



UNIVERSITY OF
OXFORD



University of Essex

RECONSTRUCTING ANCESTRAL GENOMES TO UNDERSTAND MAJOR TRANSITIONS

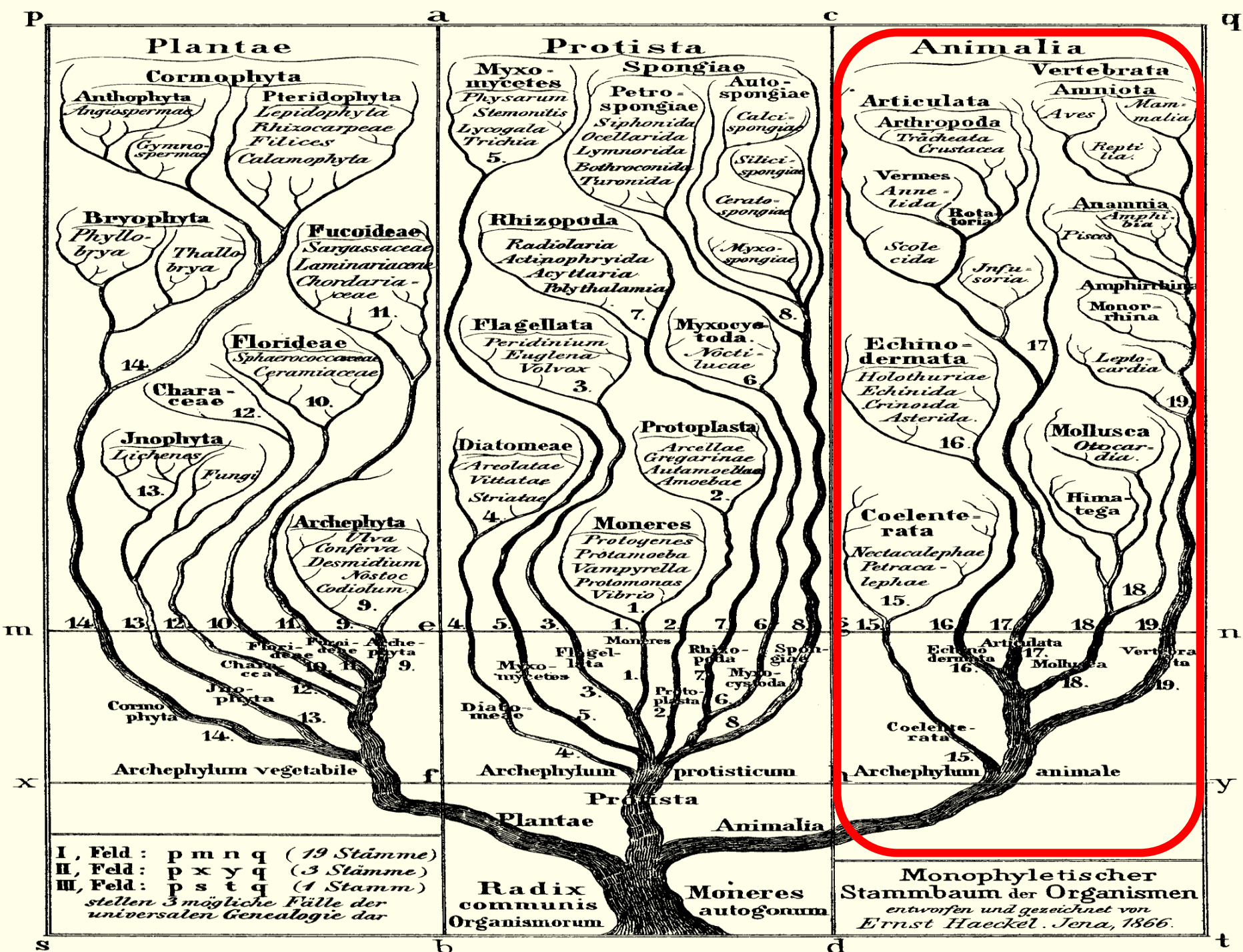
Jordi Paps



@JordiPaps



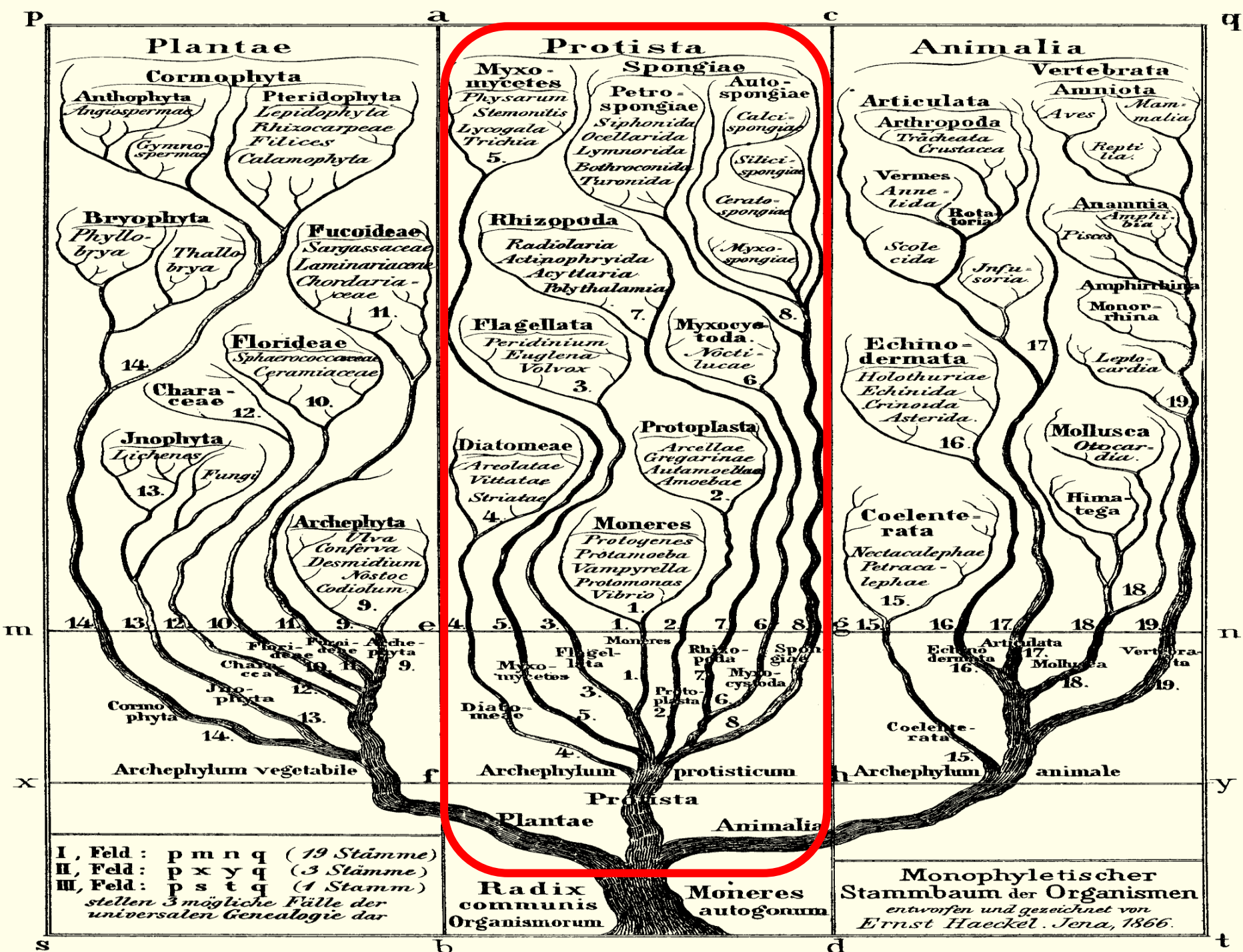
University of
BRISTOL



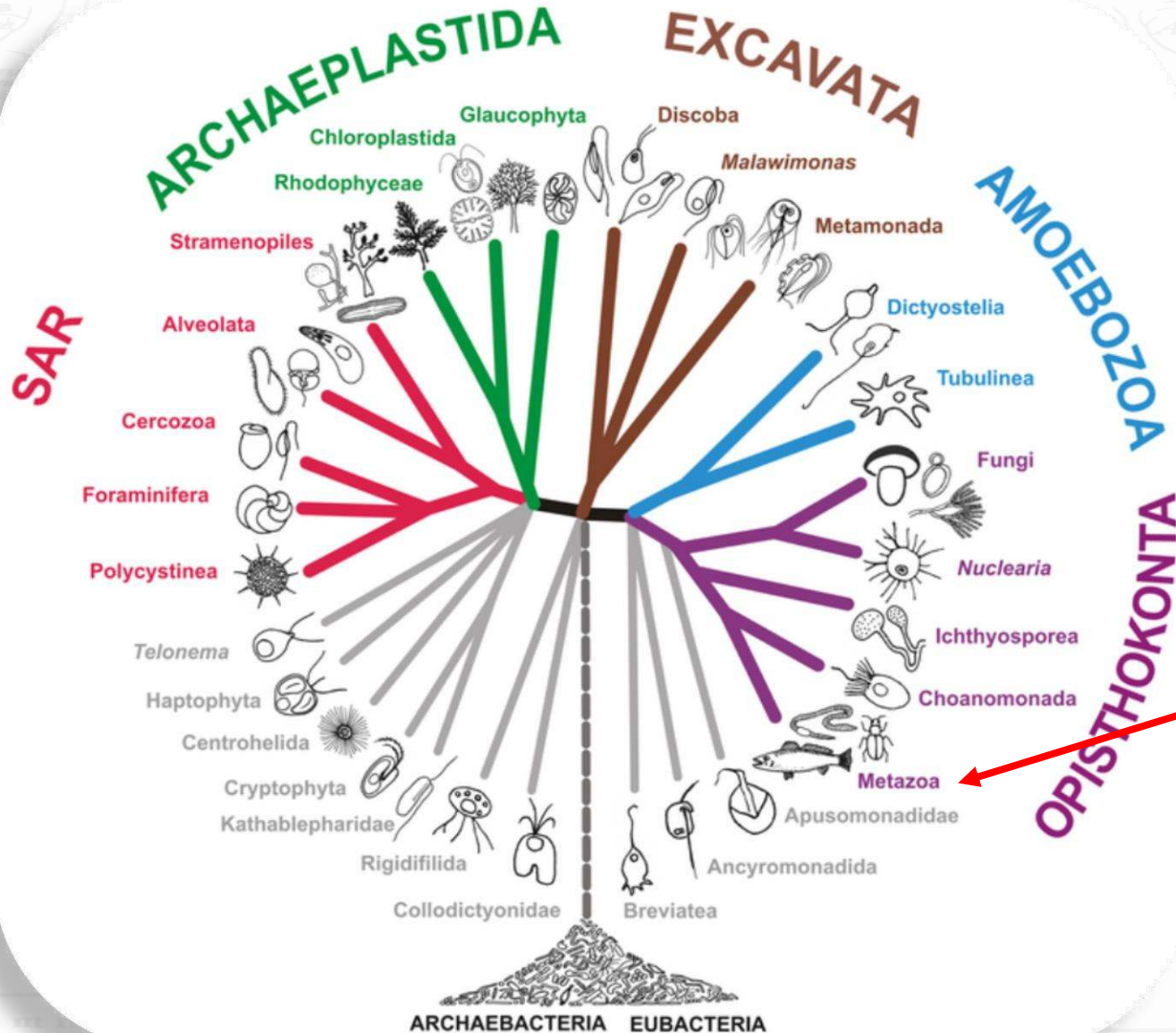
Animals



THE ORIGIN OF THE ANIMAL KINGDOM

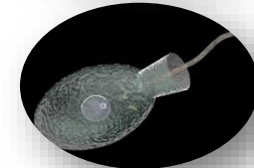
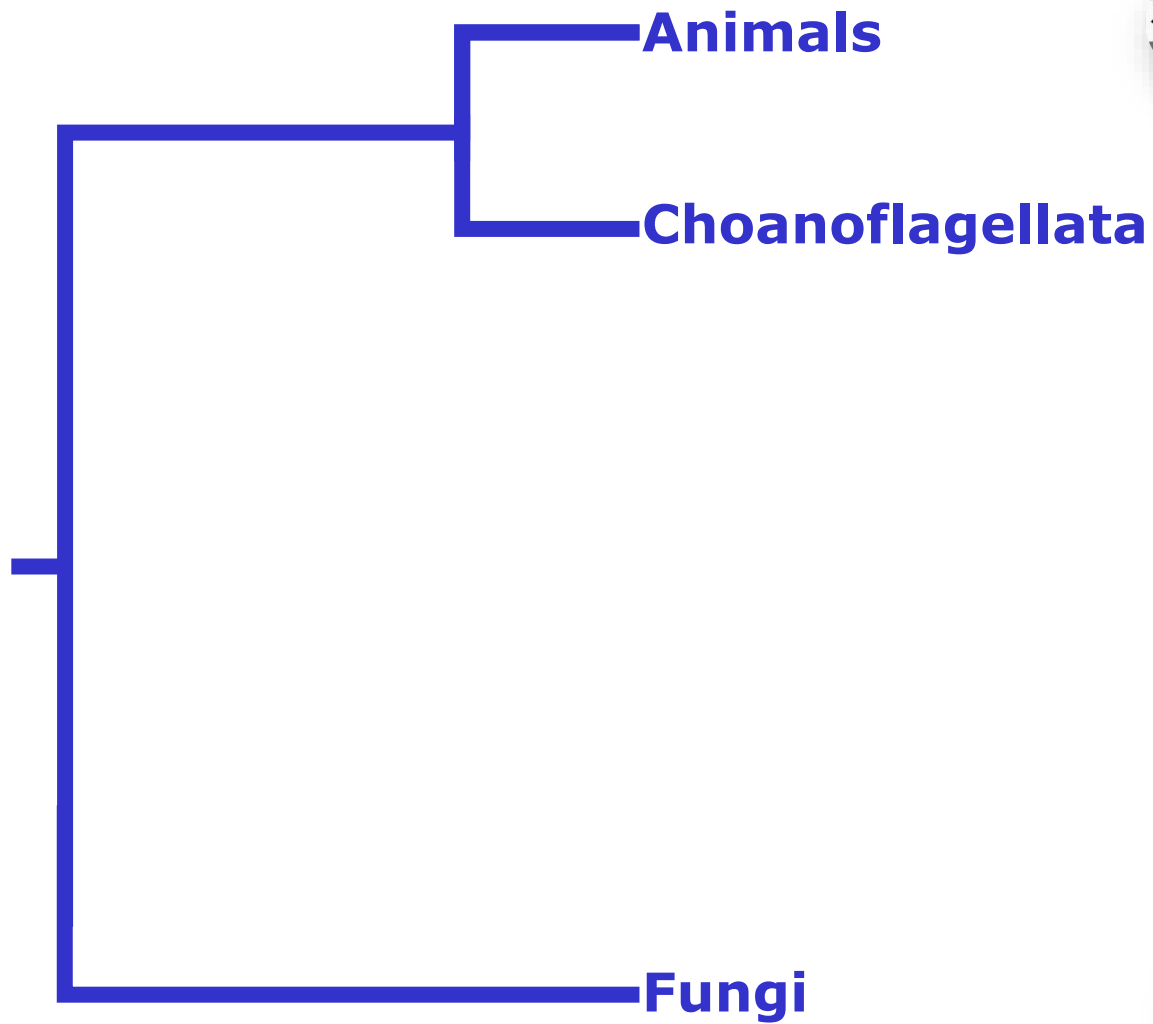


Where?



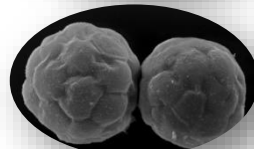
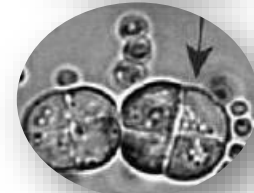
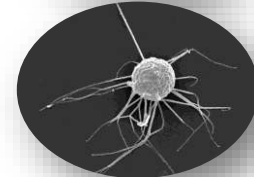
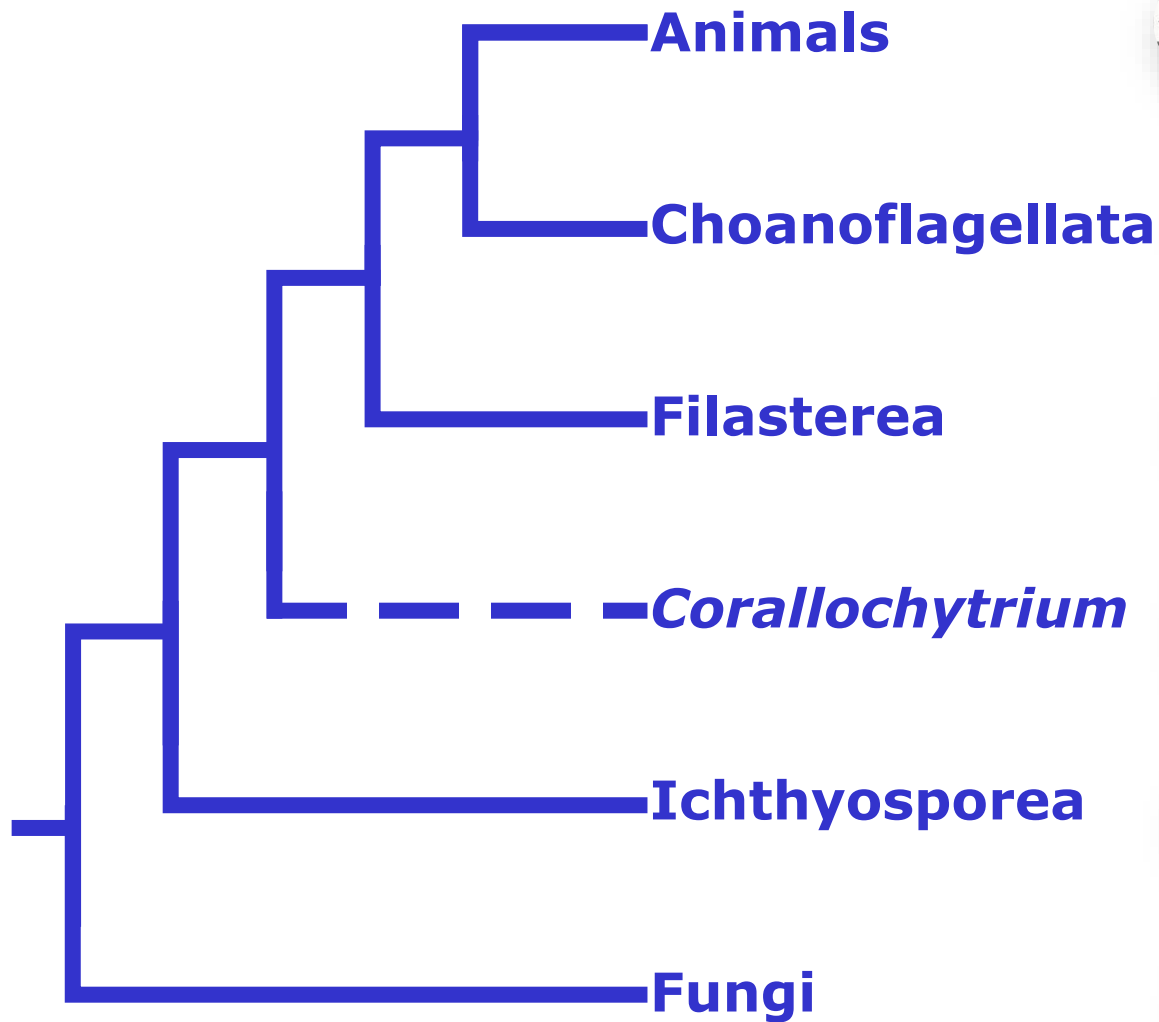
Opisthokonta

Animal Origins



Opisthokonta

Animal Origins



Choanoflagellata

Animal Origins

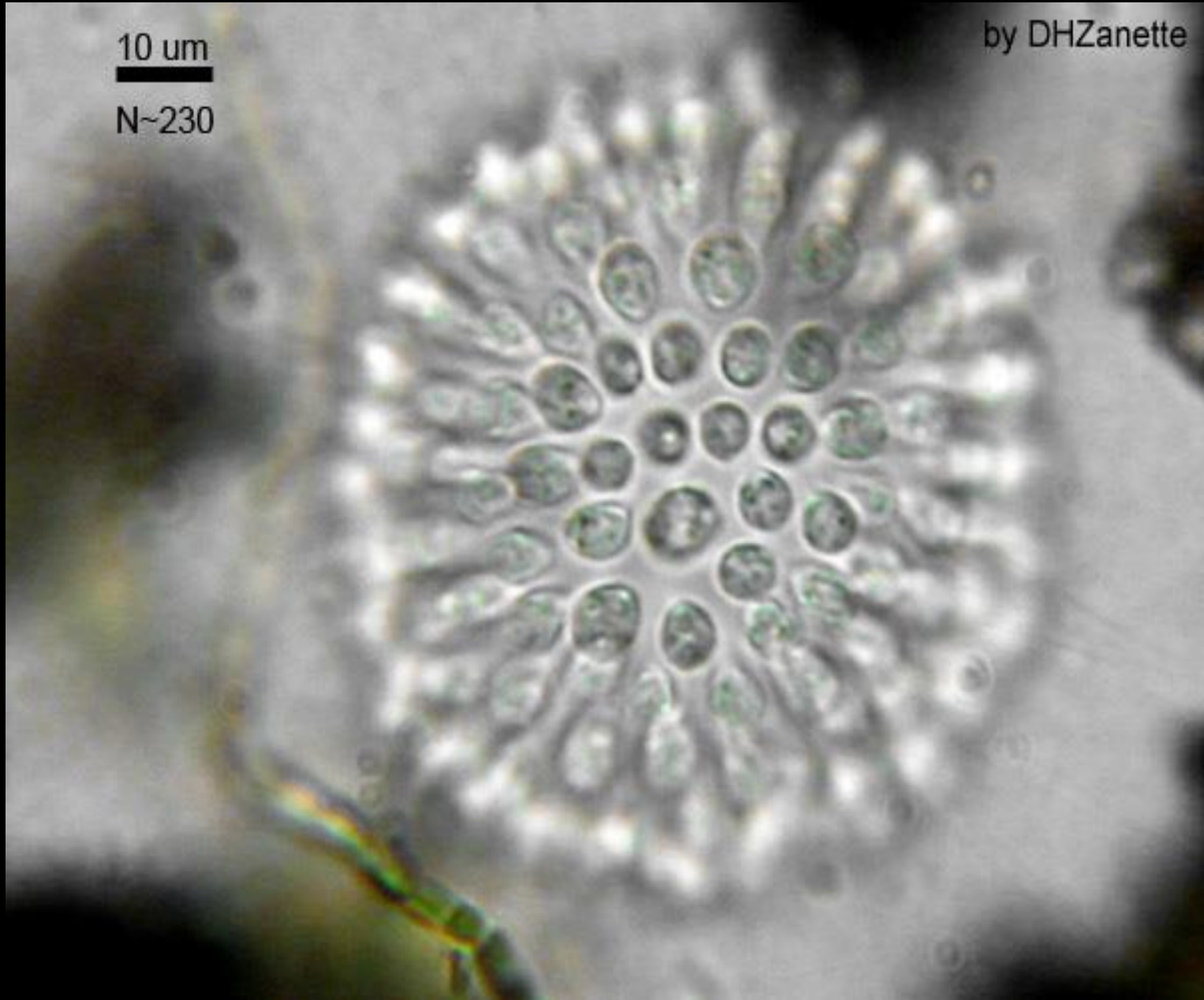
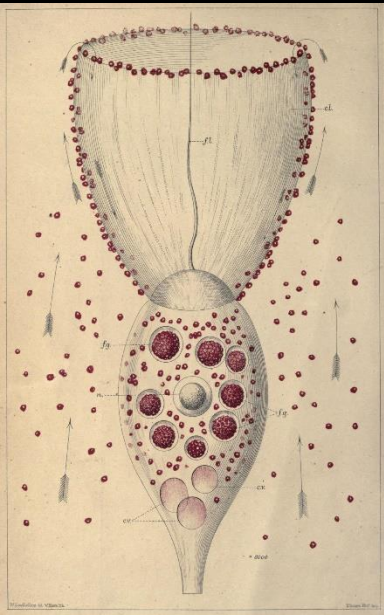
Nicole King



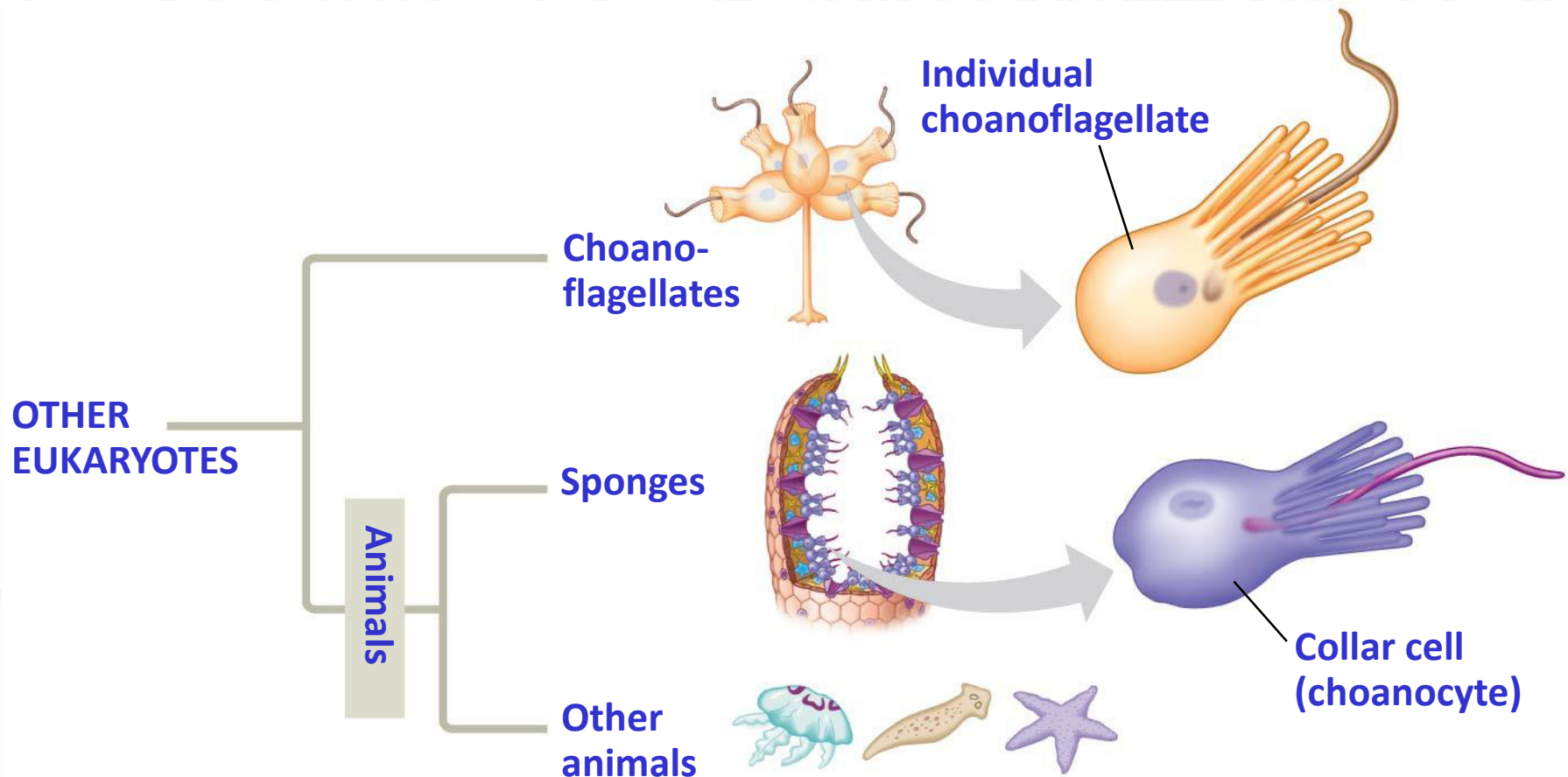
10 μ m

N~230

by DHZanette

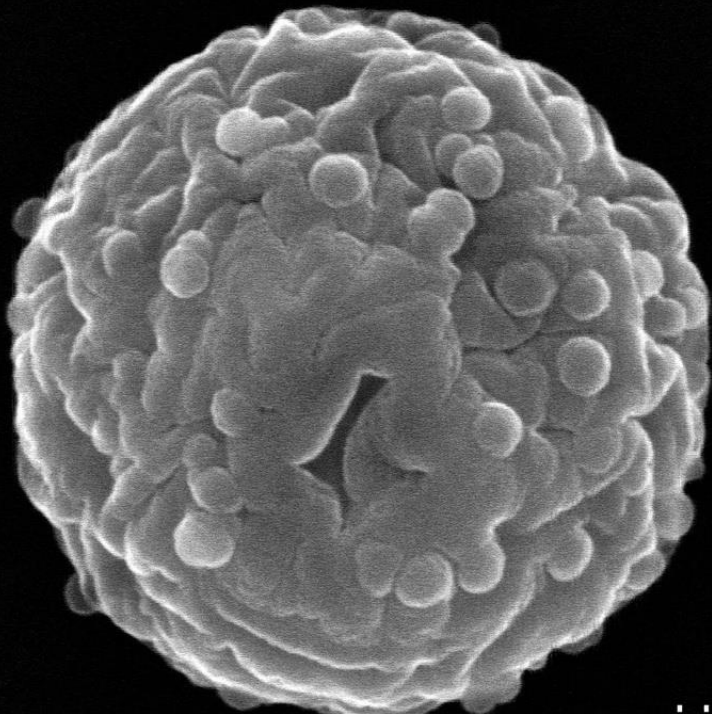


Origin of Multicellular Animals



- **Morphological and molecular** evidence points to **choanoflagellates** as the closest living relatives to animals

Filasterea (*Capsaspora*)



SCT-UB 20.0kV X25.0k

Ichthyosporea (*Sphaeroforma*)



