

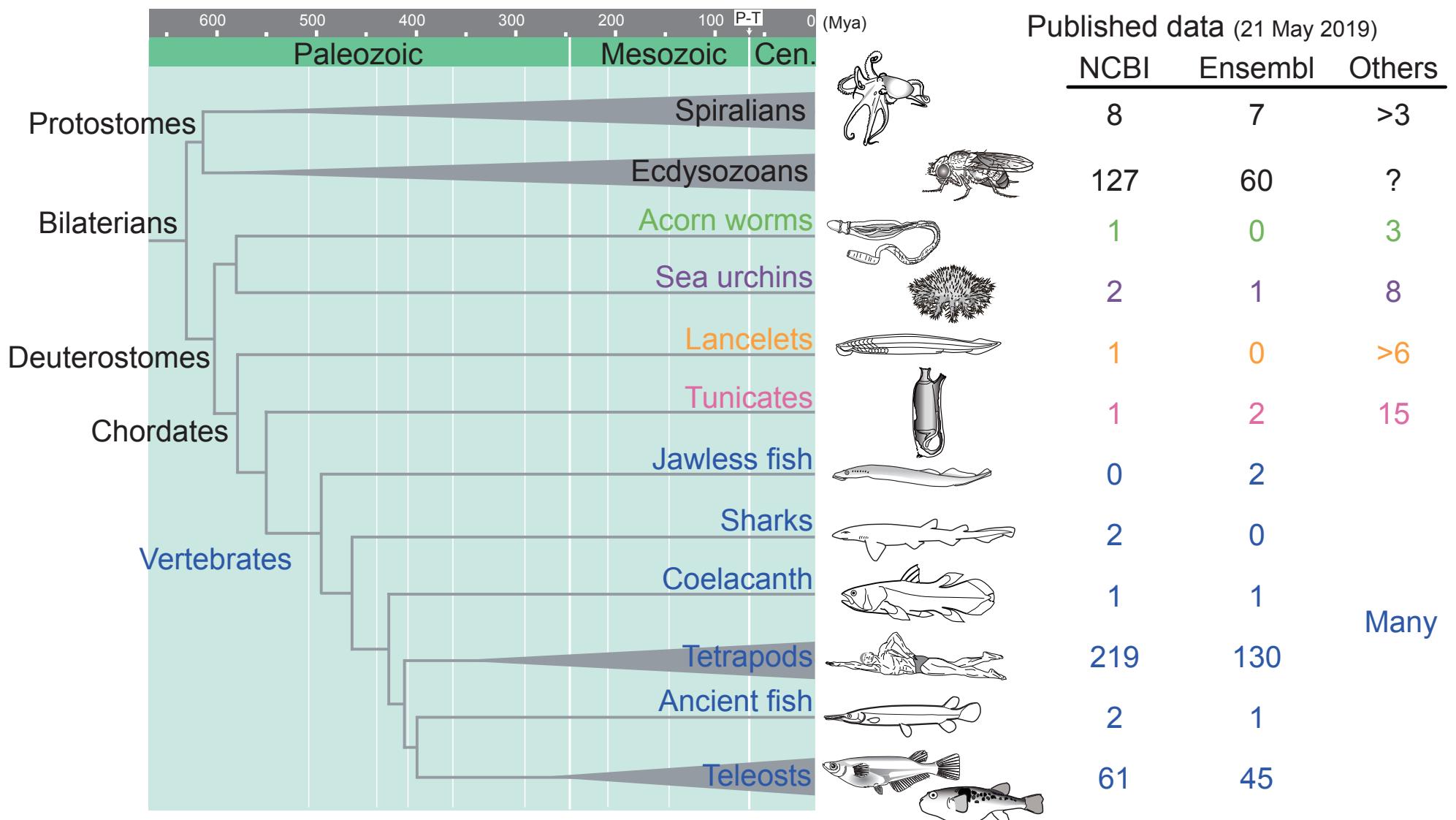
# ORTHOSCOPE: a web tool for estimating the origins and the functions of bilaterian protein-coding genes

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.

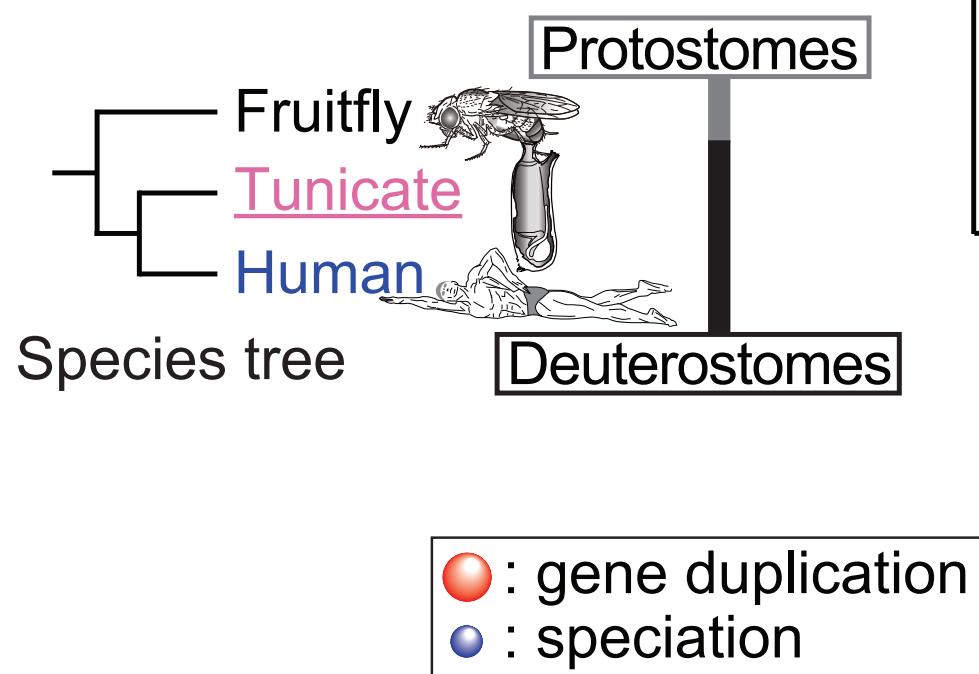
10 June 2019

# 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



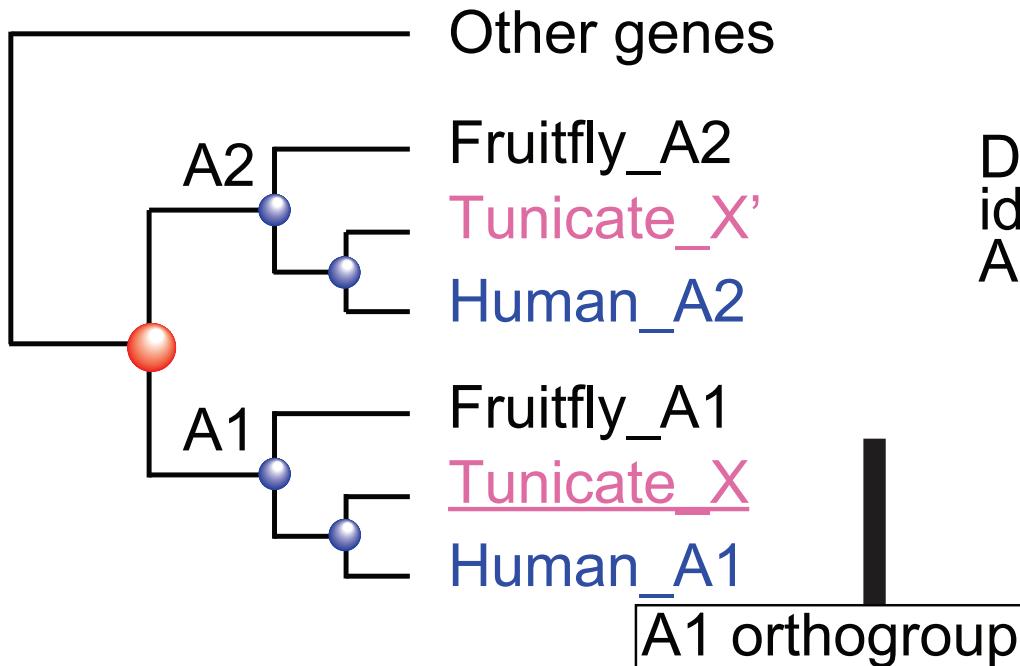
Given that orthologs are genes derived by speciation, they are used to infer gene functions, as the first step. However,

# 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).



ORTHO SCOPE

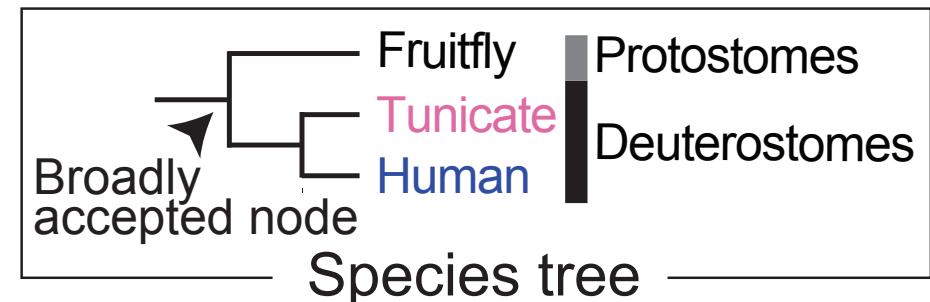
# Orthogroup identification in ORTHOSCOPE

