

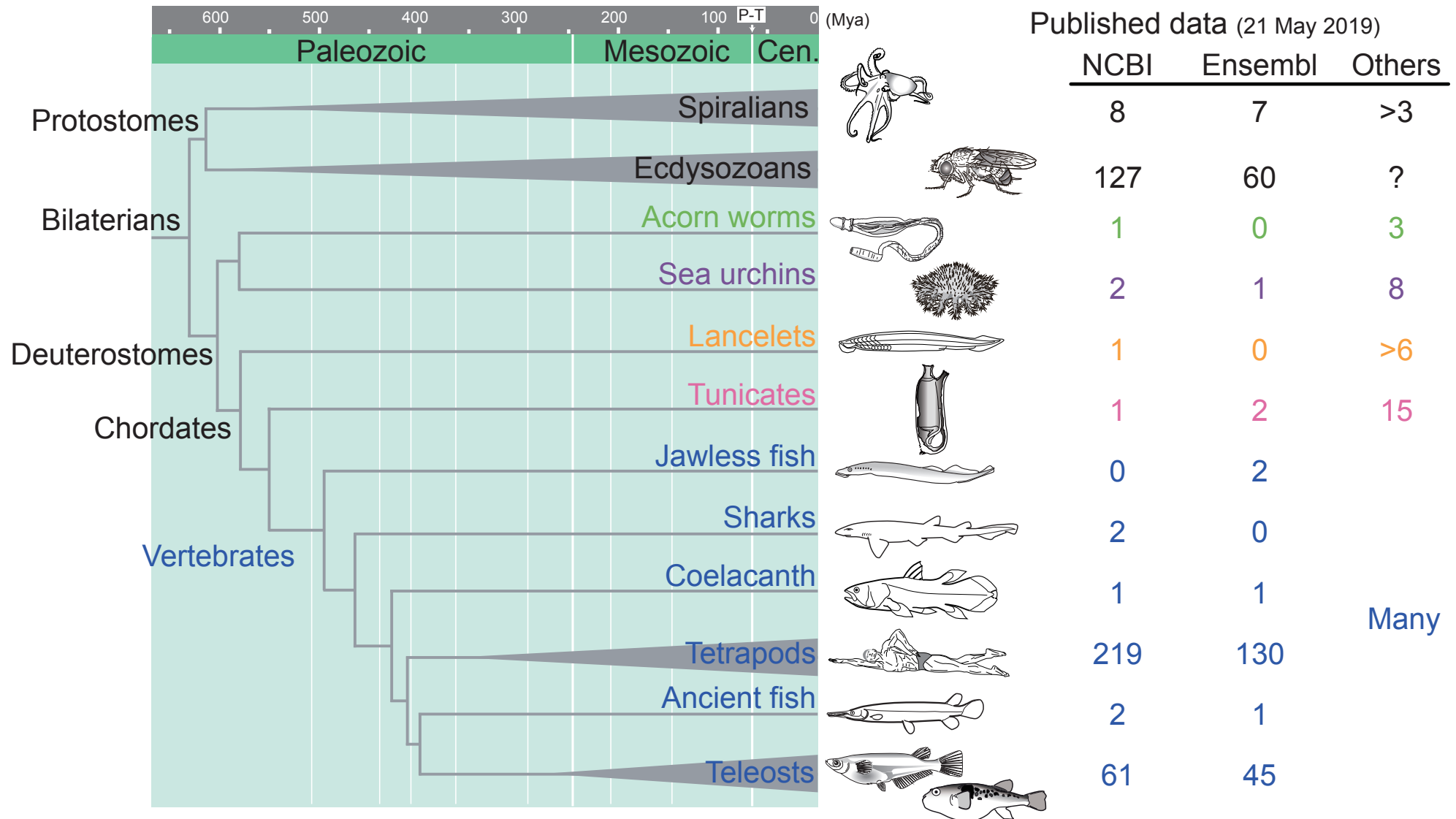
ORTHOSCOPE: a web tool for estimating the origins and the functions of bilaterian protein-coding genes

Jun Inoue

Inoue J, Satoh N. 2019. ORTHOSCOPE: an automatic web tool for phylogenetically inferring bilaterian orthogroups with user-selected taxa. *Molecular Biology and Evolution* 36: 621–631.

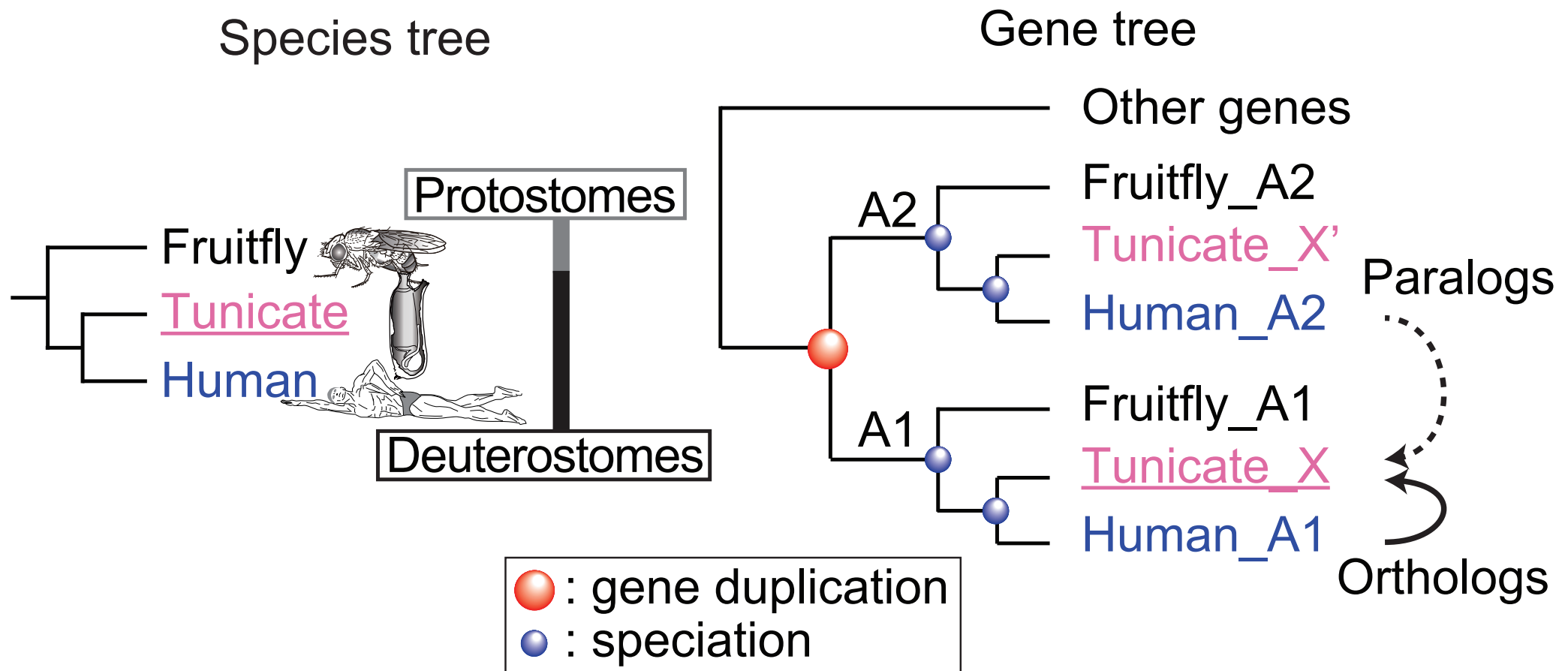
5 July 2026

Genome data are accumulating



- Major lineages => purposeful taxonomic sampling.
- Genome data are scattered among databases.
- Gene functions are not known in non-model organisms.

Inferring functions using orthologs

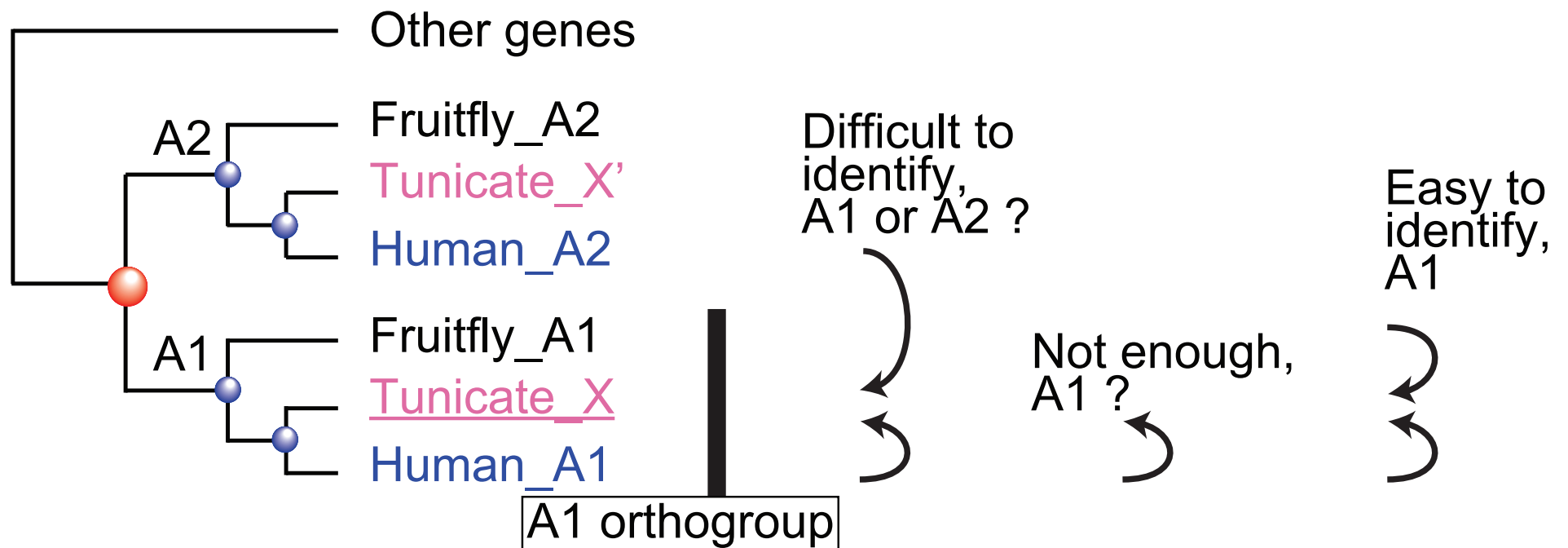


Given that orthologs are genes derived by speciation, they are used to infer gene functions, as the first step. However,

Inferring functions using orthogroups

Orthogroup:
a set of genes derived from
a single gene in the last common ancestor
of all the species being considered.

=> Considering **Tunicate** and **Human**,



Orthogroup clearly shows candidate function.

Aim

- No tool for inferring gene functions based on gene trees.
 - ➔ A web tool for infer gene functions by identifying orthogroups.
- Genome data are scattered among databases.
 - ➔ Integrate databases (gene models).



