* 258: 104–116.
 35. Lönnqvist L, Reinhardt DP, Sakai LY, Peltonen L (1998) Evidence for furin-type activity-mediated C-terminal processing of profibrillin-1 and interference in the processing by certain mutations. *Hum Mol Genet* 7: 2039–2044.
 36. Raghunath M, Putnam EA, Ritty T, Hamstra D, Park ES, et al. (1999) Carboxy-terminal conversion of profibrillin to fibrillin at a basic site by PACE/furin-like activity required for incorporation in the matrix. *J Cell Sci* 112: 1093–1100.
 37. Ritty TM, Broekelmann T, Tisdale C, Milewicz DM, Mecham RP (1999) Processing of the fibrillin-1 carboxyl-terminal domain. *J Biol Chem* 274: 8933–8940.
 38. Milewicz D, Pyeritz RE, Crawford ES, Byers PH (1992) Marfan syndrome: defective synthesis, secretion, and extracellular matrix formation of fibrillin by cultured dermal fibroblasts. *J Clin Invest* 89: 79–86.
 39. Pfaff M, Reinhardt DP, Sakai LY, Timpl R (1996) Cell adhesion and integrin binding to recombinant human fibrillin-1. *FEBS Lett* 384: 247–250.
 40. Sakamoto H, Broekelmann T, Chersesh DA, Ramirez F, Rosenbloom J, et al. (1996) Cell-type specific recognition of RGD- and non-RGD-containing cell binding domains in fibrillin-1. *J Biol Chem* 271: 4916–4922.
 41. Bax DV, Bernard SE, Lomas A, Morgan A, Humphries J, et al. (2003) Cell adhesion to fibrillin-1 molecules and microfibrils is mediated by alpha5 beta1 and alpha5 beta3 integrins. *J Biol Chem* 278: 34605–34616.
 42. Jovanovic J, Takagi J, Choulier L, Abrescia NG, Stuart DI, et al. (2007) Alpha v beta 6 is a novel receptor for human fibrillin-1: comparative studies of molecular determinants underlying integrin-RGD affinity and specificity. *J Biol Chem* 282: 6743–6751.
 43. Robertson I, Jensen S, Handford P (2011) TB domain proteins: evolutionary insights into the multifaceted roles of fibrillins and LTBP. *Biochem J* 433: 263–276.
 44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410.
 45. Lipman DJ, Pearson WR (1985) Rapid and sensitive protein similarity searches. *Science* 227: 1435–1441.
 46. Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85: 2444–2448.
 47. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
 48. von Heijne G, Abrahamsen L (1989) Species-specific variation in signal peptide design. Implications for protein secretion in foreign hosts. *FEBS Lett* 244: 439–446.
 49. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340: 783–795.
 50. Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27: 1164–1165.
 51. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8: 275–282.
 52. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
 53. Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* 57: 758–771.
 54. Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE) New Orleans, LA, USA*. pp 1–8.
 55. Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164–166.
 56. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425.
 57. Hubmacher D, Tiedemann K, Reinhardt DP (2006) Fibrillins: From biogenesis of microfibrils to signaling functions. *Curr Top Dev Biol* 75: 93–123.
 58. Cardy CM, Handford PA (1998) Metal ion dependency of microfibrils supports a rod-like conformation for fibrillin-1 calcium-binding epidermal growth factor-like domains. *J Mol Biol* 276: 855–860.
 59. Reinhardt DP, Mechling DE, Boswell BA, Keene DR, Sakai LY, et al. (1997) Calcium determines the shape of fibrillin. *J Biol Chem* 272: 7368–7373.
 60. Sage H, Gray WR (1979) Studies on the evolution of elastin-I. Phylogenetic distribution. *Comp Biochem Physiol B* 64: 313–327.
 61. Carta L, Pereira L, Arteaga-Solis E, Lee-Arteaga SY, Lenart B, et al. (2006) Fibrillins 1 and 2 perform partially overlapping functions during aortic development. *J Biol Chem* 281: 8016–8023.
 62. Arteaga-Solis E, Gayraud B, Lee SY, Shum L, Sakai LY, et al. (2001) Regulation of limb patterning by extracellular microfibrils. *J Cell Biol* 154: 275–281.
 63. Wess TJ, Purslow PP, Sherratt MJ, Ashworth J, Shuttleworth CA, et al. (1998) Calcium determines the supramolecular organization of fibrillin-rich microfibrils. *J Cell Biol* 141: 829–837.
 64. Wallis DD, Putnam EA, Cretoiu JS, Carmical SG, Cao SN, et al. (2003) Profibrillin-1 maturation by human dermal fibroblasts: proteolytic processing and molecular chaperones. *J Cell Biochem* 90: 641–652.
 65. Saharinen J, Keski-Oja J (2000) Specific sequence motif of 8-Cys repeats of TGF-beta binding proteins, LTBP, creates a hydrophobic interaction surface for binding of small latent TGF-beta. *Mol Biol Cell* 11: 2691–2704.