## Candidate selection

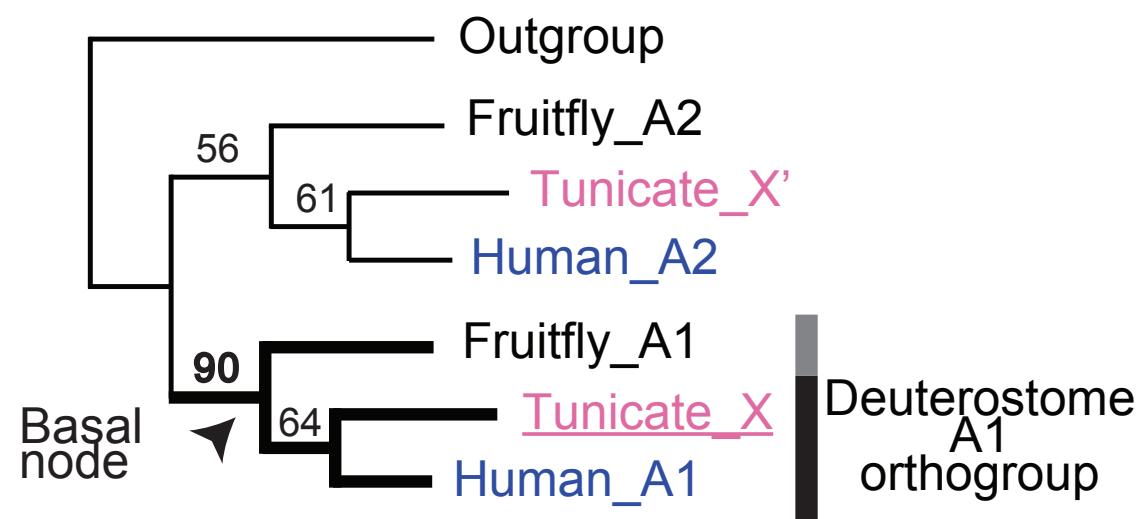
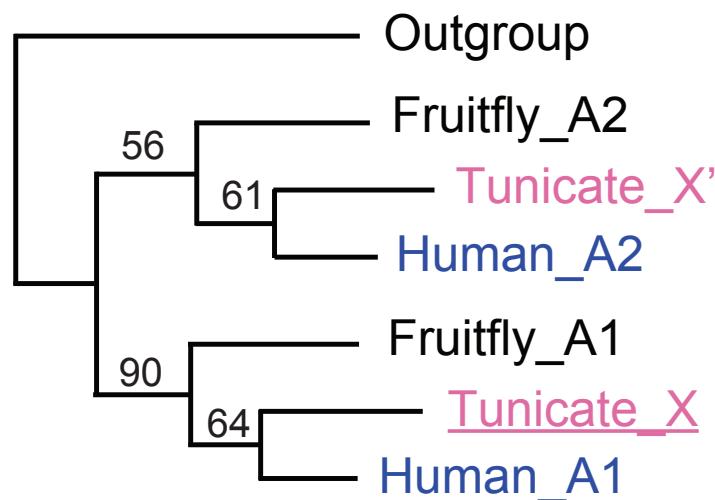
BLAST Tunicate\_X

vs Other animals  
(Gene models)

Gene tree estimation

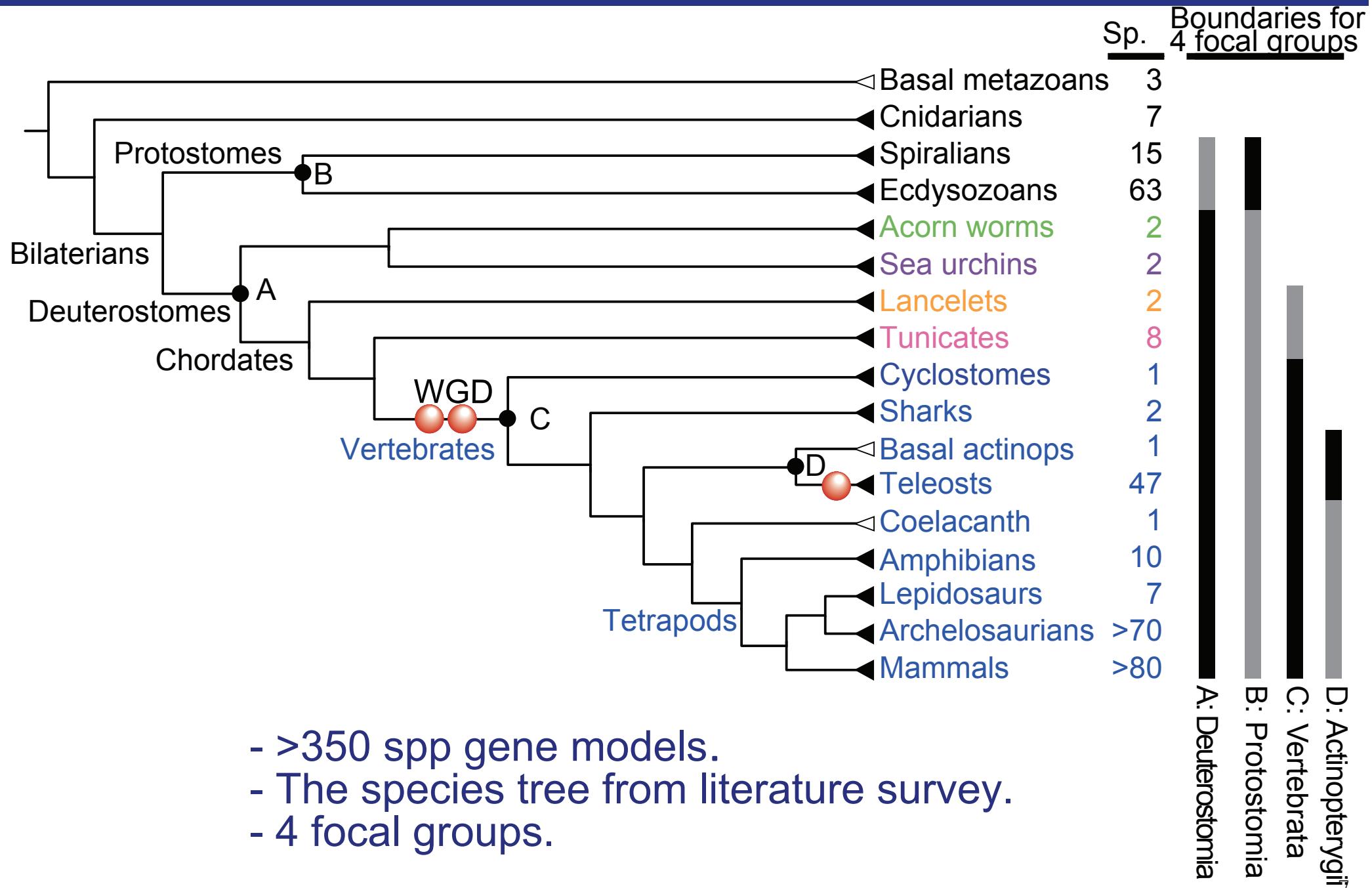


Comparison with  
the species tree



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

# ORTHO SCOPE database



# Metazoans

Eukaryotes

Metazoans

Bilaterians

Deuterostomes

Representatives  
from major lineages

Acorn worms

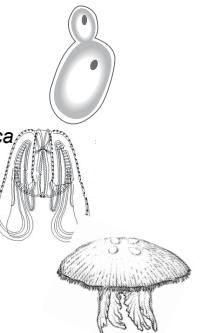
Sea urchins

Lancelets

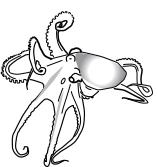
Tunicates

Vertebrates

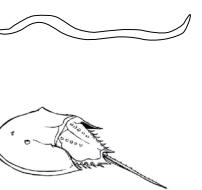
Ichthyosporea  
Salpingoecidae



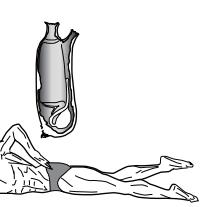
Saccharomyces cerevisiae  
Sphaeroforma arctica  
Capsaspora owczarzaki  
Salpingoeca rosetta  
Monosiga brevicollis  
Mnemiopsis leidyi  
Amphimedon queenslandica  
Trichoplax adhaerens  
Thelohanellus kitaeui  
Hydra vulgaris  
Exaiptasia pallida  
Nematostella vectensis  
Orbicella faveolata  
Stylophora 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



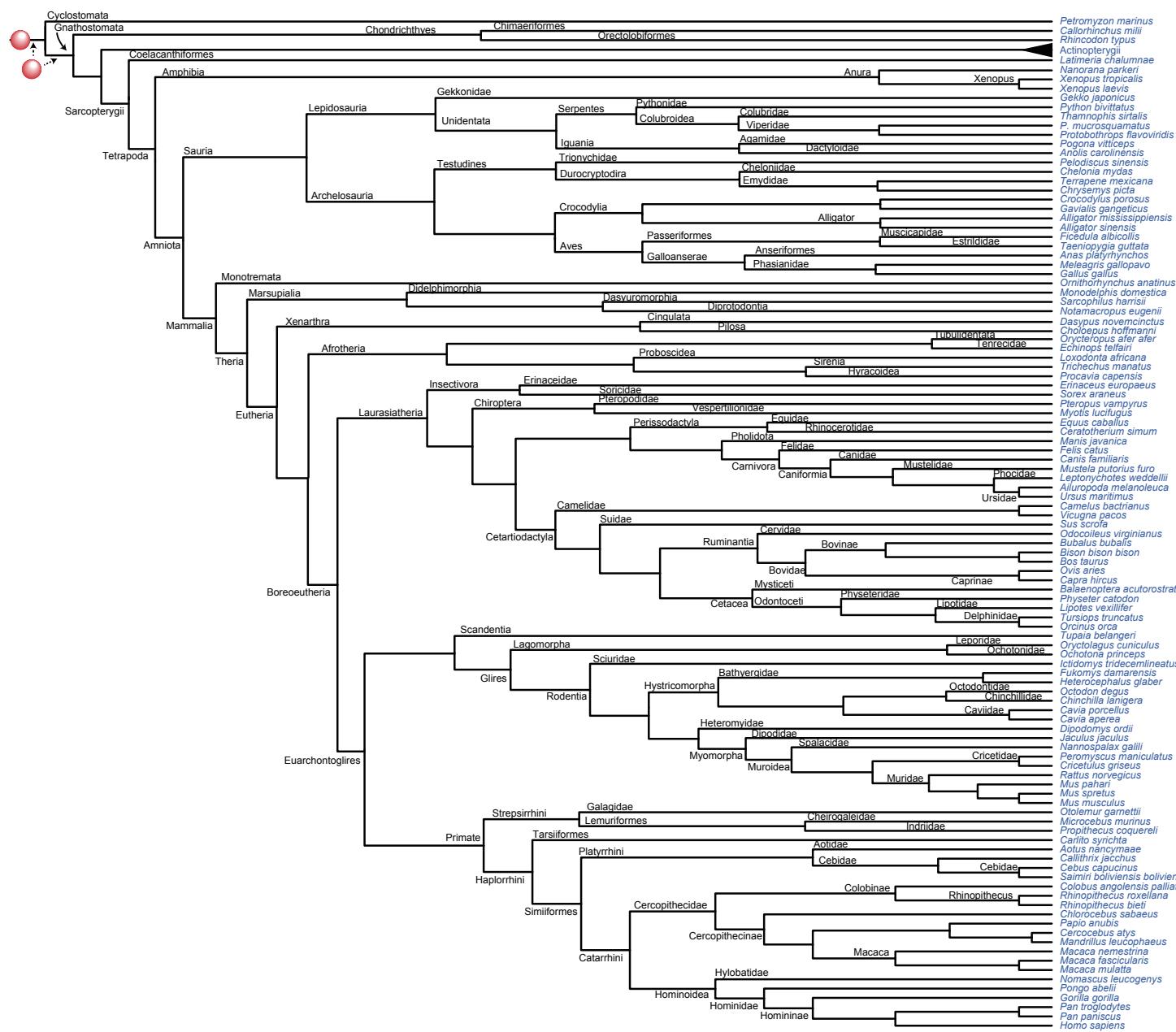
Polychaeta  
Clitellata  
Nemertea  
Brachiopoda  
Cephalopoda  
Gastropoda  
Heterobranchia  
Crassostrea  
Bivalvia  
Pterioidea