ORTHOSCOPE

Orthogroup identification in ORTHOSCOPE

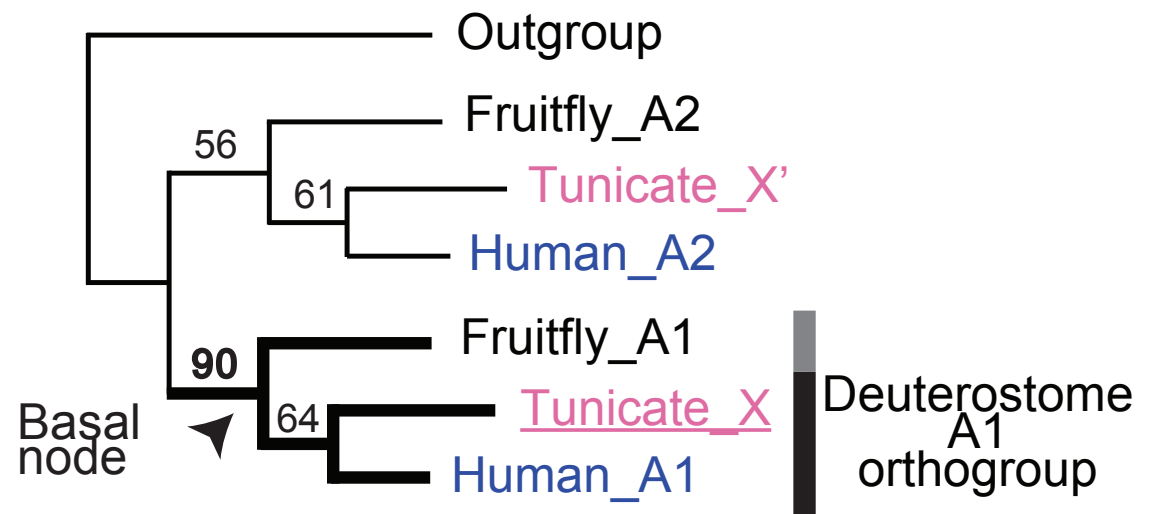
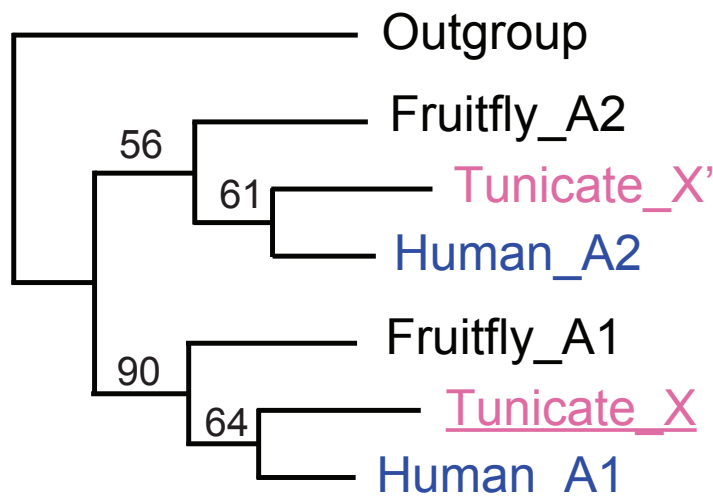
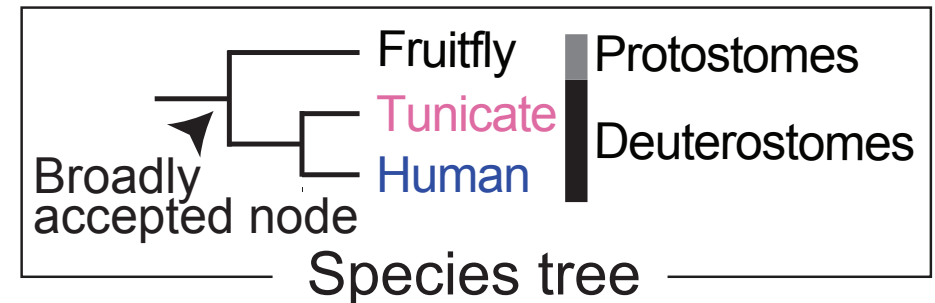
Candidate selection

BLAST Tunicate_X vs Other animals (Gene models)

Gene tree estimation

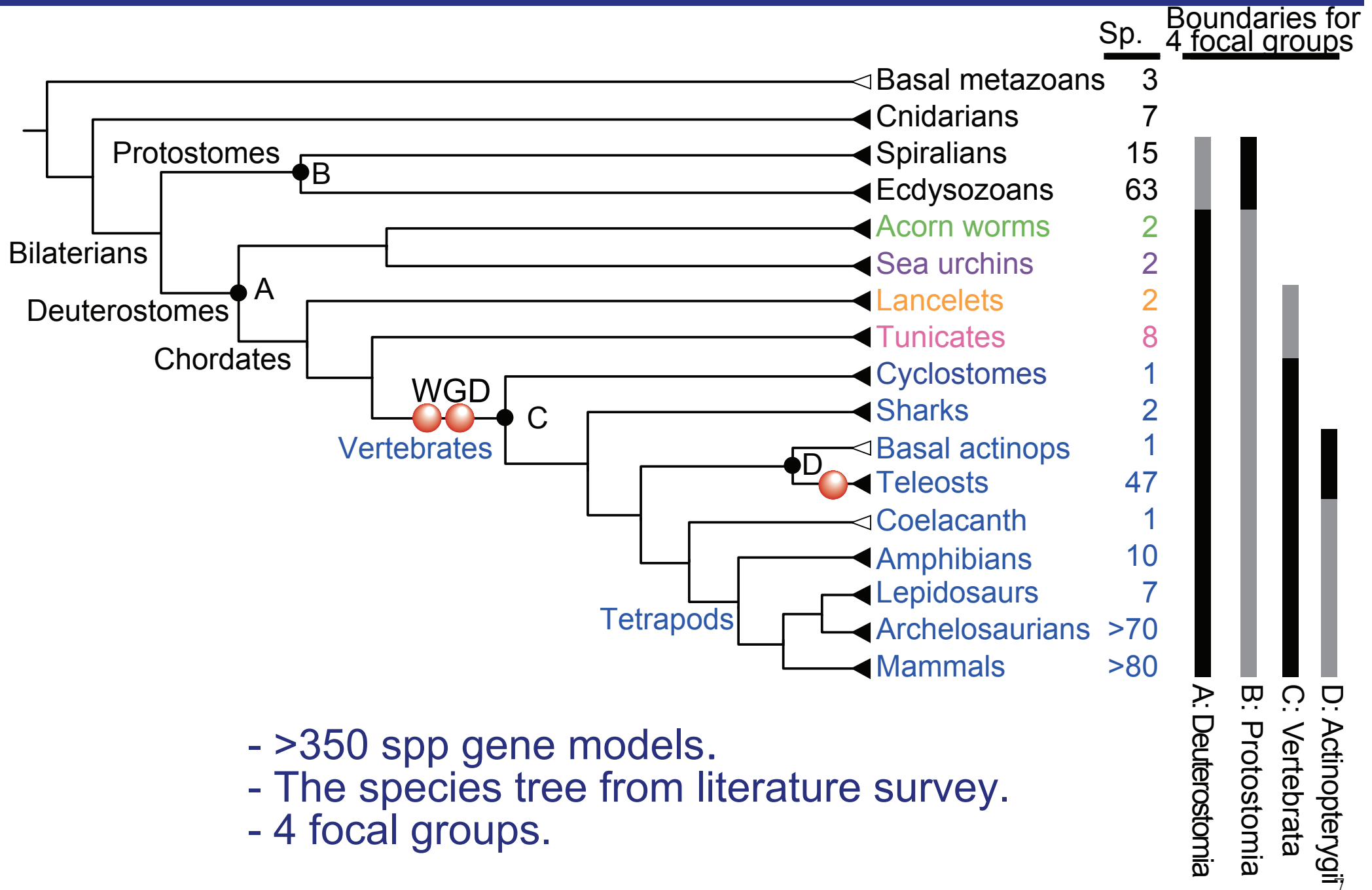


Comparison with the species tree



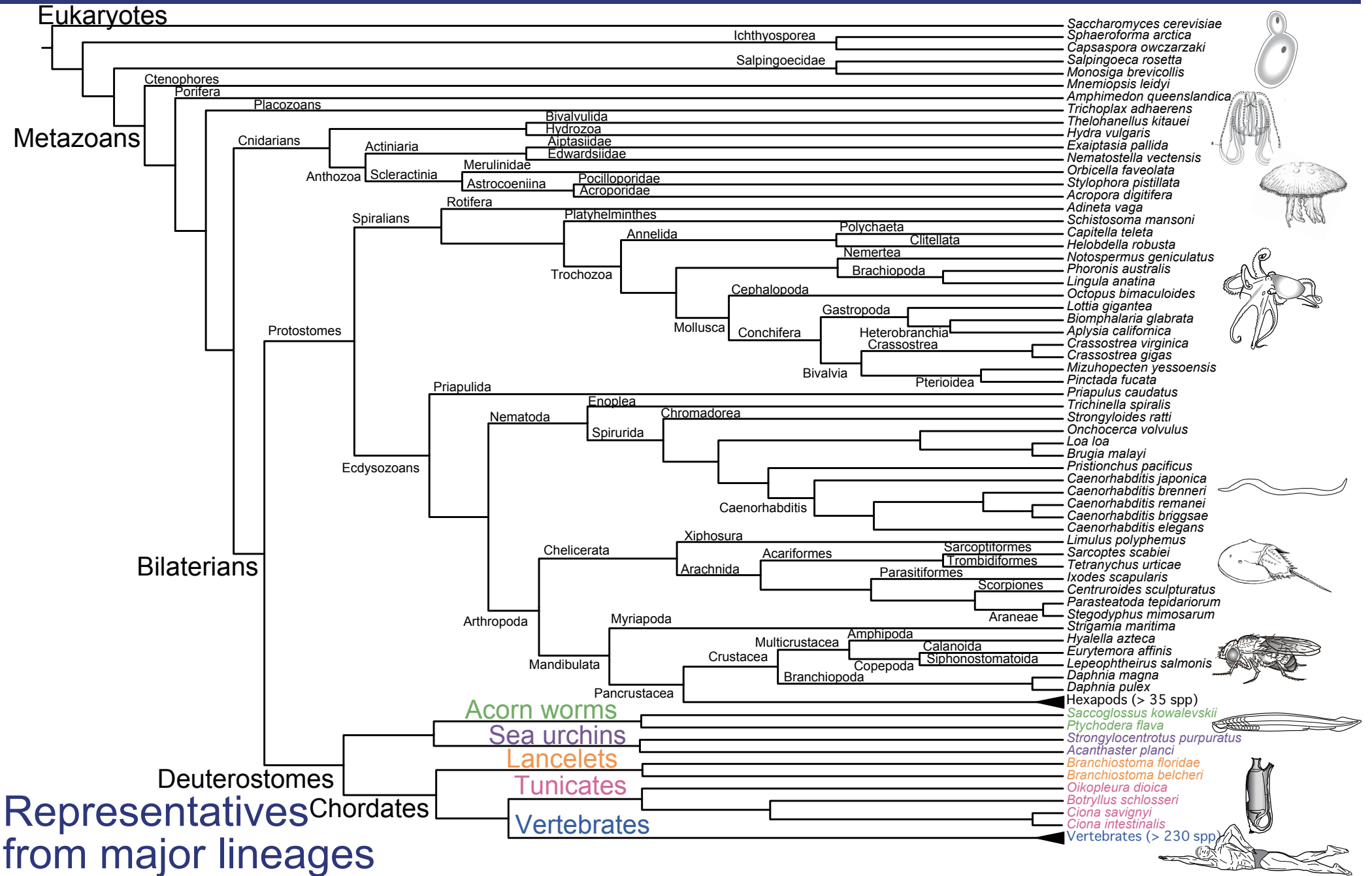
- With broadly accepted nodes, identify orthogroup.
- BS value of basal node for accuracy evaluation.

