

Evolution of Genomes



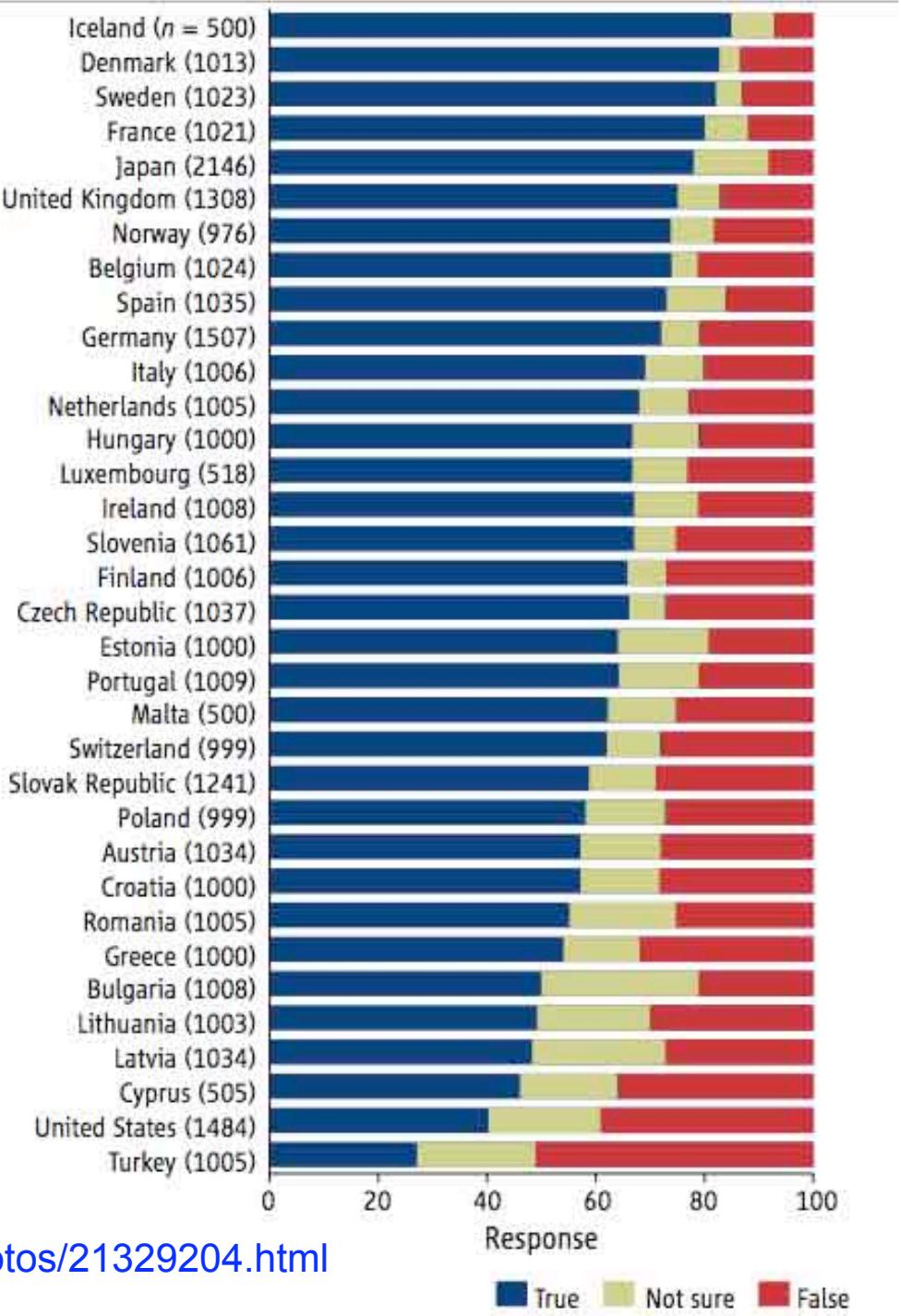
Billie J. Swalla

University of Washington

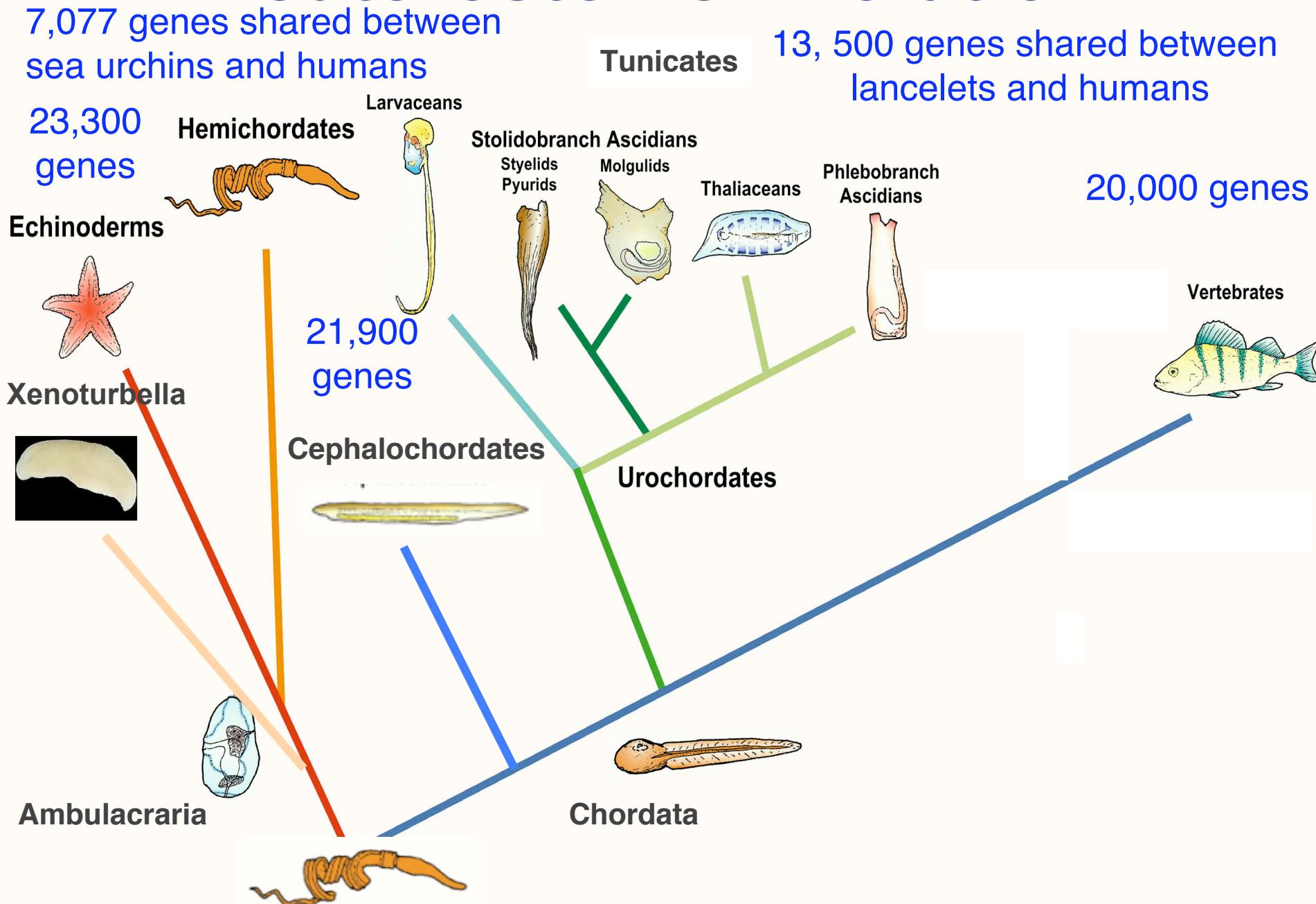
bjswalla@u.washington.edu

Question:

"Human beings, as we know them, developed from earlier species of animals."



Deuterostome Evolution

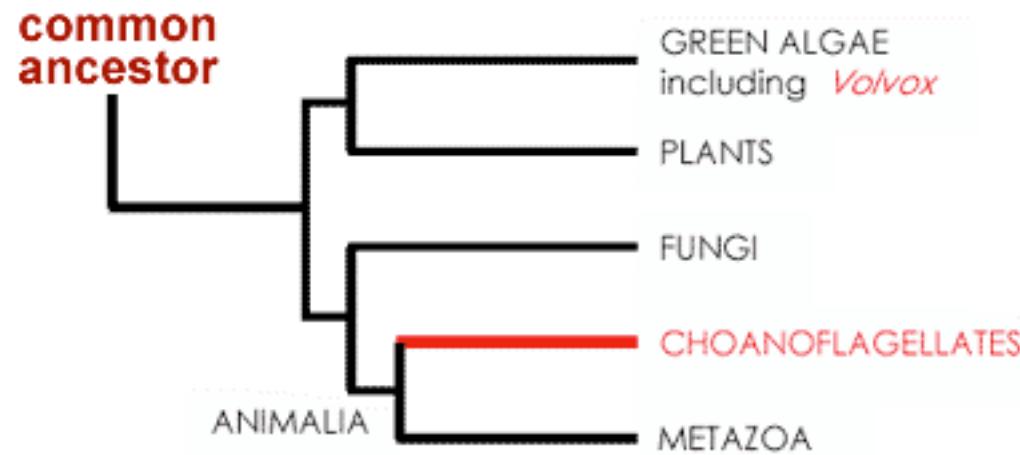


Shared Ancestral Traits

1. Universal Genetic Code

2nd→	U	C	A	G	3rd↓
1st↓	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr Stop Stop	Cys Cys Stop Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

Choanoflagellates



These organisms have cells that are characterized by a single **flagellum** surrounded by a distinctive **collar**.

Choanoflagellates can exist in simple colonial forms.

The collar cells are located on the periphery of cell aggregate, while ameobiod cells are located in the interior.

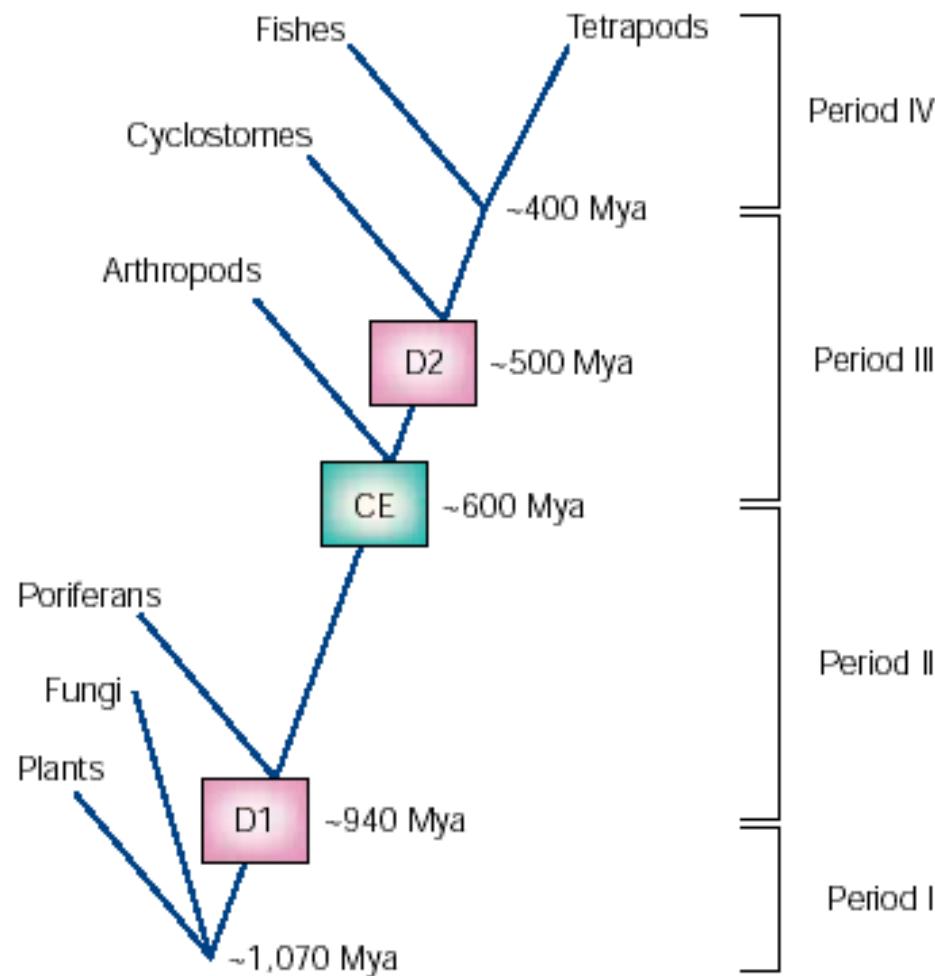
The most primitive metazoans, the sponges or **porifera**, contain cells that are very similar in structure to choanoflagellate

