Article

Not peer-reviewed version

Orthologs at the Base of the Olfactores Clade

Wilfred Donald Stein *

Posted Date: 17 April 2024

doi: 10.20944/preprints202404.1133.v1

Keywords: evolution; orthologs; tunicates; olfactores; neural crest; type II cadherins; crystalline; connexins

 doi:10.20944/preprints202404.1133.v1

Article

Orthologs at the Base of the Olfactores Clade

Wilfred D. Stein

Silberman Institute of Life Sciences, Hebrew University, Jerusalem, Israel 91904; wdstein@mail.huji.ac.il

Abstract: Tunicate orthologs in the human genome comprise just 84 genes of the 19,872 protein-coding genes, yet they stand at the base of the Olfactores clade - which radiated to generate thousands of tunicate and vertebrate species. What were the powerful drivers among the 84 that enabled this process? Many of these orthologs are present in gene families. We discuss the biological role of each family and the orthologs' quantitative contribution to the family. Most important was the evolution of a second type of cadherin. This, a Type II cadherin, had the property of detaching the cell containing that cadherin from cells that expressed the Type I class. The set of such Type II cadherins could now detach and move away from their Type I neighbours, a process which would eventually evolve into the formation of the neural crest, " the fourth germ layer", providing a wide range of possibilities for further evolutionary invention. A second important contribution were key additions to the broad development of the muscle and nerve protein and visual perception tool-kits. These developments in mobility and vision provided the basis for the development of the efficient predatory capabilities of the Vertebrata.

Keywords: evolution; orthologs; tunicates; olfactores; neural crest; type II cadherins; crystallins; connexins

1. Introduction

In a 2010 paper, Domazet-Loso and Tautz [1] proposed a classification of the animal kingdom into 19 phylostrata ranging from phylostratum 1, the first unicellular organisms, to phylostratum 19, the primates. In an alternative cladistic scheme – depicted in Figure 1, the phylostratum number is shown in parentheses after the name of the appropriate grouping:

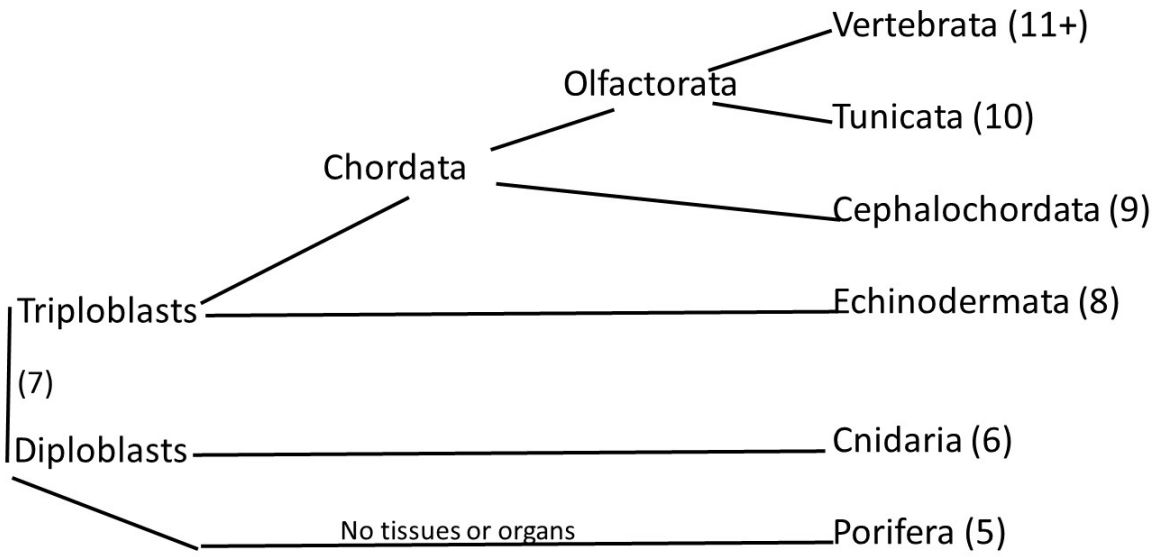


Figure 1. Cladistic diagram of animal evolution with the tunicates as the sister clade of the vertebrates, together forming the Olfactores clade [2].

 doi:10.20944/preprints202404.1133.v1

The vertebrates and the tunicates share many of the genes that appeared over evolutionary time up to and including the emergence of the tunicates. Of these genes, those that were not present in earlier-appearing animals are the tunicate orthologs, the tunicate contributions to the vertebrate genome. By searching for orthologs, Domazet-Loso and Tautz [1] assigned many of the genes of the human genome to one or other of the 19 phylostratum levels. But finding orthologs is not an easy matter. Indeed, the Search for Orthologs Consortium comprises some dozen research groups that use various algorithms in their searches to handle the difficult problem of comparing tens of thousands of genes among thousands of living species. The results reported by these different research groups are by no means consistent. Liebeskind et al., [3] adopted a consensus approach where, for each gene, the modal value of the reports of the different research groups was chosen to represent that gene's phylostratum level. Also reported was the step value, which is the median of the difference, for each gene, between the modal value and that reported by each ortholog research group. Litman and Stein [4] extended this work and published a list of all the protein-coding genes of the human genome with each gene recorded with its phylostratum number.

The tunicates, with which this paper is concerned, fall in phylostratum 10 and the orthologs that they share with humans lie at the base of the Olfactores clade that consists of the tunicates and the vertebrates [2]. Phylostratum 10 in the Litman and Stein study [4] contained 84 genes. This number is almost an order of magnitude smaller than the number of genes found for any of the other phylostrata. We were intrigued by the fact that this small number of genes would lie at the base of the Olfactores clade which branched out to the thousands of tunicate and vertebrate species. In what way did such a small number of genes begin that vast onward evolutionary process? But first we had to be sure that we had a valid set of tunicate orthologs and we proceeded to build a hand-curated set. Many of these orthologs are found to be present in gene families. In what follows, we will discuss these families in turn. The biological role of each family is described and also the extent to which the tunicate orthologs contribute quantitatively to the family.

2. Methods

Searches for orthologs were performed using the protein BLAST (Basic Local Alignment Search Tool) program of the NCBI (National Center for Biotechnology Information of the National Library of Medicine): https://blast.ncbi.nlm.nih.gov/Blast.cgi with the following search parameters: Max target sequences 1000, Expect threshold 200000, Word size 2, Max matches in a query range 0, Matrix BLOSUM62, Gap Costs Existence: 11 Extension: 1, No compositional adjustments, No Low complexity regions filter.

To be recognised by us as an ortholog, a sequence found when searching the tunicates database had to be annotated with the same name as the probe sequence from the human genome. A sequence annotated as "-like" was rejected. In addition, the probe sequence had to be absent when the Branchiostomatidae database, or that of the Echinodermata or earlier -appearing clades were searched. In a few cases, noted explicitly in Table S1, the tunicate sequence had been annotated as "name-a" instead of "name-1", and similarly for "b", "c", and "d". In other cases, again explicitly listed in Table S1, the tunicate sequence's name was a more generic synonym but presented with a convincing Expect Value.

Sequence comparisons between two proteins and dot plots of the comparisons were made using the BLAST 2-sequences tool of the BLAST program, using the same search parameters as above. All one-on-one Expect values recorded in this paper were the results of a BLAST 2-sequences comparison between the two proteins.

Alignments between protein sequences were established and dot plots generated using the COBALT (Constraint-Based Alignment Tool) aligner at the NCBI https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?CMD=Web.

Properties of the orthologs listed in Table S1 were taken from the listings in GeneCards https://genealacart.genecards.org/Result. GeneCards proved useful also when the aliases of genes discussed in the literature had to be interpreted to provide HGNC symbols.

 doi:10.20944/preprints202404.1133.v1

For finding the Unique Gene IDs of the tunicate-shared orthologs for Table S1, we searched the ANISEED database at:

https://www.aniseed.fr/aniseed/gene/?choice=find_gene&module=aniseed&action=gene:index

Gene Ontology data for the genes not found to be present in gene families was extracted from The Database for Annotation, Visualization and Integrated Discovery (DAVID) at https://david.ncifcrf.gov/summary.jsp

3. Results

Table 1 lists the 84 genes reported in [4] as being in phylostratum 10, each recorded with its HGNC symbol together with its step value, and the number of research groups (of the 13 considered) whose reported phylostratum number agreed with the modal value (here 10):

Table 1. Orthologs from Phylostratum 10: Consortium's Spread as steps (see text).

HGNC	Steps	equal to mode	HGNC	Steps	equal to mode
ADGRL3	3.15	4	*MGAT4D*	4.08	4
ANKRD16	4.77	4	*MSMP*	1.15	6
ANKRD42	2.38	8	*MUM1*	2.10	5
ARHGEF38	2.67	4	*MYBPC3*	2.08	6
ASB2	3.92	5	*MYL5*	5.01	3
ATRAID	0.46	11	*NEBL*	2.16	6
BCL6B	2.95	4	*NECTIN3*	1.85	4
C18ORF21	0.62	9	*NEXN*	2.46	5
CALML6	2.47	8	*NPC1L1*	4.85	5
CASR	1.92	7	*OLFML2A*	1.54	5
CATIP	3.18	4	*OTULIN*	1.55	7
CBLN1	1.38	4	*PDX1*	2.77	4
CCDC78	2.62	4	*PHF21A*	2.46	4
CEP63	3.40	5	*PODN*	3.46	4
CEP83	2.46	7	*PRR14*	2.27	6
CKAP2	2.54	5	*RBM14*	3.61	3
CLDN1	0.83	7	*RHBDL2*	4.54	4
CLDN18	1.25	5	*RNF4*	4.85	3
CLDN19	0.92	7	*SCLT1*	3.18	6
CLDN7	1.23	6	*SERINC5*	4.38	5
EFNA1	1.92	4	*SLC35G2*	3.31	4
FABP2	1.00	8	*SLC35G3*	4.86	3
FAM155B	1.54	5	*SLC35G4*	5.13	3
FAM187A	1.23	6	*SLC35G5*	5.50	3
FAM3D	2.38	4	*SLC35G6*	4.25	4
FBXO24	2.23	6	*SLC6A14*	3.54	3
FOXH1	2.95	5	*SLC6A6*	3.46	4
GJA10	2.02	5	*SNRNP48*	2.31	8
GJA3	1.16	7	*SPATA21*	4.50	3
GJC2	2.38	6	*STXBP6*	1.77	8
GNA15	4.00	3	*TGFB1*	2.15	4
GTF2IRD2	3.00	3	*TGFB2*	1.54	6
GTF2IRD2B	4.30	3	*TLCD2*	3.62	3
GVQW1	4.05	2	*TMEM218*	0.62	9
H1FOO	3.23	3	*TNNC1*	4.08	6
HMMR	2.92	8	*TNNC2*	4.08	4
HNF1A	1.15	7	*TSPAN12*	2.69	5
HSPB1	3.23	3	*WSB1*	2.31	4

 doi:10.20944/preprints202404.1133.v1

IKZF5	2.85	5	*ZBED5*	2.46	4
IL17C	1.78	4	*ZBED8*	2.83	4
KIZ	1.23	6	*ZC3H7B*	1.38	8
LRRC29	3.09	4	*ZNF91*	4.88	2

The mean for the step value of these phylostratum 10 genes is 2.743 and its standard distribution 1.240. It would not be surprising therefore to find that many of those in the list of 84 were incorrectly assigned to phylostratum 10, being in reality one or two levels higher or lower. It was thought necessary, therefore, to curate the list by doing a BLAST search - for each gene on the list, to check that it did appear annotated as such in the tunicates and did not appear as such in the branchiostomatidae nor in the earlier-appearing phylostrata. Only 25 genes of the 84 survived this curation.

Inasmuch as genes with phylostratum numbers reported as being above 10 might be incorrectly listed, although in reality being in phylostratum 10, we searched the literature for genes reported as being present in tunicates. If they were not already in the list of 25 curated 10s, they were subjected to a BLAST search as just described - for their presence in tunicates and not in earlier animals. 59 such genes were identified and are listed together with the original 25 in Table S1, the total number of curated tunicate-vertebrate -shared orthologs being thus 84 (coincidentally the same number as in the original list). For each gene in the list, the Expect value is listed for the 2-sequence comparison between that tunicate sequence and the corresponding human gene as well as additional information concerning that gene.

The list of 84 orthologs in Table S1 is almost certainly not a complete accounting. Many genes that the previous estimates have identified as coming from the large contribution from the fish might be revealed to be tunicate-derived orthologs if systematically studied by our hand-curation methodology.

3.1. The Muscle Protein Genes

The Proteins of Muscle

Muscular tissue can be found already in the Porifera where the sponges have the capacity for co-ordinated whole-body contractions that enable them to expel sediments [5]. Their myocytes contain a myosin homologous to the myosins of higher organisms but the myosin of the myocytes contracts and expands in response to changes in the concentration of calcium and is not ATP-dependent, nor activated by any neural connection ([6]. The muscular system is already well-developed in the sea anemone where muscles can be found in the tentacles as well as in the body column [7]. The sea anemone being diploblastic, its muscles, of course, are not mesodermal origin. In the triploblastic sea urchin, a complex muscular jaw apparatus is present that can scrape and tear the animal's food items [8]. With the emergence of the cephalochordates such as the lancelet, (amphioxus sp. of the branchiostomes), muscles line the notochord to allow the active swimming behaviour of the animal. In the tunicates, swimming is further developed and in the Appendicularians, which retain the larval form for all their lifespan, the tail is used to sweep food particles into the oral apparatus [9].

At each stage of evolutionary development, the range of molecules involved in muscle formation is, of course, increased. These molecules include actins, the light chain and heavy chain myosins, titin, meromyosins, myosin binding proteins, troponins and tropomyosin. In most of these classes, tunicate orthologs contribute – in some cases, decisively -as we will now illustrate.

The muscle protein orthologs are listed in Table 2, which is excerpted from Table S1 and serves as an example of what can be found in that large Table.

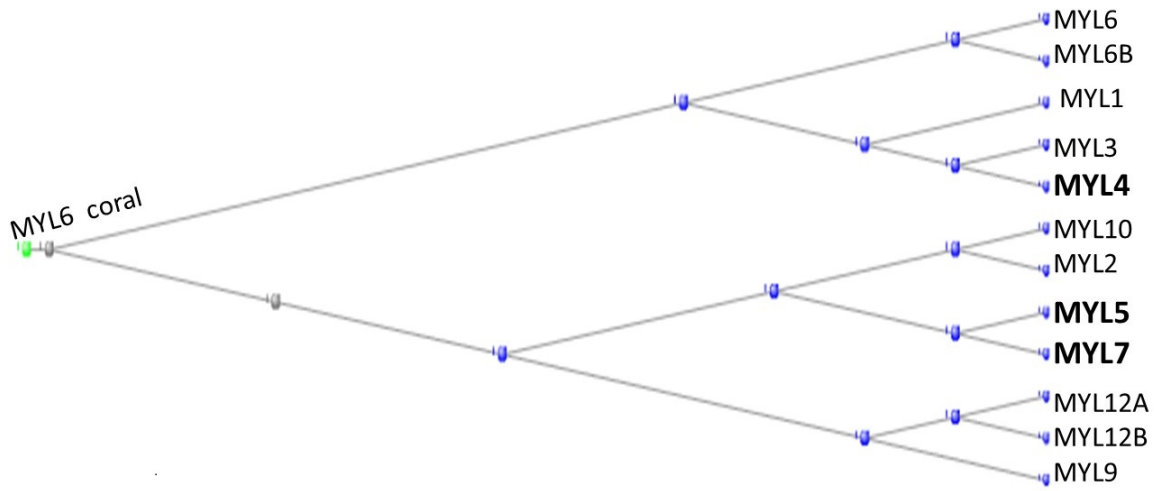
 doi:10.20944/preprints202404.1133.v1

Table 2. Descriptive data on muscle-related proteins excerpted from Table S1 of Supplementary Materials.