ORTHOSCOPE database



- >350 spp gene models.
- The species tree from literature survey.
- 4 focal groups.

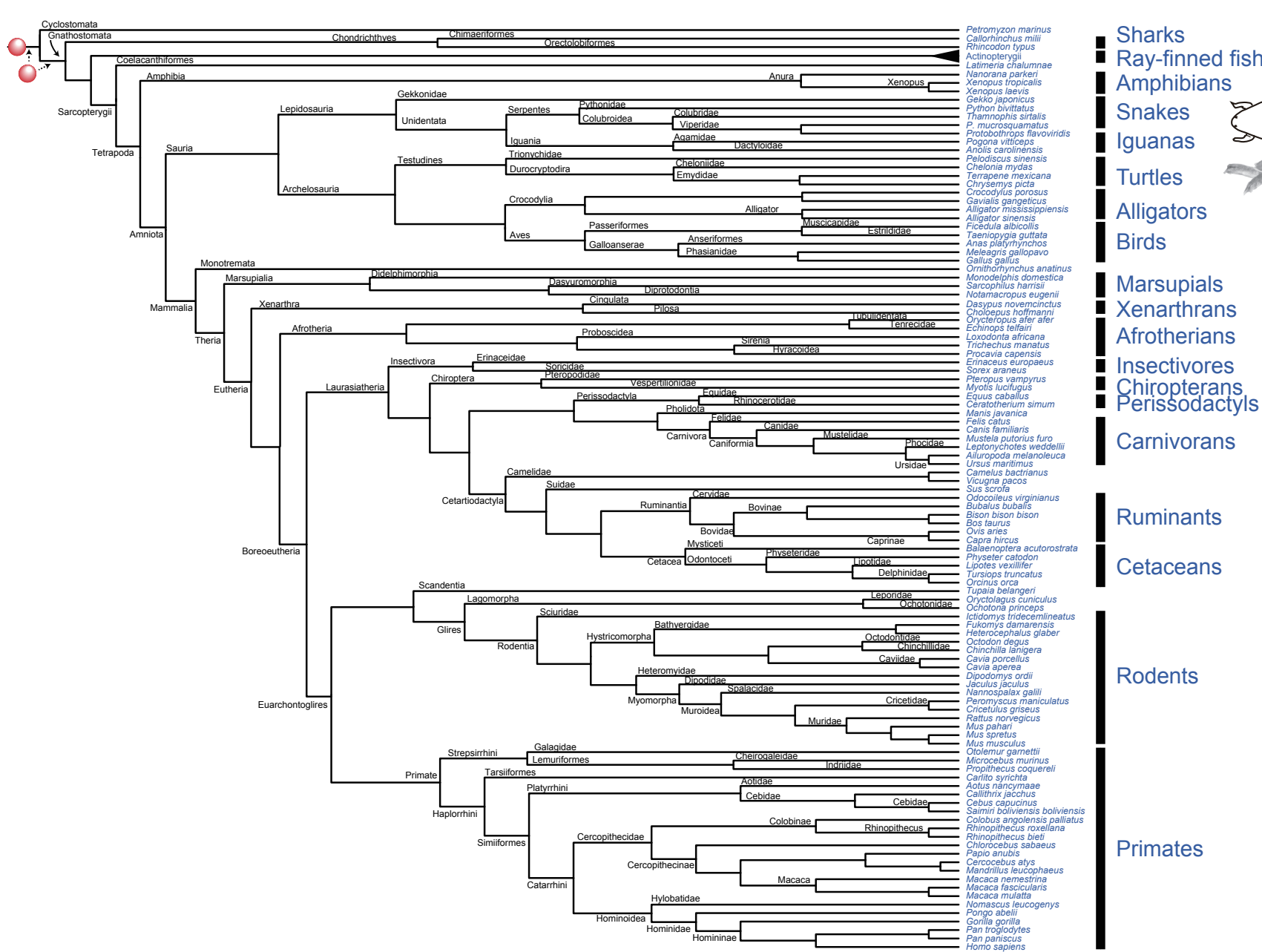
Metazoans



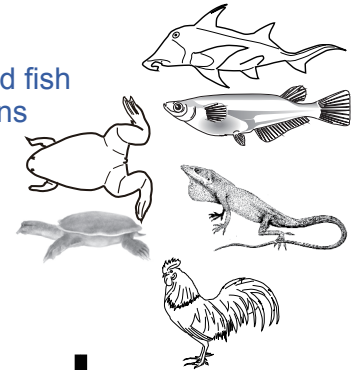
- Saccharomyces cerevisiae*
- Sphaeroforma arctica*
- Capsaspora owczarzaki*
- Salpingoeca rosetta*
- Monosiga brevicollis*
- Mnemiopsis leidyi*
- Amphimedon queenslandica*
- Trichoplax adhaerens*
- Thelohanellus kitauei*
- Hydra vulgaris*
- Exaiptasia pallida*
- Nematostella vectensis*
- Orbicella faveolata*
- Stylophora pistillata*
- Acropora digitifera*
- Adineta vaga*
- Schistosoma mansoni*
- Capitella teleta*
- Helobdella robusta*
- Notospermus geniculatus*
- Phoronis australis*
- Lingula anatina*
- Octopus bimaculoides*
- Lottia gigantea*
- Biomphalaria glabrata*
- Aplysia californica*
- Crassostrea virginica*
- Crassostrea gigas*
- Mizuhopecten yessoensis*
- Pinctada fucata*
- Priapulus caudatus*
- Trichinella spiralis*
- Strongyloides ratti*
- Onchocerca volvulus*
- Loa loa*
- Brugia malayi*
- Pristionchus pacificus*
- Caenorhabditis japonica*
- Caenorhabditis brenneri*
- Caenorhabditis remanei*
- Caenorhabditis briggsae*
- Caenorhabditis elegans*
- Limulus polyphemus*
- Sarcoptes scabiei*
- Tetranychus urticae*
- Ixodes scapularis*
- Centruroides sculpturatus*
- Parasteatoda tepidariorum*
- Stegodyphus mimosarum*
- Strigamia maritima*
- Hyalella azteca*
- Eurytemora affinis*
- Lepeophtheirus salmonis*
- Daphnia magna*
- Daphnia pulex*
- Hexapods (> 35 spp)**
- Saccoglossus kowalevskii*
- Ptychodera flava*
- Strongylocentrotus purpuratus*
- Acanthaster planci*
- Branchiostoma floridae*
- Branchiostoma belcheri*
- Oikopleura dioica*
- Botryllus schlosseri*
- Ciona savignyi*
- Ciona intestinalis*
- Vertebrates (> 230 spp)**

Representatives
from major lineages

Vertebrates



Mammals



Use of gene models

Gene ID	Gene name	Transcript ID	Length	
ENSG00000188782	CATSPER4	ENST00000338855	1516	←
ENSG00000188782	CATSPER4	ENST00000456354	1912	
ENSG00000188782	CATSPER4	ENST00000518899	1831	
ENSG00000181467	RAP2B	ENST00000323534	8351	←
ENSG00000164458	TBXT	ENST00000296946	2436	←
ENSG00000164458	TBXT	ENST00000366871	2250	
ENSG00000164458	TBXT	ENST00000366876	1497	
ENSG00000164458	TBXT	ENST00000461348	902	

The longest transcript from a single locus is used.

https://orthoscope.jp

ORTHOSCOPE: deuterostome orthogroup

Gene tree and orthogroup estimation using a species tree (< 5 min with 20-30 spp).

[Instruction](#)

Support: Safari(latest), Firefox, Chrome

Ver.1

Analysis group

[Actinopterygii](#)

[Vertebrata](#)

Deuterostomia

[Protostomia](#)

Status

Ready.

(<15 min)

Execute

Mode

☐ Tree search only

☒ Search/rearrangement

Upload file

Coding or amino acid sequence set
(fasta)

Species tree
(newick)

ファイルを選択 ファイル未選択

ファイルを選択 ファイル未選択

☐ Amino acid ☒ DNA [Example](#)

If not selected, [this tree](#) is used.

Coding sequences

Upload coding sequences (amino acid or DNA)