Genetic evidence suggests that these are the sister group to metazoans

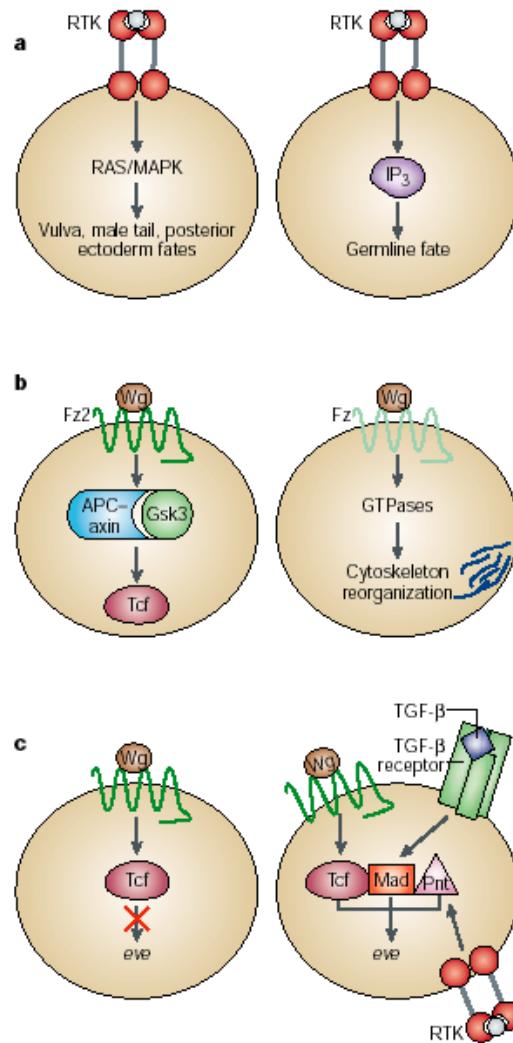
Which of the Signaling Pathways do Choanoflagellates have?

- ❖ RTKs
- ❖ Adhesion molecules
- ❖ Have frizzled and β catenin, but no wnts.
- ❖ No TGF β s

Evolution of Developmental Signaling Pathways



Evolution of Developmental Signaling Pathways



Placozoans



These organisms are made up of thousands of cells,
but only 4 cell types.

They are transparent, flat and round,
usually only several mm long.

Trichoplax adhaerens



Oliver Vines 2006

Comparison of Basal Metazoan Genomes

	<i>Monosiga brevicollis</i>	<i>Amphimedon queenslandica</i>	<i>Trichoplax adhaerens</i>	<i>Nematostella vectensis</i>	<i>Drosophila melanogaster</i>
genome size (Mb)	41.6	167	98	450	180
no. of genes	9100	?	11 514	18 000	14 601
no. of cell types	1	12	4	20	50
no. of miRNA	0	7	7	3	49
no. of metazoan transcription factors/families	?/5	57/?6	35/9	min. 87/10	min. 87/10
no. of <i>bHLH</i> genes	0	16	27	68	59

Erwin D. H. (2009) Phil. Trans. R. Soc. B 364: 2253-2261

Which of the Signaling Pathways do sponges have?

Table 1. Animal signaling-pathway components in the sponge *O. carmela*

Pathway components [†]	Functional roles	Phylogenetic distribution of signalling gene families*								
		Hs	Mm	Gg	Sp	Dm	Ce	Sp	Sc	Amb
Wnt										
<i>Wnt</i>	Ligand	●	●	●	●	●	●	○	○	○
<i>Frizzled</i>	Receptor	●	●	●	●	●	●	○	○	●
<i>Dkk</i>	Antagonist	●	●	●	○	○	○	○	○	○
<i>Dixdc</i>	Positive regulator	●	●	●	●	○	○	○	○	○
<i>β-catenin</i>	Downstream effector	●	●	●	●	●	●	●	●	●
<i>Dvl</i>	Downstream effector	●	●	●	●	●	●	○	○	○
<i>Nemo</i>	Downstream effector	●	●	●	●	●	●	○	○	○
TGFβ										
<i>Follistatin</i>	Antagonist	●	●	●	●	●	○	○	○	○
<i>Mad</i>	Downstream effector	●	●	●	●	●	●	○	○	○
Hedgehog										
<i>Hh</i>	Ligand	●	●	●	●	●	●	○	○	○
<i>Ptc</i>	Receptor	●	●	●	●	●	●	○	○	○
<i>Disp</i>	Ligand export	●	●	●	●	●	●	○	○	○
<i>Sufu</i>	Downstream effector	●	●	●	●	●	○	○	○	○
Receptor tyrosine kinase										
<i>Egfr</i>	Receptor	●	●	●	●	●	●	○	○	○
<i>Igfr</i>	Receptor	●	●	●	●	●	●	○	○	○
<i>Fgfr</i>	Receptor	●	●	●	●	●	●	○	○	○
<i>Epha</i>	Receptor	●	●	●	●	●	●	○	○	○
<i>Ret</i>	Receptor	●	●	●	○	●	●	○	○	○
<i>Musk</i>	Receptor	●	●	●	○	●	●	○	○	○
<i>Ddr</i>	Receptor	●	●	●	●	●	●	○	○	○
Jak/Stat										
<i>Jak</i>	Receptor cofactor	●	●	●	●	●	○	○	○	○
<i>Stat</i>	Transcription factor	●	●	●	●	●	●	○	○	●
<i>Stam</i>	Downstream effector	●	●	●	●	●	●	●	●	○
<i>Pias</i>	Antagonist	●	●	●	●	●	●	○	●	○
Notch/Delta										
<i>Delta</i>	Ligand	●	●	●	●	●	●	○	○	○
<i>Notch</i>	Receptor	●	●	●	●	●	●	○	○	○

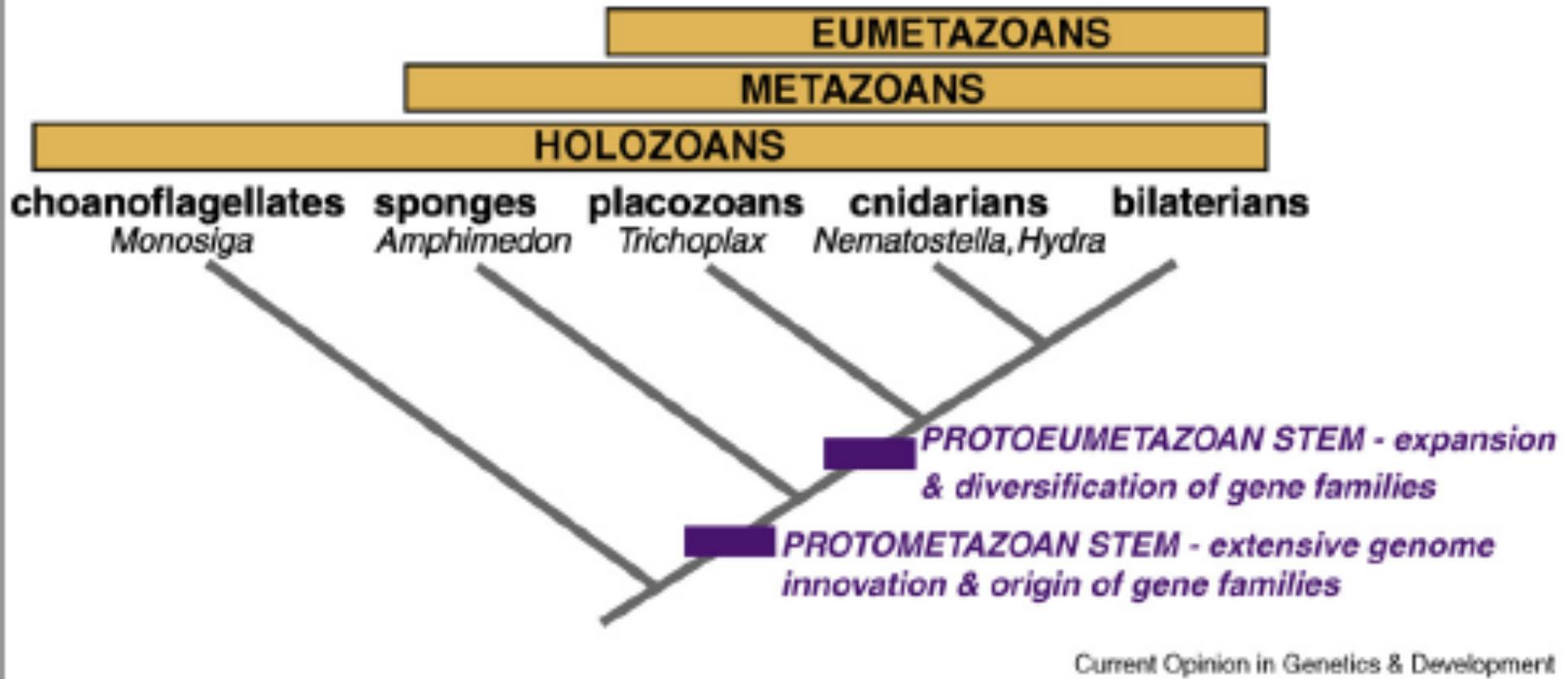
Nichols SA, Dirks W, Pearse JS, King N. (2006) Early evolution of animal cell signaling and adhesion genes. Proc Natl Acad Sci 103(33):12451-6

Which of the Adhesion Molecules do sponges have?

