Abstract

Background: The syndecans are the major family of transmembrane proteoglycans in animals and are known for multiple roles in cell interactions and growth factor signalling during development, inflammatory response, wound-repair and tumorigenesis. Although syndecans have been cloned from several invertebrate and vertebrate species, the extent of conservation of the family across the animal kingdom is unknown and there are gaps in our knowledge of chordate syndecans. Here, we develop a new level of knowledge for the whole syndecan family, by combining molecular phylogeny of syndecan protein sequences with analysis of the genomic contexts of syndecan genes in multiple vertebrate organisms.

Results: We identified syndecan-encoding sequences in representative Cnidaria and throughout the Bilateria. The C1 and C2 regions of the cytoplasmic domain are highly conserved throughout the animal kingdom. We identified in the variable region a universally-conserved leucine residue and a tyrosine residue that is conserved throughout the Bilateria. Of all the genomes examined, only tetrapod and fish genomes encode multiple syndecans. No syndecan-1 was identified in fish. The genomic context of each vertebrate syndecan gene is syntenic between human, mouse and chicken, and this conservation clearly extends to syndecan-2 and -3 in T. nigroviridis. In addition, tetrapod syndecans were found to be encoded from paralogous chromosomal regions that also contain the four members of the matrilin family. Whereas the matrilin-3 and syndecan-1 genes are adjacent in tetrapods, this chromosomal region appears to have undergone extensive lineage-specific rearrangements in fish.

Conclusion: Throughout the animal kingdom, syndecan extracellular domains have undergone rapid change and elements of the cytoplasmic domains have been very conserved. The four syndecan genes of vertebrates are syntenic across tetrapods, and syteny of the syndecan-2 and -3 genes is apparent between tetrapods and fish. In vertebrates, each of the four family members are encoded from paralogous genomic regions in which members of the matrilin family are also syntenic between tetrapods and fish. This genomic organization appears to have been set up after the divergence of urochordates (Ciona) and vertebrates. The syndecan-1 gene appears to have been lost relatively early in the fish lineage. These conclusions provide the basis for a new model of syndecan evolution in vertebrates and a new perspective for analyzing the roles of syndecans in cells and whole organisms.
Background
Proteoglycans are important and ancient mediators of cell interactions in metazoan organisms. In the simplest multicellular animals, sponges, extracellular proteoglycans contribute to adhesion between cells and processes of self-recognition and host defence [1,2]. The syndecans are the only major family of transmembrane proteoglycans and are conserved in metazoa from nematode worms to man. The genomes of invertebrates such as C. elegans and D. melanogaster contain a single syndecan gene, whereas mammalian genomes contain four syndecan genes [3,4]. In all the organisms in which they have been studied, syndecans have important roles in cell interactions, adhesion, migration and receptor signaling. The single syndecans of C. elegans and D. melanogaster are required for proper axon guidance during development of the nervous system [5-8]. C. elegans syndecan also functions in vulval development [9]. The vertebrate syndecans each have distinct tissue expression patterns and distinct functional attributes. For example, syndecan-1 is expressed in epithelia and certain mesenchymal cells. Syndecan-1 null mice show deficits in cell migration behaviors associated with inflammatory response and wound-healing [10,11]. In single cells, syndecan-1 activates formation of filopodia and lamellipodia that mediate cell motility [12,13]. Syndecan-2 is expressed by mesenchymal, liver and neuronal cells. It participates in the development of the left/right body axis during Xenopus development [14], has been identified to function in the assembly of neuronal dendritic spines and in ECM assembly in mammalian systems, and is essential for angiogenesis in zebrafish [15-17]. Syndecan-3 is mostly expressed in the nervous system and is important in hippocampal function, particularly in controlling feeding behavior and hippocampus-dependent memory [18,19]. Syndecan-4 is broadly expressed by many tissues and cell types and has specific signaling roles as an integrin co-receptor in the assembly of focal adhesions. It also contributes to angiogenesis and wound-healing in mice [4,20,21].

In mammals, syndecans have significant clinical relevance in the pathologies of infection, cancer and wound repair. Syndecan-1 is a receptor for human immunodeficiency virus in cell culture [22] and contributes to the pathology of infection by Pseudomonas aeruginosa and Staphylococcus aureus [23-25]. Syndecan-1 deficient mice have increased resistance to wingless-1-mediated tumorigenesis [26]. Altered expression of syndecan-1 has been documented in many human carcinomas, lymphomas and in multiple myeloma; in multiple myeloma syndecan-1 has been proposed for clinical use as a prognostic marker [27-34]. Other studies have implicated syndecans in tissue repair. Thus, the lack of syndecan-1 or -4 in mice results in impaired inflammatory responses, wound-healing, and angiogenesis [10,21,35]. Studies in animal models have demonstrated increased expression of syndecan-1, -3 and -4 after cardiac injury, indicating possible roles in cardiac remodeling [36-38].

In structure, all syndecans consist of an extracellular domain with sites for attachment of glycosaminoglycan (GAG) sidechains, a single transmembrane domain and a short cytoplasmic domain. Through their GAG chains, syndecans act as co-receptors for the cell-surface binding of heparin-binding growth factors such as bFGFs (basic fibroblast growth factor) and Wnt (wingless) [3,4,26,39]. The activity of invertebrate syndecans in axon guidance depends on co-receptor activity in the slit/Robo pathway [5,8]. Direct protein-protein interactions of the extracellular domain have also been implied for syndecan-1 and -4 [40]. Based on comparisons of mammalian, C. elegans and D. melanogaster syndecans, the cytoplasmic domains are recognized to contain two conserved regions, designated C1 and C2, that are present in all these syndecans, and a central variable, (V), region that is unique to each form of syndecan. The V regions themselves are very highly conserved between species orthologues of syndecans -1 to -4, suggestive of important specific functions [3,4]. In general, the cytoplasmic domains participate in the assembly of juxtamembrane complexes that regulate cell signaling and the organizational state of the actin cytoskeleton. The C2 region binds PDZ-containing proteins that form multiprotein scaffolds by oligomerization and which also mediate syndecan recycling [41]. Cask, src tyrosine kinase and synectin are known binding partners of the C1 region. The V regions have been most intensively studied in syndecan-2 and syndecan-4, and have specific binding partners that contribute to cell signaling that regulates cytoskeletal structures [3,4,15]. On the basis of their protein sequences, the four mammalian syndecans group into two pairs: syndecan-1 and -3 have higher sequence identity with each other, as do syndecan-2 and syndecan-4 [3,4].