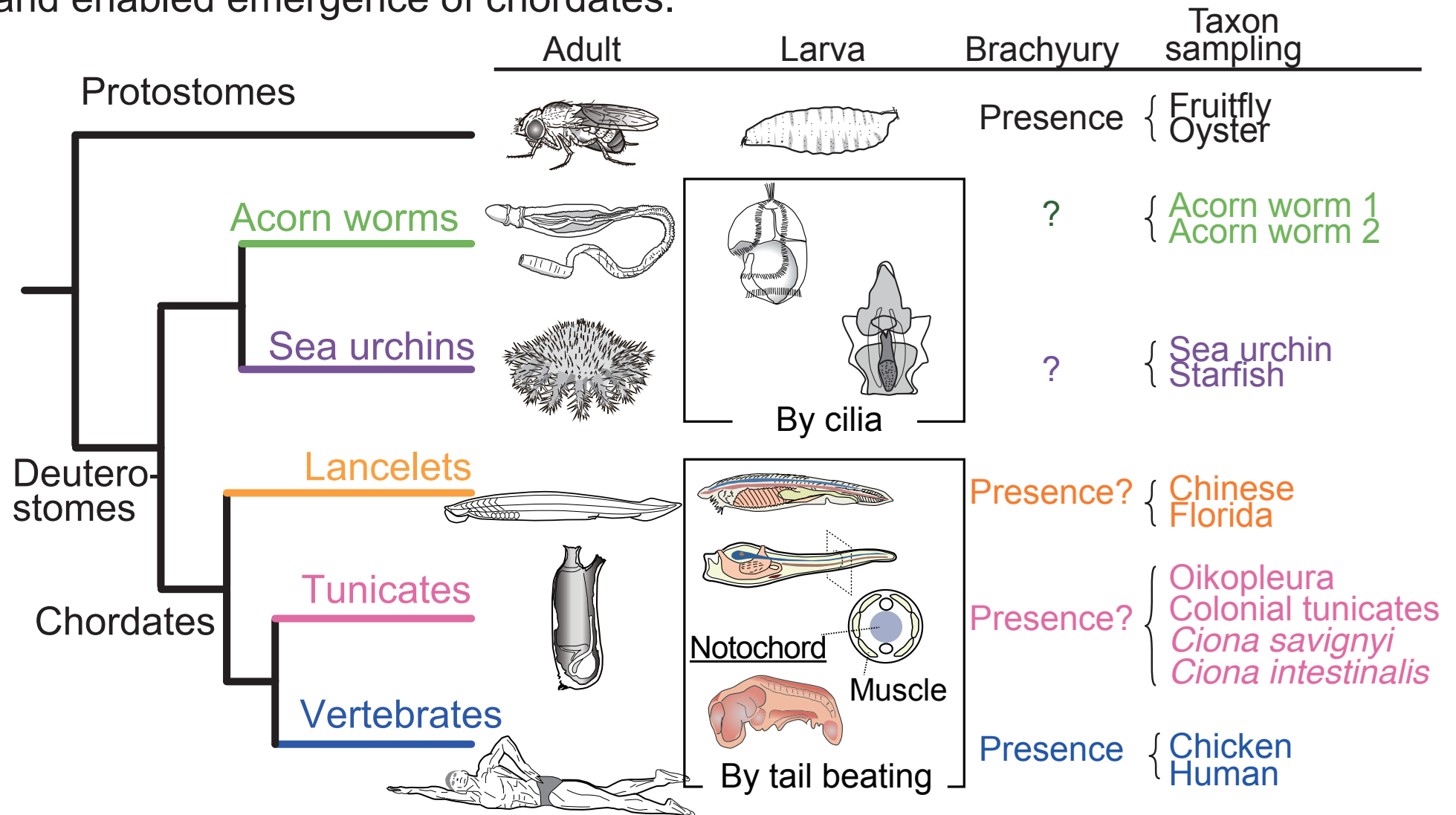
ORTHOSCOPE analysis - 1

Genes enabled emergence of chordates:
Brachyury gene

Inoue J, Yasuoka Y, Takahashi H, Satoh N. 2017.
The chordate ancestor possessed a single copy of
the *Brachyury* gene for notochord acquisition.
Zoological Letters. 3: 4.

Taxon sampling

Brachyury plays an essential role in notochord formation and enabled emergence of chordates.



How many *Brachyury* genes in chordate ancestor?

ORTHOSCOPE: deuterostome orthogroup

Gene tree and orthogroup estimation using a species tree (< 5 min with 20-30 spp).

Instructions

Support: Safari(latest), Firefox, Chrome

Ver.1.0.2

Focal group

[Actinopterygii](#)

[Vertebrata](#)

Deuterostomia

[Protostomia](#)

Status

Ready.

Execute

Mode

☐ Tree Search Only

☒ Search/Rearrangement

Upload file

Coding or amino acid sequence set
(fasta)

ファイル未選択

☐ Amino acid ☒ DNA [Example](#)

Species tree
(newick)

ファイル未選択

If not selected, [this tree](#) is used.

```
DeuterostomeBra.fas.txt Free Mode
~/Downloads/DeuterostomeBra.fas.txt
1 >Homo-sapiens_XXX1
2 ATGAGCTCCCCTGGCACCGAGAGCGCGGGAAAGAGCCTGCAGTACCGAGTGGACCACCTGCTGAGC
3 >Branchiostoma-floridiae_XXX
4 ATGAAGCAGACACCGGACCAGTTCTCGGTTAGCCACCTCCTGAGCGCCGTGGAGAGCGAGATCTCG
5 >Homo-sapiens_ENST00000389554.7_T-box-brain-1
6 ATGCAGCTGGAGCACTGCCTTCTCCTTCTATCATGCTCTCCAAGAAATTTCTCAATGTGAGCAGC
```

L: 1 C: 1 Text File - Unicode (UTF-8) - Unix (LF) - Saved: 2018/05/31 11:11:20 4,267 / 12 / 7 100%

1st query

Multiple query sequences can be used.

	☒ <i>Drosophila melanogaster</i>	Fruit fly	EnsMet38
	☐ <i>Drosophila melanogaster</i> 1	Fruit fly	RefSeq89
Xenacoelomorpha			
Xenoturbella	☐ <i>Xenoturbella bocki</i>	Xenoturbella	TSA
Nemertodermatida			
Acoela	☐ <i>Symsagittifera ruscifensis</i>	Acoela	TSA
	☐ <i>Hofstenia miamia</i>	Acoela	TSA
	☐ <i>Isodiametra pulchra</i>	Acoela	TSA
	☐ <i>Prosagittifera nankaiensis</i>	Acoela	OIST-S
Deuterostomia			
Hemichordata	☒ <i>Saccoglossus kowalevskii</i>	Helical acorn worm	OIST-R
	☐ <i>Saccoglossus kowalevskii</i> 1	Helical acorn worm	RefSeq89
	☒ <i>Ptychodera flava</i>	Yellow acorn worm	OIST-S
Echinodermata	☒ <i>Strongylocentrotus purpuratus</i>	Purple urchin	EnsMet38
	☐ <i>Strongylocentrotus purpuratus</i> 1	Purple urchin	RefSeq89
	☒ <i>Acanthaster planci</i>	Crown-of-thorns starfish (Okinawa)	OIST-S
	☐ <i>Acanthaster planci</i> 1	Crown-of-thorns starfish (Okinawa)	RefSeq89
	☐ <i>Acanthaster planci</i> <i>GBR</i>	Crown-of-thorns starfish (Great Barrier Reef)	OIST-S
Cephalochordata	☐ <i>Asymmetron lacayanum</i>	Lancelet	TSA
	☒ <i>Branchiostoma belcheri</i>	Belcher's lancelet	LanceletDB
	☐ <i>Branchiostoma belcheri</i> 1	Belcher's lancelet	RefSeq89
	☐ <i>Branchiostoma lanceolatum</i>	Mediterranean amphioxus	AmphioxusDB
	☒ <i>Branchiostoma floridae</i>	Florida lancelet	JGI
	☐ <i>Branchiostoma floridae</i> 1	Florida lancelet	RefSeq89
Urochordata	☒ <i>Dikoppleura dioica</i>	Pelagic tunicate	OikaBase
	☐ <i>Salpa thompsoni</i>	Salpa	TSA
	☐ <i>Molgula occidentalis</i>	—	aNISEED
	☐ <i>Molgula oculata</i>	Sea grape	aNISEED
	☐ <i>Botryllodes leachi</i>	Leache's ascidian	aNISEED
	☒ <i>Botryllus schlosseri</i>	Golden star tunicate	Scanford
	☒ <i>Ciona savignyi</i>	Pacific transparent sea squirt	Ens91
	☐ <i>Ciona savignyi</i> 1	Pacific transparent sea squirt	aNISEED
	☒ <i>Ciona intestinalis</i>	Transparent sea squirt	Ens91
	☐ <i>Ciona intestinalis</i> 1	Transparent sea squirt	RefSeq89
Vertebrata			
Myxiniiformes	☐ <i>Epiplatys burgeri</i>	Inshore hagfish	Ens91

High quality gene models
as defaults.

result3300

010_candidates_nucl.txt

020_treeRearranged.pdf

020_tree.pdf

100_2ndTree.zip

000_summary.txt

010_candidates_prot.txt

010_aln_nucl.html

今日 11:19

--

フォルダ

102 KB

8 KB

8 KB

8 KB

44 KB

36 KB

1.5 MB

BBEdit

Adobe

Adobe

ZIPアー

BBEdit

BBEdit

LIBRE

~/Downloads/result3300/000_summary.txt

Results

>Queries_used_in_the_analysis

Homo-sapiens_ENST00000296946.6_T-T-brachyury-transcrip

Branchiostoma-floridiae_BRAFL121413_NONE

Homo-sapiens_ENST00000389554.7_TBR1-T-box-brain-1

>BootstrapValue_OrthogroupMonophyly

The orthogroup monophyly is supported by 100% bootstra

>Orthogroup

Acanthaster-planci_oki15-190_NONE

Botryllus-schlosseri_botctg002459g63408_NONE

Branchiostoma-belcheri_BRBE102770F-t1_NONE

Branchiostoma-belcheri_BRBE102780R-t1_NONE

Branchiostoma-floridiae_BRAFL121413_NONE

Branchiostoma-floridiae_BRAFL279431_NONE

Ciona-intestinalis_ENSCINT000000001477.3_brachyury-tr

Ciona-savignyi_ENSCSAVT000000003855.1_NONE

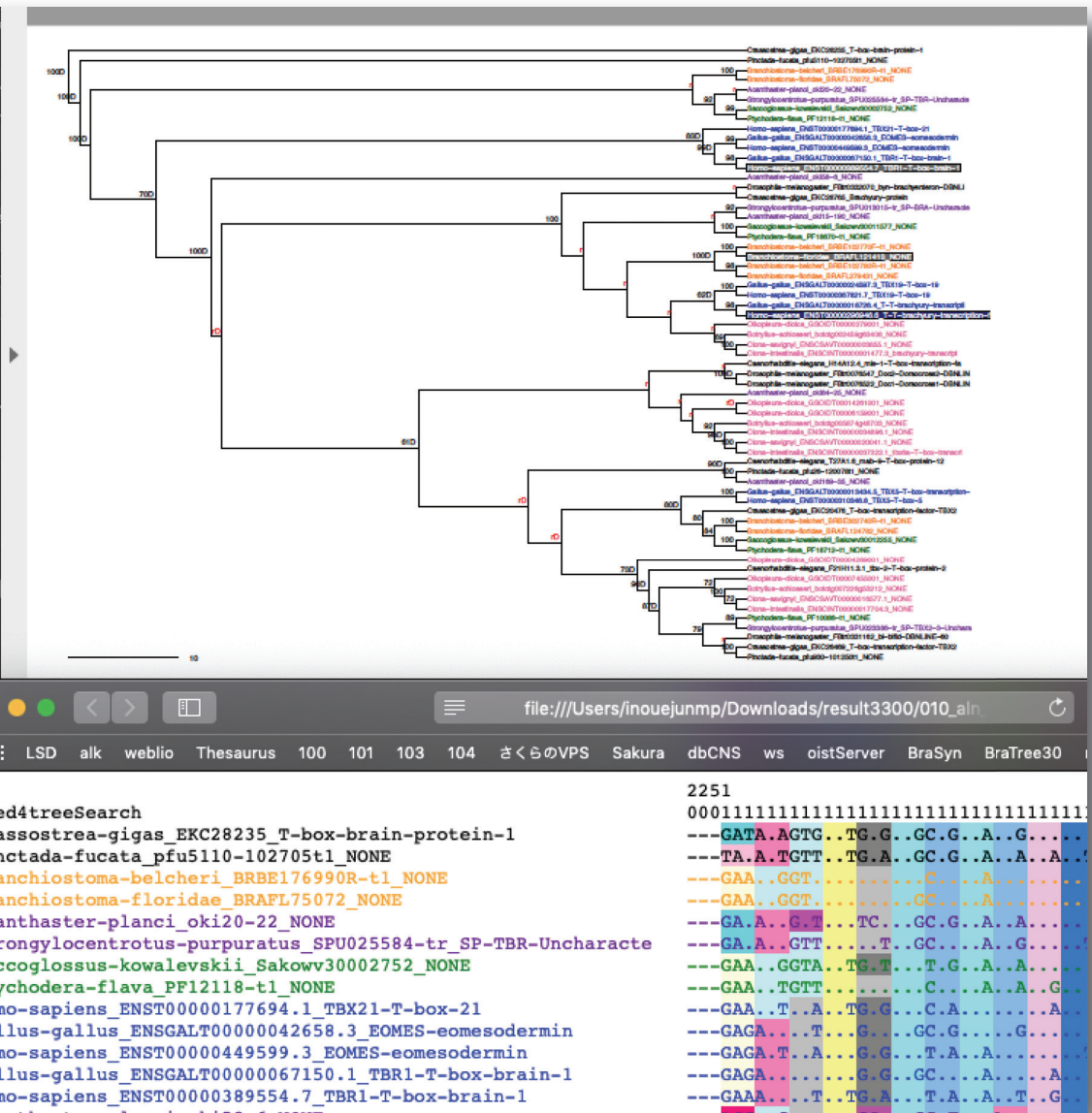
Crassostrea-gigas_EKC28765_Brachyury-protein