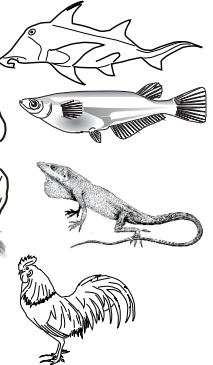
Priapulida  
Enoplea  
Chromadorea  
Spirurida  
Caenorhabditis  
Xiphosura  
Chelicerata  
Arachnida  
Myriapoda  
Mandibulata  
Pancrustacea  
Nematoda  
Priapulus caudatus  
Trichinella spiralis  
Strongyloides ratti  
Onchocerca volvulus  
Loa loa  
Brugia malayi  
Pristionchus pacificus  
Caenorhabditis japonica  
Caenorhabditis brenneri  
Caenorhabditis remanei  
Caenorhabditis briggsae  
Caenorhabditis elegans  
Limulus polyphemus  
Sarcophaga scabiei  
Tetranychus urticae  
Ixodes scapularis  
Centruroides sculpturatus  
Parasteatoda tepidariorum  
Stegodyphus mimosarum  
Strigamia maritima  
Hyalaea azteca  
Eurytemora affinis  
Lepeophtheirus salmonis  
Daphnia magna  
Daphnia pulex  
Hexapods (> 35 spp)  
Saccoglossus kowalevskii  
Ptychoderia flava  
Strongylocentrotus purpuratus  
Acanthaster planci  
Branchiostoma floridae  
Branchiostoma belcheri  
Oikopleura dioica  
Botryllus schlosseri  
Ciona savignyi  
Ciona intestinalis  
Vertebrates (> 230 spp)



# Vertebrates



- Sharks
  - Ray-finned fish
  - Amphibians
  - Snakes
  - Iguanas
  - Turtles
  - Alligators



- Marsupials  
Xenarthrans  
Afrotherians  
Insectivores  
Chiropterans  
Perissodactyls  
Carnivorans  
Ruminants  
Cetaceans



## Mammals

Rodents

Primates

# Use of gene models

| <b>Gene ID</b>  | <b>Gene name</b> | <b>Transcript ID</b> | <b>Length</b> |   |
|-----------------|------------------|----------------------|---------------|---|
| 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 was 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  
Amino acid DNA [Example](#) ファイルを選択 ファイル未選択  
If not selected, [this tree](#) is used.

**Coding sequences**

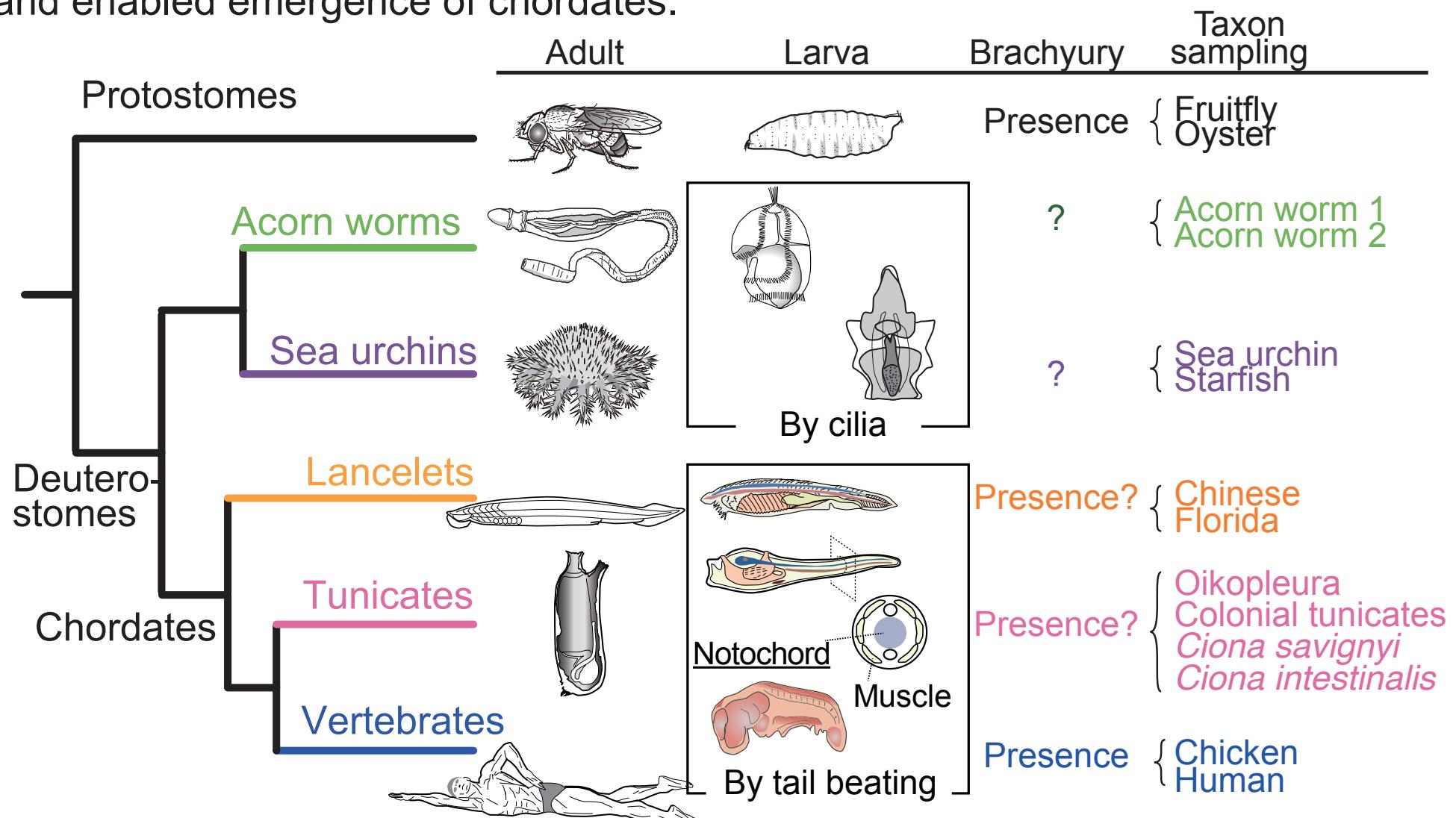
Upload coding sequences (amino acid or DNA)

## 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 demonstration

# 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)

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.

```
DeuterostomeBra.fas.txt
~/Downloads/DeuterostomeBra.fas.txt
1 >Homo-sapiens_XXX1
2 ATGAGCTCCCTGGCACCGAGAGCGCGGGAAAGAGCCTGCAGTACCGAGTGGAACCACTGCTGAGC
3 >Branchiostoma-floridae_XXX
4 ATGAAGCAGACACCGGACCAGTTCTCGGTTAGCCACCTCCTGAGCGCCGTGGAGAGCGAGATCTCG
5 >Homo-sapiens_ENST00000389554.7_T-box-brain-1
6 ATGCAGCTGGAGCACTGCCTTCTCCTTCTATCATGCTCTCCAAGAAATTCTCAATGTGAGCAGC
```

L: 1 C: 1    Text File ▾  Unicode (UTF-8) ▾  Unix (LF) ▾  Saved: 2019/05/31 11:11:20    4,767 / 12 / 7    100% ▾

1st query

Multiple query sequences can be used.