To date, the relationships of invertebrate and vertebrate syndecans have been considered only at the level of their protein sequences. With regard to these analyzes, there are gaps in our knowledge of chordate syndecans, as the most intensive studies have focused on mammalian and amphibian syndecans [e.g., 3,4,10,11,14,18,20,21]]. Invertebrate syndecans have been experimentally studied in D. melanogaster and C. elegans, but little is known about other invertebrate syndecans. It is now recognized that the genomes of D. melanogaster and C. elegans have evolved rapidly with extensive gene loss [42,43]. Thus, one purpose of our study was to capitalize on the recent sequencing of the genomes of the basal chordate, Ciona intestinalis, the chicken and three species of fish, along with the expanded knowledge from the genomes of the sea urchin Stronglyocentrotus purpuratus, the amphioxus Branchios-
toma floridae, multiple representatives of the Cnidaria, and the many expressed sequence tag projects directed to invertebrate species, to obtain a comprehensive view of this important proteoglycan family at protein sequence level. A second purpose has been to develop a new level of knowledge for the whole syndecan family, by analyzing the genomic contexts of syndecan genes in multiple organisms. The comparative genomics of gene families is a powerful approach to understanding the relationships between the members of multigene families, either within a single genome or through comparison of the conservation of genomic context across the genomes of multiple organisms. In addition, ideas about the origins of gene families can be clarified through the study of the genomes of modern animals from different phyla. The insights obtained from comparative genomics are relevant to making choices of model organisms for experimental purposes and may advance understanding of the roles of gene families at systems level [44].

Table 1: Accession numbers of Syndecans.

| Species              | Syndecan-1 | Syndecan-2 | Syndecan-3 | Syndecan-4 | References |
|----------------------|------------|------------|------------|------------|------------|
| H. sapiens           | NP_002988  | NM_002998  | NM_14654   | NM_002999  | 96–99      |
| M. musculus          | NM_011519  | NM_011519  | NM_011520  | NM_011521  | 100, 101   |
| G. gallus            | XM_419972  | NM_001001462 | NM_205383  | NM_001007869 | 101–104    |
| D. rerio             | Not found  | NM_173223  | XM_695352  | XM_691930  | 17         |
| T. nigroviridis      | Not found  | GSTENP000-18305001 | GSTENT003-1504001 | GSTENP0001-1022001 | 75         |
| X. laevis            | AAB81324   | AAB81325   | AAB81326   | BU915423, CF285173 | 49, 105    |
| X. tropicalis        | Scaffold 100* | NM_001017163 | Scaffold 411* | Scaffold 334* | 50         |
| B. floridiae         | B1386782, BW711533, BW936816 and others | 53         |
| C. intestinalis      | KYOTOGRAIL2005.125.42, Gene collection ID: CiGC35101 | 54, 76     |
| C. savignyi          | AB021960   |           |           |           | 106        |
| A. crassispina       | AB063328   |           |           |           | 107        |
| S. purpuratus        | XM_776754  |           |           |           | 108        |
| D. melanogaster      | NM_166449  |           |           |           |           |
| M. japonicus         | CI998451   |           |           |           |           |
| R. appendiculatus    | CD794925, CD788877 | 85         |
| E. scolopes          | DW251986   |           |           |           |           |
| C. elegans           | P50605     |           |           |           | 8, 9, 83   |
| S. mediterranea      | DN308360   |           |           |           | 84         |
| H. magnapapillata    | DT612491   |           |           |           |           |
| N. vectensis         | S0645      |           |           |           | 43         |
| A. tenuis            | BAE46797   |           |           |           | 86, 88     |
| A. palmata           | DR9839838, DR984551 | 88         |

All numbers are GenBank Accession numbers, except for T. nigroviridis numbers from the Genoscope database and N. vectensis from Stellabase. * scaffold numbers refer to X. tropicalis genome assembly v4.1 (JGI).

Table 2: Sequence conservation of representative invertebrate, fish and amphibian syndecans.

| Organism               | Human-S1 | Human-S2 | Human-S3 | Human-S4 |
|------------------------|----------|----------|----------|----------|
| C. savignyi            | 0.229    | 0.201    | 0.219    | 0.170    |
| A. crassispina         | 0.190    | 0.176    | -        | 0.131    |
| D. melanogaster        | 0.204    | 0.192    | 0.179    | 0.175    |
| C. elegans             | 0.182    | 0.150    | -        | 0.112    |
| T. nigroviridis, sdc2  | 0.230    | 0.260    | 0.180    | 0.190    |
| T. nigroviridis, sdc3  | 0.241    | 0.301    | 0.335    | 0.206    |
| T. nigroviridis, sdc4  | 0.210    | 0.240    | 0.187    | 0.292    |
| X. laevis, sdc1        | 0.363    | 0.207    | 0.267    | 0.170    |
| X. laevis, sdc2        | 0.224    | 0.485    | 0.222    | 0.247    |
| X. laevis, sdc3        | 0.260    | 0.238    | 0.554    | 0.196    |

Amino acid sequences of human syndecans were used to conduct BlastP searches against the Uniprot database. Conservation scores of relevant retrieved hits (representative of invertebrate, urochordate, fish and amphibian) were then calculated [92]. Dashes indicate searches for which no hit with syndecan-3 was retrieved.
Figure 1
Conservation of syndecan cytoplasmic domains.
Amino acid sequences of syndecan cytoplasmic domains from different species were aligned in ClustalW. The alignment is presented in Boxshade 3.2. Black shading shows identical amino acids; gray indicates conservative substitutions and no shading indicates unrelated amino acids. V = variable region and the very highly-conserved leucine and tyrosine residues are indicated by asterisks. The central hypervariable region (13) is also indicated. Key: Ac, Anthocidaris crassispina; Ap, Acropora palma; At, Acropora tenuis; Bf, Branchiostoma floridae; Ce, Caenorhabditis elegans; Ci, Ciona intestinalis; Cs, Ciona savignyi; Dm, Drosophila melanogaster; Dr, Danio rerio; Es, Euprymna scolopes; Gg, Gallus gallus; Hm, Hydra magnipapillata; Hs, Homo sapiens; Mj, Marsupenaeus japonicus; Mm, Mus musculus; Nv, Nemastoma vectensis; Rp, Rhipephalus appendiculatus; Sm, Schmidtea mediterranea; Sp, Strongylocentrotus purpuratus; Tn, Tetraodon nigroviridis; Xi, Xenopus laevis; Xc, Xenopus tropicalis. All syndecans are abbreviated as S.

Results
Molecular phylogeny of syndecans across the animal kingdom

Our searches of recently completed genome sequences and the database of expressed sequence tags produced an expanded dataset of syndecans that demonstrates remarkable conservation of this proteoglycan family across the animal kingdom. Syndecans were identified in multiple species of Cnidaria as well as throughout the Bilateria. The genomes of Ciona intestinalis, Strongylocentrotus purpuratus and amphioxus (Branchiostoma floridae) encode single syndecans and multiple syndecans are encoded in fish and other vertebrate genomes (Table 1). However, no syndecan-1 sequence was identified in fish, notwithstanding extensive searches of three fish genomes and the large database of expressed sequence tags from many teleost fish species.

The sequence relationships of selected invertebrate and fish syndecans to human syndecans were examined by calculating WU-BLASTP conservation scores [45]. X. laevis syndecans were included for comparison. Whereas each of the X. laevis and T. nigroviridis syndecans was closely related to one of the human syndecans (e.g., T. nigroviridis syndecan-3 has the highest score with human syndecan-3, etc), each of the invertebrate syndecans had similar low conservation scores with all the human syndecans and in some cases no hit was obtained (Table 2). This suggests that the sequences of invertebrate syndecans are similarly related to all four of the vertebrate syndecans.

