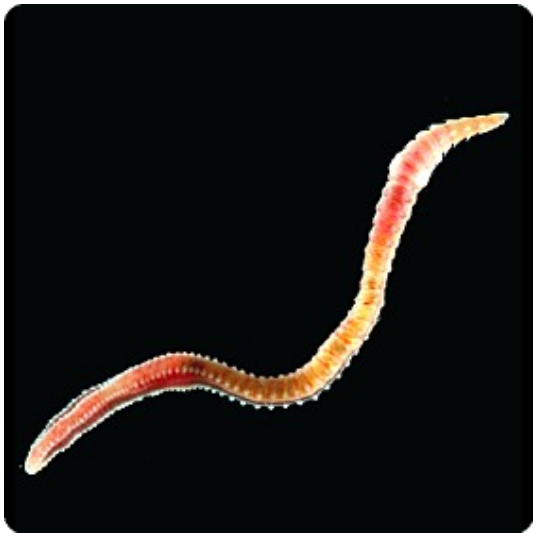
Comparative genomics:

What animal genomes can tell us about their past (and future?)




About myself

Until 2009 Studies U Karlsruhe/U Heidelberg, DE

MSc thesis UC Berkeley, USA

(Dan Rokhsar)

2009 – 2013 PhD

EMBL Heidelberg and U Konstanz, DE

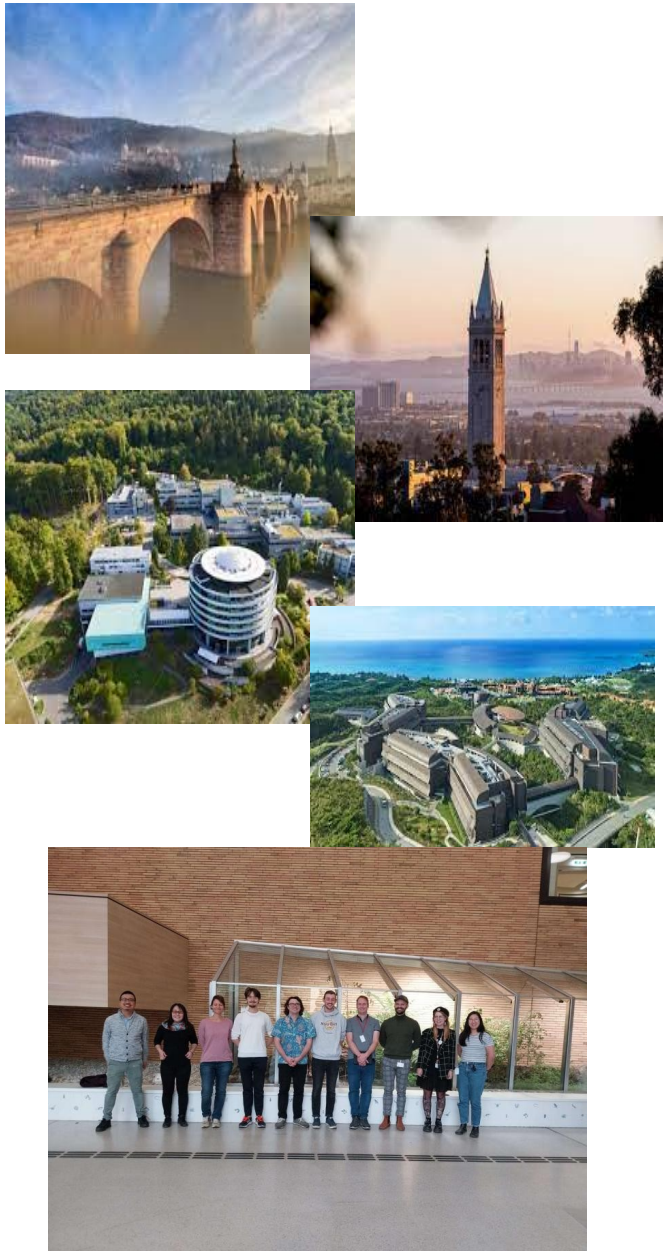
(Detlev Arendt, Axel Meyer)

2013 – 2017 Postdocs

U Heidelberg (Thomas Holstein)

OIST Okinawa, Japan (Dan Rokhsar, Sydney Brenner)

Since 2017: Assis/Assoc Prof U Vienna, AT



Personal motivation:
“evolution/implications of “complexity” ”

- Around 2008: Early animal genomic sequencing accomplishments

“ancestral complexity” - much cool

MSc project: hydra and sponge genomes

“The dynamic genome of Hydra”

Back then: no bioinfo courses! Pascal, Delphi (I do write manuals! molbiosoft.de), then perl

PhD:

Evo-devo from a quantitative genomics perspective
Attempt to overlay genome
(micro/macro-)evolutionary trends onto cell type evolution. *Platynereis* foregut as a model system.

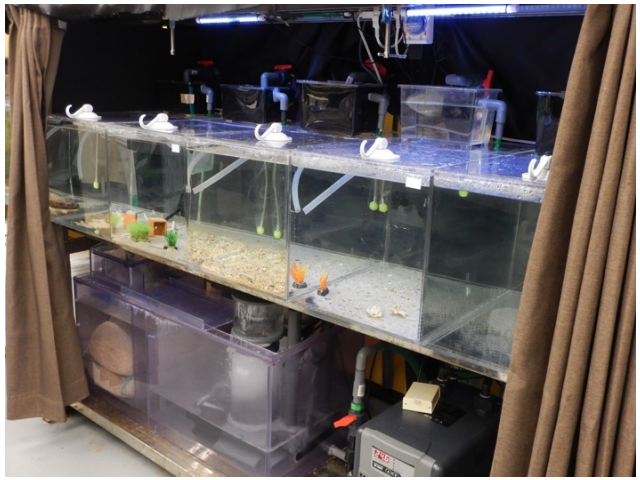
- too “difficult”

- Back to genomics
The three spiralian genomes paper

Postdocs:

Early branching animal genome architecture:
From micro- to macro-synteny (and function)

Changes in genome architecture as measured by synteny



The system at OIST that almost burned down!

Press: “omg, aliens”

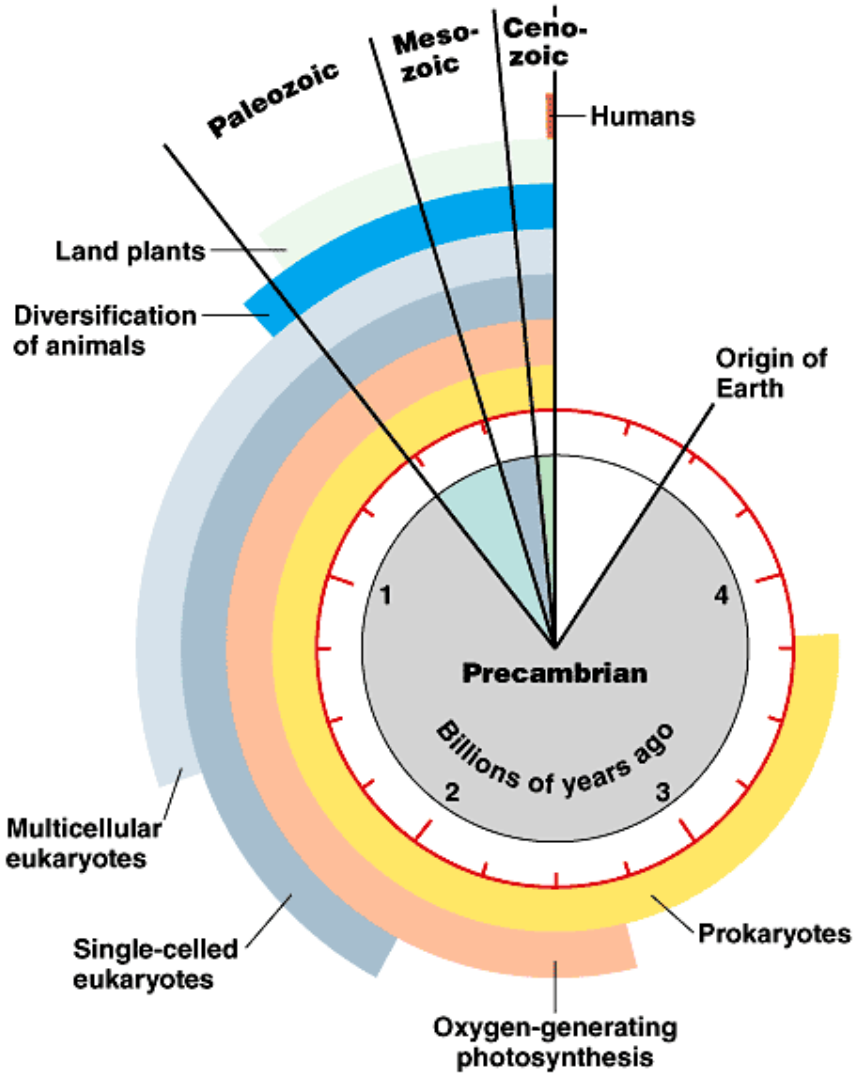
Tenure-track at U Vienna:

Genome architecture (synteny), its evolutionary properties (also repetitive element forces) and impact of its change onto gene regulation

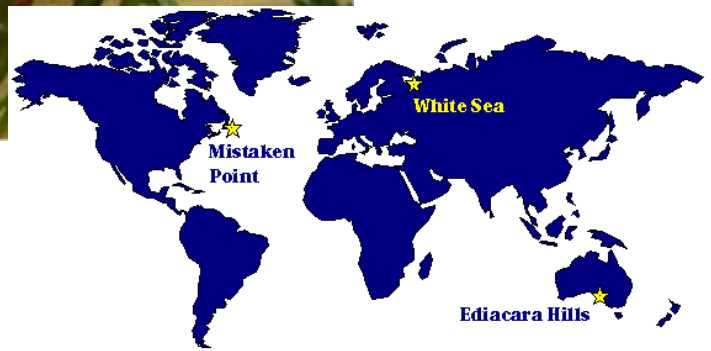
I. (Somewhat) early animal genomics and initial insights
II. Current approaches and bottlenecks (from chromosomes to 3D structure?)
III. Future theoretical genomics, predictions and applications

I. (Somewhat) early animal genomics and initial insights

‘Recent’ appearance of animal multicellularity



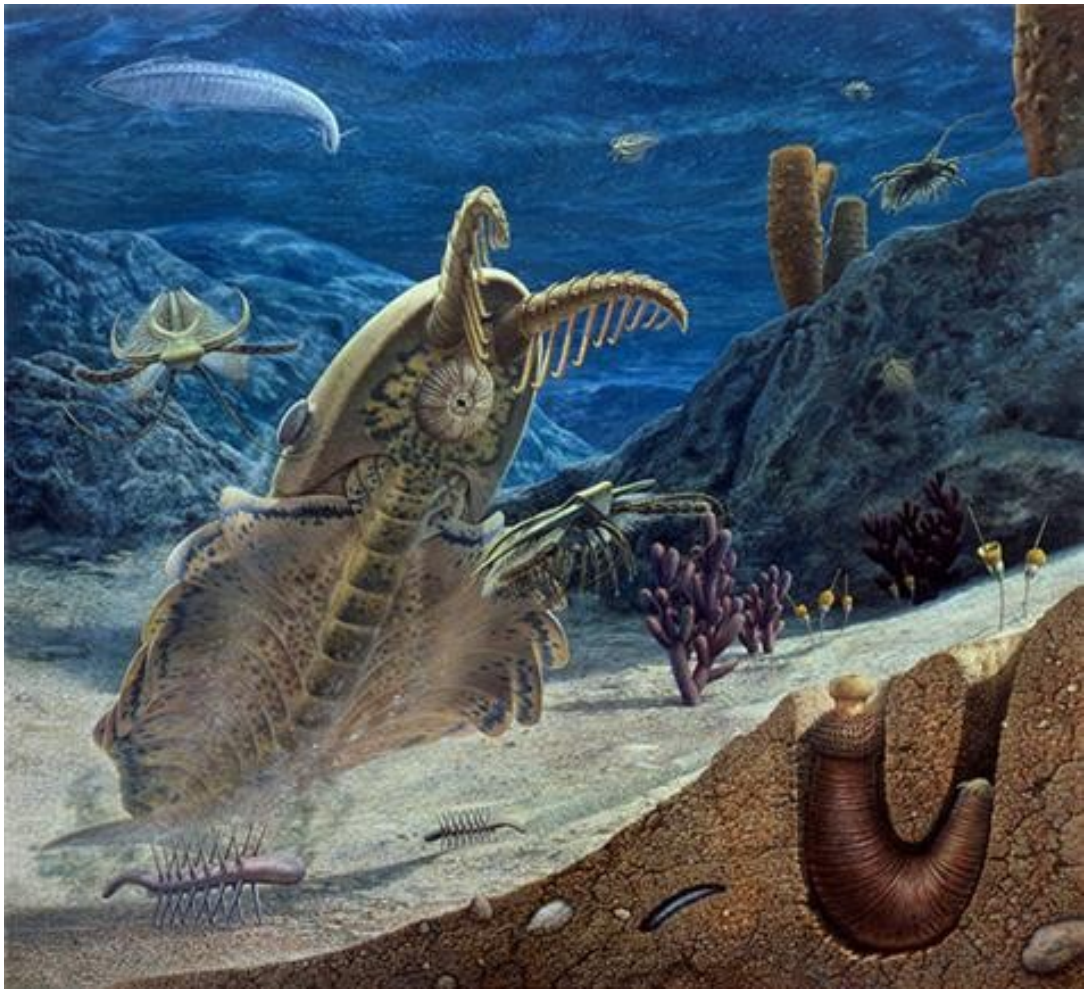
The cradle of animal life: Ediacaran/Vendian