|                        |  |  |                             |
|------------------------|--|--|-----------------------------|
|                        | <input checked="" type="checkbox"/> <i>Drosophila melanogaster</i>       | Fruit fly  | EnsMet38                    |
|                        | <input type="checkbox"/> <i>Drosophila melanogaster</i> 1                | Fruit fly  | <a href="#">RefSeq89</a>    |
| <b>Xenacoelomorpha</b> |  |  |                             |
| Xenoturbella           | <input type="checkbox"/> <i>Xenoturbella bocki</i>                       | Xenoturbella                                     | <a href="#">TSA</a>         |
| Nemertodermatida       |  |  |                             |
| Acoela                 | <input type="checkbox"/> <i>Symsagittifera roscoffensis</i>              | Acoela   | <a href="#">TSA</a>         |
|                        | <input type="checkbox"/> <i>Hofstenia miamia</i>                         | Acoela   | <a href="#">TSA</a>         |
|                        | <input type="checkbox"/> <i>Isodiametra pulchra</i>                      | Acoela   | <a href="#">TSA</a>         |
|                        | <input type="checkbox"/> <i>Praesagittifera naikaiensis</i>              | Acoela   | OIST-S                      |
| <b>Deuterostomia</b>   |  |  |                             |
| Hemichordata           | <input checked="" type="checkbox"/> <i>Saccoglossus kowalevskii</i>      | Helical acorn worm                               | <a href="#">OIST-R</a>      |
|                        | <input type="checkbox"/> <i>Saccoglossus kowalevskii</i> 1               | Helical acorn worm                               | <a href="#">RefSeq89</a>    |
|                        | <input checked="" type="checkbox"/> <i>Ptychodera flava</i>              | Yellow acorn worm                                | OIST-S                      |
| Echinodermata          | <input checked="" type="checkbox"/> <i>Strongylocentrotus purpuratus</i> | Purple urchin                                    | EnsMet38                    |
|                        | <input type="checkbox"/> <i>Strongylocentrotus purpuratus</i> 1          | Purple urchin                                    | <a href="#">RefSeq89</a>    |
|                        | <input checked="" type="checkbox"/> <i>Acanthaster planci</i>            | Crown-of-thorns starfish<br>(Okinawa)            | OIST-S                      |
|                        | <input type="checkbox"/> <i>Acanthaster planci</i> 1                     | Crown-of-thorns starfish<br>(Okinawa)            | <a href="#">RefSeq89</a>    |
|                        | <input type="checkbox"/> <i>Acanthaster planci GBR</i>                   | Crown-of-thorns starfish<br>(Great Barrier Reef) | OIST-S                      |
| Cephalochordata        | <input type="checkbox"/> <i>Asymmetron lucayanum</i>                     | Lancelet   | <a href="#">TSA</a>         |
|                        | <input checked="" type="checkbox"/> <i>Branchiostoma belcheri</i>        | Belcher's lancelet                               | <a href="#">LanceletDB</a>  |
|                        | <input type="checkbox"/> <i>Branchiostoma belcheri</i> 1                 | Belcher's lancelet                               | <a href="#">RefSeq89</a>    |
|                        | <input type="checkbox"/> <i>Branchiostoma lanceolatum</i>                | Mediterranean amphioxus                          | <a href="#">Amphiencode</a> |
|                        | <input checked="" type="checkbox"/> <i>Branchiostoma floridae</i>        | Florida lancelet                                 | <a href="#">JGI</a>         |
|                        | <input type="checkbox"/> <i>Branchiostoma floridae</i> 1                 | Florida lancelet                                 | <a href="#">RefSeq89</a>    |
| Urochordata            | <input checked="" type="checkbox"/> <i>Oikopleura dioica</i>             | Pelagic tunicate                                 | <a href="#">OikoBase</a>    |
|                        | <input type="checkbox"/> <i>Salpa thompsoni</i>                          | Salpa  | <a href="#">TSA</a>         |
|                        | <input type="checkbox"/> <i>Molgula occidentalis</i>                     | —  | <a href="#">aNISEED</a>     |
|                        | <input type="checkbox"/> <i>Molgula oculata</i>                          | Sea grape  | <a href="#">aNISEED</a>     |
|                        | <input type="checkbox"/> <i>Botrylloides leachii</i>                     | Leache's ascidian                                | <a href="#">aNISEED</a>     |
|                        | <input checked="" type="checkbox"/> <i>Botryllus schlosseri</i>          | Golden star tunicate                             | <a href="#">Stanford</a>    |
|                        | <input checked="" type="checkbox"/> <i>Ciona savignyi</i>                | Pacific transparent sea squirt                   | <a href="#">Ens91</a>       |
|                        | <input type="checkbox"/> <i>Ciona savignyi</i> 1                         | Pacific transparent sea squirt                   | <a href="#">aNISEED</a>     |
|                        | <input checked="" type="checkbox"/> <i>Ciona intestinalis</i>            | Transparent sea squirt                           | <a href="#">Ens91</a>       |
|                        | <input type="checkbox"/> <i>Ciona intestinalis</i> 1                     | Transparent sea squirt                           | <a href="#">RefSeq89</a>    |
| <b>Vertebrata</b>      |  |  |                             |
| Myxiniformes           | <input type="checkbox"/> <i>Eptatretus burgeri</i>                       | Inshore hagfish                                  | <a href="#">Ens94</a>       |

High quality gene models  
as defaults.

result3300

```

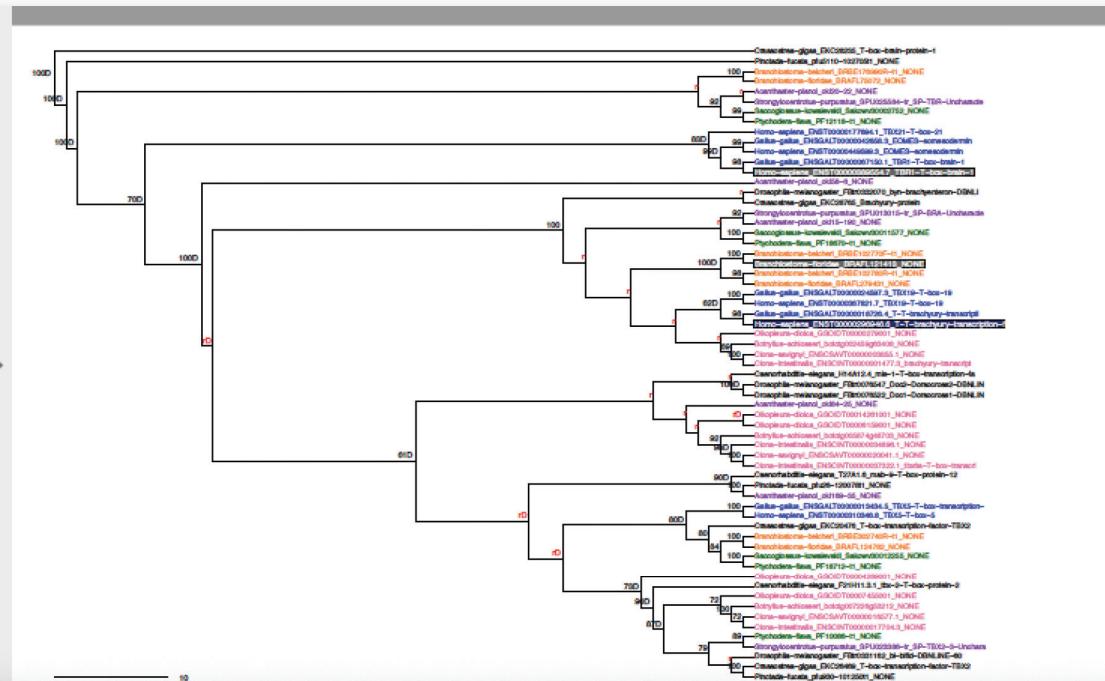
010_candidates_nucl.txt 今日 11:19 -- フォルタ
020_treeRearranged.pdf 今日 11:19 102 KB BBEdit.
020_tree.pdf 今日 11:19 8 KB Adobe.
100_2ndTree.zip 今日 11:19 8 KB ZIPアーカイブ
000_summary.txt 今日 11:19 44 KB BBEdit.
010_candidates_prot.txt 今日 11:19 36 KB BBEdit.
010_align.aln.txt 今日 11:19 4.2 MB BBEdit.

000_summary.txt
#####
Results #####
#####
>Queries_used_in_the_analysis
Homo-sapiens_ENST00000296946.6_T-T-brachyury-transcrip...
Branchiostoma-floridae_BRAFL121413_NONE
Homo-sapiens_ENST00000389554.7_TBR1-T-box-brain-1

>BootstrapValue_OrthogroupMonophyly
The orthogroup monophyly is supported by 100% bootstrap

>Orthogroup
Acanthaster-planci_oki15-190_NONE
Botryllus-schlosseri_botctg002459g63408_NONE
Branchiostoma-belcheri_BRBE102770F-t1_NONE
Branchiostoma-belcheri_BRBE102780R-t1_NONE
Branchiostoma-floridae_BRAFL121413_NONE
Branchiostoma-floridae_BRAFL279431_NONE
Ciona-intestinalis_ENSCINT00000001477.3_brachyury-tr...
Ciona-savignyi_ENSCSAVT00000003855.1_NONE
Crassostrea-gigas_EKC28765_Brachyury-protein
Drosophila-melanogaster_FBtr0332070_byn-brachyenter...
Gallus-gallus_ENSGALT00000018726.4_T-T-brachyury-tr...
Gallus-gallus_ENSGALT00000024597.3_TBX19-T-box-19
Homo-sapiens_ENST00000296946.6_T-T-brachyury-transcrip...