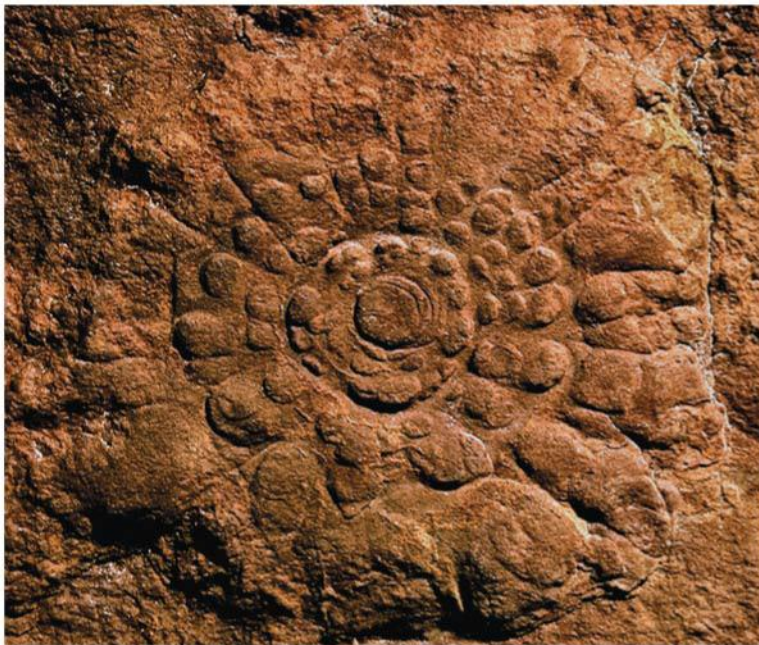
Iñaki Ruiz-Trillo



[illegible]

Ediacaran biota

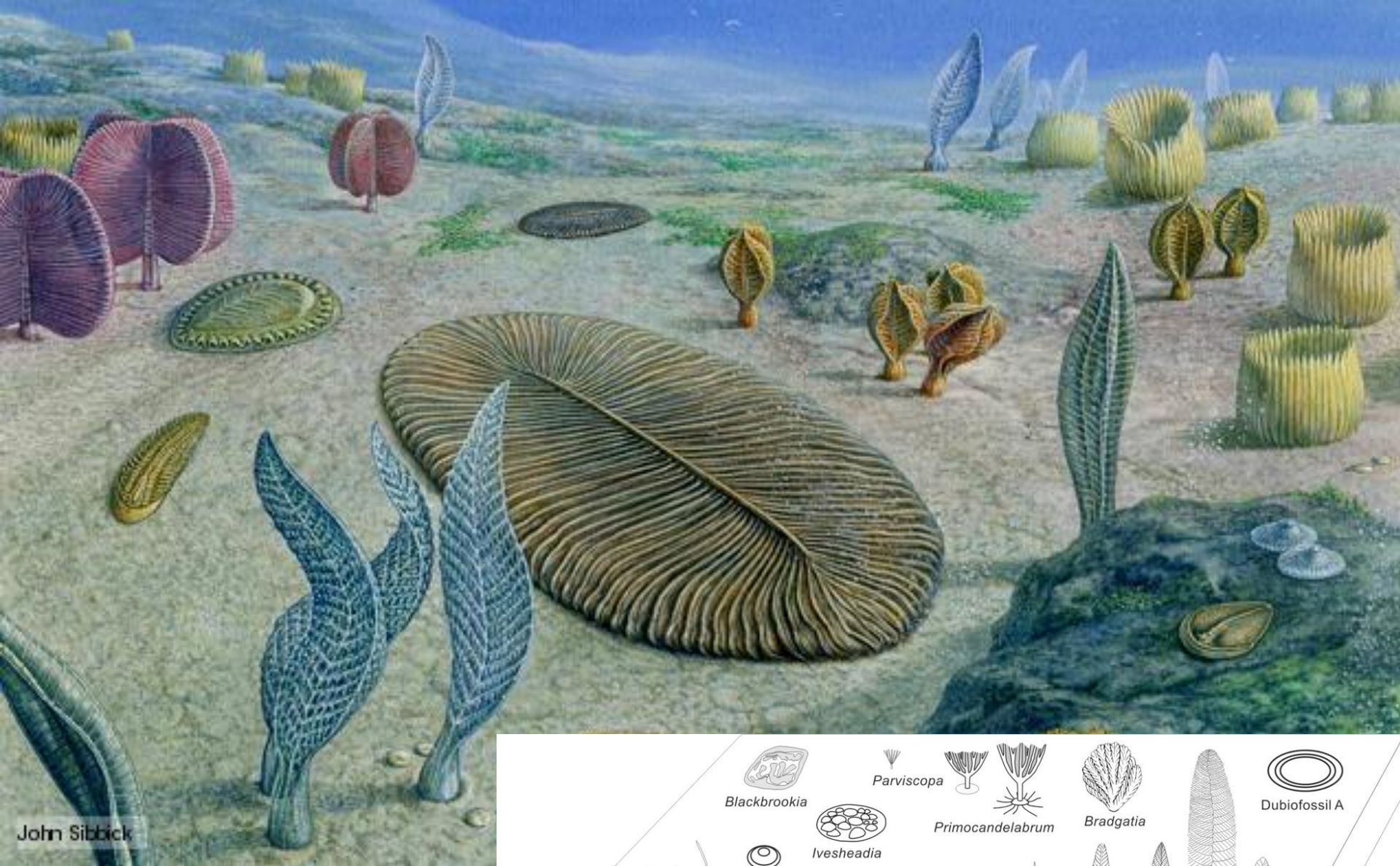
- **Ediacaran biota** (560 million years ago) composed by **multicellular organisms**. Strong debate about their nature: **animals or not?**



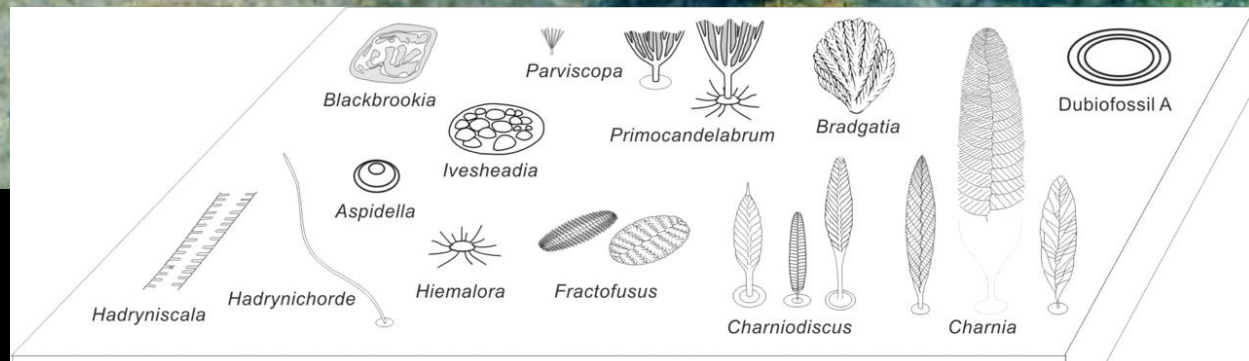
(a) *Mawsonites spriggi*



(b) *Spriggina floundersi*



John Sibbick



Bonavista assemblage

The image displays a phylogenetic tree of life, with major groups labeled in blue. The tree is rooted at the bottom and branches upwards. The groups are organized into several main categories: Bacteria, Eukarya, and Archaea. The Bacteria branch is on the left, Eukarya in the center, and Archaea on the right. The Eukarya branch further divides into Plantae, Fungi, and Animalia. The Archaea branch includes groups like Ciliophora, Metazoa, and others. The tree is annotated with key evolutionary milestones in millions of years ago (Mya): 4,000 Mya (Origin of life), 2,000 Mya (Oldest eukaryote fossils), 560 Mya (Ediacaran biota), and 542 Mya (Cambrian explosion). The tree also shows the emergence of major groups like Bacteria, Eukarya, Archaea, and the three domains of life.

When?

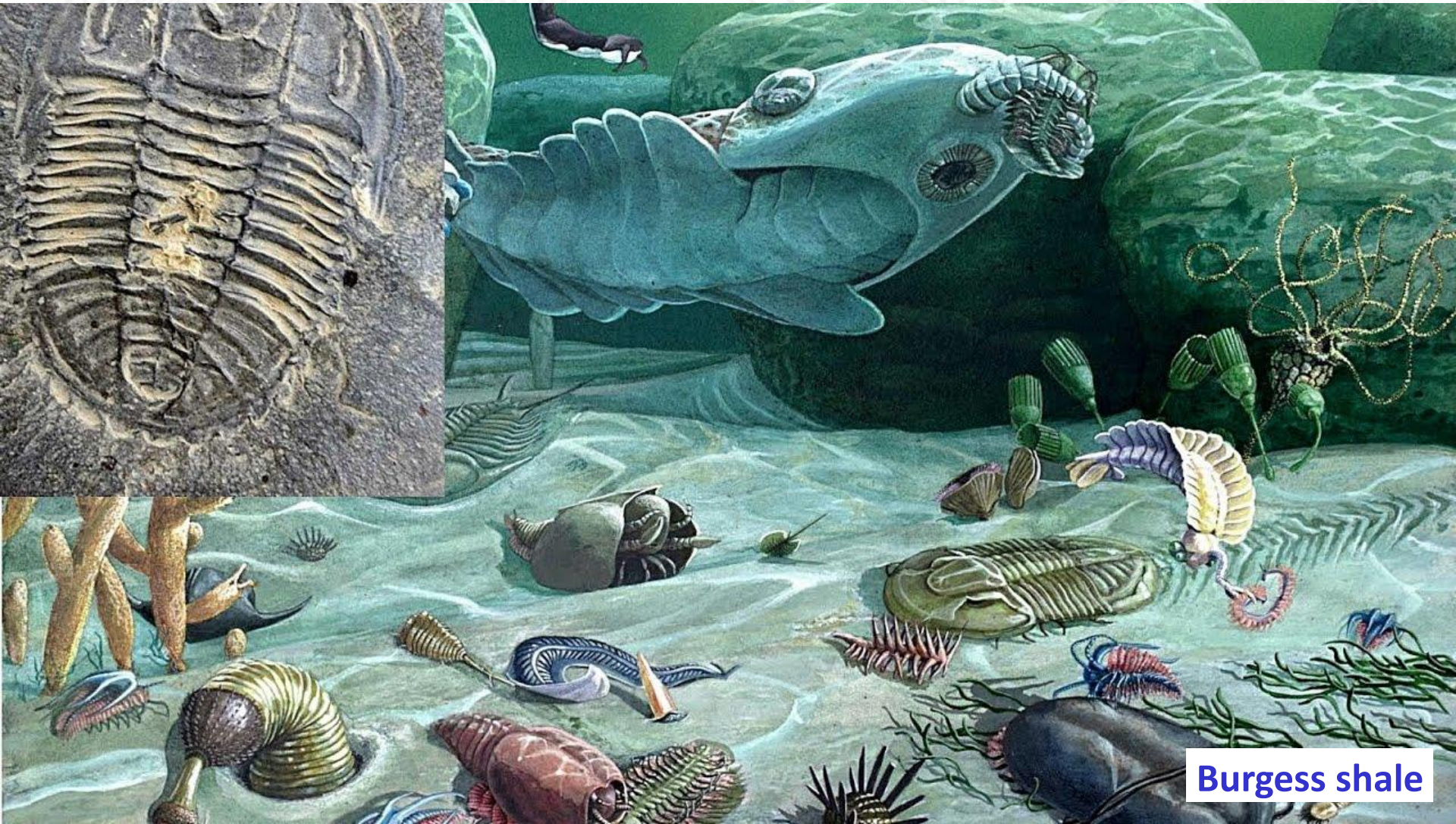
542 Mya Cambrian explosion

560 Mya Ediacaran biota

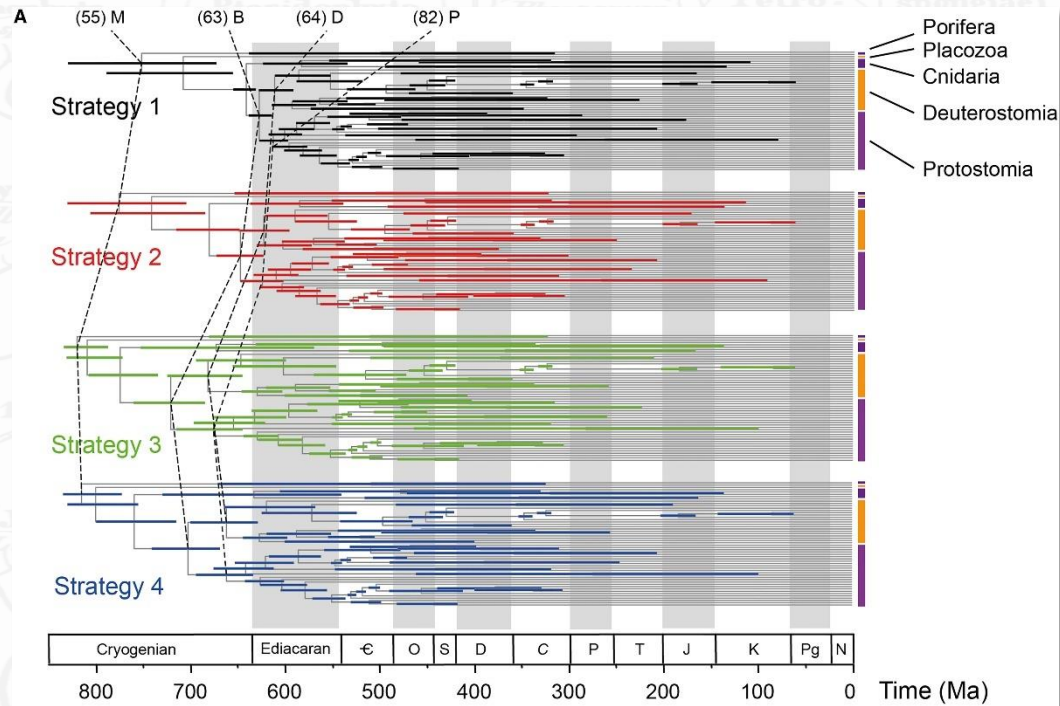
2,000 Mya Oldest eukaryote fossils

4,000 Mya Origin of life

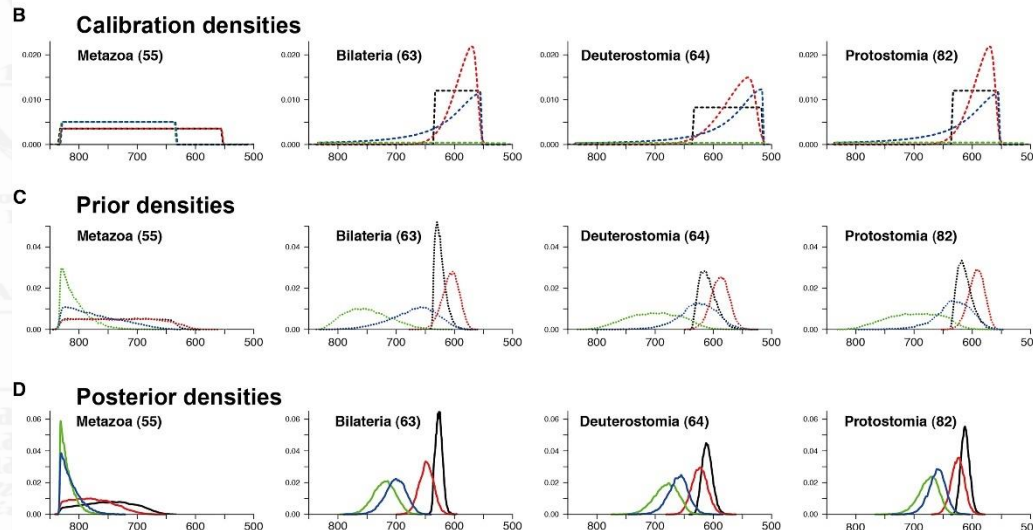
The Cambrian explosion (535 to 525 Mya): sudden fossil appearance of all the major groups of extant groups animals



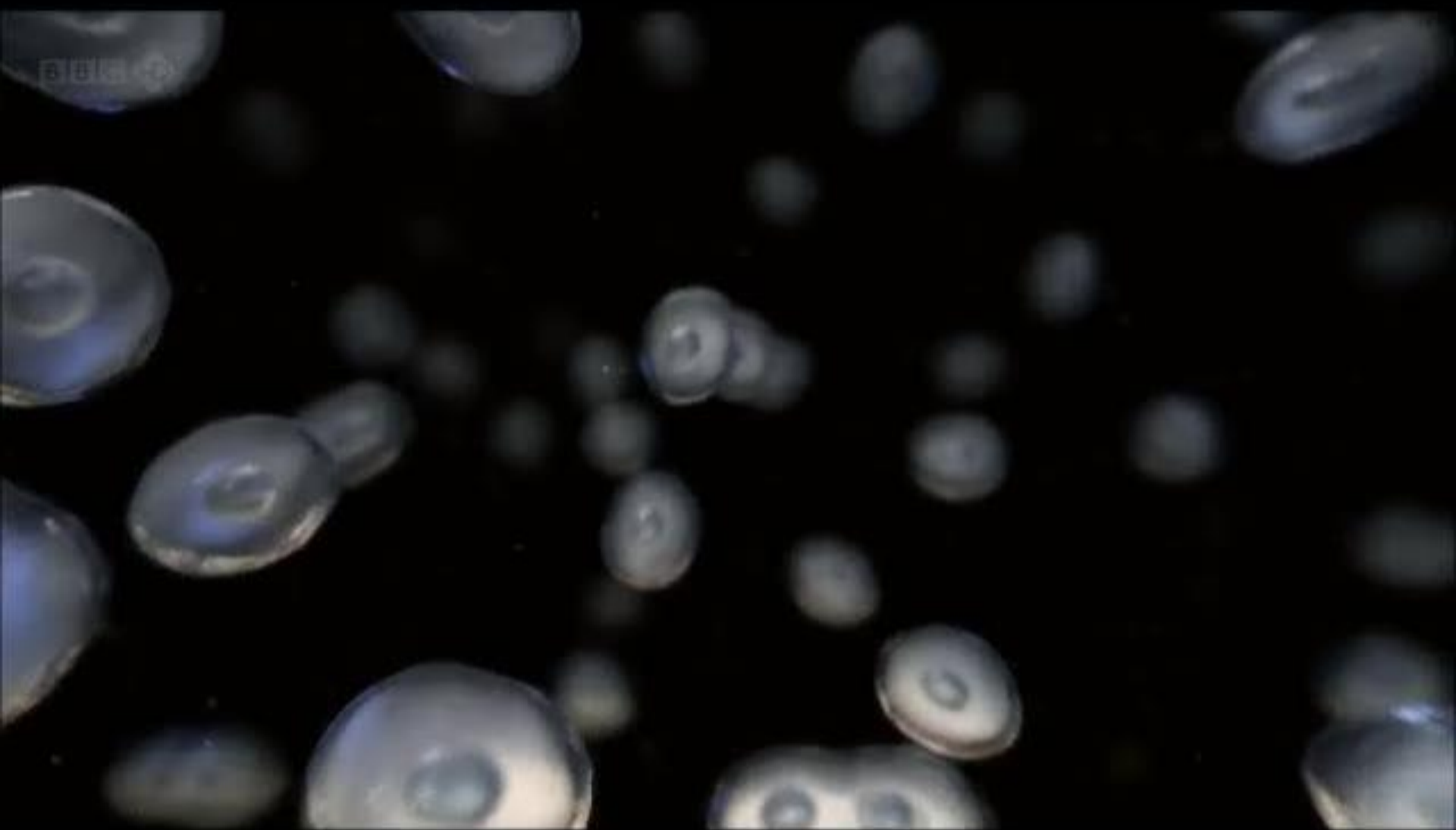
When?

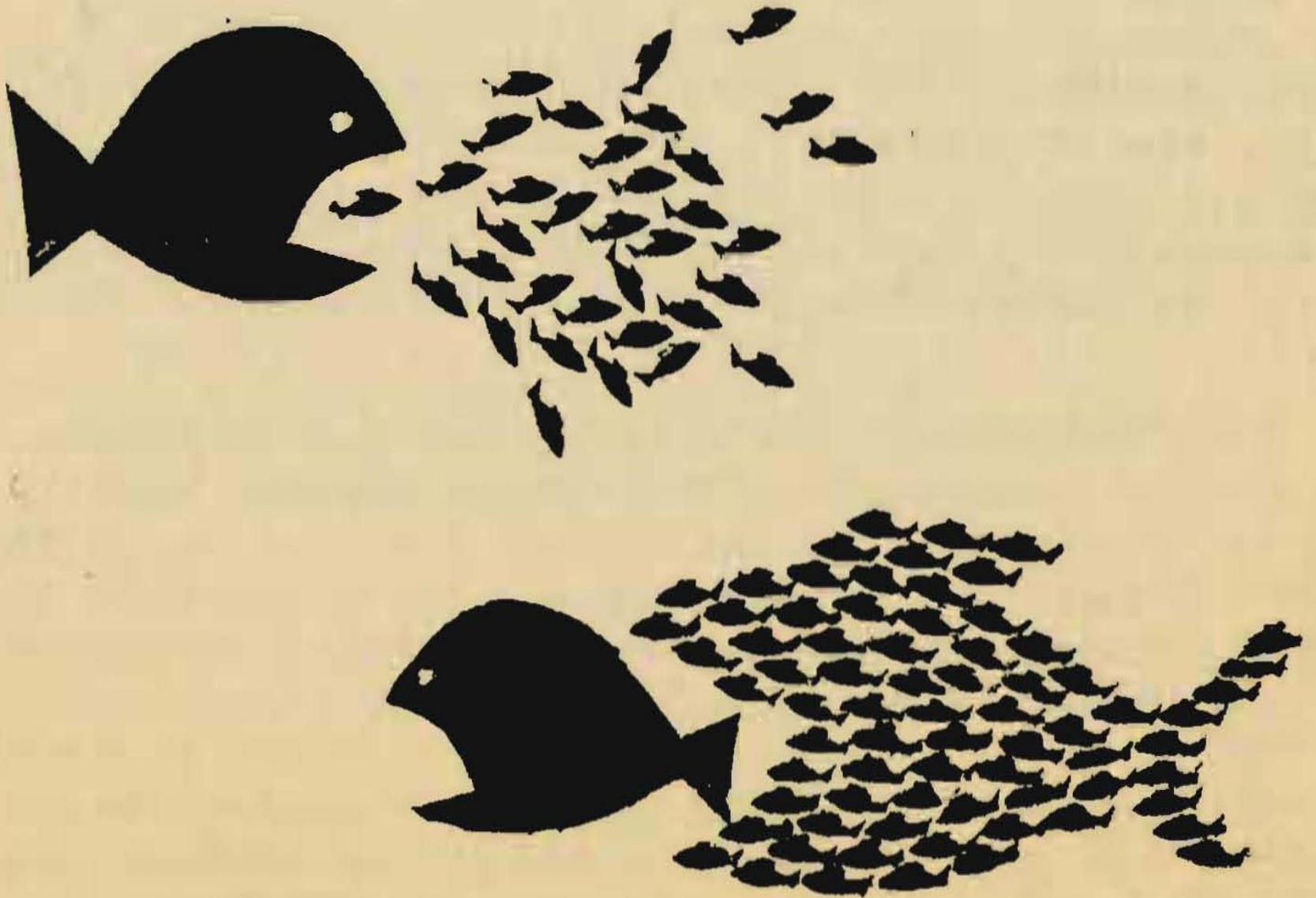


833–650 Ma



Dos Reis et al 2015



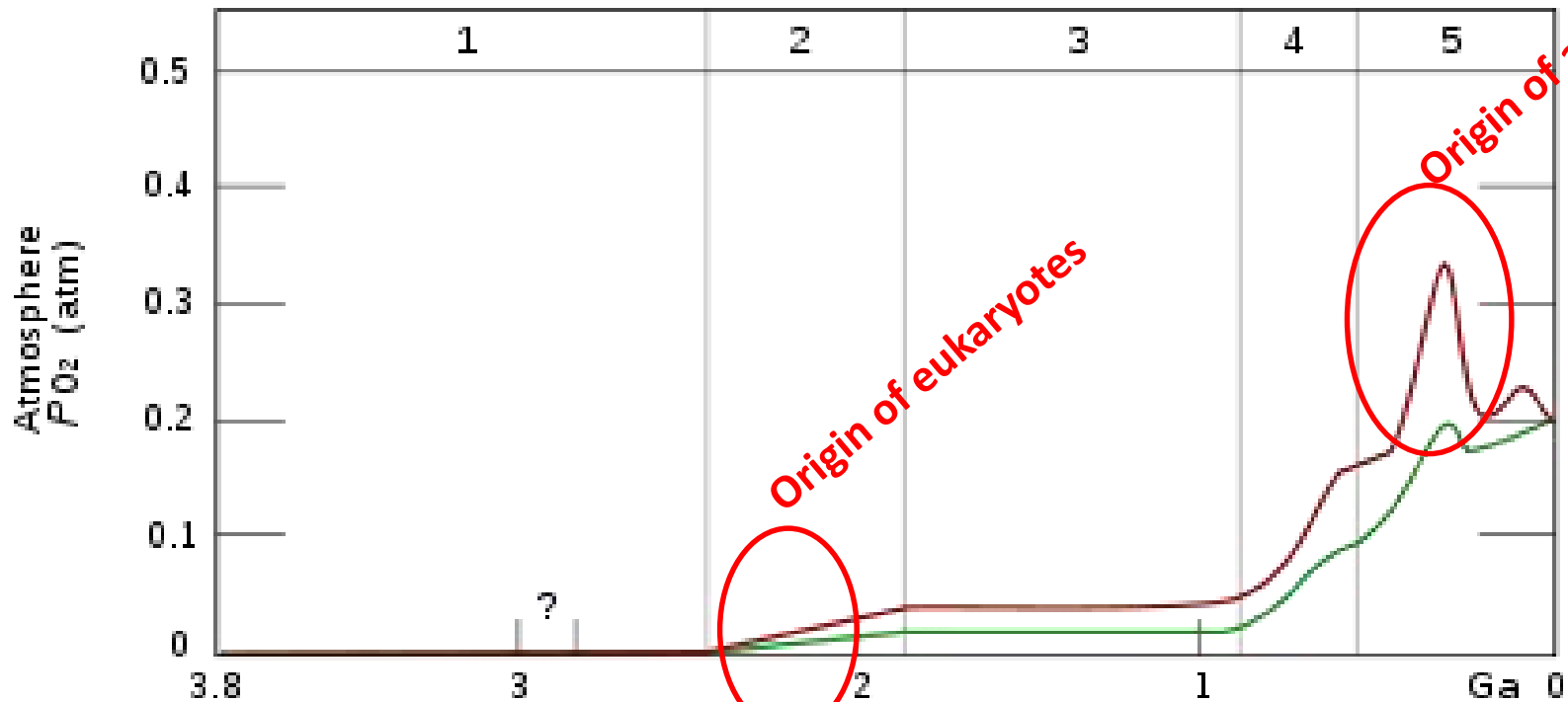


Environment

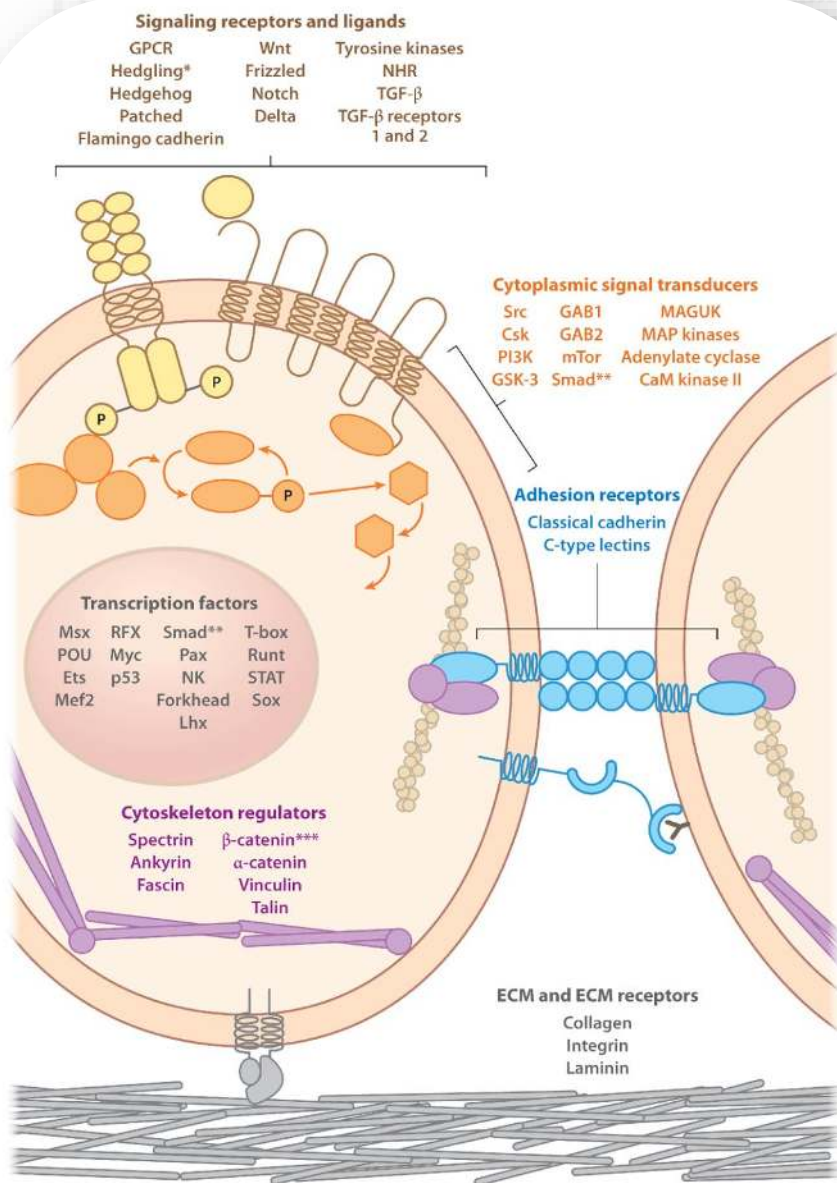
Triggers



Stages



Genome



- **Gene regulation** (TFs, and signalling pathways)
- **Cell adhesion**
- **Cell type specialization**
- **Cell cycle**
- **Immunity**

HOW TO MAKE A BABY?

How to make a baby...