Table 2. Eumetazoan cell-adhesion machinery in the sponge *O. carmela*

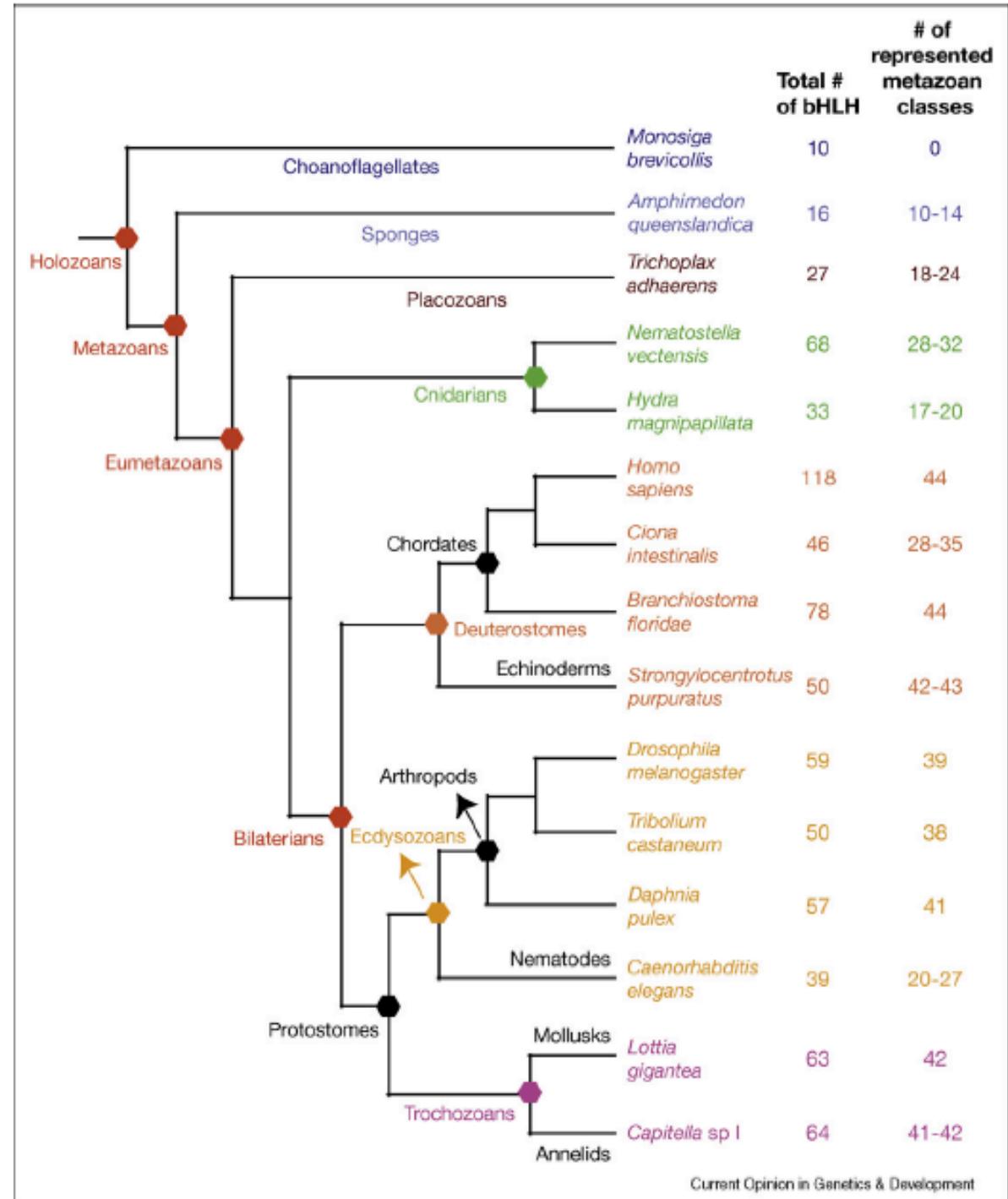
Adhesion genes homologs [†]	Phylogenetic distribution of cell adhesion gene families*								
	Bilateria				Fungi		Amb		
	Hs	Mm	Gg	Sp	Dm	Ce	Sp	Sc	Dd
Cell-contact and -adhesion proteins									
Contactin	●	●	●	●	●	●	○	○	○
β-integrin	●	●	●	●	●	●	○	○	○
α-integrin	●	●	●	●	●	●	○	○	○
ADAM									
Protocadherin	●	●	●	●	●	●	○	○	○
Fat	●	●	●	●	●	●	○	○	○
NCAM	●	●	●	●	●	●	○	○	○
Selectin	●	●	●	●	●	●	○	○	○
Tetraspanin									
Plexin	●	●	●	●	●	●	○	○	○
Heparanase	●	●	●	●	●	●	○	○	○
Neurexin	●	●	●	●	●	●	○	○	○
Crumbs						●	○	○	○
ECM molecules									
Agrin	●	●	●	●	●	●	○	○	○
Cthrc	●	●	●	●	●	●	○	○	○
Col11a2	●	●	●	●	●	●	○	○	○
Col4									
Fibulin	●	●	●	●	●	●	○	○	○
Frst	●	●	●	●	●	●	○	○	○
α-laminin	●	●	●	●	●	●	○	○	○
β-laminin									
Netrin	●	●	●	●	●	●	○	○	○
Perlecan	●	●	●	●	●	●	○	○	○
Tenascin									
Thrombospondin	●	●	●	●	●	●	○	○	○
Spondin	●	●	●	●	●	●	○	○	○
Fibrillin	●	●	●	●	●	●	○	○	○
40S ribosomal protein SA	●	●	●	●	●	●	●	●	●
Nardilysin	●	●	●	●	●	●	○	○	○
Adhesion-related cytoskeletal linkers									
Ankyrin	●	●	●	●	●	●	○	○	○
α-actinin	●	●	●	●	●	●	●	●	●
Paxillin	●	●	●	●	●	●	○	○	○
β-catenin									
p130 ^{cas}	●	●	●	●	●	●	○	○	○
Rho1 GTPase	●	●	●	●	●	●	○	○	○
Spectrin	●	●	●	●	●	●	○	○	○
Talin	●	●	●	●	●	●	○	○	●
Vinculin	●	●	●	●	●	●	○	○	○
Fascin									
Filamin	●	●	●	●	●	●	○	○	○
Parvin	●	●	●	●	●	●	○	○	●
LASP-1	●	●	●	●	●	●	○	○	○

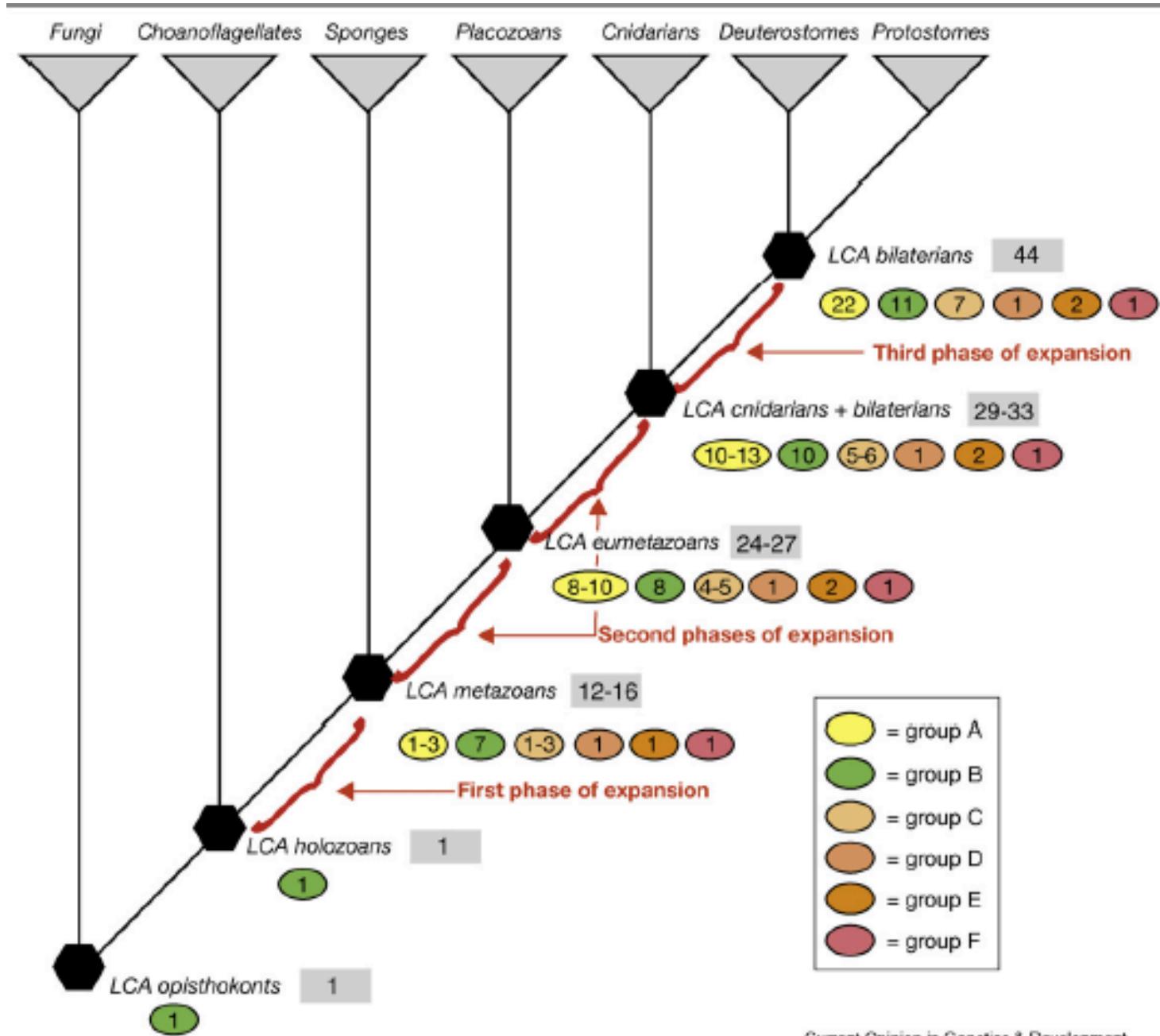
Nichols SA, Dirks W, Pearse JS, King N. (2006) Early evolution of animal cell signaling and adhesion genes. Proc Natl Acad Sci 103(33):12451-6



Current Opinion in Genetics & Development

Degnan et al. (2009)
Current Opinion in Genetics
and Development
19:1-9





Current Opinion in Genetics & Development

Current Opinion in Genetics and Development (2009) 19:1-9

Unexpected complexity of the *Wnt* gene family in a sea anemone

Arne Kusserow¹, Kevin Pang², Carsten Sturm¹, Martina Hrouda³,
Jan Lentfer¹, Heiko A. Schmidt⁴, Ulrich Technau^{1,*}, Arndt von Haeseler^{4,5},
Bert Hobmayer³, Mark Q. Martindale² & Thomas W. Holstein^{1,6}

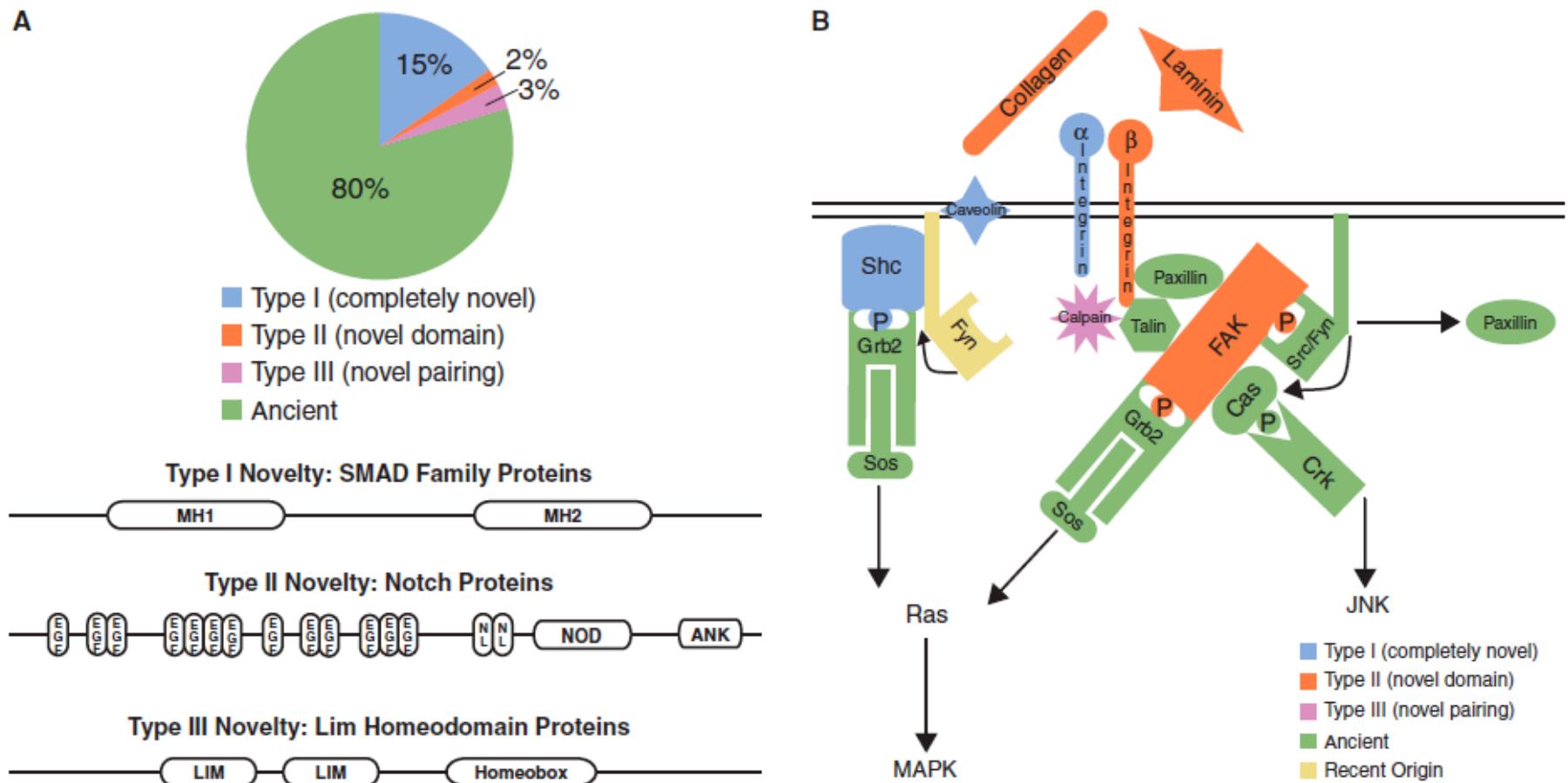
Sea Anemone Genome Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization

