

SUPPLEMENTARY DATA

Supplementary file 1. Accession numbers and references for the adhesome components discussed here.

Table S1. Representation of ECM components and major matrix proteases in basal metazoa and protostomes.

Table S2. Representation of cell-matrix adhesion receptors in basal metazoa and protostomes.

Table S3. Major ECM innovations within deuterostomes.

Supplementary File 1. Accession numbers for Extracellular Matrix Adhesome Components discussed in this article.

This dataset includes the accession numbers of components identified positively by BLAST searches based on mammalian and/or known protostomal protein sequences of the selected adhesome components and an e-score cutoff of 1e-20. Components on this list were verified by reverse BLAST search. Components that were identified throughout the basal metazoa and protostomes (laminin, etc) are not included on the deuterostome list. Databases searched included the Peptide, Nucleotide and expressed sequence tag divisions of GenBank, all at NCBI; Ensembl Proteins; JGI Eukaryotic Genomes; Smed DB; the elephant shark genome at <http://esharkgenome.imcb.a-star.edu.sg/>; the lamprey preliminary genome assembly at ENSEMBL; the Broad Institute Origins of Multicellularity Initiative www.broadinstitute.org/annotation/genome/multicellularity_project, and the Hydra genome at <http://hydrazome.metazome.net/>. Referenced publications are cited at the end of the list.

NON-METAZOA

AMOEBOZOA

Dictyostelium discoideum (Eichinger et al., 2005)

Similar to integrin beta proteins A-E; the main similarities are an extracellular vWF_A domain and talin-binding by the cytoplasmic domain), DDB0187447, DDB0187821, DDB0187822, DDB0219318, DDB0187788 (Cornillon et al., 2006, 2008)
CD36-like, XP_647472.1 (has c. 300aa insert in central region)

APUSOZOA

Thecamonas trahens (formerly *Amastigomonas* sp.)

Integrin alpha (Sebé-Pedrós et al., 2010)

Integrin beta (Sebé-Pedrós et al., 2010)

OPISTHOKONTS

Independent lineage within Opisthokonts

Capsaspora owczarzaki

Integrin alpha Co1 CAOG_01284.1, alpha Co2 CAOG_02006.1, alpha Co3 CAOG_05059.1, alpha Co4 CAOG_04087.1 (Sebé-Pedrós et al., 2010) (Broad Institute Origins of Multicellularity)
Integrin beta Co1-3, ADI46542.1, ADI46543.1, ADI46544.1 respectively (Sebé-Pedrós et al., 2010) (Broad Institute Origins of Multicellularity)
CD36-like CAOG_04346 (Broad Institute Origins of Multicellularity)

Choanoflagellates

Monosiga brevicollis (unicellular) (King et al., 2008)

Collagen IV-like, XP_001748906 (has interspersed vWF_A domains but no non-collagenous domains)

Collagen I-like, XP_001749460 (no non-collagenous domains)

Three proteins with C-propeptide like domains, XP_001744748, XP_001747370, XP_001744747 (all lack the cysteines needed for disulphide bonds between the chains, Exposito et al., 2008)

Fibrillin-like, XP_001748060 (lacks TB domains)

CD36-like, XP_001747074. (this protein model is much longer than other CD36 sequences and also contains a "HIT family domain". In view of the homology of the rest of the model to CD36 (BLASTP e-score $1e-32$), the inclusion of this domain may represent a modeling artifact)

Usherin-like, XP_001747261 (e-score of $9e-54$ vs. *H. sapiens* usherin isoform A, but contains no FN-III domains)

***Salpingoeca rosetta* (unicellular and colonial)**

Fibrillar collagen-like, PTSG_01978 (no non-collagenous domains) (Broad Institute Origins of Multicellularity)

Collagen IV-like, PTSG_00277 (has interspersed vWF_A domains but no non-collagenous domains) (Broad Institute Origins of Multicellularity)

Proteins with C-propeptide like domains, PTSG_00533.1, PTSG_00534.1 (Broad Institute Origins of Multicellularity)

Fibrillin-like, PTSG_00996.1 (Broad Institute Origins of Multicellularity)

Usherin PTSG_02226 (Broad Institute Origins of Multicellularity) (e-score of $2e-132$ vs. human usherin isoform B, FN type III domains are present)

Ministeria vibrans

Integrin beta-like, AM904781* (partial sequence; Shalchian-Tabrizi et al., 2008)

METAZOA

Porifera

Fibrillar collagen CAA49472, (*Ephydatia muelleri*) (Exposito et al., 1993)

BMP/tolloid-like GW173395* (*Amphimedon queenslandica*)

Collagen IV CAA65082, CAA65083 (*Pseudocortidium jarrei*) (Boute et al., 1996)

Laminin EB741489*, EC377287* (LN α -like), EB741492* (LN β -like), (*Oscarella carmela*)

Perlecan EC371537* (*Oscarella carmela*)

Fibrillin-like AM763182*, EC375110* (*Oscarella carmela*)

Thrombospondin AM764134* (*Oscarella lobularis*)

Agrin EB741461* (*Oscarella carmela*)

SPARC G840P35RA13.TO (*Oscarella carmela*)

Integrin alpha CAA65943.1 (*Geodia cydonium*) (Pancer et al., 1997)

Integrin betaPo1 AAB66911 (*Ophlitaspongia tenuis*) (Brower et al., 1997)