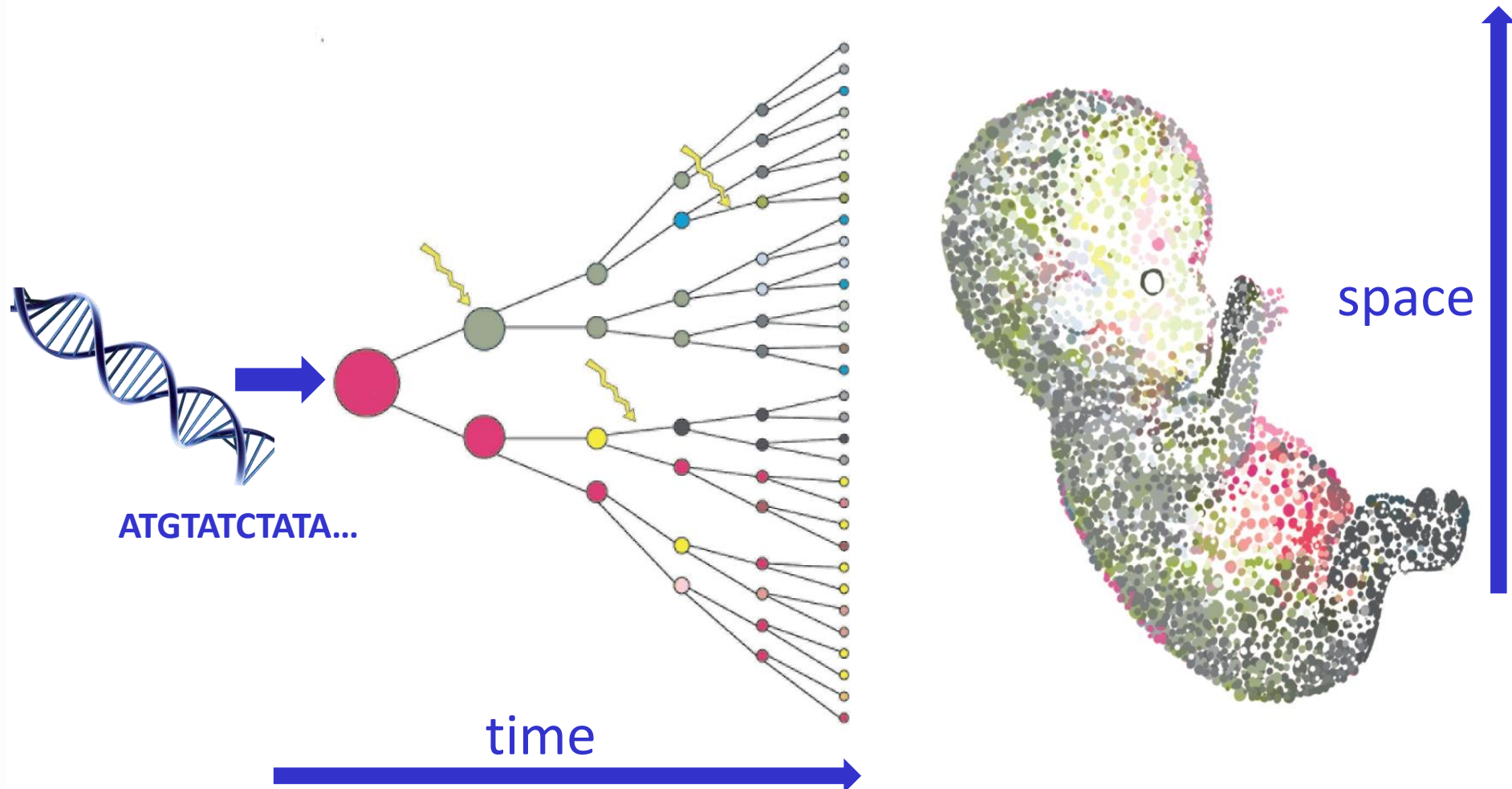
I, Feld: p m n q (19 Stämme)
 II, Feld: p x y q (3 Stämme)
 III, Feld: p s t q (1 Stamm)
 stellen 3 mögliche Fälle der
 universalen Genealogie dar

Radix
 communis
 Organismorum

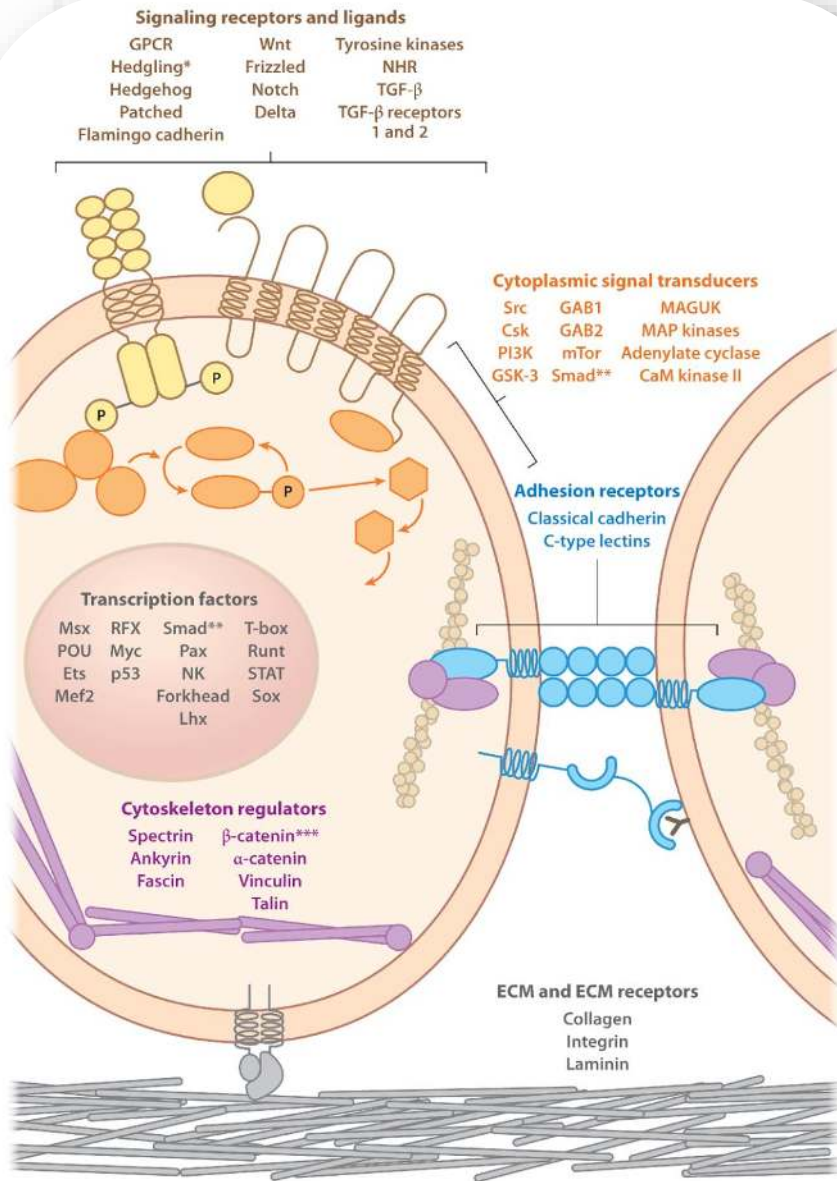
Moneres
 autogonum

Monophyletischer
 Stammbaum der Organismen
 entworfen und gezeichnet von
 Ernst Haeckel, Jena, 1866.

Development converts info from 1D (DNA) to 4D



Genome



• **Gene regulation** (TFs, and signalling pathways)

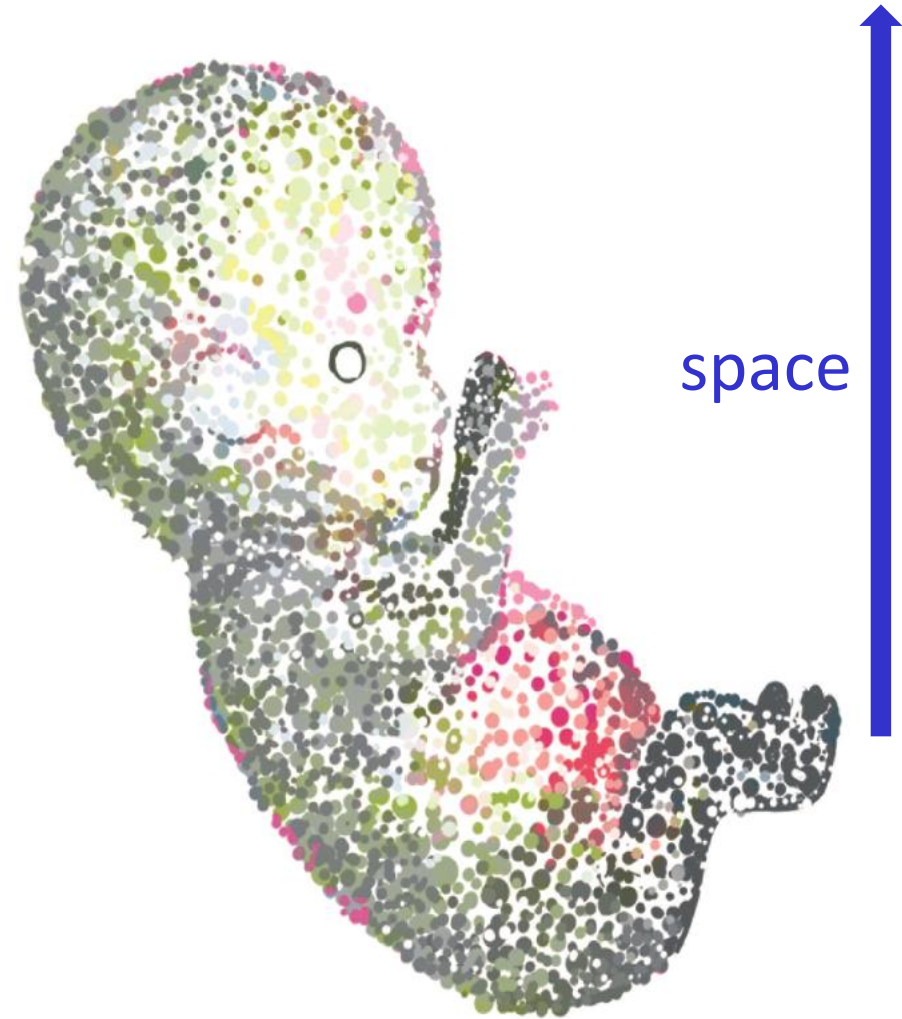
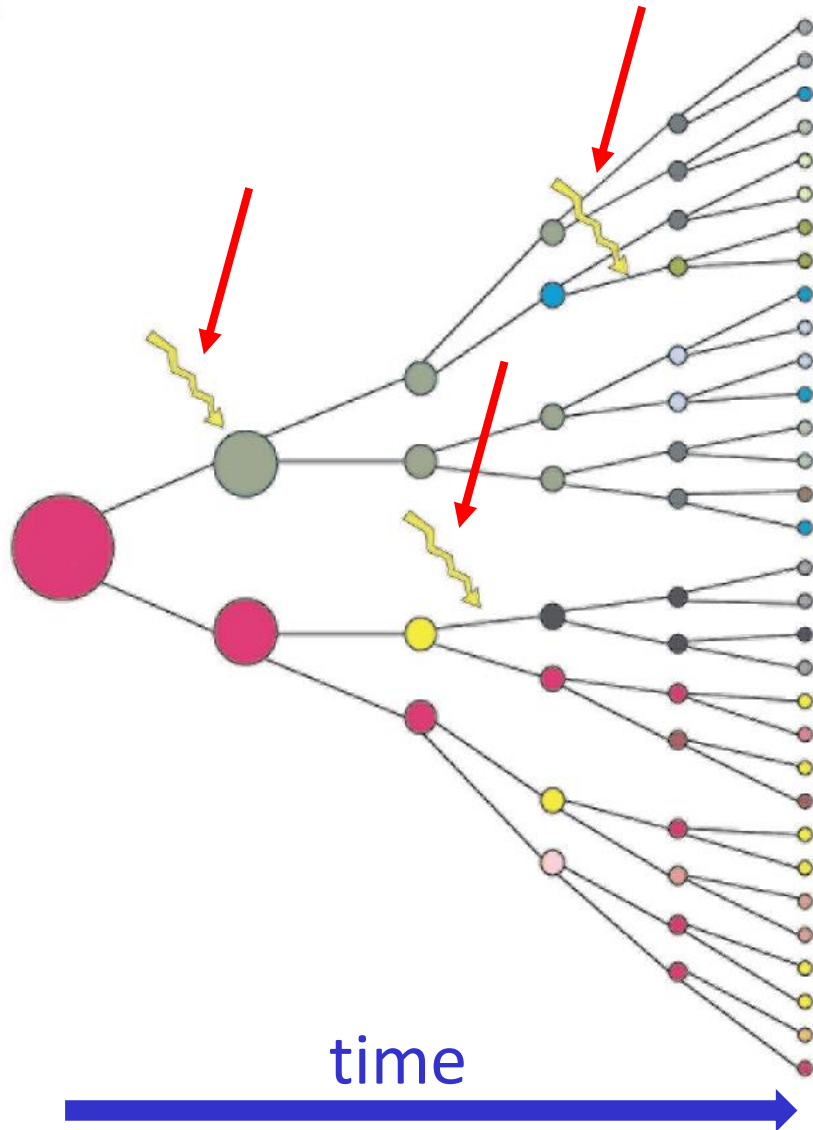
• **Cell adhesion**

• **Cell type specialization**

• **Cell cycle**

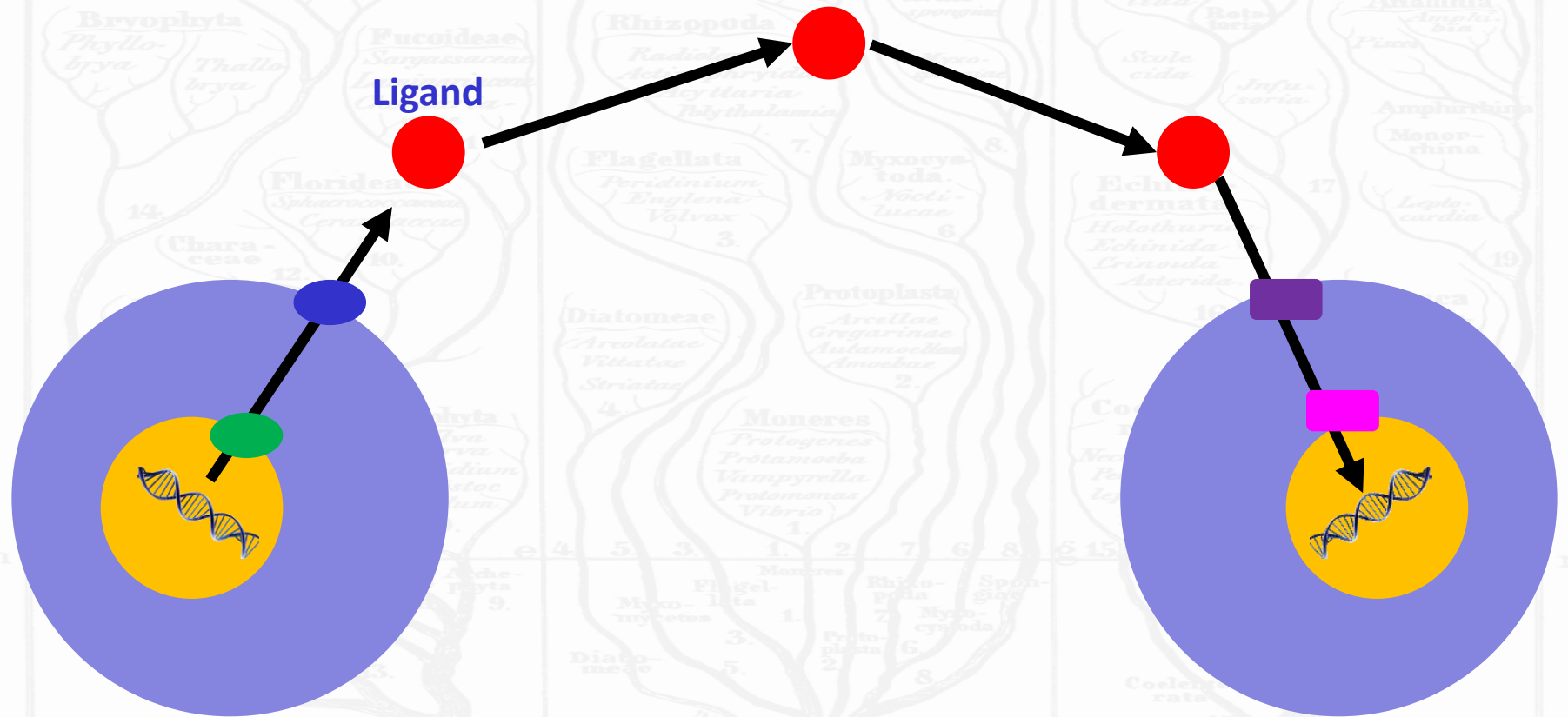
• **Immunity**

Differential gene expression



Signalling pathways: cell communication

Genome



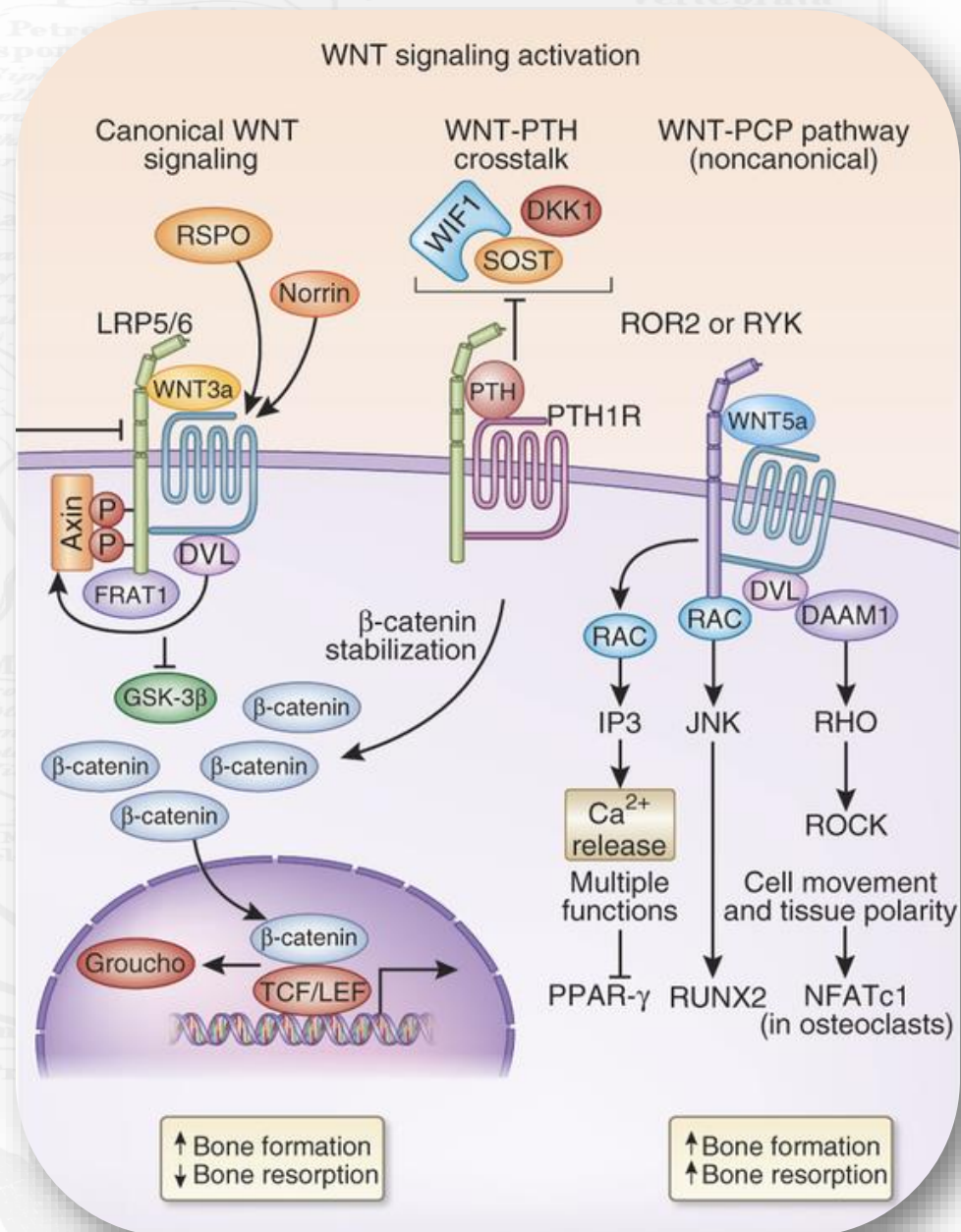
I, Feld: p m n q (19 Stämme)
 II, Feld: p x y q (3 Stämme)
 III, Feld: p s t q (1 Stamm)
 stellen 3 mögliche Fälle der
 universalen Genealogie dar

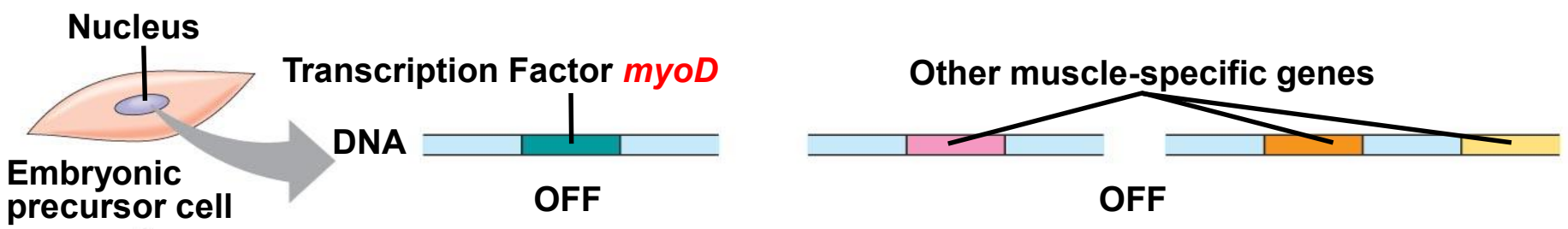
Radix
 communis
 Organismorum

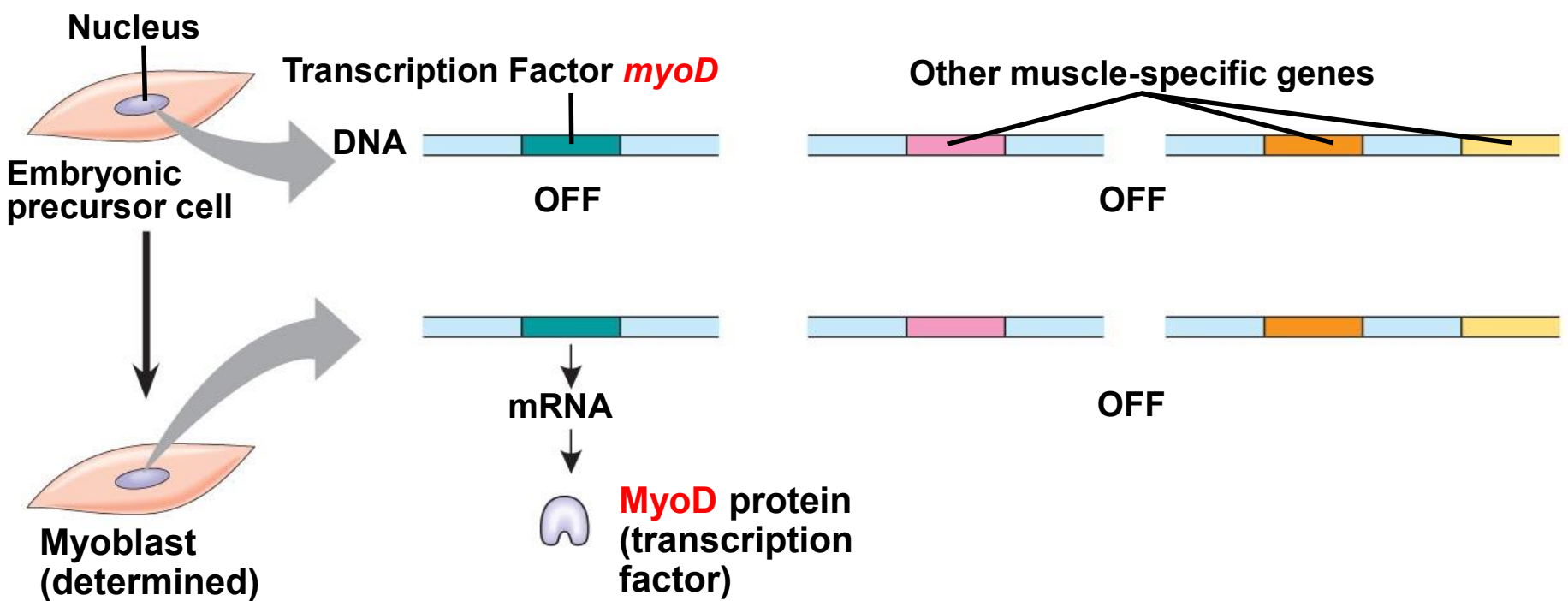
Moneres
 autogenum

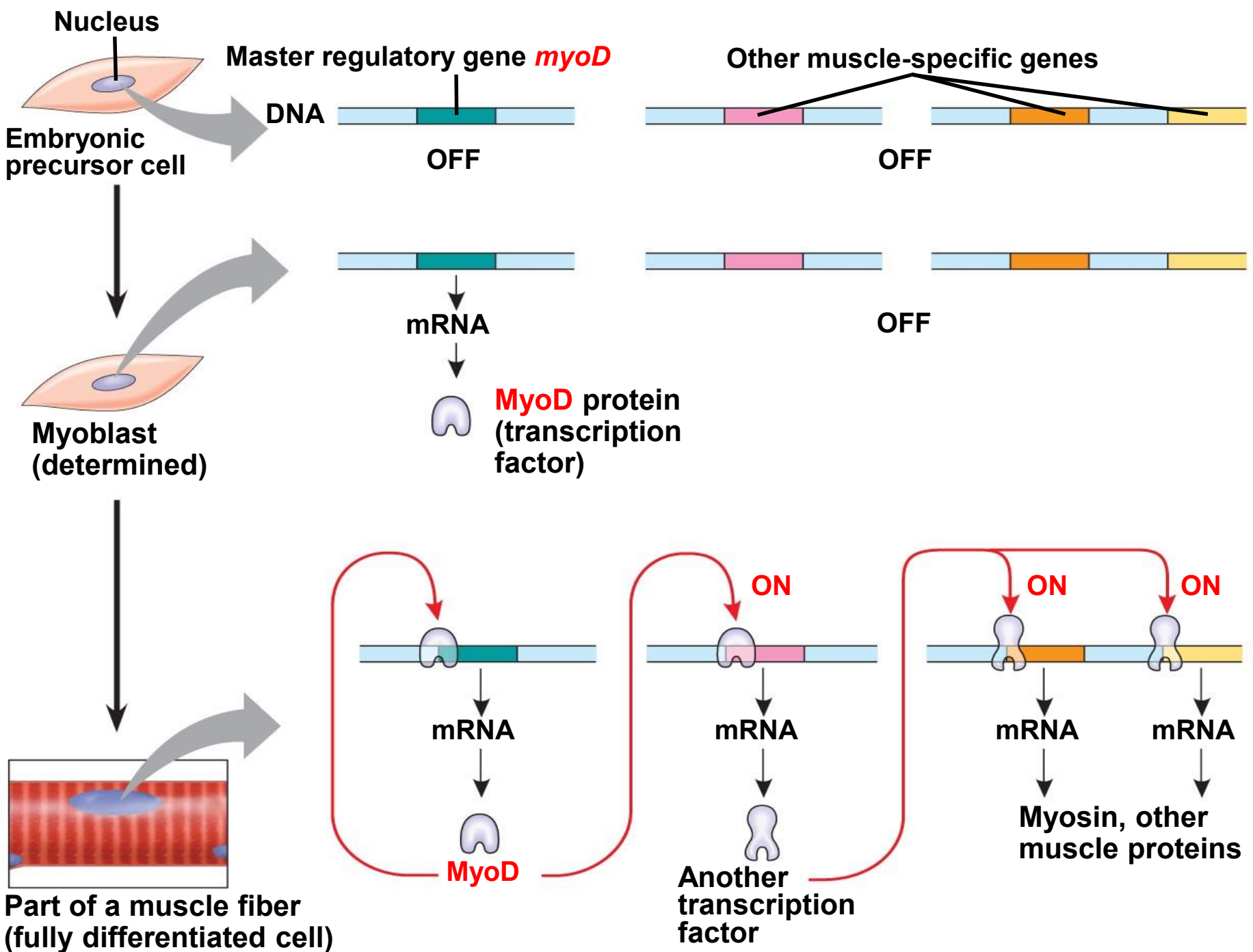
Monophyletischer
 Stammbaum der Organismen
 entworfen und gezeichnet von
 Ernst Haeckel, Jena, 1866.