The cytoplasmic domains have been recognized as the most conserved region of syndecan protein sequences [3,4]. To examine the conservation of cytoplasmic domains across the new dataset, we aligned the sequences by the CLUSTALW progressive local alignment method [46]. The alignment demonstrated near complete conservation of the C1 region in all syndecans, with some variation at the second and third positions in the vertebrate syndecan-2 group, D. rerio syndecan-4, and many of the invertebrate syndecans (Figure 1). The C2 region is almost universally conserved in Bilateria as EFYA, with the exception of Ciona syndecans that contain EYEA and fish syndecan-4s that contain EYEA. Greater variability of the C2 motif is evident in the Cnidaria (Figure 1). All of these variants are predicted to be functional PDZ-binding motifs [47].

The alignment also demonstrated that the invertebrate syndecans from different phyla contain distinctive V regions, all of which are different from those of syndecan-1, -2, -3 and -4. In vertebrate syndecans, the syndecan-4 V region is well-conserved between fish and tetrapods. In the syndecan-2 group, the V region of D. rerio syndecan-2 is very similar to that of tetrapod syndecan-2s, whereas in
T. nigroviridis the central hypervariable region is distinct at 3 residues. The V regions of syndecan-3s are well-conserved between fish, chicken, mouse and human, yet the V region of the Xenopus syndecan-3s have unique features that result in a separate clustering in the sequence alignment (Figure 1). Most strikingly, the alignment highlighted the universal conservation of the first leucine residue in the V regions of all animals. A single tyrosine residue is also very well-conserved across the Bilateria but is not present in any of the V regions from the Cnidaria (asterisks in Figure 1).

To compare all the full-length sequences in the dataset, we used the TCOFFEE alignment algorithm that combines pairwise/global and local alignment methods into a single model and is more accurate than CLUSTALW for sequences of lower identity [48]. The Phylip distance matrix output was used to prepare an unrooted tree. This analysis confirmed the absence of syndecan-1 from the fish species. Fish and amphibian syndecan-2, -3 and -4 all grouped clearly with their tetrapod counterparts. Each set of syndecan-1, -2, 3- and -4 sequences formed a distinct group within the phylogenetic diagram, and the syndecan-2 and -4 groups appeared more closely related to each other than to other syndecans. X. laevis syndecan-1 has unusual histidine-rich sequence features [49] and therefore forms an deep-branching node with tetrapod syndecan-1s; nevertheless the syndecan-1 grouping was clear (Figure 2; the sequence of X. tropicalis syndecan-1 is similar to that of X. laevis; X. tropicalis syndecan-1 is present on scaffold 100 in genome assembly 4.1) [50].

To evaluate the robustness of these groupings, further analyses of the full-length sequences and the cytoplasmic domains were conducted using the program PHYML, that estimates phylogenies based on the maximum-likelihood principle [51], with inclusion of bootstrap analysis [52]. The unrooted tree derived from the full-length sequences confirmed the distinct groupings of the sets of syndecan-1, -2, -3 and -4 and also demonstrated the syndecan-1/syndecan-3 and syndecan-2/syndecan-4 pairings. These topologies were supported by robust bootstrap values (at least 67%), with the exception of the G. gallus /X. laevis syndecan-1 node) (Figure 3A). However, the positioning of invertebrate syndecans relative to vertebrate syndecans was not well supported, likely due to the extent of divergence of their extracellular domain sequences (Figure 3A). Most of the subgroupings within the invertebrates appeared biologically meaningful and were well-supported by the bootstrap analysis, with the exception of the placements of the amphioxus sequence at a node with Cnidarian syndecans and C. elegans at a node with echinoderm syndecans (Figure 3A). It is likely that these anomalies are an artifact of long branch attraction. At present, there are no additional cephalochordate or nematode syndecan sequences to include in the dataset. The same general topologies for the vertebrate syndecans were supported in the tree prepared from the cytoplasmic domains. Again, not all nodes within the invertebrate syndecan group were biologically meaningful -M. japonicus syndecan was placed as a outgroup of the Cnidaria – and the central nodes between vertebrate and invertebrate syndecans were not robust (Figure 3B). In summary, the multiple molecular phylogenetic methods of analysis of syndecan protein sequences support the pairings of syndecan-1 with syndecan-3, and syndecan-2 with syndecan-4, and indicate that the invertebrate syndecans are distantly and equivalently related to the four vertebrate syndecans.

**Vertebrate syndecan genes show extensive conservation of synteny**
To obtain a clearer perspective on the relationships of the four vertebrate syndecans, we evaluated the vertebrate syndecan gene family at genomic level by comparing the conservation of the gene neighbors of each syndecan gene between the mapped genomes of human, mouse, chicken and, where available, fish. In each genome, each syndecan gene is located on a different chromosome. For each family member, a set of conserved gene neighbors could be identified. For the syndecan-1 gene, six local neighboring genes (HS1BP, RHOB, PUM2, LAPTMAA, MATN3 and WDR35) and, more remotely, KCNS3, were conserved in all three species (Figure 4A).

| Species       | Matrilin-1 (Accession) | Matrilin-2 (Accession) | Matrilin-3 (Accession) | Matrilin-4 (Accession) | References |
|---------------|------------------------|------------------------|------------------------|------------------------|------------|
| H. sapiens    | NM_002379              | XM_209824              | NM_002381              | NM_003833              | 55–56, 109–11 |
| M. musculus   | NM_010769              | NM_016762              | NM_010770              | NM_013592              | 55–56, 110  |
| G. gallus     | NM_001030375           | XM_422012              | NM_205072              | XM_425698              | 80, 112    |
| D. rerio      | BC045465               | -                      | NM_001004007(3a)       | NM_213549              | 62         |
| T. nigroviridis| GSTENT000045-8001      | GSTENT000045-8001      | GSTENT000347-51001(3a)| GSTENT000045-8001      | 75         |

All numbers are GenBank Accession numbers, except for T. nigroviridis numbers from the Genoscope database.

**Table 3: Accession numbers of Matrilins.**

http://www.biomedcentral.com/1471-2164/7/83
For the syndecan-2 gene, STK3 and PGCP were conserved as neighbouring genes in human, chicken and mouse. TSPYL5, LAPTM4B, MATN2 and KCNS2 genes were also conserved between human and mouse and PTDSS1, GDF16 and TP53INP were also conserved between human and chicken (Figure 4B). In comparing the locus of the syndecan-2 gene in *T. nigroviridis*, PGCP was also conserved as a neighboring gene (Figure 4B). MATN2 was not present in this region, but additional searches identified that it is indeed encoded on chromosome 8 (discussed further below). In *D. rerio*, the syndecan-2 gene on chromosome 5 has no gene neighbors in common with those of tetrapods or *T. nigroviridis*, and we infer that this gene underwent a lineage-specific transposition (data not shown).