HGNC	Unique ID*	Uniprot	Expect value**
MYBPC1	Cirobu.g00000146	Q00872	0
MYBPC2	Cirobu.g00000146	Q14324	0
MYBPC3	Cirobu.g00000146	Q14896	0
MYH1	Boschl.g00071919	P12882	0 as myosin heavy chain, cardiac muscle isoform *Ciona intestinalis*
MYH13	Harore.g00014293	Q9UKX3	0 as embryonic muscle myosin heavy chain *Halocynthia roretzi*
MYH15	Harore.g00003537	Q9Y2K3	0 as embryonic muscle myosin heavy chain *Halocynthia roretzi*
MYH4	Cirobu.g00005587	Q9Y623	0 as myosin heavy chain, cardiac muscle isoform *Ciona intestinalis*
MYH8	Harore.g00010698	P13535	0 as embryonic muscle myosin heavy chain *Halocynthia roretzi*
MYL4	Cirobu.g00008856	P12829	2E-74 as smooth muscle isoform X3 of Ciona intestinalis
MYL5	Cirobu.g00008931	Q02045	2.00E-82
MYL7	Cirobu.g00009534	Q01449	3E-73 as smooth muscle isoform X1 of Ciona intestinalis
MYOM1	Phmamm.g00001992	P52179	7.00E-93
MYBPC1	Cirobu.g00000146	Q00872	0
MYBPC2	Cirobu.g00000146	Q14324	0
MYBPC3	Cirobu.g00000146	Q14896	0
PALLD	Cirobu.g00008036	Q8WX93	1.00E-109
SYNPO2	Boleac.g00003200	Q9UMS6	2.00E-13
TNNT1	Cirobu.g00006671	P13805	8E-85 as TroponinT, slow skletal muscle of *Ciona intestinalis*

*From the ANISEED database – see Methods ; **Found using the BLAST 2-sequence program – see Methods. Subject specified if no identical match in BLAST search.

The light chain myosins have 12 representatives in the human genome with HGNC symbols beginning with MYL. Figure 2 depicts a phylogram of these proteins, rooted at the annotation of the coral ortholog of MYL6.



 doi:10.20944/preprints202404.1133.v1

Figure 2. Phylogram of the MYL proteins, rooted at the Coral ortholog of MYL6. The orthologs contributed by the tunicates are shown in bold type.

In addition to that protein, MYL2 and MYL6B have also orthologs in the corals. MYL4, MYL5 and MYL7 (in bold in the figure) have orthologs in the tunicates, while all the other proteins have orthologs either in the Echinodermata or in the Cephalochordata, leaving it to those 4 tunicate proteins to complete the family.

In a comprehensive analysis of the proteins in skeletal and cardiac muscles, Lindskog et al., (2015) found that MYL2 and MYL3 were present in both tissues, MYL1 only in skeletal muscle, while MYL4 and MYL7 were present only in cardiac muscle [10].

In the fourteen-membered heavy chain myosin class (having HGNC symbols beginning with MYH), the tunicate orthologs contribute five examples (Figure 3 below).

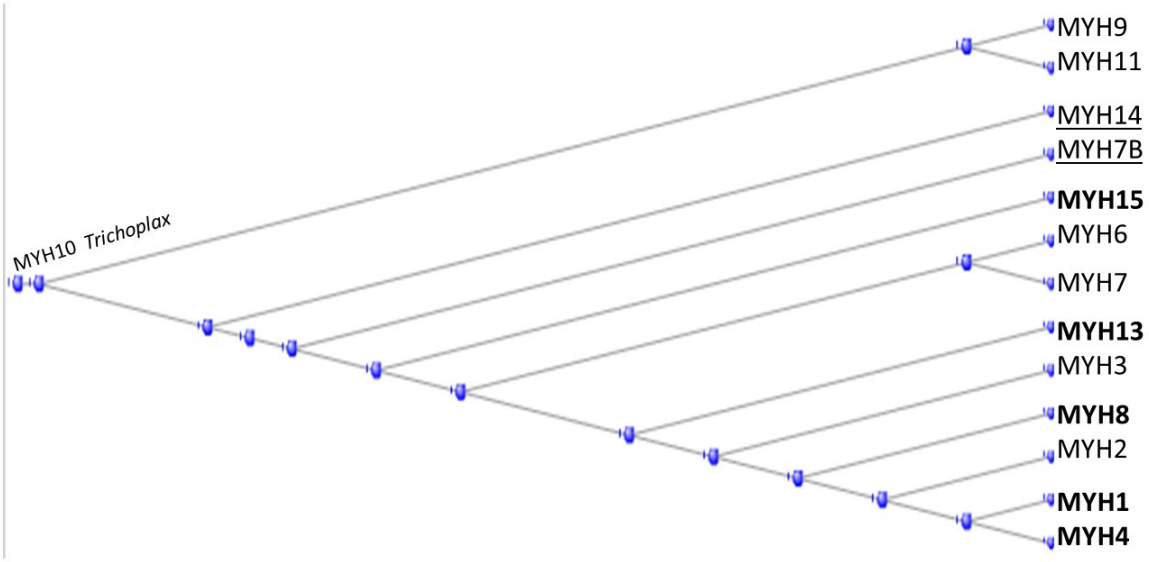


Figure 3. Phylogram of the MYH proteins, rooted at the annotation of the Trichoplax ortholog of MYH10. Tunicate orthologs in bold type. Fish orthologs underlined.

MYH14 and MYH7B (underlined) appeared only with the fish but the tunicates had contributed a substantial portion of the earlier orthologs.

Lindskog et al., (2015) found that MYH7 and MYK7b were present in both skeletal and cardiac tissues, MYH1, MHY2, MYH4 AND MYH8 only in skeletal muscle, while MYH6 was only in cardiac muscle [10].

The tunicates contributed all three of the myosin binding protein C family (the MYBPCs). MYBPC1 is present in slow skeletal muscle, MYBPC2 in fast skeletal muscle and MYBPC3 in cardiac muscle.

 doi:10.20944/preprints202404.1133.v1

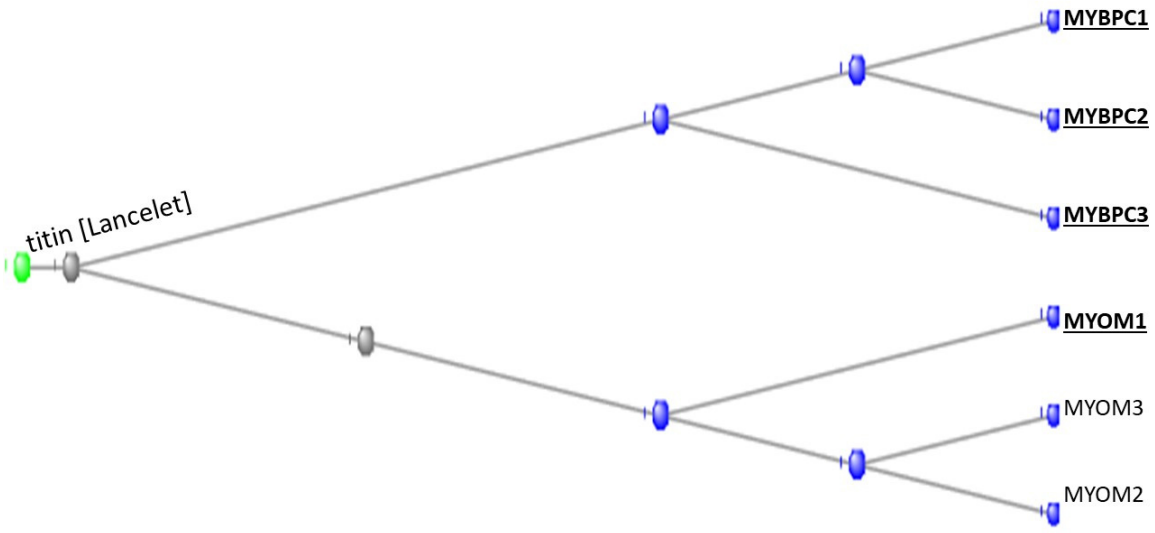


Figure 4. Phylogram of the MYBPC and MYOM proteins, rooted at the Lancelet ortholog of the giant muscle protein Titin, which appeared with the Branchiostomata. The orthologs contributed by the tunicates are shown in bold type.

The tunicates contributed one of the three myomesin proteins (the MYOMs). The myomesins are present in the M band of striated muscles and act by stabilising the sarcomere during contraction of the muscle, acting as a molecular spring [11].

Two other proteins in the list presented by Lindskog et al., 2015 [10] are tunicate orthologs and thus appear, annotated, in Table S1: One is synaptopodin 2 (SYNPO2), the first to appear in evolution of the two synaptopodins, which are molecules that regulate actin fibres. The second is troponin T type 1 (TNTT1), the last to appear of the three TNTTs that regulate muscle contraction in response to alterations in intracellular calcium ion concentration.

3.2. The Gap Junction Proteins

The human genome contains 21 Gap Junction proteins (also known as connexins). They are divided by sequence similarity into five subgroups labelled as GJAP's through GJEP's or with the corresponding Greek letters, α through ε. Figure 5 shows a phylogram of these proteins:

 doi:10.20944/preprints202404.1133.v1

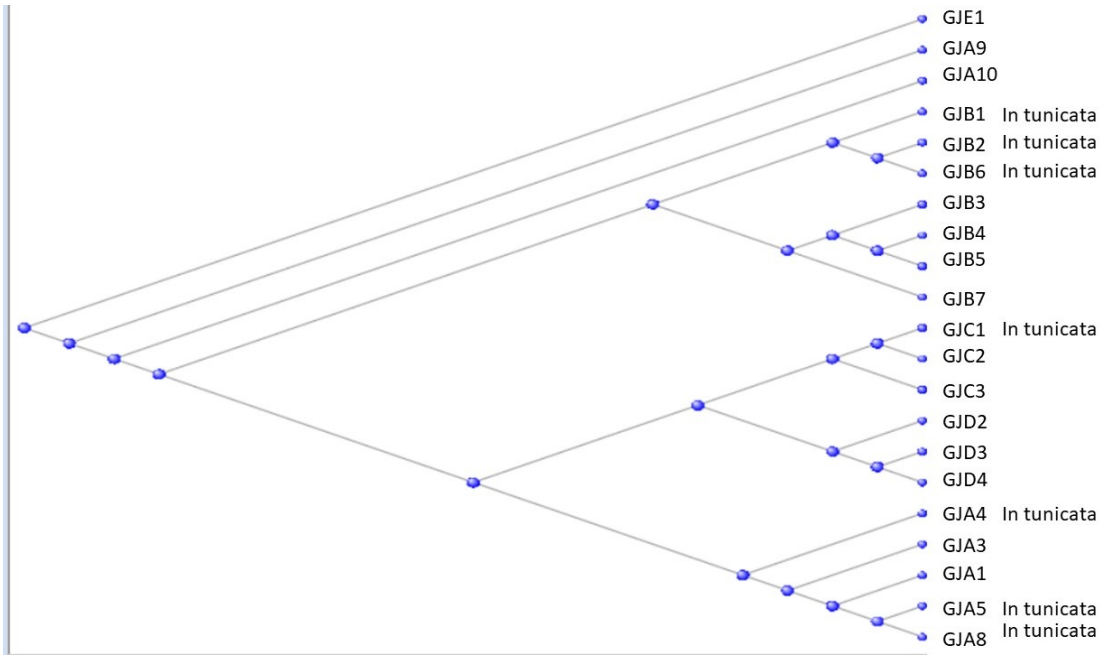


Figure 5. Phylogram of the Gap Junction proteins. Those found with orthologs in the Tunicata are labelled on the figure.

(see also [12,13] for fuller discussions of the evolution and biology of the Gap Junction proteins). Seven of the gap junction proteins are found in Table S1 of the orthologs at the base of the Olfactores, and these are labelled accordingly in the figure. Results of BLAST searches against the tunicate genomes are listed in Table 3:

Table 3. Gap Junction Genes in the Tunicate Genome Annotation**.

HGNC Symbol	Uniprot Symbol	Expect Value*	XP_026690575.1
GJA4	P35212	4.00E-41	XP_002119855.2
GJA5	P36382	1.00E-32	XP_004227068.2
GJA8	P48165	3.00E-57	CAB3249107.1
GJB1	P08034	6.00E-34	CAB3249120.1
GJB2	P29033	7.00E-35	XP_026696017.1
GJB6	O95452	3.00E-47	XP_002130872.1
GJC1	P36383	3.00E-67	XP_026691633.1
*The Expect values were from 2-sequence BLAST analyses against the corresponding human homolog			** from BLAST search. Click to retrieve the FASTA sequence

As can be seen in Figure 5, the A, B, C, and D subgroups all have representatives from the tunicates which, in each case, may be suggested to be the progenitors of that subgroup.

The Branchiostomatidae genomes do not contain gap Junction proteins so the gap junction proteins may be considered as inventions of the founders of the Olfactores clade. Gap junction proteins are transmembrane proteins that connect one cell with its neighbour. Among many other roles in the body, gap junction proteins are found in the Schwann cells. These Schwann cells form a sheath that wraps around a neuron and provides the insulating myelin coating to the nerves of the bony fishes and higher organisms. Gap junction proteins connect each Schwann cell with its neighbour in the sheath and thus provide a route that enables rapid cell to cell movement of solutes from the nerve cell interior to the extracellular surface. A build-up of potassium within the nerve cell is thus mitigated and nutrient flow from the extracellular medium to the interior of the nerve is facilitated. This transport has been estimated to be a million times faster than would such movement otherwise be in the absence of the gap junction proteins. The myelin coating that the Schwann cells

 doi:10.20944/preprints202404.1133.v1

provide insulates the nerve and allows electrical signals to pass from the interior to the exterior only where the Schwann cell coating is absent -at the internodes. The electrical signal thus jumps ("saltatory movement") from internode to internode, speeding up manyfold the rate of signal transmission. For example, whereas unmyelinated axon conduction velocities range from about 0.5 to 10 m/s, myelinated axons can conduct at velocities up to 150 m/s. [14]. In a giraffe, a signal that takes some 60 milliseconds to travel from the foot to the brain [15] will instead take some 6 seconds. It is to be noted that the cyclostomes such as the lamprey do not produce myelin, although they possess Schwann cells [16]. Indeed, BLAST searches of the lamprey genome for some of the genes that a GeneCards search suggested as being involved in myelin production (MPZ, MBP, PMP2, and MAG) proved fruitless. Thus myelin is an invention of the jawed fish.