650-540 million years ago
No predation
Mostly sessile
Sponge/cnidarian grade organization

Image: Ryan Somma
Map: https://ucmp.berkeley.edu/vendian/vendianloc.htm

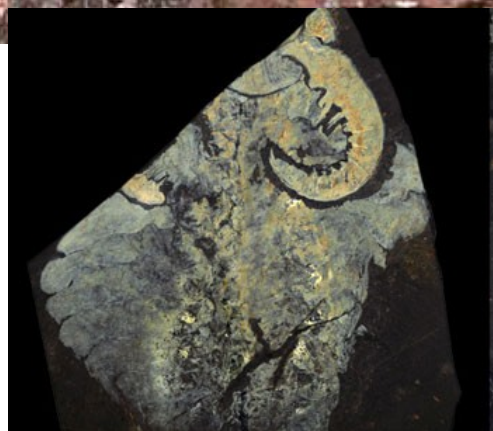
The ‘explosion’: Cambrian



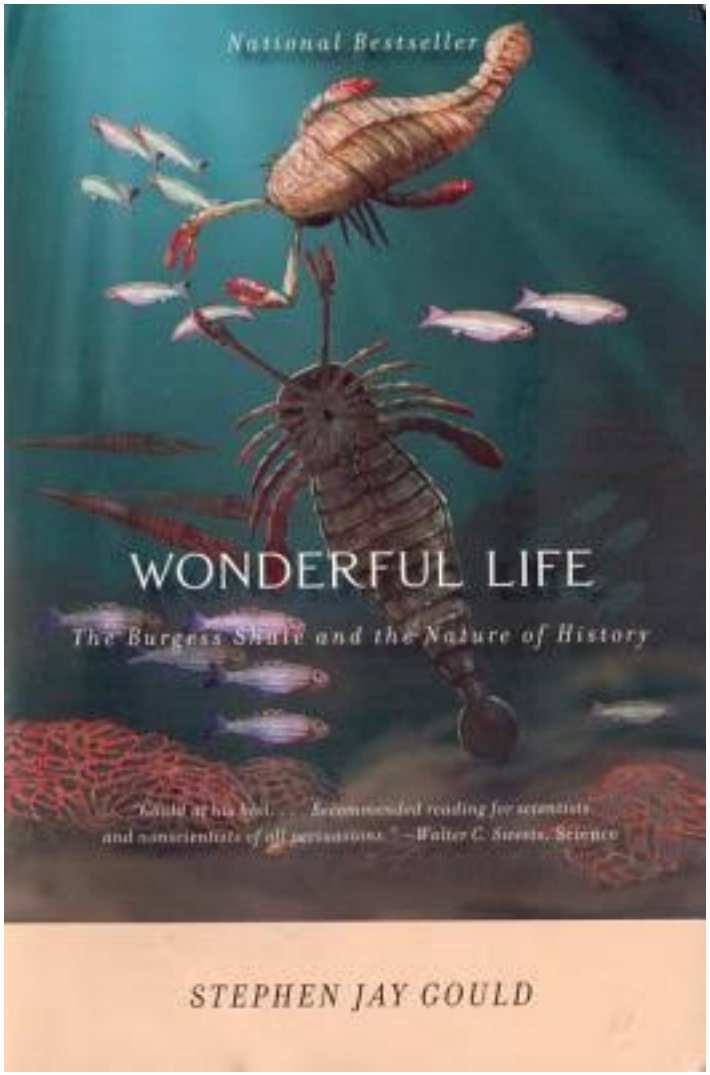
540-520 million years ago
True bilateral symmetry
High predation
Origin of the major animal phyla

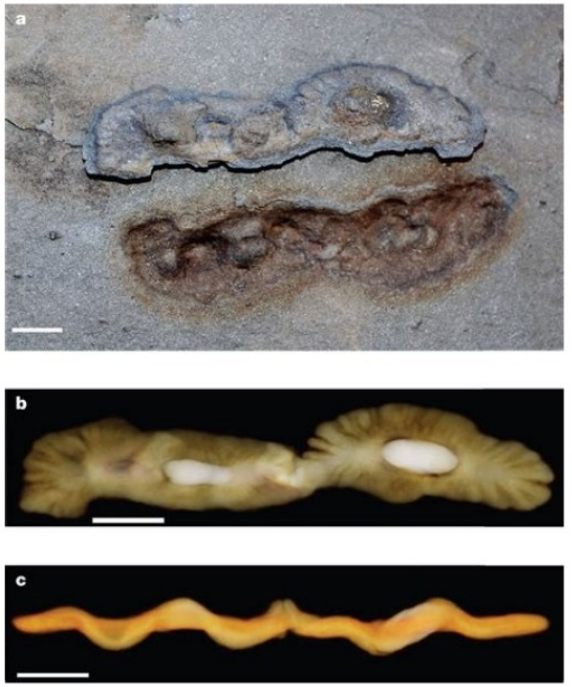
Image: John Sibbick

Burgess shale



Wonderful life? (how often?)





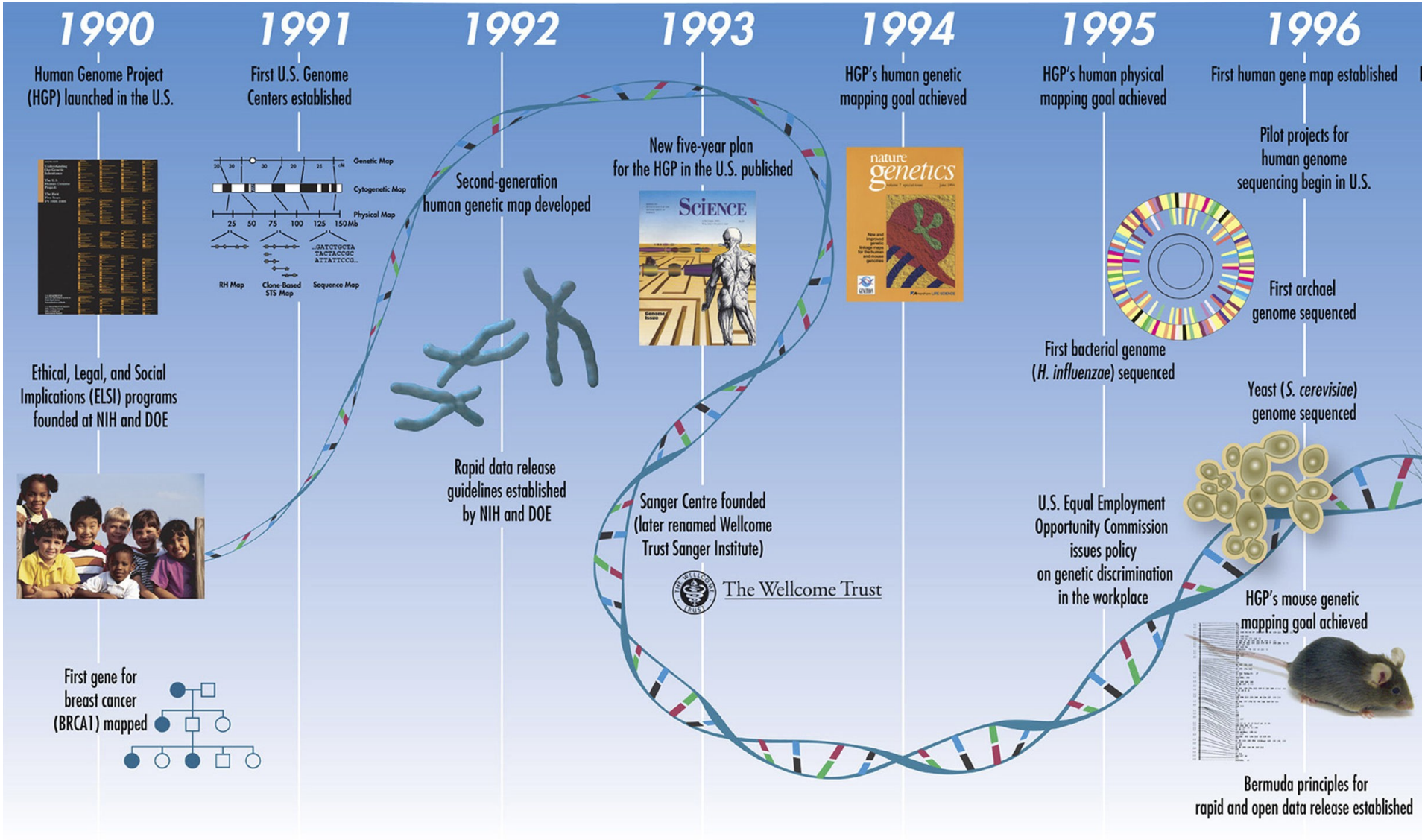
Large colonial organisms with coordinated growth in oxygenated environments 2.1 Gyr ago

Abderrazak El Albani, Stefan Bengtson, Donald E. Canfield, Andrey Bekker, Roberto Macchiarell
Arnaud Mazurier, Emma U. Hammarlund, Philippe Boulvais, Jean-Jacques Dupuy, Claude Fontain
Franz T. Fürsich, François Gauthier-Lafaye, Philippe Janvier, Emmanuelle Javaux, Frantz Ossa Oss
Anne-Catherine Pierson-Wickmann, Armelle Riboulleau, Paul Sardini, Daniel Vachard, Martin
Whitehouse & Alain Meunier

Enter “genome era”