**PUM1** and **MATN1** were conserved gene neighbors of the syndecan-3 gene in human, mouse, chicken and *T. nigroviridis*, and **MATN1** also mapped adjacent to *D. rerio* **SDC3** (Figure 4C). The most extensive local synteny was between human and mouse: **TDE2l**, **FABP3**, **LAPTM5** and **PTPRU** were all conserved as local gene neighbors (Figure 4C). Of these genes, **FABP3** was also identified to be syntenic with **SDC3** in *T. nigroviridis*. The gene encoding the tyrosine kinase Lck was also local to **SDC3** in human and mouse. In the two fish genomes, genes encoding tyrosine kinases were also adjacent to **SDC3**, however the gene products were more similar to other members of the tyrosine kinase family (Figure 4C).

For the syndecan-4 gene, strong conservation of the genomic regions was apparent between human, mouse and chicken: **KCNS1**, **MATN4** and **RBPSUHL** were conserved between all three species and additional local genes were conserved between human and mouse (Figure 4D). **SDC4** is not yet mapped to a chromosome in either the *T. nigroviridis* or *D. rerio* genomes.

The high conservation of the genomic context of vertebrate syndecan genes prompted us to examine whether this synteny extends to the urochordate, *Ciona intestinalis*. The urochordate group are basal in the chordate lineage and have not undergone whole genome duplications [53]. The single syndecan gene of *C. intestinalis* is encoded in a single matrilin (Gene Cluster 14563), however, this is located on scaffold 259 on chromosome 1. (This gene product has highest identity to matrilin-3. The Gene Cluster 00035/06628 that is assigned as "matrilin-2" at the Ghost website has highest identity to fibrillin). Of the other gene families consistently represented on the same chromosomes as tetrapod syndecan genes, only *Gdf16* was identified on *C. intestinalis* chromosome 2 (scaffolds 213 and 58). Ten local gene neighbours of the syndecan gene of *C. intestinalis* are not syntenic with any human or chicken syndecan gene (data not shown). Thus, there are many differences in the genomic context of *C. intestinalis* syndecan.

### Paralogous locations of syndecan and matrilin genes in the human genome

Overall, the data on the genomic contexts of vertebrate syndecan genes demonstrated two striking points. First, in considering the individual family members, there is clear conservation of synteny between fish and tetrapods for each of the syndecan-2, -3 and -4 genes. This indicates that these loci must each have been present in the last common ancestor of fish and tetrapods. Secondly, a deeper level of homology is apparent, in that the syndecan-1, -2, -3 and -4 genes in the organisms we examined all have conserved gene neighbours that are paralogous members of the same gene families. The clearest example of this are the matrilins, that are components of extracellular matrix and which in tetrapods also comprise a gene family of four members [55,56]. With the exception of chicken syndecan-2, a matrilin gene is located near to every examined tetrapod syndecan gene. The same trend was apparent for the fish syndecan-3 genes (Figure 4). Members of the **LAPTM4** and **KCNS** gene families were also present in the local genomic region of most of the tetrapod syndecan genes. The two members of the pumilio gene family were conserved adjacent to the syndecan-1 and -3 genes (Figure 4).

These findings strongly suggest that, in each genome, the four syndecan genes are located in paralogous genomic regions that have been conserved throughout the evolution of vertebrates. The existence of such regions in vertebrate genomes provides evidence of whole genome duplication events that took place early in chordate evolution [57-60]. In the lineage of each of the organisms examined, subsequent gene loss or localized rearrangements have blurred the initial four-fold replication of the region in distinct ways over time [61]. In the human genome, the rate of DNA rearrangement is slower than in rodents [61] and paralogous regions have been identified globally by BLASTP-based searches of the genome against itself [57,58]. To substantiate the view obtained from our analysis of local genes on an individual basis, we examined whether the human genome contains evidence of chromosomal paralogies in the regions of the four syndecan genes according to this unbiased independent method. Searches were made through the database of paralogs in the human genome, v5.28 [57]. Strikingly, the genomic region of each human syndecan gene was found to be related to the genomic region of all the other family members. Members of the matrilin family were included in all blocks, and other gene families identified in our local searches were represented in individual paired blocks along with genes that are more distant on each
chromosome (Figure 5). Together, these findings provide strong evidence that the four vertebrate syndecan genes have evolved as a consequence of rounds of duplication of a single ancestral chromosomal region.

The absence of syndecan-1 in fish: analysis of the genomic contexts of matrilin genes in T. nigroviridis and D. rerio

Because synteny of syndecan-2 and -3 genes is conserved between tetrapods and pufferfish, and because all four syndecan gene loci show evidence of paralogy in tetrapods, we inferred that the absence of syndecan-1 from fish is a derived characteristic. This raised the question of whether other parts of the conserved chromosomal region are present in fish, even though the syndecan-1 gene itself is missing. We addressed this question through the proximity of syndecan and matrilin genes. In all tetrapods examined, MATN3 is close to SDC1 (Figure 4). All four forms of matrilin are present in T. nigroviridis (Table 3) and both T. nigroviridis and D. rerio matrils include two matrilin-3 paralogues (Table 3) [62]. The existence of such paralogous pairs is consistent with the strong evidence for an additional whole genome duplication early in the evolution of teleost fish [63]. We examined the genomic contexts of matrilin genes in T. nigroviridis and D. rerio. In T. nigroviridis, all four matrilins, including MATN3, are encoded within genomic regions syntenic to those of tetrapod matrilins (Figure 6A). MATN2 is encoded at a distinct point on chromosome 8 from SDC2, yet the adjacent genes include LAPTM4B and KCNS2, that are conserved neighbours in tetrapods. Similarly, KCNS2 is adjacent to MATN4 (Figure 6A). We hypothesized that fish matrilin-3s should be encoded within a genomic region similar to that of tetrapod matrilin-3 and syndecan-1. Importantly, the order of genes adjacent to T. nigroviridis MATN3 matched closely with those adjacent to tetrapod MATN3: however, in T. nigroviridis, PLUM2 and LAPTM4A are adjacent to each other, suggesting a specific loss of SDC1. Although WDR35 is not present on chromosome 14, we found that this gene is located on chromosome 10 adjacent to the T. nigroviridis parologue of MATN3a (Figure 6A).

The genomic contexts of D. rerio matrilin genes, (some of which were previously mapped by the radiation hybrid method [62] and all of which are now physically mapped in Zebrafish genome assembly Zv5), demonstrated the same points, although the conservation of local neighboring genes was not as extensive for MATN1 as in T. nigroviridis (Figure 6B). In the case of MATN4, the local neighboring genes did not match those of tetrapod MATN4, but RBPSUHL is conserved as a neighboring gene between D. rerio and T. nigroviridis (Figure 6B). As in T. nigroviridis, MATN3a and MATN3b both have gene neighbors that match those of tetrapod MATN3. However, the neighbors of each MATN3 parologue represent the opposite sides of the conserved genomic region of tetrapods, indicating a break of the chromosome after the additional genome duplication [62]. We did not identify PLUM2 in the D. rerio genome (assembly Zv5). We examined the genomic regions of fish matrilin-3s, as shown in Figure 6, more closely for evidence of syndecan-1 coding sequence. BLAT search of the Tetraodon genome with the mRNA nucleotide sequences of human, chicken, or X. laevis syndecan-1 did not yield significant findings: in each case, only short, 10–20 nucleotide regions on many different chromosomes were identified as homologous. Syndecan-1 protein sequence searches identified only the cytoplasmic domain of syndecan-2 on chromosome 21. Pairwise BLAST searches of syndecan-1 coding sequences against the Tetraodon or Danio genomic sequence regions did not identify significant matches. Thus, the syndecan-1 coding sequence is not represented in this genome. In conclusion, our findings provide further evidence for partial synteny of the genomic contexts of syndecan and matrilin genes between fish and tetrapods. Most importantly, the fish genomes provide evidence of complex genomic rearrangements in the region of the matrilin-3 gene that have involved the loss of the syndecan-1 encoding sequence.