Drosophila-melanogaster_FBtr0332070_byn-brachyentero

Gallus-gallus_ENSGALT000000018726.4_T-T-brachyury-tra

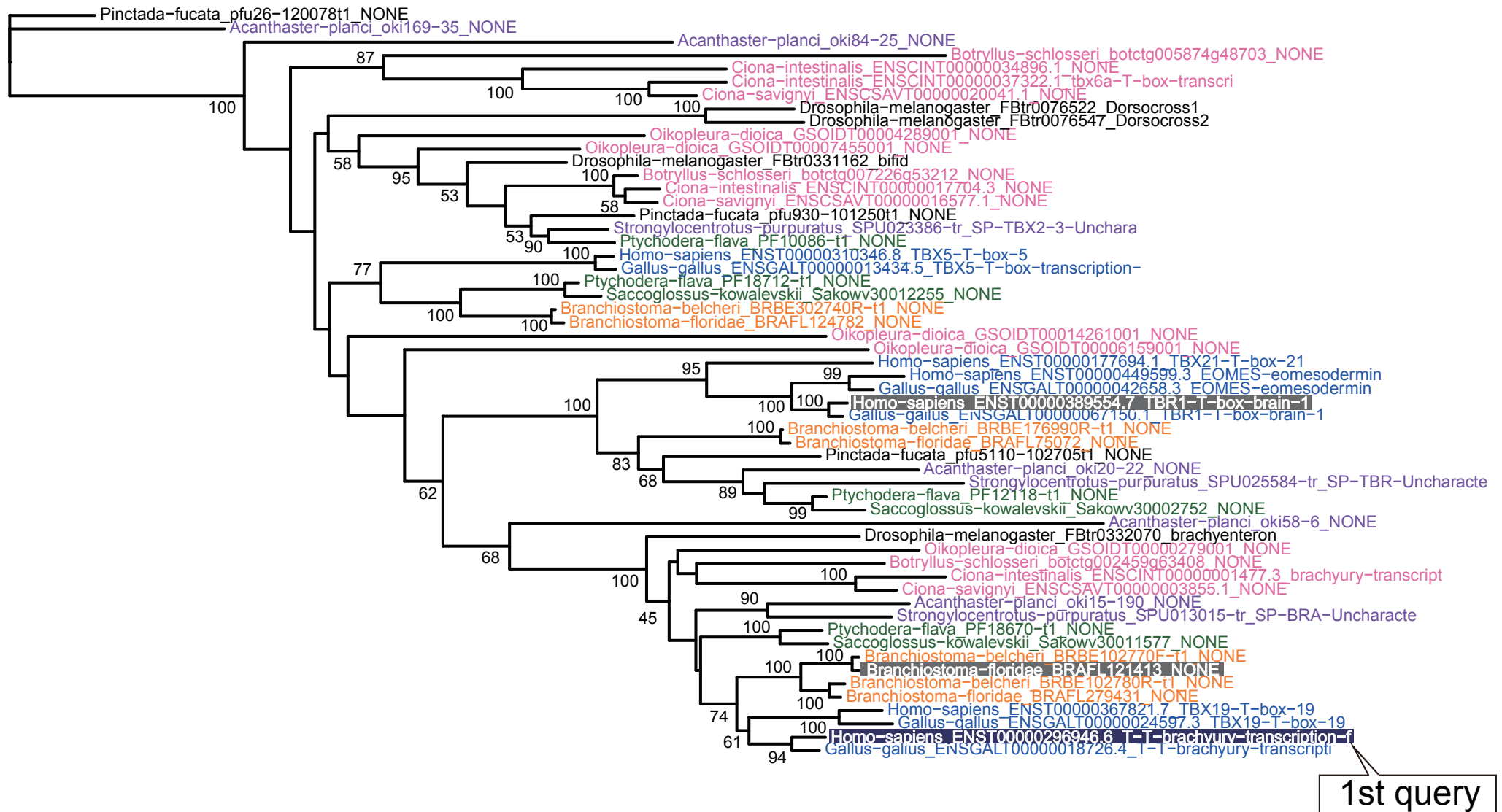
Gallus-gallus_ENSGALT000000024597.3_TBX19-T-box-19

Homo-sapiens_ENST00000296946.6_T-T-brachyury-transcrip

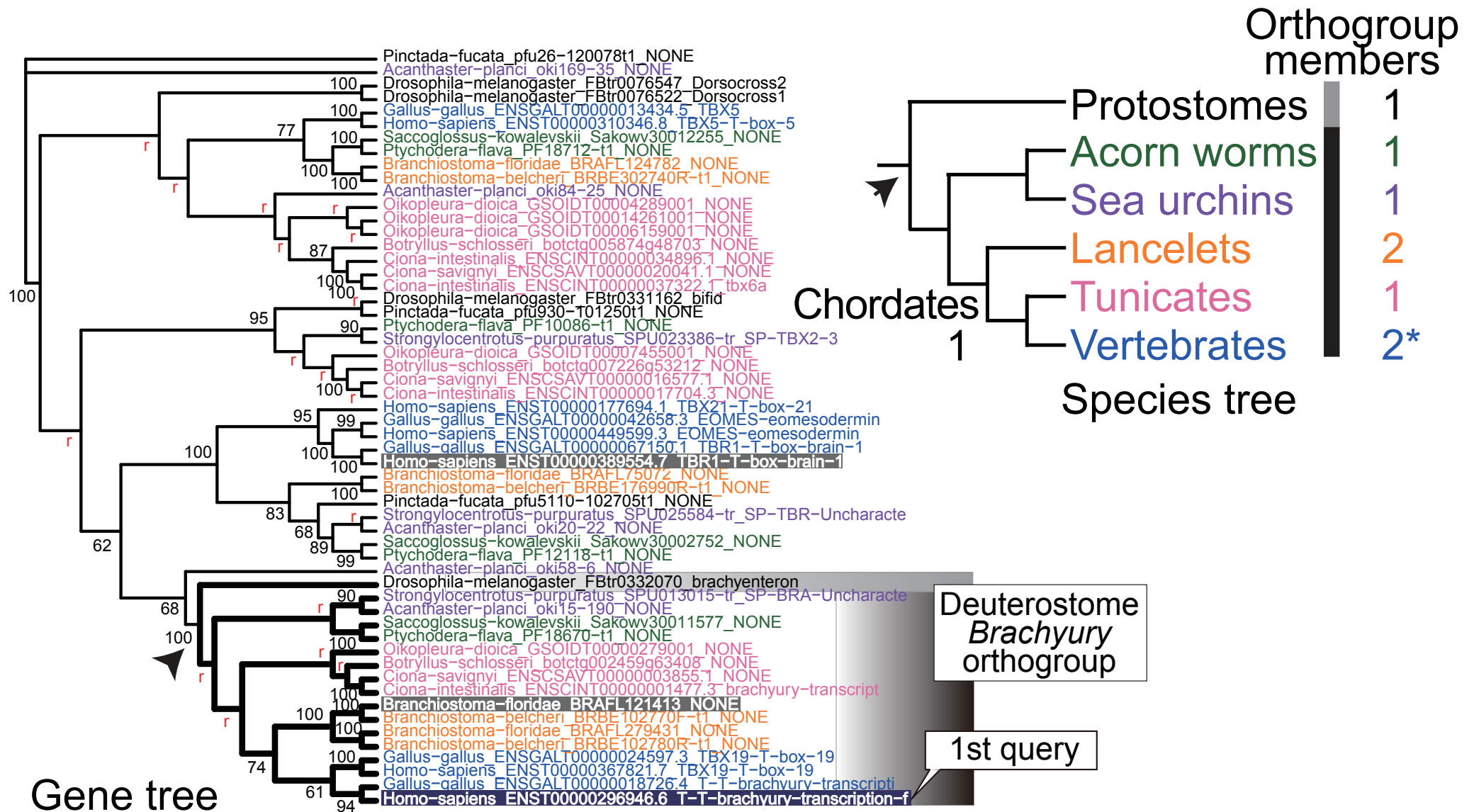


Results can be downloaded.

Results: NJ tree



Gene tree with identified orthogroup



Only 1 *Brachyury* gene in chordate ancestor.
=> New gene was not created for notochord acquisition.

Transcript ID

Drosophila-melanogaster_FBtr0332070_brachyenteron

Ensembl (<http://www.ensembl.org>)

Search

All species for

e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)