Nicholas H. Putnam,¹ Mansi Srivastava,² Uffe Hellsten,¹ Bill Dirks,² Jarrod Chapman,¹
Asaf Salamov,¹ Astrid Terry,¹ Harris Shapiro,¹ Erika Lindquist,¹ Vladimir V. Kapitonov,³
Jerzy Jurka,³ Grigory Genikhovich,⁴ Igor V. Grigoriev,¹ Susan M. Lucas,¹ Robert E. Steele,⁵
John R. Finnerty,⁶ Ulrich Technau,⁴ Mark Q. Martindale,⁷ Daniel S. Rokhsar^{1,2*}

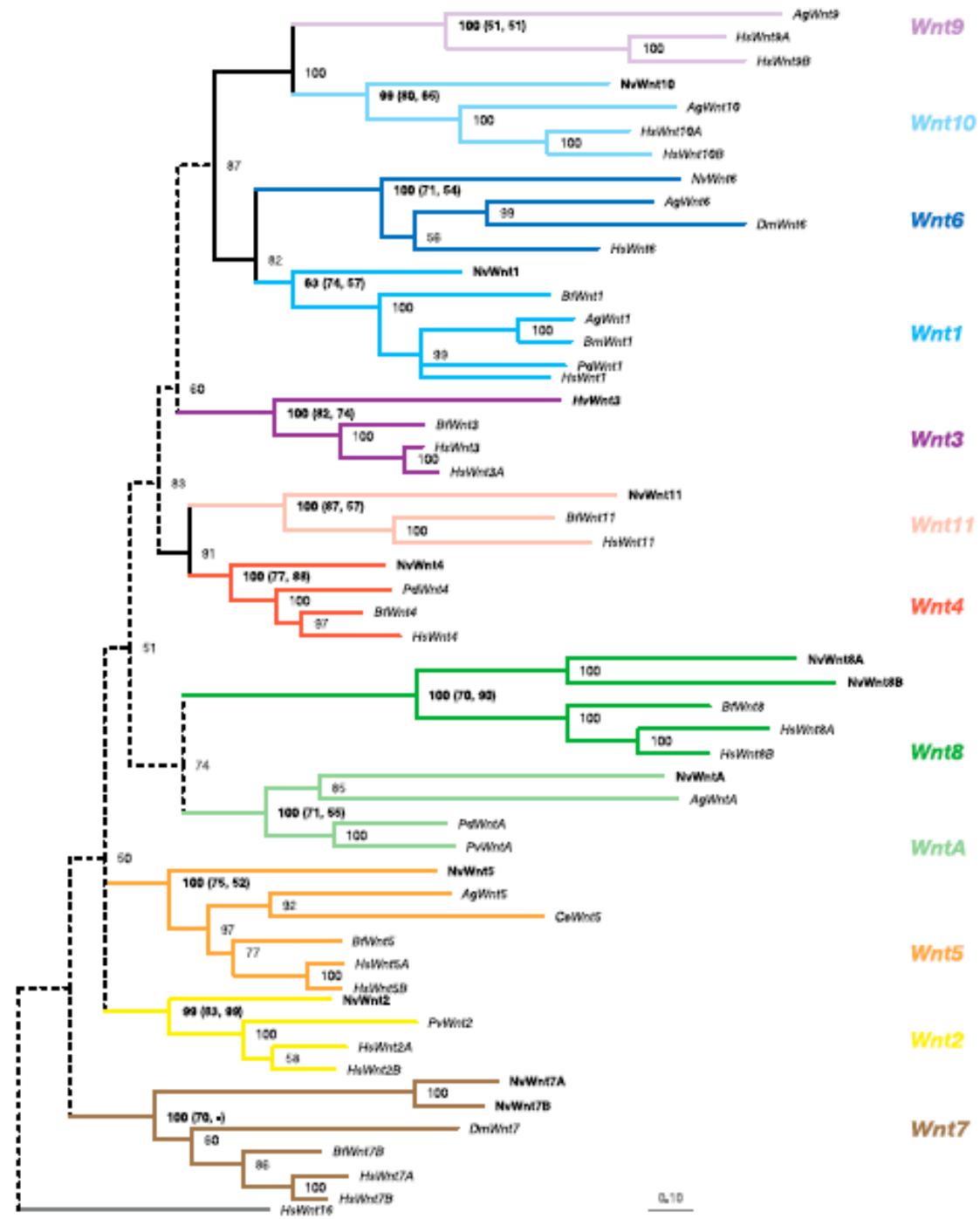
Nematostella vectensis - sea anenome



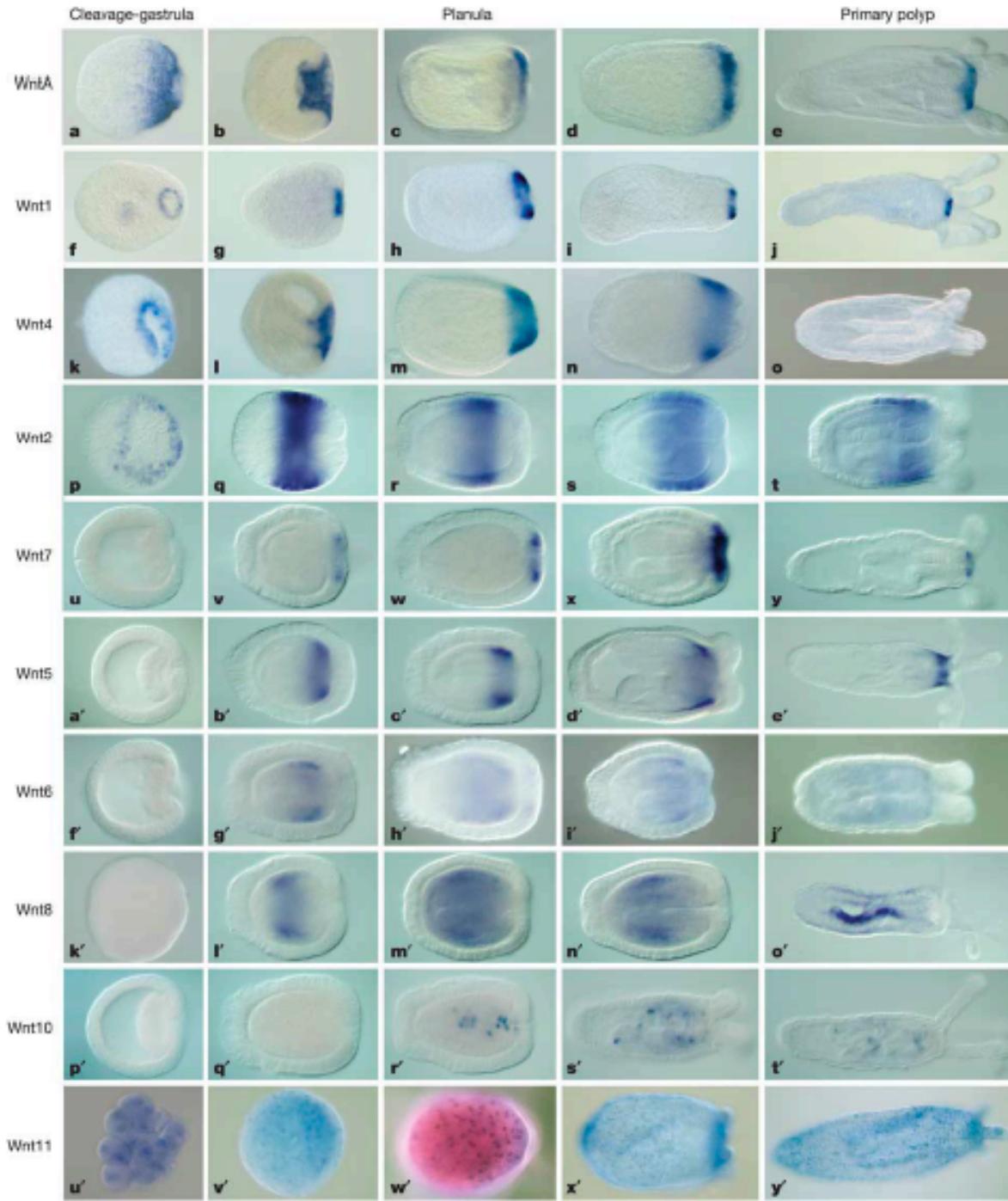
Is most of the *Nematostella* genome recent or ancient?



Putnam et al. (2007) Science 317: 86-94.