Syndecan AM7632658* (*Oscarella lobularis*)

Glypican GW164369* (*Amphimedon queenslandica*)

CD36 CAD91339 (*Suberites domuncula*) (Muller et al., 2004)

MMP EC37690*, EC376317* (*Oscarella carmela*)

ADAMTS EC372914* (*Oscarella carmela*), GW175932* (*Amphimedon queenslandica*)

Placozoa

Trichoplax adhaerens (Srivastava et al., 2008)

Col IV (a1)-like XP_002116296

Col IV (a2)-like, XP_002116198

Laminin XP_002114273 (LN α -like), XP_002111840 (LN β -like), XP_002109259 (LN γ -like)

Perlecan XP_002113826

Nidogen XP_002113519

Fibrillin XP_002113297

Thrombospondin XP_002107975

Agrin XP_002113830
 Integrin alpha Ta1 XP_002111197, Integrin alpha Ta2 XP_002111224
 Integrin beta Ta1 XP_002110650, Integrin beta Ta2 XP_002111182
 Glypican XP_002117928
 Dystroglycan XP_002112652
 CD36 XP_002112871
 ADAMTS XP_002108291

Cnidaria

Hydra magnipapillata (Chapman et al., 2010)

Fibrillar collagen XP_002161962
 Collagen IV XP_002157001
 Laminin XP_002160645 (LN α), XP_002161535 (LN β)
 Perlecan XP_002168817
 Fibrillin XP_001630852
 Thrombospondin XP_002164610
 SPARC XP_002167014 (lacks acidic domain) (Koehler et al., 2009)
 Integrin alpha XP_002161020.1 (partial)
 Integrin beta XP_002164638.1 (partial)
 Syndecan Hma2.224712 (Hydra genome project)
 Glypican XP_002157574
 Dystroglycan XP_002164217
 DDR XP_002168413
 CD36 XP_002169936
 MMP XP_002163794
 ADAMTS XP_002166940

Nematostella vectensis (starlet sea anemone) (Putnam et al., 2007)

Fibrillar collagen XP_001635016
 Collagen IV XP_001626265, FC261970*
 Laminin XP_001628586 (LN β), XP_001621795, XP_001623203 (LN γ)
 Nidogen XP_001625225
 Perlecan XP_001627394
 Fibrillin XP_001630852
 Thrombospondins Nv22035, Nv168100, Nv85341, Nv30790 (JGI)
 SPARC XP_001629356, XP_01629356, XP_001626442, XP_001641541 (all lack acidic domain, Koehler et al., 2009)
 Integrin alpha XP_001641435 (NvItg α 1), (Knack et al., 2008)
 Integrin beta XP_001641468 (NvItg β 1); XP_001627336 (NvItg β 2); XP_001637894 (NvItg β 3); XP_001621822 (NvItg β 4) (Knack et al., 2008)
 Syndecan FC288353*
 Glypican XP_001624312
 Dystroglycan XP_001629936
 CD36 XP_001626798
 MMP XP_001633230
 ADAMTS XP_00163643

Nematode***C. elegans*** (*C. elegans* Genome Sequencing Consortium, 1998)

Collagen IV NP_001022662, NP_510664

Laminin NP_492775.2 (LN α), NP_500734.2 (LN β) NP_509204.2 (LN γ)

Perlecan NP_497044

Nidogen NP_506228

Fibrillin NP_498670

SPARC NP_500039

Agrin NP_001022152

Integrin alpha P34446 (α Pat2), NP_499032 (α Ina-1)Integrin beta NP_497787 (β Pat3)

Syndecan, NP_741894

Glypican NP_510582

DDR NP_508572

Dystroglycan NP_509826

NG2/CSPG4 NP_491802

CD36 NP_508919

MMP NP_497596, NP_741156, NP_503790

ADAMTS NP_501792, NP_510116, NP_505017, NP_741569, NP_510291,

NP_001024532

Platyhelminth***Schmidtea mediterranea*** (planarian worm) (SmedDB, Robb et al. 2008; searched vs MAKER transcripts)

Fibrillar collagen mk4.008247, mk4010742 (collagen V-like)

Laminin mk4.008300 (LN α -like), mk4.000288 (LN β -like), mk4.000676 (LN γ -like)

Perlecan DN297419*

Fibrillin mk4.001126 (2 gene products in the model). EE669859*, EE673994*

SPARC DN296182*

Agrin mk4.004859

Integrin alpha – not identified, insufficient information

Integrin beta DN315080*, EE672330*

Syndecan DN297550*

CD36 mk4.000557

MMP DN314012*

ADAMTS mk4.000736

Protostomes**Annelid*****Capitella teleta (worm)*** (JGI Eukaryotic genomes)\$

Fibrillar collagen 224336, 18662

Collagen IV 226710, 227631

Laminin 219723, 157479 (LN α -like), 183991 (LN β -like), 172041, 62597 (LN γ -like),

Perlecan 90300

Nidogen 173985

Fibrillin 229950, 219783

Thrombospondin 172540

SPARC 160588

Agrin 71935, 223648
 Integrin alpha 222780, 218861
 Integrin beta 93073, 93285
 Syndecan, EY526330*, EY605833*
 Glypican 32704
 DDR 150374, 155564
 Dystroglycan 183589
 NG2/CSPG4 71887
 CD36 178424, 189109
 MMP 164703
 ADAMTS 141382, 158288