FB2019_03, released Jun 11, 2019

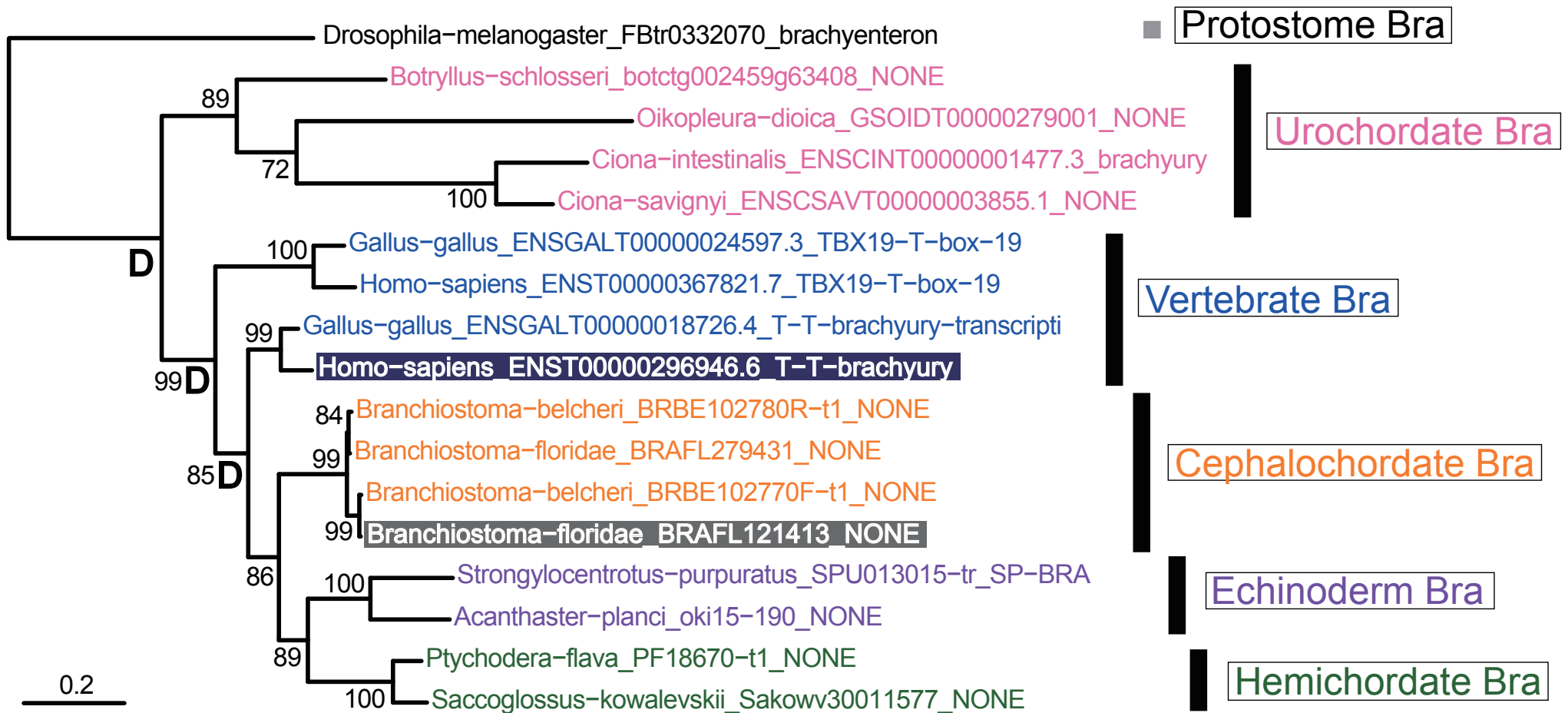
Gene: Dmel\byn

Home Tools Downloads Links Community Species About Help Archives J2G Jump

General Information

Symbol	Dmel\byn	Species	<i>D. melanogaster</i>
Name	brachyenteron	Annotation Symbol	CG7260
Feature Type	protein_coding_gene	FlyBase ID	FBgn0011723
Gene Model Status	Current	Stock Availability	6 publicly available
Gene Snapshot	brachyenteron (byn) encodes a T-domain transcriptional activator that acts in developmental specification, in particular it specifies posterior gut structures and a subset of posteriorly derived visceral muscles. [Date last reviewed: 2019-03-07]		

Detailed analysis using orthogroup members



Identify gene duplication nodes (D).

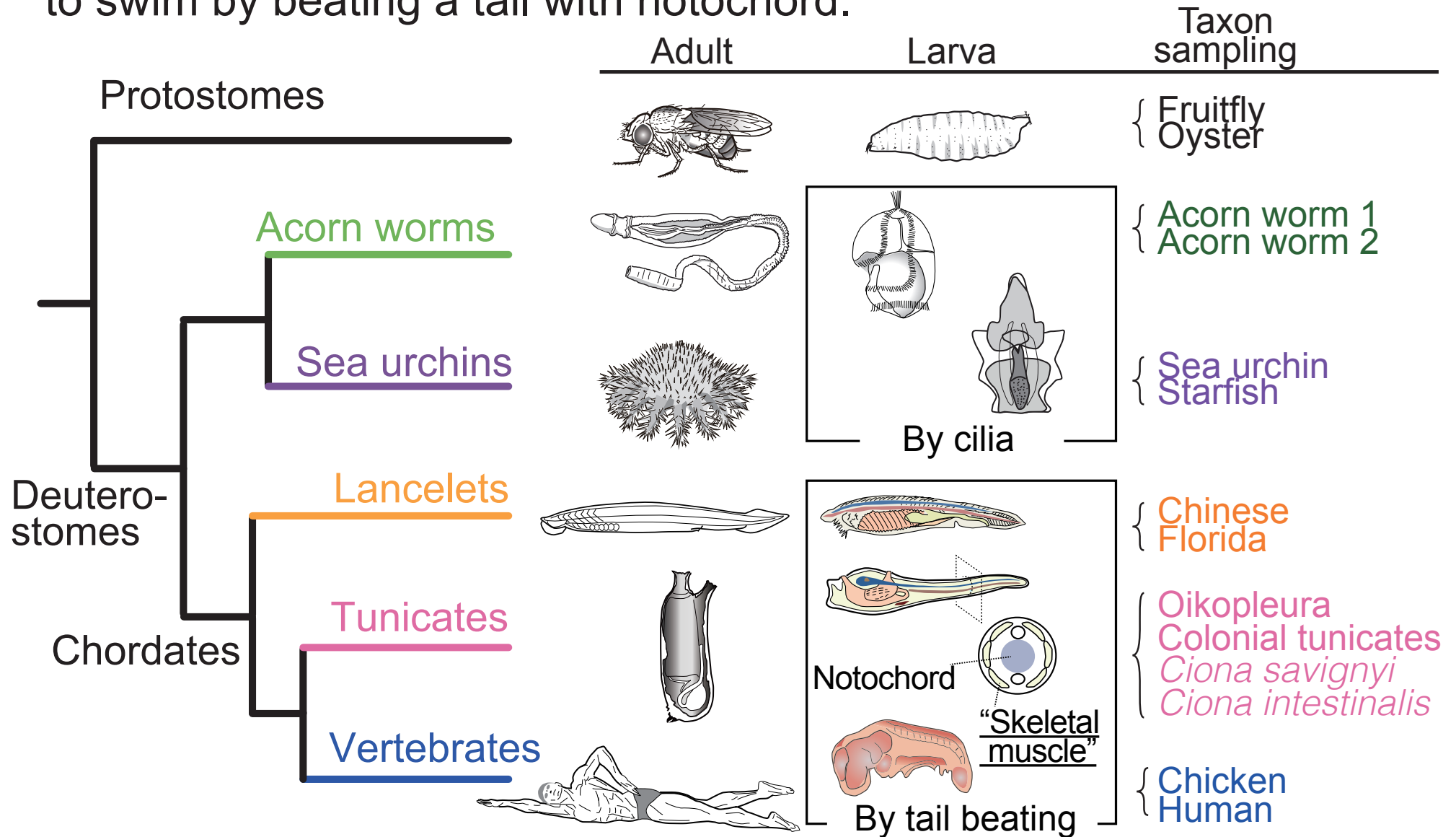
ORTHOSCOPE analysis - 2

Genes enabled emergence of chordates:
Muscle genes

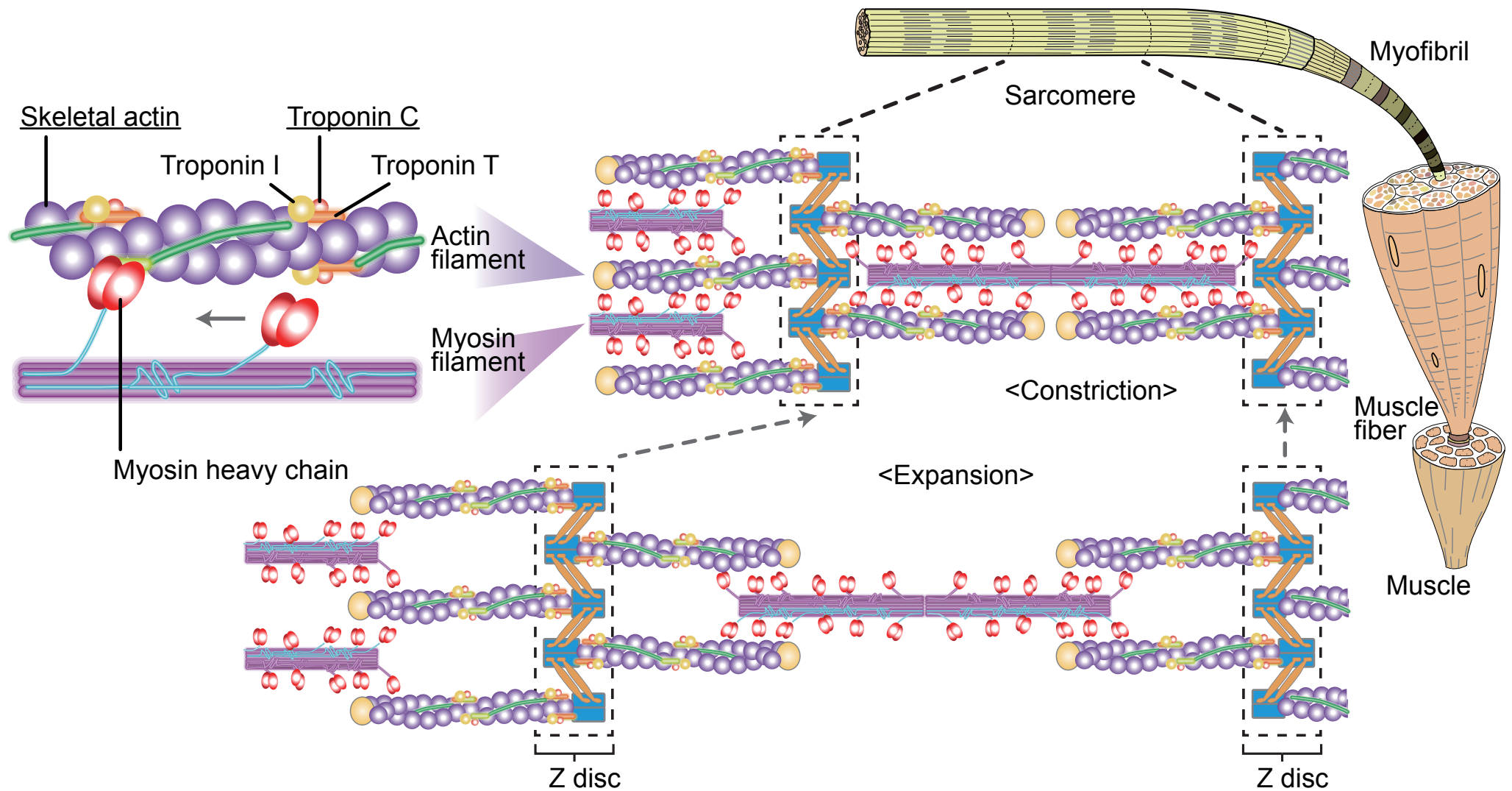
Inoue J, Satoh N. 2018.
Deuterostome genomics: Lineage-specific protein
expansions that enabled chordate muscle evolution.
Molecular Biology and Evolution. 35:914-924.