Kusserow et al.
(2005)
Nature
433: 156-160



Kusserow et al.
(2005)
Nature
433: 156-160

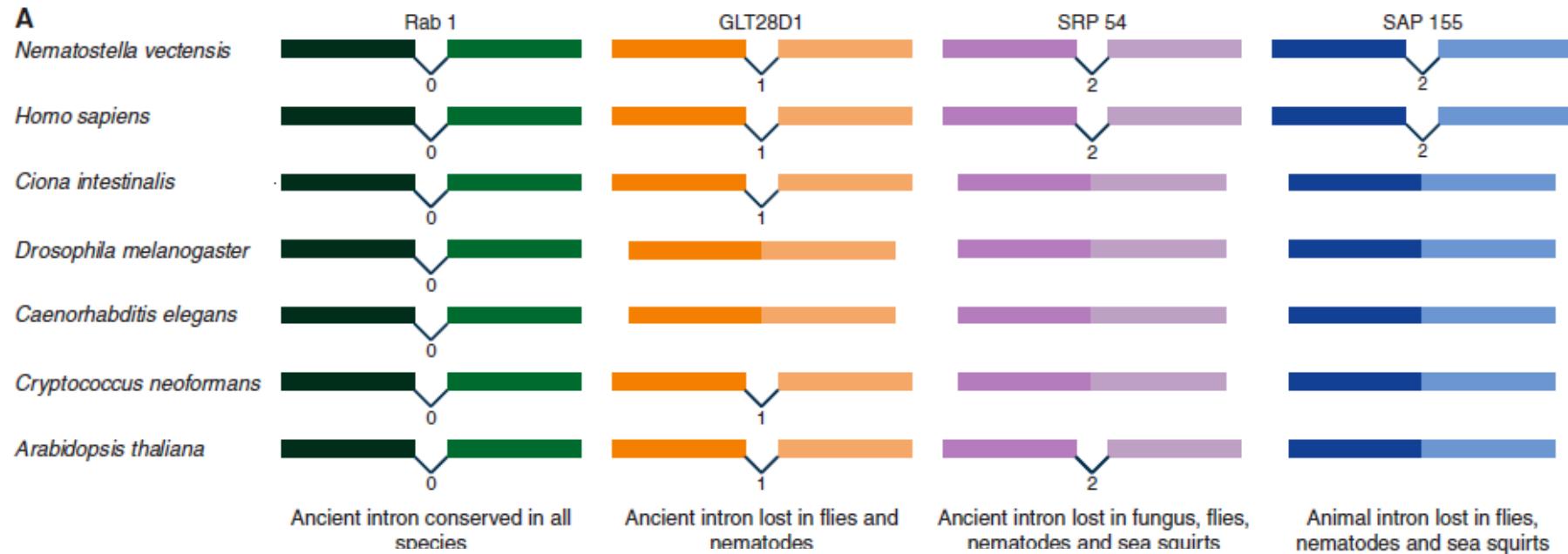
Families	Genes	Distribution of the Activators of the Wnt signals throughout the animal kingdom							
		Cnidarians	Ecdysozoans	Lophotrochozoans	Echinoderms	Urochordates	Cephalochordates	Vertebrates	
		Sea anemone (<i>Nematostella vectensis</i>)	Fly (<i>Drosophila melanogaster</i>)	Annelid and Molluscs (<i>Platynereis dumerilii</i> and <i>Parietula vulgaris</i>)	Sea Urchin (<i>Strongylocephalus purpuratus</i>)	Ascidian (<i>Clavelina intestinalis</i>)	Amphioxus (<i>Branchiostoma floridae</i>)	Human (<i>Homo sapiens</i>)	
wnt1		[teal]	[teal]		[teal]	[teal]	[teal]	[teal]	
wnt2/13		[yellow]	X	[yellow]	X	[yellow]	[yellow]	[yellow]	
wnt3		[purple]	X		[purple]	[purple]	[purple]	[purple]	
wnt4		[red]	X	[red]	[red]	[red]	[red]	[red]	
wnt5		[orange]	[orange]	[orange] (D. japonica)	[orange]	[orange]	[orange]	[orange]	
wnt6		[blue]	[blue]		[blue]	[blue]	[blue]	[blue]	
wnt7		[brown]	[brown]	[brown]	[brown]	[brown]	[brown]	[brown]	
wnt8		[green]	X		[green]	X	[green]	[green]	
wnt9/14/15			[purple]	[purple]	[purple]	[purple]	[purple]	[purple]	
wnt10		[light blue]	[light blue]	[light blue]	[light blue]	[light blue]		[light blue]	
wnt11		[black]	X		X	X	[black]	[black]	
wnt16		[pink]	(C. elegans-egl20)		[pink]	[pink]		[pink]	
wntA		[olive green]	(A. gambiae)	[olive green]		X		X	

Fig. 5. Survey of the Wnt family of secreted signaling molecules in selected metazoans. Each square indicates a single Wnt gene identified either through genome analyses or independent studies, and squares with a question mark

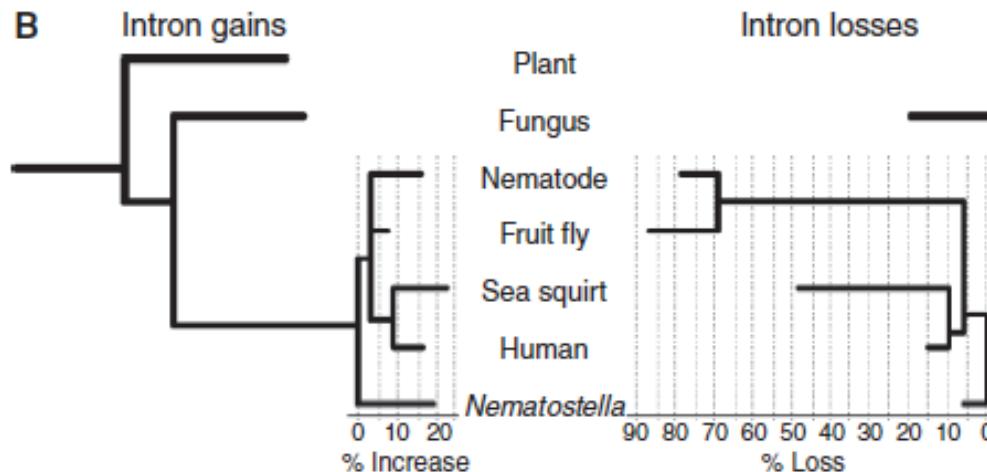
indicate uncertainty of the orthology. Letter X's represent absence of members of that subfamily in the corresponding annotated genome; empty spaces have been left for species for which genomic databases are not yet available. [From (30)]

Sea Urchin Genome Sequencing
Consortium (2006)
Science
314: 941-952

Presence and Position of Introns

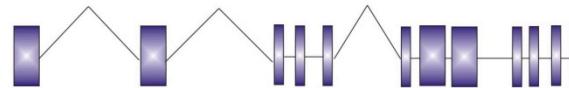


Putnam et al. (2007) Science 317: 86-94.



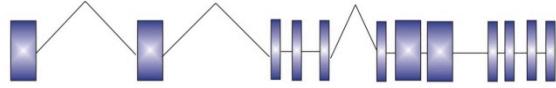
Vertebrate P2X

Homo sapiens P2X4
2028 mRNA
388 aa



24,245 gene length
12 exons

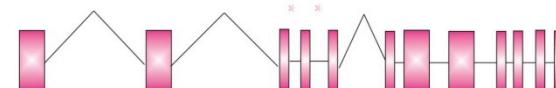
Mus musculus P2X4
1995 mRNA
388 aa



21,488 gene length
12 exons

Invertebrate P2X

Aplysia californica
1302 mRNA
434 aa



26,883* gene length
12 exons

Daphnia pulex
1206 mRNA
402 aa



5,189 gene length
8 exons(3 combined)

Nematostella vectensis
1206 mRNA
402 aa



4,397 gene length
8 exons(3 combined)