Figure 2
Sequence relationships of invertebrate and vertebrate syndecans. The amino acid sequences of all the full-length syndecans in the dataset were aligned in TCOFFEE. The Phylib distance matrix output was used to prepared an unrooted phylogenetic diagram in DRAWTREE. Key is the same as in Figure 1. Scale bar = 0.1 substitution/site.
Discussion
We report here the results of molecular phylogenetic and comparative genomic analyses of members of the syndecan family of proteoglycans. The presence of syndecans in organisms from Cnidaria to mammals establishes firmly that the syndecan family is ancient in the animal lineage. The Cnidaria, comprising corals, jellyfish, sea anemones and hydroids, are typified by diploblasty, ability for sexual and asexual reproduction and relatively simple body plans. The divergence of the Cnidaria and Bilateria is thought to have taken place between 650–1000 million years ago (MYA) [43]. Nevertheless, genome and EST project have revealed a high level of conservation of Cnidarian genes and their coding sequences with those of vertebrates, including integrins and components of the extracellular matrix [42,64]. The identification of syndecan-encoding sequences in multiple species of Cnidaria provides a strong indication that syndecans, like integrins, have participated as mediators of extracellular interactions throughout animal evolution. The data also establish firmly that the encoding of multiple syndecans in a single genome is a vertebrate-specific attribute: the presence of syndecans in organisms from Cnidaria to mammals establishes firmly that the syndecan family is ancient in the animal lineage.

The molecular phylogenetic studies demonstrated that syndecan extracellular domains have diverged rapidly in sequence – e.g., each of the invertebrate syndecans has an extracellular domain that is distinct from other species and from all the vertebrate syndecans – whereas the cytoplasmic domains contain highly conserved elements. Identification of syndecans across phyla thus relies heavily on recognition of the cytoplasmic domains in database searches. Within all cytoplasmic domains, the C1 and C2 regions are extremely well-conserved. With regard to the distinctive V regions, our study identified two residues that are extremely highly conserved: the first leucine residue, (corresponding to L289 of mouse syndecan-1) that is universally conserved and the tyrosine residue (corresponding to Y300 of mouse syndecan-1) that is conserved throughout the Bilateria but not in Cnidaria (Figure 1). Current knowledge of vertebrate syndecans suggests that these residues could serve either structural or functional roles. The syndecan-4 cytoplasmic domain forms a homodimer in which L186 is important for the stability of the structure because it makes three contacts with the C1 and V regions of the partner cytoplasmic domain [65]. Although other syndecan cytoplasmic domains are not known to form dimers, it is conceivable that the corresponding leucines could make similar contacts with heterologous binding proteins. Y192 in syndecan-4 also makes three contacts within the homodimer [65]. Thus, in syndecan-4 the most critical roles of L186 and Y192 are structural. In contrast, Y300 in syndecan-1 has been implicated in several signaling roles: it appears necessary for the alignment of syndecan-1 with actin stress fibres [66] and a phosphomimetic Y300E mutant inhibits the assembly of lamellipodia and actin-and-fascin bundles by activated syndecan-1 [13]. It is possible that the presence of this additional tyrosine conferred additional signaling properties on syndecans in the Bilateria. The perspective provided by the new syndecan protein sequence dataset will help guide further experimental analysis of syndecan cytoplasmic domains.

To understand the relationships between the four tetrapod syndecans and gain more insight into the origins of fish syndecans, we turned to a comparative genomic analysis of the vertebrate syndecans. The conservation of synteny is a powerful way to resolve the relationships between paralogous and orthologous members of multigene families in different species [67]. Previous phylogenetic coding sequence analyses have provided indications of paralogy within the family, but the data were not conclusive [68,69]. The loci of the syndecan-1, -2, -3 and -4 genes each showed striking synteny in human, mouse and chicken. With reference to the syndecan genes alone, only partial synteny was detected in fish: whereas in T. nigroviridis and D. rerio the context of SDC3 was clearly syntenic, synteny of SDC2 was partial in T. nigroviridis and absent in D. rerio. Synteny in fish could not be addressed for the absent syndecan-1 and the unmapped syndecan-4. As discussed below, we achieved resolution of this question by the combined analysis of syndecan and matrilin genes in fish.

The identification of conserved neighboring genes in multiple genomes also revealed a deeper level of homology between all four loci within a single genome, in that most of the conserved neighboring genes also corresponded to paralogous members of gene families. This indication that the four syndecan genes are located in paralogous chromosomal regions was confirmed on the basis of an independent, computationally-based method of analysis of the human genome [57]. When the loci of the four mouse syndecan genes were identified by inter-species backcross mapping, it was noted that the syndecan-1, -2 and -3 genes were syntenic with three members of the myc gene family [70]. There is now strong evidence that paralogous regions exist in vertebrate genomes as a result of whole genome duplications that took place in a vertebrate ancestor after the divergence of the amphioxus lineage [53,57-59]. It appears that the genomic regions of tetrapod syndecan genes have been very well-conserved subsequent to these duplication events. At protein sequence level, there is evidence of pairing within the syndecan family, such that syndecan-1 and -3 are more closely related, as are syn-
decan-2 and -4 [3,4]. These pairings are also evident at genomic level: the members of a two gene family, PUM1 and PUM2, are encoded adjacent SDC3 and SDC1, respectively, and STK3 and STK4 are encoded adjacent SDC2 and SDC4 (Figure 4 and Figure 5). Furthermore, the matrilin family also contains pairs: matrilin-1 and -3 are more closely related, as are matrilin-2 and -4 [55,56]. Collectively, these findings are incorporated into a model for the expansion of the syndecan family from a single syndecan in an ancestral chordate to the current multigene status in modern tetrapods and fish (Figure 7).