Some myelin-associated genes (MYRF, PLP1, and PMP2) are present in the Tunicate genome (see also [17]) and, as we have seen, seven Gap Junction genes are present in the tunicates. Their function does not seem to be known, but a search of ANISEED, the ascidian database (see Methods), for the Gap Junction gene GJC1 returned

REG00001500 (Cirobu.REG.KhC2.5771334-5773564|Msx)

In which for this gene a "CRM (cis-regulatory module) is active in the region of the primordial pharynx, in the sensory vesicle and in the neck connecting the sensory vesicle, in the visceral ganglion and in the caudal muscles. At earlier stages, it is active along the neural tube in the tailbud, along the neural folds in the neurula, in the precursors of sensory organs and in the ventral epidermis in both stages, and also in the mesoderm that will give rise to the muscles". Presumably this gene and its congeners were co-opted by the ancestor of the Schwann cell to enable the formation of the myelinated nerves of the jawed fish and their descendants.

We searched the genome of the Branchiostomatidae for a gene that could be the ancestor of the gap junction genes of the tunicates. We found that GJAP8 had, as top hit with the Branchiostomatidae, a gene denoted as the "predicted Titin homologue" annotated as XP_019638343.1. This had an Expect Value against GJAP8 of $9e^{-11}$. The protein had a length of 1894 amino acids. Figure 6 depicts the alignment of this protein against the seven tunicate Gap Junction genes, Figure A showing the alignment against the whole Titin sequence while Figure B shows the alignment against the 891 portion of the sequence that spans those of the shorter GAP Junction sequences.

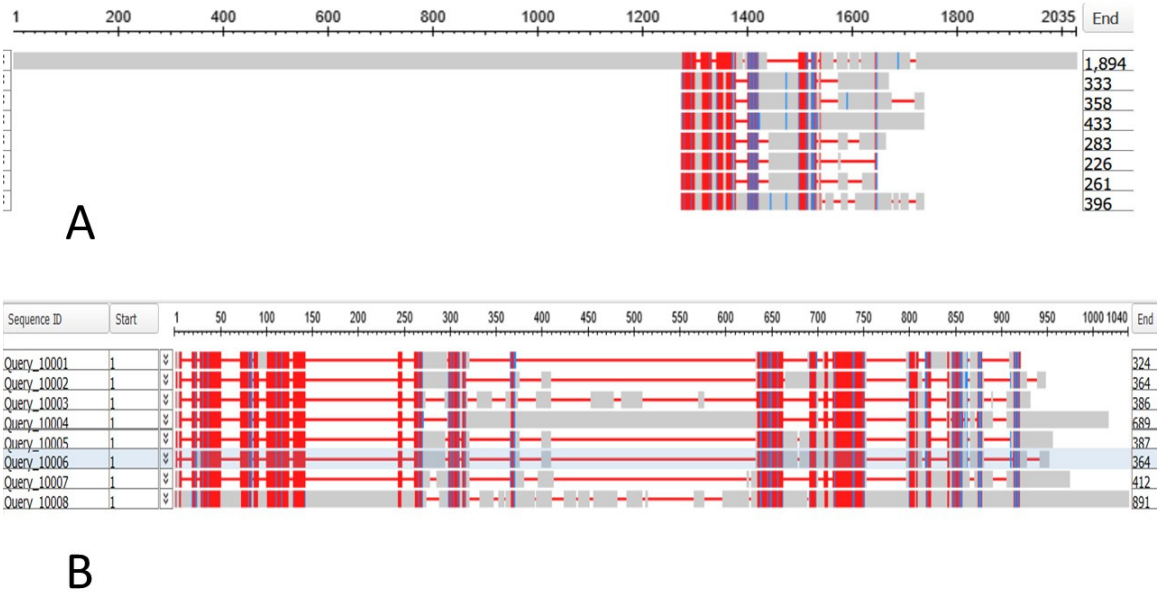


Figure 6. Figure 6A depicts the alignment of XP_019638343.1, the annotation for the sequence of the titin homolog of *Branchiostomata Belcheri* against the seven tunicate Gap Junction genes listed in Table 4; Figure 6B shows the alignment against the 891 portion of the sequence that spans those of the shorter GAP Junction sequences.

 doi:10.20944/preprints202404.1133.v1

Figure 7 shows a phylogram with the 891-long sequence of the Titin homologue and these seven tunicate Gap Junction proteins. The data suggest that a portion of a member of the Branchiostomatidae titins might have evolved to produce the Gap Junction proteins of the Olfactores.

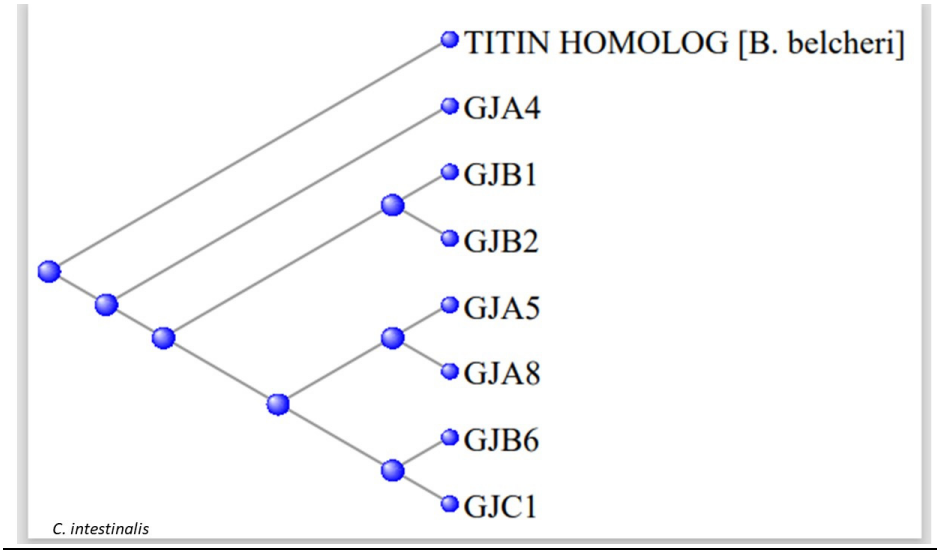


Figure 7. Phylogram of the seven tunicate Gap Junction proteins rooted at the 891-long sequence of the Titin *Branchiostomata belcheri* homolog.

3.3. The Cadherins

The human genome contains 23 cadherin (CDH) genes, with HGNC symbols *CDH1* through *CDH26* (three are absent). Of these 23, seven (*CDH1*, *CDH2*, *CDH7*, *CDH8*, *CDH11*, *CDH16* and *CDH18*) are found as orthologs in the Tunicates while another (*CDH23*) is an ortholog in the Branchiostomatidae. The 23 CDH genes are divided into two types, I and II. These *CDH* genes code for correspondingly-named membrane-bound CDH proteins, each containing multiple cadherin repeats of some 110 amino-acids. The Type II cadherins are missing a long sequence of amino acids in the N-terminal region. Refer to Figures 8 and 9.

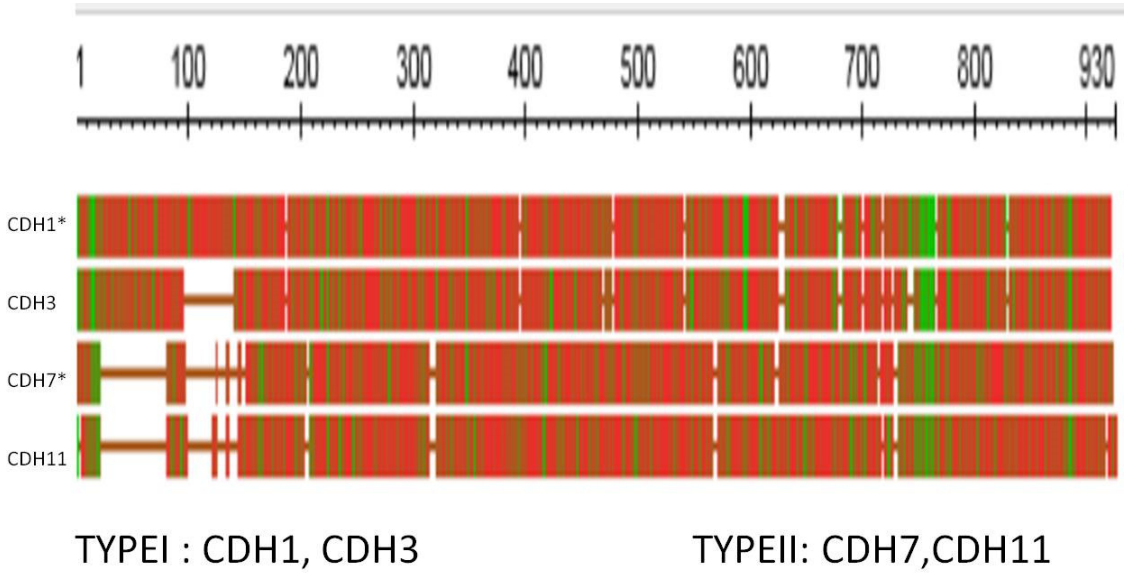


Figure 8. The upper two rows depict cadherins of type I; the lower rows depict those of type II. In each case an ortholog found in the tunicates is marked with an asterisk.

 doi:10.20944/preprints202404.1133.v1


```
P12830.3   1    M--GPWSRSLSALLLLLQVSSWLCQEPEPCHPGFDAESYTFTVPRRHLERGRVLGRVnFEDCTGRQRTAYFSLDTRFKVG  78
P22223.2   1    M--G-LPRGPLASLLLLQVCWLQCAASEPCRAVFREAEVTLEAGGAEQEPGQALGKV-FMGCPGQEPALFSTDNDDFTVR  76
Q9ULB5.2   1    MklGKVEFCHFLQLIALFLCF-------------------------------------------------------------  21
P55287.2   1    M---KENYCLQAALVCLGMLC-------------------------------------------------------------  18

P12830.3   79   TDGVITVKRPLRFHNPQIH[22]GHHH[9]SG[4]LLTFPNSSPGLRRQKRDWVIPPISCPENEKGPFPKNLVQIKSNKD  183
P22223.2   77   NGETVQERRSLKERNP---    ----   --   LKIFP-SKRILRRHKRDWVVAPISVPENGKGPFPQRLNQLKSNKD  136
Q9ULB5.2   22   -SGMSQAELSRSRSKPYF-    ---Q   SG   -----RS-----RTKRSWVWNQFFVLEEYMGSDPLYVGKLHS--D  74
P55287.2   19   -HSHAFAPERRGHLRPSFH    GHHE   KG   -----KEGQVLQRSKRGWVWNQFFVIEEYTGPDPVLVGRLHS--D  80

P12830.3   184  K---EGKVFYSITGQGADTPPVGVFIIERETGWLKVTEPLDRERIATYTLFSHAVSS-NGNAVEDPMEILITVTDQNDNK  259
P22223.2   137  R---DTKIFYSITGPGADSPPEGVFAVEKETGWLLLNKPLDREEIAKYELFGHAVSE-NGASVEDPMNISIIVTDQNDHK  212
Q9ULB5.2   75   VDKGDGSIKYILSGEGASS----IFIIDENTGDIHATKRLDREEQAYYTLRAQALDRLTNKPVEPESEFVIKIQDINDNE  150
P55287.2   81   IDSGDGNIKYILSGEGAGT----IFVIDDKSGNIHATKTLDREERAQYTLMAQAVDRDTNRPLEPPSEFIVKVQDINDNP  156

P12830.3   260  PEFTQEVFKGSVMEGALPGTSVMEVTATDADDDVNTYNAAIAYTILSQDPELPDKNMFTINRNTGVISVVTTGLDRESFP  339
P22223.2   213  PKFTQDTFRGSVLEGVLPGTSVMQVTATDEDDAIYTYNGVVAYSIHSQEPKDPHDLMFTIHRSTGTISVISSGLDREKVP  292
Q9ULB5.2   151  PKFLDGPYTAGVPEMSPVGTSVVQVTATDADDPTYGNSARVVYSILQGQP------YFSVEPKTGVIKTALPNMDREAKD  224
P55287.2   157  PEFLHETYHANVPERSNVGTSVIQVTASDADDPTYGNSAKLVYSILEGQP------YFSVEAQTGIIRTALPNMDREAKE  230
```

Figure 9. As in Figure 8. Expanded first section.

The upper two rows in Figures 8 and 9 depict cadherins of type I; the lower rows depict those of 8type II. In each case an ortholog found in the tunicates is marked with an asterisk. Figure 8 depicts the full sequences while Figure 9 shows the N-terminal sequences expanded. A cadherin from *Homo sapiens* and its tunicate ortholog show very close alignment. The green-shaded region of the sequences in Figure 8 marks the membrane-embedded portion of the proteins.

A phylogram of all the cadherins from *Homo sapiens* is presented in Figure 10.

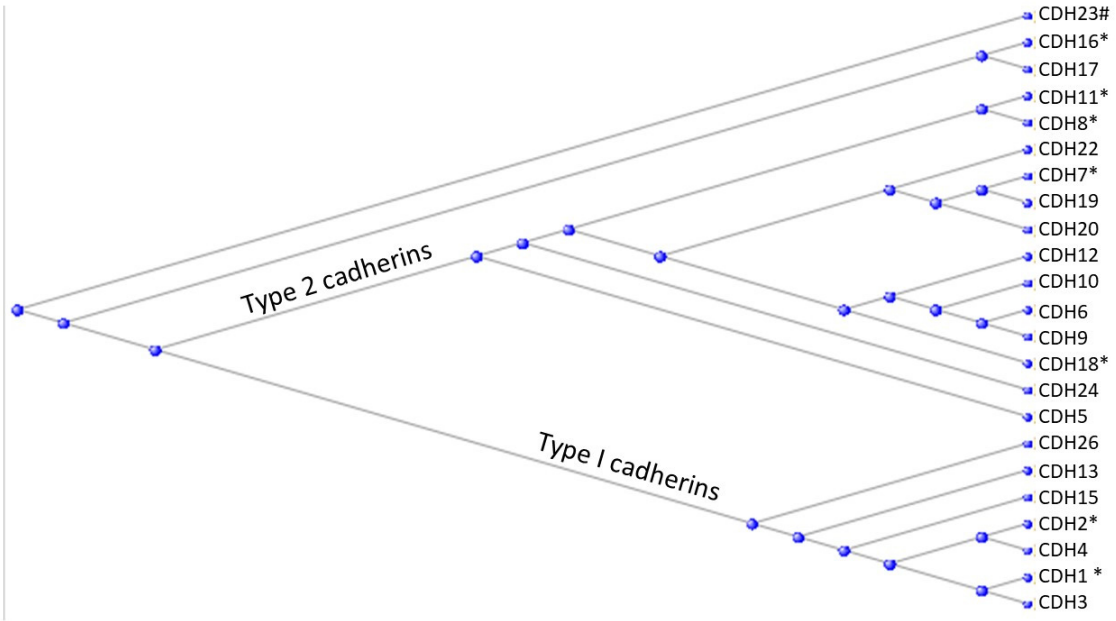


Figure 10. LEGEND: Phylogram of the cadherin proteins of *Homo sapiens*. Those with tunicate orthologs are marked with an asterisk *. The single cadherin with an ortholog in the Branchiostomatidae is marked with a hashtag #,.