Taxon sampling

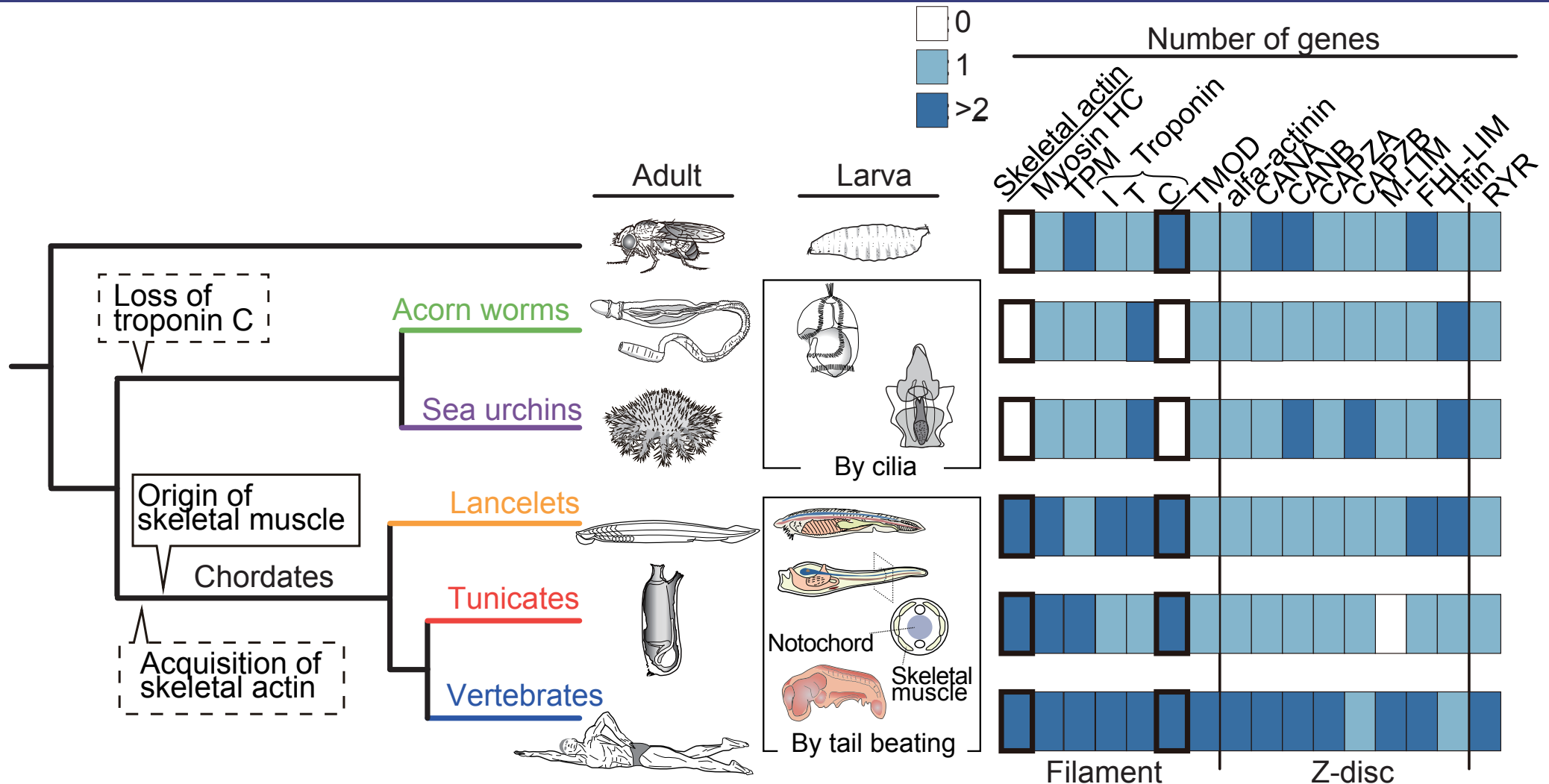
“Skeletal muscle” enabled chordate ancestors to swim by beating a tail with notochord.



Skeletal muscle of vertebrates

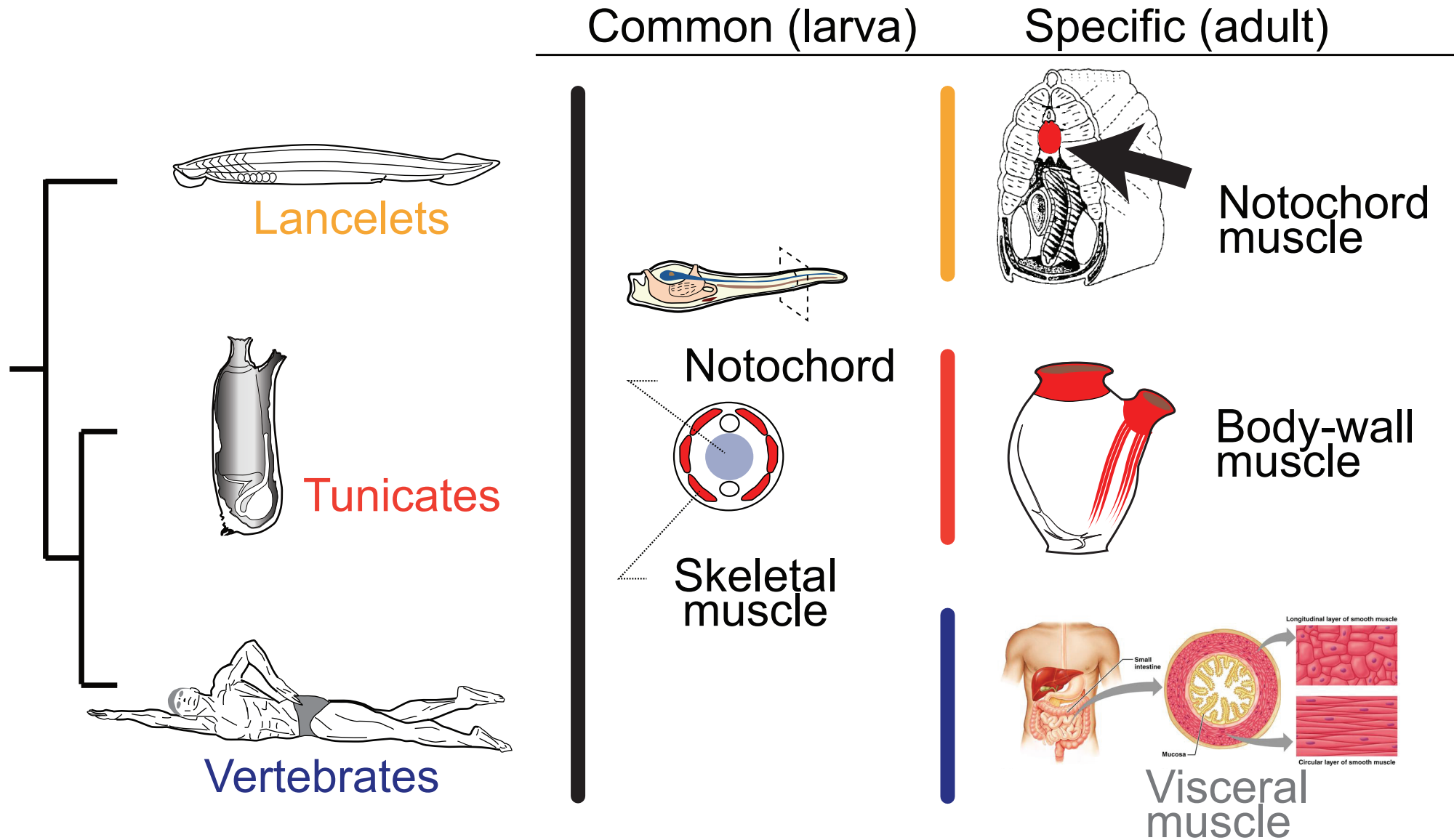


Results: Presence or absence of orthogroup members



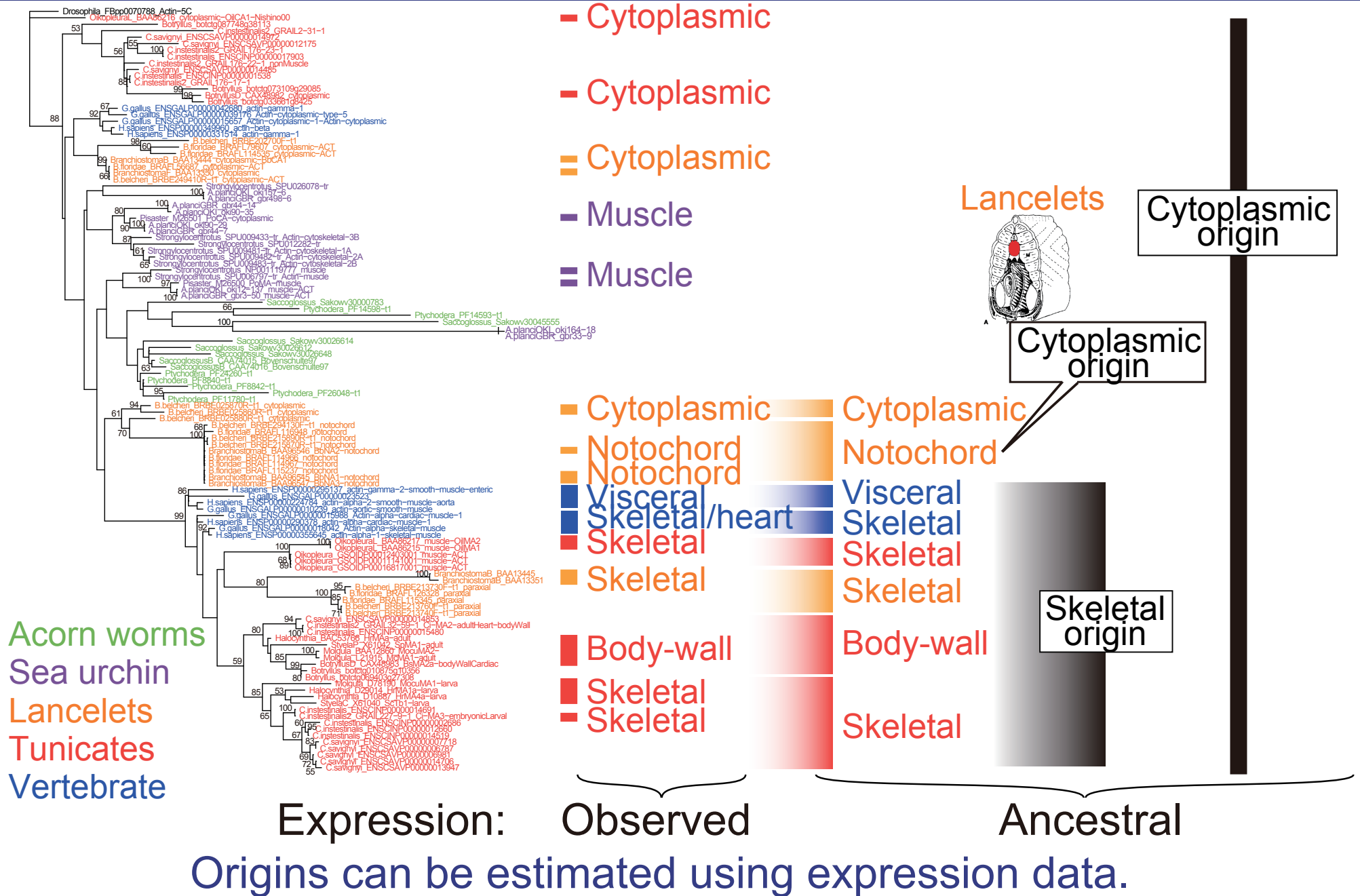
- Skeletal muscle was acquired by using 2 proteins
Skeletal actin and Troponin C.