```

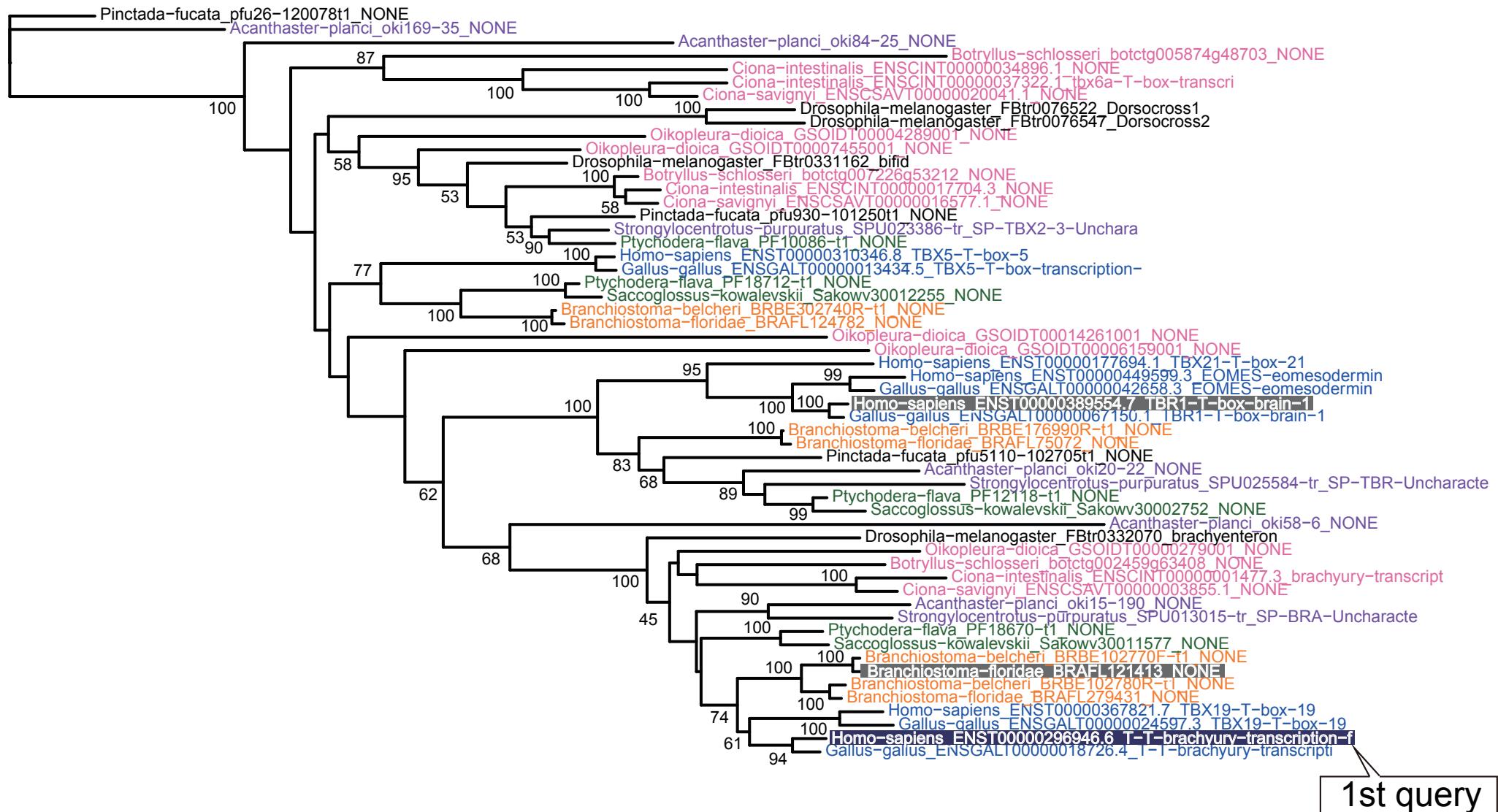


file:///Users/inouejunmp/Downloads/result3300/010\_align

|  | LSD | alk | weblio | Thesaurus | 100 | 101 | 103 | 104 | さくらのVPS | Sakura | dbCNS | ws | oistServer | BraSyn | BraTree30 |  |
|--|-----|-----|--------|-----------|-----|-----|-----|-----|---------|--------|-------|----|------------|--------|-----------|--|
| Used4treeSearch  |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Crassostrea-gigas_EKC28235_T-box-brain-protein-1               |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Pinctada-fucata_pfu5110-102705t1_NONE                          |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Branchiostoma-belcheri_BRBE176990R-t1_NONE                     |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Branchiostoma-floridae_BRAFL75072_NONE                         |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Acanthaster-planci_oki20-22_NONE                               |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Strongylocentrotus-purpuratus_SPU025584-tr_SP-TBR-Uncharact... |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Saccoglossus-kowalevskii_Sakowv30002752_NONE                   |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Ptychodera-flava_PF12118-t1_NONE                               |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Homo-sapiens_ENST00000177694.1_TBX21-T-box-21                  |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Gallus-gallus_ENSGALT00000042658.3_EOMES-eomesodermin          |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Homo-sapiens_ENST00000449599.3_EOMES-eomesodermin              |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Gallus-gallus_ENSGALT00000067150.1_TBR1-T-box-brain-1          |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |
| Homo-sapiens_ENST00000389554.7_TBR1-T-box-brain-1              |     |     |        |           |     |     |     |     |         |        |       |    |            |        |           |  |

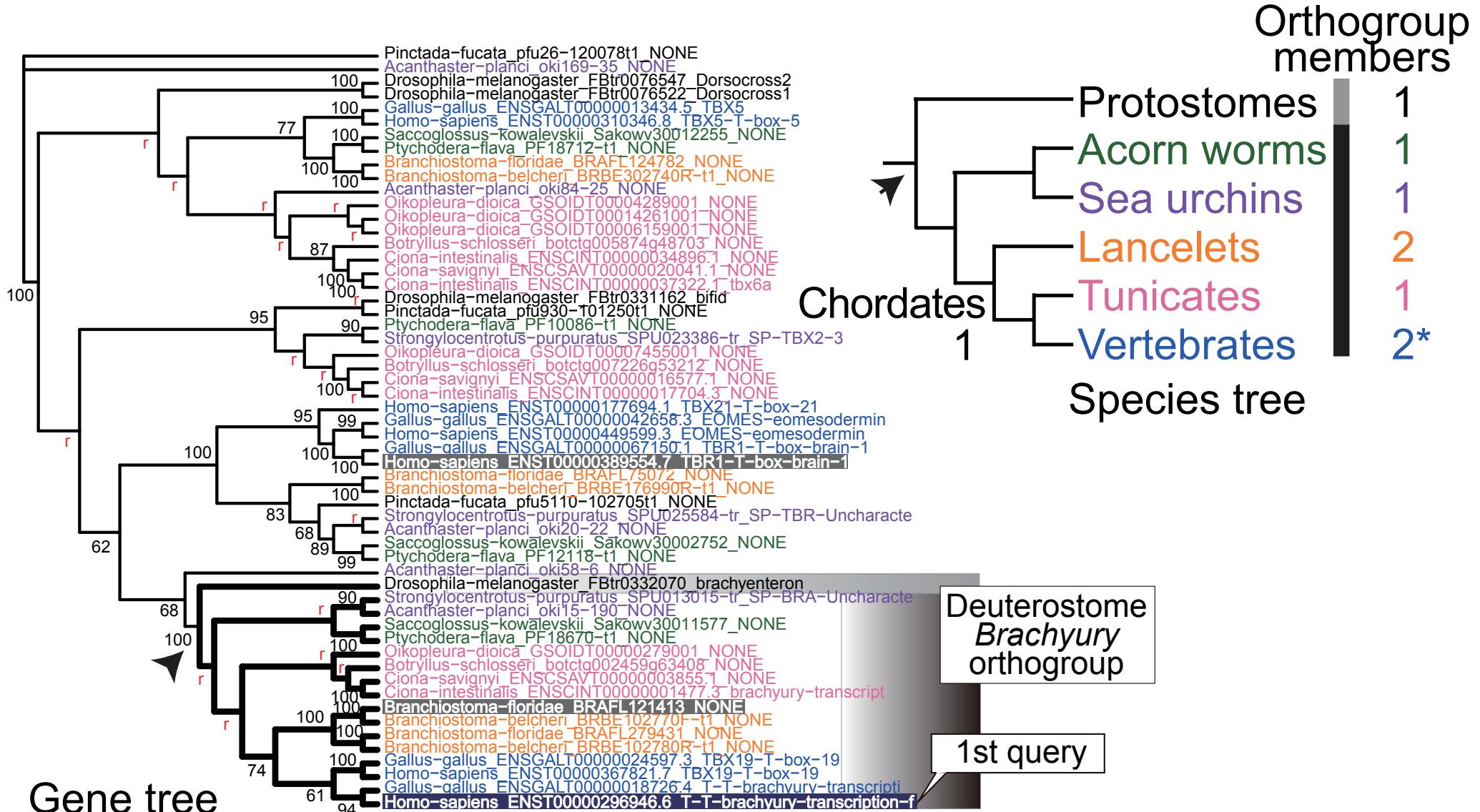
Results can be downloaded.

# Results: NJ tree



1st query

# Gene tree with identified orthogroup



# Transcript ID

Drosophila-melanogaster \_FBtr0332070 \_brachyenteron

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

Search

All species for FBtr0332070 Go

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

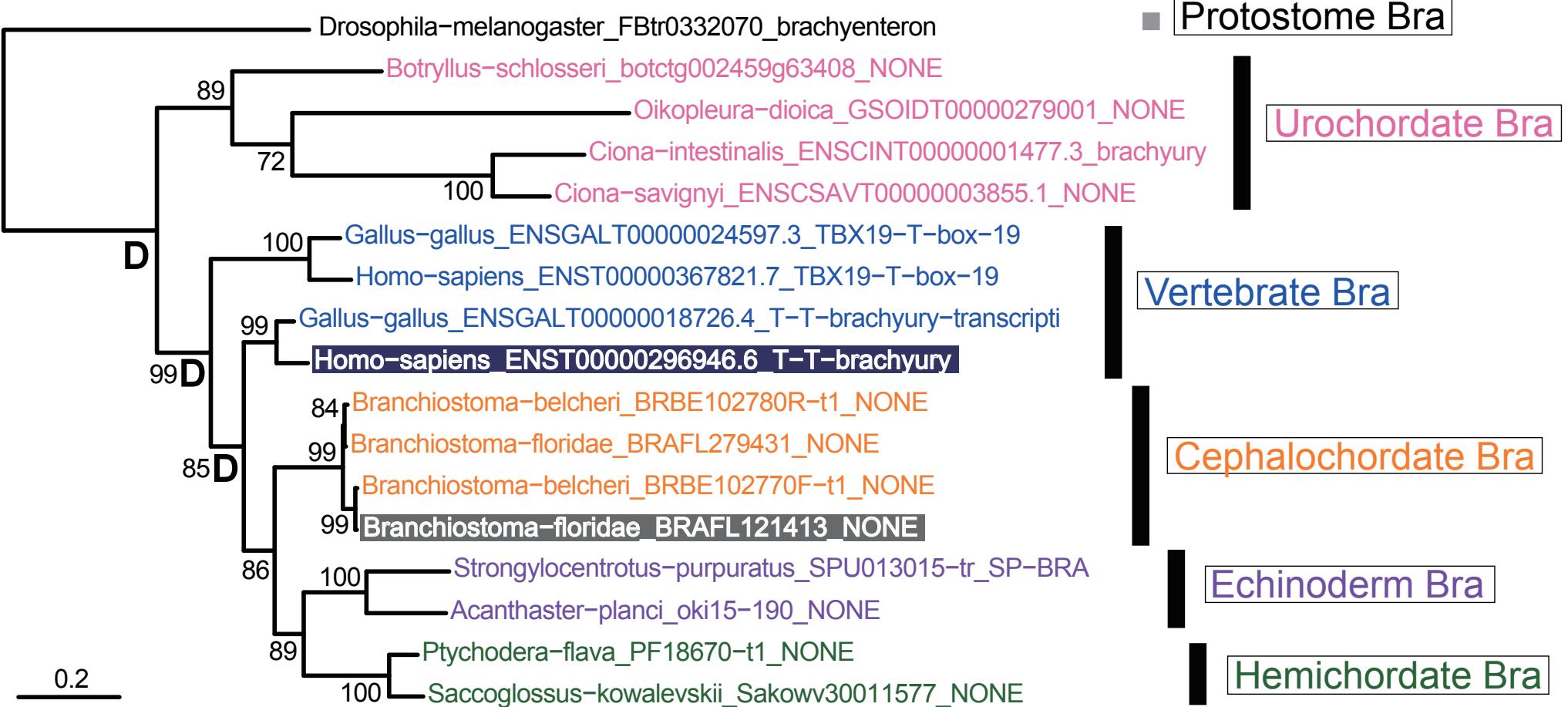
GTGGGCAATCCTAAAGATAGCCAAATATTATTGTTCAGATACTCAC  
AGCGTCAAGTGAGATGCCCTGGAGTGTTGAAATCAGTGAAATT  
ATTCGGCAGGGGAGCTTTTGGGTTGATGAAACAGGGGGCATCTTCATAGA  
AATATAA  
 FlyBase FB2019\_03, released Jun 11, 2019 Gene: Dmel\byn

Home Tools Downloads Links Community Species About Help Archives J2G ▾ Jump

### General Information

| Symbol            | Dmel\byn  | Species            | D. melanogaster      |
|-------------------|---|--------------------|----------------------|
| 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