Mollusc

Lottia gigantea (JGI Eukaryotic genomes)\$
 Fibrillar collagen 194209, 194221
 Collagen IV 188217, 154576
 Laminin 227981 (LN α), 209765, 144813 (LN β)
 Perlecan 125410
 Nidogen 208735
 Fibrillin 232356, 218536
 Thrombospondin 91969
 Agrin 71215
 SPARC 109908
 Integrin alpha 239025, 168627, FC738676*
 Integrin beta 174238, 196574
 Syndecan FC773329*, FC736857*
 Glypican 139272
 DDR 116145
 Dystroglycan 224800
 NG2/CSPG4 120635
 CD36 165058, 159641
 MMP 158923
 ADAMTS 157969, 152088

Arthropods

Insect: Drosophila melanogaster (fruit fly) (Adams et al., 2000)
 Collagen IV AAF52204, AAN10519, AAN10520
 Laminin AAF50672 (LN α), AAN10647 (LN β)
 Perlecan NP_001027037
 Nidogen NP_610575
 Fibrillin NP_787974
 Thrombospondin NP_523495
 SPARC NP_651509
 Glypican NP_524071
 Integrin alpha Q24247 (α PS1), P12080 (α PS2) O44386 (α PS3) NP_611025 (α PS4),
 NP_611808 (α PS5)
 Integrin beta P11584 (β PS), Q27591 (β nu)
 Syndecan NP_476965
 Glypican NP_523983
 DDR NP_001014474

Dystroglycan NP_523756, NP_725523
 NG2/CSPG4 NP_609881
 CD36 NP_787957
 MMP NP_726473, NP_995788
 ADAMTS ACY56893, AA084907, NP_001163761, NP_788751, NP_001163760,
 NP_788752

Crustacean: *Daphnia pulex* (JGI Eukaryotic genomes)\$ (Bauer et al., 2007)

Fibrillar collagen 188996, 226717
 Collagen IV 226325
 Laminin 237253, 49658 (LN α), LN β - no hits found
 Perlecan 232498, 186952
 Nidogen 194034
 Fibrillin 240066
 Thrombospondin 213032
 Agrin 254090
 SPARC 210519
 Integrin alpha 60396, 197169
 Integrin beta 211620, 202093
 Syndecan FE368002*
 Glypican 20775
 DDR 65119
 Dystroglycan 228542
 NG2/CSPG4 249084
 CD36 254798
 MMP 214039
 ADAMTS 20190

Deuterostomes

Echinoderm

***Strongylocentrotus purpuratus* (sea urchin)** (Sea Urchin Genome Sequencing consortium et al., 2006)
 FACIT collagen, IX-like XP_00182505

Hemichordate

***Saccoglossus kowalevskii* (acorn worm)** (Baylor College of Medicine Human Genome Sequencing Center, acorn genome project)
 FACIT collagen IX-like, XP_002730406
 CCN-like XP_002731449 (predicted as a transmembrane protein)

Chordates

Cephalochordate

***Branchiostoma floridae* (Amphioxus)** (Putnam et al., 2008)
 Tenascin, genome scaffold 155:1236900-1282536 (UCSC chrUn: 431,946,615-431,985,644) (Tenascin gene with up to 46 FN3 domains; Tucker et al., 2006, and additional analysis)
 No fibronectin identified
 FACIT collagen, collagen IX-like XP_002609608
 CCN XP_002600149

Hyaluronan synthase XP_002586937, XP_02586938, XP_002586939, XP_002589932, XP_002585658
 Matrilin XM_002593611 (mat1-like), XM_002593611 (mat4-like)

Urochordate

***Ciona intestinalis*, *Ciona savignyi* (sea squirts)** (Dehal et al., 2002, Satou et al., 2002, Vinson et al., 2005)

Tenascin, ENSCING00000003482 (*Ciona intestinalis*)

Fibronectin-like of *Ciona intestinalis* CI0100130823 (Huxley-Jones et al., 2007)

Fibronectin-like of *Ciona savignyi* SNAP00000093593 + SNAP00000093601 + SNAP00000046474 (Tucker and Chiquet-Ehrismann, 2009)

FACIT collagen, collagen IX-like NP_001027707 (*Ciona intestinalis*, Vizzini et al., 2002)

CCN XM_002131886, XM_002127121, XM_002130707 (*Ciona intestinalis*)

Matrilin XP_002123463 (mat1-like), XP_002123195 (mat2-like) (*Ciona intestinalis*)

Craniate: Hyperoartia

***Petromyzon marinus* (lamprey)** (ENSEMBL preliminary genome assembly)

TN-C-like, partial sequences from genome, Contig44637:6129-6557, Contig58813:6704-6853

TN-R-like, partial sequences from genome, Contig31899:9943-10107, Contig65177:3656-3817

Fibronectin, insufficient information, no identification at this time

Hyaluronan synthase, partial sequences from genome, Contig30839, Contig 6935

Vitronectin, insufficient information, no identification at this time

FACIT collagen, collagen IX-like, FD725681*

CCN insufficient information from genome, 1 EST FD727930*

Matrilin insufficient information from genome, 1 EST FD714309.1* (matrilin-4)