A surprising and important outcome of the molecular phylogenetic and genomic analyses was the absence of syndecan-1 from the genomes of multiple species of fish (D. rerio, T. nigroviridis and T. rubripes). No syndecan-1-like sequences were identified in the three fish genomes, using either amphibian, chicken, or human syndecan-1 as the query sequences. Searches of dbEST, which includes ESTs from additional bony fish species, also did not provide any evidence for a fish syndecan-1. However, all four matrilins are represented in fish (Fig. 6) [62]. Through analysis of the genomic contexts of fish matrilin genes, we were able to strengthen the evidence for conservation of the respective genomic regions between fish and tetrapods. The fish-specific genome duplication (FSGD) that is estimated to have taken place around 320 MYA [63,71] is assumed to have given rise to the paralogous MATN3a and MATN3b genes. We examined in detail the genomic region around the MATN3 genes, as the expected locus of the syndecan-1 gene. The comparison of the loci of MATN3 genes in T. nigroviridis and D. rerio revealed that the genomic region appears to have become split after the genome duplication. An alternative possibility that cannot be excluded on current data is that different sets of genes have been retained alongside MATN3a and MATN3b in each chromosome after the FSGD. Other gene losses or rearrangements (e.g. loss of PUM2 from the D. rerio genome, relocation of WDR35 to a different chromosome in T. nigroviridis; Figure 6) appear lineage-specific and therefore, we infer, were of more recent occurrence. From the fossil record, the zebrafish and pufferfish lineages are estimated to have diverged around 284–296 MYA [71]. Thus the loss of SDC1 appears, in evolutionary terms, to have occurred relatively soon after FSGD. It is possible that SDC1 was lost before, or at the time of, FSGD, however, the available genomes only sample a portion of the fish lineage. Information on the genomes of more basally diverging species of bony fish and a cartilaginous fish would be required to distinguish between these possibilities. Our working model for the current status of fish syndecan genes incorporates the notion of extensive gene loss after FSGD (Figure 7).

Our findings have several practical implications for future studies of syndecan function. While the zebrafish is an excellent vertebrate model organism for most developmental and disease processes [72], it would not be the model of choice for physiological in vivo studies of syndecan-1. There is evidence from human, flies and C. elegans that adjacent genes can show correlated expression and, in some instances, function in the same pathway [73,74]. The possibility of co-expression or functional association between syndecans and matrilins has not been considered, yet it is intriguing that both are ECM-associated proteins with roles in cell interactions during development and in disease.
urochordates (Ciona) and vertebrates. The syndecan-1 gene appears to have been lost relatively early in the fish lineage. These conclusions provide the basis for a new model of syndecan evolution in vertebrates and a new perspective for experimental analysis of the roles of syndecans in cells and whole organisms.

Methods
Syndecan dataset and molecular phylogeny
A. accession numbers
The accession numbers of syndecan and matrilin family members from different organisms, except T. nigroviridis and C. intestinalis, were obtained by BLASTP and TBLASTN searches of GenBank and dbEST at NCBI. The accession numbers of T. nigroviridis syndecans were obtained by BLAST searches of the T. nigroviridis genome assembly [75]. C. intestinalis syndecan was identified by BLAST searches of the genome and the Gene Clusters...
Ancestral chordate 2, with other assemblies were searched: genome databases, and dbEST. The following genome TBLASTN alignment algorithms at GenBank, their search for syndecans in other species using BLASTP or TBLASTN searches of dbEST: the species included in Table sub-phyla for our dataset.

1 were selected as representatives of additional phyla and TBLASTN searches of dbEST: the species included in Table Syndecans from additional species were identified by [54,76]. The complete datasets are listed in Table 1 and 3. The genomic contexts of matrilin genes in [77,78], [80], [79], [83] and Yu, J., Holland, L.Z., Shin-i, T., Kohara, Y., Satou, Y. and Satoh, (Euprymna scolopes), chelicerate (Rhipicephalus appendiculatus) [85], and crustacean (Marsupenaeus japonicus) syndecans were identified in dbEST and the complete sequence of amphioxus (Branchiostoma floridae) syndecan was compiled from over-lapping ESTs in dbEST [[53] and Yu, J., Holland, L.Z., Shin-i, T., Kohara, Y., Satou, Y. and Satoh,

**Figure 6**
**Genomic contexts of matrilin genes in T. nigroviridis and D. rerio.** The genomic contexts of matrilin gene family members from (A), T. nigroviridis and (B), D. rerio, were analysed for conservation of neighboring genes with human, mouse and chicken, as shown in Fig. 3, and with each other. Partial synteny was identified. Each diagram represents the order of genes along a portion of the indicated chromosomes. Each horizontal line represents a gene: red lines represent genes that are syntenic and black lines represent non-conserved intervening genes. The numbers indicate the exact location in nucleotides of each region on its respective chromosome in each species. HUGO gene names are given, and the matrilin genes are indicated in red.

[54,76]. The complete datasets are listed in Table 1 and 3. Syndecans from additional species were identified by TBLASTN searches of dbEST: the species included in Table 1 were selected as representatives of additional phyla and sub-phyla for our dataset.

**B. BLAST searches**
The full-length amino acid sequences of syndecan family members from either mouse or human were used to search for syndecans in other species using BLASTP or TBLASTN alignment algorithms at GenBank, their genome databases, and dbEST. The following genome assemblies were searched: *H. sapiens* (NCBI build 35.1) [77,78], *M. musculus* (NCBI build 34 of May 2005) [79], *G. gallus* (NCBI build 1.1, March 2004) [80], *T. rubripes* [81] and *D. melanogaster* (build 4.1, February 2005) [82] at NCBI Blast 2.2.12. *D. rerio* (Zv5 assembly of August 2005) was searched at EBI. The *C. elegans* genome [83] was searched at Sanger worm informatics website (WS149; assembly of Sept, 2005). The *T. nigroviridis* genome [75] (assembly of April, 2004), *X. tropicalis* genome assembly v4.1 [50] and *C. intestinalis* genome and EST database [54,76] were also accessed. Sequences including the cytoplasmic domains of representatives of planarian (*Schmidtea mediterranea*) [84], molluscan

**Figure 7**
**Model for the evolution of syndecans in fish and tetrapods.** The upper panel of the model shows the hypothesized ancestral genomic context of a single syndecan gene, originating in an ancestral chordate (chordate 2) subsequent to the divergence of the Urochordate lineage. The model assumes that four-fold paralogy was then set up in an ancestral vertebrate as a result of two rounds of whole-genome duplication. This process also sets up pairing within each set of paralogues. The lower right hand panel represents how the initial complete paralogy has degenerated through gene rearrangements in modern tetrapods. Fish underwent an additional round of genome duplication that would have generated additional paralogous pairs [58]. The lower left-hand panel represents the situation in two modern fish, in which only the two matrilin-3 paralogues have been retained from FSGD. The syndecan-1 locus appears to have been lost early in the fish lineage. Other gene rearrangements or losses appear specific to the zebrafish or pufferfish lineages.
N. Expressed genes in Branchiostoma floridae project. Cnidarian syndecan sequences, from Acropora tenuis [86], Acropora palmata (Schwarz, J.A., Brokstein, P., Manohar, C., Coffroth, M.A., Szmant, A. and Medina, M., Coral-Sym
correlation in the physically-mapped genomes of human [93],

The chromosomal locations of syndecan genes were iden
tification of synteny between syndecan genes

The newly-identified syndecans from representative invertebrates, urochordate, amphibian and fish were compared with the well-characterized human syndecans using WU-BLAST 2.0 against the Uniprot dataset [90,91], using a coding sequence cutoff value of e = 10^-7. Because of the low sequence identity of syndecan extracellular domains, syndecan-1 and -3 were not reproducibly recognized by these criteria. All syndecan genes are included in all blocks in Figure 5.