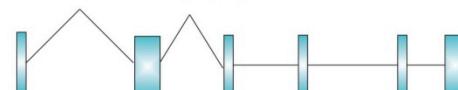
Monosiga brevicollis
1443 mRNA
418 aa



2,970 gene length
15 exons

Other P2X

Hydra magnipapillata
1077 mRNA
359 aa



18,200 gene length
6 exon

Nematostella vectensis
1191 mRNA
398 aa



6854 gene length
7 exons

Dictyostelium discoideum
906 mRNA
302 aa



1009 gene length
4 exons

Pleurobrachia bachei
1018 mRNA
339 aa



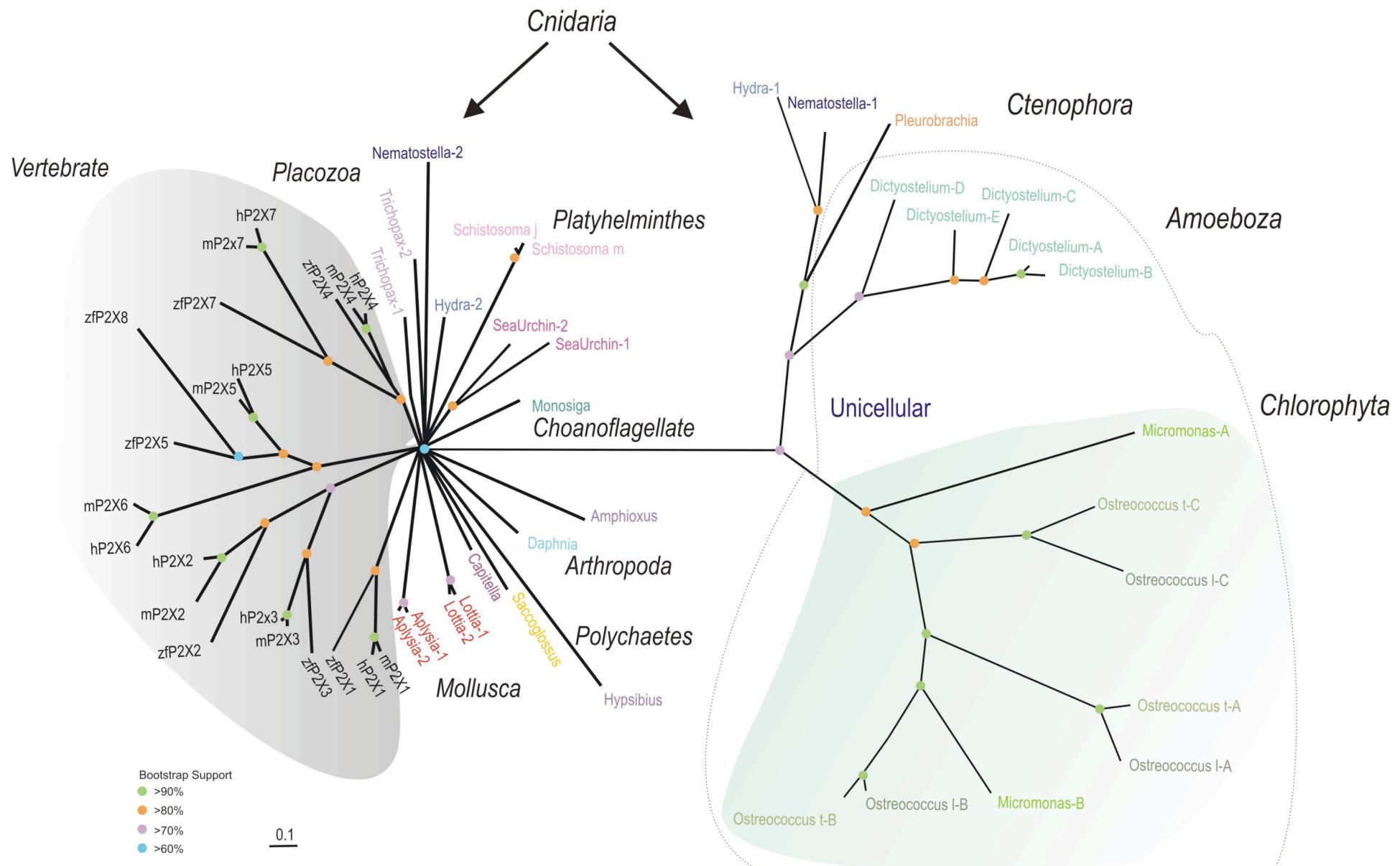
1029 gene length
2 exons

Ostreococcus tauri
1779 mRNA
593 aa



1779 gene length
0 exons

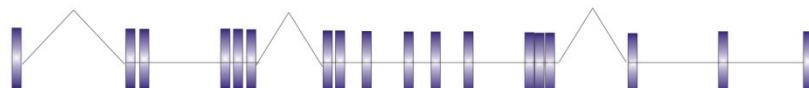
Andrea Kohn



Andrea Kohn

Glutamate Receptors

Human GRIK2



671,053 bp in genome on Chromosome 6
17exons 4897 bp mRNA 869 aa protein

Mouse Grik2



682,480 bp in genome on Chromosome 10
16 exons 4861 bp mRNA 869 aa protein

C. elegans glr-4



9,137 bp in genome on Chromosome 2
15 exons 2959 bp mRNA 951 aa protein

Drosophila gluR



5,300 bp in genome on Chromosome 2L
12 exons 3305 bp mRNA 1000 aa protein

Plant AtGLR2



6 exons

3,638 bp in genome on Chromosome 2
6 exons 3005 bp mRNA 951 aa protein

Pleurobranchia GluR4

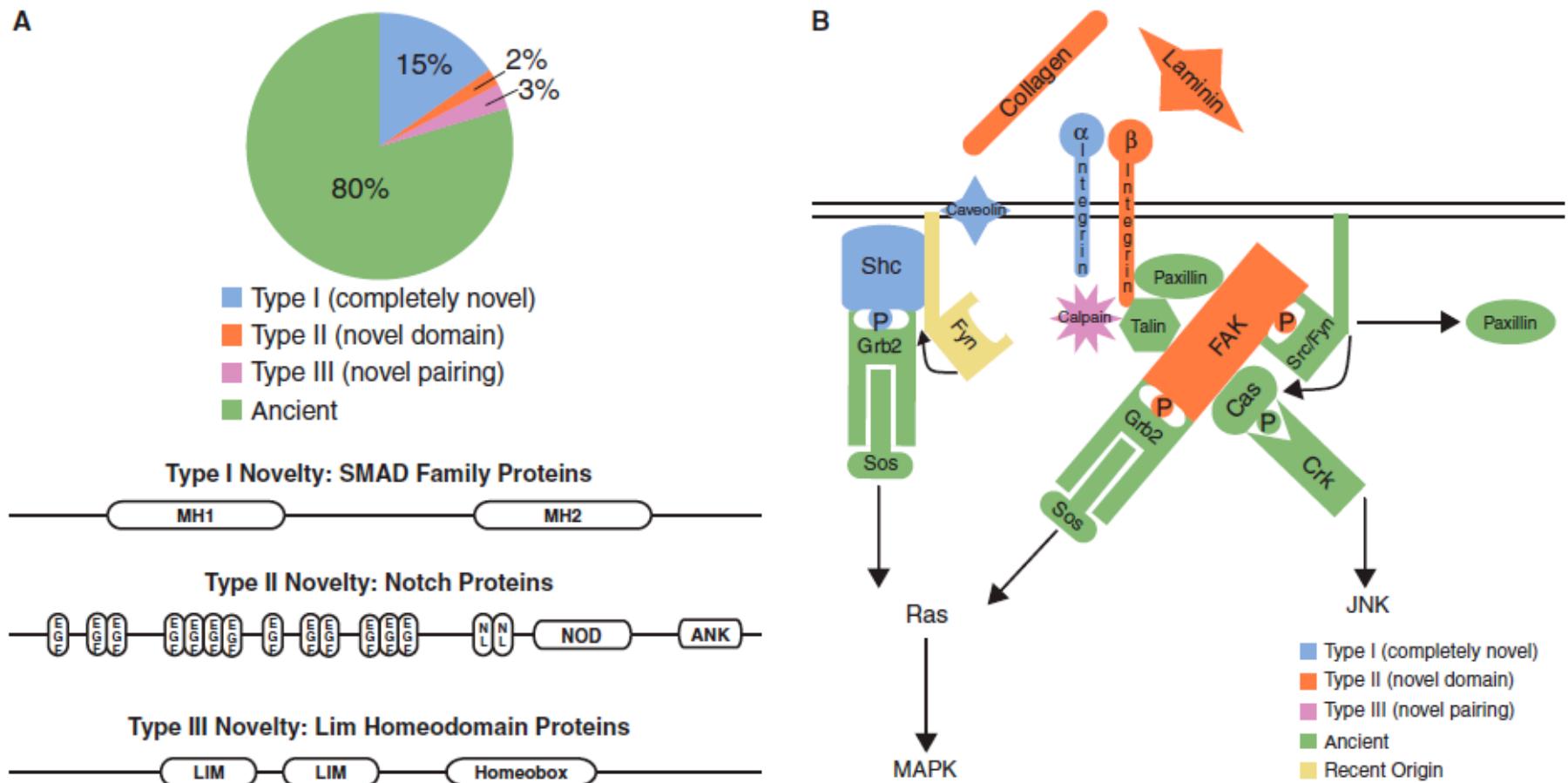


1 exons

2918 bp in genome on
1 exons 2685 bp mRNA 861 aa protein

Andrea Kohn

Is most of the *Nematostella* genome recent or ancient?



Putnam et al. (2007) Science 317: 86-94.

Synteny of Genes

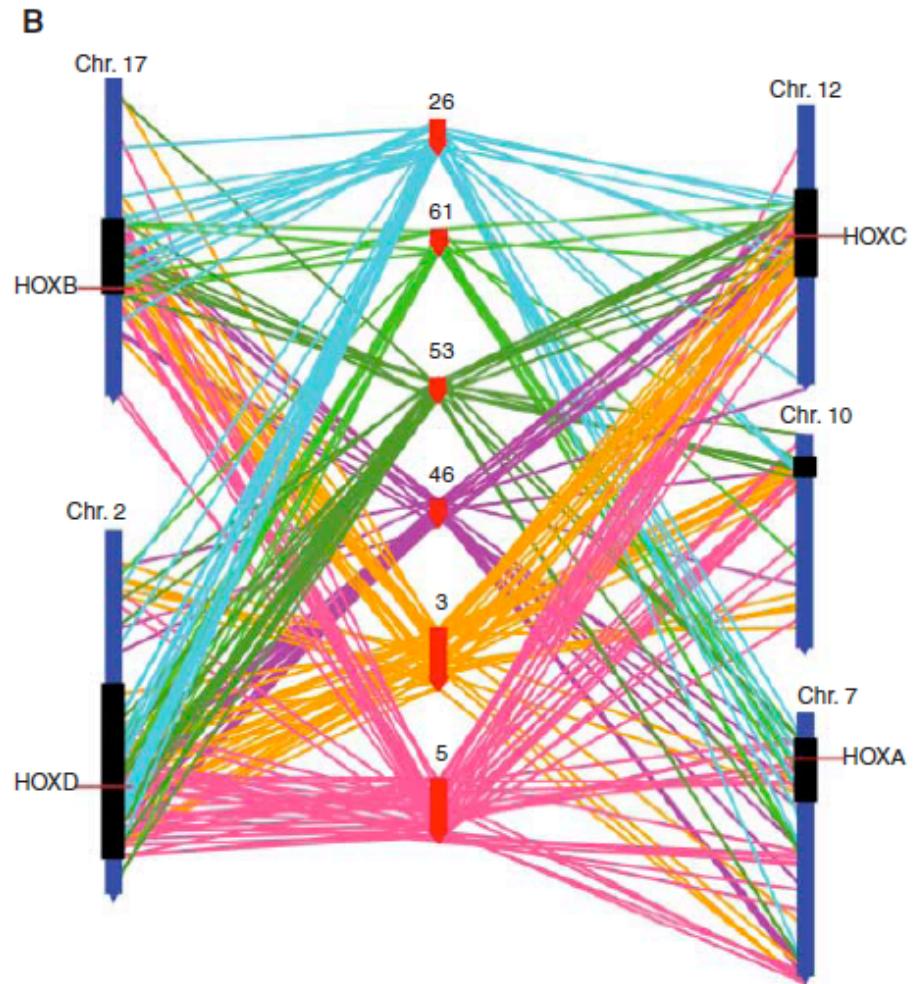
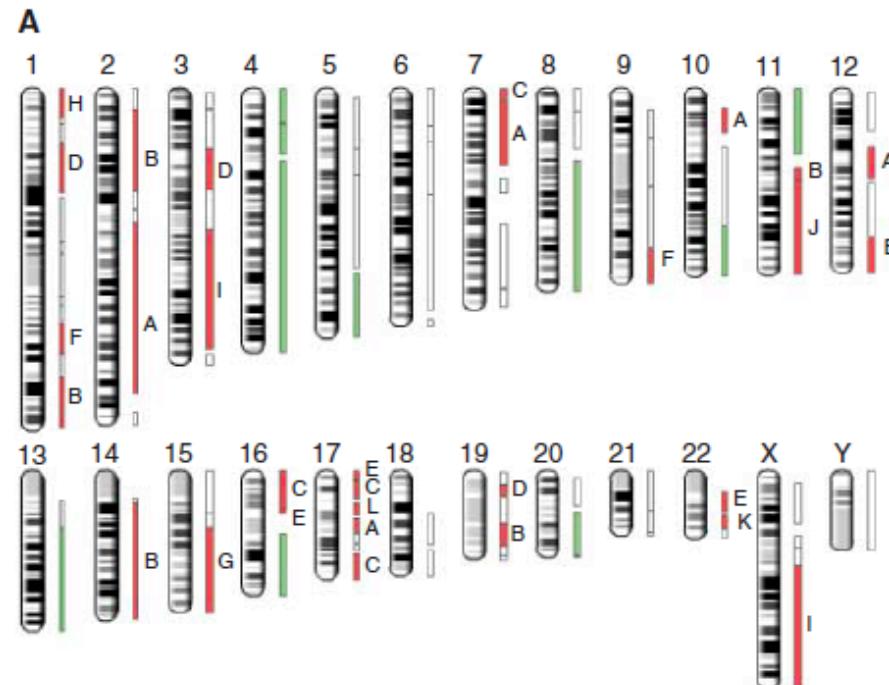


Fig. 3. Conserved synteny between the human and anemone genomes. (A) The human genome, segmented into 98 regions whose linkage has not been broken during chordate evolution. Colored segments indicate statistically significant conservation of linkage between human and *Nematostella*. Red segments are members of the 12 compact PALs labeled A to L. Green segments fall into the diffuse 13th PAL (32). White segments do not show significant conservation of linkage. **(B)** Conserved linkage between human chromosomal segments and *Nematostella* scaffolds in the first PAL (which includes the human Hox clusters). *Nematostella* scaffolds 26, 61, 53, 46, 3,

Putnam et al. (2007) Science 317: 86-94.

Novel Genes in *Nematostella* genome

Type I Novelty - Animal genes with no homologies outside metazoans (15%) - wnts, TGF- β , etc.

Type II Novelty - Combine animal-specific domains with ancient eucaryotic genes (2%)

Type III Novelty - Ancient motifs arranged in novel combinations (3%) - LIM domains, Homeodomains, etc.

Evolution of Signaling Pathways

Putnam et al. (2007) Science 317: 86-94.