Signalling pathways: Wnt

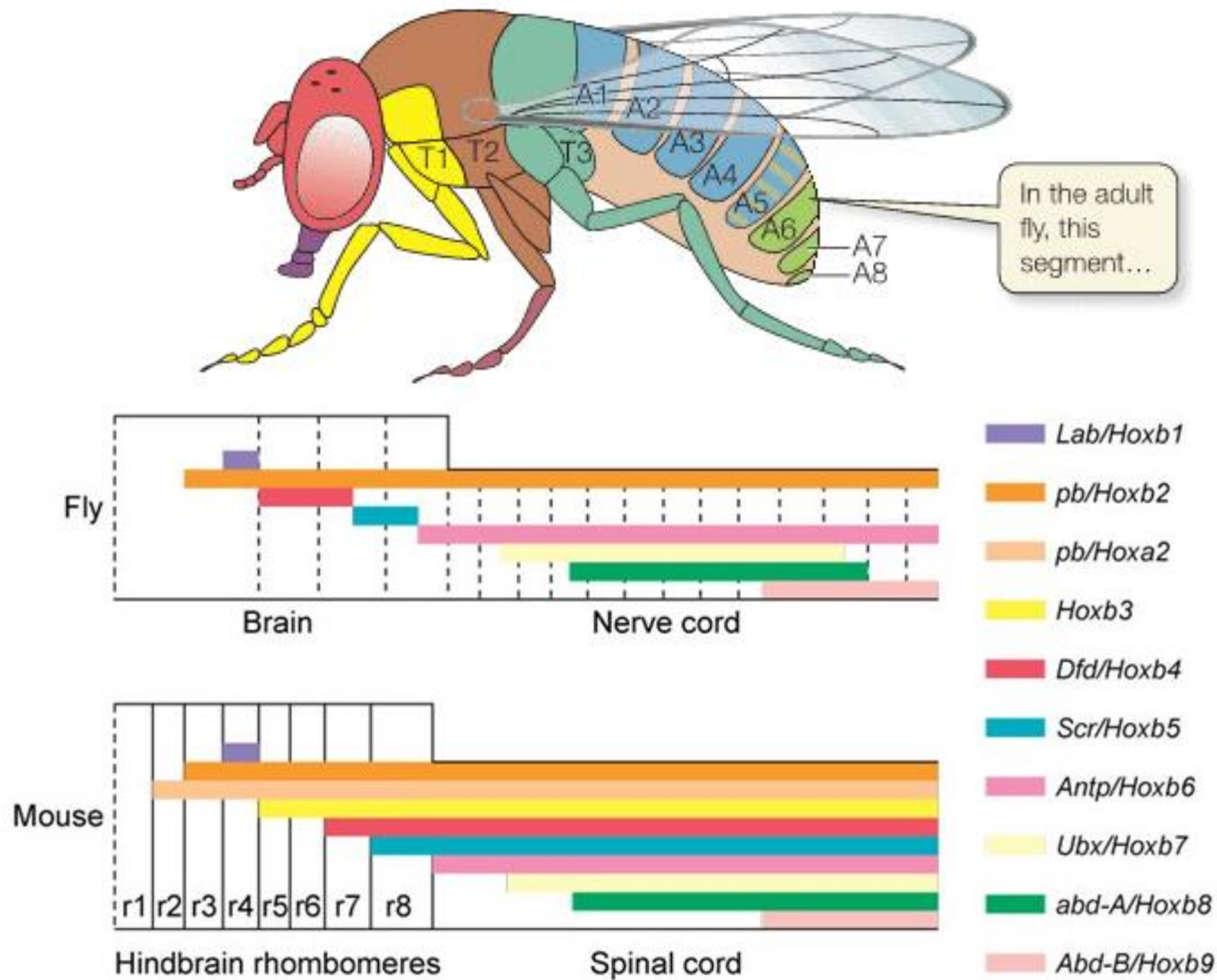




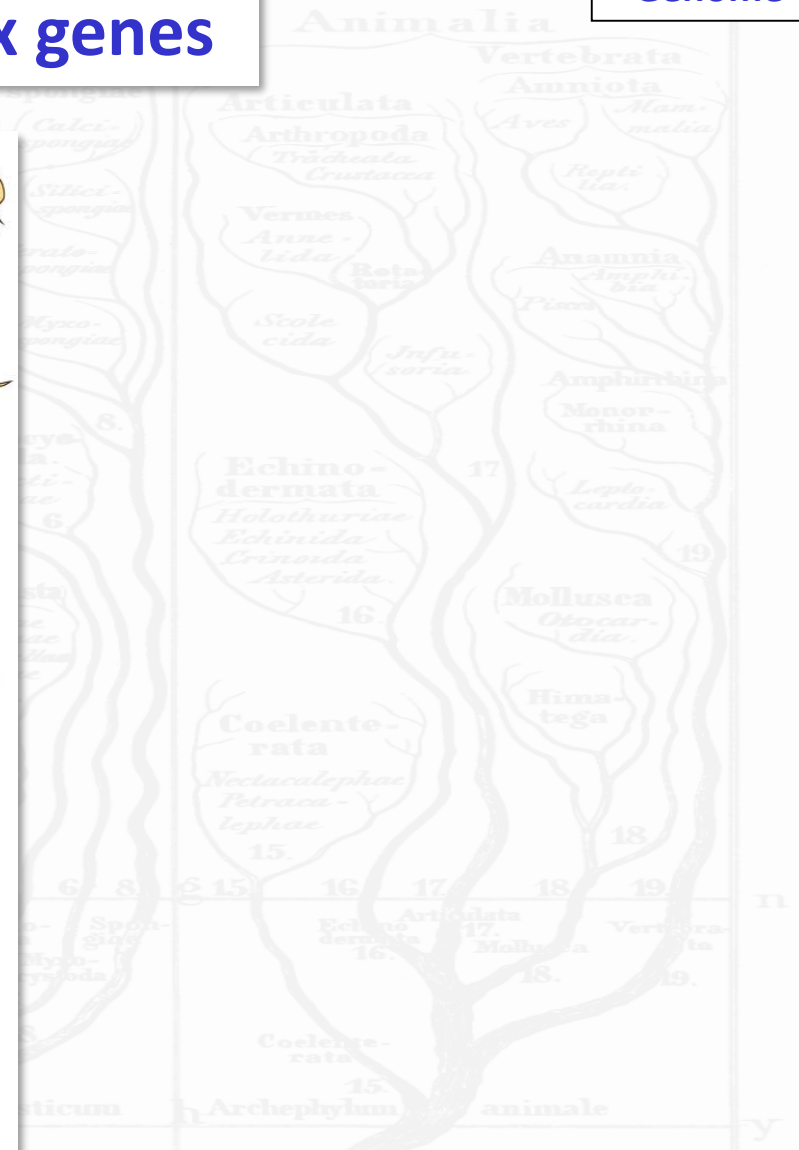
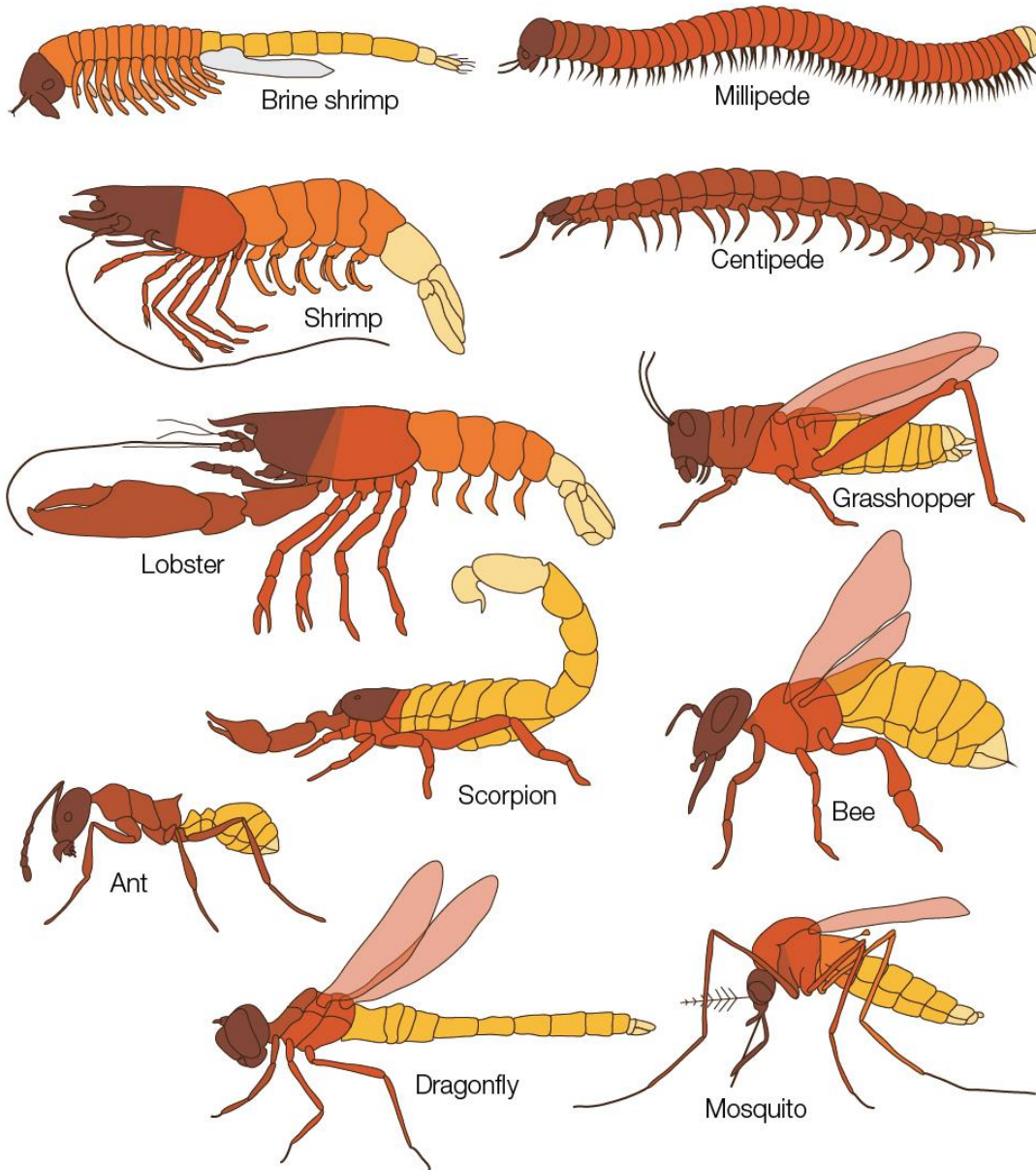




Transcription Factors (TFs): Hox genes



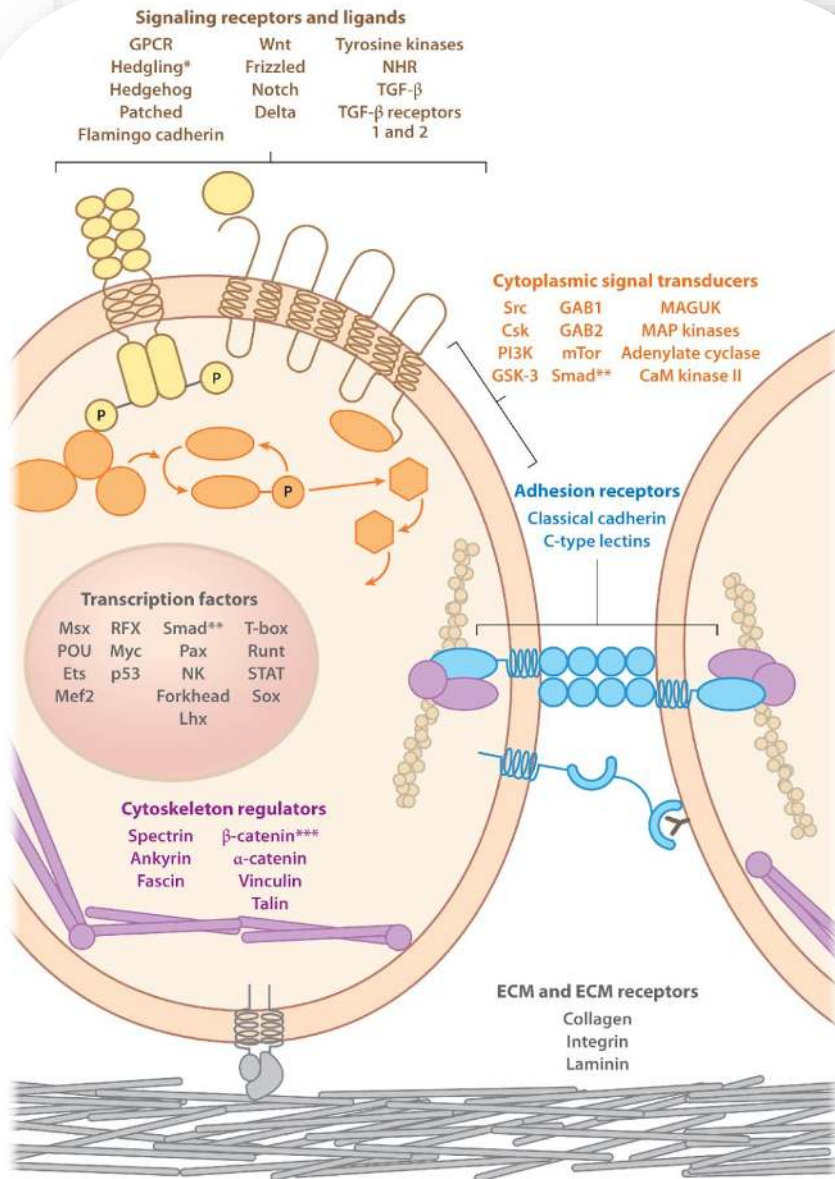
Transcription Factors (TFs): Hox genes



Learn genetics

<http://learn.genetics.utah.edu/content/basics/hoxgenes/>

Genome



• **Gene regulation** (TFs, and signalling pathways)

• **Cell adhesion**

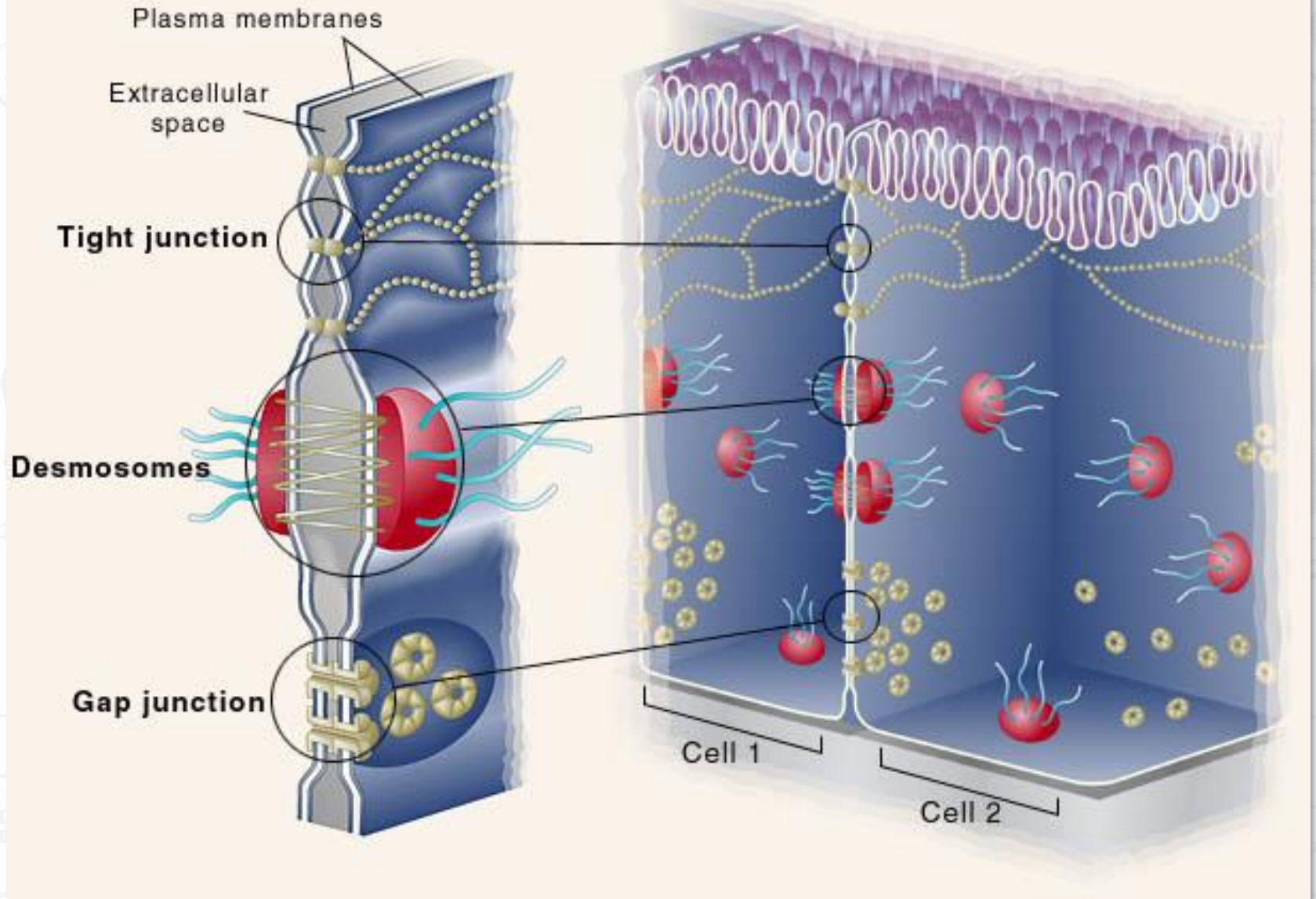
• **Cell type specialization**

• **Cell cycle**

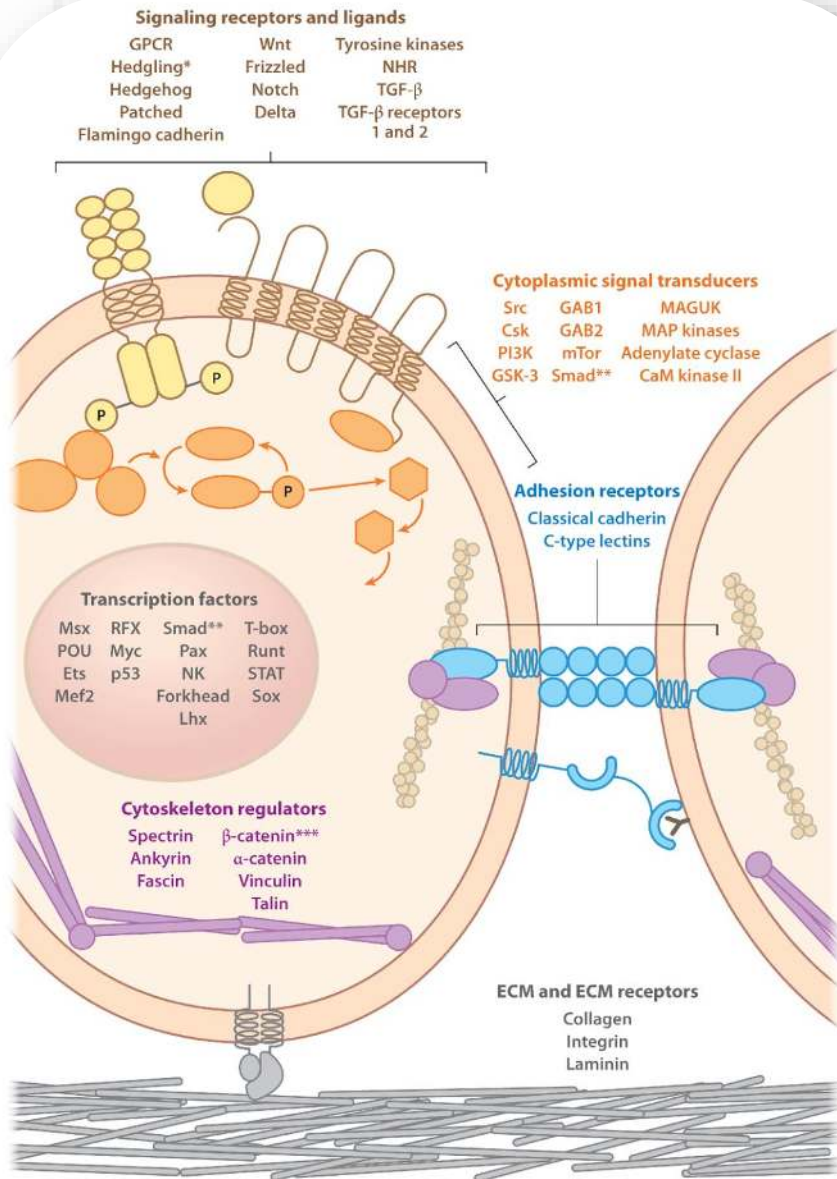
• **Immunity**

Cell adhesion

Specialized cell junctions

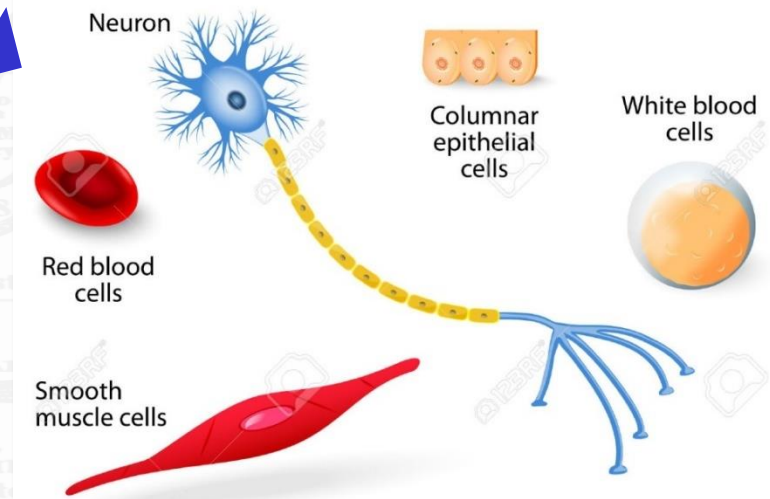
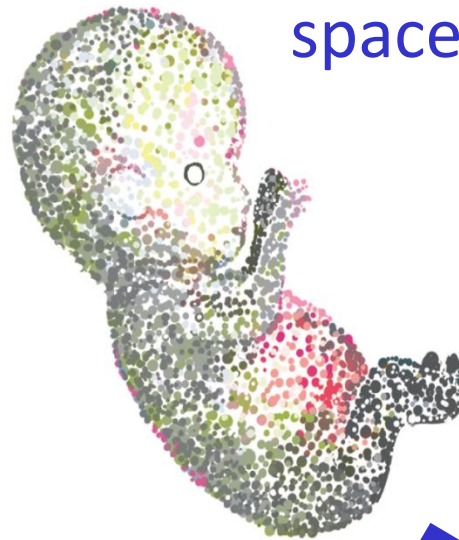
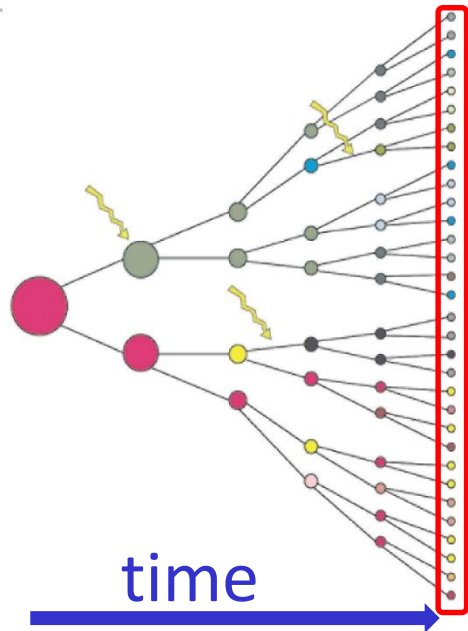


Genome

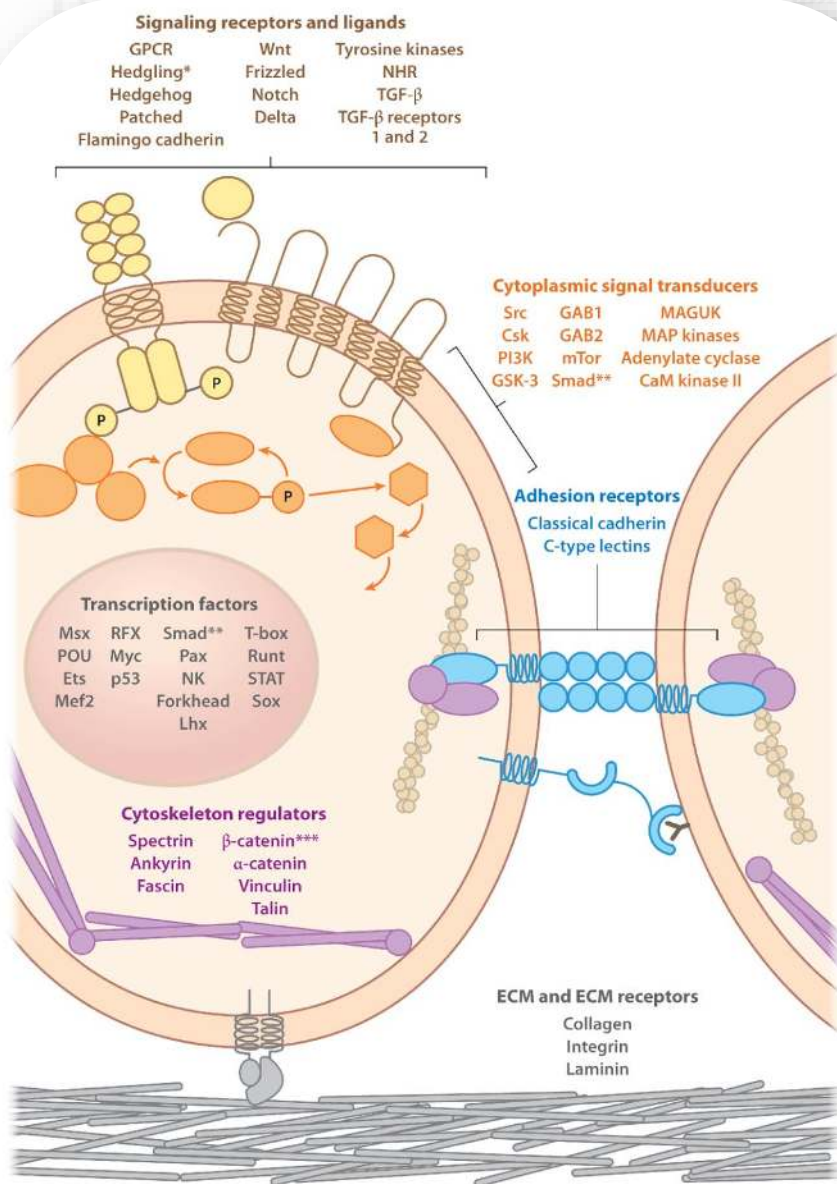


- **Gene regulation** (TFs, and signalling pathways)
- **Cell adhesion**
- **Cell type specialization**
- **Cell cycle**
- **Immunity**

Cell types



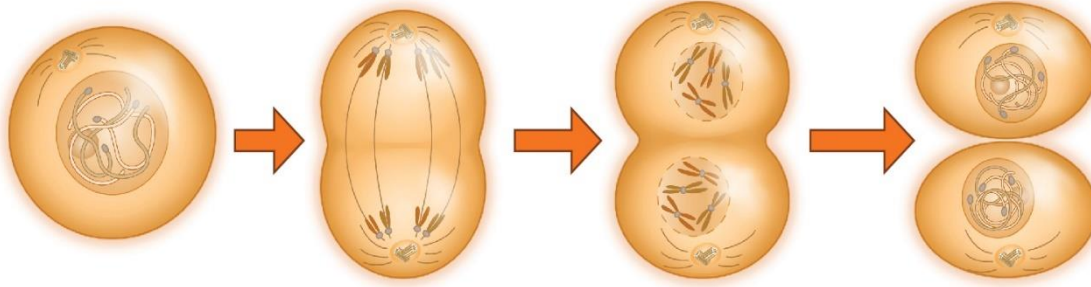
Genome



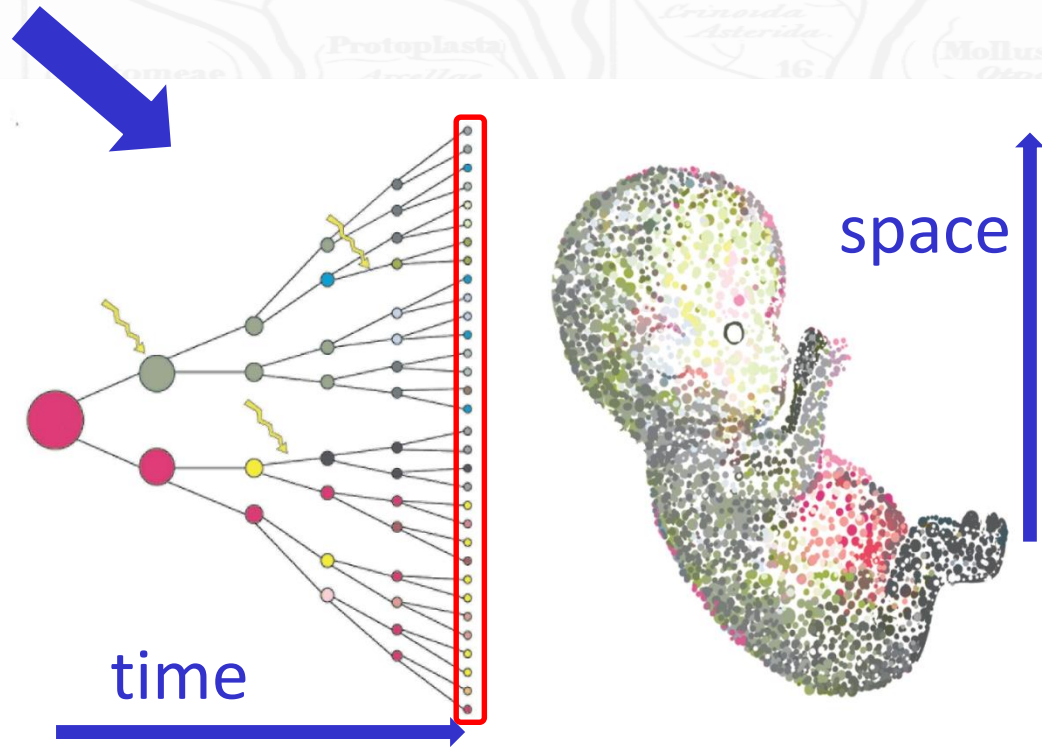
- **Gene regulation** (TFs, and signalling pathways)
- **Cell adhesion**
- **Cell type specialization**
- **Cell cycle**
- **Immunity**

Cell cycle

Cell division

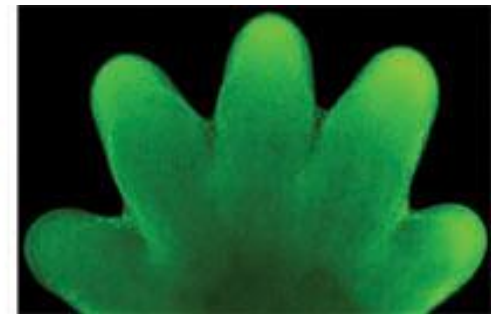
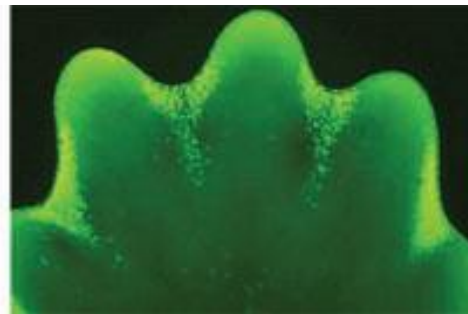
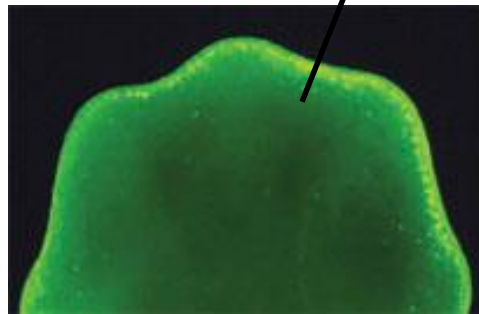


© 2013 Encyclopædia Britannica, Inc.