C. multiple alignment
Predicted amino acid sequences of the cytoplasmic domains of syndecans from organisms representing different phyla were aligned in CLUSTALW [46] and are presented in Boxshade 3.2. Predicted amino acid sequences of full-length syndecans were aligned in TCoffee (version_2.11) with default parameters using pairwise methods [48].

D. tree building
Several parallel methods were used to assess molecular phylogenetic relationship of syndecans, using either the full-length sequences or the cytoplasmic domains. An unrooted phylogenetic tree was constructed from the Phylip distance matrix output of a TCoffee alignment [48] in DRAWTREE, and is presented in Phylodendron (D.G. Gilbert, version 0.8d) with the choice of intermediate node positions and node lengths for tree growth [89]. Aligned amino acid sequences of the full-length and cytoplasmic domains of syndecans were analyzed using PHYML, a maximum likelihood reconstruction method, with the WAG substitution model [51]. Bootstrap proportion was used to assess the strength of the topologies for this method [52]. The output data was used to obtain a consensus tree based upon majority extended rule. The consensus outtree file (in newick format) was used to prepare an unrooted tree in Phylopendron.

E. WU-BLAST
The newly-identified syndecans from representative invertebrates, urochordate, amphibian and fish were compared with the well-characterized human syndecans using WU-BLAST 2.0 against the Uniprot dataset [90,91]. The conservation scores were calculated by dividing the bit score of each hit by the maximum bit score value obtained from the sequence used to conduct the WU-BLAST search [92].

Identification of synteny between syndecan genes
The chromosomal locations of syndecan genes were identified in the physically-mapped genomes of human [93], mouse [94], chicken [95], T. nigroviridis [75] and D. rerio (EBI assembly Zv5) by TBLASTN searches of each syndecan against the relevant genome. Genes located on either side of the target syndecan gene, (25 on each side, total 50), on the same contig were identified and compared between organisms. In the case of different gene nomenclatures in different organisms, the encoded protein sequences were used in BLASTP or TBLASTN searches against the human genome assembly, to identify the corresponding human gene and its locus. The HUGO gene names are given in the figures.

Assessment of paralogy
To identify whether the syndecan and matrilin genes are present in paralogous regions of the human genome, the "dataset of paralogons in the human genome, v5.28" was searched [57]. These blocks were defined by McLysaught et al., using a coding sequence cutoff value of e = 10^-7. Because of the low sequence identity of syndecan extracellular domains, syndecan-1 and -3 were not reproducibly recognized by these criteria. All syndecan genes are included in all blocks in Figure 5.

Abbreviations
BLAST, Basic Local Alignment Search Tool; dbEST, database of Expressed Sequence Tags; EBI, European Bioinformatics Institute; ECM, Extracellular matrix; EST, Expressed Sequence Tag; FSGD, Fish Specific Genome Duplication; HUGO, Human Genome Organisation; MYA, Million of Years Ago; NCBI, National Center for Biotechnology Information; PDZ protein, PSD-95, Disks-large and ZO1 protein.

Authors’ contributions
RC contributed data acquisition and analysis, prepared 5 figures and 3 tables, and contributed in the writing of the manuscript. JCA contributed guidance of the project and data analysis, prepared 2 figures and contributed in the writing and final organization of the manuscript.

Acknowledgements
We thank Dr. Ken Wolfe for advice and permission to modify human paralogon block diagrams, and Dr. Mario Caccamo, the Wellcome Trust Sanger Institute, for clarification of mapping queries in zebrafish genome assembly Zv5. This research was supported by a grant from the NIH, GM068073, to JCA.

References
1. Fernandez-Busquets X, Burger MM: Cell adhesion and histocompatibility in sponges. Microsc Res Tech 1999, 44:204-218.
2. Bucior I, Burger MM: Carbohydrate-carbohydrate interaction as a major force initiating cell-cell recognition. Glycoconj J 2004, 21:111-123.
3. Bernfield M, Gott M, Park PW, Reizes O, Fitzgerald ML, Lincecum J, Zako M: Functions of cell surface heparan sulfate proteoglycans. Annu Rev Biochem 1999, 68:729-777.
4. Couchman JR: Syndecans: proteoglycan regulators of cell-surface microdomains? Nat Rev Mol Cell Biol 2003, 4:926-937.
5. Steigemann P, Molitor A, Feltl S, Jackle H, Vorbruggen G: Heparan sulfate proteoglycan syndecan promotes axonal and myotube guidance by slit/robo signalling. Curr Biol 2004, 14:225-230.
Fox AN, Zinn K: The heparan sulfate proteoglycan syndecan is an in vivo ligand for the Drosophila LAR receptor tyrosine phosphatase. Proc Natl Acad Sci USA 2005, 102:701-706.

Rawson JM, Dimirtroff B, Johnson KG, Rawson JM, Ge X, Van Vactor D, Selleck SB: The heparan sulfate proteoglycans Dally-like and Syndecan have distinct functions in axon guidance and visual-system assembly in Drosophila. Curr Biol 2005, 15:833-838.

Rhinier C, Gygi S, Frohli E, Hengartner MO, Hajnal A: Syndecan reg- ulates cell migration and axon guidance in C. elegans. Develop ment 2005, 132:4621-4633.

Mimini AN, Labarca M, Hurtado C, Brandon E: Caenorhabditis elegans syndecan (SDN-1) is required for normal egg laying and associates with the nervous system and the vulva. J Cell Sci 2004, 117:15179-15190.

Li Q, Park PW, Wilson CL, Parks WC: Syndecan-1 shedding by Pseudomonas aeruginosa enhances virulence. J Invert Pathol 2004, 83:301-309.

Dhodapkar MV, Abe E, Theus A, Lacy M, Langford JK, Barlogie B, Sanderson RD: Syndecan-1 is a multifunctional regulator of myeloma pathobiology: control of tumor cell survival, growth, and bone cell differentiation. Blood 1998, 91:2679-2688.

van den Brule A, Antoniou E, Eriksson M, Makartova R, Altham H, Kinnula V, Leppa S: Soluble syndecan-1 and serum basic fibroblast growth factor are new prognostic factors in lung cancer. Cancer Res 2002, 62:5210-5217.

Yang Y, Taccoby S, Liu W, Langford JK, Pumphrey CY, Theus A, Epstein J, Sanderson RD: Soluble syndecan-1 promotes growth of myeloma tumors in vivo. Blood 2002, 100:610-617.

Leivonen M, Lundin J, Nordling S, von Boguslawski K, Haglund C: Prognostic value of syndecan-1 expression in breast cancer. Oncology 2004, 67:1-18.

Mueeda T, Alexander C, Friedl A: Induction of syndecan-1 expression in stromal fibroblasts promotes proliferation of human breast cancer cells. Cancer Res 2004, 64:612-621.

Juuti A, Nordling S, Lundin J, Louhimo J, Haglund C: Syndecan-1 expression—a novel prognostic marker in pancreatic cancer. Oncology 2002, 63:22-27.

Lovell R, Dunn JA, Begum G, Barth NJ, Plant T, Moss PA, Drayton MT, Pratt G: Soluble syndecan-1 level at diagnosis is an independent prognostic factor in multiple myeloma and the extent of fall from diagnosis to plateau predicts for overall survival. Br J Haematol 2003, 129:249-255.