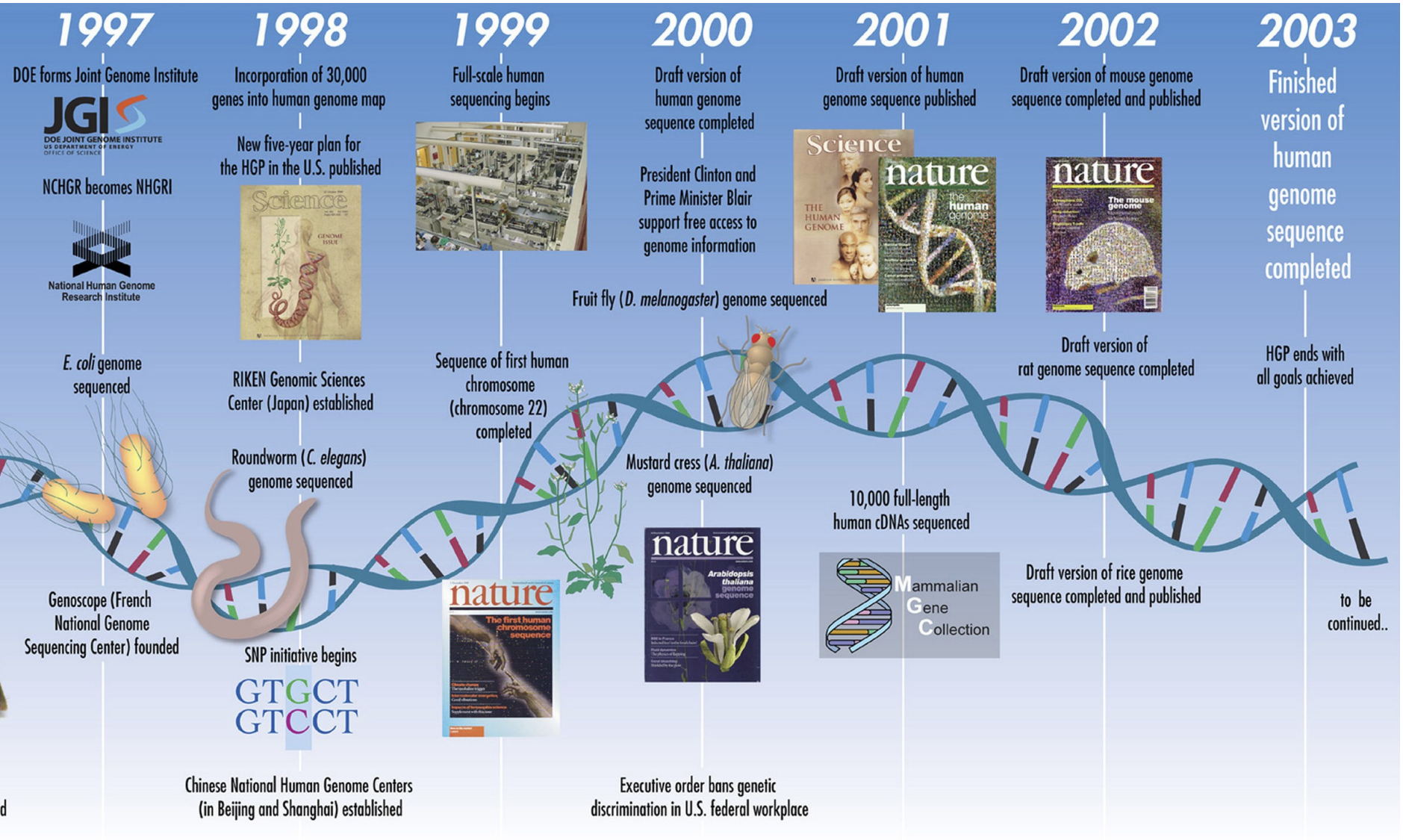
Working draft of human genome

Human Genome Project



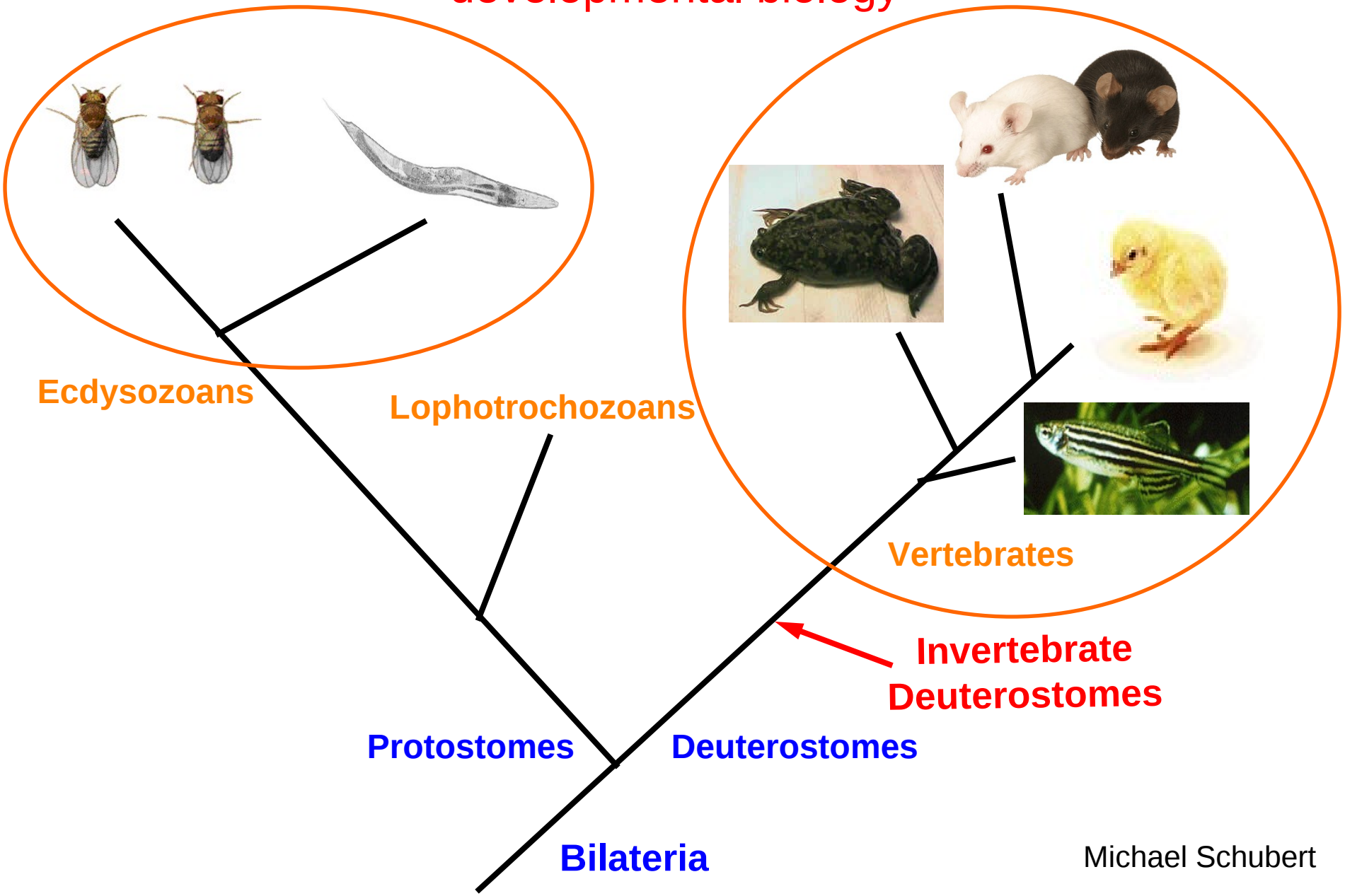
https://www.mun.ca/biology/scarr/Human_Genome_Project_timeline.html

Human Genome Project



https://www.mun.ca/biology/scarr/Human_Genome_Project_timeline.html

Pre-genomic era: Phylogeny of the classical model systems in developmental biology



Michael Schubert

Previous notion – basal animals are 'simple'

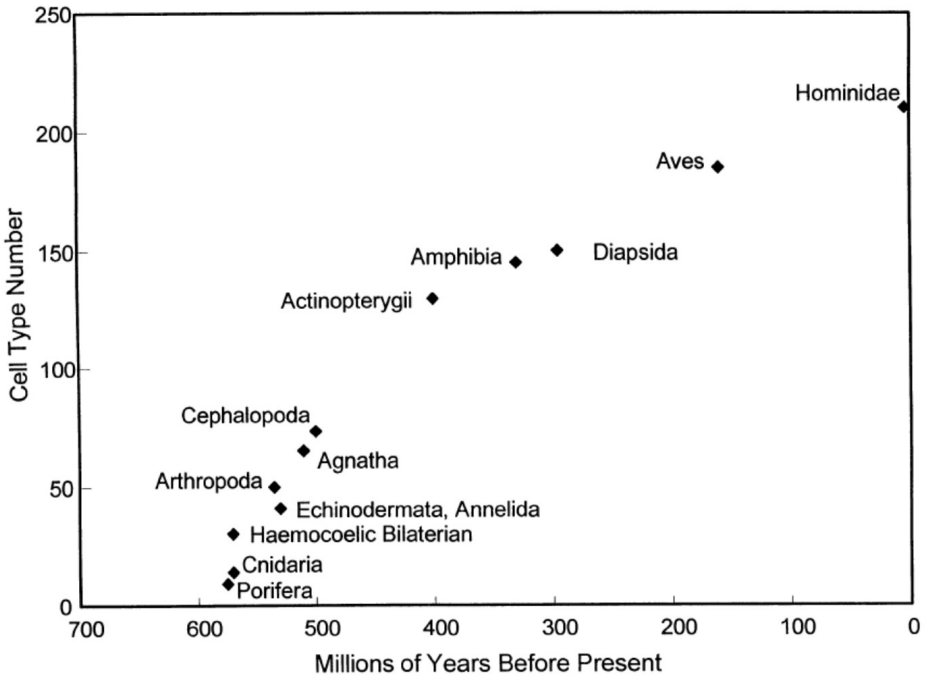
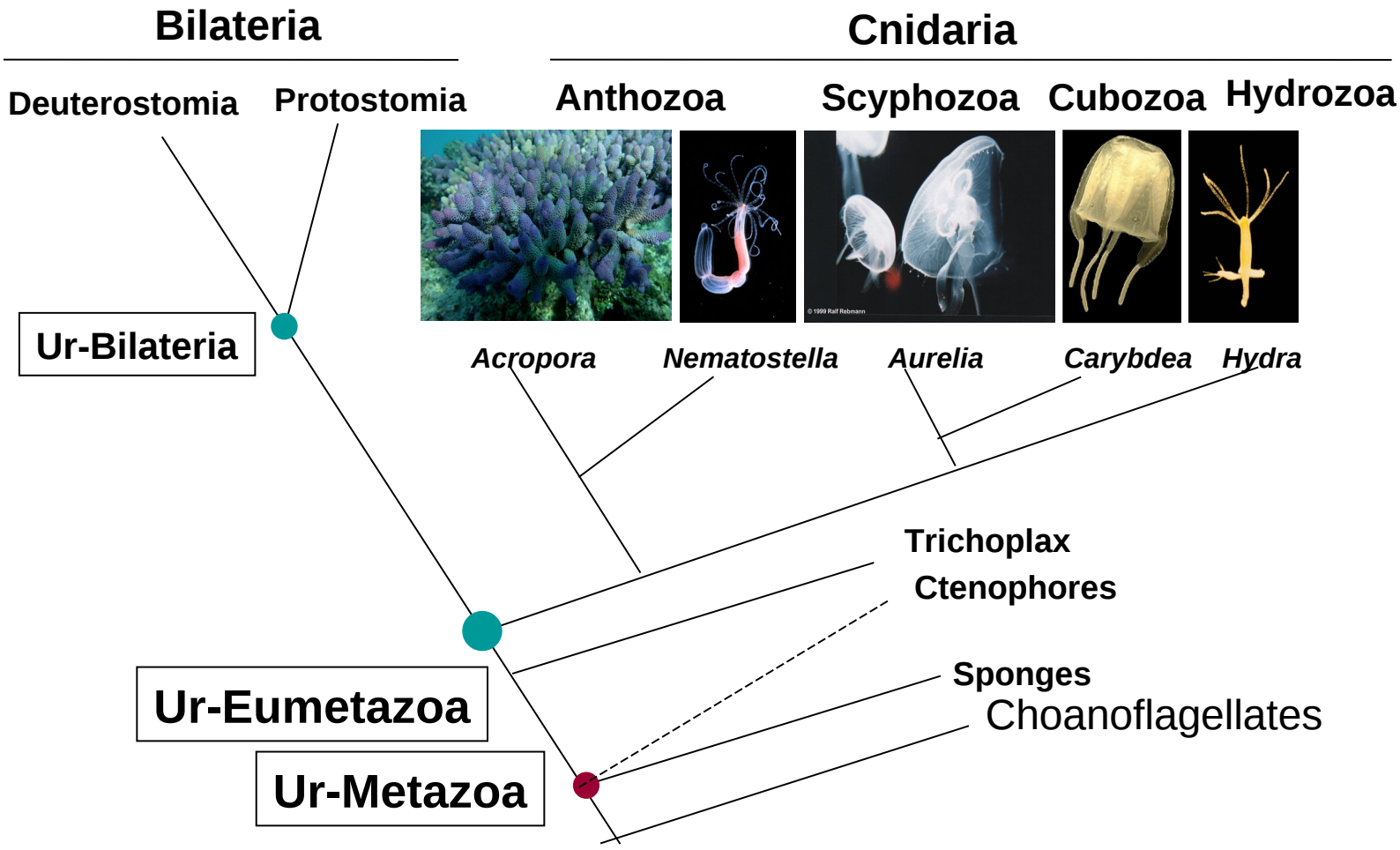