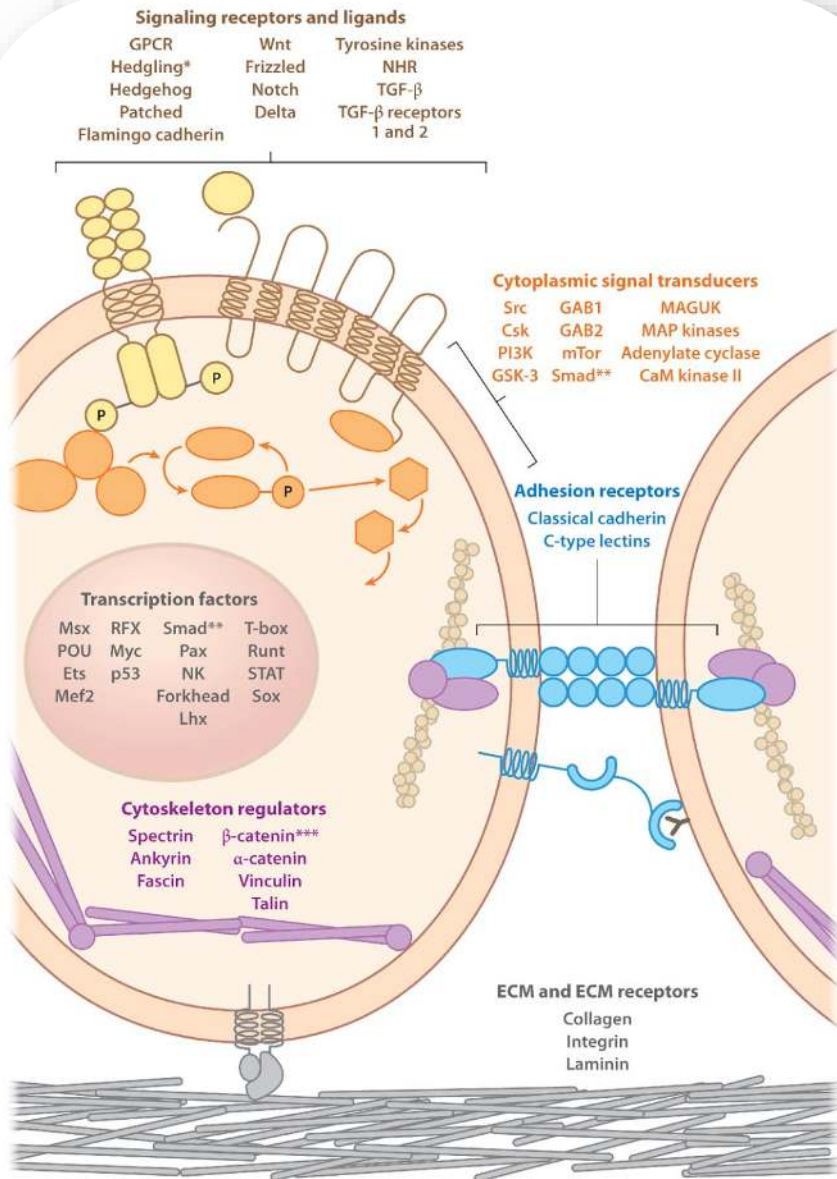
Programmed cell death: apoptosis

Interdigital tissue

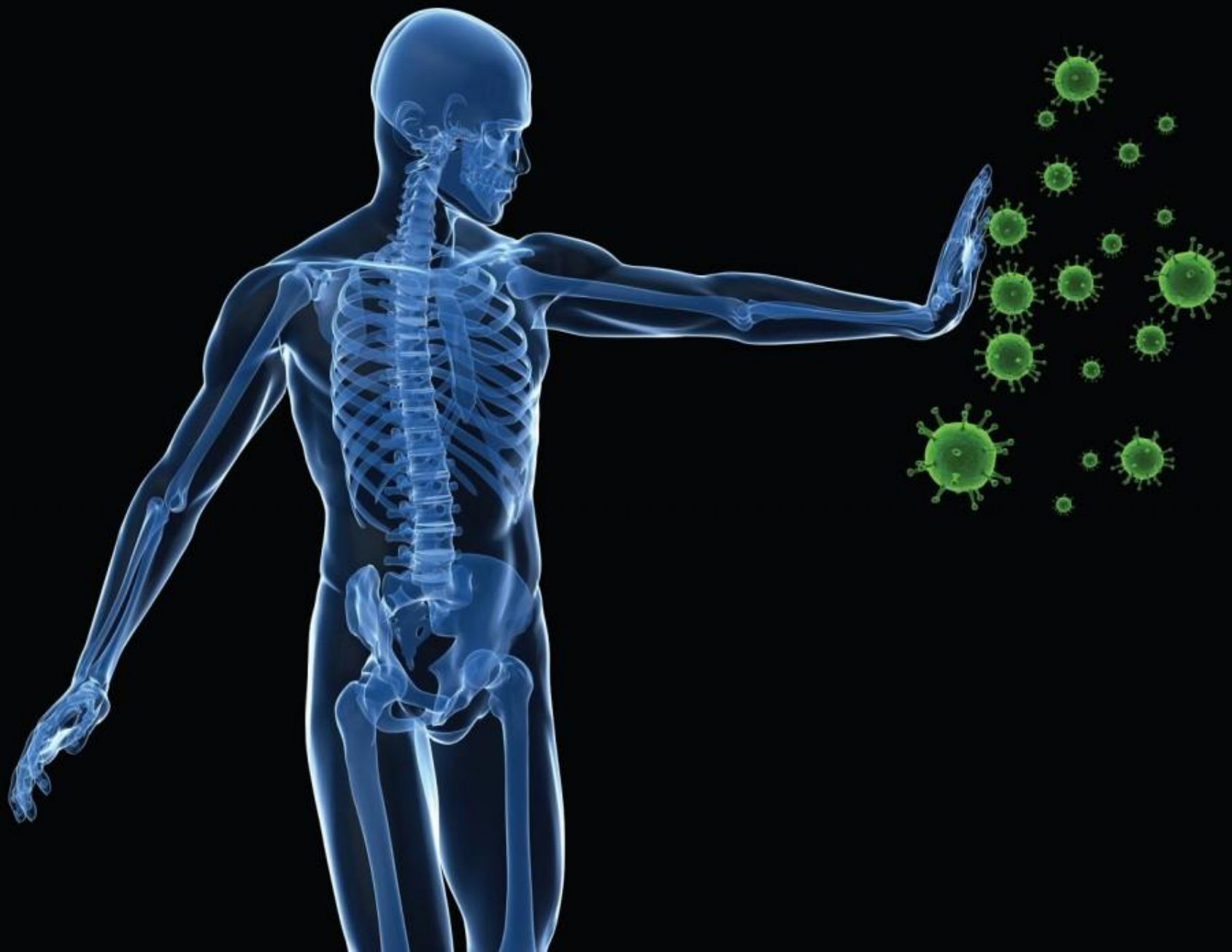


1 mm

Genome



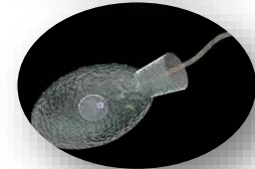
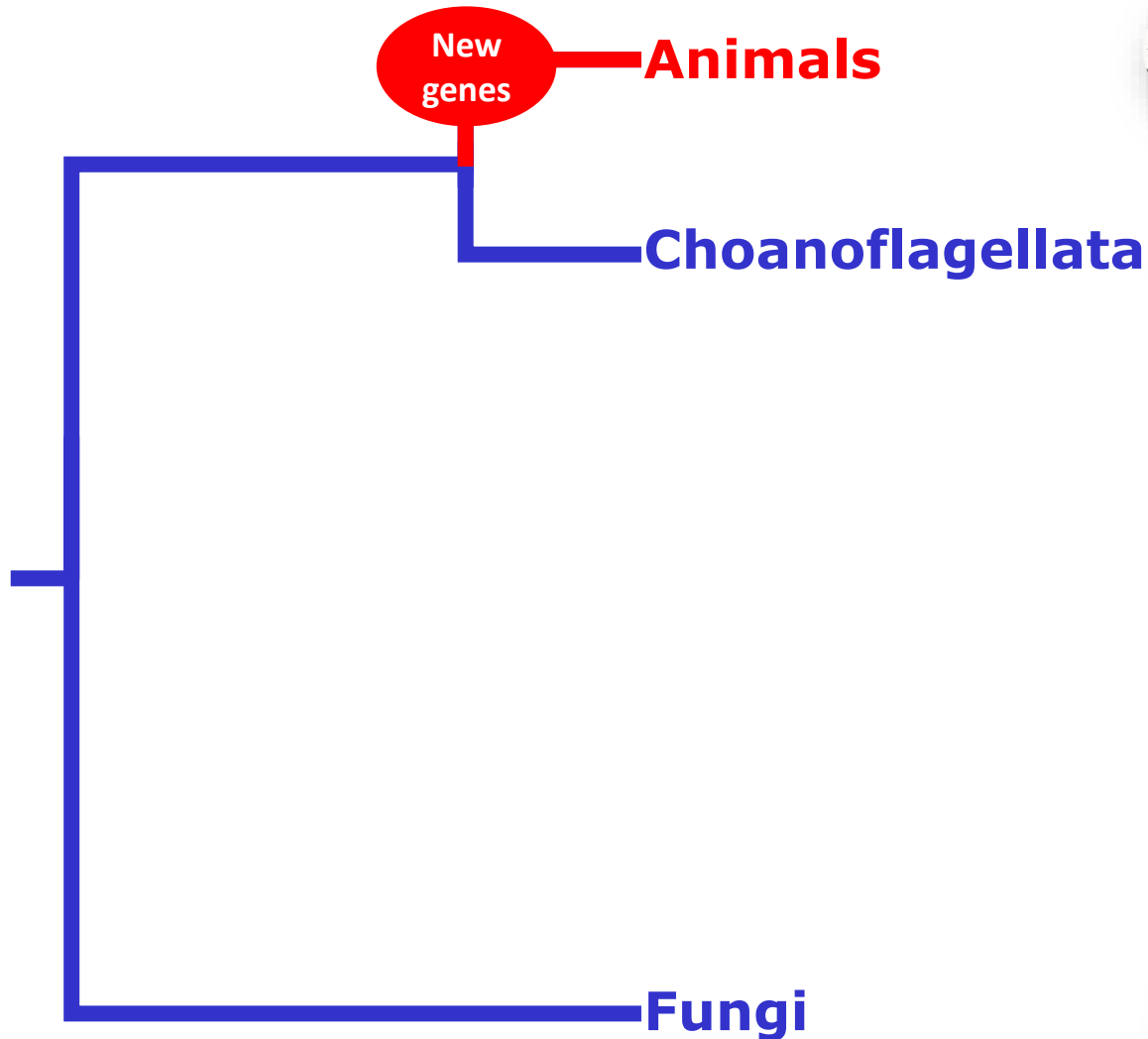
- **Gene regulation** (TFs, and signalling pathways)
- **Cell adhesion**
- **Cell type specialization**
- **Cell cycle**
- **Immunity**





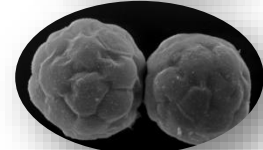
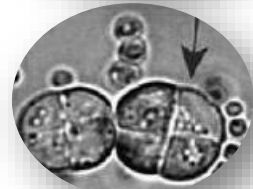
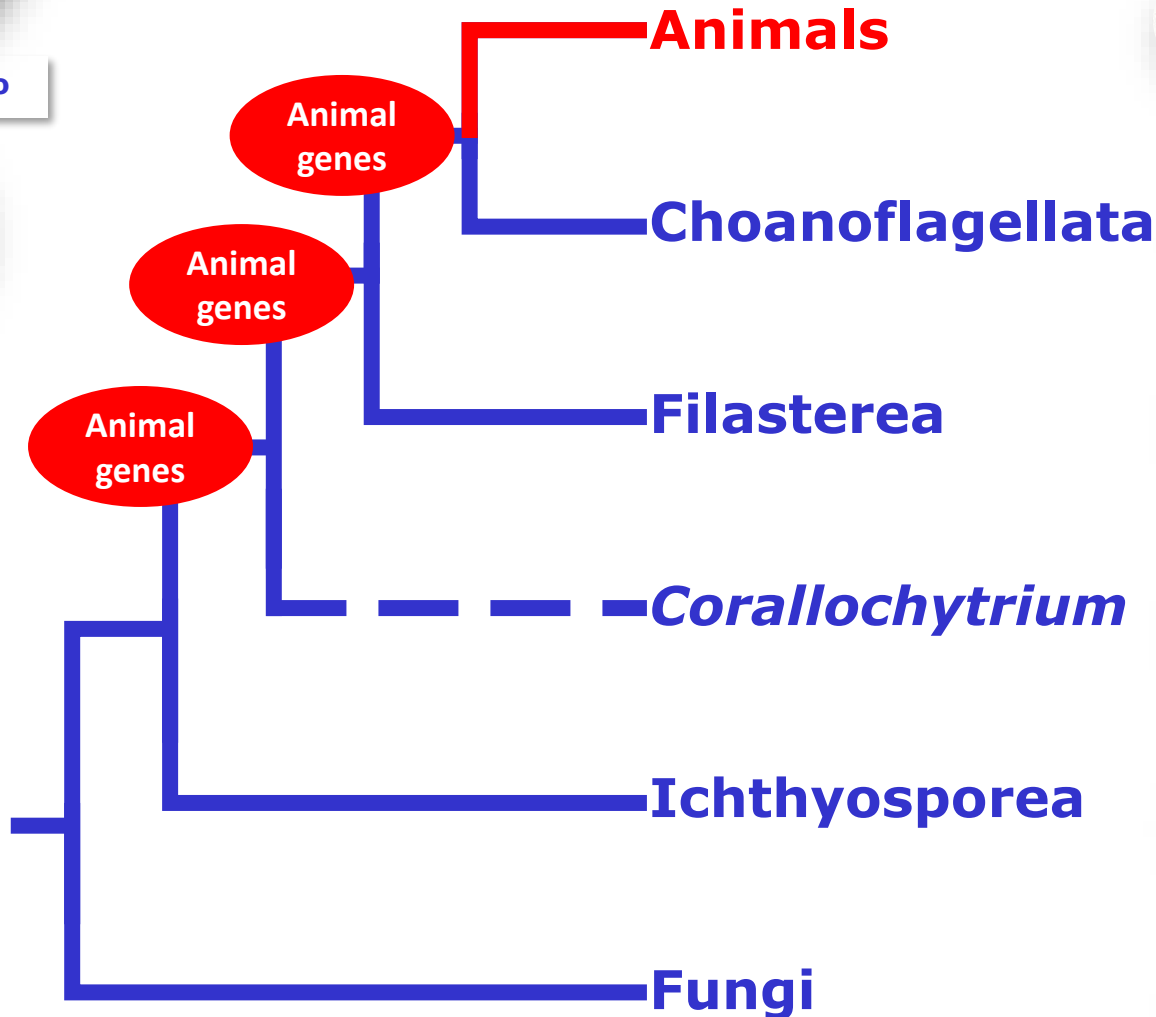
Genomic Novelties

Old vs New



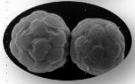
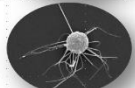
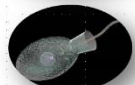
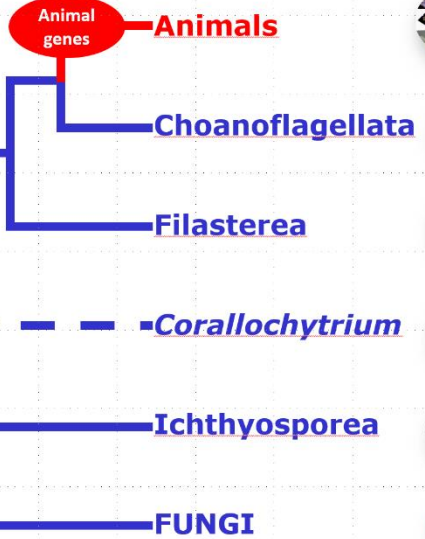
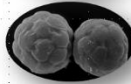
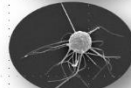
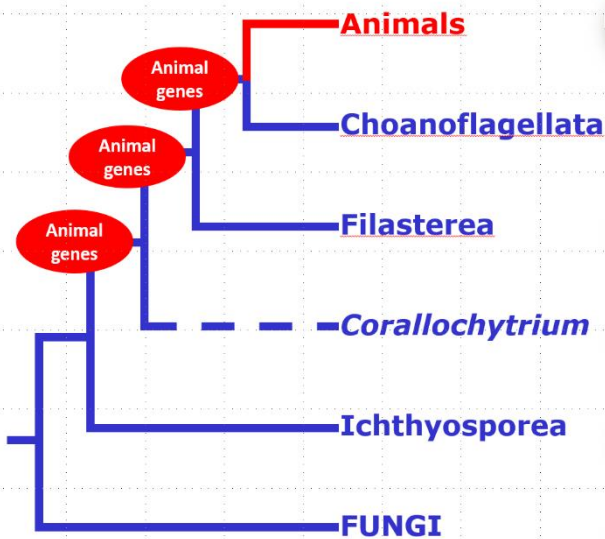