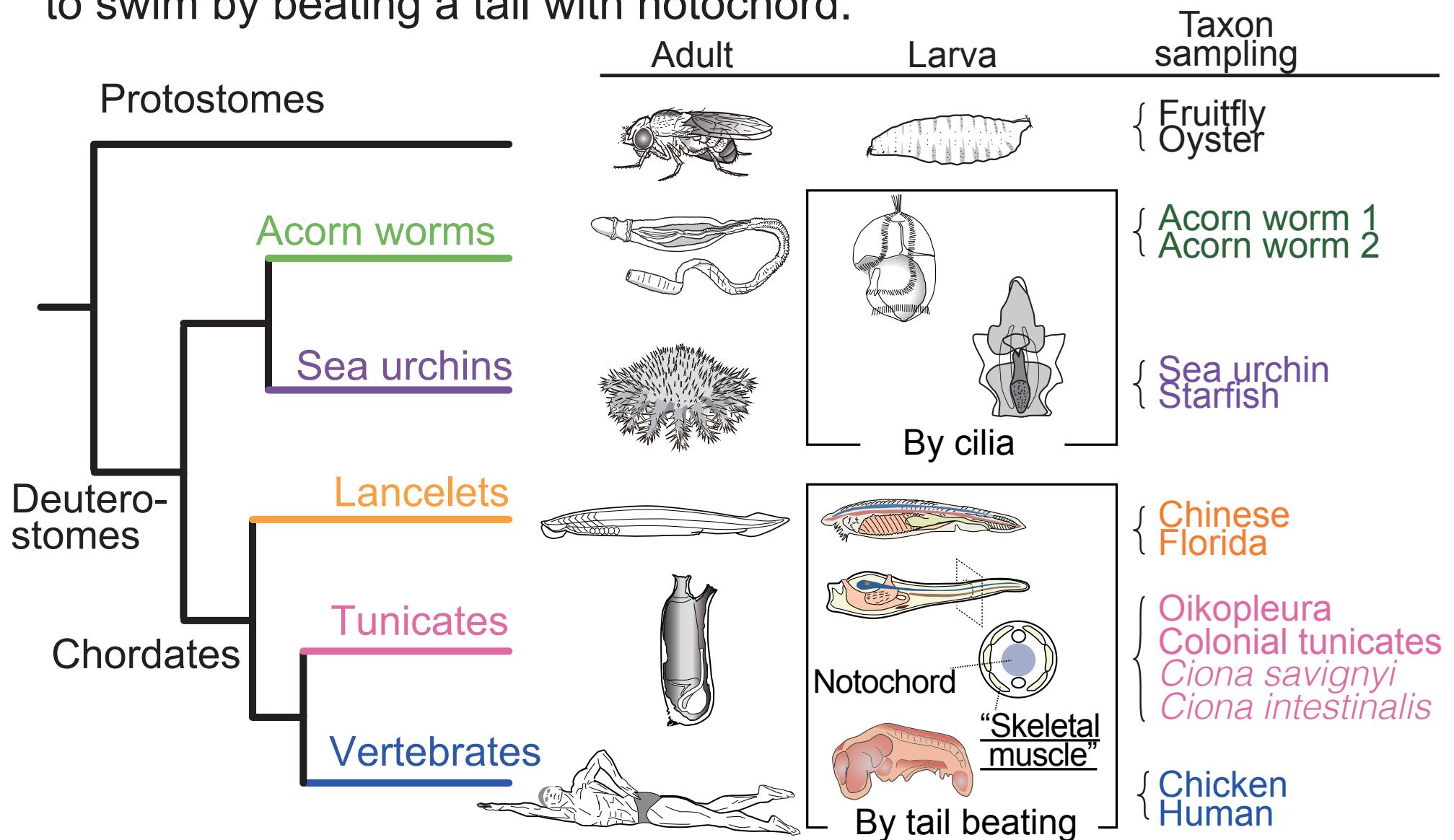
Identify gene duplication nodes (D).

## Genes enabled emergence of chordates: Muscle genes

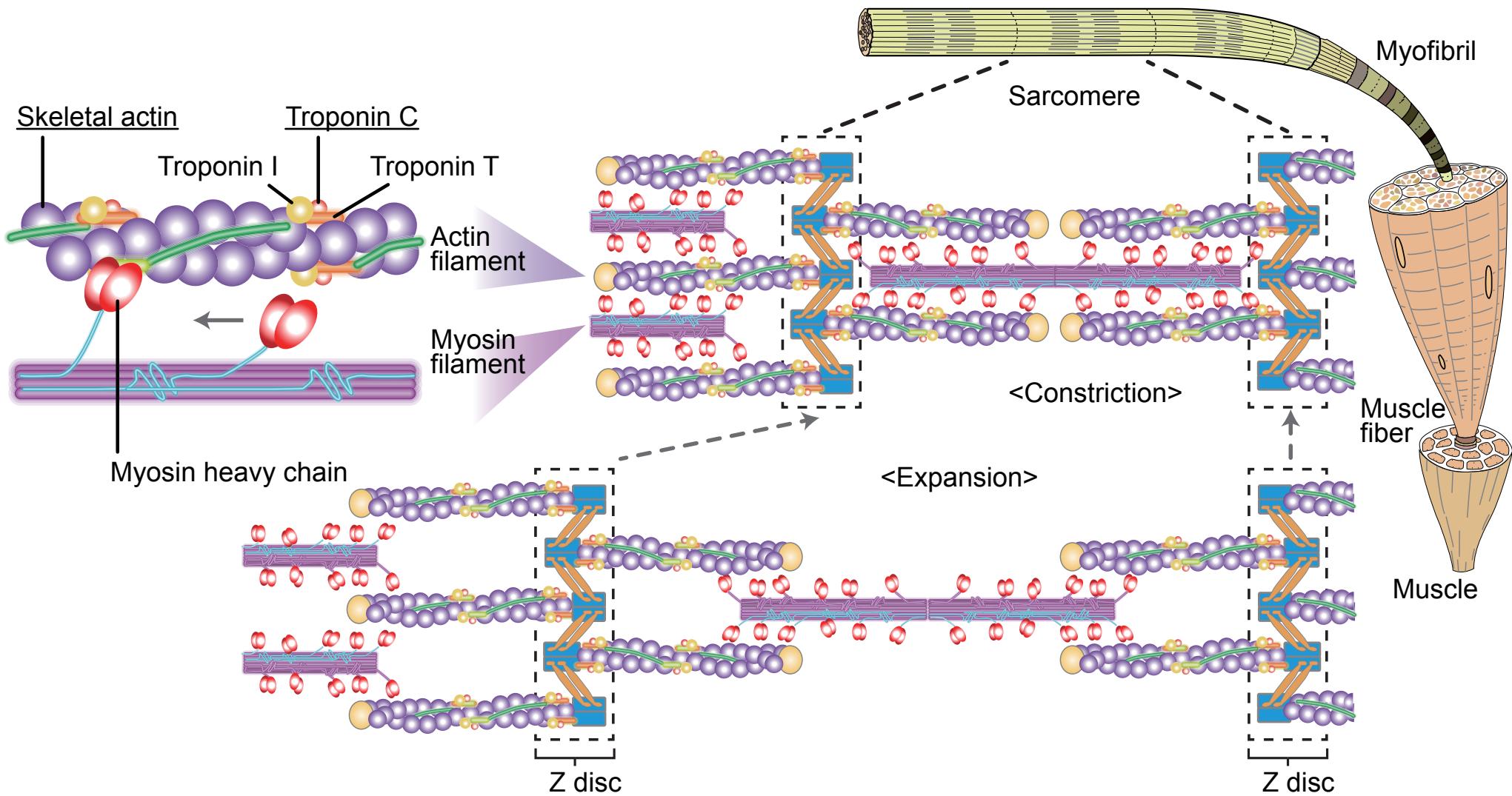
Inoue J, Satoh N. 2018.  
Deuterostome genomics: Lineage-specific protein  
expansions that enabled chordate muscle evolution.  
*Molecular Biology and Evolution*. 35:914-924.

# Taxon sampling

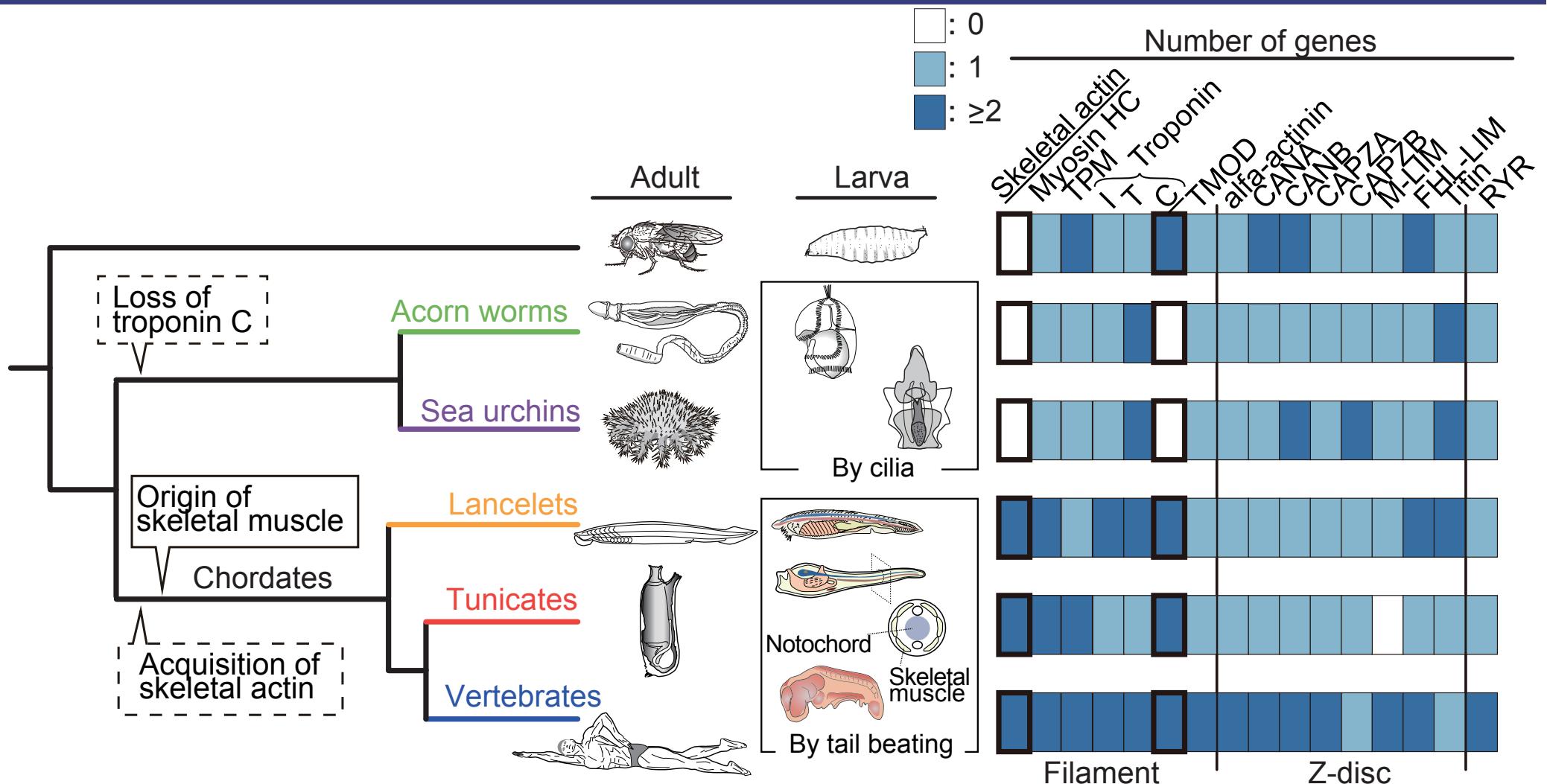
“Skeletal muscle” enabled chordate ancestors to swim by beating a tail with notochord.