Craniate: Gnathostome: cartilaginous fish

***Callorhynchus milli* (elephant shark)** (Venkatesh et al. 2007)\$

<http://esharkgenome.imcb.a-star.edu.sg/>

TN-C-like, partial sequences from genome, AAVX01052536.1, AAVX01336473, AAVX01207200

TN-R-like, partial sequences from genome, AAVX01644962, AAVX01459514.1, AAVX01001598.1, AAVX01144994.1, AAVX01251000.1, AAVX01370913.1

Fibronectin, partial sequences from genome, AAVX01073350.1, AAVX01073441.1, AAVX01173383.1

Hyaluronan synthases, partial sequences from genome. AAVX01121459 (HAS2-like), also AAVX01006632, AAVX01117413, HAS3-like AAVX01095215 (HAS3-like)

CCN – insufficient information

Vitronectin AAVX01477766 (e-score 2e-21), AAVX01097045 (e-score 5e-15)

FACIT collagen, e.g. collagen IX-like AAVX01214773

Matrilin AAVX01035295.1, AAVX01027437.1 (matrilin-4like)

Versican AAVX01149243.1

Aggrecan AAVX01409029

Craniate: Gnathostomes: ray-finned fish and tetrapods

Bony fish

Takifugu rubripes (marine puffer fish) (Aparicio et al., 2002)

Tenascin-Ca, ENSTRUT0000008995
 Tenascin-Cb, ENSTRUT00000017437
 Tenascin-R, ENSTRUT00000008977
 Tenascin-X, ENSTRUT00000041256
 Tenascin-W, ENSTRUT00000006069
 Fibronectin 1, ENSTRUT00000003179
 Fibronectin 1b, ENSTRUT00000044055
 CCN ENSTRUT00000034720, ENSTRUT00000025553, ENSTRUT00000047434
 Hyaluronan synthase JGI Protein ID 709251 [scaffold 22:1935751-1937741](#) (HAS1),
 JGI Protein ID 571120 [scaffold 133:635916-639428](#) (HAS2), JGI Protein ID 592658
[scaffold 2739:9368-11891](#) (HAS3)
 CD44 JGI Protein ID 120971 [scaffold 279:248145-254658](#)
 Vitronectin JGI Protein ID 728631 [scaffold 71:917923-920251](#) (vitronectin a), JGI
 Protein ID 594143 [scaffold 9:2460537-2462775](#)(vitronectin b)
 FACIT collagen e.g., collagen IX. JGI Protein ID 139763 [scaffold 24:1632310-1642513](#)
 (alpha-1 IX), JGI Protein ID 149937 [scaffold 257:124147-131000](#) (alpha-3 IX)
 Matrilin JGI Protein ID 576258 [scaffold 45:911050-913637](#) (matrilin-1), JGI Protein ID
 744708 [scaffold 2179:10548-18098](#) (matrilin-2), JGI Protein ID 730629
[scaffold 94:779373-784718](#) (matrilin-3a), JGI Protein ID 724156 [scaffold 34:476474-498040](#)
 (matrilin-3b), JGI Protein ID 744708 [scaffold 2179:10548-18098](#) (matrilin-4)
 Aggrecan JGI Protein ID 579364 [scaffold 1:641058-660360](#) (aggrecan a), JGI Protein
 ID 731418 [scaffold 104:327667-334279](#) (aggrecan b)
 Versican Protein ID (JGI) 723209 [scaffold 27:937450-958030](#) (versican b)

Tetraodon nigriviridis (freshwater green-spotted pufferfish) (Jaillon et al., 2004)

Tenascin-Ca, CAG01316
 Tenascin-Cb, CAG05242
 Tenascin-R, CAG07653
 Tenascin-X, GSTENT00034161001
 Tenascin-W, CAG07652
 Fibronectin 1, GSTENT10013160001, GSTENT10013159001, GSTENT10013158001
 GSTENT10017308001, GSTENT10017309001, GSTENT10017310001
 Fibronectin 1b, GSTENT10017311001
 CCN CAG01808, ENSTNIT00000021487, ENSTNIT00000012697
 Hyaluronan synthase GSTENT10019574001 (HAS1), GSTENT10019361001 (HAS2)
 GSTENT10029049001 (HAS3)
 CD44 GSTENT10005623001
 Vitronectin ENSTNIT00000019453 (vitronectin a), ENSTNIT00000008425 (vitronectin b)
 FACIT collagen e.g., collagen IX. ENSTNIT00000019915 (alpha-1 IX),
 ENSTNIT00000015781 (alpha-3 IX)
 Matrilin GSTENT10013768001 (matrilin-1), GSTENT10004112001 (matrilin-2 partial),
 GSTENT10004111001 (matrilin-2 partial), GSTENT10006657001 (matrilin-3a),
 ENSTNIT00000000599 (matrilin-3b), ENSTNIT00000008057 (matrilin-4)
 Aggrecan GSTENT10017592001 (aggrecan a, partial), GSTENT10017593001
 (aggrecan a, partial), GSTENT10017594001 (aggrecan a, partial),
 GSTENT10006265001 (aggrecan b, partial), GSTENT10006264001 (aggrecan b,
 partial)

Versican GSTENT10016256001 (versican a), GSTENT10004225001 (versican b, partial), GSTENT10004231001 (versican b, partial)