As can be seen, each of the two cadherin types contains proteins (asterisked) which have orthologs in the tunicates and might perhaps be considered as founder members of that type. The single cadherin ortholog that derives from the Branchiostomatidae (marked with a hashtag #) lies separate from the cadherins in the two types I and II.

 doi:10.20944/preprints202404.1133.v1

The cadherins are well-named, being proteins that cause cells to adhere to one another under the influence of calcium (Ca). An x-ray crystallographic study of cadherins (Patel] et al., (2005) showed that it was largely the N-terminal sequences of the molecules that determined the binding between two cadherins [18]. The N-terminal sequence of one molecule formed an anchor that fitted into a pocket in the second molecule, whose N-terminal sequence in turn inserted into a pocket on the first molecule. The process is called "swapping" of the two N-terminal sequences, although of course no chemical bonds are broken in this swapping. The different structures of the anchor (see the N-terminal sequences of the two types of cadherins in Figure CAD2) and pocket between type I and type II cadherins render the two types incompatible with each other. This incompatibility is of profound significance in the embryological development. Cells from an epithelial sheet of cells held together by cadherins of one type (say Type I) will detach from the sheet if they are temporally programmed to express cadherins of Type II. This process, accompanied by the programmed lowering of expression of Type I cells, is of crucial significance in the formation and detachment of the neural crest [19].

Table 4 lists some frequently-studied cadherins with the commonly used names that they are referred to in the literature, the tissue in the body where each protein is usually found, and the HGNC symbol for each protein.

Table 4. Names and tissue distributions of frequently-studied Cadherins.

Familiar name	Tissue distribution	HGNC	Type	Has Tunicate ortholog ?
Cadherin E	Epithelial	CDH1	I	Yes
Cadherin N	Neural	CDH2	I	Yes
Cadherin P	Placental	CDH3	I	
Cadherin R	Retinal	CDH4	I	
Cadherin VE	Epithelial	CDH5	II	
Cadherin K	Brain; kidney	CDH6	II	
Cadherin OB	Osteoblast	CDH11	II	Yes
Cadherin BR	Brain	CDH12	II	
Cadherin T and H	Heart	CDH13	I	
Cadherin M	Muscle	CDH15	I	

Cells bearing cadherins of one name will readily form homotopic interactions but less frequently heterotopic combinations with other cadherins, and this leads to sorting-out of different cell types in embryological development.

In the animal body, the cadherins exert their adhesion function in many epithelial tissues but also, importantly, in neural tissue and the formation and detachment of the neural crest. Cells from the neural crest migrate throughout the embryo, giving rise to an array of cell types that characterize the vertebrate clade, including the peripheral sensory nervous system and most of the craniofacial skeleton.

In the cephalochordate Amphioxus, two cadherin paralogs are found [20], one is expressed in the mesoderm and the other in the ectoderm, but both are found expressed in neural tissue, at different developmental periods. These two paralogs are, of course, of the same type and can only demonstrate adhesion and not repulsion. In Amphioxus there are no cells that have been identified as homologous to neural crest [21]. The tunicates possess cadherins of both types so that both cohesive and repulsive interactions are now possible. In the tunicates [22], a CDH5-like protein is expressed at the tailbud stage in the nerve cord and in the peripheral neurons of the tail, while CDH7-like is widely expressed in the epidermis and also in the nerve cord, the sensory vesicle and the visceral ganglion. Neural crest-like cells are indeed found in the tunicates as migratory cells that produce pigment (as do vertebrate cells of neural crest origin) and in addition express many of the genes that regulate the development of vertebrate neural crest [23].

 doi:10.20944/preprints202404.1133.v1

The role of the cadherins expanded greatly as the number of cadherins themselves expanded and diversified during the evolution of the vertebrates, with multiple roles in embryological morphogenesis [24].

An additional set of proteins involved in cell adhesion and in particular in cortical development and synapse formation [25] are the FLRTs (Fibronectin Leucine Rich Transmembrane proteins) of which FLRT2 and FLRT3 were tunicate innovations and which, in their developmental role, interact with latrophilin (HGNC symbol ADGRL3), also first found in the tunicates. Latrophilin and the FLRTs also interact with teneurin proteins (the TNFMs) to form trimeric complexes, important during synapse formation and in guiding the migration of young neurons [26].

3.4. The Claudins

The human genome contains 24 claudin genes symbolised by *CLDN*, enumerated up to *CLDN34* with some absent members. These genes, of course, code for the correspondingly named proteins. The origin of the world claudin [27] is from the Latin *claudere* meaning "to close", very appropriate since their function is to form the Tight Junctions between adjacent cells in an epithelium or endothelium [28]. These tight junctions ensure that passage through them, from one side of an epithelium to the other, is tightly controlled in a selective permeability, with the different claudins having different selectivities [28]. Orthologs of the claudins are found in the tunicates (three of them) while another is found in the earlier-appearing Branchiostomatidae, but not in the even earlier Echinodermta. Figure 11 depicts a phylogram of the human claudin proteins with the three tunicate-shared orthologs marked with an asterisk*and the single Branchiostomatidae-shared ortholog marked with the hashtag #.

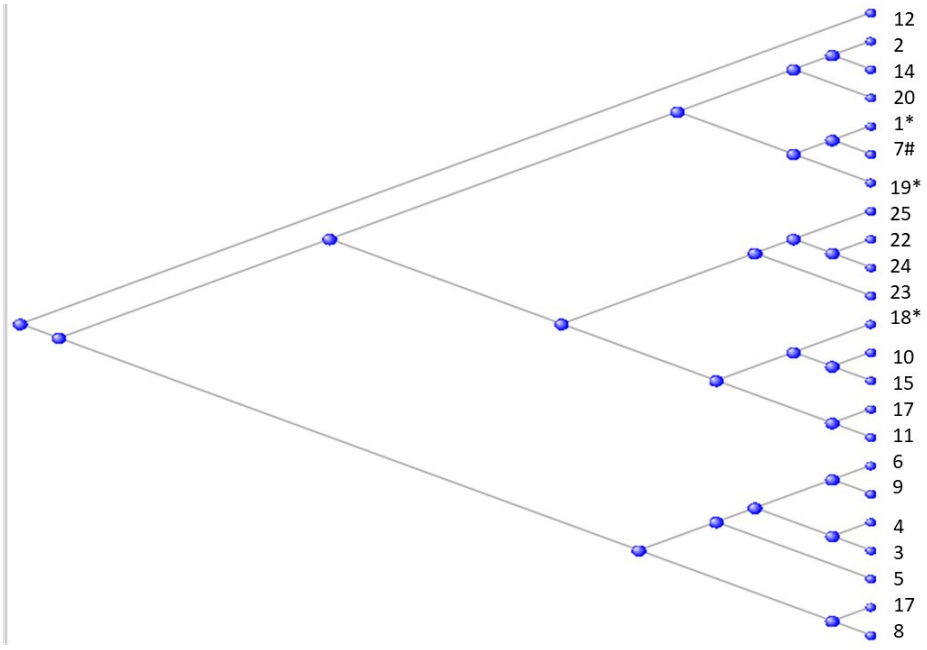


Figure 11. Phylogram of the human claudin proteins with the three tunicate-shared orthologs marked with an asterisk*and the single Branchiostomatidae-shared ortholog marked with the hashtag #.

Figure 12 shows just a portion of the phylogram rerooted from CLDN7, the Branchiostomatidae-shared member. On this figure the lamprey-shared sequence is marked with a dollar sign $ while two sequences present in the sharks are marked with the ampersand &. The figure suggests how *CLDN7* could have been the founder of the tunicate-shared sequences and of the sequences found in the fish. Figure 13 shows a comparison of the sequence of the CLDN18 orthologs from the tunicate *Ciona intestinalis* and the lamprey *Petromyzin mainus*, that of the lamprey being shown as the lower sequence of the two in each row. The two sequences, analysed in a BLAST one against one comparison returned an Expect Value of $2E^{-09}$.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2024 doi:10.20944/preprints202404.1133.v1

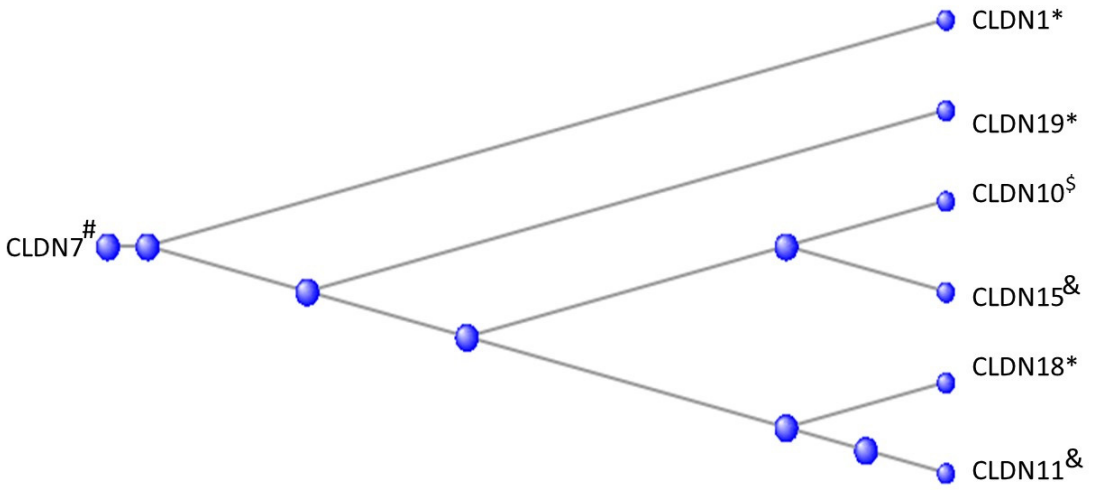


Figure 12. A portion of the phylogram of Figure 10, rerooted from CLDN7, the Branchiostomatidae-shared member. The lamprey-shared sequence is marked with a dollar sign $ while two sequences present in the sharks are marked with the ampersand &.

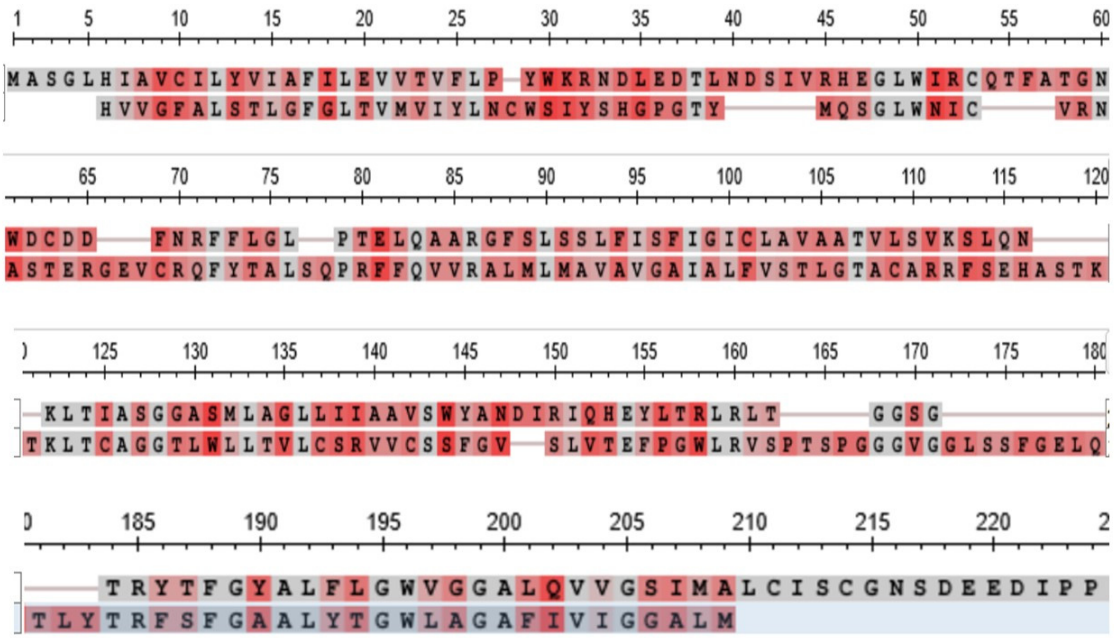


Figure 13. Comparison of the sequence of the CLDN18 orthologs from the tunicate *Ciona intestinalis* and the lamprey *Petromyzin mainus.* The sequnce of the lamprey ortholog is shown as the lower sequence of the two in each row.

Defects in claudin genes are associated with congenital deafness in children. Studying the expression of claudin genes in the Zebrafish, Kollmar] et al., (2001) found claudins in the otic and lateral-line placodes of the fish [29]. They suggested, on the basis of their findings, that claudins could have an additional role in vertebrate morphogenesis. Perhaps the role of the early claudins in tunicates might relate to a role in tunicate morphogenesis in addition to their function as aiding to form barriers between the external and internal environments, as indicated by electron microscopy studies.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2024 doi:10.20944/preprints202404.1133.v1

3.5. The Ephrons and Ephrins

The human genome contains 14 ephron genes, symbolised by *EPHA1* to *EPHA10* (*EPHA9* is absent) and *EPHB* to *EPHB4* [30]. These code for corresponding membrane-bound proteins. Table 5 includes these genes together with the animal species where their latest shared orthologs appear.

Table 5. Latest orthologs of Ephron and Ephrin genes.

Ephron Genes		Ephrin Genes	
HGNC symbol	Latest ortholog in:	HGNC symbol	Latest ortholog in:
EPHA2	Tunicata	EFNA1	Tunicata
EPHA3	Elasmobranchii	EFNA2	Tunicata
EPHA4	Tunicata	EFNA3	Tunicata
EPHA5	Branchiostomata	EFNA4	Tunicata
EPHA6	Elasmobranchii	EFNA5	Tunicata
EPHA7	Elasmobranchii	EFNB1	Tunicata
EPHA8	Elasmobranchii	EFNB2	Branchiostomato
EPHA10	Osteichthyes	EFNB3	Elasmobranchii
EPHB2	Branchiostomata		
EPHB3	Elasmobranchii		
EPHB4	Cnidaria		

Figure 14A displays a phylogram of the relation between the genes of the A series, the phylogram being rooted at EPHA5, while Figure 14B displays the corresponding phylogram for the genes of the B series, the phylogram being rooted at EPHB2. The ephron proteins are membrane-bound receptors for the 8 ligand Ephrin proteins, symbolized by EFNA1 through to EFNA5 and EFNB1 through EFNB3, coded for by the corresponding *EFN* genes [30].