Iñaki Ruiz-Trillo



Old vs new

Old vs New



COMPARE GENOMES TO PROFILE THE ROLE OF NOVELTY IN ANIMAL ORIGINS



Peter WH Holland

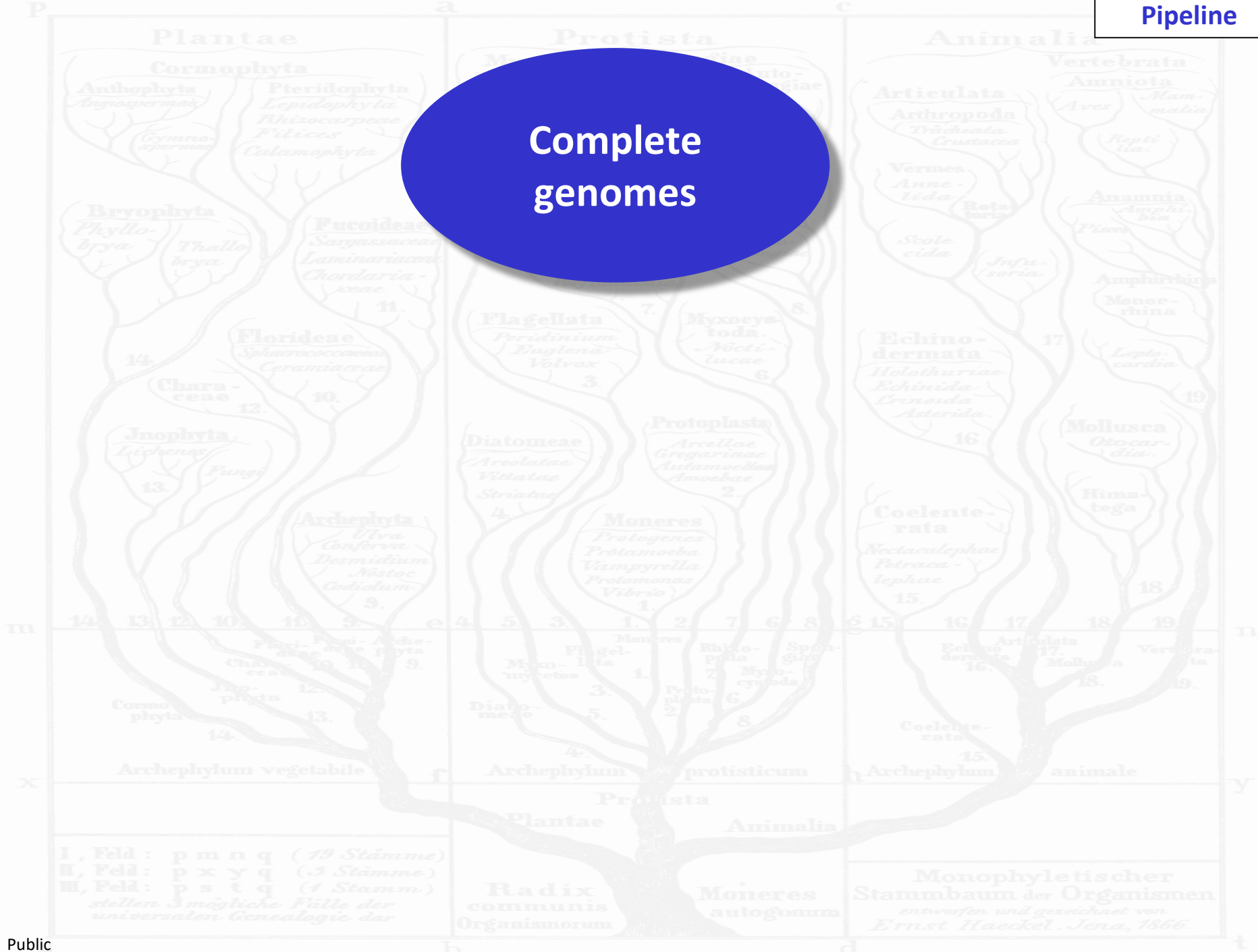
THE PIPELINE

Complete
genomes

Define groups
of homologous
genes

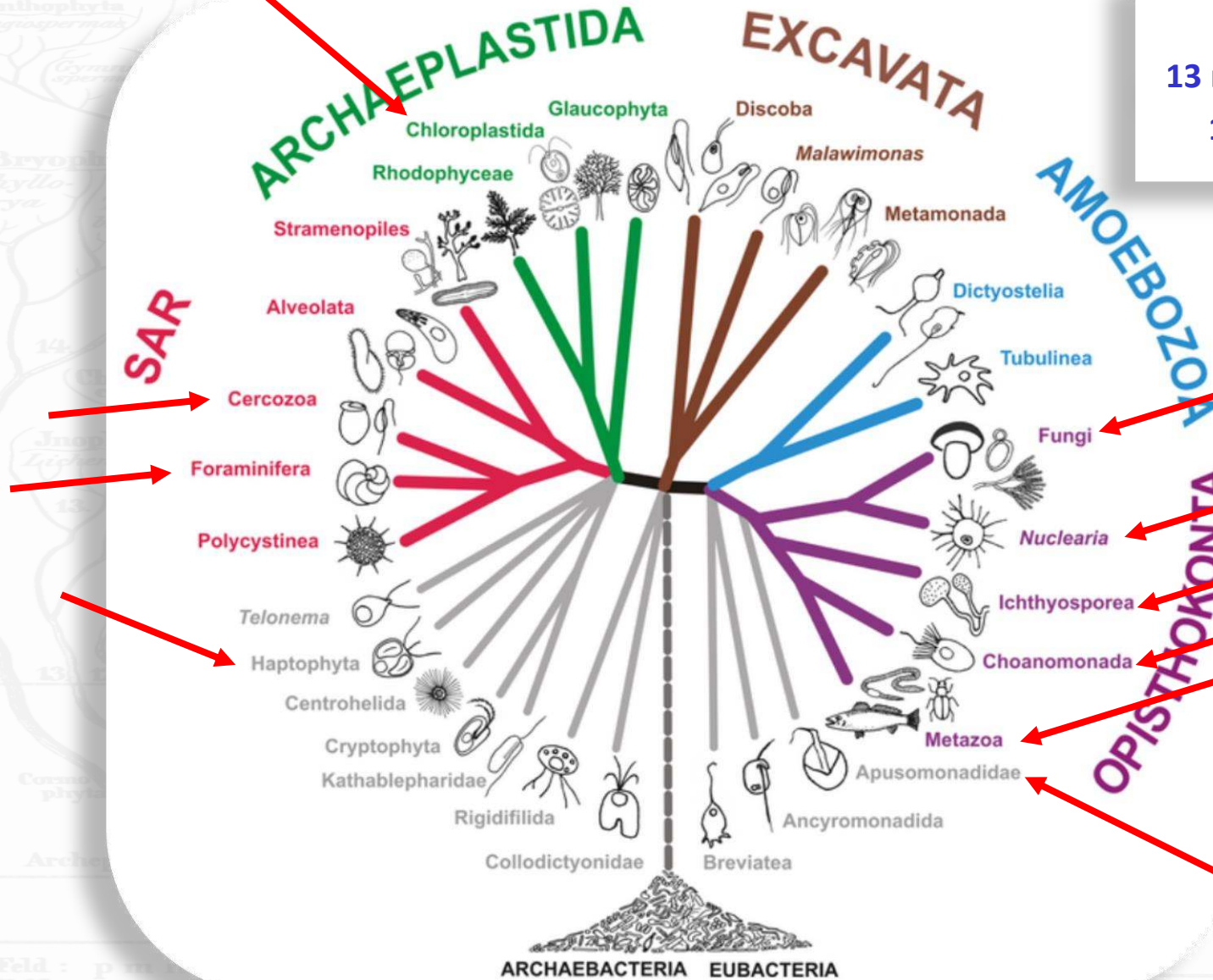
Map genes
distribution

Complete
genomes



Eukaryotes

62 Genomes
13 metazoan phyla
10 outgroups

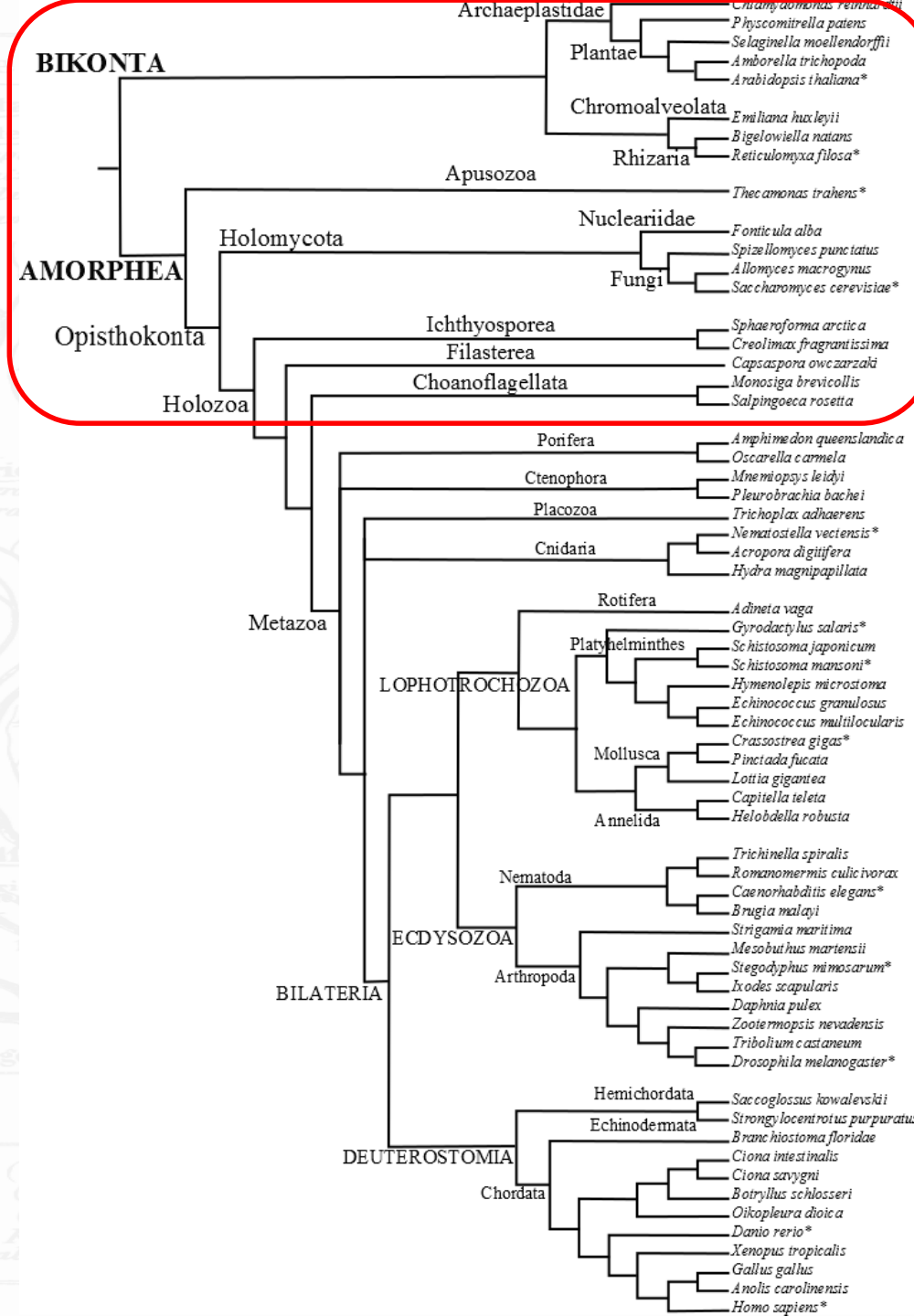


Animals Origins

62 Genomes

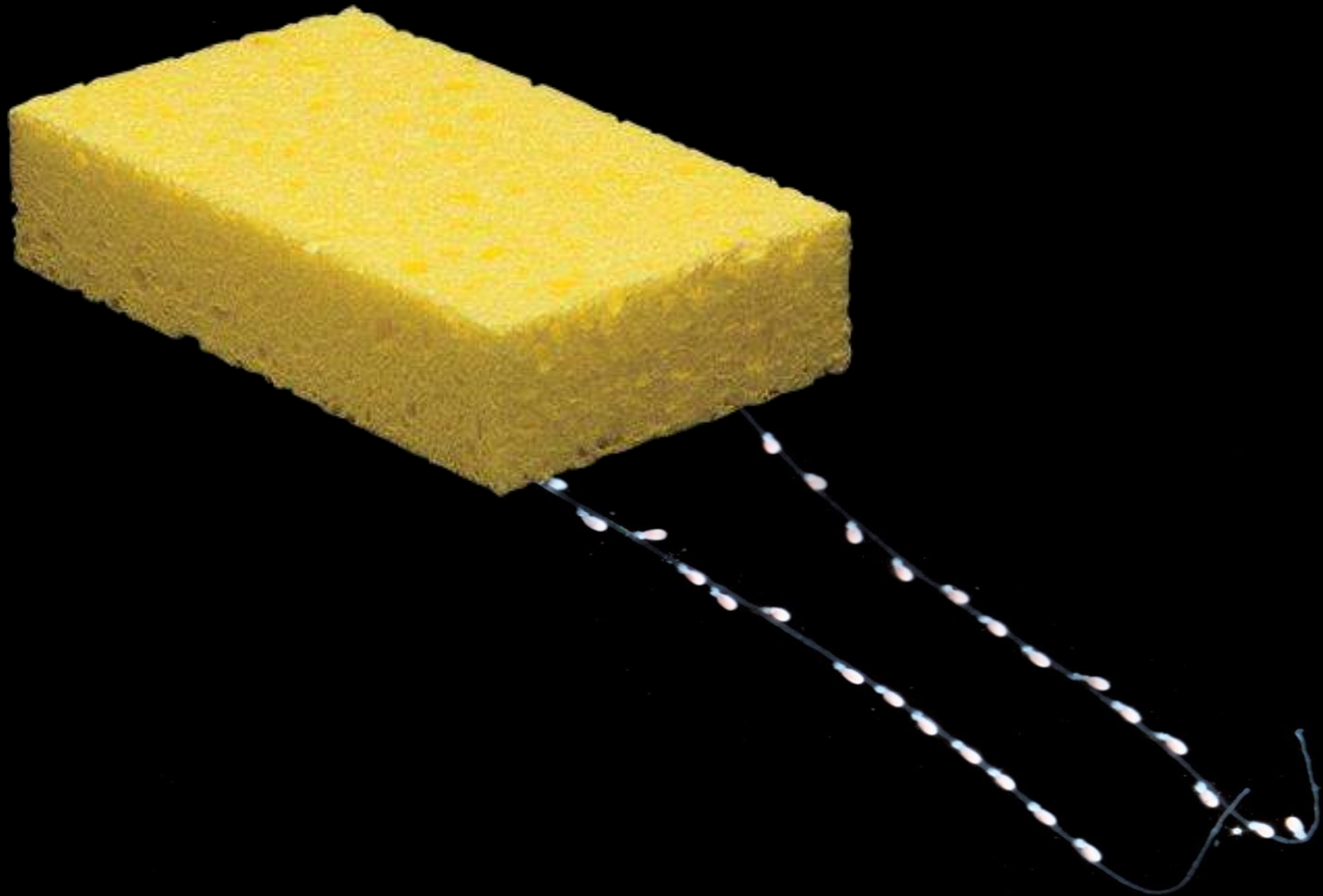
13 metazoan phyla

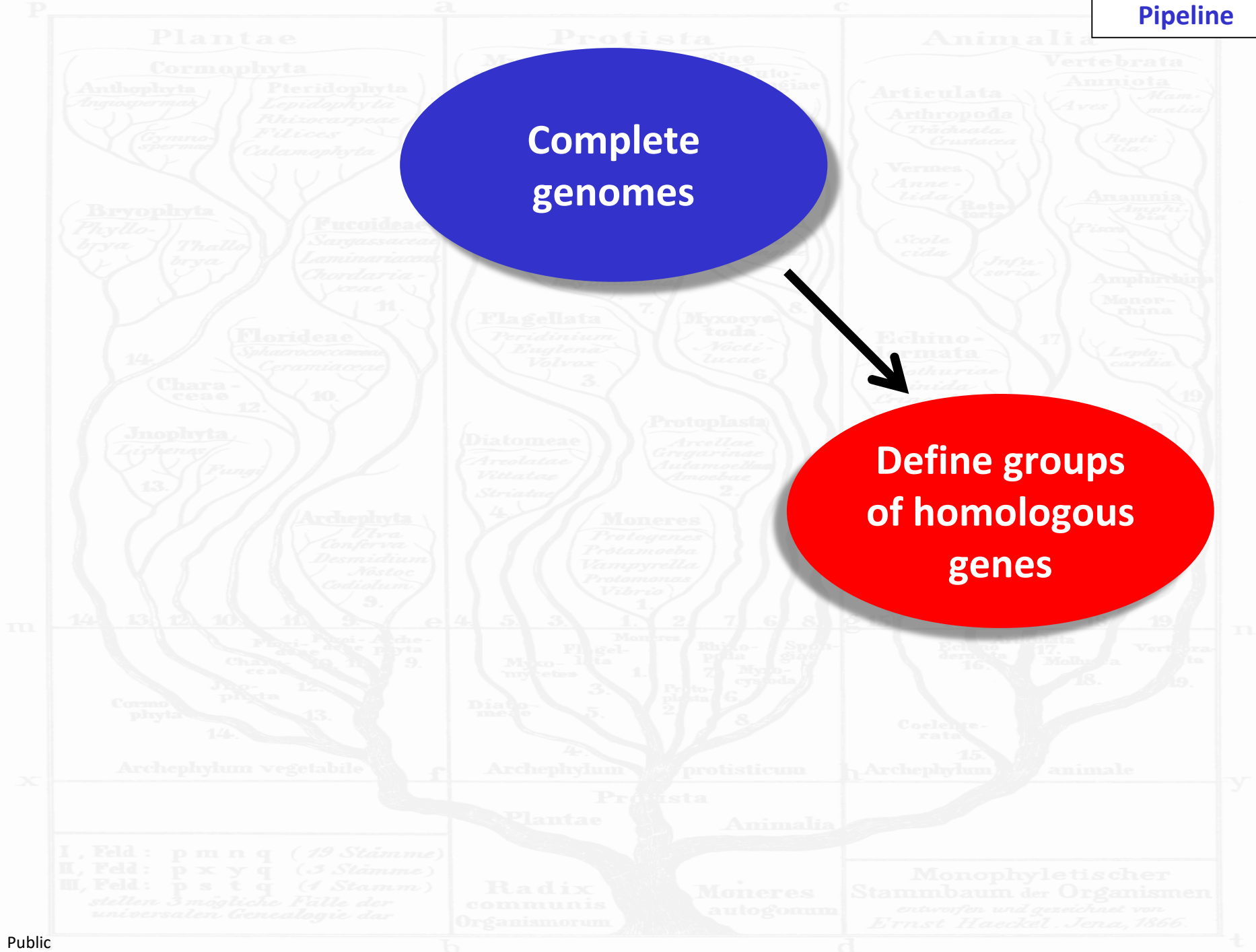
10 outgroup lineages



Animal root: spongctenophorifera

Pipeline





Homology assignment

	<u>Speed</u>	<u>Accuracy</u>	<u>Orthology/Paralogy</u>
<u>Phylostratigraphy</u> one spp query one-way BLAST	+++	+	None (assumes one-to-one)
<u>Reciprocal Best BLAST</u> All spp vs all Two-way BLAST RBB/RBD/RBH...	++	++	BLAST e-value (i.e. Inparanoid) Gene trees
<u>Reciprocal BLAST + HMM</u> All spp vs all Two-way BLAST MCL	+	+++	BLAST e-value (i.e. OrthoMCL) Gene trees

Markov Clustering

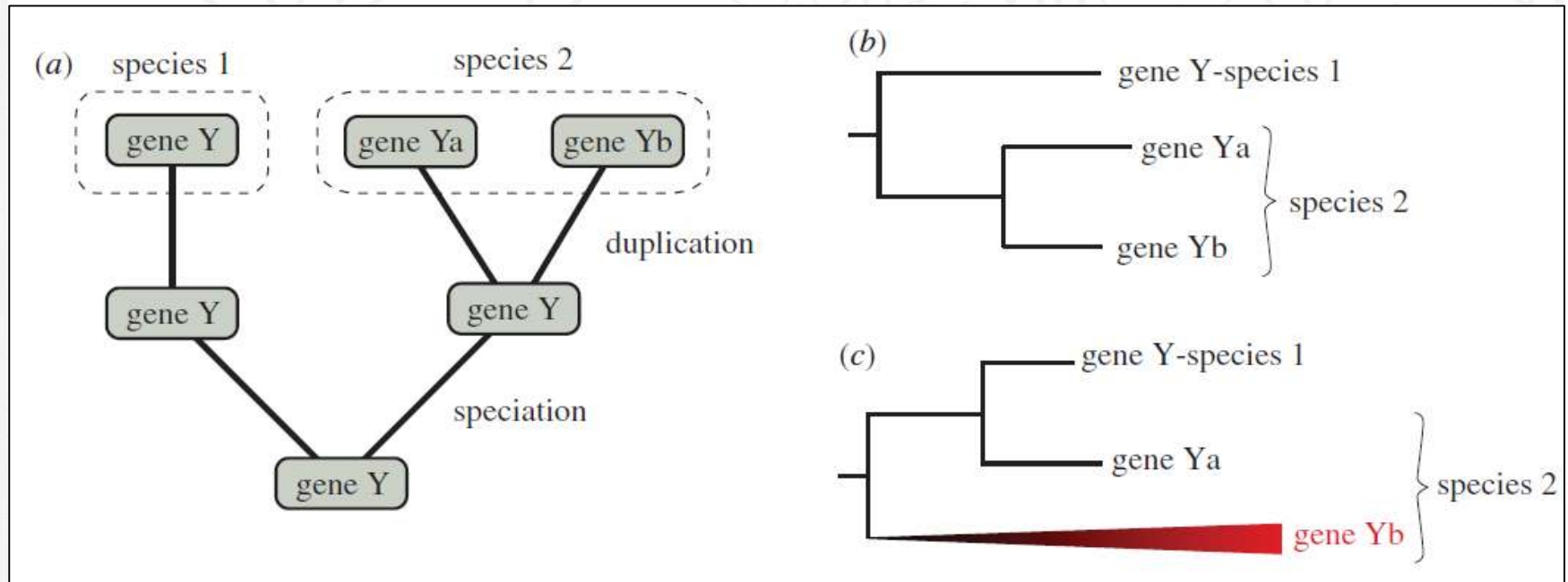
Pipeline

**Reciprocal
Best BLAST
+
MCL**

- 1,500,000 genes in the 62 genomes
- 2,000,000,000,000 one vs one comparisons
- 268,440 homology groups (HG)

Why no orthology/paralogy?

Assymmetric evolution



PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Review

Cite this article: Holland PWH, Marlétaz F, Maeso I, Dunwell TL, Paps J. 2017 New genes from old: asymmetric divergence of gene duplicates and the evolution of development. *Phil. Trans. R. Soc. B* 372: 20150480. <http://dx.doi.org/10.1098/rstb.2015.0480>



New genes from old: asymmetric divergence of gene duplicates and the evolution of development

Peter W. H. Holland¹, Ferdinand Marlétaz^{1,2}, Ignacio Maeso^{1,3}, Thomas L. Dunwell¹ and Jordi Paps^{1,4}

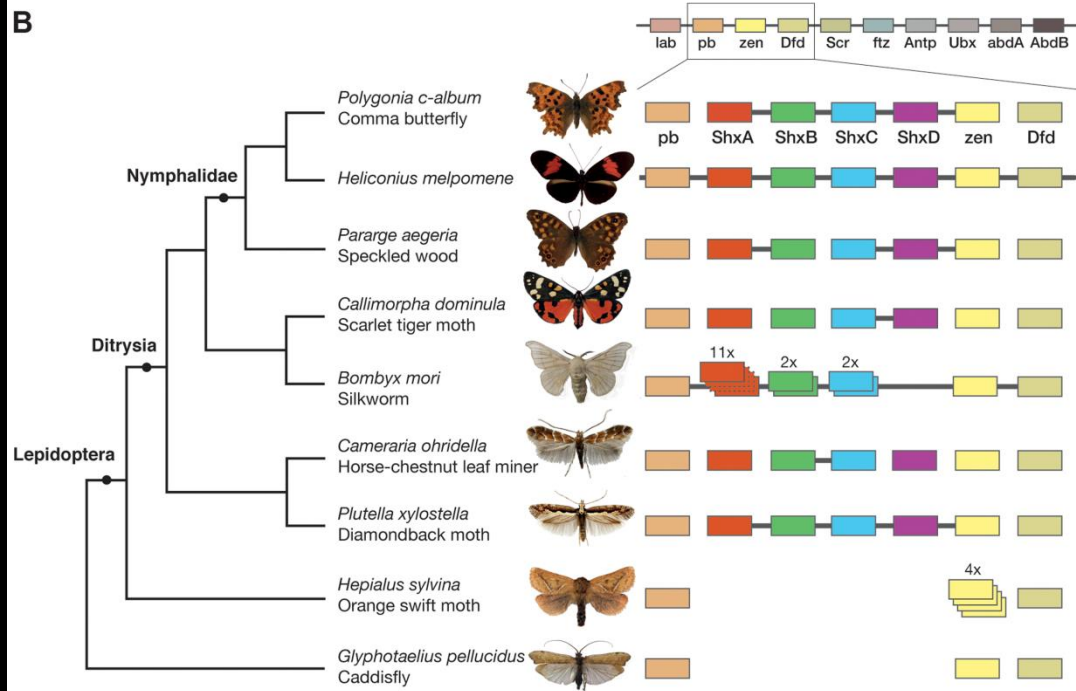
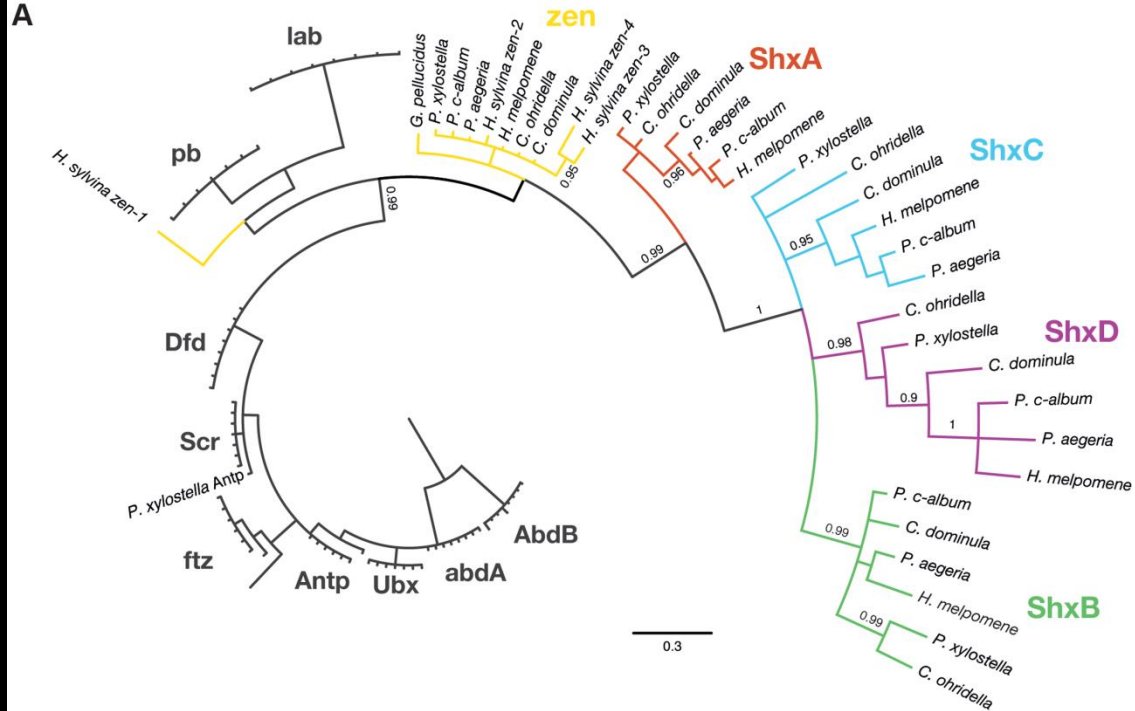
¹Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

²Molecular Genetics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan

³Centro Andaluz de Biología del Desarrollo, Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, 41013 Sevilla, Spain

⁴School of Biological Sciences, University of Essex, Colchester, Essex, UK

© PWHH, 0000-0003-1533-9376



Homeobox Superfamily

↓
Class ANTP

↓
Subclass HOXL

↓
Family *Hox1*

- All human genome: 9,415 homology groups
- All fruit fly genome: 7,681 homology clusters

Pipeline

Classes	Families
ANTP(767)	HOXL subclass Cdx(22) Evx(17) Gbx(15) Gsx(13) Hox1(21) Hox2(13) Hox3(25) Hox4(24) Hox5(19) Hox6-8(55) Hox9-13(15)(96) Meox(14) Mnx(15) Pdx(6)
	NKL subclass Abox(5) Ankx(1) Barhl(16) Bari(4) Barx(10) Bsx(10) Dbx(15) Dlx(35) Emx(22) En(19) Hhex(10) Hlx(9) Hx(1) Lbx(14) Lcx(1) Msx(22) Msxlx(3) Nanog(15) Nedx(5) Nk1(17) Nk2.1(18) Nk2.2(15) Nk3(14) Nk4(19) Nk5/Hmx(25) Nk6(19) Nk7(4) Noto(10) Ro(5) Tlx(19) Vax(12) Ventx(18)
PRD(458)	Alx(16) AprdA(1) AprdB(1) AprdC(1) AprdD(2) AprdE(1) Argfx(3) Arx(14) CG11294(3) Dmbx(8) Dprx(8) Drgx(9) Dux(42) Esx(2) Gsc(14) Hbn(3) Hesx(5) Hopx(5) Isx(5) Leutx(1) Mix(7) Nobox(5) Obox(36) Otp(12) Otx(24) Pax2/5/8(12) Pax3/7(24) Pax4/6(24) Phox(17) Pitx(20) Prop(8) Prrx(14) Rax(14) Repo(4) Rhox(39) Sebox(3) Shox(12) Tprx(5) Uncx(16) Vsx(18)
LIM(104)	Isl(14) Lhx1/5(17) Lhx2/9(19) Lhx3/4(15) Lhx6/8(18) Lmx(21)
POU(112)	Hdx(6) Pou1(6) Pou2(24) Pou3(29) Pou4(21) Pou5(12) Pou6(15)
HNF(23)	Ahnfx(1) Hmbox(10) Hnf1(12)
SINE(48)	Six1/2(18) Six3/6(17) Six4/5(13)
TALE(147)	Atale(1) Irx(44) Meis(23) Mlxx(8) Pbx(29) Pknox(14) Tgif(28)
CUT(58)	Acut(1) Cmp(5) Cux(17) Onecut(24) Satb(11)
PROS(16)	Prox(16)
ZF(83)	Adnp(9) Azfh(1) Tshz(16) Zeb(14) Zfhx(21) Zhx/Homez(22)
CERS(26)	Cers(26)

Zhong and Holland 2011

Complete
genomes

Define groups
of homologous
genes

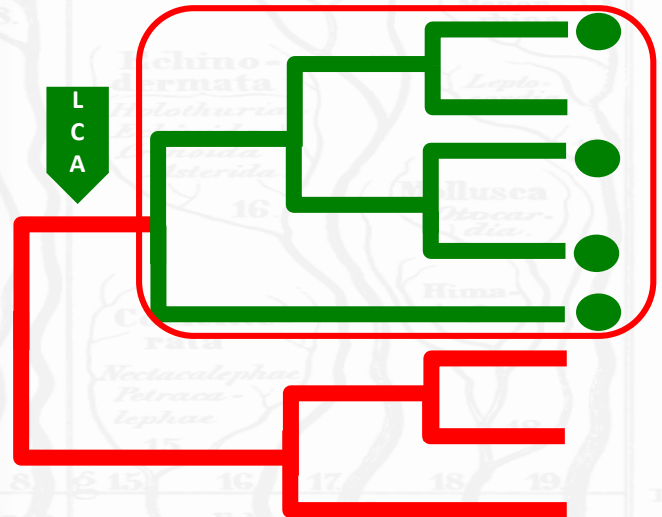
Map genes
distribution

Phylogenetically-Aware Parsing Script

jordi@zoo-paralog: ~/ACTG/0_Analyses/03_Analyses

Tree:

```
Eukaryota
Amorphea
Opisthokonta
Holozoa
Metazoa
  Diploblastica_(paraphyletic)
  Cnidaria
    Anthozoa
      Acropora_digitifera_(Adig) => 24
      Nematostella_vectensis_(Nvec) => 23
    Medusozoa
      Hydra_magnipapillata_(Hmag) => 25
  Placozoa
    Trichoplax_adhaerens_(Tadh) => 22
  Ctenophora
    Tentaculata_Lobata
      Mnemiopsis_leidy_(Mley) => 20
    Tentaculata_Cydippida
      Pleurobrachia_bachei_(Pbac) => 21
  Porifera
    Homoscleromorpha
      Oscarella_carmela_(Ocar) => 19
    Demospongiae
      Amphimedon_queenslandica_(Aque) => 18
Bilateria
  Deuterostomia
    Hemichordata
      Enteropneusta
        Saccolossus_kowaleski_(Skow) => 50
    Chordata
      Cephalochordata
        Branchiostoma_floridae_(Bflo) => 52
      Urochordata
        Botryllus_schlosseri_(Bsch) => 55
        Ciona_intestinalis_(Cint) => 53
        Ciona_savignyi_(Csav) => 54
        Oikopleura_dioica_(Odio) => 56
      Vertebrata
        Anolis_carolinensis_(Acar) => 60
        Gallus_gallus_(Ggal) => 59
        Danio_rerio_(Drer) => 57
        Homo_sapiens_(Hsap) => 61
        Xenopus_tropicales_(Xtro) => 58
    Echinodermata
      Echinoidea
        Strongylocentrotus_purpuratus_(Spur) => 51
  Ecdysozoa
    Nematoda
      Chromadorea
        Caenorhabditis_elegans_(Cele) => 40
        Brugia_malayi_(Bmal) => 41
      Enoplea
        Romanomermis_culicivox_(Rcul) => 39
        Trichinella_spiralis_(Tspi) => 38
```



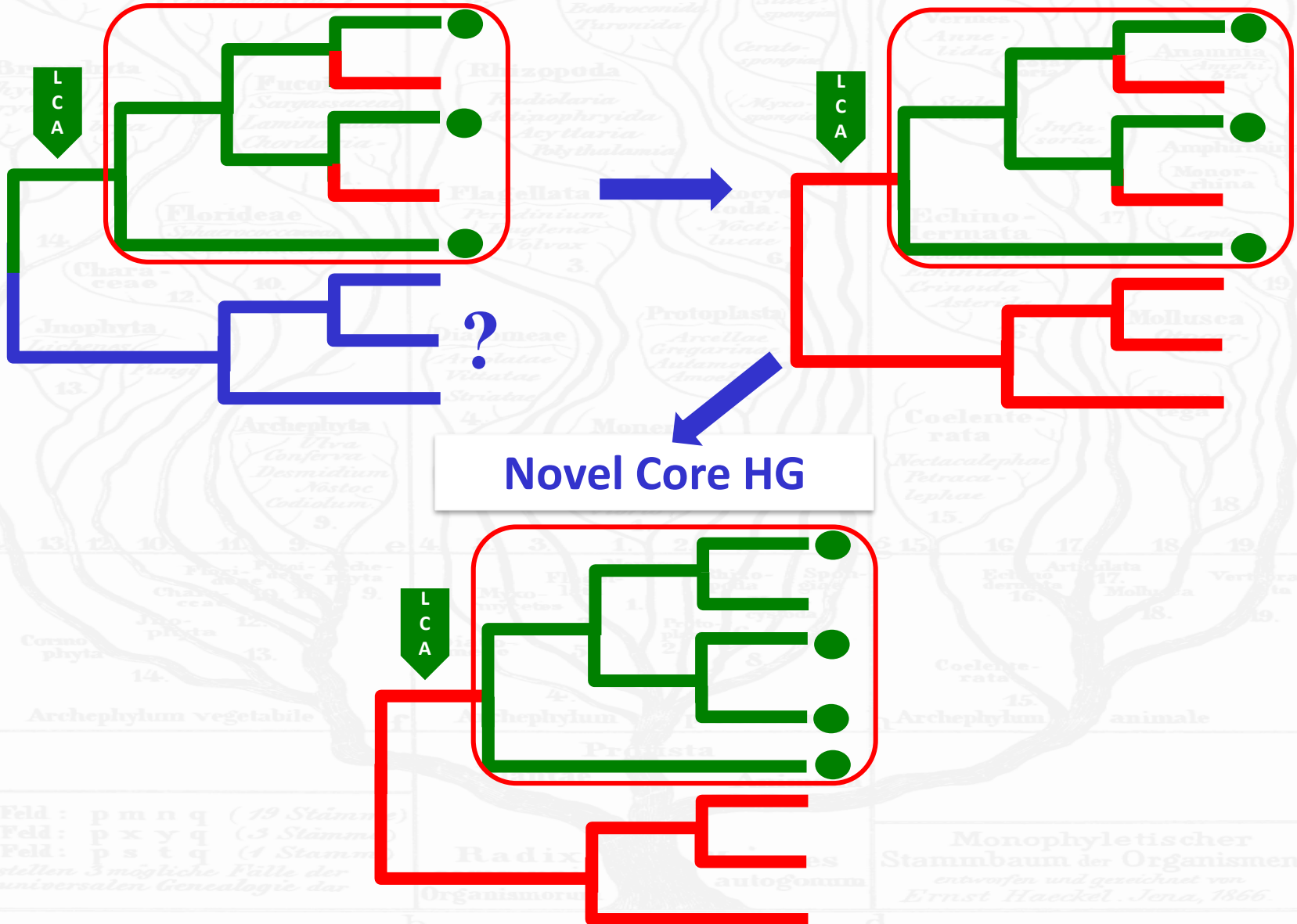
Monophyletischer
Stammbaum der Organismen
entworfen und gezeichnet von
Ernst Haeckel, Jena, 1866.

Types of gene groups

Ancestral HG

Novel HG

Novel Core HG



Limitations

- Based on **protein-coding** genes, neglecting the role of **non-coding genes** (e.g., non-coding RNAs), **transposable elements**, and **regulatory elements** (e.g., enhancer, promoters, etc.).
- Based on **BLAST**, ignoring **gene fusions** and **fissions**, **alternative splicing**, etc. Limited to the **detection power** of the algorithm.
- Robust taxon sampling**, but still far from complete.



THE GENOME OF THE FIRST ANIMAL

Clade
Ancestral

Holozoa

Metazoa

Metazoa

6331

Deuterostomia (12 species)



Ecdysozoa (12)



Lophotrochozoa (12)



Cnidaria (3)



Placozoa (1)



Ctenophora (2)



Porifera (2)



Choanoflagellata (2)



Filasterea (1)



Ichthyosporea (2)



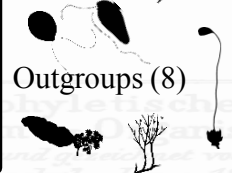
Holomycota (4)



Apusozoa (1)



Outgroups (8)





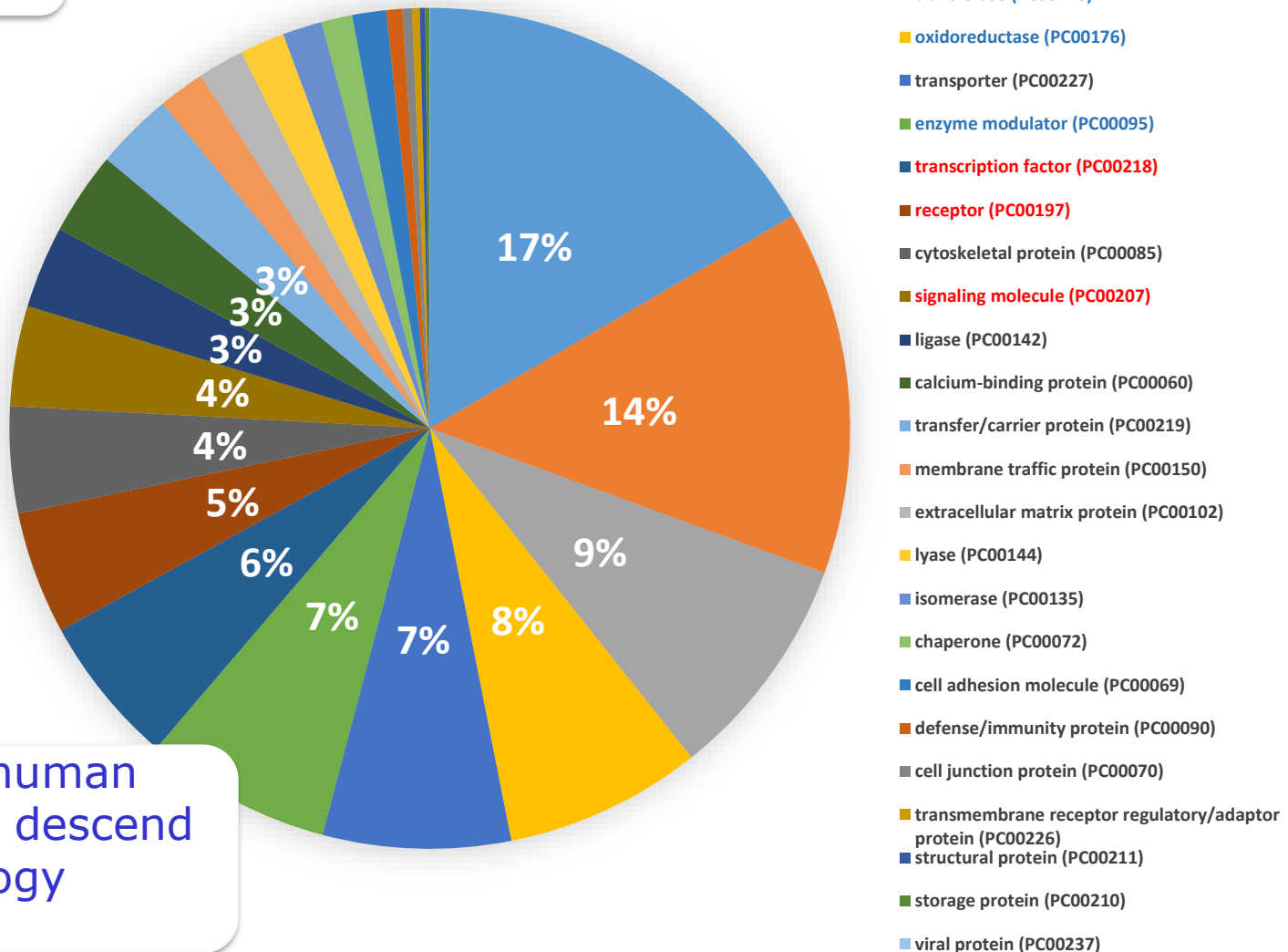
6331 homology groups in the first animal

Animal results

•Gene regulation

•Metabolism

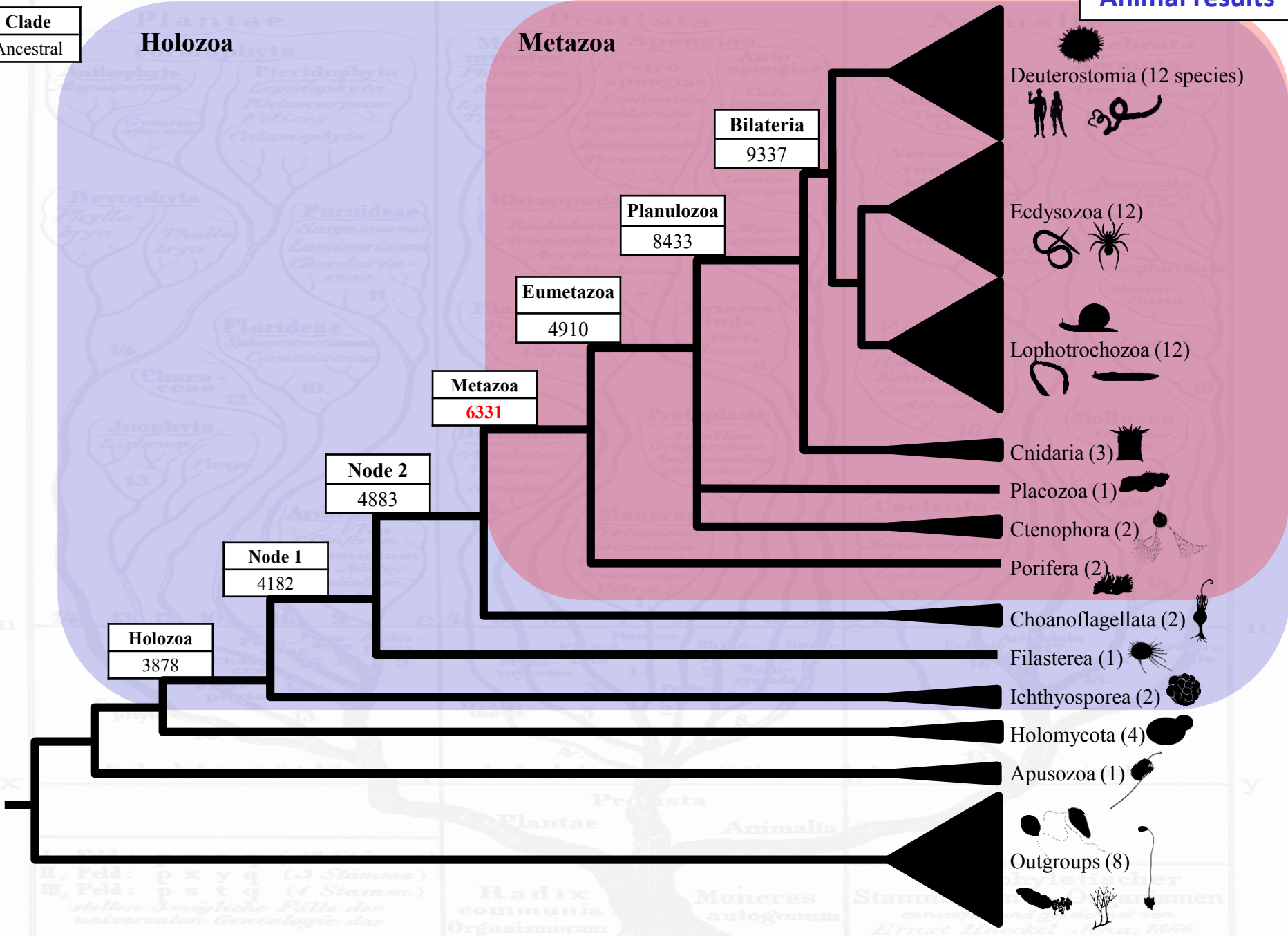
•60% of modern human and fruit fly genes descend from these homology groups.