Danio rerio (zebrafish) (The Zebrafish genome sequencing project group at the Wellcome Trust Sanger Institute www.sanger.ac.uk/Projects/D_rerio/)

Tenascin-C, DQ096731 (Schweitzer et al., 2005)

Tenascin-R, XP_002660896

Tenascin-X, XP_002665649

Tenascin-W, NP_571111

Fibronectin 1, NP_571595

Fibronectin 1b, AY725818 (Sun et al., 2005)

CCN zCyr61-c5: paralog of mammalian Cyr61 on zebrafish chromosome 5, GQ273493;

zCyr61-c8: paralog of mammalian Cyr61 on zebrafish chromosome 8, GQ273499; zCyr61-

c23: paralog of mammalian Cyr61 on zebrafish chromosome 23, NM_001001826;

zCTGF-c19: paralog of mammalian CTGF on zebrafish chromosome 19, GQ920789;

zCTGF-c20: paralog of mammalian CTGF on zebrafish chromosome 20, NM_001015041;

zWISP1-c16: paralog of mammalian WISP1 on zebrafish chromosome 16, GQ273496;

zWISP1-c19: paralog of mammalian WISP1 on zebrafish chromosome 19, GQ273497;

zWISP2-c23: paralog of mammalian WISP2 on zebrafish chromosome 23, GQ273495;

zWISP3-c20: paralog of mammalian WISP3 on zebrafish chromosome 20, GQ273498.

Hyaluronan synthase NP_001157502 (HAS1), NP_705936 (HAS2), NP_775327 (HAS3)

CD44 XP_002667016

Vitronectin NP_001018508 (vitronectin a), NP_001132933 (vitronectin b)

FACIT collagen e.g. collagen IX. NP_998429 (alpha-1 IX), XP_695491 (alpha-3 IX)

Matrilin NP_001093210 (matrilin-1), NP_998714 (matrilin-2), NP_001004007 (matrilin-

3a), NP_001012385 (matrilin-3b), CAG27565 (matrilin-4)

Aggrecan XP_686182 (aggrecan a), ENSDART00000046249 (aggrecan b), [BM812126*](#)

(aggrecan b)

Versican XP_002662132 (versican a), NP_999853 (versican b)

Tetrapods

Xenopus tropicalis/Xenopus laevis (toads) (Hallsten et al., 2010)

Tenascin-C, ENSXETT00000051641, ENSXETT00000051644

Tenascin-R, NP_001107287

Tenascin-X, ENSXETT00000011268

Tenascin-W, ENSXETT00000043349

Fibronectin, AAH72841

CCN CTGF (AAH94492), Wisp-2 (AAH87808), Nov (NP_001079127), Cyr61

(NP_001079908)

Hyaluronan synthase NP_001120590 (HAS1), NP_001120557 (HAS2), AAP58398

(HAS3)

CD44 NP_988948

Vitronectin ACI29973 (from *X. laevis*)

FACIT collagen e.g. collagen IX. NP_001086796 (alpha-1 IX; from *X. laevis*),

NP_001090661 (alpha-2 IX), NP_001120497 (alpha-3 IX)

Matrilin NP_001025613 (matrilin-1), AAH63920 (matrilin-2), XP_002938060

(matrilin-4; from *X. laevis*)

Aggrecan XP_002932297

Versican NP_001104185 (from *X.laevis*)

Gallus gallus (chicken) (International Chicken Genome Sequencing Consortium, 2004)

Tenascin-C, AAA49086
 Tenascin-R, NP_990607
 Tenascin-X, CAA67509
 Tenascin-W, CAJ77765
 Fibronectin, XP_421868
 CCN CTGF (NP_989605), Wisp-1 (AAY21159), Nov (CAA41975), Cyr-61 (NP_001026734)
 Hyaluronan synthase NP_990137 (HAS2), XP_425137 (HAS3)
 CD44 NP_990191
 Vitronectin NP_990392
 FACIT collagen e.g., collagen IX. NP_001094381 (alpha-1 IX), XP_001233987 (alpha-2 IX), NP_990636 (alpha-3 IX)
 Matrilin NP_001025546 (matrilin-1), XP_424219 (matrilin-2), NP_990403 (matrilin-3), XP_425698 (matrilin-4)
 Aggrecan XP_001232950
 Versican NP_990118

Mus musculus (mouse) (Mouse Genome Sequencing Consortium et al., 2002)

Tenascin-C, NP_035737
 Tenascin-R, AAI38044
 Tenascin-X, BAA24436
 Tenascin-W, AAI38336
 Fibronectin, NP_034363
 Hyaluronan synthase NP_032241 (HAS1), NP_032242 (HAS2), NP_032243 (HAS3)
 CD44 NP_033981(isoform a)
 Vitronectin AAA40558
 FACIT collagen e.g., collagen IX. NP_031766 (alpha-1 IX), NP_031767 (alpha-2 IX), NP_034066 (alpha-3 IX)
 Matrilin AAH47140 (matrilin-1), AAH92298 (matrilin-2), AAH71224 (matrilin-3), AAH36558 (matrilin-4)
 CCN CTGF (AAD18058), Wisp-1 (NP_061353), Wisp-2 (NP_058569), Nov (NP_058569), Cyr61 (NP_034646), Wisp-3 (NP_001120848)
 Aggrecan AAC37670
 Versican NP_001074718 (versican isoform 1)