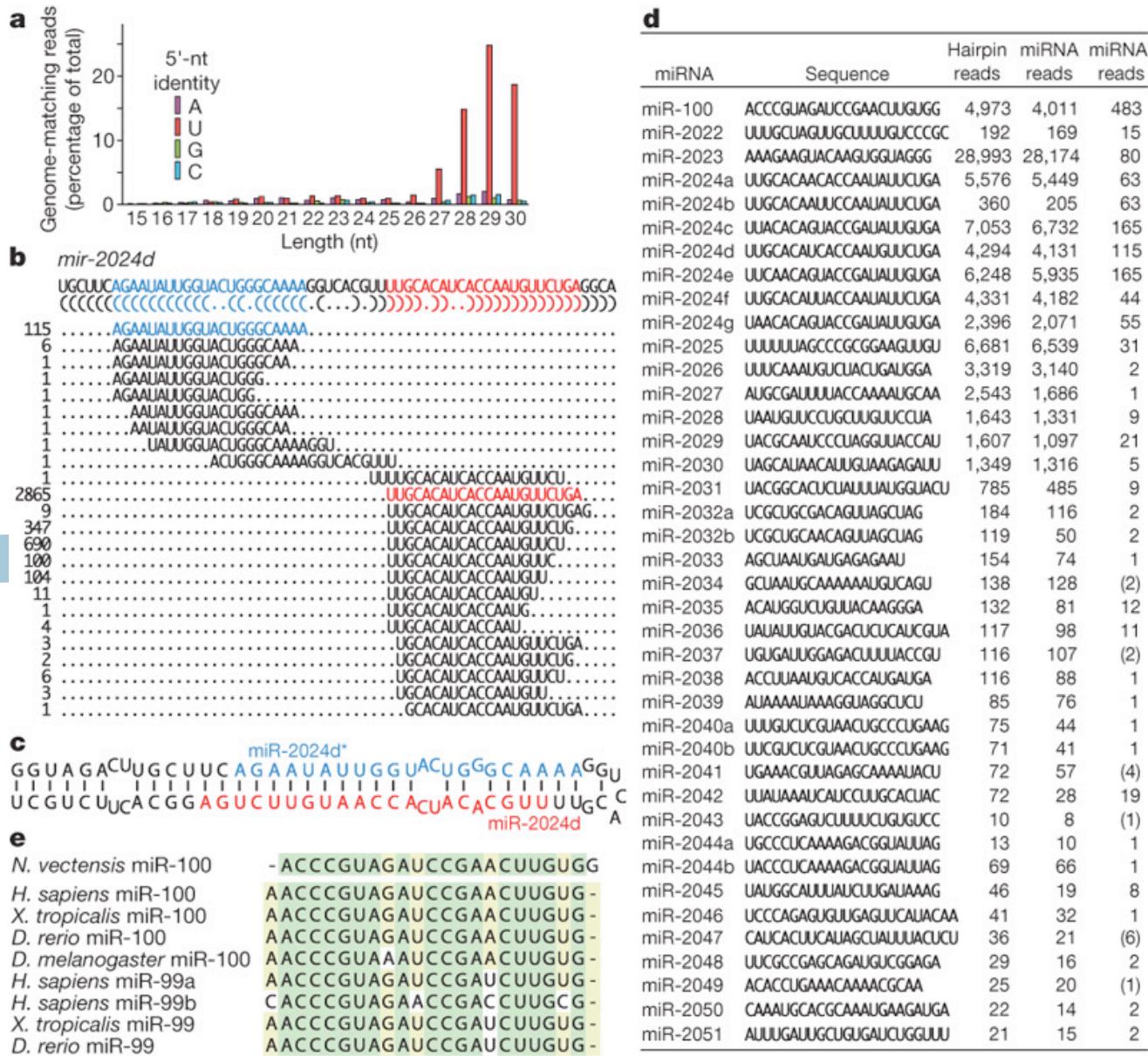
Pathway	Type I novelty	Type II novelty	Type III novelty	Ancient gene
Integrin signaling	Integrin-alpha; caveolin	Collagen; Integrin-β; Fak; Jun	Calpain	Talin; vinculin; paxillin; Ras; Grb2; SoS; Rap; ERK; MEK, Crk
Wnt signaling	Wnt; secreted frizzled related factors; frizzled; strabismus/van gogh	Dickkopf; arrow; dishevelled; axin		β-catenin; GSK3; APC; TCF/LEF; groucho
TGFβ signaling	Dpp/BMP; activin; gremlin; chordin; follistatin; R-SMAD; I-SMAD; co-SMAD	Type I receptors: TGFβR1, BMPR1A; ATF/JunB; snoN	Tolloid/BMP1	Type II receptors: ACVR2, BMPR2
Notch signaling	Numb; hairy/E(spl)	Notch		Jagged; deltex; fringe; presenilin; ADAM10; nicastrin; furin; Aph1; PEN2; mastermind
Ephrin signaling		Ephrin; Fak	Eph (receptor)	Abl/SYK
Insulin signaling	Insulin	Insulin receptor substrate; phosphoinositide-3-kinase, catalytic	Insulin receptor/IGF; phosphoinositide-3-kinase, class 2	phosphoinositide-3-kinase, class 3; phosphoinositide-3-kinase, regulatory subunit; 3-phosphoinositide-dependent protein kinase-1; PTEN
FGF signaling	FGF; Shc	Raf homolog serine/threonine-protein kinase; Ras GTPase activating protein	FGFR; RAS protein activator; phospholipase C-γ; phosphoinositide-3-kinase, class 2; protein kinase C ϵ	MAPK; phosphoinositide-3-kinase, class 3; Grb2; Protein kinase C; SoS; Rac
Cytokine signaling	Inositol 1,4,5-triphosphate receptor; SOCS; arrestin; guanine nucleotide binding protein γ; regulator of G-protein signaling; REL/NFKB; NFAT	Adenylate cyclase 5/6; STAT5; ATF/Jun	CDC42 binding protein kinase	MAPK; Rho kinase; Rho

Evolution of Cells and Tissues

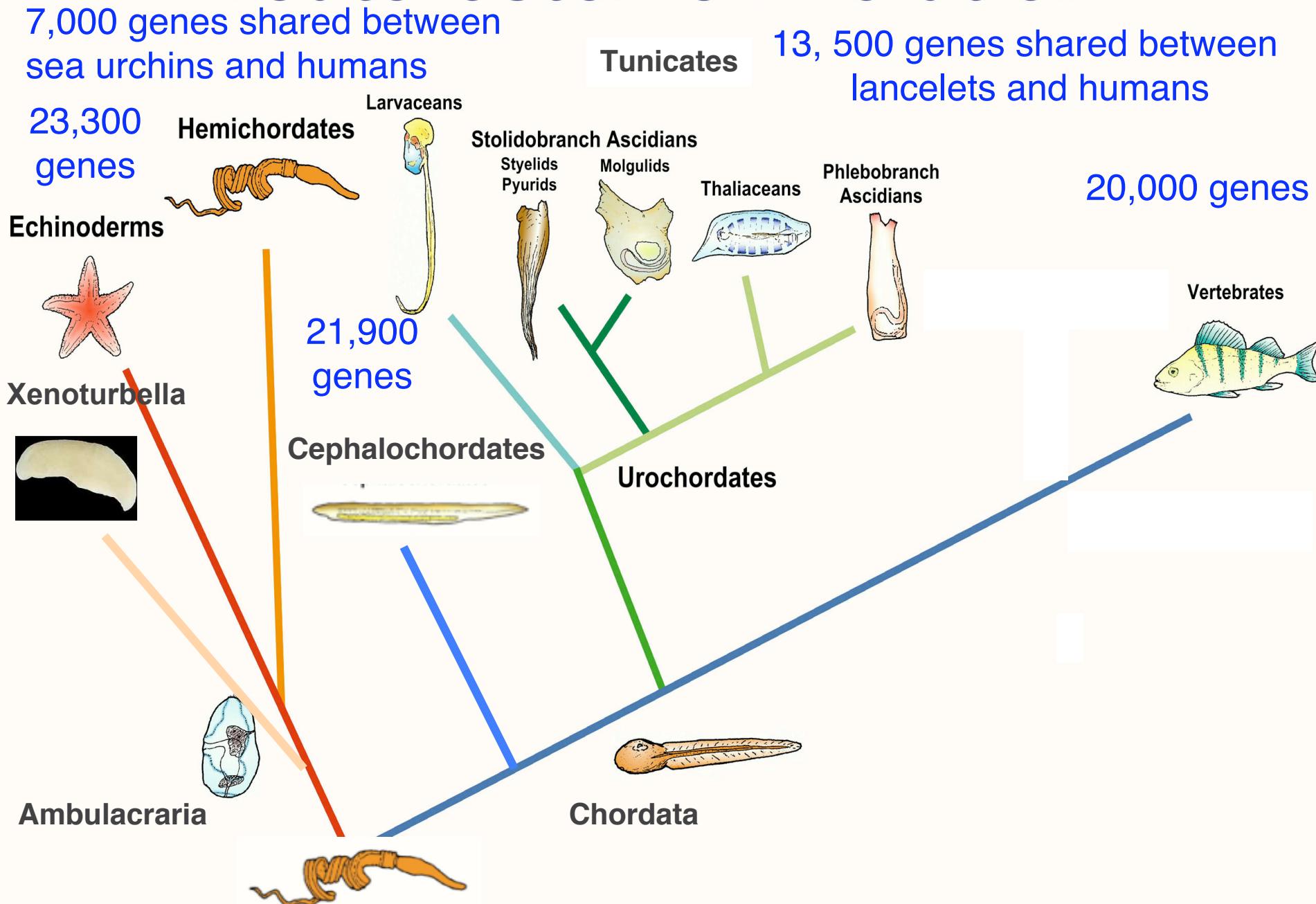
Putnam et al. (2007) Science 317: 86-94.

Process	Type I novelty	Type II novelty	Type III novelty	Ancient gene
Neurogenesis	Hes; Gcm; Ephrin; netrin; semaphoring; dachsund; ski oncogene	Notch; NGFR; Dsh; Arx; CREB/ATF; neuralized	Neuropilin; Lhx; ephrin receptor	Single-minded/HIF; achaete-scute; elav; Emx; Otp; Jagged; Deltex; Irx; Gli, Otx/Phox; stonal/neuroD/neuroG; reticulon
Synaptic transmission	Nitric oxide synthase (neuronal) adapter protein; DOPA-β monoxygenase; calcium channel voltage-dependent β; syntrophin; synaptophysin; dystrophin; potassium large conductance calcium-activated channel, subfamily M β	Cholinergic receptor, nicotinic; neurexin	K-voltage gated channel; discs large	Glutamate receptor; synaptotagmin; intersectin; synapsin; neuroligin/CES; syntaxin; glutamate transporter
Extracellular matrix	Netrin; dermatopontin; semaphorin; glycan; stereocilin	Collagen; spondin; laminin	Nidogen; stabilin; neuropilin; matrix metalloprotease; thrombospondin	Leprecan; microfibrillar-associated protein
Cell junction	par-6	Tight junction protein		Salvador
Muscle contraction	Voltage-dependent calcium channel β, β-sarcoglycan, β-dystrobrevin	Cholinergic receptor, nicotinic; nebulin; tropomyosin; calponin/transgelin	Voltage-dependent calcium channel α2/δ subunit; inositol triphosphate receptor; calcium activated potassium channel slowpoke	Phosphorylase kinase; myosin light chain cytoplasmic; calcium channel alpha subunit; cGMP-dependent protein kinase; calcium/calmodulin-dependent kinase II; myosin regulatory light chain
Apoptosis	TNF5/10/11; Bcl2; BOK; GULP; CRADD; caspase 8/10; growth arrest and DNA-damage-inducible; DNA fragmentation factor 40-kD subunit; interleukin enhancer-binding factor 3; FMR	Neuronal apoptosis inhibitory protein; CARD9/11	NGFR; SRGAP; calpain	TNFRSF; TRAF; scavenger receptor class B; huntingtin interacting protein; programmed cell death 1/5; Bcl2-associated athanogene; Akt; SUMO; defender against cell death 1; apoptosis-inducing factor-like mitochondrion-associated inducer of death; death-associated protein kinase
Transcription factors	L3MBT; T-Box; Nuclear hormone receptor; SMAD; dachsund; gcm; NFAT; nuclear respiratory factor; SKI family; sprouty; AP-2; onecut; MAF-related	CBP/p300; ETO/MTG8/Nervy; groucho; Jun; Myt1; runt; STAT	Hairless; nuclear protein 95; LIM homeobox; CCAAT enhancer binding; aryl hydrocarbon receptor related	Zic; Gli; homeobox; bHLH; achaete-scute; sox; retinoblastoma binding protein 5/8; NFKB-related; Krueppel C2H2 type zinc finger; Irx; Deltex; ataxin

The miRNAs of *N. vectensis*.