Clade
Ancestral

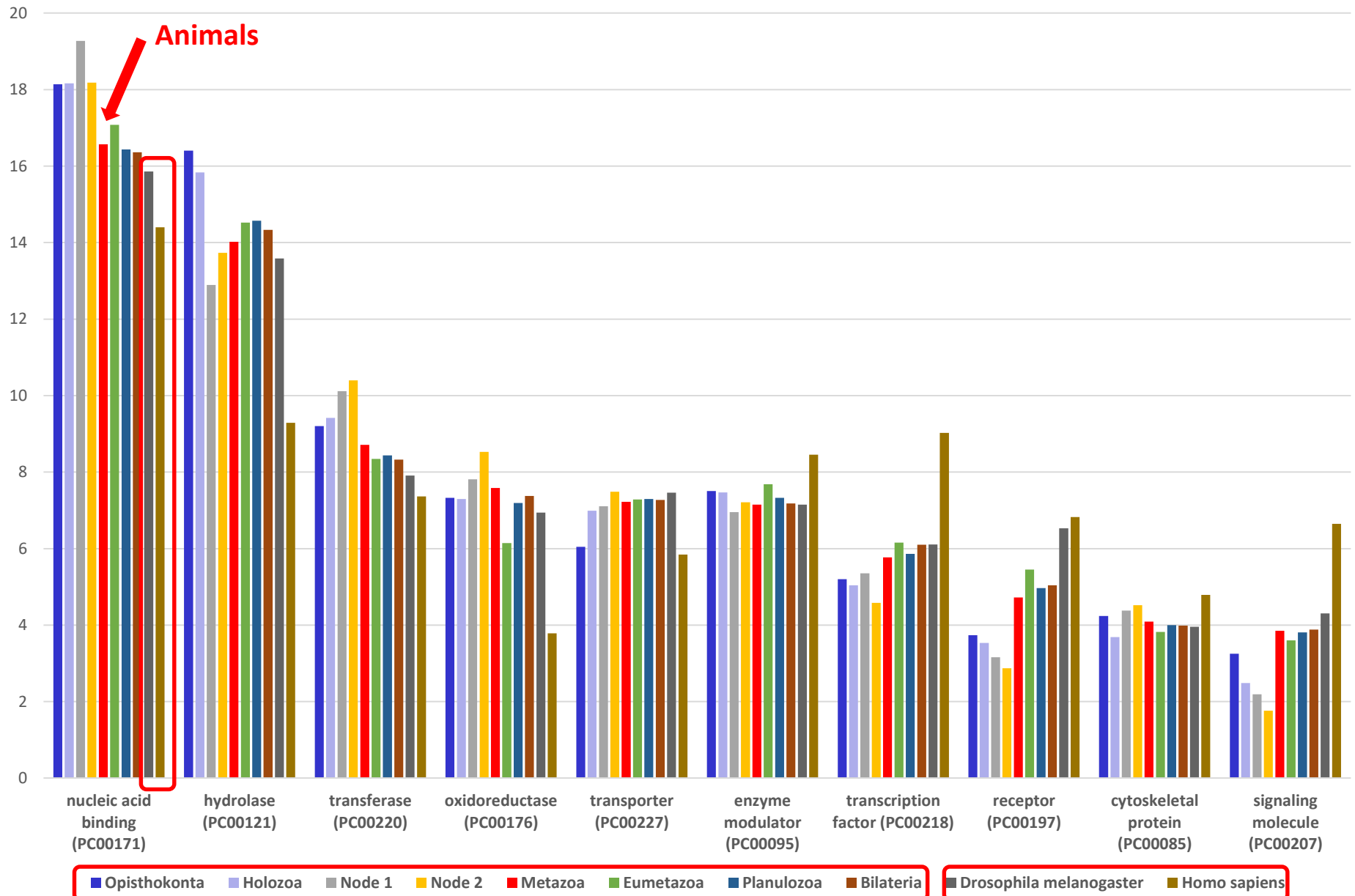
Holozoa

Metazoa

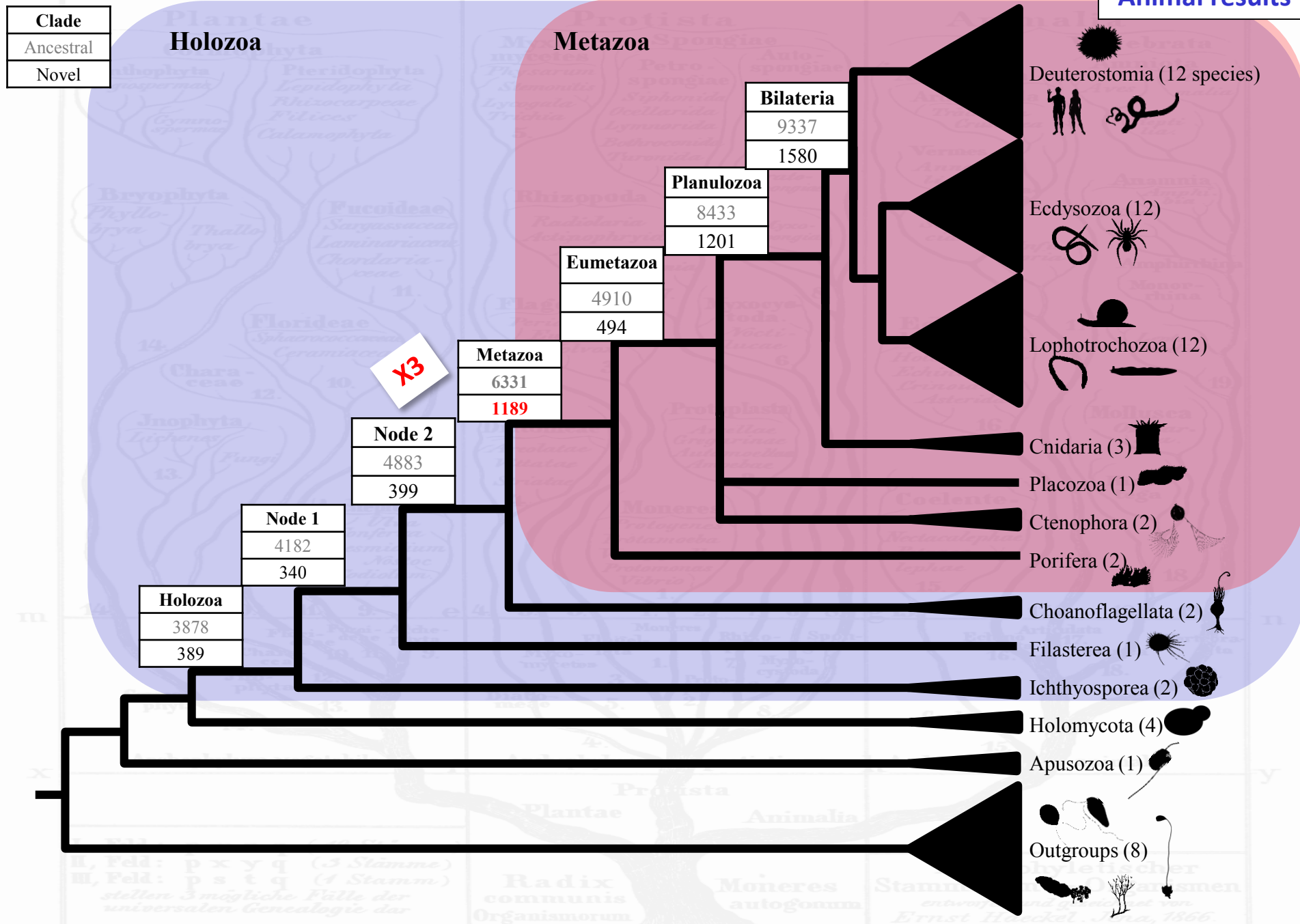


Evolution of Ancestral HG

Animal results

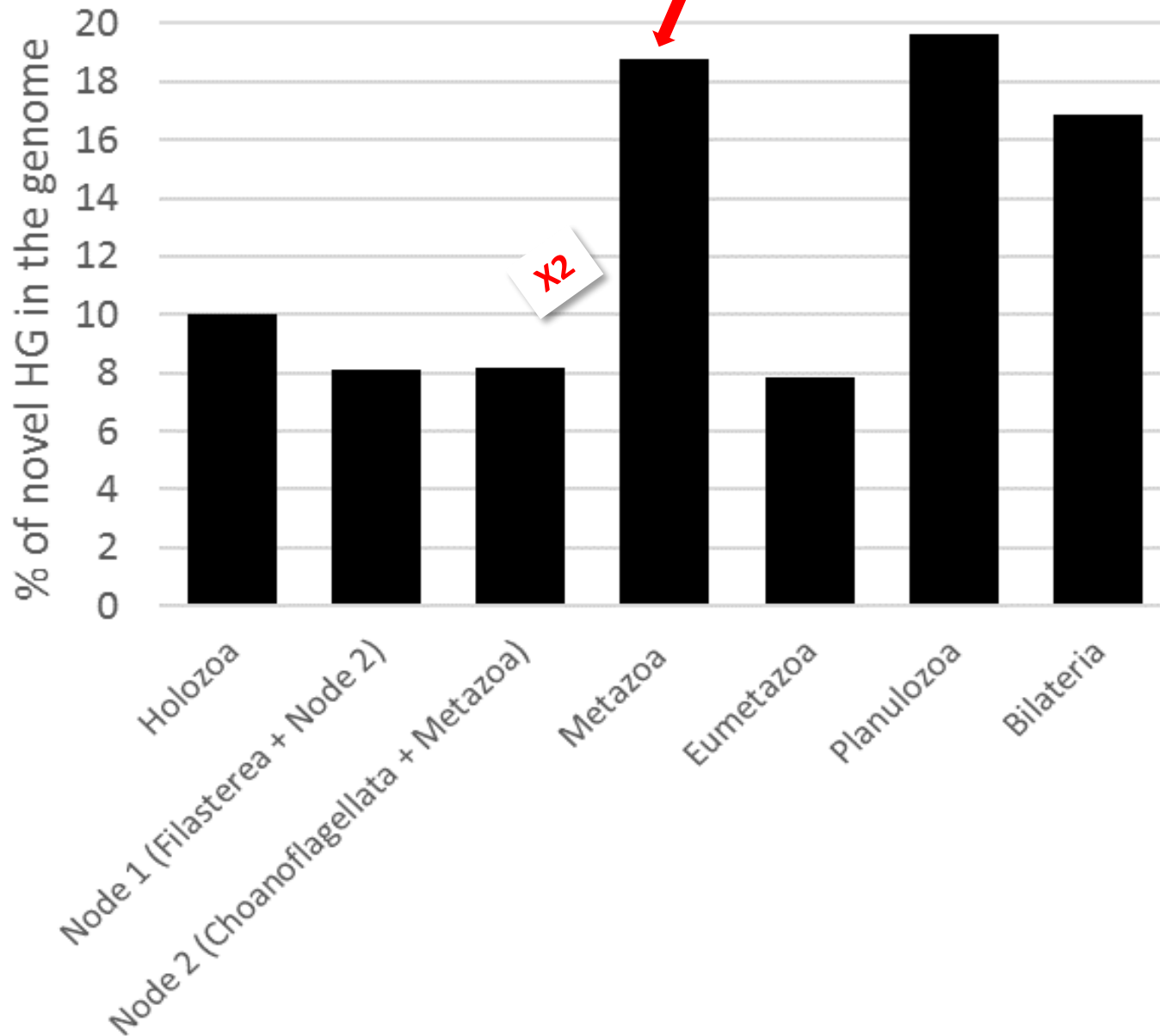


A GENOMIC EXPLOSION OF NOVELTY

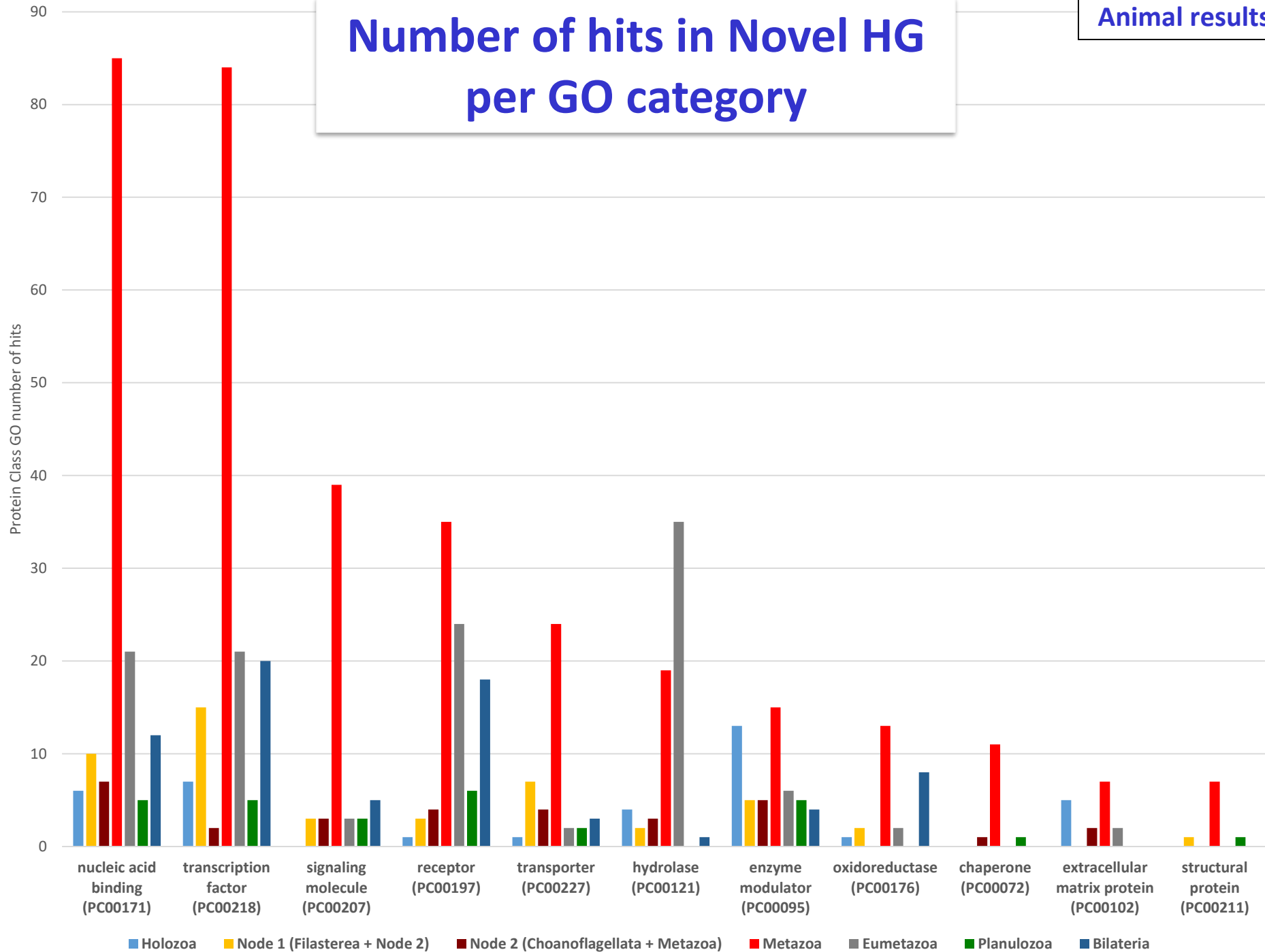


% Novelty

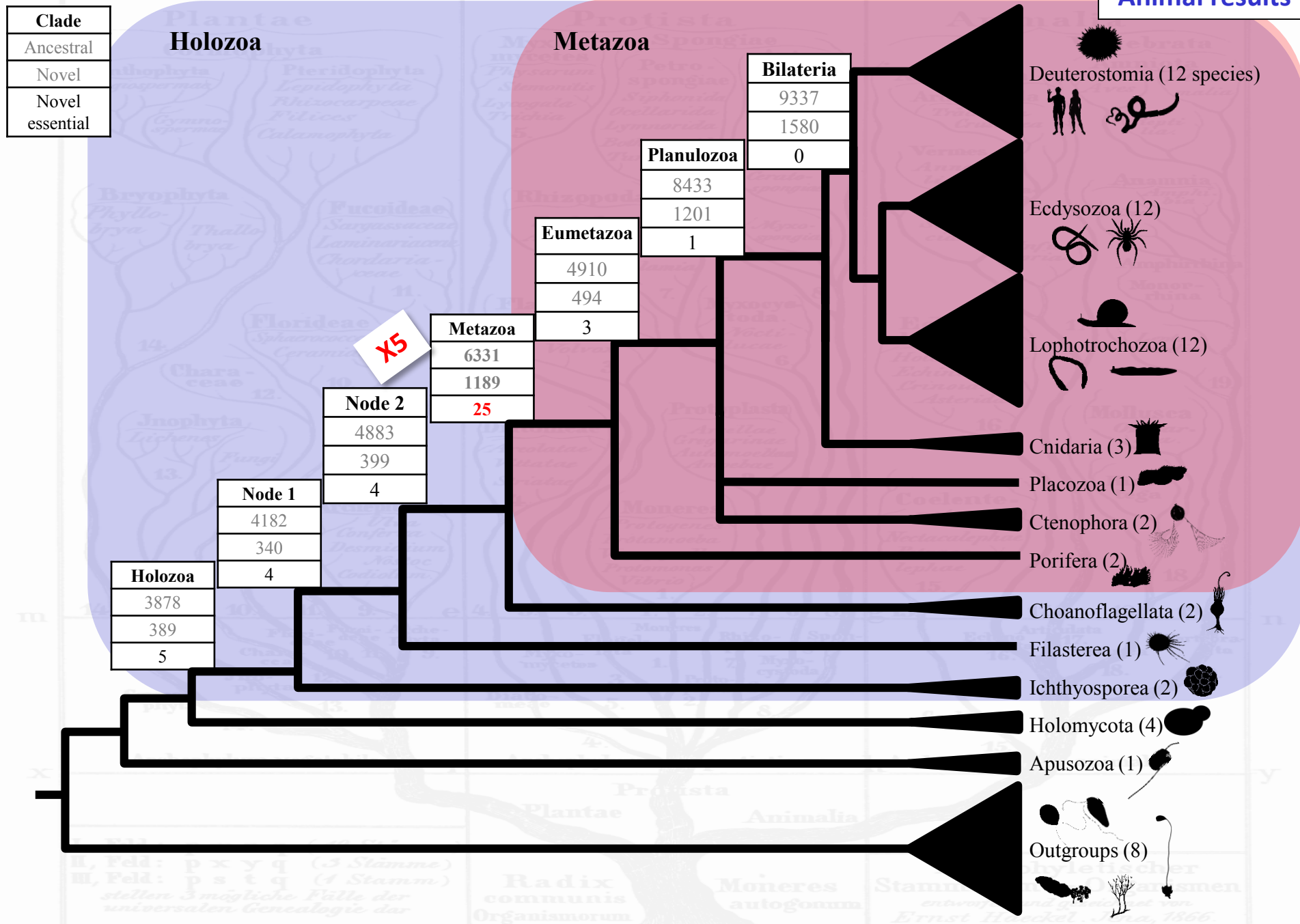
Animal results



Number of hits in Novel HG per GO category



25 ESSENTIAL NEW ANIMAL HOMOLOGY GROUPS



Summary of metazoan novelties

Animal results

8 HG involved in signalling pathways

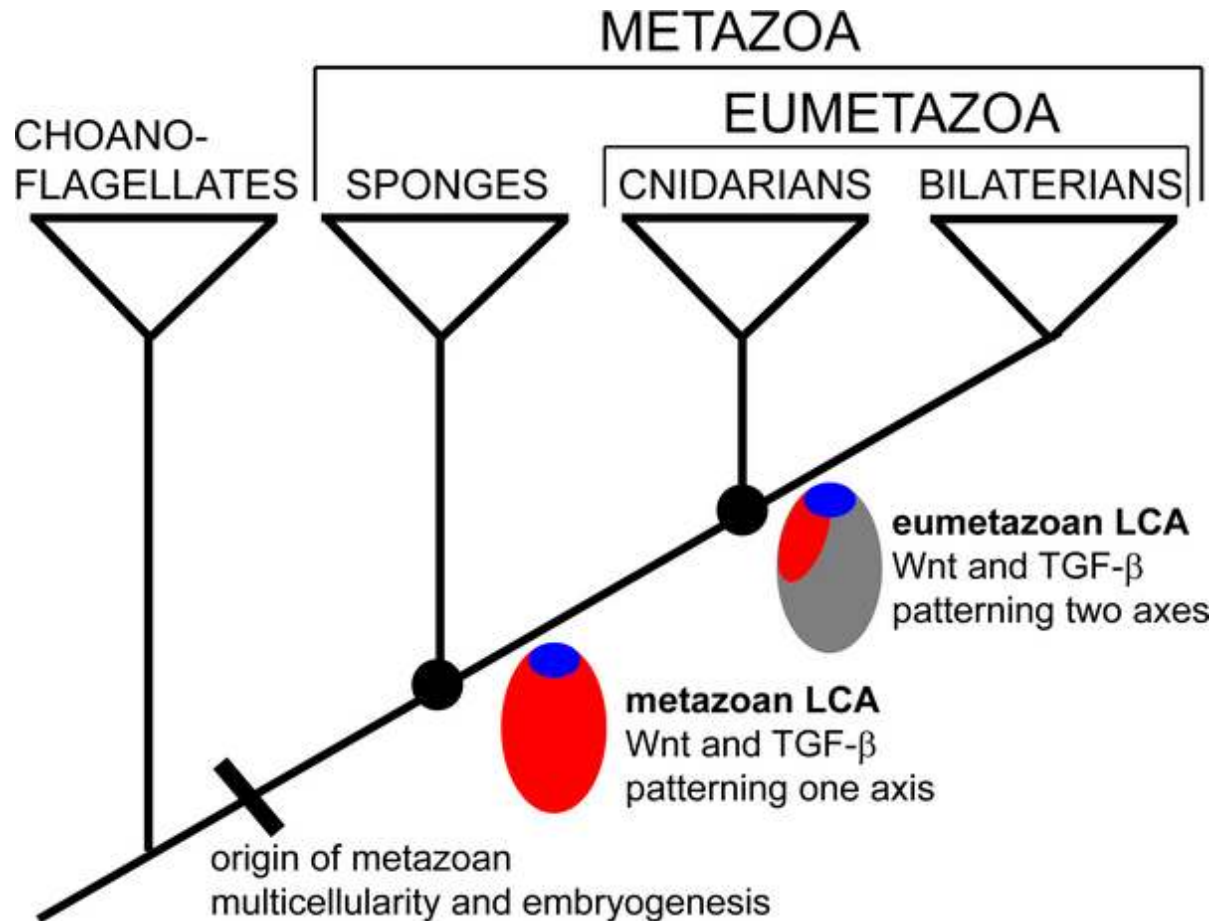
Signalling pathways

Fruitfly genes examples

Wnt	• Wnt	<i>wingless</i>
	• Frizzled	<i>frizzled, smoothened</i>
	• pangolín/TCF-LEF	<i>pangolin</i>
	• armadillo/beta-catenin	<i>armadillo</i>
TGF-Beta/BMP	• TGF-Beta/BMP	<i>decapentaplegic, screw, activin beta</i>
	• SMAD	<i>mothers against decapentaplegic</i>
	• TFG-Beta Receptor	<i>punt, saxophone, wishful thinking</i>
	• MAPKKK	<i>sunday driver</i>

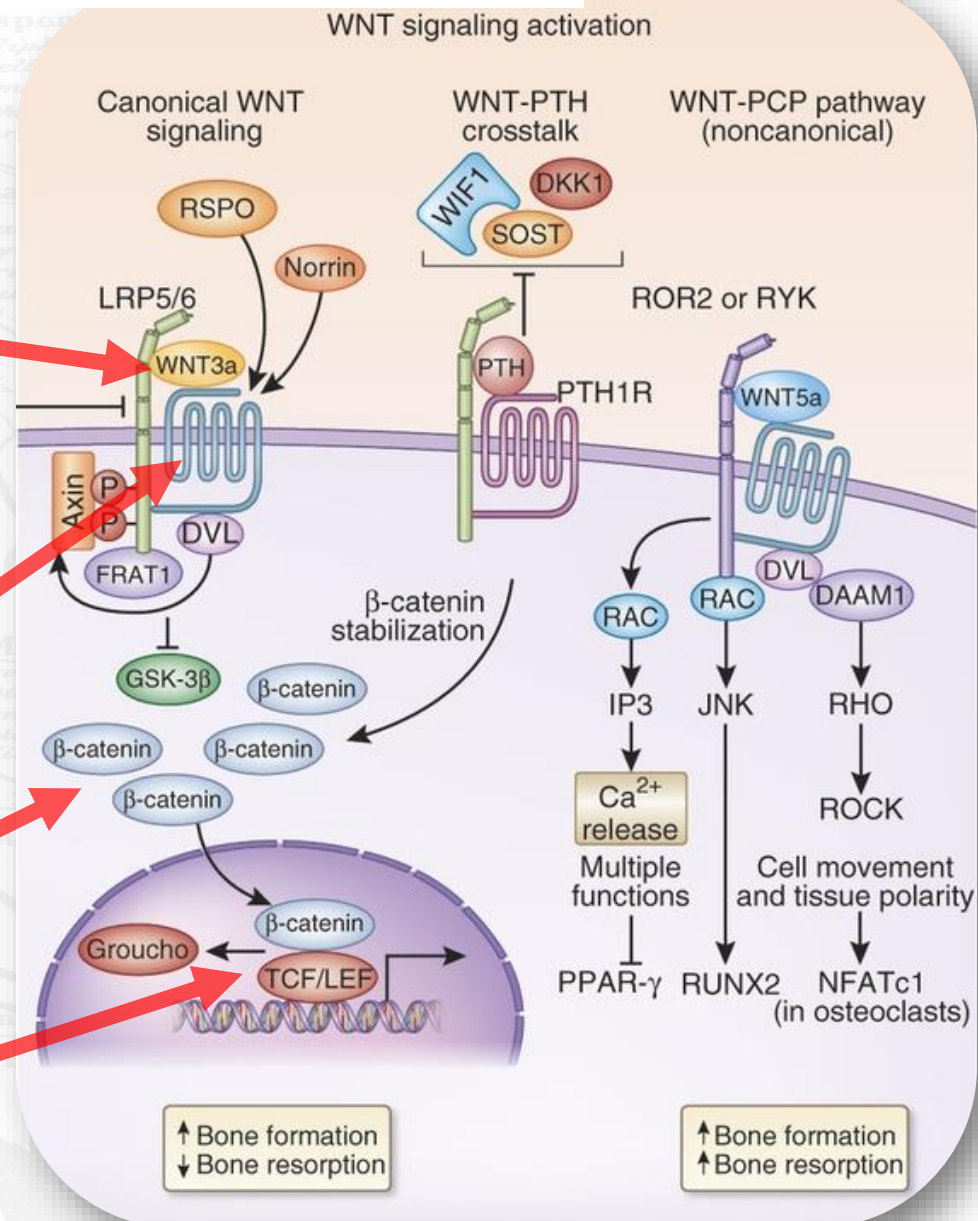
Wnt and TFG-Beta

Animal results



Signalling pathways: Wnt

Wnt	Ligand
Frizzled	GPCR receptor of Wnt, activates <i>Dishevelled</i> in cytosol
Armadillo/beta-catenin	Translocation to nucleus
Pangolin LEF/TCF	Gene activation



TGF-Beta / BMP

Animal results

**TGF-β / BMP
Growth
differentiation
factor / inhibin /
myostatin**

Cell signalling proteins, involved in **growth, development, tissue homeostasis** and **immune system**. Cellular processes in adult and embryo (cell growth, differentiation, apoptosis, cellular homeostasis...), growth factors such as cytokines and as metabologen.