Homo sapiens (Venter et al., 2001)

Tenascin-C, CAA55309
 Tenascin-R, NP_003276
 Tenascin-X, NP_061978
 Tenascin-W, NP_071376
 Fibronectin, NP_997641
 Hyaluronan synthase NP_001514 (HAS1), NP_005319 (HAS2), NP_619515 (HAS3)
 CD44 ACI46596
 Vitronectin AAH05046.1
 FACIT collagen e.g., collagen IX, NP_001842 (alpha-1 IX), NP_001843 (alpha-2 IX), NP_001844 (alpha-3 IX)
 Matrilin CAI19322 (Matrilin-1), EAW91764 (Matrilin-2), AAI39908 (Matrilin-3), CAB46380 (Matrilin-4)
 Aggrecan AAH36445

Versican NP_004376.2 (Variant 1), NP_001119808.1 (Variant 2), NP_001157569.1 (Variant 3), NP_001157570.1 (Variant 4)
 CCN Wisp-1 (NP_003873), Wisp-2 (NP_543028), Wisp-3 (NP_937882), CTGF (NP_001892), Nov (NP_002505), Cyr61 (NP_001545)

KEY

*Sequence is from an expressed sequence tag

\$JGI genomes were searched against the model transcripts; only the top verified transcript hits from JGI or the elephant shark genome sequence project are included on the list.

REFERENCES

Adams JC et al. (2003) Characterisation of Drosophila thrombospondin defines an early origin of pentameric thrombospondins. *J Mol Biol* 328: 479-494.

Adams MD et al. (2000) The genome sequence of Drosophila melanogaster. *Science* 287: 2185-95.

Aparicio S et al., (2002) Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science* 297: 1301-1310.

Bauer DJ. (2007) The Daphnia Genomics Consortium Meeting: the genome biology of the model crustacean Daphnia. *Expert Rev. Proteomics* 4: 601-602.

Bentley AA, Adams JC (2010) The evolution of thrombospondins and their ligand-binding activities. *Mol Biol. Evol.* 27: 2187-97.

Boute N, Exposito JY, Boury-Esnault N, Vacelet J, Noro N, Miyazaki K, Yoshizato K and Garrone R. (1996) Type IV collagen in sponges, the missing link in basement membrane ubiquity *Biol. Cell* 88: 37-44.

Brower DL, Brower SM, Hayward DC, Ball EE. (1997) Molecular evolution of integrins: genes encoding integrin beta subunits from a coral and a sponge. *Proc Natl Acad Sci U S A.* 94: 9182-7.

C. elegans Sequencing Consortium. (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282: 2012-2018.

Chakravarti, R. and Adams, J.C. (2006) Comparative genomics of the syndecans defines an ancestral genomic context associated with matrilins in vertebrates. *BMC Genomics* 7: 83.

Chapman JA et al. (2010) The dynamic genome of Hydra. *Nature* 464: 592-96.

Cornillon S, Gebbie L, Benghezal M, Nair P, Keller S, Wehrle-Haller B, Charette SJ, Brückert F, Letourneur F, Cosson P. (2006) An adhesion molecule in free-living Dictyostelium amoebae with integrin beta features. *EMBO Rep.* 7: 617-21.

Cornillon S, Froquet R, Cosson P. (2008) Involvement of Sib proteins in the regulation of cellular adhesion in Dictyostelium discoideum. *Eukaryot Cell* 7: 1600-5.

Dehal P, et al. (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298: 2157-2167.

Eichinger L, et al. (2005) The genome of the social amoeba *Dictyostelium discoideum*. *Nature* 435: 43-57.

Exposito JY, van der Rest M, Garrone R (1993) The complete intron/exon structure of *Ephydatia mulleri* fibrillar collagen gene suggests a mechanism for the evolution of an ancestral gene module. *J. Mol. Evol.* 37: 254-259.

Exposito JY, et al. (2008) Demosponge and sea anemone fibrillar collagen diversity reveals the early emergence of A/C clades and the maintenance of the modular structure of type V/XI collagens from sponge to human. *J. Biol. Chem.* 283: 28226-28235.

Fanjul-Fernández M, Folgueras AR, Cabrera S, López-Otín C. (2010) Matrix metalloproteinases: evolution, gene regulation and functional analysis in mouse models. *Biochim Biophys Acta* 1803:3-19.

Hellsten U, et al. (2010) The genome of the Western clawed frog *Xenopus tropicalis*. *Science* 328: 633-36.

Huhtala M, Heino J, Casciari D, de Luise A, Johnson MS. (2005) Integrin evolution: insights from ascidian and teleost fish genomes. *Matrix Biol.* 24:83-95.

Hutter H, Vogel BE, Plenefisch JD, Norris CR, Proenca RB, Spieth J, Guo C, Mastwal S, Zhu X, Scheel J, Hedgecock EM. (2000) Conservation and novelty in the evolution of cell adhesion and extracellular matrix genes. *Science* 287: 989-94.

Huxley-Jones J, Robertson DL, Boot-Handford RP. (2007) On the origins of the extracellular matrix in vertebrates. *Matrix Biol.* 26: 2-11.

Hynes RO, Zhao Q. (2000) The evolution of cell adhesion. *J. Cell Biol.* 150: F89–F96.

International Chicken Genome Sequencing Consortium. (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695-716.