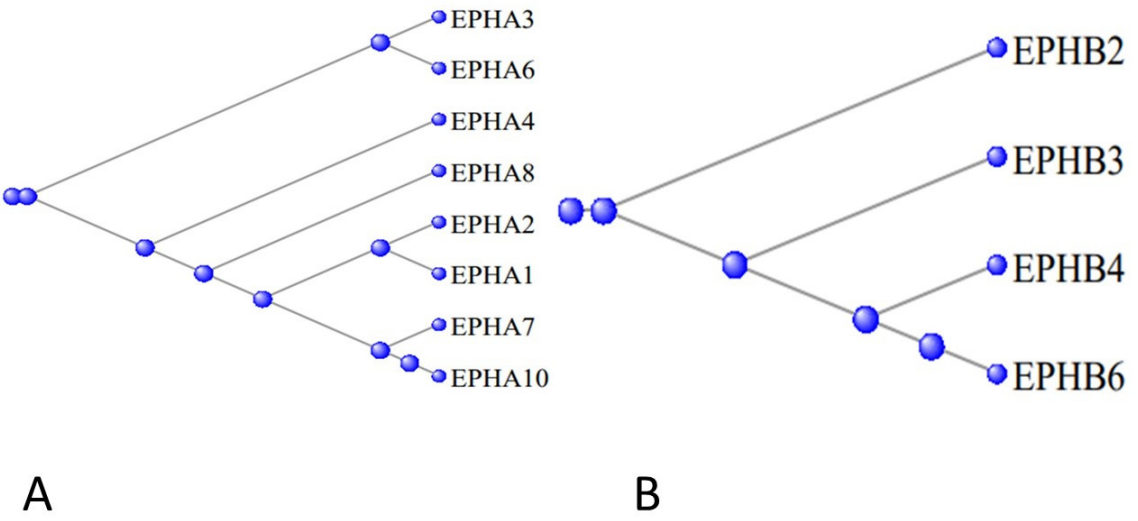


Figure 14. A - Phylogram of the relation between the genes of the A series, the phylogram being rooted at EPHA5. B_ the corresponding phylogram for the genes of the B series, the phylogram being rooted at EPHB2.

Table 5 includes also these genes together with the animal species where their latest shared orthologs appear.

Figure 15 displays a phylogram of the relation between these genes, the A and B series being clearly separated into their two subgroups.

 doi:10.20944/preprints202404.1133.v1

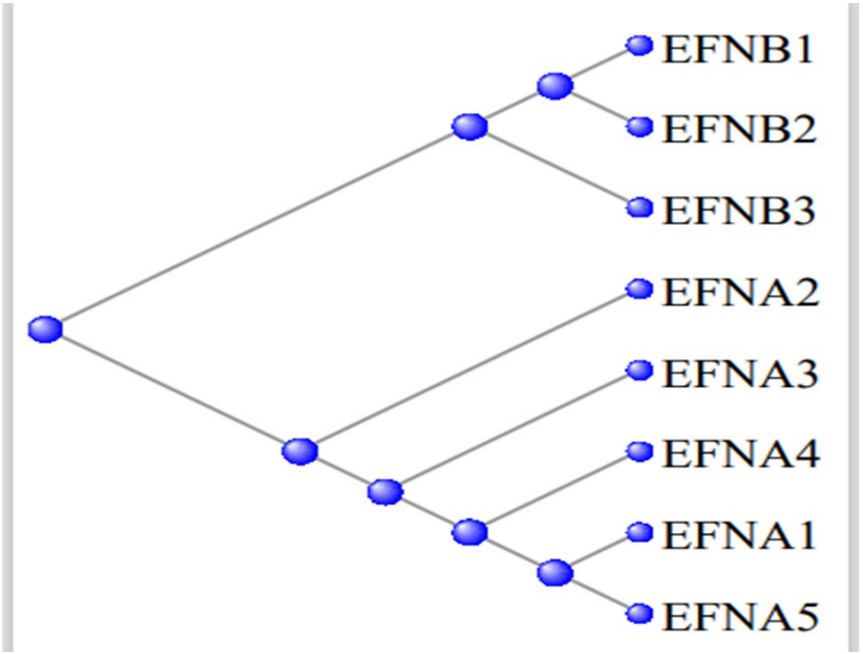


Figure 15. Phylogram of the relation between the Ephrin genes - the A and B series separate into their two subgroups.

Since both the ephrons and the ephrins are membrane-bound, their interactions are only cell to cell The proteins of the A series of the receptor ephrons bind to the proteins of the A series of the ligand ephrins and repel the proteins of the B series of the ephrins. B series ephrons bind B series ephrins and repel those of the A series. These attractive and repulsive interactions between cells determine many morphogenetic processes in animal development [31]. Repulsive effects determine the boundaries between different cell types. These repulsive responses are likely to involve depolymerization of the actin cytoskeleton, leading to the collapse of the filopedia and retraction of the cells involved. Actin-cytoskeletal interactions are the probable base of the attractive responses. A combination of repulsive and attractive effects guides the movement of nerve cells during the embryonic formation of the brain and ensures appropriate movement of cells in the intestinal crypt, retaining the Paneth at the base of the crypt while guiding the movement of cells of the crypt towards the lumen of the intestine.

The ancestor of the Olfactores contributed none of the 14 Ephron genes but as many as 4 of the 8 Ephrin genes, a substantial contribution. Two Ephron genes (*EPHA4* and *EPHA2*) were already present in the Cnidarians as well as one ephrin gene (*EFNB1*). In the tunicates themselves, these genes are involved in (among other processes) neural induction [32], endoderm invagination [33], asymmetric cell divisions in the vegetal hemisphere of the developing brain [34], delimiting the number of pigment cells in the central nervous system [35] and neural tube patterning [36].

3.6. The MAGE Genes

There are some 40 *MAGE* genes (the name derives from Melanoma Antigen Gene) in the human genome, divided into two subfamilies, Type I and Type II, according to their sequence and chromosome location [37]. The Type II family can again be subdivided according to their ancestry – those originating from an ortholog in the tunicates and being mostly named in the form *MAGEDx* and those originating from an ortholog in the branchiostomata, most being named *MAGEEx*. The phylogram in Figure 16 shows how the Type II MAGE proteins divide into these two subfamilies, together with a separate protein NDN, necdin, also a member of the MAGE family.

 doi:10.20944/preprints202404.1133.v1

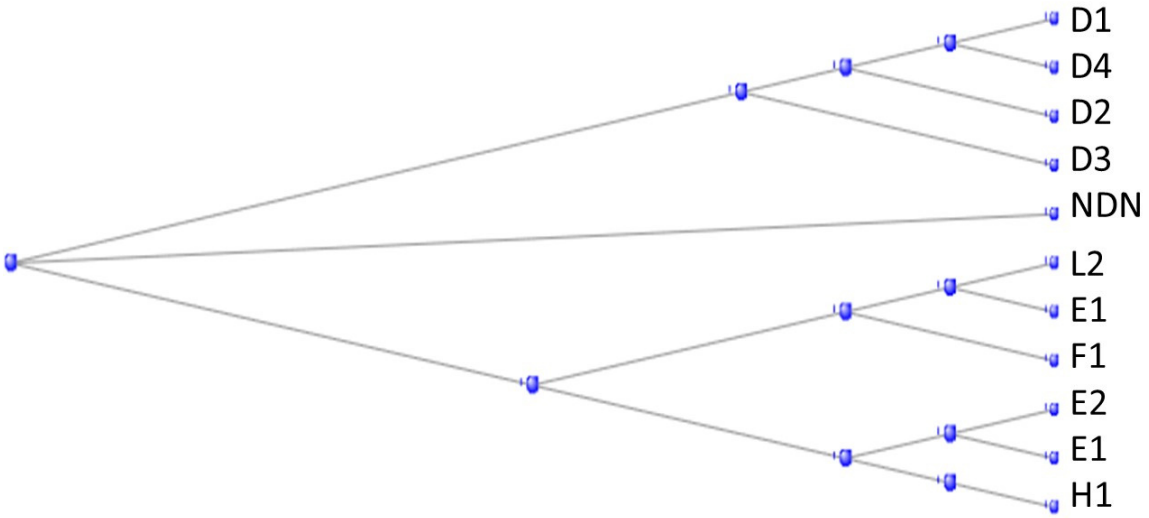


Figure 16. Phylogram of the Type II MAGE proteins, together with a separate protein NDN, necdin, also a member of the MAGE family.

Figure 17 shows a phylogram of the proteins coded by genes descended from the tunicate gene Cirobu.g00007177 (also known as "melanoma-associated antigen D2 isoform X2 [*Ciona intestinalis*]"),

Figure 17. Phylogram of the D series of MAGE proteins rooted at the tunicate gene Cirobu.g00007177.

Figure 18 shows the corresponding phylogram for those descended from the Branchiostoma gene

 doi:10.20944/preprints202404.1133.v1

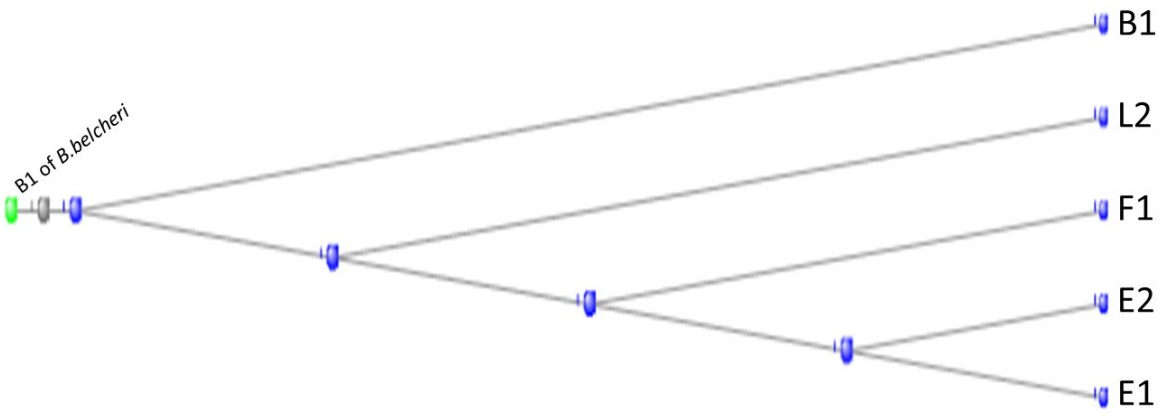


Figure 18. Phylogram of the E series of MAGE proteins rooted at the B1 ortholog of *Branchiostomata bbelcheri*.

The two subfamilies of Fig. MAG1 correspond with the tunicate-descended and the branchiostomata-descended groups of Figures 17 and 18, respectively.

The biological roles of the MAGE proteins are tabulated in Table S1.

3.7. The Crystallins

The Gamma Crystallins

The human genome contains 18 crystallin proteins. They are divided by sequence similarity into three subgroups, the α, β and γ crystallins and, interacting together, they form the transparent lens of the eye. The six gamma crystallins are descended from a single ortholog in the Tunicates, namely, the gene designated in the ANISEED database (see Methods) with the unique ID of Cirobu.g00014781. This gene can appear in BLAST searches of the Tunicates as one of three annotated sequences, two from *Ciona intestinalis*: XP_002126888.1 (named as gamma-crystallin S), and 2BV2_A (Named as Chain A, Ciona betagamma-crystallin) and one from *Styela clava*, XP_039266482.1 (named as gamma-crystallin N-A-like). In what follows, the sequence 2BV2_A has been used. The HGNC symbols for the six gamma crystallins of Homo sapiens are CRYGA, CRYGB, CRYGC, CRYGD, CRYGN, and CRYGS. (The three crystallin genes *CRYBG1*, *CRYBG2* and *CRYBG3*, coding for proteins that also present in the vertebrate lens, but some ten-fold longer than the gamma crystallins, will not be discussed here. Supplementary Figure S1, part A depicts a COBALT-based alignment of these proteins compared with the six gamma crystallins and the single tunicate crystallin, while part B is a phylogram from these data. The gamma crystallins and the CRYBG proteins clearly separate into two families. The *CRYBG* genes are not found in the genome of the tunicates, although *CRYBG2* is present in the Branchiostomatidae. These genes have curiously-named synonyms of the form *AIMx* where *AIM* stands for Absent In Melanoma [38].

Figure 19 depicts a COBALT-derived comparison of the six gamma crystallins together with the Ciona crystallin.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 April 2024 doi:10.20944/preprints202404.1133.v1

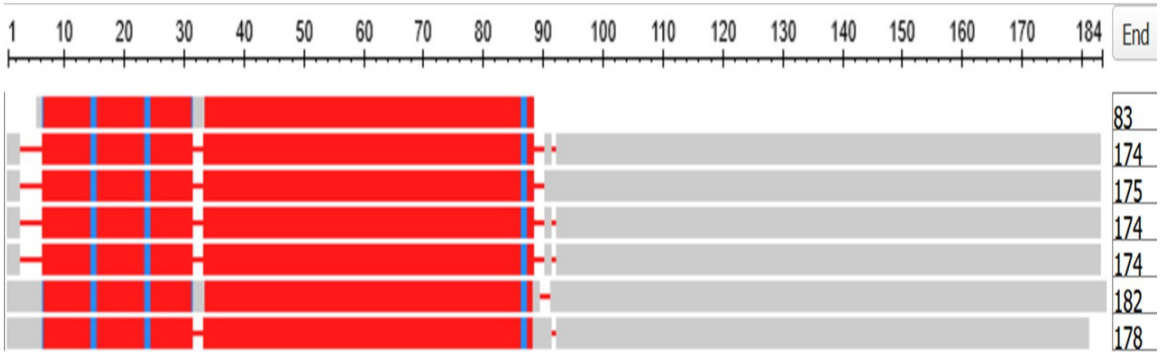


Figure 19. Multiple alignment of the six gamma crystallins, compared with the single crystallin of the tunicates. The figure was generated by the COBALT program of the National Library of Medicine (see Methods). From top to bottom, the proteins depicted are.

Chain A, betagamma-crystallin of *Ciona intestinalis*, CRYGA, CRYGB, CRYGC, CRYGD, CRYGN, and CRYGS of *Homo sapiens*. The length of each sequence in amino-acids is listed under the heading END.

