Electronic Supplementary Material:

Figure S1  Pages 2-3
Figure S2  Pages 4-5
Figure S3  Pages 6-29
Figure S4  Pages 30-34
Figure S5  Pages 35-36
Figure S1. Maximum Likelihood phylogenetic tree of the SPARC family. SPOCK family was used as outgroup. aLRT results are shown on each node. Scale bar indicates number of substitutions per site. See table S1 for accession numbers and name abbreviations.
Figure S2. (A) Domain organization of SPARC, SPOCK and SMOC families members in amphioxus and zebrafish. (B) Alignment of SPARC family members. Only the two regions involved in collagen binding are shown. Positions indicated correspond to human SPARC. There are five essential residues for collagen binding in human SPARC that are: R149, N156, L242, M245, and E246 (Sasaki, T., et al., Crystal structure and mapping by site-directed mutagenesis of the collagen-binding epitope of an activated form of BM-40/SPARC/osteonectin. Embo J, 1998. 17(6): p. 1625-34). The structural analysis of SPARC and collagen binding showed that the "Phe pocket" of the SPARC-collagen interface is delineated by the residues F146, M150, W153 and L242 and that a salt-bridge is formed between R149 and E246 (Hohenester, E., et al., Structural basis of sequence-specific collagen recognition by SPARC. Proc Natl Acad Sci U S A, 2008. 105(47): p. 18273-7). Except for N156, the residues important for SPARC-collagen interaction are overall well conserved between all the SPARC family members in all the species studied.
Figure S3. Maximum Likelihood phylogenetic trees of all the gene families used in the synteny conservation analysis. aLRT results are shown on each node. Scale bar indicates number of substitutions per site. Tip labels correspond to Ensembl or JGI accession numbers of the sequences. Amphioxus sequences are highlighted in yellow.
CENPE
KIF3C
KIF3B
KIF3A
KIF11
**Figure S4.** Detailed schemes of synteny conservation between amphioxus and human, and between human and zebrafish. The dotted lines are connecting orthologous genes. Paralogous and orthologous genes are represented by rectangles of the same color in each page. For the sake of simplicity we did not schematize the synteny conservation between amphioxus and zebrafish. Names and positions on the scaffolds/chromosomes are given. Bf: *Branchiostoma floridae*, Hs: *Homo sapiens*, Dr: *Danio rerio*. Gene names are according to Ensembl nomenclature. Detailed data including all the accession numbers are given in table S3.
HsChr5 134-180Mb

DrChr14 6-52Mb

1cm=2Mb
Figure S5

Figure S5. Amphioxus SPARC/SPARCL1 (A-G), amphioxus SPARCB (A'-F') and zebrafish SPARCB (A''-C'') in situ hybridizations. Amphioxus SPARC/SPARCL1 is expressed in the paraxial mesoderm of late gastrula (A-B), early neurula (C-D) and late neurula (E-F) stage embryos. In larvae, expression is observed in paraxial and ventral mesoderm as well as in the notochord (G). Amphioxus SPARCB is expressed in the anterior notochord of late neurulae (A'-B'), then in the notochord and pharyngeal endoderm of neurula stage embryos before mouth opening (C'). This expression pattern persists in the larvae (D'-F'). SPARCB expression was analysed in zebrafish embryos at 24 (A''), 48 (B'') and 72hpf (C''). Only a weak SPARCB expression can be detected in the lens between 24 and 48hpf. Panels A, C, E, G, A', C', D', E', A''-C'' are lateral views with anterior to the left. Panels B, D, F, B' are dorsal views with anterior to the left. Panel F' is a dorsal view of the pharyngeal region of the larva presented in D' and E'.