Jaillon O et al. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431: 946-957.

King N et al. (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoa. *Nature* 451: 783-788.

Knack BA, Iguchi A, Shinzato C, Hayward DC, Ball EE, Miller DJ. (2008) Unexpected diversity of cnidarian integrins: expression during coral gastrulation. *BMC Evol Biol* 8:136.

- Koehler A, Desser S, Chang B, MacDonald J, Tepass U & Ringuette M. (2009) Molecular evolution of SPARC: absence of the acidic module and expression in the endoderm of the starlet sea anemone, *Nematostella vectensis*. *Dev. Genes Evol.* 219: 509-521.
- Mouse Genome Sequencing Consortium et al. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520-62.
- Mueller WEG, Thakur NL, Ushijima H, Thakur AN, Krasko A, Le Pennec G, Indap, MM, Perovic-Ottstadt S, Schroeder HC, Lang G, Bringmann G. (2004) Matrix-mediated canal formation in primmorphs from the sponge *Suberites domuncula* involves the expression of a CD36 receptor-ligand system. *J. Cell. Sci.* 117: 2579-2590.
- Pancer Z, Kruse M, Müller I, Müller WE. (1997) On the origin of Metazoan adhesion receptors: cloning of integrin alpha subunit from the sponge *Geodia cydonium*. *Mol Biol Evol.* 14: 391-8.
- Putnam NH, et al. (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317: 86-94.
- Putnam NH, et al. (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453: 1064-1071.
- Robb SM, Ross E, Sánchez Alvarado A. (2008) SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Res* 36: D599-606.
- Satou Y, et al. (2002) A cDNA resource from the basal chordate *Ciona intestinalis*. *Genesis* 33: 153-154.
- Schweitzer J, Becker T, Lefebvre J, Granato M, Schachner M, Becker CG. (2005) Tenascin-C is involved in motor axon outgrowth in the trunk of developing zebrafish. *Dev Dyn.* 234: 550-66.
- Sea Urchin Genome Sequencing Consortium, et al. (2006) The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314: 941-952.
- Sebé-Pedrós A, Roger AJ, Lang FB, King N, Ruiz-Trillo I. (2010) Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proc Natl Acad Sci U S A.* 107: 10142-7.
- Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, Cavalier-Smith T. (2008) Multigene phylogeny of choanozoa and the origin of animals. *PLoS One* 3: e2098.
- Srivastava M, et al. (2008) The *Trichoplax* genome and the nature of placozoans. *Nature* 454: 955-960.
- Sun L, Zou Z, Collodi P, Xu F, Xu X, Zhao Q. (2005) Identification and characterization of a second fibronectin gene in zebrafish. *Matrix Biol.* 24: 69-77.

Tucker RP, Hess J, Drabikowski K, Ferralli J, Chiquet-Ehrismann R, Adams JC. (2006) Phylogenetic analysis of the tenascin gene family: evidence of origin early in the chordate lineage. *BMC Evolutionary Biology* 6: 60.

Tucker RP, Chiquet-Ehrismann R. (2009a) Evidence for the evolution of tenascin and fibronectin early in the chordate lineage. *Int. J. Biochem. Cell Biol.* 41: 424-434.

Venkatesh B, et al. (2007) Survey sequencing and comparative analysis of the elephant shark (*Callorhynchus milii*) genome. *PLoS Biol* 5: e101.

Venter CJ et al. (2001) The sequence of the human genome. *Science* 291: 1304-1351.

Vinson JP, Jaffe DB, O'Neill K, Karlsson EK, Stange-Thomann N, S, Mesirov JP, Satoh N, Satou Y, Nusbaum C, Birren B, Galagan JE, Lander ES (2005) Assembly of polymorphic genomes: Algorithms and application to *Ciona savignyi* *Genome Res.* 15: 1127-1135.

Vizzini A, Arizza V, Cervello M, Cammarata M, Gambino R and Parrinello N. (2002) Cloning and expression of a type IX-like collagen in tissues of the ascidian *Ciona intestinalis*. *Biochim. Biophys. Acta* 1577: 38-44.

Wagener R, Ehlen HW, Ko YP, Kobbe B, Mann HH, Sengle G, Paulsson M. (2005) The matrilins--adaptor proteins in the extracellular matrix. *FEBS Lett.* 579: 3323-29.

Weigel PH, DeAngelis PL. (2007) Hyaluronan synthases: a decade-plus of novel glycosyltransferases. *J Biol Chem.* 282: 36777-81.

Whittaker CA, Bergeron KF, Whittle J, Brandhorst BP, Burke RD, Hynes RO. (2006) The echinoderm adhesome. *Dev. Biol.* 300: 252-266.