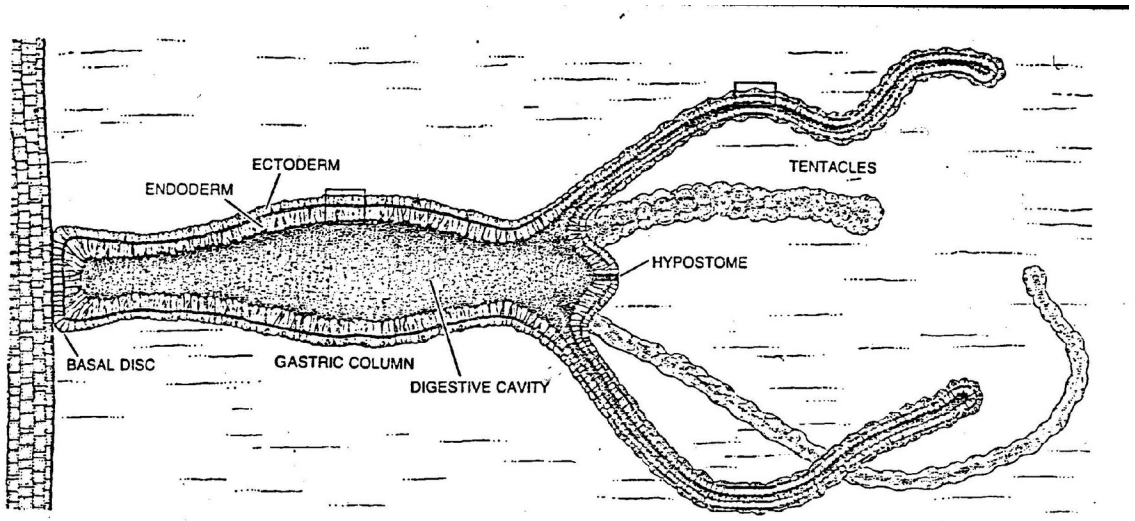
FIGURE 1. Estimated cell-type numbers of primitive members of selected metazoan taxa, inferred from counts of cells in living individuals, plotted against the estimated time of origin of the body plan of each taxon. Only taxa that are believed to have been rather near the upper bound of cell-type numbers when originating are included. For sources on which cell-type number estimates are based, and sources for ages, see Appendix.

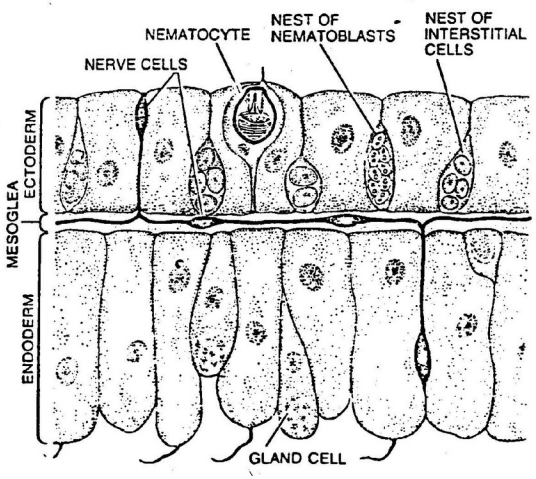
Reconstructing the common ancestor of Cnidaria and Bilateria

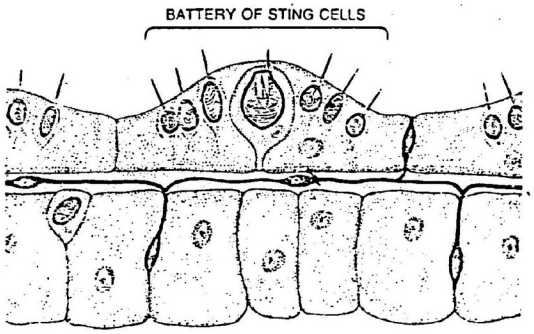


Cnidarians are simple (?)

2 cell layers, one apparent body axis, 20 (?) cell types

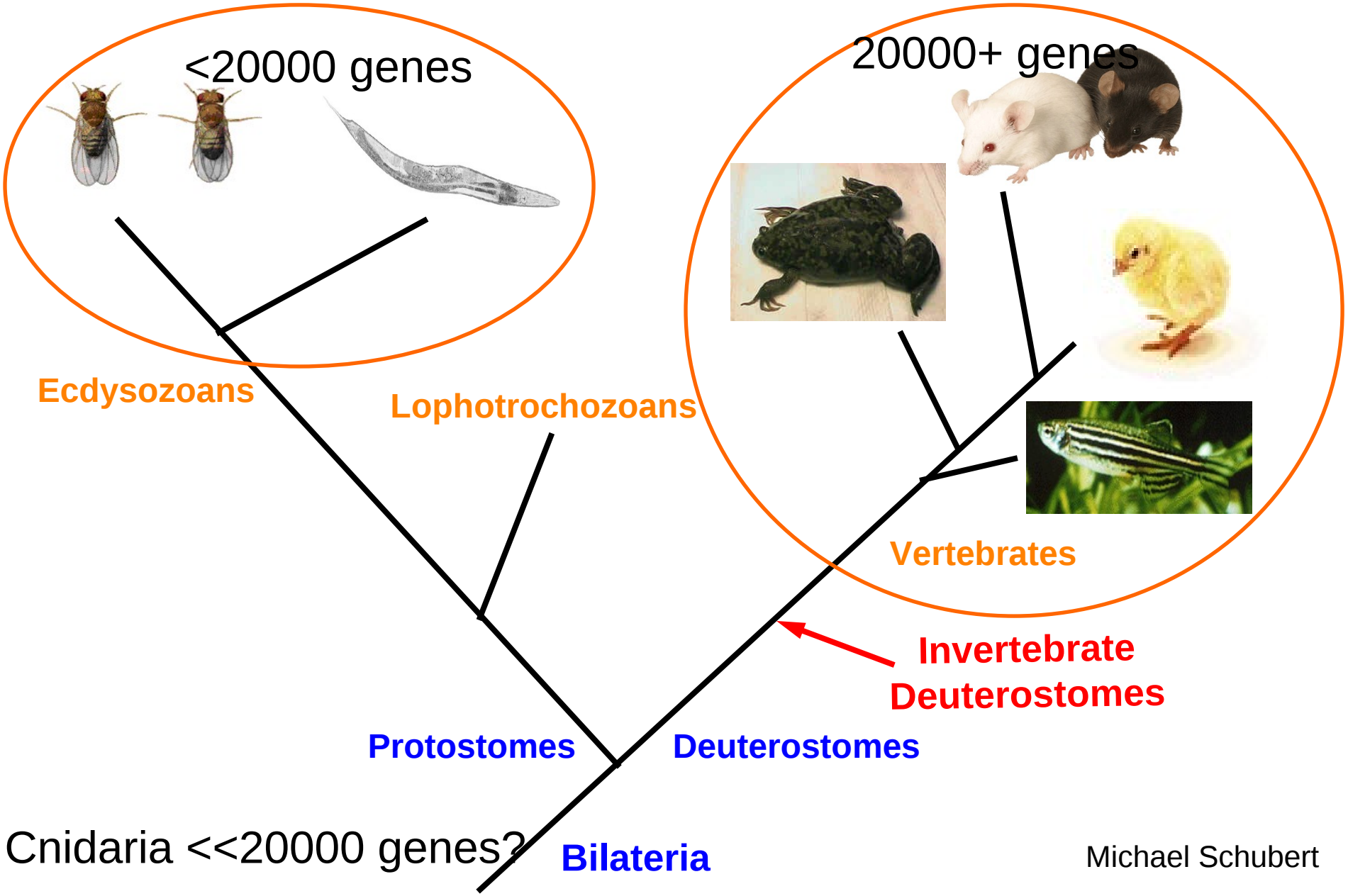






Gierer et al ,1974

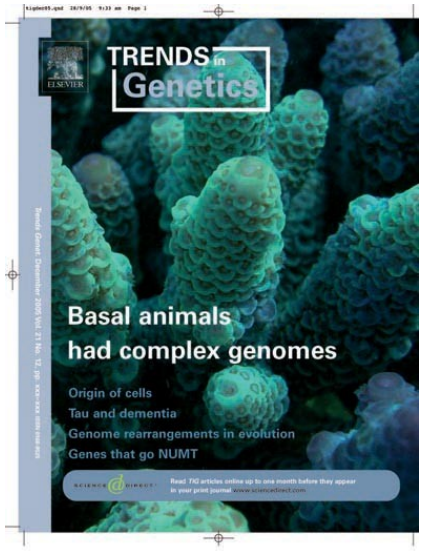
Pre-genomic era predictions



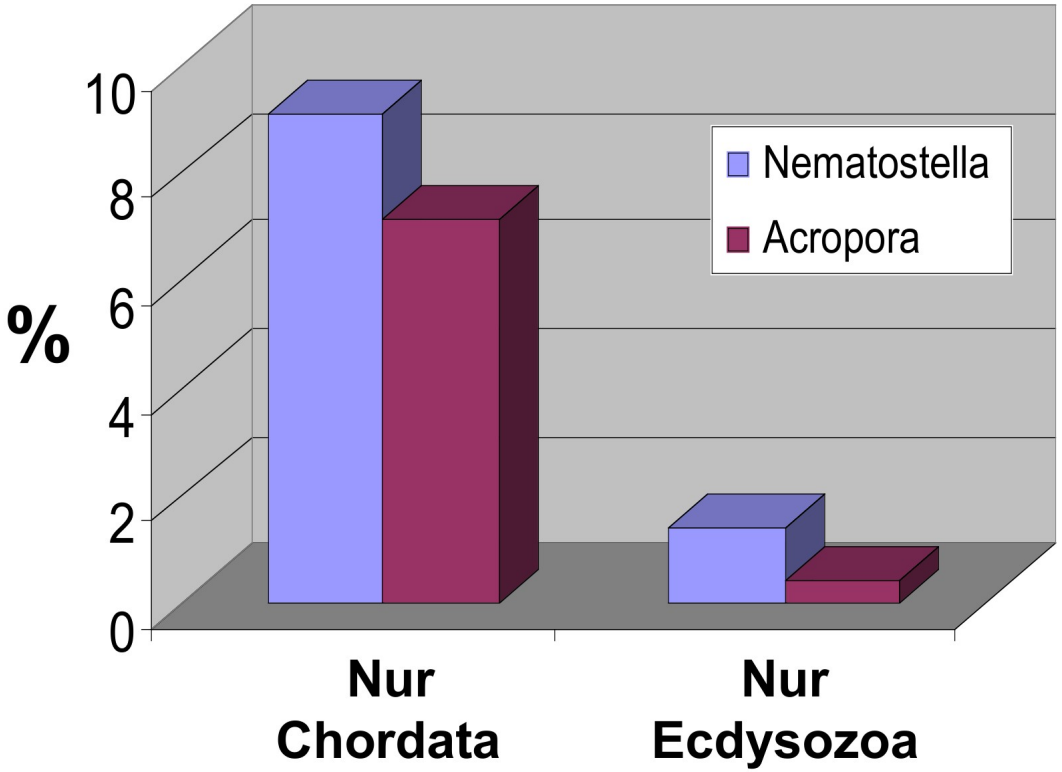
Michael Schubert

Transcriptome-screen in Anthozoa:

Cnidaria have a complex transcriptome and share many genes only with chordates



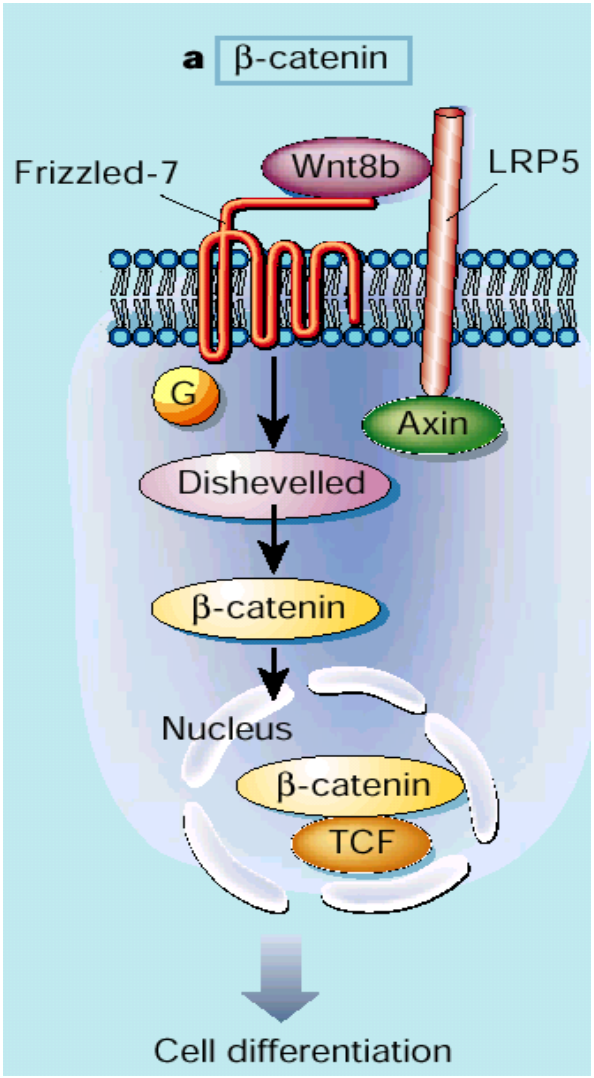
Technau et al., TiG, 2005



Genes shared with anthozoans

→ **Significant gene loss in some Ecdysozoa**

The canonical Wnt signaling pathway: One pathway, many different ligands



Nematostella has 12 out of 13 *Wnt*-Sub-families

Evidence for significant loss of *Wnt*-genes in insects and nematodes

	Wnt1	Wnt2	Wnt3	Wnt4	Wnt5	Wnt6	Wnt7	WntA	Wnt8	Wnt9	Wnt10	Wnt11	Wnt16
Ecdysozoa													
Fly		X	X					X	X		X	X	X
Nematode		X	X			X	X	X	X	X	X	X	
Beetle		X	X	X						X			
Lophotrochozoa													
Polychaete													
Mollusc													
Deuterostomia													
Amphioxus								X					
Human								X					

Kusserow et al., *Nature*, 2005

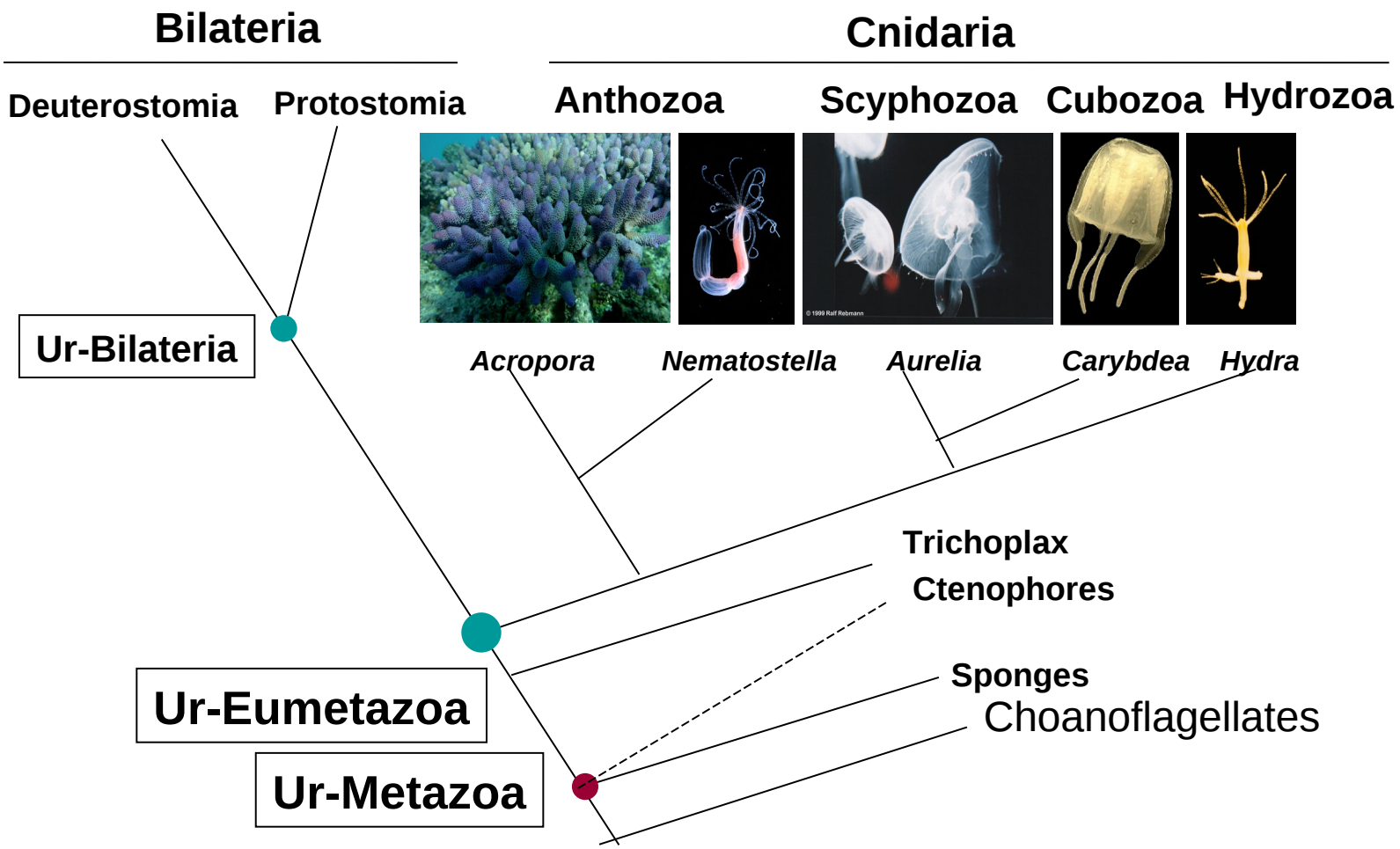
Nematostella has 12 out of 13 *Wnt*-Sub-families

Evidence for significant loss of *Wnt*-genes in insects and nematodes

	Wnt1	Wnt2	Wnt3	Wnt4	Wnt5	Wnt6	Wnt7	WntA	Wnt8	Wnt9	Wnt10	Wnt11	Wnt16
Ecdysozoa													
Fly		X	X					X	X		X	X	X
Nematode		X	X			X	X	X	X	X	X	X	
Beetle		X	X	X						X			
Lophotrochozoa													
Polychaete													
Mollusc													
Deuterostomia													
Amphioxus								X					
Human								X					
Cnidaria													
Ur-Eumetazoa													

Kusserow et al., *Nature*, 2005

From mid 2000's:
Genomes of "Phylogenetically informative species"

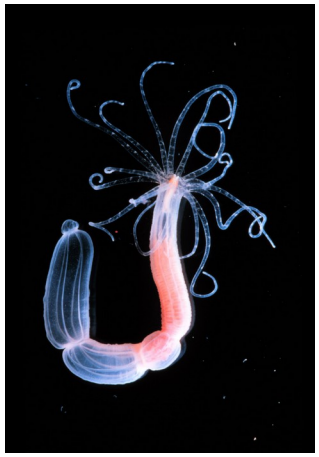


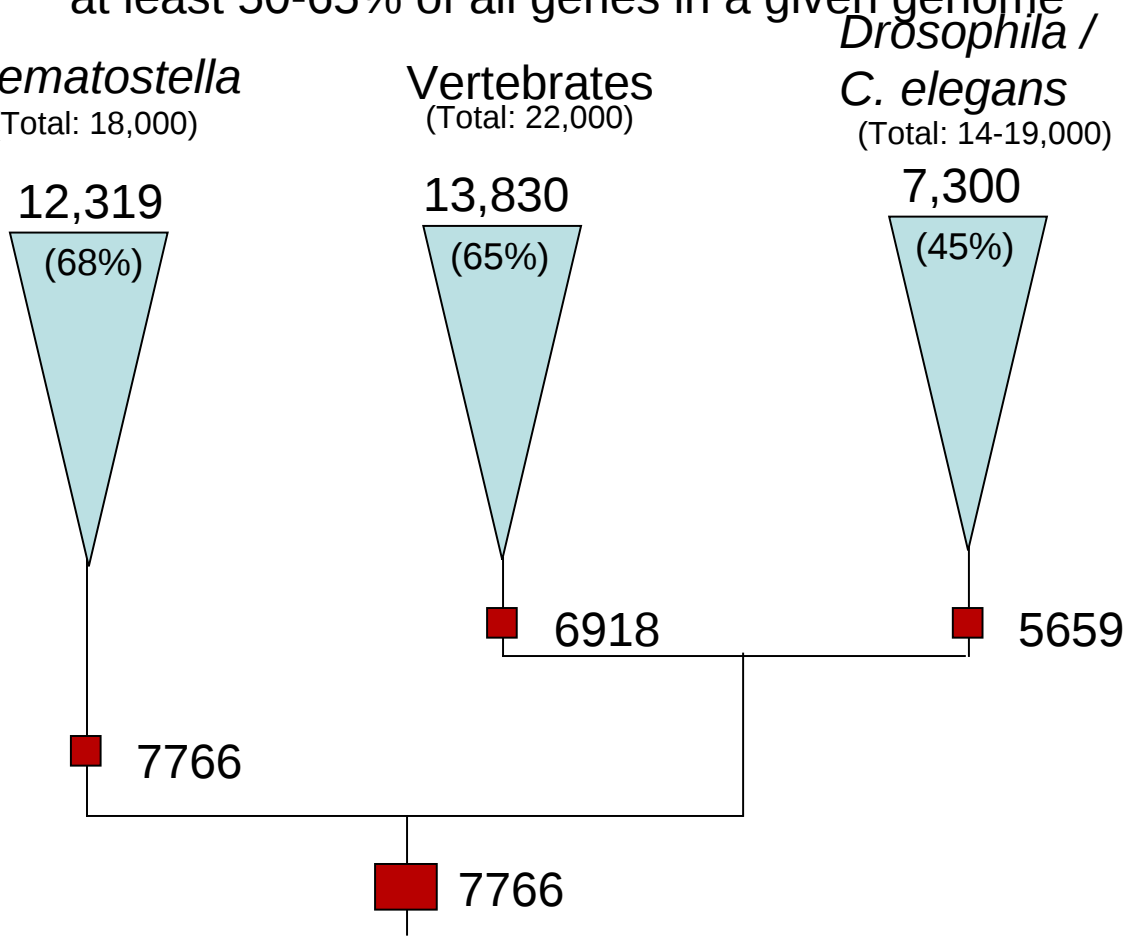
Sequencing: Sanger ~million $ per project
(JGI, JCVI, Broad etc.)