Figure 20 depicts the Phylogram that COBALT produced from the data, here rooted at the Ciona crystallin

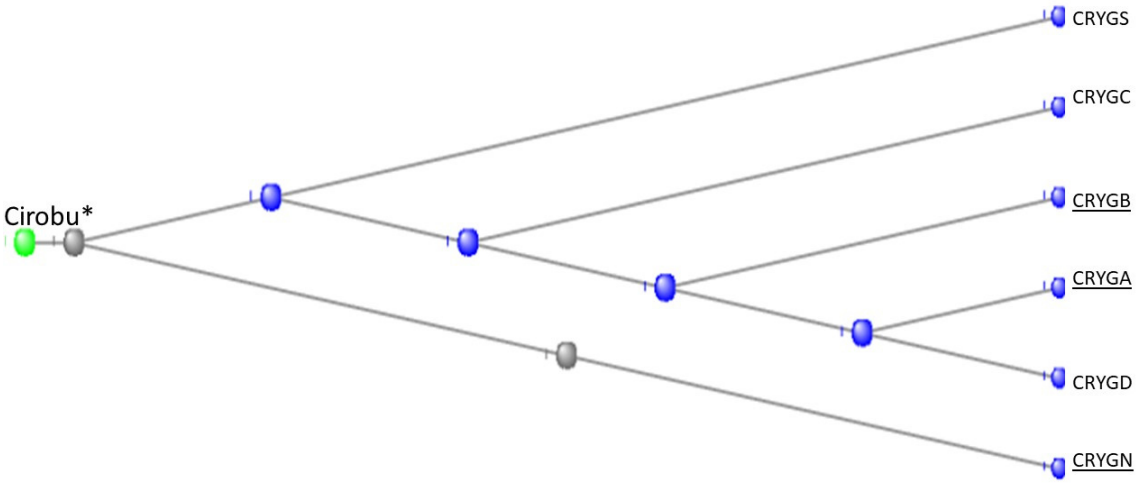
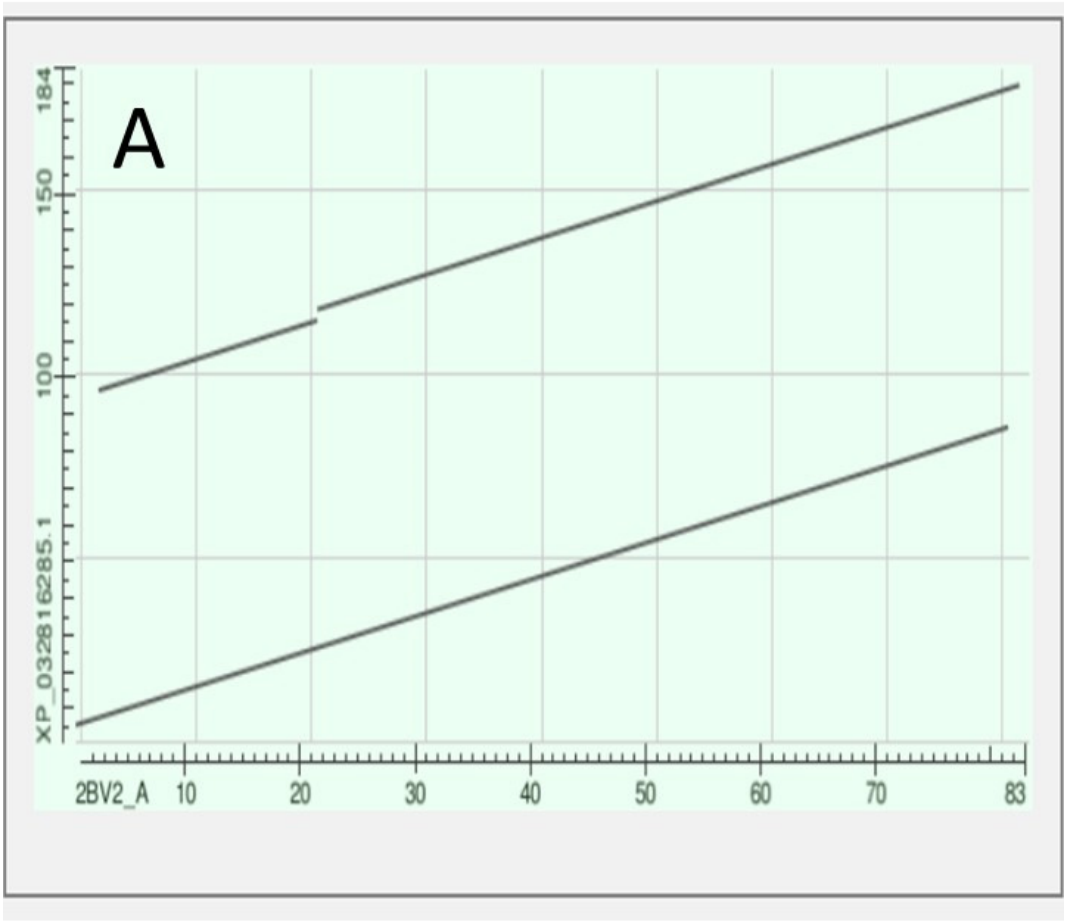
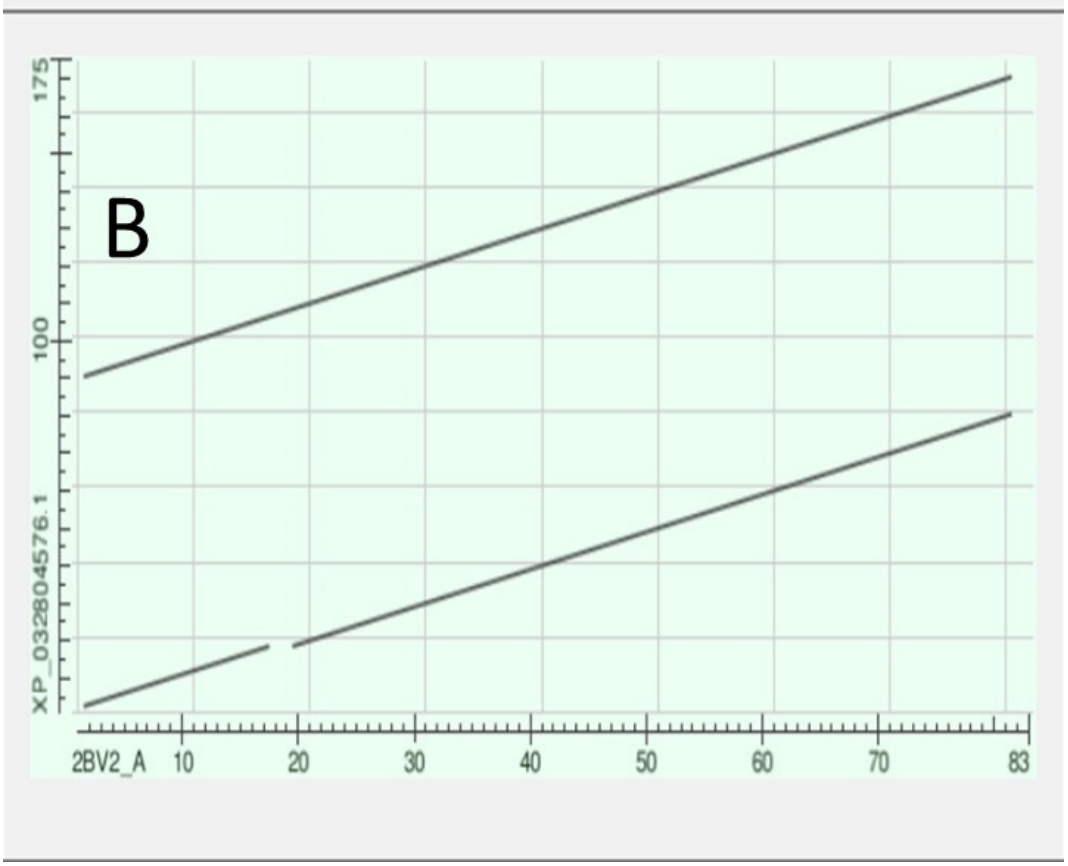


Figure 20. Phylogram of the six gamma crystallins together with the Tunicate crystallin. The figure was generated by the COBALT program of the National Library of Medicine (see Methods), being rooted at the Tunicate crystallin. (see also [39–43], for fuller discussions of the evolution and biology of the lens and of the crystallin proteins).

Note in Figure 19 that the tunicate crystallin is almost exactly half the length of the crystallins of *H. sapiens*, suggesting that the tunicate gene was doubled as it evolved into its vertebrate descendants. This suggestion is confirmed if the sequences of the tunicate crystallin and two vertebrate crystallins are compared as in the dotplots of Figure 21:

 doi:10.20944/preprints202404.1133.v1





 doi:10.20944/preprints202404.1133.v1

Figure 21. Dotplots produced by the BLAST program (see Methods) with the tunicate sequence on the x-axis and, on the y-axis, in A the CRYGN and in B the CRYGS of the lamprey *Petromyzon marinus*.

The 83 residue sequence of the tunicate protein is repeated in the second half of both vertebrate proteins with a minor modification: a small break in the second half of the sequence of CRYGN (part A of the figure) and in the first half of the sequence for CRYGS (part B). This difference between the two plots suggests that the two vertebrate proteins arose by two independent gene duplications, compatible with the positions of these two vertebrate proteins in the phylogram of Figure 20. It would appear from the phylogram that the further speciation of the crystallins originated with CRYGS. These two crystallin proteins, CRYGN and CRYGS, are both found in our table of the orthologs at the base of the Olfactores, since they appear to have evolved independently.

It was mentioned at the beginning of this section that in vertebrates the crystallins are found in the lens of the eye. In a tunicate, however, the crystallin protein is not found where one might expect it to be: in the ocellus or eyespot of the organism. Rather the crystallin is found in the palp, an organ in the snout of the larva of the sessile tunicates whose function is to provide, together with lectins and other components, a mucilage that fixes the tunicate to the sea bed [44]. That the tunicate crystallin is not expressed in the organism's eyespot suggested to Horie et al., 2008: "that the lens of the Ciona pigmented ocellus is not homologous with that of the vertebrate eye" [45].

The tunicate crystallin can bind calcium [46], thereby stabilising the protein against temperature denaturation. The vertebrate crystallins have lost these calcium-binding sites. Figure 22 shows a comparison of part of the amino-acid sequence of the lamprey's CRYGN (upper row) and the tunicate crystallin (lower row). The wide arrows point to the sites at which calcium binds in the tunicate protein [46].

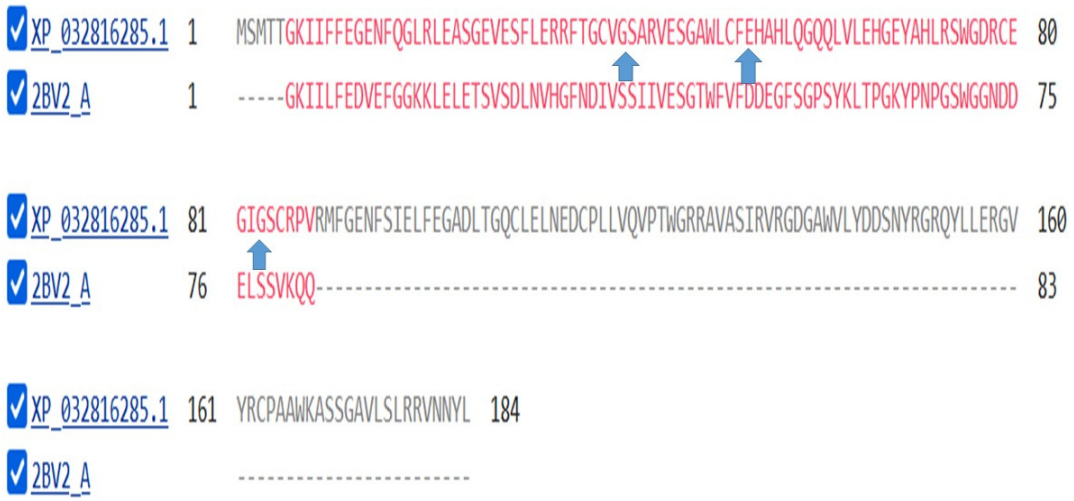


Figure 22. Alignment of the amino-acid sequences of the lamprey CRYGN protein and the tunicate crystallin. The wide arrows show the mutations that occurred between the lamprey and tunicate in the sites that bind calcium in the latter organism (sequence data from [46]).

These mutations and the duplication of the tunicate's sequence led to the dramatic evolution of the protein from a mucilage in the tunicates into the vertebrate proteins that confer upon the lens its transparency.

3.8. The Distal-Less Genes

The human genome contains six members of the distal-less family with HGNC symbols *DLX1* through *DLX6*. Of these, *DLX1*, *DLX4*, and *DLX6* first appear in the Cnidaria, with *DLX 1* and *4*

 doi:10.20944/preprints202404.1133.v1

present in corals and *DLX6* in a sea anemone. *DLX5* first appears in the Echinodermata, in a sea urchin, while *DLX2* and *DLX3* were the contribution of the tunicates (Figure 23)

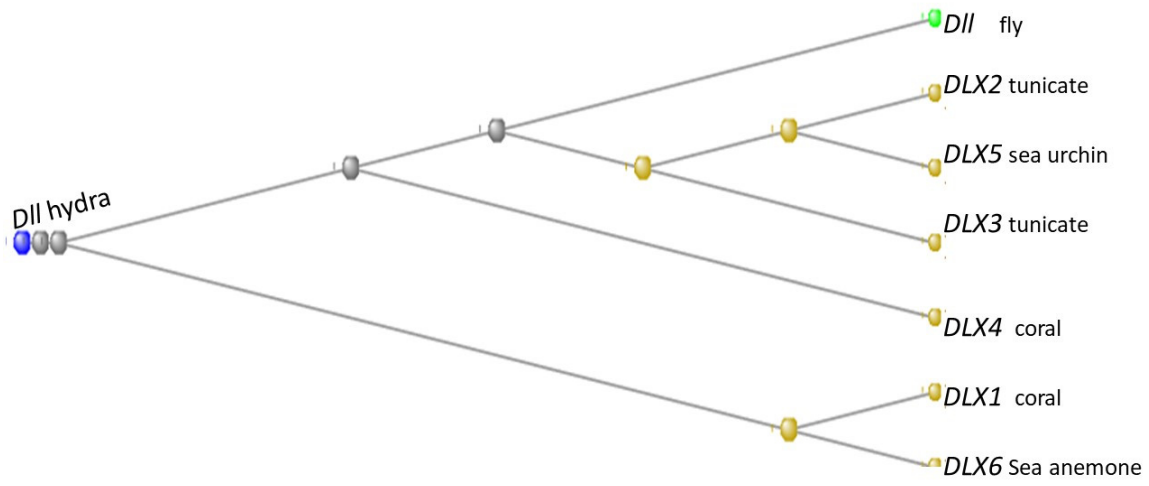


Figure 23. Phylogram of the six *DLX* genes of the human genome together with the gene Dll of *Drosophila melanogaster*, rooted at the *Dll* gene of *Hydra vulgaris*. The organisms in which the orthologs with the human gene first appeared is indicated next to each gene. The phylogram was built using the COBALT program (see Methods).