Taxonomy/Species	Fibrillar collagen	Collagen IV	Laminin	Perlecan	Nidogen	Fibrillin	Thrombospondin	Aggrin	SPARC	MMP	ADAMTS
Porifera: <i>several species</i>	+	+(Homosclera only)	+	+	#	+	+	+	+	+	+
Placozoa: <i>Trichoplax adhaerens</i>	-	+	+	+	+	+	+	+	-	-	+
Cnidaria: <i>Nematostella vectensis</i>	+	+	+	+	+	+	+	-	+	+	+
<i>Hydra magnipapillata</i>	+	+	+	+	-	+	+	-	+	+	+
Platyhelminth: <i>Schmidtea mediterranea</i>	+	#	+	+	-	+	-	+	+	+	+
Nematode: <i>Caenorhabditis elegans</i>	-	+	+	+	+	+	-	+	+	+	+
Annelid: <i>Capitella teleta</i>	+	+	+	+	+	+	+	+	+	+	+
Mollusc: <i>Lottia gigantea</i>	+	+	+	+	+	+	+	+	+	+	+
Arthropods: <i>Daphnia pulex</i> (crustacean)	+	+	+	+	+	+	+	+	+	+	+
<i>Drosophila melanogaster</i> (insect)	-	+	+	+	+	+	+	-	+	+	+

Table S1. Representation of ECM components and major ECM protease categories in basal metazoa and protostomes. Based on current sequenced genomes, EST datasets, cDNA sequences and data from Boute et al., 1996, Hynes and Zhao 2000, Hutter et al., 2001, Adams et al., 2003, Exposito et al., 2008, Fanjul-Fernández et al., 2009, Koehler et al., 2009, Bentley and Adams, 2010. All references are in Supp. File 1. #Insufficient information from genome or ESTs.

Taxonomy/Species	Integrin α	Integrin β	CD36	Syndecan	Glypican	DDR	Dystro- glycan	NG2/ CSPG4
Porifera: several species	+	+	+	+	+	-	-	-
Placozoa: <i>Trichoplax adhaerens</i>	+	+	+	-#	+	-*	+	-
Cnidaria: <i>Nematostella vectensis</i>	+	+	+	+	+	-*	+	-
<i>Hydra magnipapillata</i>	+	+	+	+	+	+	+	-
Platyhelminth: <i>Schmidtea mediterranea</i>	-#	+	+	+	-	-	-	-
Nematode: <i>Caenorhabditis elegans</i>	+	+	+	+	+	+	+	+
Annelid: <i>Capitella teleta</i>	+	+	+	+	+	+	+	+
Mollusc: <i>Lottia gigantea</i>	+	+	+	+	+	+	+	+
Arthropods: <i>Daphnia pulex</i> (crustacean)	+	+	+	+	+	+	+	+
<i>Drosophila melanogaster</i> (insect)	+	+	+	+	+	+	+	+

Table S2. Representation of ECM adhesion receptors in basal metazoa and protostomes.

Based on current sequenced genomes, EST datasets, cDNA sequences and data from Brower et al., 1997, Pancer et al., 1997, Hynes and Zhao, 2000, Hutter et al., 2001, Hutala et al., 2005, Chakravarti and Adams, 2006, Exposito et al., 2008, Chapman et al., 2010. All references are in Supp. file 1. DDR = discoidin domain receptor. *A DDR-like discoidin domain and receptor tyrosine kinase domain are encoded by separate predicted proteins, XP_002109287 and XP_002109288, respectively (*T. adhaerens*) and XP_001629823 and XP_001637553 (*N. vectensis*). #Insufficient information from genome or ESTs.

Taxonomy/Species/common name	FACIT collagen	CCN	Matrilin	Tenascin	Fibronectin	Hyaluronan synthase	Vitronectin	Versican	Aggrecan
Echinoderm: <i>Strongylocentrotus purpuratus</i> (sea urchin)	+	-	-	-	-	-	-	-	-
Hemichordate: <i>Saccoglossus kowalevskii</i> (acorn worm)	+	+	-	-	-	-	-	-	-
Urochordate: <i>Ciona intestinalis</i> (sea squirt)	+	+	+	+	FN-like	-	-	-	-
Cephalochordate: <i>Branchiostoma floridae</i> (Amphioxus)	+	+	+	+	-	+	-	-	-
Craniate: <i>Petromyzon marinus</i> (lamprey)	+#	-#	+#	+	-#	+	-#	-	-#
Craniate, Gnathostome: <i>Callorhynchus milii</i> (elephant shark)	+	-#	+	+	+	+	+	+	+
Craniate, Gnathostome, ray-finned fish: <i>Takifugu rubripes</i> (puffer fish)	+	+	+@	+*	+	+	+	+	+
<i>Tetraodon nigriviridis</i> (freshwater puffer)	+	+	+@	+*	+^	+	+	+	+
<i>Danio rerio</i> (zebrafish)	+	+	+@	+	+^	+	+	+	+
Craniate, Gnathostome, tetrapods: <i>Xenopus tropicalis</i> (toad)	+	+	+	+	+	+	+	+	+
<i>Gallus gallus</i> (chicken)	+	+	+	+	+	+	+	+	+
<i>Mus musculus</i> (mouse)	+	+	+	+	+	+	+	+	+
<i>Homo sapiens</i> (human)	+	+	+	+	+	+	+	+	+

Table S3. Major ECM innovations in Deuterostomes.

Based on current sequenced genomes, EST datasets, cDNA sequences and data from Schweitzer et al., 2005, Sun et al., 2005, Wagener et al., 2005, Chakravarti and Adams, 2006, Whittaker et al., 2006, Tucker et al., 2006, Huxley-Jones et al., 2007, Weigel et al., 2007, Tucker and Chiquet-Ehrismann et al., 2009. All references are listed in Supp. File 1.

*has two TN-C paralogues, ^ has two FN paralogues, @ has two matrilin-3 paralogues. #Insufficient information from genome.