The eumetazoan ancestor contained at least 7k gene families

(orthologs are present in modern day species)

Descendants of the ancestral eumetazoan gene set account for at least 50-65% of all genes in a given genome



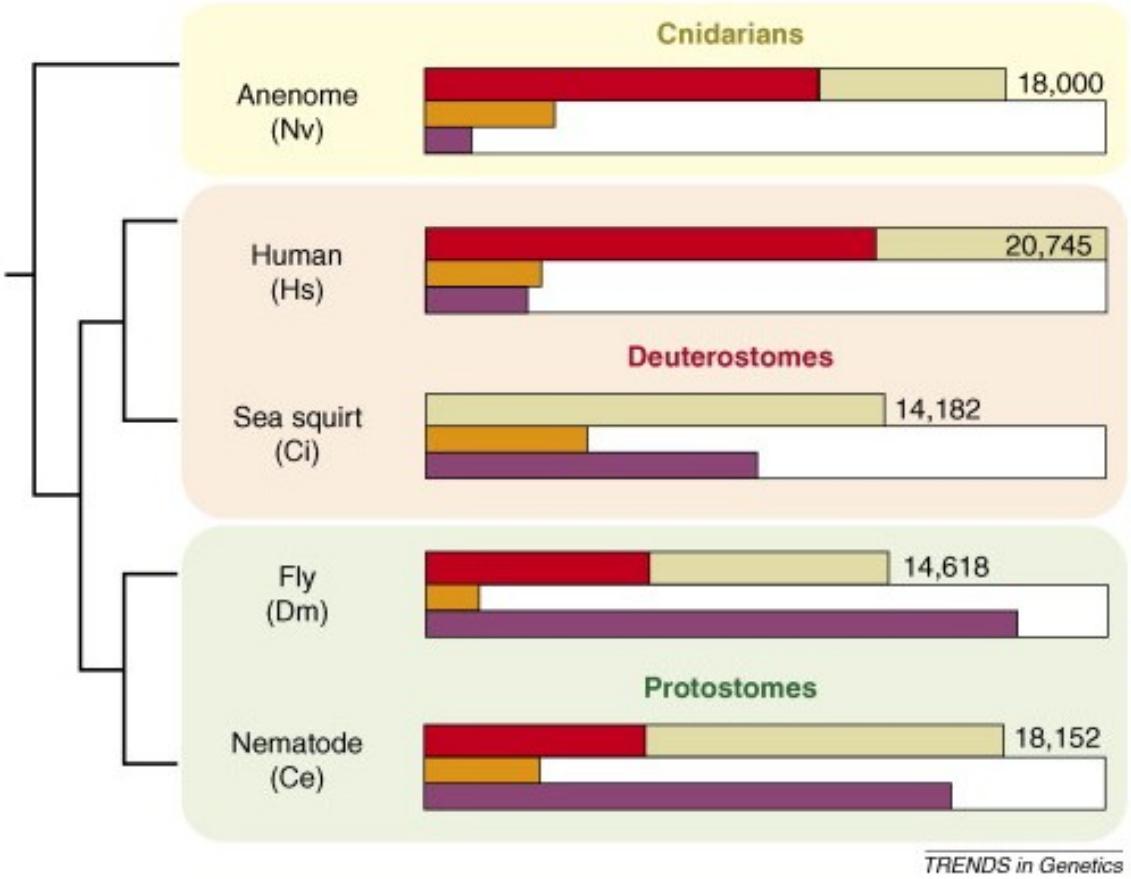


Ancestral Eumetazoan gene set

Putnam et al. Science 2007

Nematostella and humans have preserved 80% of the ancestral introns*

Significant intron losses in model ecdysozoans



Ancestral Introns

Intron gain

Intron loss

* Of 5175 introns in highly conserved protein sequence positions

Miller and Ball, TiG, 2008
Putnam et al., Science, 2007

Intron conservation

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]*

Sea anemones are seemingly primitive animals that, along with corals, jellyfish, and hydras, constitute the oldest eumetazoan phylum, the Cnidaria. Here, we report a comparative analysis of the draft genome of an emerging cnidarian model, the starlet sea anemone *Nematostella vectensis*. The sea anemone genome is complex, with a gene repertoire, exon-intron structure, and large-scale gene linkage more similar to vertebrates than to flies or nematodes, implying that the genome of the eumetazoan ancestor was similarly complex. Nearly one-fifth of the inferred genes of the ancestor are eumetazoan novelties, which are enriched for animal functions like cell signaling, adhesion, and synaptic transmission. Analysis of diverse pathways suggests that these gene "inventions" along the lineage leading to animals were likely already well integrated with preexisting eukaryotic genes in the eumetazoan progenitor.

Gene order/colocalization:
Synteny conservation between chromosomes

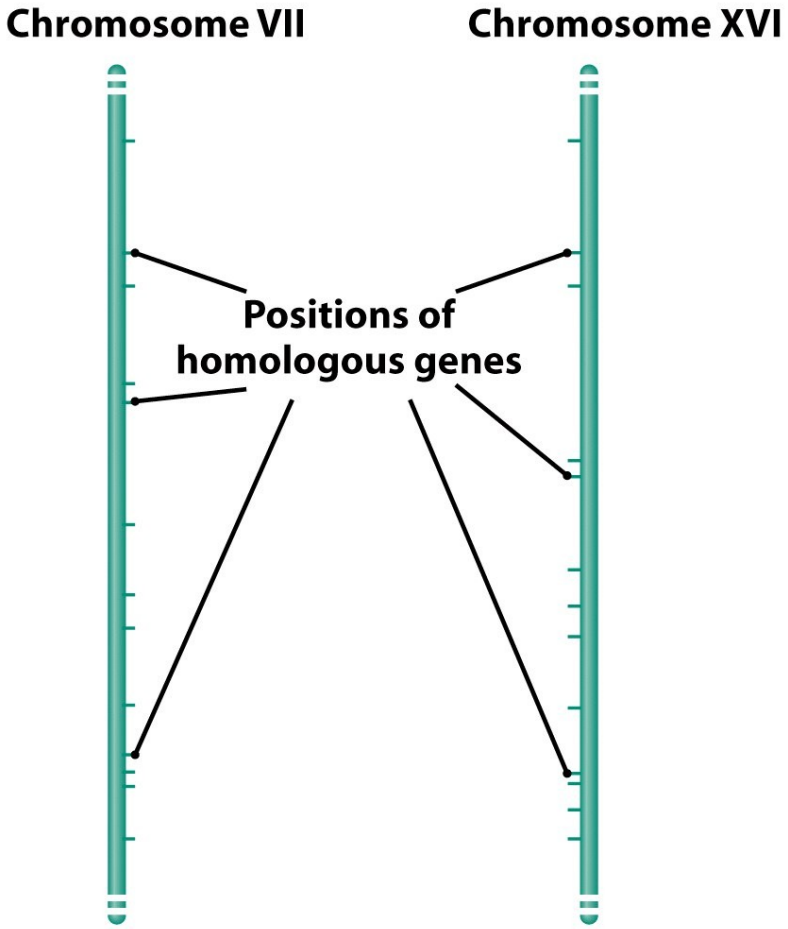
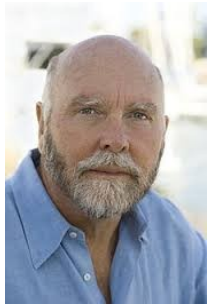
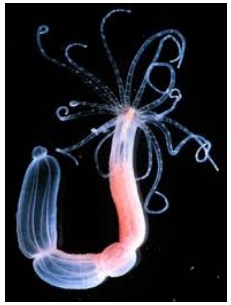
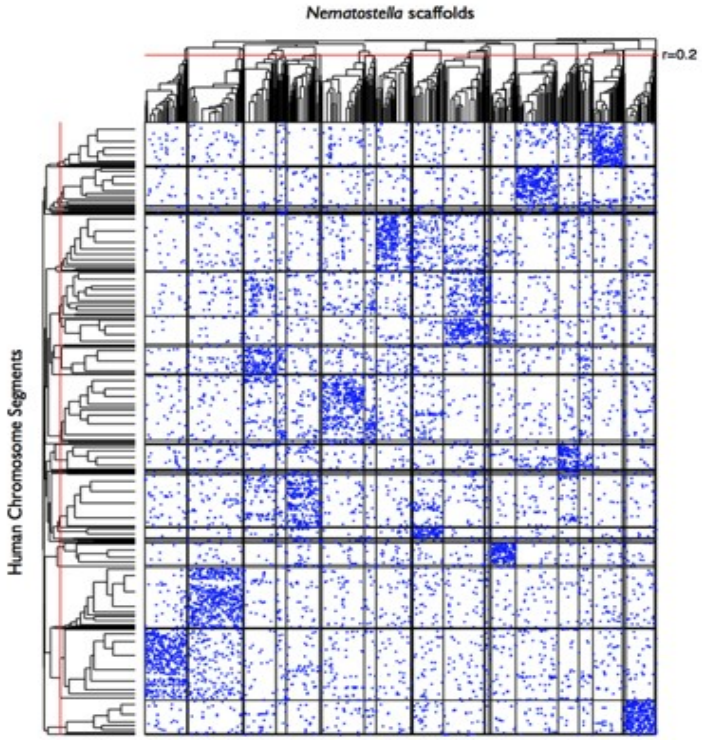


Figure 18.15 *Genomes 3* (© Garland Science 2007)

“Bags of genes”: Syntenic signal (gene colocalization in the same ‘territories’, chromosomes?)



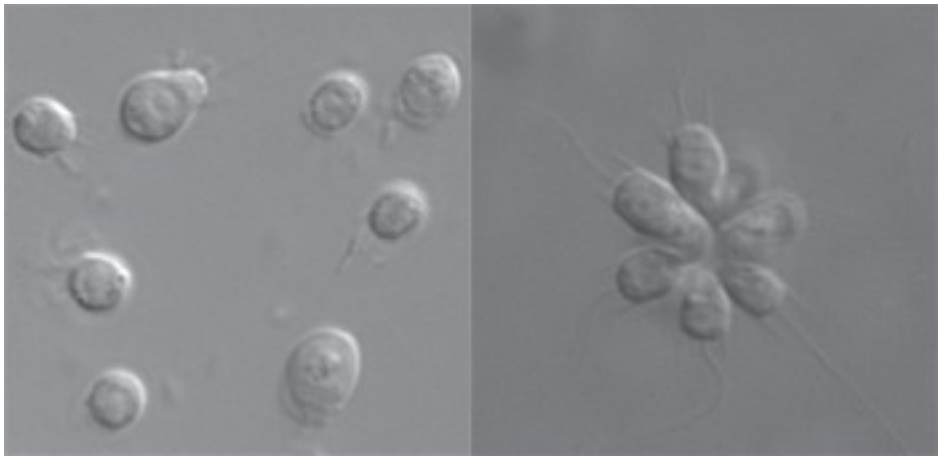




→ Large gene family complement
→ Complex gene structure
→ Preservation of synteny

What about unicellular to metazoan transition?