Lineage-specific muscles

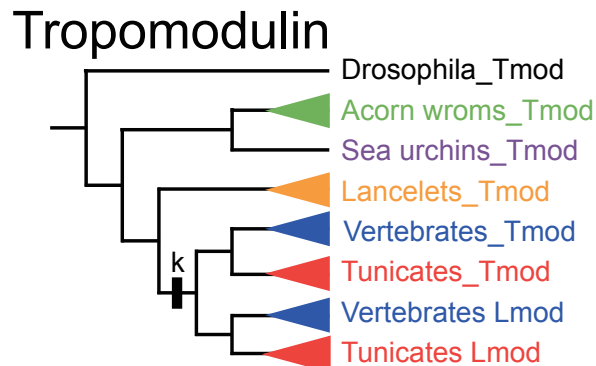
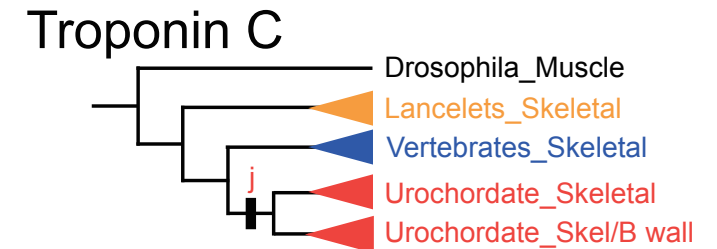
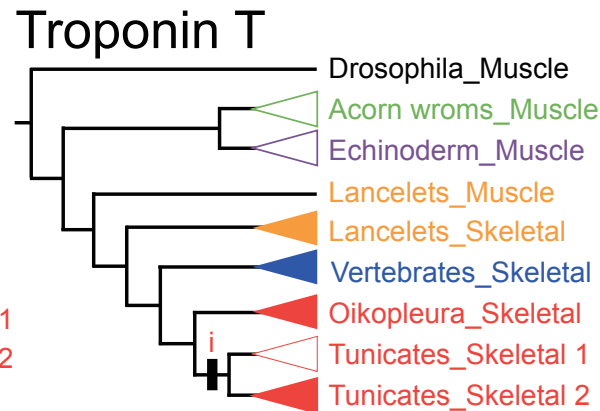
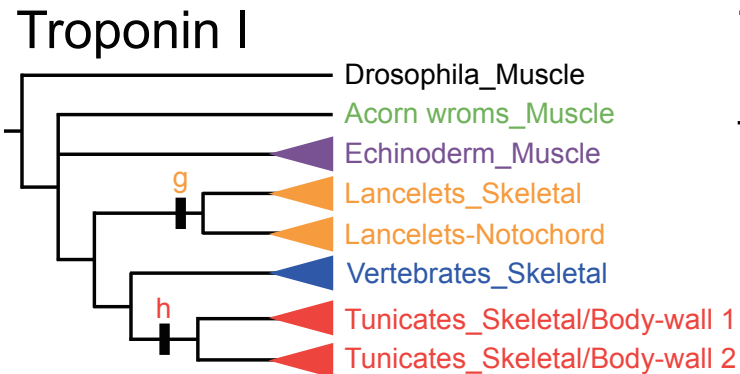
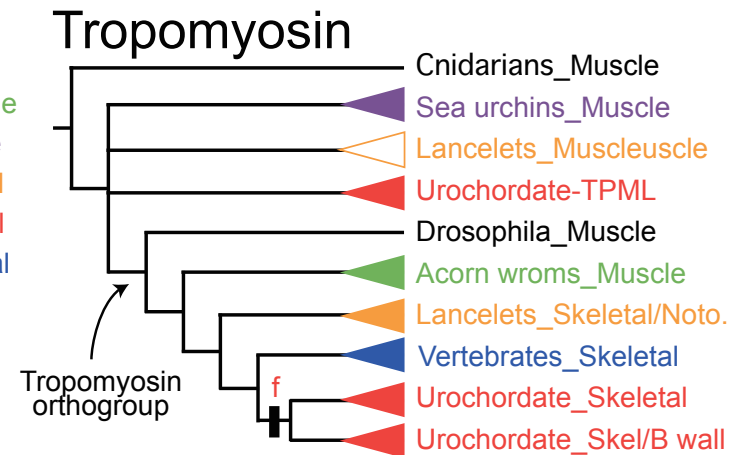
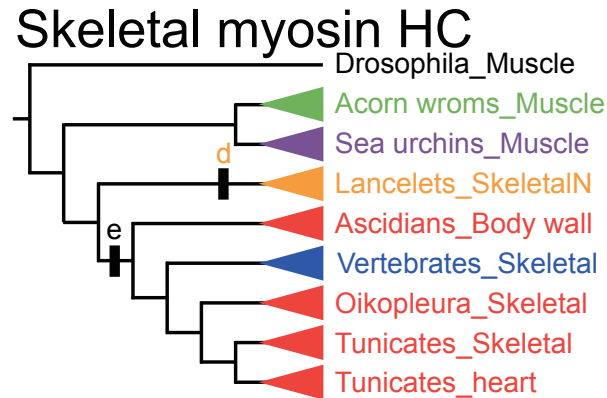
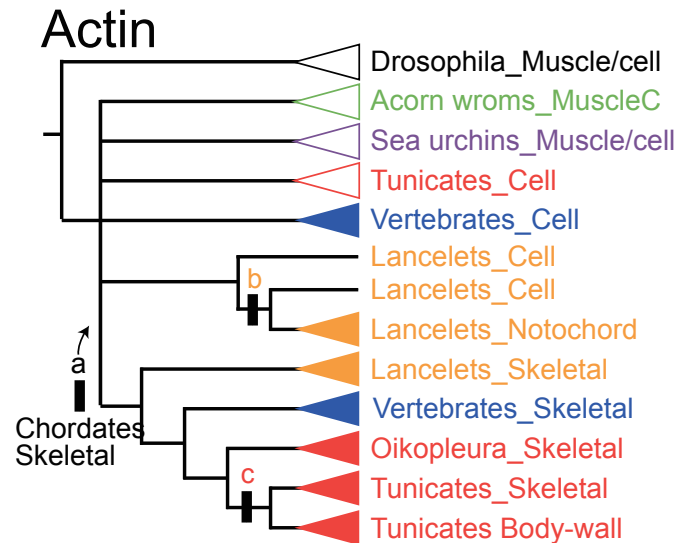


Origin of each protein?

Results: Estimation of actin origins



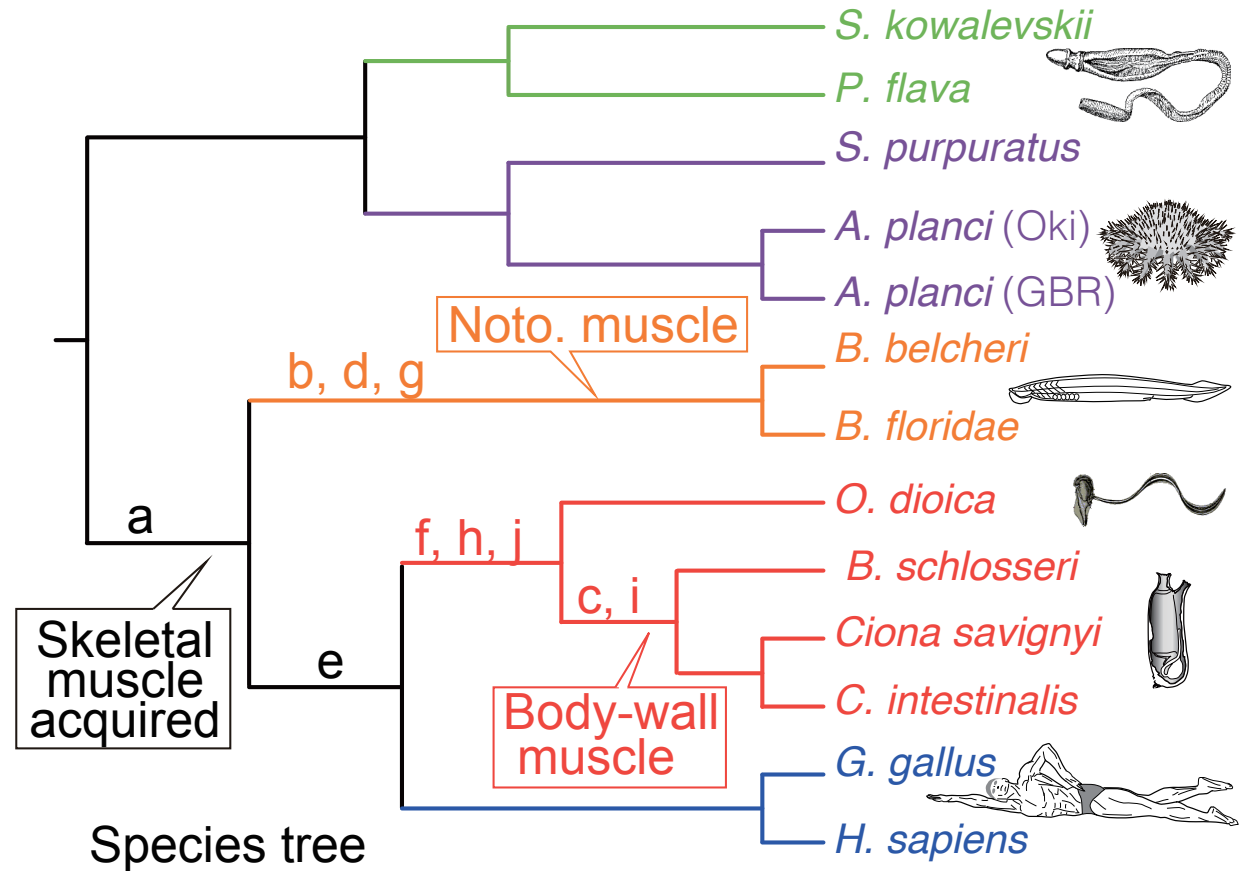
Other gene trees



Estimate origin of each protein.

Origins of specific-muscle proteins

Ancestral Dup.		
Lancelets Notochord muscle		
Noto. actin	Cytoplasmic	b
Skeletal MHC	Skeletal	d
Troponin I	Skeletal	g
Tunicates Body-wall muscle		
Skeletal MHC	Skeletal	e
Tropomyosin	Skeletal	f
Troponin I	Skeletal	h
Troponin C	Skeletal	j
Body-wall actin	Skeletal	c
Troponin T	Skeletal	i



- Most proteins of specific muscles were derived from the skeletal type.
- Most duplications were placed near lineages specific-muscles emerged, except for tunicate Skeletal MHC (e).

End of presentation