# Skeletal muscle of vertebrates

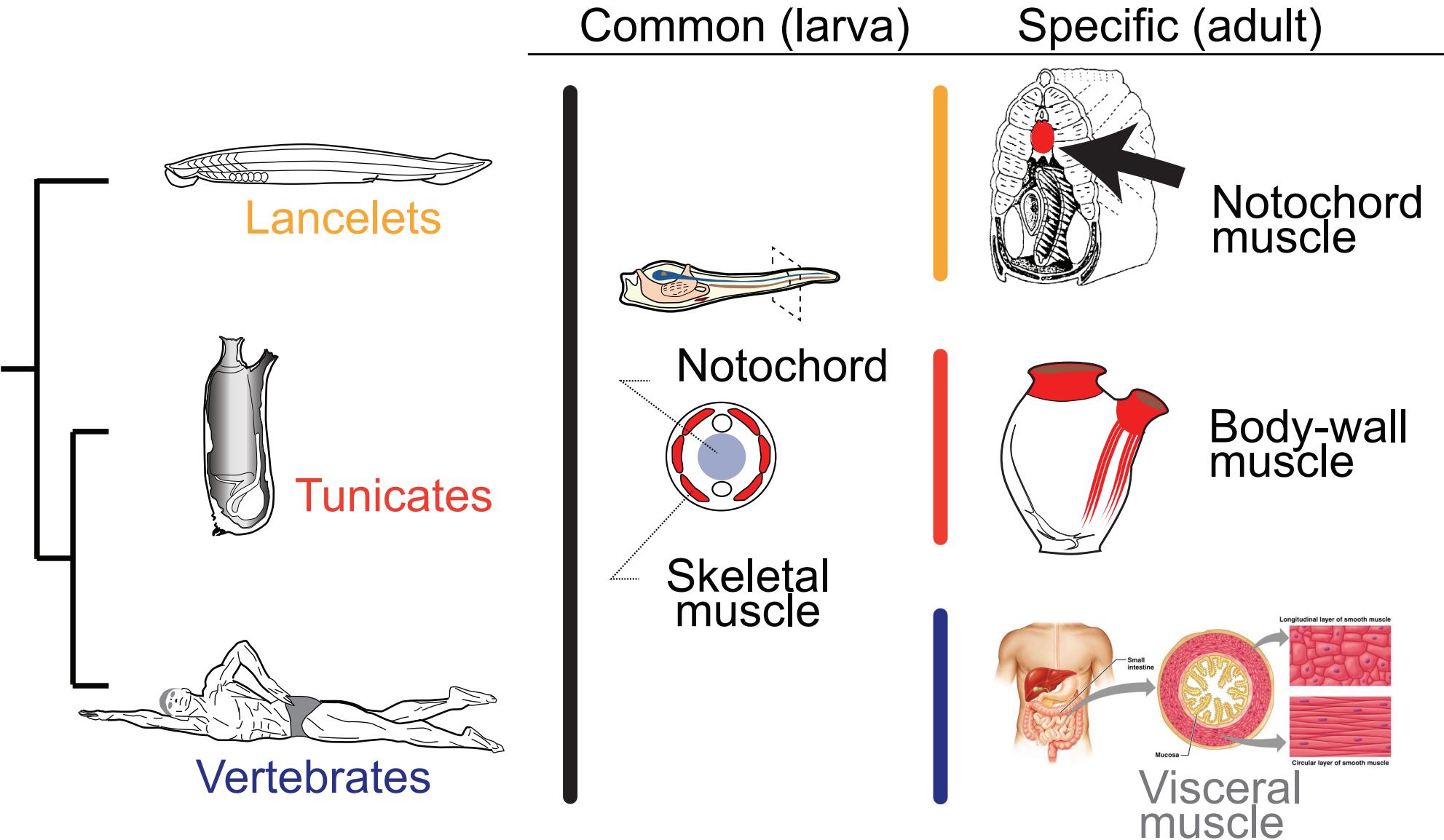


# Results: Presence or absence of orthogroup members



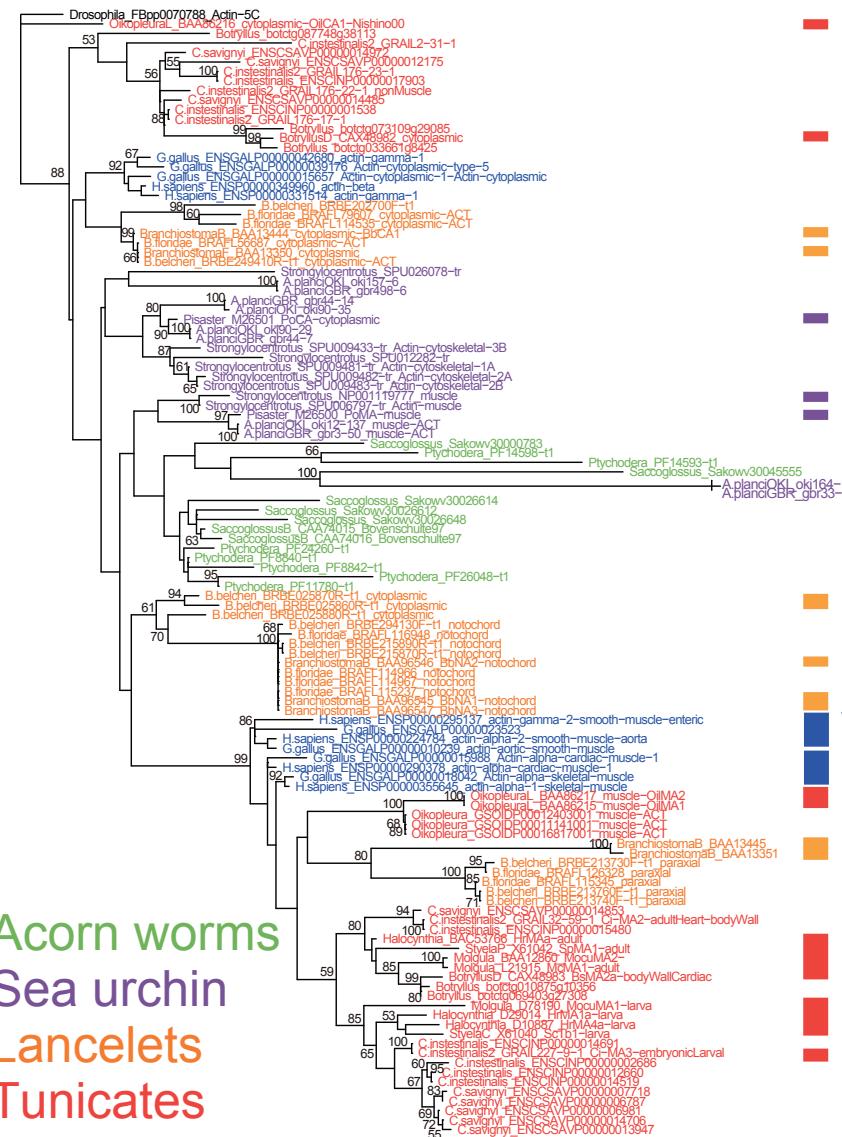
- Skeletal muscle was acquired by using 2 proteins  
Skeletal actin and Troponin C.

# Lineage-specific muscles



Origin of each protein?

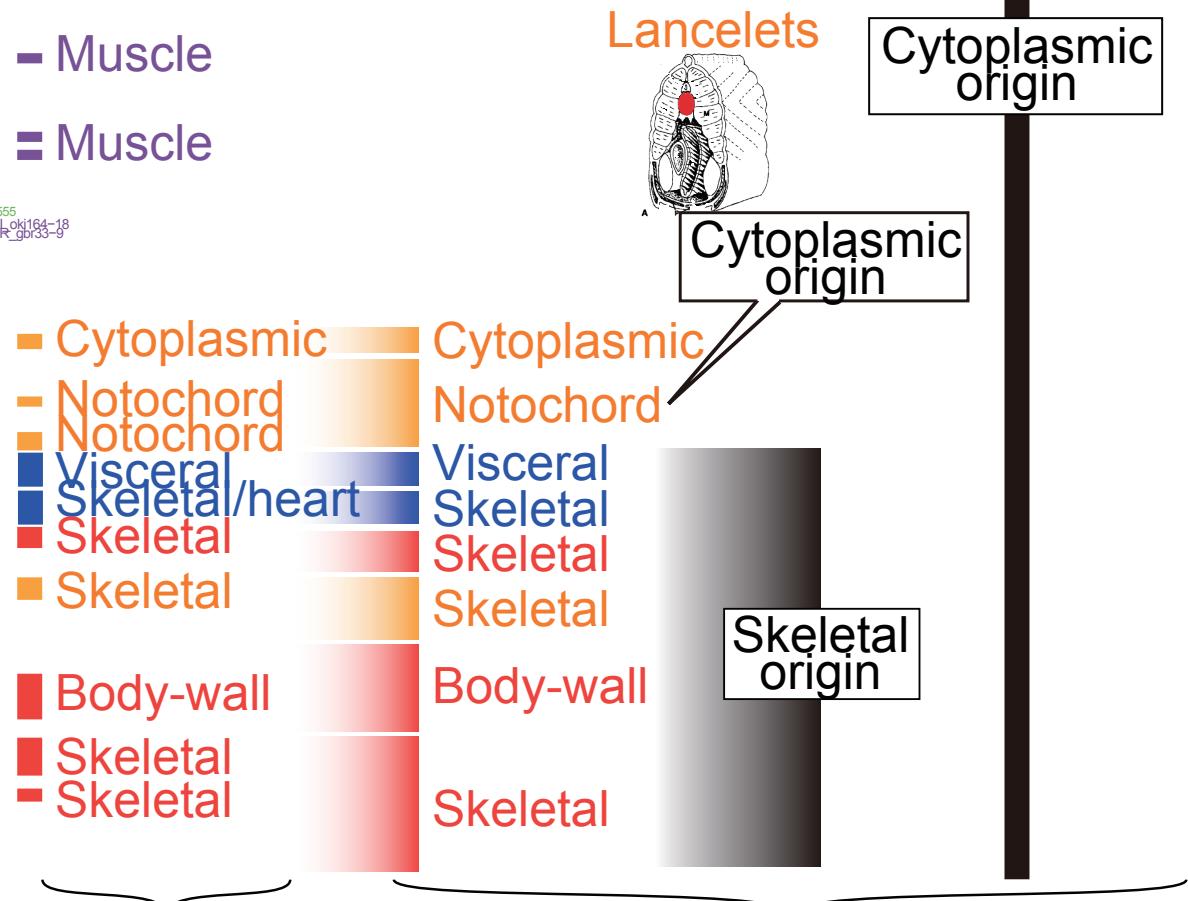
# Results: Estimation of actin origins



## Expression: Observed

Origins can be estimated using expression data.