Our closest unicellular relatives



Nicole King

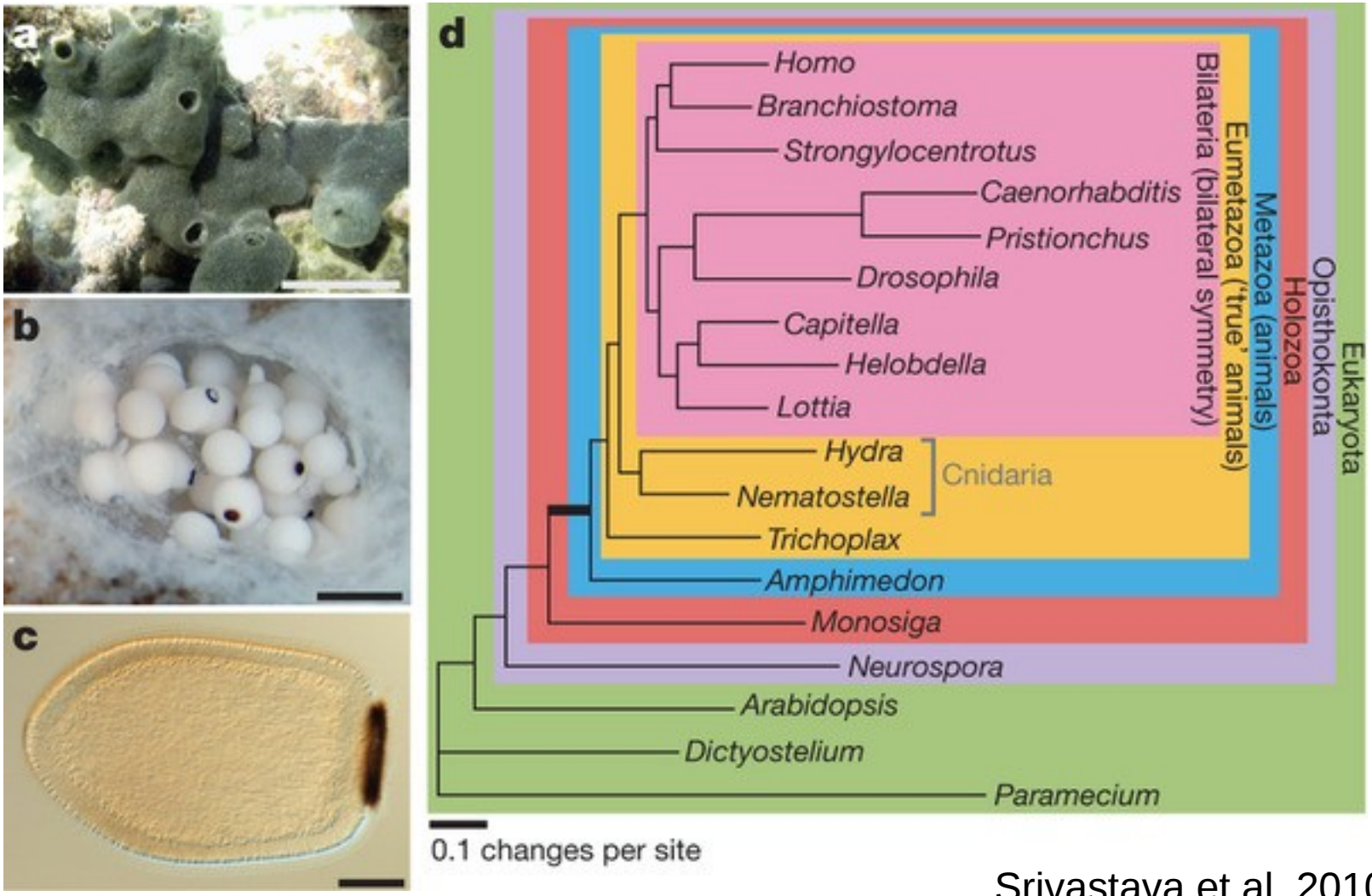
Uniconta: Monosiga sp.

DSCC

Sponge

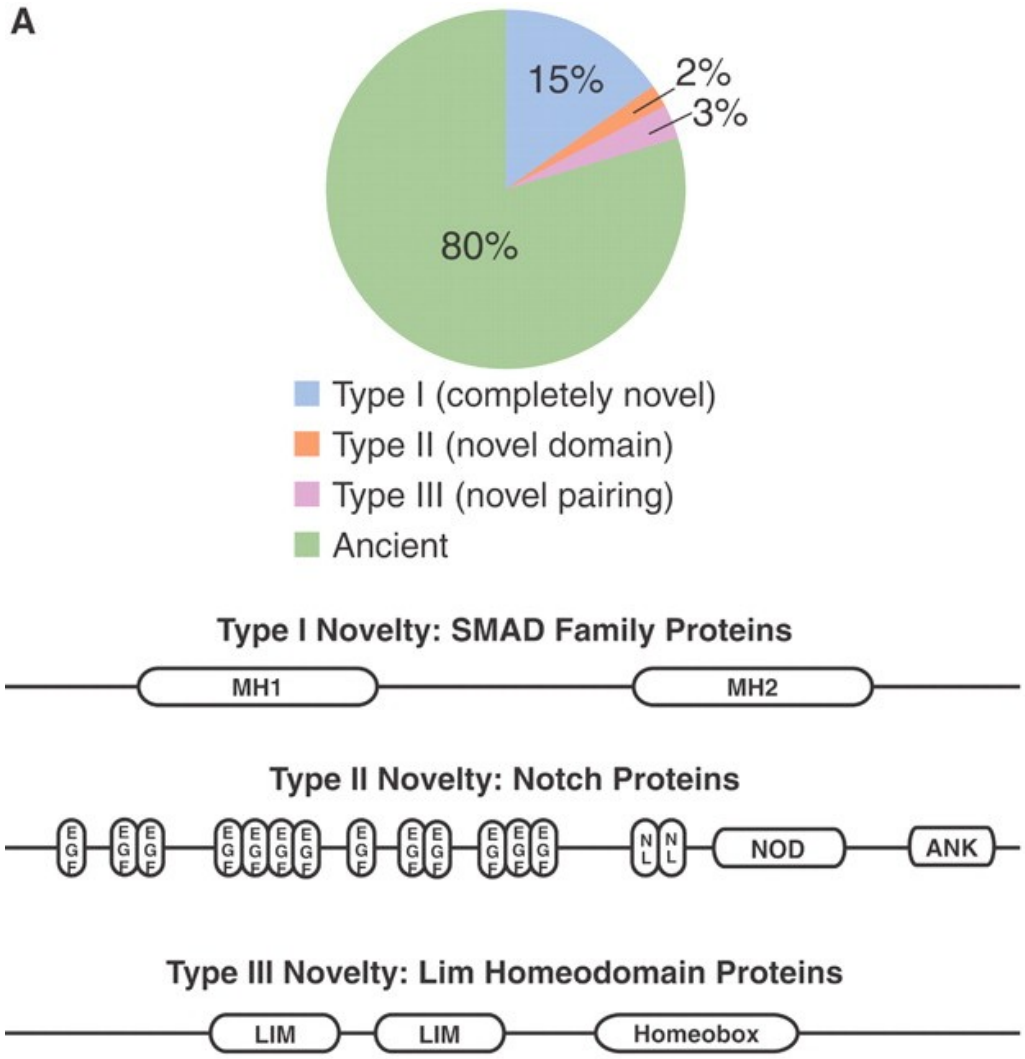
-> cell adhesion evolution
-> cell signaling

Unicont position on the phylogenetic tree

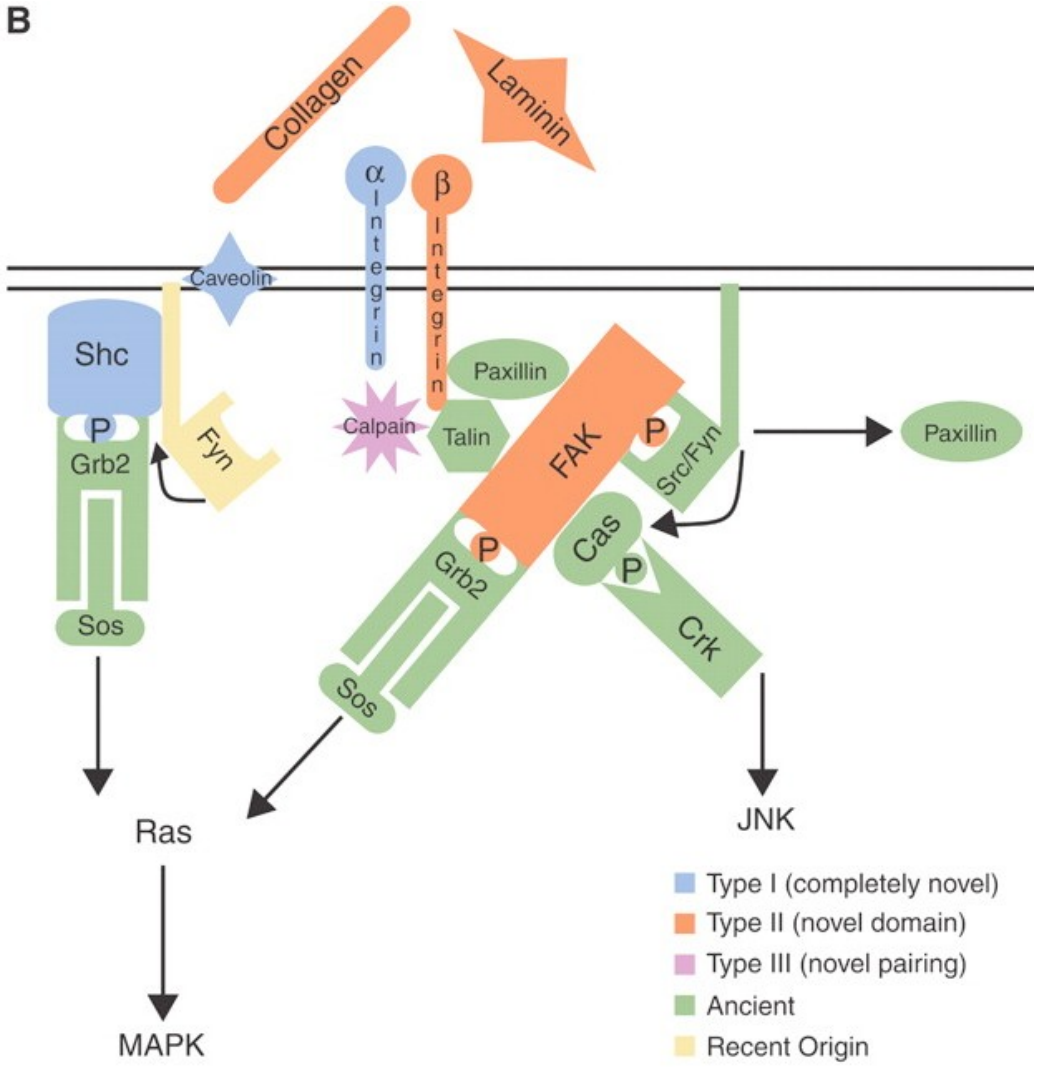


Srivastava et al, 2010

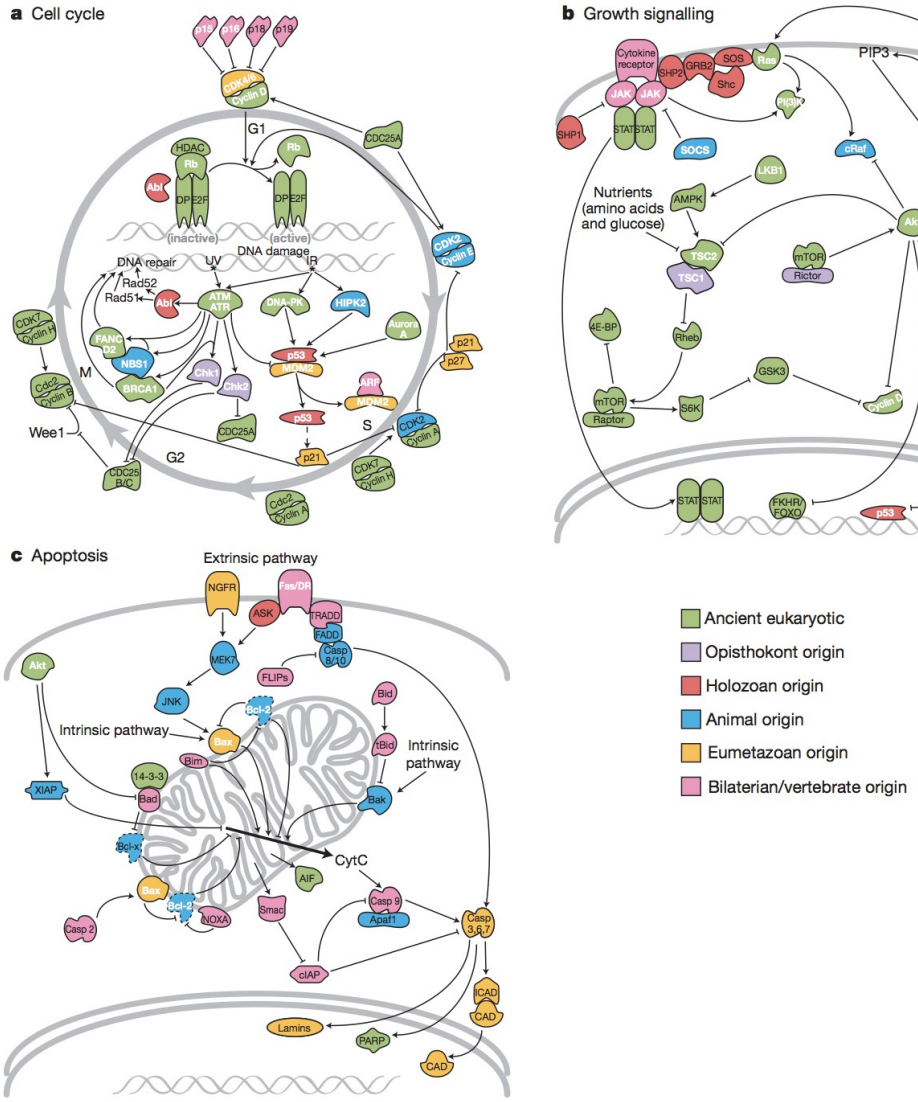
Nature of novel genes in metazoan evolution