A multiple alignment of these genes is shown as Figure 24:

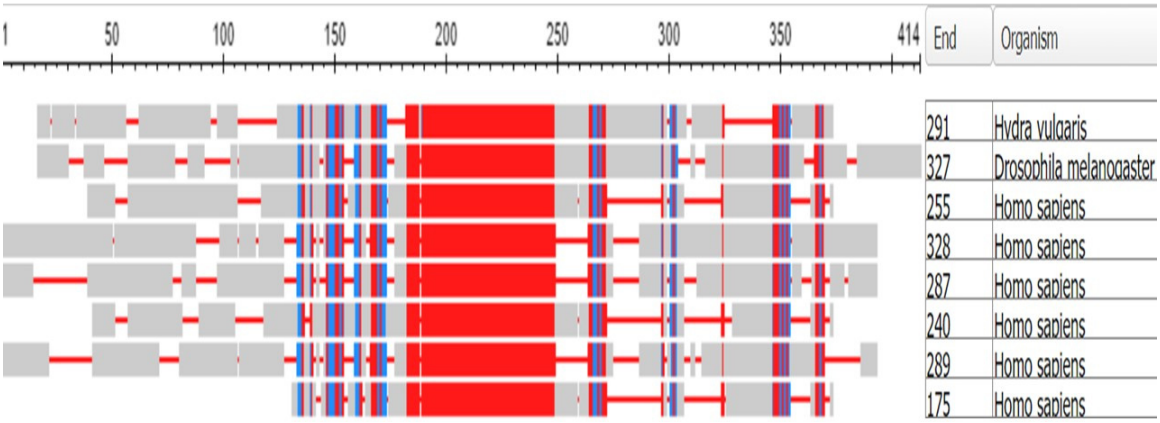


Figure 24. Multiple alignment of the six DLX proteins of the human genome together with the gene Dll of *Drosophila melanogaster*. The rows from the top: Dll of *Hydra*, Dll of the fly, and then DLX 2,5,3,1,6,and 4 of *Homo sapiens*. The alignment was made using the COBALT program (see Methods). The number of amino-acid residues in each protein is listed under "End".

A member of the gene family was first noticed in the fruit fly, *Drosophila melanogaster*, as the gene *Dll* when a mutant lacking the distal ends of the limbs was found (reviewed in [47]), and the gene was postulated to control pattern formation along the proximal-distal axis of the limbs. A similar role

 doi:10.20944/preprints202404.1133.v1

in controlling pattern formation along the limb proximal-distal axis is found in the Crustacea [48]. The phenomenon of an absence of distal regions of the limbs is also found in mice, where targeted inactivation of the genes *DLX5* and *DLX6* led to severe malformation of the distal portion of the limbs. The human disease Split-hand/split-foot malformation (SHFM) is a human limb defect characterized by missing digits and fusion of remaining digits. Ullah et al., (2016) showed that a mutation in *DLX6* is associated with the type I form of SHFM [49].

In the mice, disruption of *DLX* genes led to gross malformation of the craniofacial skeleton. These craniofacial defects, associated with distal-less gene mutations in the mouse, suggest that distal-less controls more than simply proximal-distal pattern information. Indeed, *Dll* in its first appearance in Hydra (where of course no proximal- distal axis is present) is expressed in the head, battery, and the stem-like peduncle [50]. [In Hydra, potent stinging cells are grouped with other neurons in what are called "battery complexes" on the hydra's tentacles]. Indeed, the mutation in the Drosophila *Dll* gene is associated also with defects in the antenna of the fly, a sensory organ. In the sea urchin (which too has no proximal-distal axis), ectodermal expression of distal-less is evident at the tips of spines but is not associated with other skeletal elements [51]. In the myriapod *Glomeris marginata,* the distal-less gene is expressed in the limb appendages but also, and to a large extent, in the mouth parts of the organism [52]. In the tunicate *Ciona intestinalis,* DLX3 is expressed in the atrial siphon and in addition in the adhesive organ, which is located in the head, at the anterior end of the body, and is used by the larvae at the beginning of metamorphosis to attach to a solid substrate [53]. Finally, in Amphioxus (the lancelet *Branchiostoma floridae)*, a context where the proximal-distal axis is again not relevant, Holland] et al., (1996) [54] showed that the animal's single *Dll* gene is expressed in the anterior three/fourths of the cerebral vesicle and implicated in the "establishment of the dorsoventral axis, specification of migratory epidermal cells early in neurulation and the specification of forebrain". They add the intriguing suggestion that: "Such a multiplicity of Distal-less functions probably represents an ancestral chordate condition and, during craniate evolution, when this gene diversified into a family of six members, the original functions evidently tended to be parcelled out among the descendant genes". Interestingly, in the parallel bilaterian evolutionary path to the insects, as we saw in the case of the Myriapod, the ancestral DLX gene again took on the function of regulating pattern formation in the proximal-distal axis and as well as specifying head development and aspects of sensory perception. Thus, the *DLX* genes have evolved to be interpreters that express their function by being directed to various locations in the body. In accordance with signalling systems present at this location, the DLX genes drive and direct the expression of genes that act together to form the particular organ or organs there found.

3.9. DAVID Analysis of Individual Orthologs Not Discussed in Previous Subsections.

In the list of the 84 tunicate orthologs assembled in Table S1, 36 have not yet been covered in the text so far. To tease out what might be their role in the tunicate lifestyle and perhaps in that of the vertebrates as well, we subjected the 36 to an analysis using the DAVID program (see Methods). As it describes itself, DAVID is "The Database for Annotation, Visualization and Integrated Discovery (DAVID) [that] provides a comprehensive set of functional annotation tools for investigators to understand the biological meaning behind large lists of genes." We applied the functional annotation clustering tool of DAVID (that attempts to find common themes in groups of genes selected from a submitted list). Table 6 that follows is extracted from the full output of the DAVID clustering.

Table 6. Abridgement of DAVID cluster analysis of 36 genes not discussed in main text.

	Term	Genes	FDR*
Cluster 2			
	heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules	*NECTIN3, CBLN1, NECTIN1*	0.358759

 doi:10.20944/preprints202404.1133.v1

	identical protein binding	*TFAP2E, COL23A1, HSPB1, NECTIN3, CBLN1, NECTIN1*	0.671885
Cluster 4	Synapse	*NECTIN3, CBLN1, NECTIN1*	0.935728
	homophilic cell adhesion via plasma membrane adhesion molecules	*PALLD, NECTIN3, NECTIN1*	0.739434
	DOMAIN:Ig-like C2-type 1	*PALLD, NECTIN3, NECTIN1*	0.656496
	DOMAIN:Ig-like C2-type 2	*PALLD, NECTIN3, NECTIN1*	0.656496
	Immunoglobulin domain	*PALLD, NECTIN3, NECTIN1*	0.877976
	Immunoglobulin-like fold	*PALLD, NECTIN3, NECTIN1*	0.984375
Cluster 6			
	Z disc	*PALLD, SYNPO2, HSPB1*	0.481232
	actin cytoskeleton	*CALD1, PALLD, SYNPO2*	0.564483
	actin binding	*CALD1, PALLD, SYNPO2*	0.614614
	focal adhesion	*PALLD, SYNPO2, HSPB1*	0.679141
Cluster 7			
	Signaling pathways regulating pluripotency of stem cells	*FZD3, ZIC3, FZD6*	0.180791
	Developmental protein	*FZD3, ZIC3, FZD6, PITX3*	0.819281
Cluster 8			
	Zinc	*ZMAT1, ZIC3, ASXL3*	1
	Zinc-finger	*ZMAT1, ZIC3, ASXL3*	0.988184
	Metal-binding	*ZMAT1, ZIC3, ASXL3*	1

*False Discovery Rate (maximum is 1).

Of these listed clusters, cluster 2 and cluster 4 continue the emphasis on adhesion between cells that was discussed previously in the sections on the cadherins and the ephrons and ephrins, while cluster 6 extends the list of genes concerned with muscular contraction. Cluster 7 contains genes concerned with embryological development while cluster 8 concerns the important zinc- finger genes with their widespread roles in gene regulation.

Information abstracted from the summaries in the GeneCards database (see Methods) is listed for each of these 36 genes in Table S1, together with all the orthologs that are assembled in that table.

4. Discussion and Conclusions

In the list that partitioned the 19,653 protein-coding genes of the human gene into their appropriate phylostratum level, [Litman] and Stein (2018) found, as stated earlier, that 84 of these genes fell into phylostratum 10, associated with the tunicates [4]. Thus these tunicate-shared genes, which stand at the base of the Olfactores clade, are a very minor fraction of the genes of the human genome. This tunicate contribution was preceded by a large contribution of genes from the Branchiostomatidae, 406, and was followed by an even larger number from the fishes, 3650. The hand-curated list of phylostratum 10 genes that are listed in Table S1 of this paper again contains only 84 genes. That there are orthologs that are found shared with the vertebrates in the tunicates, and yet are not present in the Branchiostomatidae, accords with the suggestion of [2] Delsuc], 2008 that the tunicates and vertebrates are sister clades of the Olfactores and succeeded the Branchiostomatidae in evolutionary time.

We were intrigued that this small number of 84 genes at the base of the Olfactores could open the vast evolutionary changes that led, on the one hand, to the tunicates themselves and on the other, to the vertebrates.

 doi:10.20944/preprints202404.1133.v1

Perhaps the most important contribution of the tunicates was the evolution of a second type of cadherin. This, a Type II cadherin, had the property of detaching the cell containing that cadherin from cells that expressed the Type I class. The set of such Type II cadherins could now detach and move away from their Type I neighbours, a process which would eventually evolve into the formation of the neural crest. Cells from the neural crest migrate throughout the embryo giving rise to an array of cell types that characterize the vertebrate clade, including the peripheral sensory nervous system and most of the craniofacial skeleton. Indeed, the neural crest can be considered as the fourth germ layer, providing a wide range of possibilities for further evolutionary invention [55].

A second important contribution was the broad development of the muscle and nerve protein tool-kits. The tunicates contributed all three of the MYBPC genes that populate the M-band of the sarcomere, leading to a great improvement in the functioning of muscle and the tunicates contributed numerous genes to the other classes of muscle proteins. The tunicates "invented" the first gap junction genes, so important in the Schwann cells and the development of rapidly-transmitting myelinated axons. The crystalline genes of the lens of the eye are again of tunicate origin being found in the tunicate palp's mucilage that enables the larval tunicate to attach to the sea bottom. These developments in mobility and vision provided the basis for the development of the efficient predatory capabilities of the Vertebrata. The evolutionary change from the lifestyle of the Branchiostomatidae to the shared portion of the lifestyles of the Tunicata and vertebrata is perhaps not so great a leap and the 84 orthologs that lie at the base of the Olfactores clade were clearly adequate to bridge this gap.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contribution: As this is a single-authored paper, WDS conceived the project, performed all the data retrieval and analysis, and wrote the paper.

Funding: No funding was needed.

Institutional Review Board Statement: As this was a database only study, no animals were used and thus no IRB statement is necessary

Acknowledgments: the author is very grateful to Dr Patrick Lemaire and Dr Clare Hudson for an introduction to the ANISEED database and for help with its use and to Dr Salvatore D'aniello for a helpful discussion of the swimming behavior of *Amphioxus*. Chana Stein suggested changes to, and edited early drafts of, the manuscript. Dr Thomas Litman, as always, provided much help and encouragement in this study.

Conflicts of Interest: There are no conflicts of interest between the author and any institution or company.

References

1. Domazet-Lošo T, Tautz D. Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa. *BMC B*iol 2010, 8:66 http://www.biomedcentral.com/1741-7007/8/66
2. Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalo-chordates are the closest living relatives of vertebrates. *Nature* 2006;439:965-8.
3. Liebeskind BJ, McWhite D, Marcotte EM. Towards consensus gene ages. *Genome Biol Evol* 2016;8:1812–1823. doi:10.1093/gbe/evw113
4. Litman T, Stein WD. Obtaining estimates for the ages of all the protein-coding genes and most of the ontology-identified noncoding genes of the human genome, assigned to 19 phylostrata. *Sem Oncology* 2019;46:3–9
5. Colgren J, Nichols SA. MRTF specifies a muscle-like contractile module in Porifera. *Nature Communications* 2022;13:4134 | https://doi.org/10.1038/s41467-022-31756-9
6. Zullo L, Bozzo M, Daya A, Di Clemente A, Mancini FP, Megighian A, Nesher N, Röttinger E, Shomrat T, Tiozzo S, Zullo A, Candiani S. The diversity of muscles and their regenerative potential across animals. *Cells* 2020: 9:1925
7. Jahnel SM, Walz M, Technau U. Development and epithelial organisation of muscle cells in the sea anemone Nematostella vectensis. *Front Zoolog* 2014;11:44
8. Wilkie IC, Candia Carnevali MD, Andrietti F. Mechanical properties of sea-urchin lantern muscles: a comparative investigation of intact muscle groups in Paracentrotus lividus (Lam.) and Stylocidaris affinis (Phil.) (Echinodermata, Echinoidea). *J Comp Physiol B* 1998;168:204-212