TGFβ ligands bind to a **type II receptor** -> **phosphorylates a type I receptor** -> phosphorylates receptor-regulated **SMADs** (R-SMADs) which can now bind the coSMAD SMAD4 -> R-SMAD/coSMAD act as TFs in the nucleus

**TGF Beta Recept
/ Activin Recept /
BMP Recept**

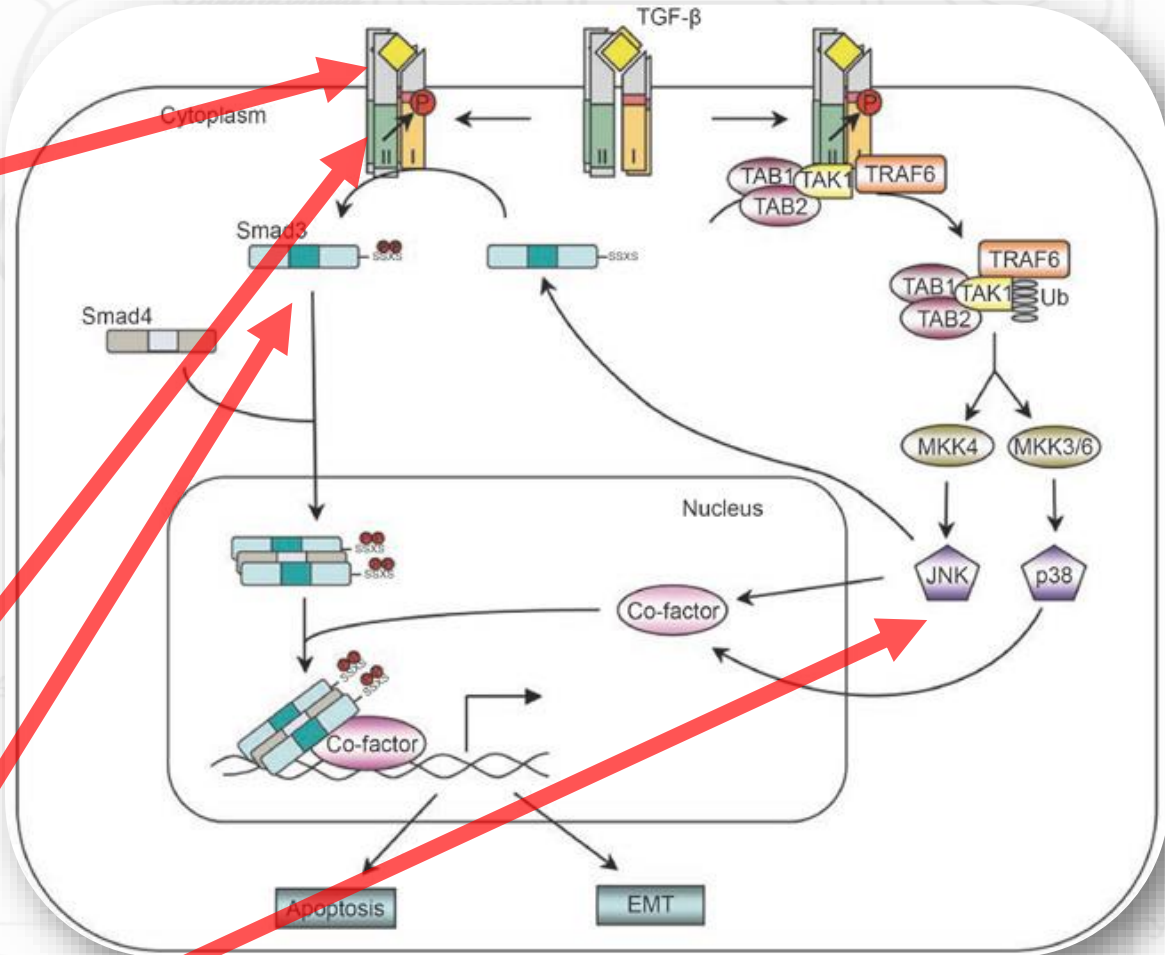
Transmembrane receptor

SMAD

Signal transducer of extracellular TGF beta ligands)

**MAP kinase
kinase kinase
(MAPK8IP3,
Sunday driver)**

serine/threonine-specific protein kinase, **JNK**



7 HG comprising Transcription Factors

	Transcription Factors	Fruitfly genes examples
Homeobox	NKL subclass of the ANTP Class	<i>tinman, distalless, ems</i>
	SIX Class	<i>sine oculis, optix</i>
	POU Class	<i>pou proteins, nubbi, ventral veins lacking</i>
bHLH	hes/hairy	<i>hairy, hey, deadpan, clockwork orange</i>
	bHLH-PAS	<i>single-minded, spineless, trachealess</i>
	twist/hand	<i>twist, hand, target of Poxn, taxi, atonal</i>
	ETS	<i>ets65A, anterior open, pointed</i>

Homeobox genes

Classes	Families
ANTP(767)	HOXL subclass Cdx(22) Evx(17) Gbx(15) Gsx(13) Hox1(21) Hox2(13) Hox3(25) Hox4(24) Hox5(19) Hox6-8(55) Hox9-13(15)(96) Meox(14) Mnx(15) Pdx(6)
	NKL subclass Abox(5) Ankx(1) Barhl(16) Bari(4) Barx(10) Bsx(10) Dbx(15) Dlx(35) Emx(22) En(19) Hhex(10) Hlx(9) Hx(1) Lbx(14) Lcx(1) Msx(22) Msxlx(3) Nanog(15) Nedx(5) Nk1(17) Nk2.1(18) Nk2.2(15) Nk3(14) Nk4(19) Nk5/Hmx(25) Nk6(19) Nk7(4) Noto(10) Ro(5) Tlx(19) Vax(12) Ventx(18)
PRD(458)	Alx(16) AprdA(1) AprdB(1) AprdC(1) AprdD(2) AprdE(1) Argfx(3) Arx(14) CG11294(3) Dmbx(8) Dprx(8) Drgx(9) Dux(42) Esx(2) Gsc(14) Hbn(3) Hesx(5) Hopx(5) Isx(5) Leutx(1) Mix(7) Nobox(5) Obox(36) Otp(12) Otx(24) Pax2/5/8(12) Pax3/7(24) Pax4/6(24) Phox(17) Pitx(20) Prop(8) Prrx(14) Rax(14) Repo(4) Rhox(39) Sebox(3) Shox(12) Tprx(5) Uncx(16) Vsx(18)
LIM(104)	Isl(14) Lhx1/5(17) Lhx2/9(19) Lhx3/4(15) Lhx6/8(18) Lmx(21)
POU(112)	Hdx(6) Pou1(6) Pou2(24) Pou3(29) Pou4(21) Pou5(12) Pou6(15)
HNF(23)	Ahnfx(1) Hmbox(10) Hnf1(12)
SINE(48)	Six1/2(18) Six3/6(17) Six4/5(13)
TALE(147)	Atale(1) Irx(44) Meis(23) Mlxx(8) Pbx(29) Pknox(14) Tgif(28)
CUT(58)	Acut(1) Cmp(5) Cux(17) Onecut(24) Satb(11)
PROS(16)	Prox(16)
ZF(83)	Admp(9) Azfh(1) Tshz(16) Zeb(14) Zfhx(21) Zhx/Homez(22)
CERS(26)	Cers(26)

Transcripts polyadenylation

- **Cytoplasmic Polyadenylation Element Binding (CPEB)**

Binds the **polyA tail of mRNA** and **regulates their translation**

Cell adhesion

- **Alpha-catenin**
- **Fermitin**
- **Liprin**

Activation of the **integrin pathway** and other **cell-extracellular matrix** interactions via **focal adhesions**

Cell cycle

- **RUN**
- **MADD/GEF**

Cell **proliferation**

Cell **apoptosis**

Receptors

- **Nuclear Hormone Receptors**
- **Neurotransmitter Receptors**

Synaptic exocytosis

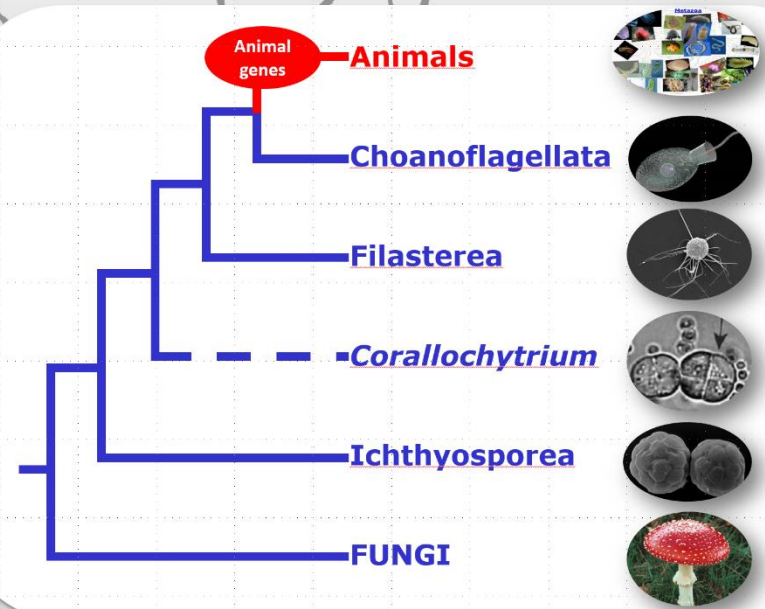
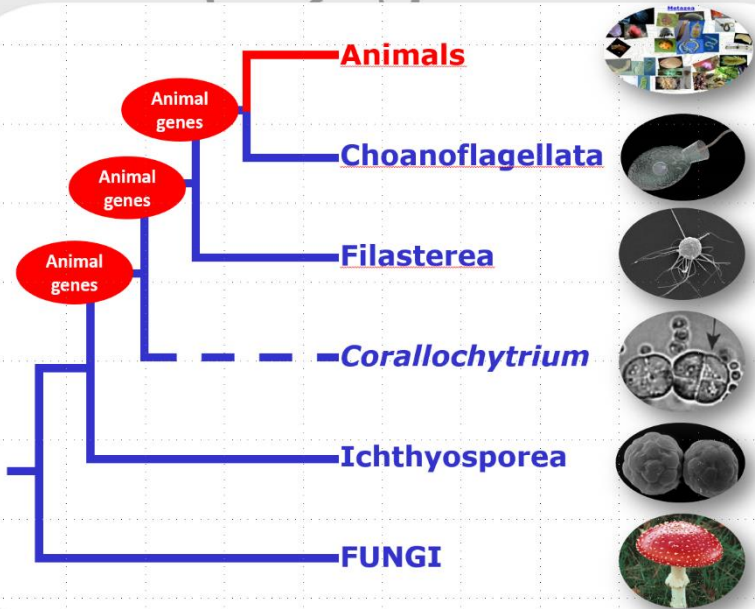
- **Calcium activated protein for secretion (CAPS)**
- **Rab3-interacting molecules (RIM)**

Regulate the **fusion of vesicles** with the cell membrane, **neuroendocrine cells** and **presynaptic neurons** respectively



Old vs new

Animal results



ANIMAL ORIGINS: STEW VS POPCORN

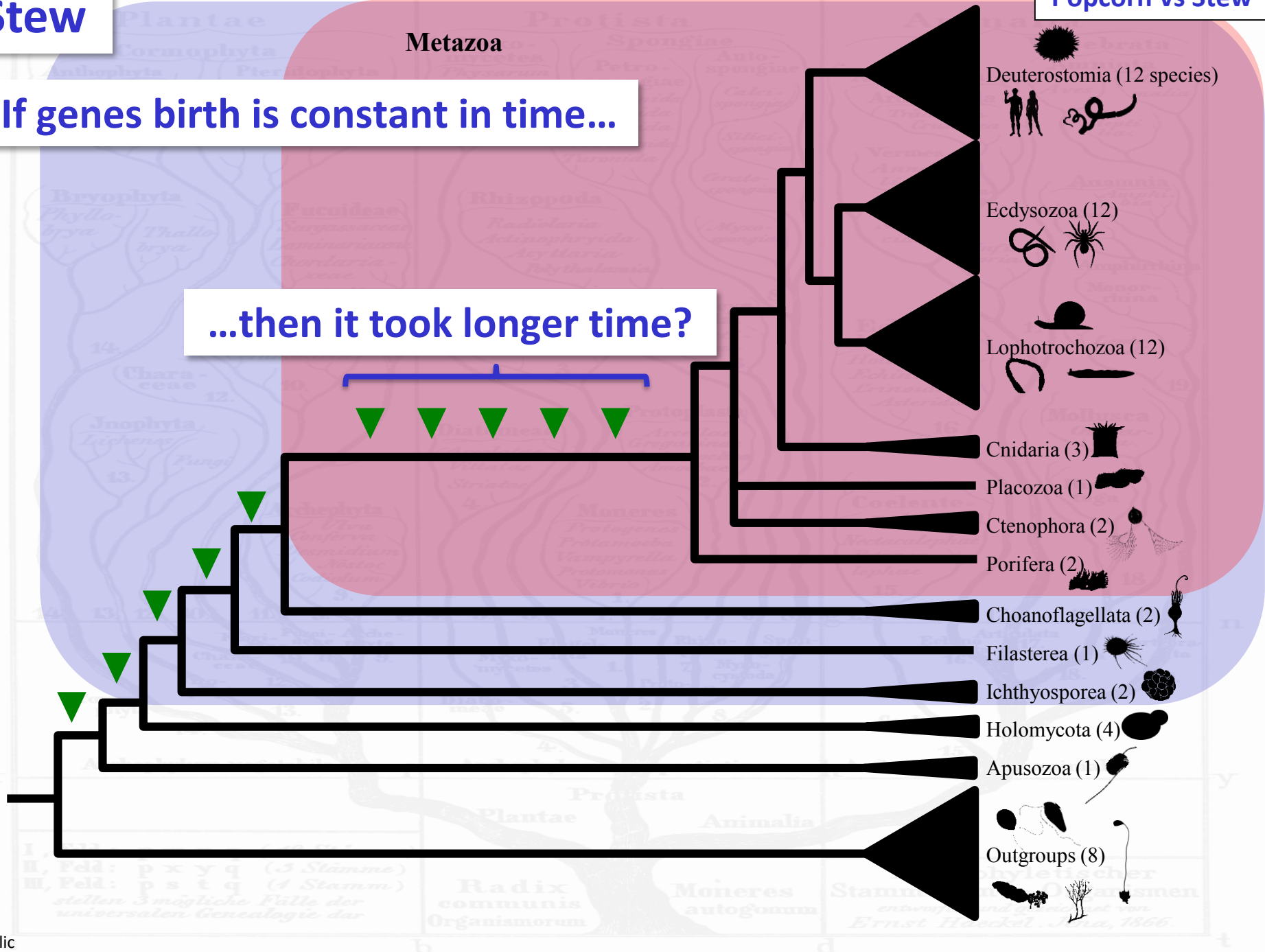
Stew

Popcorn vs Stew

•If genes birth is constant in time...

...then it took longer time?

Metazoa



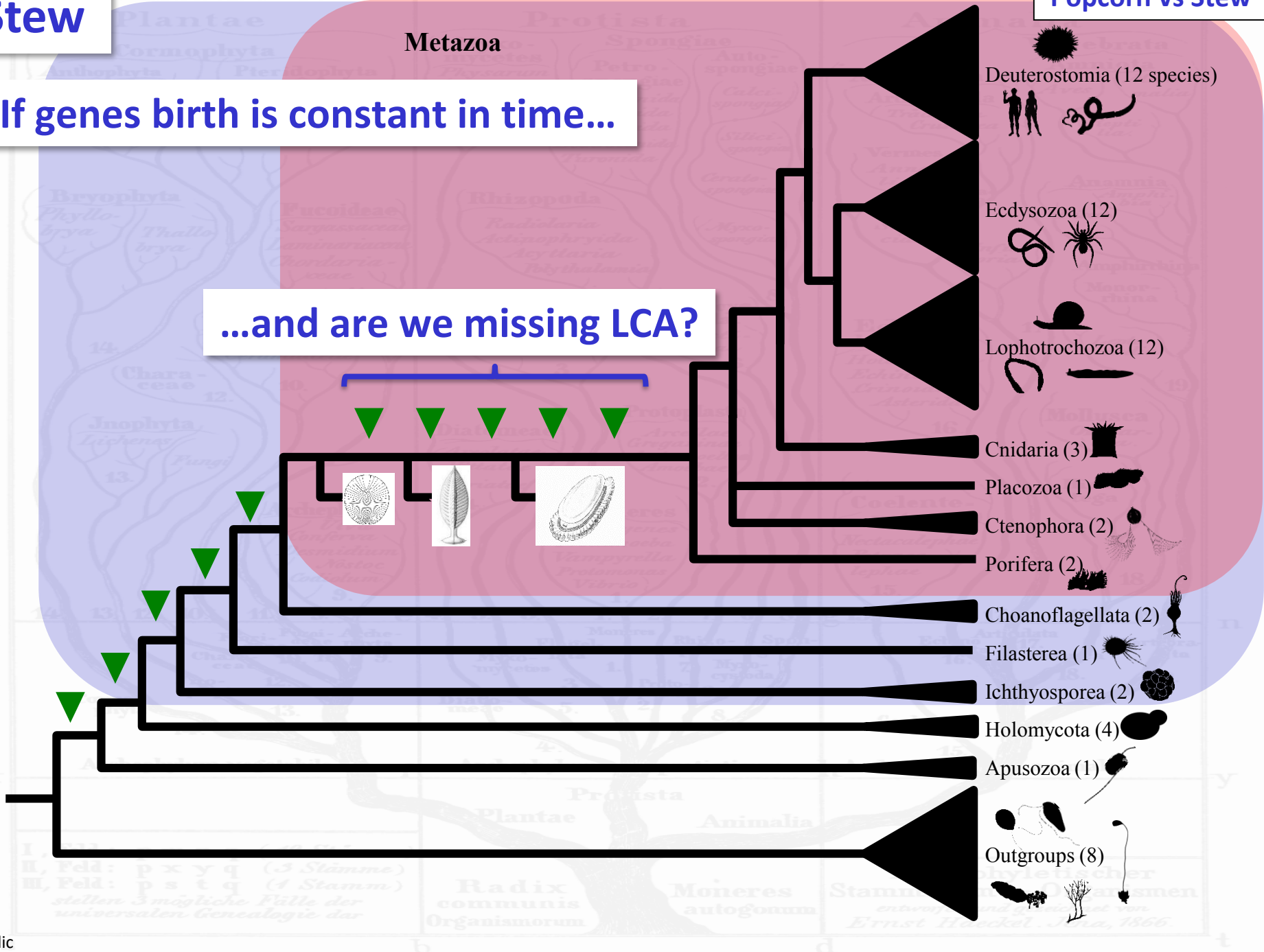
Stew

Popcorn vs Stew

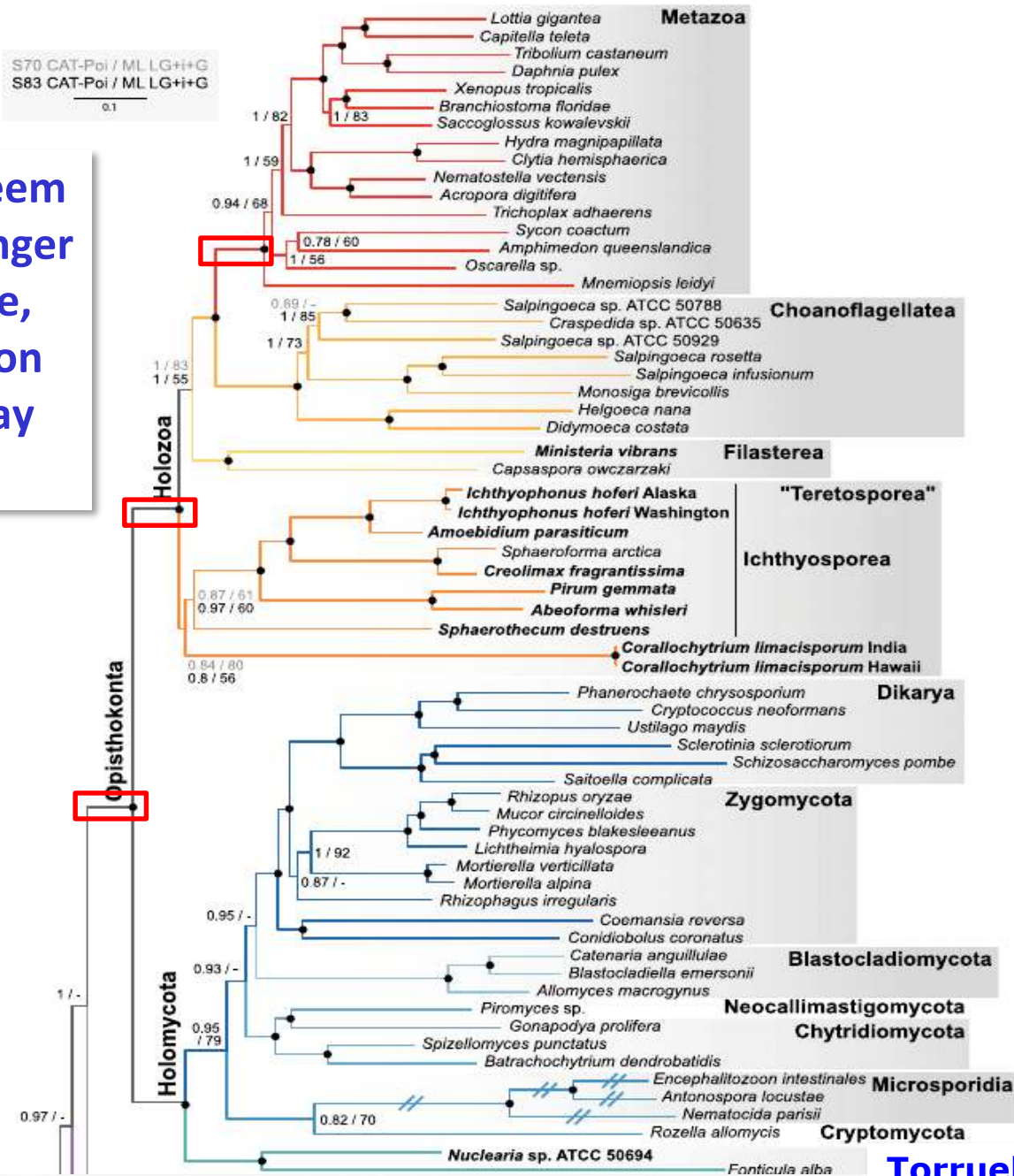
•If genes birth is constant in time...

...and are we missing LCA?

Metazoa



Trees do not seem to support a longer period of time, but substitution saturation may mask it.



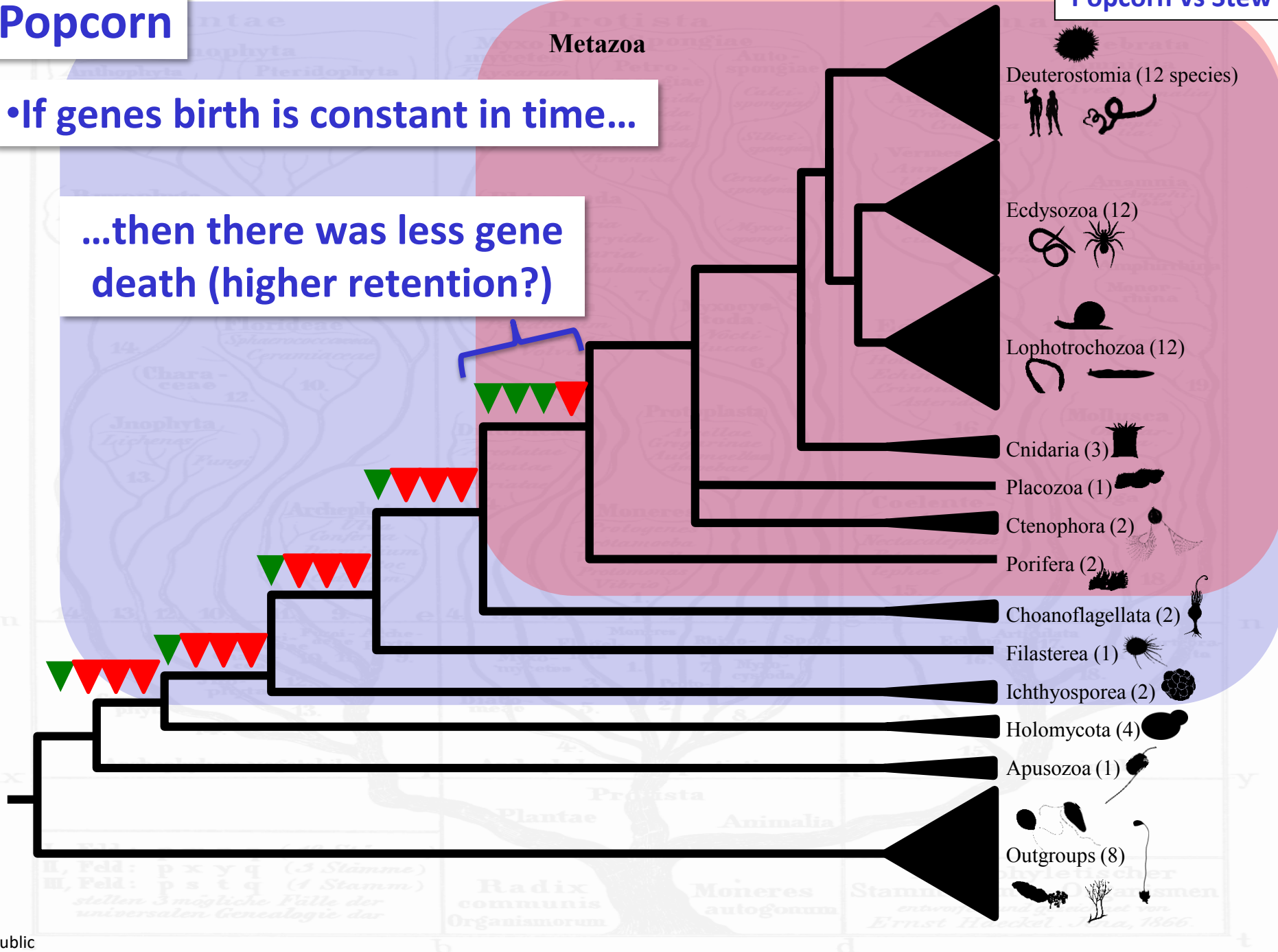
Popcorn

•If genes birth is constant in time...

...then there was less gene death (higher retention?)

Metazoa

Popcorn vs Stew



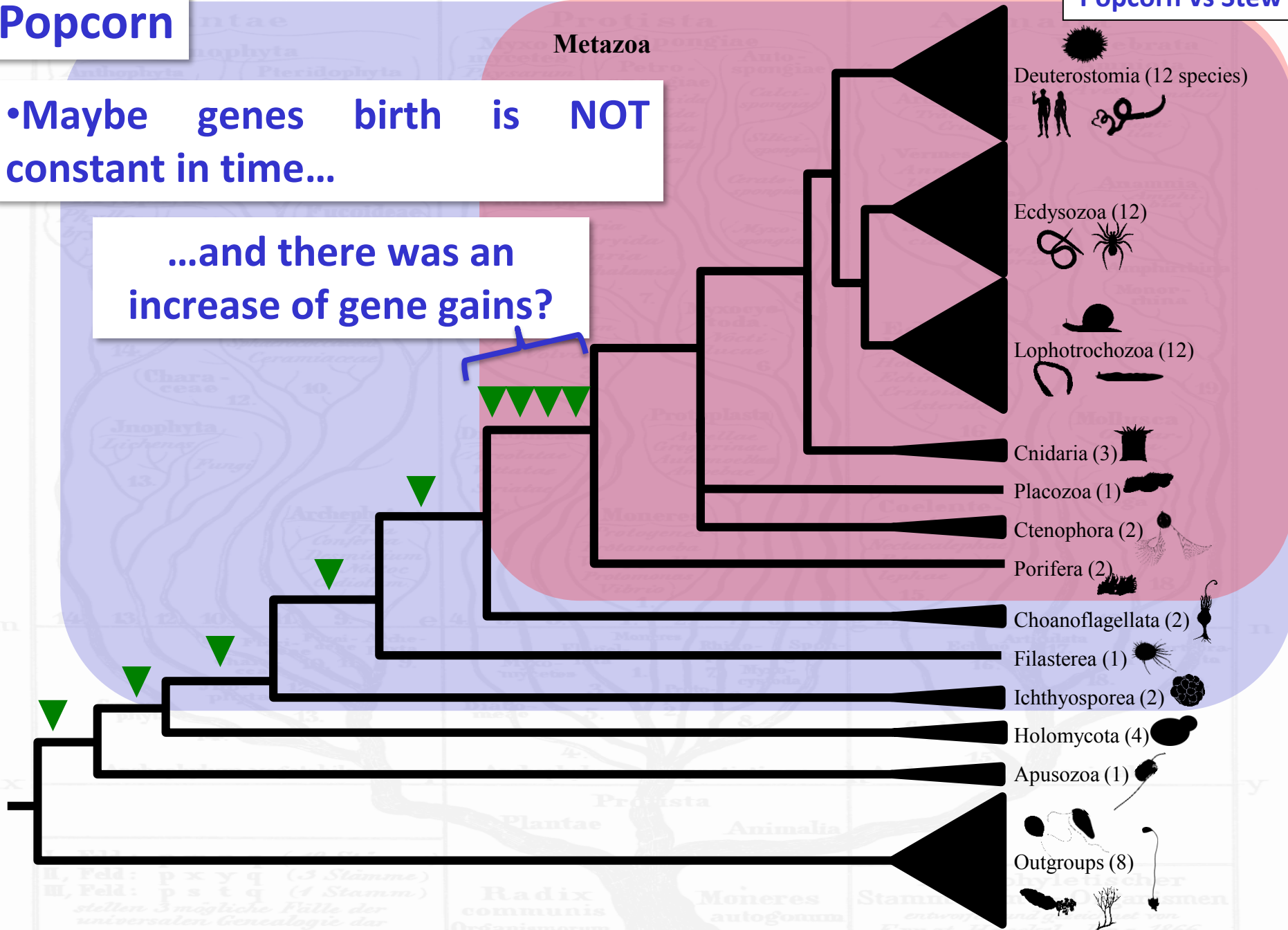
Popcorn

• Maybe genes birth is NOT constant in time...

...and there was an increase of gene gains?

Metazoa

Popcorn vs Stew



To sum up...

Conclusions

- The reconstruction of the first animal genome shows a (*boring*) list of **6381** homology groups. Predominant functions include **gene regulation** and **metabolism**, its composition is similar to other **LCA** and modern metazoan genomes.
- The origin of animals, compared to other nodes, shows a **surge in the level of gene novelties**, specially for **gene regulation**.
- There is a **core set of 25 homology groups**, never lost since the origin of animals, formed by some **iconic systems** (i.e. Wnt, Nk homeobox), but also by other **novel genes** never linked before to animal origins with relevant functions related to multicellularity.