Wiliapoulos TP, Kyrtonis MC, Papadogiannis A, Nadali G, Angelopoulou MK, Tzenou T, Dimopoulou MN, Sakantaraki MP, Kon dopidou FN, Kalpakidou C, Kokoris SI, Dimitrioudis EM, Tsatsaris P, Pizziolo G, Pangalis GA: Serum levels of soluble syndecan-1 in multiple myeloma pts falling from diagnosis to plateau predicts for overall survival. Br J Haematol 2003, 129:249-255.

Gotte M: Syndecans in inflammation. FASEB J 2003, 17:575-591.

Nikkari ST, Jarvelainen HT, Wight TN, Ferguson M, Clowes AW: Smooth muscle cell expression of extracellular matrix genes after arterial injury. Am J Pathol 1994, 144:1348-1356.

Cianfoni-Smith G, Lang E, Showalter L, Carey DJ: Syndecan-4 is a primary response gene induced by basic fibroblast growth factor and arterial injury in vascular smooth muscle cells. Arterioscler Thromb Vasc Biol 1997, 17:172-180.

Wang H, Moore S, Alavi MZ: Expression of syndecan-1 in rabbit neointima following de-endothelialization by a balloon cath- eter. Atherosclerosis 1997, 131:141-147.

Filla MS, Dam P, Raapraeger AC: The cell surface proteoglycan syndecan-2 mediates fibroblast growth factor-2 binding and signaling. J Cell Physiol 1998, 174:310-321.

McFall AJ, Raapraeger AC: Characterization of the high affinity cell-binding domain in the cell surface proteoglycan synde- can-4. J Biol Chem 1998, 273:28270-28276.

Zimmermann P, Zhang Z, Degeest G, Mortier E, Leenaerts I, Nahorski SR: Syndecan-1 regulates chemokine mobilization and transepithelial cell migration. J Cell Sci 2001, 114:1981-1990.

Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994, 22:4673-4680.

Harris BZ, Lim WA: Mechanism and role of PDZ domains in signaling complex assembly. J Cell Sci 2001, 114:3219-3231.

Notredame C, Higgins D, Heringa J: T-Coffee: A new method for multiple sequence alignments. J Mol Biol 2000, 302:205-217.

Teel AL, Yost HJ; Embryonic expression patterns of Xenopus syndecans. Mech Dev 1996, 59:115-127.

Klein SL, Strasburg RL, Wagner L, Pontius J, Cliffton SW, Richardson P: Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative. Dev Dyn 2002, 225:384-391.
93. McPherson JD, Marra L, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK: A physical map of the human genome. Nature 2001, 409:934-941.

94. Gregory SG, Sekhon M, Schein J, Zhao S, Suyemitsu T, Satou Y, Chiba S, Satoh N: Characterization of human syndecan-3 (MATN3). Genomics 1998, 53:391-394.

95. Gould SE, Upholt WB, Kosher RA: Baciu PC, Acaster C, Goetinck PF: Molecular cloning and characterization of chicken syndecan-2 proteoglycan. J Biol Chem 1992, 265:481-490.

96. Mali M, Jaakkola P, Arvilommi AM, Jalkanen M: Wallis JW, Aerts J, Groenen MA, Crooijmans RP, Layman D, Graves TA, Scheer DE, Kremizki C, Fedele MJ, Mudd NK, Cardenas M, Higginbotham J, Carter J, McGrane R, Gaige T, Mead K, Walker J, Albracht D, Davito J, Yang SP, Leong S, Chinwalla A, Sekhon M, Wylie K, De Jong P, Osogawa K, Neftov M, Zhang H, McPherson JD, Krzywinski M, Schein J, Hillier L, Mardis ER, Wilson RK, Warren WC: A physical map of the chicken genome. Nature 2004, 432:761-764.

97. David G, van der Schueren B, Marynen P, Cassiman JJ, van den Bergh H: Molecular cloning of amphiglycan, a novel integral membrane heparan sulfate proteoglycan expressed by epithelial and fibroblastic cells. J Cell Biol 1992, 118:961-969.

98. Kojima T, Inazawa J, Takamatsu J, Rosenberg RD, Saito H: Human ryudocan core protein: molecular cloning and characterization of the cDNA, and chromosomal localization of the gene. Biochem Biophys Res Commun 1993, 190:814-822.

99. Berndt C, Casaroli-Marano RP, Vilaro S, Reina M: Cloning and characterization of human syndecan-3. J Cell Biochem 2001, 82:246-259.

100. Saunders S, Jalkanen M, O'Farrell S, Bernfield M: Molecular cloning of syndecan, an integral membrane proteoglycan. J Cell Biol 1989, 108:1547-1556.

101. Tsuzuki S, Kojima T, Katsumi A, Yamazaki T, Sugiura I, Saito H: Molecular cloning, genomic organization, promoter activity, and tissue-specific expression of the mouse ryudocan gene. J Biochem 1997, 122:17-24.

102. Chen L, Coughman JR, Smith J, Woods A: Molecular characterization of chicken syndecan-2 proteoglycan. Biochem J 2002, 366:481-490.

103. Gould SE, Upholt WB, Kosher RA: Characterization of chicken syndecan-3 as a heparan sulfate proteoglycan and its expression during embryogenesis. Dev Biol 1995, 168:438-451.

104. Baciu PC, Acaster C, Goetinck PF: Molecular cloning and genomic organization of chicken syndecan-4. J Biol Chem 1994, 269:696-703.

105. Rosenblum ND, Botelho BB, Bernfield M: Expression of a Xenopus counterpart of mammalian syndecan 2 during embryogenesis. Biochem J 1995, 309:69-76.

106. Satou Y, Chiba S, Satoh N: Expression cloning of an ascdian syndecan suggests its role in embryonic cell adhesion and morphogenesis. Dev Biol 1999, 211:198-207.

107. Tomita Y, Yamakawa K, Suyemitsu T: Cloning and characterization of cDNA for syndecan core protein in sea urchin embryos. Dev Growth Differ 2000, 42:449-458.

108. Spring J, Paine-Saunders SE, Hynes RO, Bernfield M: Drosophila syndecan: conservation of a cell-surface heparan sulfate proteoglycan. Proc Natl Acad Sci USA 1994, 91:3334-3338.

109. Muratoglu S, Krysan K, Balazs M, Sheng H, Zakany R, Modis L, Kiss I, Deak F: Primary structure of human matrilin-2, chromosome location of the MATN2 gene and conservation of an AT-AC intron in matrilin genes. Cytogeten Cell Genet 2000, 90:323-327.

110. Mates L, Korpos E, Deak F, Liu Z, Beier DR, Aszodi A, Kiss I: Comparative analysis of the mouse and human genes (Matn2 and MATN2) for matrilin-2, a filament-forming protein widely distributed in extracellular matrices. Matrix Biol 2002, 21:163-174.

111. Belluoccio D, Schenker T, Baici A, Trueb B: Characterization of human matrilin-3 (MATN3). Genomics 1998, 53:391-394.

112. Belluoccio D, Trueb B: Matrilin-3 from chicken cartilage. FEBS Lett 1997, 415:212-216.