9. Hiebert TC, Gemmell BJ, von Dassow G, Conley KR, Sutherland KR. The hydrodynamics and kinematics of the appendicularian tail underpin peristaltic pumping. *J Roy Soc Interface* 2023;20:20230404. Https://doi.org/10.1098/rsif.2023.0404
10. Lindskog C, Linné J, Fagerberg L, Hallström BM, Sundberg CJ, Lindholm M, Huss M, Kampf C, Choi H, Liem DA, Ping P, Väremo L, Mardinoglu A, Nielsen J, Larsson E, Pontén F, Uhlén M. The human cardiac and skeletal muscle proteomes defined by transcriptomics and antibody-based profiling. *BMC Genomics* 2015;16:475
11. Schoenauer R, Bertoncini P, Machaidze G, Aebi U, Perriard JC, Hegner M, Agarkova I. Myomesin is a molecular spring with adaptable elasticity. *J Mol Biol* 2005; 349: 367–379
12. Abascal F, Zardoya R. Evolutionary analyses of gap junction protein families. *Biochimica Biophysica Acta* 2013;1828: 4–14
13. Lucaciu SA, Leighton SE, Hauser A, Yee R, Laird DW. Diversity in connexin biology. *J Biol Chem* 2023;299:105263
14. Purves D, Augustine GJ, Fitzpatrick D, et al., (ed). *Neuroscience*. 2nd edition. Sunderland (MA): Sinauer Associates; 2001.
15. More HL, O'Connor SM, Brøndum E, Wang T, Bertelsen MF, Grøndahl C, Kastberg K, Hørlyck A, Funder J, Maxwell Donelan J. Sensorimotor responsiveness and resolution in the giraffe. *J Exp Biol* 2013 216, 1003-1011
16. Weil MT, Heibeck S, Topperwien M, tom Dieck S, Ruhwedel T, Salditt T, Rodicio MC, Morgan JR, Nave KA, Mobius W, Werner HB. Axonal ensheathment in the nervous system of lamprey: implications for the evolution of myelinating glia. *J Neurosci* 2018;38:6586 – 6596
17. Gould RM, Morrison HG, Gilland E, Campbell RK. Myelin tetraspan family proteins but no non-tetraspan family proteins are present in the ascidian (Ciona intestinalis) genome. *Biol Bull* 2005; 209:49-66
18. Patel SD, Ciatto C, Chen CP, Bahna F, Rajebhosale M, Arkus N, Schieren I, Jessell TM, Honig B, Price SR, Shapiro L. Type II cadherin ectodomain structures: implications for classical cadherin specificity. *Cell* 2006;124: 1255–1268
19. Taneyhill LA. To adhere or not to adhere: The role of cadherins in neural crest development. *Cell Adhesion & Migration* 2008;2:223-230
20. Oda H, Akiyama-Oda Y, Zhang S. Two classic cadherin-related molecules with no cadherin extracellular repeats in the cephalochordate amphioxus: distinct adhesive specificities and possible involvement in the development of multicell-layered structures. *J Cell Sci* 2003; 117: 2757-2767
21. York JR, McCauley DW. The origin and evolution of vertebrate neural crest cells. *Open Biol* 2020;10:190285. http://dx.doi.org/10.1098/rsob.190285
22. Noda T, Satoh N. A comprehensive survey of cadherin superfamily gene expression patterns in Ciona intestinalis. *Gene Expression Patterns* 2008;8:349–356
23. Jeffery WR, Chiba T, Krajka FR, Deyts C, Satoh N, Joly JS. Trunk lateral cells are neural crest-like cells in the ascidian Ciona intestinalis: Insights into the ancestry and evolution of the neural crest. *Dev Biol* 2008;324:152–160
24. Niessen CM, Leckband D, Yap AS. Tissue organization by cadherin adhesion molecules: Dynamic molecular and cellular mechanisms of morphogenetic regulation. *Physiol Rev* 2011;91:691–731 doi:10.1152/physrev.00004.2010.
25. Lu YC, Nazarko OV, Sando III R, Salzman GS, Li NS, Südhof TC, Araç D. Structural basis of latrophilin - FLRT - UNC5 interaction in cell adhesion. *Structure* 2015; 23:1678–1691. doi:10.1016/j.str.2015.06.024.
26. del Toro D, Carrasquero-Ordaz MA, Chu A, Ruff T, Shahin M, Jackson VA, Chavent M, Berbeira-Santana M, Seyit-Bremer G, Brignani S, Kaufmann R, Lowe E, Klein R, Seiradake E. Structural basis of teneurin-latrophilin interaction in repulsive guidance of migrating neurons. *Cell* 2020;180: 323–339
27. Furuse M, Fujita K, Hiiragi T, Fujimoto K, Tsukita S. Claudin-1 and -2: Novel integral membrane proteins localizing at tight junctions with no sequence similarity to occludin. *J Cell Biol* 1998;141:1539–1550
28. Günzel D, Yu ASL. Claudins and the modulation of tight junction permeability. *Physiol Rev* 2013; 93: 525–569, doi:10.1152/physrev.00019.2012
29. Kollmar R, Nakamura SK, Kappler JA, Hudspeth AJ. Expression and phylogeny of claudins in vertebrate primordia. *PNAS* 2001;98:1810196–10201
30. Mellott DO, Burke RD. The molecular phylogeny of eph receptors and ephrin ligands. *BMC Cell Biol* 2008, 9:27 doi:10.1186/1471-2121-9-27
31. Taylor H, Campbell J, Nobes CD. Ephs and ephrins. *Curr Biol* 2027;27: R83–R102
32. Williaume G, de Buyl S, Sirour C, Haupaix N, Bettoni R, Imai KS, Satou Y, Dupont G, Hudson C, Yasuo H. Cell geometry, signal dampening, and a bimodal transcriptional response underlie the spatial precision of an ERK-mediated embryonic induction. *Dev Cell* 2021;56: 2966–2979
33. Fiuza UM, Negishi T, Rouan A, Yasuo H, Lemaire P. A Nodal/Eph signalling relay drives the transition from apical constriction to apico-basal shortening in ascidian endoderm invagination. *Development* 2020;147:dev186965. doi:10.1242/dev.186965

34. Negishi T, Nishida H. Asymmetric and Unequal Cell Divisions in Ascidian Embryos. In: Tassan, JP, Kubiak J. (eds) Asymmetric cell division in development, differentiation and cancer. Results and Problems in Cell Differentiation, 2017;61. Springer, Cham. https://doi.org/10.1007/978-3-319-53150-2_12
35. Haupaix N, Abitua PB, Sirour C, Yasuo H, Levine M, Hudson C. Ephrin-mediated restriction of ERK1/2 activity delimits the number of pigment cells in the Ciona CNS. *Dev Biol* 2014; 394: 170–180.
36. Stolfi A, Wagner E, Taliaferro JM, Chou S, Levine M. Neural tube patterning by Ephrin, FGF and Notch signaling relays. *Development* 2011;138: 5429-5439. doi:10.1242/dev.072108
37. Gee RRF, Chen H, Lee AK, Daly CA, Wilander BA, Fon Tacer K, Potts PR. Emerging roles of the MAGE protein family in stress response pathways. *J Biol Chem* 2020;295:16121–16155
38. Ray ME, Wistow G, Su YA, Meltzer PS, Trent JM. AIM1, a novel non-lens member of the bg-crystallin superfamily, is associated with the control of tumorigenicity in human malignant melanoma. *PNAS* 1997;94:3229–3234
39. Shimeld Sebastian M, Purkiss AG, Dirks RPH, Bateman OA, Slingsby C, Lubsen NH. Urochordate-crystallin and the evolutionary origin of the vertebrate eye lens. *Curr Biol* 2005;15: 1684–1689
40. Riyahi K, Shimeld SM. Chordate βγ-crystallins and the evolutionary developmental biology of the vertebrate lens. *Comp Biochem Physiol*, Part B 2007;147:347–357
41. Kappe G, Purkiss AG, van Genesen ST, Slingsby C, Lubsen NH. Explosive expansion of bc-crystallin genes in the ancestral vertebrate. J *Mol Evol* 2010;71:219–230 DOI 10.1007/s00239-010-9379-2
42. Slingsby C, Wistow GJ, Clark AR. Evolution of crystallins for a role in the vertebrate eye lens. *Prot Science* 2013;22:367-380
43. Cvekl A, Zhao Y, Mcgreal Rebecca, Xie Q, Gu X, and Zheng D. Evolutionary origins of Pax6 control of crystallin genes. *Genome Biol Evol* 2017; 9:2075–2092. Doi:10.1093/gbe/evx153
44. Zenga F, Wunderera J, Salvenmosera W, Hessb MW, Ladurnera P, Rothbächera U. Papillae revisited and the nature of the adhesive secreting collocytes. Dev Biol 2019;448:183–198
45. Horie T, Sakurai D, Ohtsuki H, Terakita A, Shichida Y, Usukura J, Kusakabe T, Tsuda M. Pigmented and nonpigmented ocelli in the brain vesicle of the ascidian larva. *J Comp Neur* 2008;509:88 –102
46. Kozlyuk N, Sengupta S, Bierma JC, Martin RW. Calcium binding dramatically stabilizes an ancestral crystallin fold in tunicate βγ-crystallin. *Biochemistry* 2016;55: 6961−6968
47. Cohen SM, Bronner G, Kuttner F, Jurgens G, Jackie H. Distal-less encodes a homoeodomain protein required for limb development in Drosophila. *Nature* 1989; 338: 432-434
48. Williams TA, Nulsen C, Nagy LM. A complex role for distal-less in crustacean appendage development. *Dev Biol* 2002;241: 302–312. doi:10.1006/dbio.2001.0497
49. Ullah A, Hammid A, Umair M, Ahmad W. A novel heterozygous intragenic sequence variant in DLX6 probably underlies first case of autosomal dominant split-hand/foot malformation type 1. *Mol Syndromol* 2017;8:79–84 DOI: 10.1159/000453350
50. Arendt D. Many ways to build a polyp. *Trends Gen* 2019; 35: 885-887
51. Lowe CJ, Wray GA. Radical alterations in the roles of homeobox genes during echinoderm evolution. *Nature* 1997;389:718- 721
52. Prpic NM, Tautz D. The expression of the proximodistal axis patterning genes Distal-less and dachshund in the appendages of Glomeris marginata (Myriapoda: Diplopoda) suggests a special role of these genes in patterning the head appendages. *Dev Biol* 2003;260:97–112
53. Caracciolo A, Di Gregorio A, Aniello F, Di Lauro R, Branno M. Identification and developmental expression of three Distal-less homeobox containing genes in the ascidian Ciona intestinalis. *Mech Dev* 2000;99:173-176
54. Holland ND, Panganiban G, Henyey EL, Holland LZ. Sequence and developmental expression of amphidll, an amphioxus Distal-less gene transcribed in the ectoderm, epidermis and nervous system: insights into evolution of craniate forebrain and neural crest *Development* 1996;122: 2911-2920
55. Shyamala K, Yanduri S, Girish HC, Murgod S. Neural crest: The fourth germ layer. *J Oral Maxillofac Pathol* 2015;19:221-9.
56. Wang S, DeLeon C, Sun W, Quae SR, Roth BL, Südhof TC. Alternative splicing of latrophilin-3 controls synapse formation. *Nature* 2024;626:128-135 doi: 10.1038/s41586-023-06913-9
57. Huber PAJ. Caldesmon. *Int J Biochem Cell Biol* 1997;29:1047-1051,
58. Polanco J, Reyes-Vigil F, Weisberg SD, Dhimitruka I, Brusés JL. differential spatiotemporal expression of type i and type ii cadherins associated with the segmentation of the central nervous system and formation of brain nuclei in the developing mouse. *Front Mol Neurosci* 2021;14:2021 https://doi.org/10.3389/fnmol.2021.633719
59. Hashimoto H, Munro E. Differential expression of a classic cadherin directs tissue-level contractile asymmetry during neural tube closure. *Dev Cell* 2019;51: 158–172
60. Paulson AF, Prasad MS, Thuringer AH. Manzerra P. Regulation of cadherin expression in nervous system development. *Cell Adhesion & Migration* 2014;8: 19-28, DOI: 10.4161/cam.27839

61. Matsuoka H, Yamaoka A, Hamashima T, Shima A, Kosako M, Tahara Y, Kamishikiryo J, Michihara A. EGF-dependent activation of ELK1 contributes to the induction of CLDND1 expression involved in tight junction formation. *Biomedicines* 2022;10:1792. doi: 10.3390/biomedicines10081792
62. Exposito JY, Cluzel C, Garrone R, Lethias C. Evolution of collagens. *Anatom Rec* 2002;268:302–316
63. Panganiban G, Rubenstein JLR. Developmental functions of the Distal-less/Dlx homeobox genes. *Development* 2002;129, 4371-4386
64. Stolfi A, Gainous TB, Young JJ, Mori A, Levine M, Christiaen L. Early chordate origins of the vertebrate second heart field. *Science* 2010; 329: 565–568 doi:10.1126/science.1190181
65. Moser C, Gossel´ KA, Balaz M, Balazova L, Horvath C, Künzle P, Okreglicka KM, Li F, Blüher M, Stierstorferf B, Hessf E, Lamla T, Hamilton B, Klein H, Neubauerf H, Wolfrum C, Wolfrum S. FAM3D: A gut secreted protein and its potential in the regulation of glucose metabolism. *Peptides* 2023;167:171047
66. Chen W, Gao D, Xie L, Wang A, Zhao H, Guo C, Sun Y, Nie Y, Hong A, Xiong S. SCF-FBXO24 regulates cell proliferation by mediating ubiquitination and degradation of PRMT6. *Biochem Biophys Res Comm* 2020;530:75-81,
67. Satou Y, Tokuoka M, Oda-Ishii I, Tokuhiro S, Ishida T, Liu B, Iwamura Y. A manually curated gene model set for an ascidian, *Ciona robusta (Ciona intestinalis type A*). *Zool Sci* 2022;39: 253–260
68. Smith HM, Khairallah SM, Nguyen AH, Newman-Smith E, Smith WC. Misregulation of cell adhesion molecules in the Ciona neural tube closure mutant bugeye. *Dev Biol* 2021;480: 14–24
69. Abitua PB, Gainous TB, Kaczmarczyk AN, Winchell CJ, Hudson C, Kamata K, Nakagawa M, Tsuda M, Kusakabe TG, Levine M. The pre-vertebrate origins of neurogenic placodes. *Nature*. 2015; 524: 462–465. Doi:10.1038/nature14657.
70. Edvardsen RB, Seo HC, Jensen MF, Mialon A, Mikhaleva J, Bjordal M, Cartry J, Reinhardt R, Weissenbach J, Wincker P, Chourrout D. Remodelling of the homeobox gene complement in the tunicate Oikopleura dioica. *Curr Biol* Vol 15 No 1
71. Labat-de-Hoz L, Rubio-Ramos A, Correas I, Alonso MA. The MAL family of proteins: normal function, expression in cancer, and potential use as cancer biomarkers. *Cancers* 2023;15:2801. Https://doi.org/10.3390/cancers15102801
72. Chen YH, Pai CW, Huang SW, Chang SN, Lin LY, Chiang FT, Lin JL, Hwang JJ, Tsai CT. Inactivation of myosin binding protein C homolog in zebrafish as a model for human cardiac hypertrophy and diastolic dysfunction. *J Am Heart Assoc* 2013;2:e000231 doi: 10.1161/JAHA.113.000231
73. Razy-Krajka F, Stolf A. Regulation and evolution of muscle development in tunicates. *EvoDevo* 2019;10:13 https://doi.org/10.1186/s13227-019-0125-6
74. Heissler SM, Sellers JR. Myosin light chains: Teaching old dogs new tricks. *Bioarchitecture* 2014;4:169-188
75. Ramirez I, Gholkar AA, Velasquez EF., Guo X, Tofig B, Damoiseaux R, Torres JZ. The myosin regulatory light chain Myl5 localizes to mitotic spindle poles and is required for proper cell division. *Cytoskeleton* 2021; 78: 23–35. doi:10.1002/cm.21654.
76. Shaffer JF, Gillis TE. Evolution of the regulatory control of vertebrate striated muscle: the roles of troponin I and myosin binding protein-C. *Physiol Genomics* 2010;42: 406–419
77. Langea S, Pinotsisc N, Agarkovad I, Ehlere E. The M-band: The underestimated part of the sarcomere. *BBA - Mol Cell Res* 2020;1867:118440
78. Du Pasquier L. Speculations on the origin of the vertebrate immune system. Immunol Lett. 2004;92:1-2
79. Juriloff DM, Harris MJ. Insights into the etiology of mammalian neural tube closure defects from developmental, genetic and evolutionary studies. *J Dev Biol* 2018;6:22; doi:10.3390/jdb6030022
80. Medina-Martinez O, Shah R, Jamrich M. Pitx3 controls multiple aspects of lens development. *Dev Dyn* 2009; 238: 2193–2201. doi:10.1002/dvdy.21924.
81. Scales SJ, Hesser BA, Masuda ES, Scheller RH. Amisyn, a novel syntaxin-binding protein that may regulate snare complex assembly. *J Biol Chem* 2002;277:28271–28279,
82. Wei B, Jin JP. TNNT1, TNNT2, and TNNT3: Isoform genes, regulation, and structure-function relationships. *Gene* 2016; 582: 1–13. doi:10.1016/j.gene.2016.01.006.
83. Satou Y, Imai KS. Ascidian Zic Genes. In: Aruga, J. (eds) Zic family. Advances in Experimental Medicine and Biology, 2018;1046. Springer, Singapore. https://doi.org/10.1007/978-981-10-7311-3_6