- Cytoplasmic
  - Cytoplasmic
  - = Cytoplasmic
  - Muscle
  - = Muscle



# Cytoplasmic origin

# Lancelets

## Cytoplasmic origin

## Cytoplasmic

# Notochord

Viscera

Viseira  
Skeletal

Skeleta

# Skeletal

CRECITA

## Body-wall

1

Skeletal

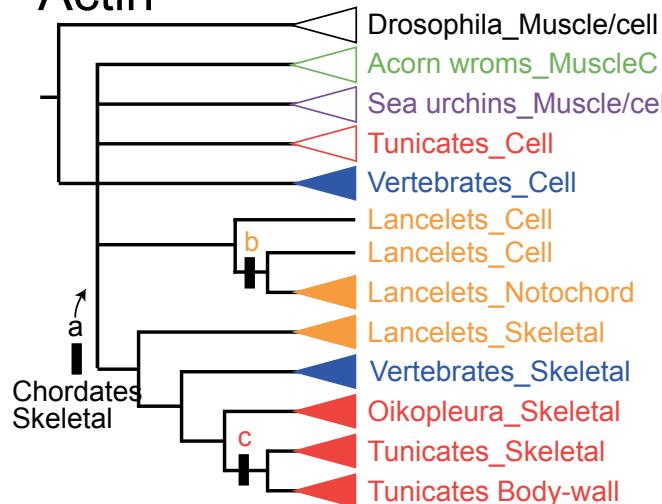
Skeletal

ANSWER

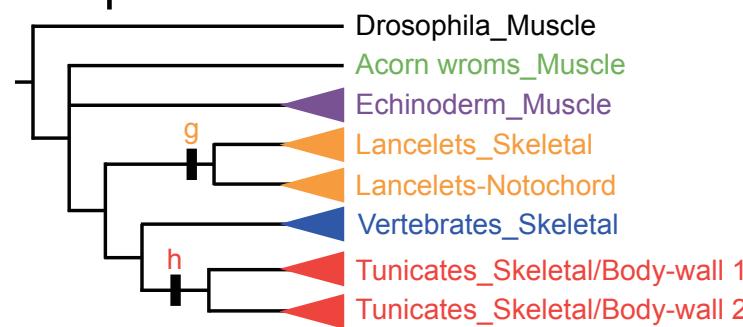
Ancestra

# Other gene trees

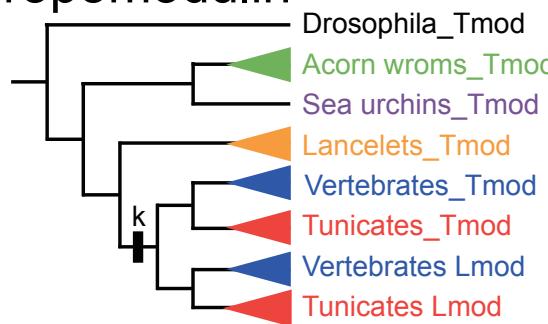
**Actin**



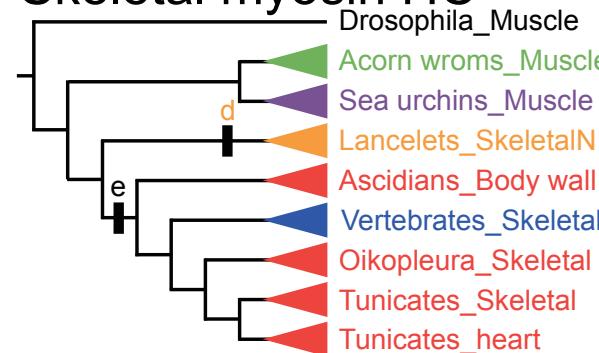
**Troponin I**



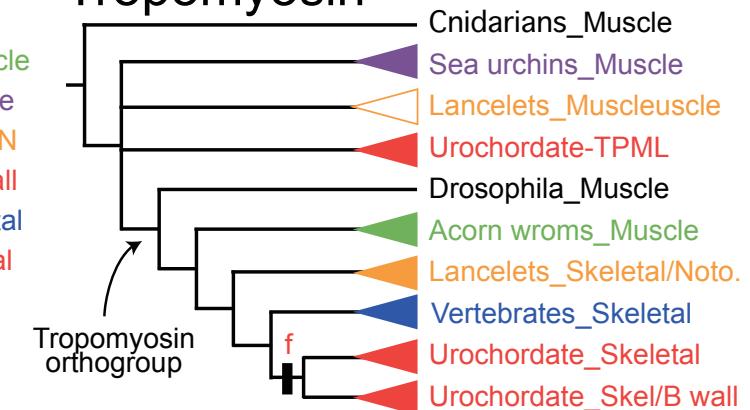
**Tropomodulin**



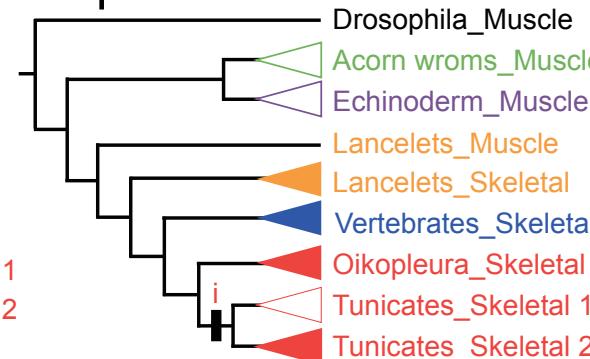
**Skeletal myosin HC**



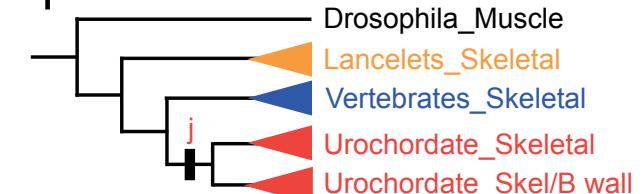
**Tropomyosin**



**Troponin T**



**Troponin C**



Estimate origin of each protein.

# Origins of specific-muscle proteins

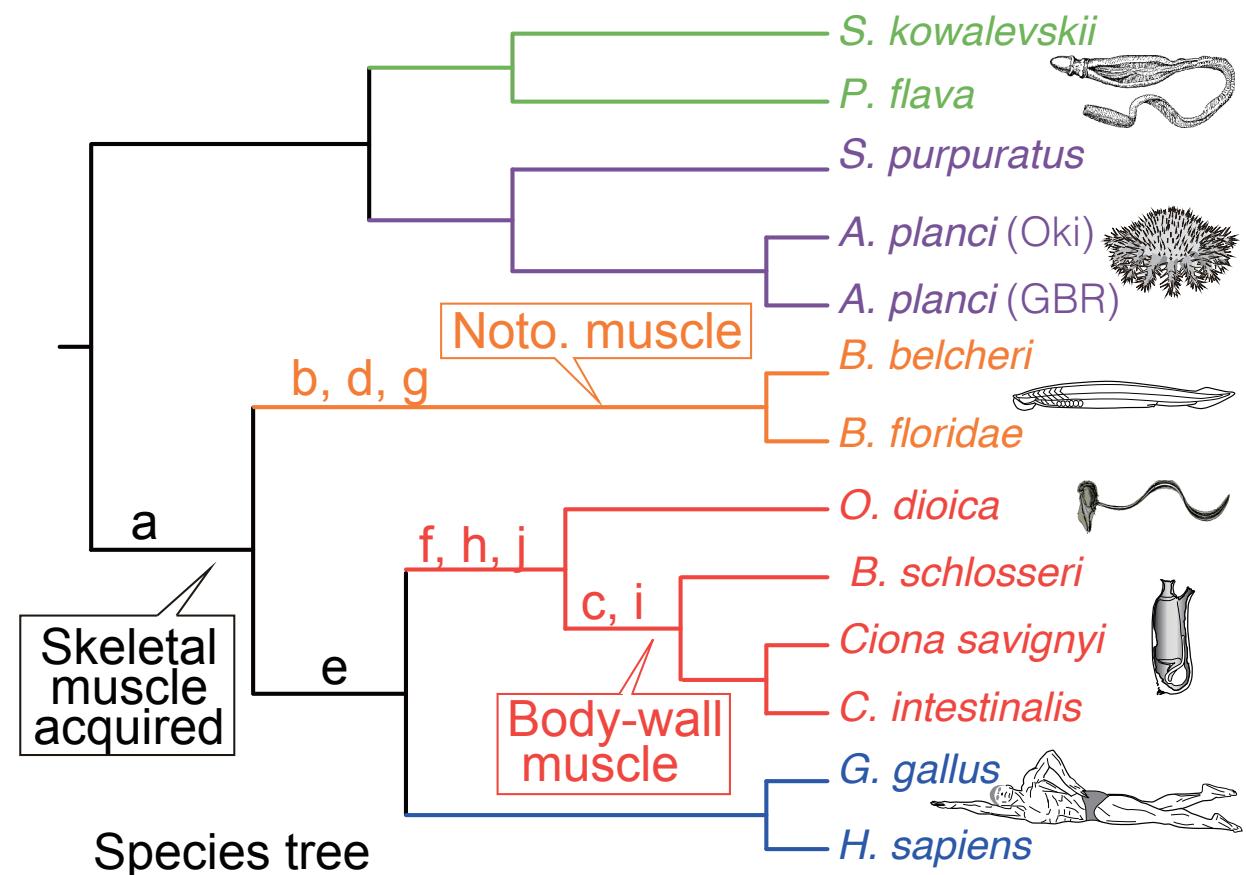
Ancestral Dup.

## 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