Deuterostome Evolution



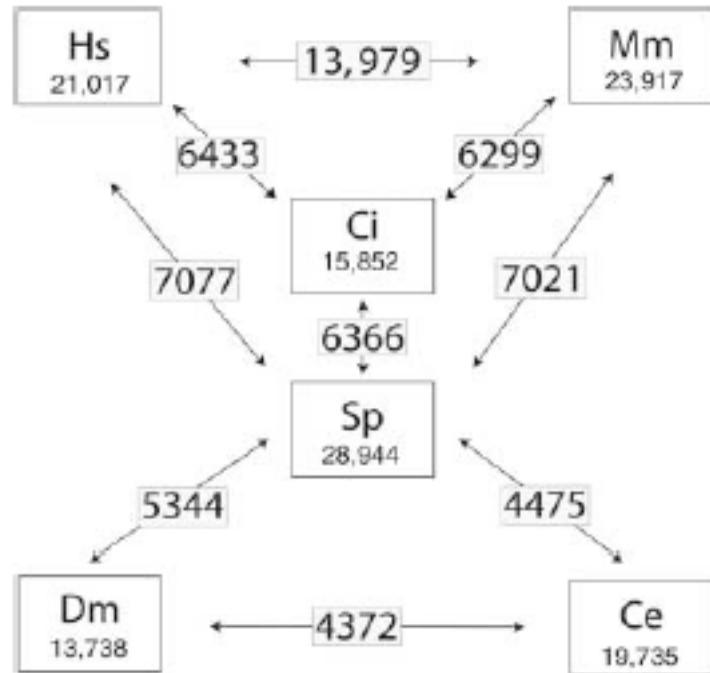


Fig. 2. Orthologs among the Bilateria. The number of 1:1 orthologs captured by BLAST alignments at a match value of $e = 1 \times 10^{-6}$ in comparisons of sequenced genomes among the Bilateria. The number of orthologs is indicated in the boxes along the arrows, and the total number of International Protein Index database sequences is shown under the species symbol. Hs, *Homo sapiens*; Mm, *Mus musculus*; Ci, *Ciona intestinalis*; Sp, *S. purpuratus*; Dm, *Drosophila melanogaster*; Ce, *Caenorhabditis elegans*.

Sea Urchin Genome Sequencing Consortium (2006)
 Science
 314: 941-952

Sea Urchin Genome
Sequencing
Consortium
(2006)
Science
314: 941-952

Process	Domain	PFAM no.	Sp	Hs	Dm	Ce	Nv
Cell cycle control	Cyclin_N	PF00134	15 (17)	21	11	7	7
	Cyclin_C	PF02984	7 (8)	12	4	5	4
	E2F_TDP	PF02319	3 (5)	11	3	4	3
	RB_A	PF01858	2	3	2	1	0
	RB_B	PF01857	2	3	2	1	0
	P53	PF00870	1	3	1	1	0
	Cullin	PF00888	7	9	8	7	4
Histone metabolism	Skp1	PF01466	1	3	5	21	1
	Histone [*]	PF00125	49	75	8	?	?
	Linker histone [*]	PF00538	5	8	2	?	?
	Nucleo-plasmin	PF03066	2	5	2	0	1
	NAP	PF00956	2	24	4	2	0
	HDAC	PF00850	8	11	5	8	3
RNA metabolism	DOT1	PF08123	1	1	1	6	1
	RRM_1	PF00076	140 (178)	245	126	99	41
	TUDOR	PF00567	15	13	15	8	7
	DEAD	PF00270	93 (125)	78	56	65	27
	LSM	PF01423	17	21	17	18	4
	KH-1	PF00013	28 (31)	36	28	28	5
	DSRM	PF00035	14 (15)	21	14	13	8
	3'-5'-Exo-nuclease	PF01612	13 (15)	5	5	9	5
	Exonuc_X-T	PF00929	9 (11)	15	7	10	5
	Caspase	PF00656	31 (33)	14	7	4	5
Apoptosis [†]	BIR	PF00653	4 (7)	8	4	1	4
	Bcl-2	PF00452	10	11	2	1	7
	TNFR_c6	PF00020	8 (9)	8	1 (no DD)	1 (no DD)	2
	NACHT	PF05729	129 (145)	18	1	1	2
	NB-ARC	PF00931	3	1	1	1	0
	DEATH	PF00531	47 (101)	30	9	6	6
	DED	PF01335	4 (5)	7	1	0	5
	CARD	PF00619	5 (10)	20	1	0	8

- Yellow box: Complexity intermediate between that in vertebrates and protostome invertebrate model organisms
- Green box: Complexity greater than that found in other model organisms
- Red box: Complexity lower than that found in other model organisms

Branchiostoma floridae - lancelet



Branchiostoma floridae - lancelet

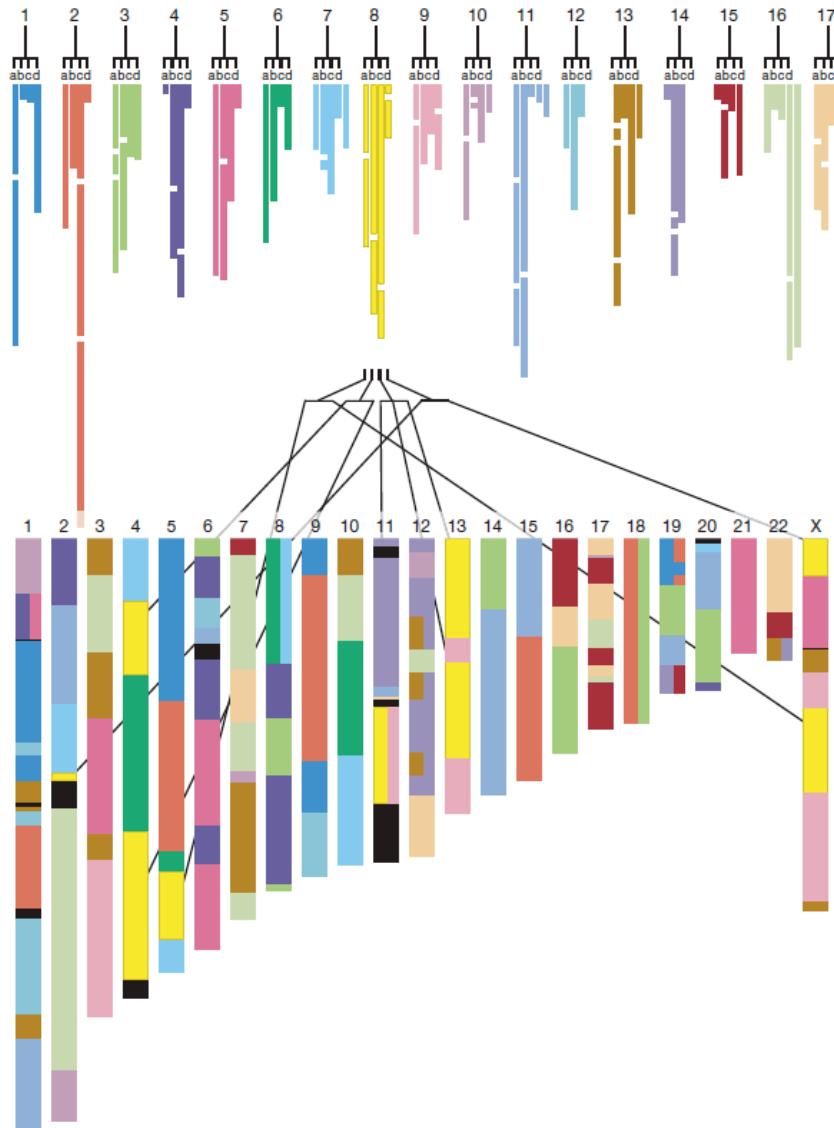
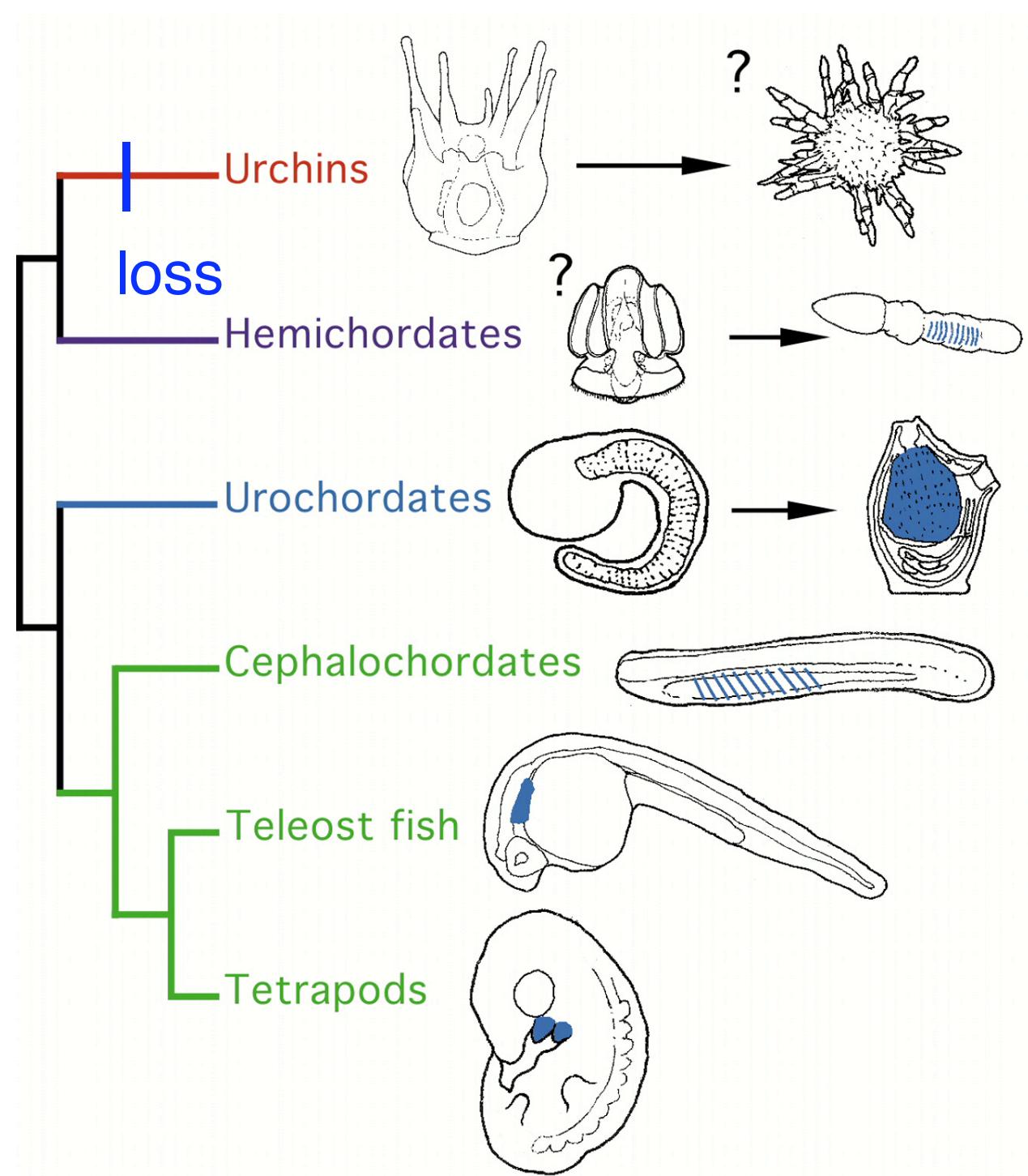


Figure 3 | Quadruple conserved synteny. Partitioning of the human chromosomes into segments with defined patterns of conserved synteny to amphioxus (*B. floridae*) scaffolds. Numbers 1–17 at the top represent the 17 reconstructed ancestral chordate linkage groups, and letters a–d represent

the four products resulting from two rounds of genome duplication. Coloured bars are segments of the human genome, shown grouped by ancestral linkage group (above), and in context of the human chromosomes (below).

Pax 1/9 expression



Brachyury T expression in metazoans

notochord

gut