Integration of novel proteins into existing protein networks

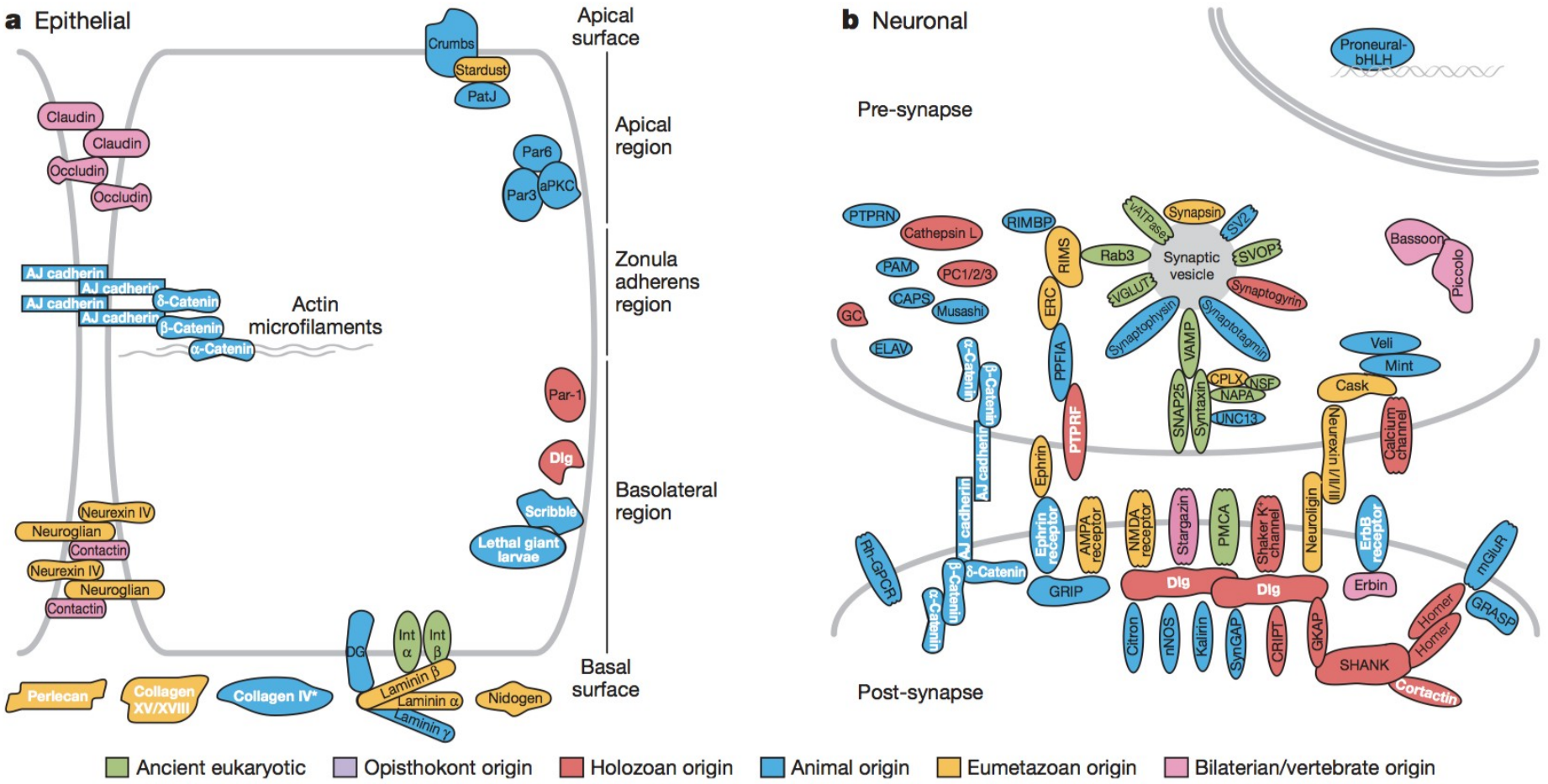


Gene novelties and evolutionary transitions



Srivastava,et al, Nature 2010

Cell signaling



Ancient eukaryotic | Opisthokont origin | Holozoan origin | Animal origin | Eumetazoan origin | Bilaterian/vertebrate origin

-> gene novelties contributed significantly to early metazoan evolution

Quick recap:

The metazoan ancestor had a fairly complex gene repertoire and genomic organisation - “**ancestral complexity**”

This complexity is observed at ALL scales of genomic organization: gene structure (introns), gene presence, gene synteny

Gene loss and genome compaction dominated in some model species

No single species represents the ancestral state – species sampling and comparison are important!

Origin of multicellularity can be linked to several key innovations (and gene novelties) in cell-cell interactions (signaling, adhesion etc)

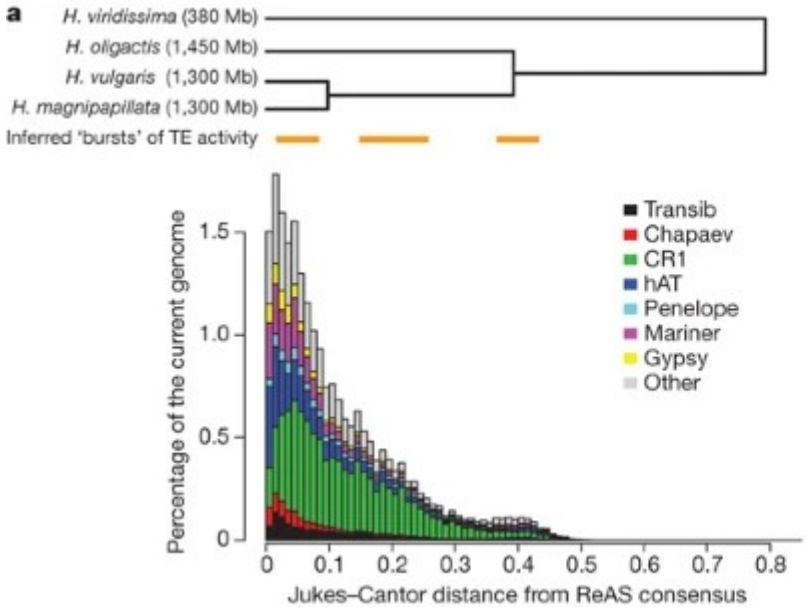
Beyond “woa, things were really complex back then”

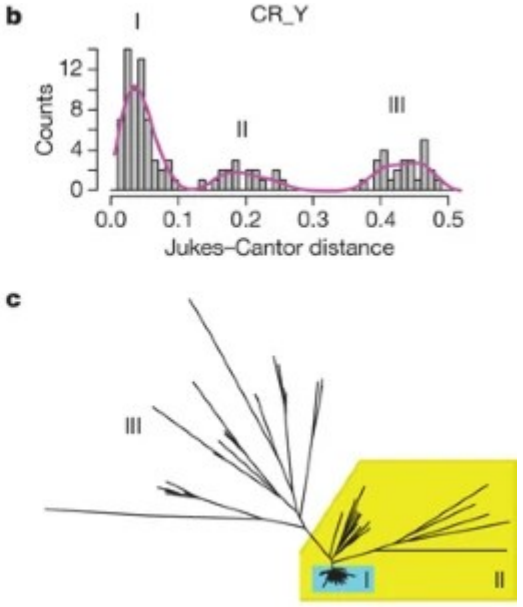
Predictions?

- Conserved gene families and especially transcription factors must have function (→ evo-devo)
- Synteny preservation beyond the observed contigs or scaffolds
- Unclear what the actual ‘innovation’ driver is in the genome? Gene novelty, duplication, loss, non-coding element evolution etc?

The value of the “second” genome?

The dynamic genome? How dynamic can genomes be??
Do repeats break ancient “architecture”?



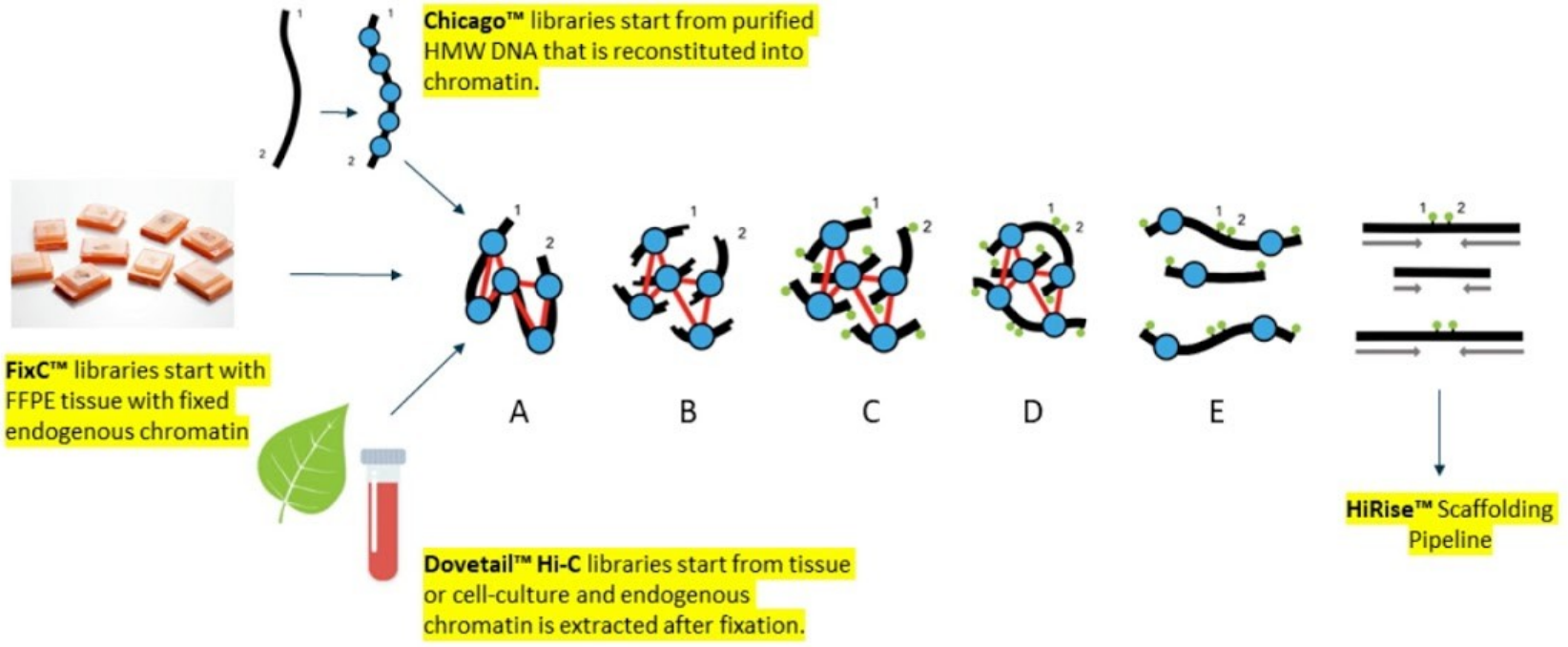


Chapman et al, 2010

II. Current approaches and bottlenecks (from chromosomes to 3D structure?)

Genome scaffolding with HiC (chromosomal conformational capture derivative)

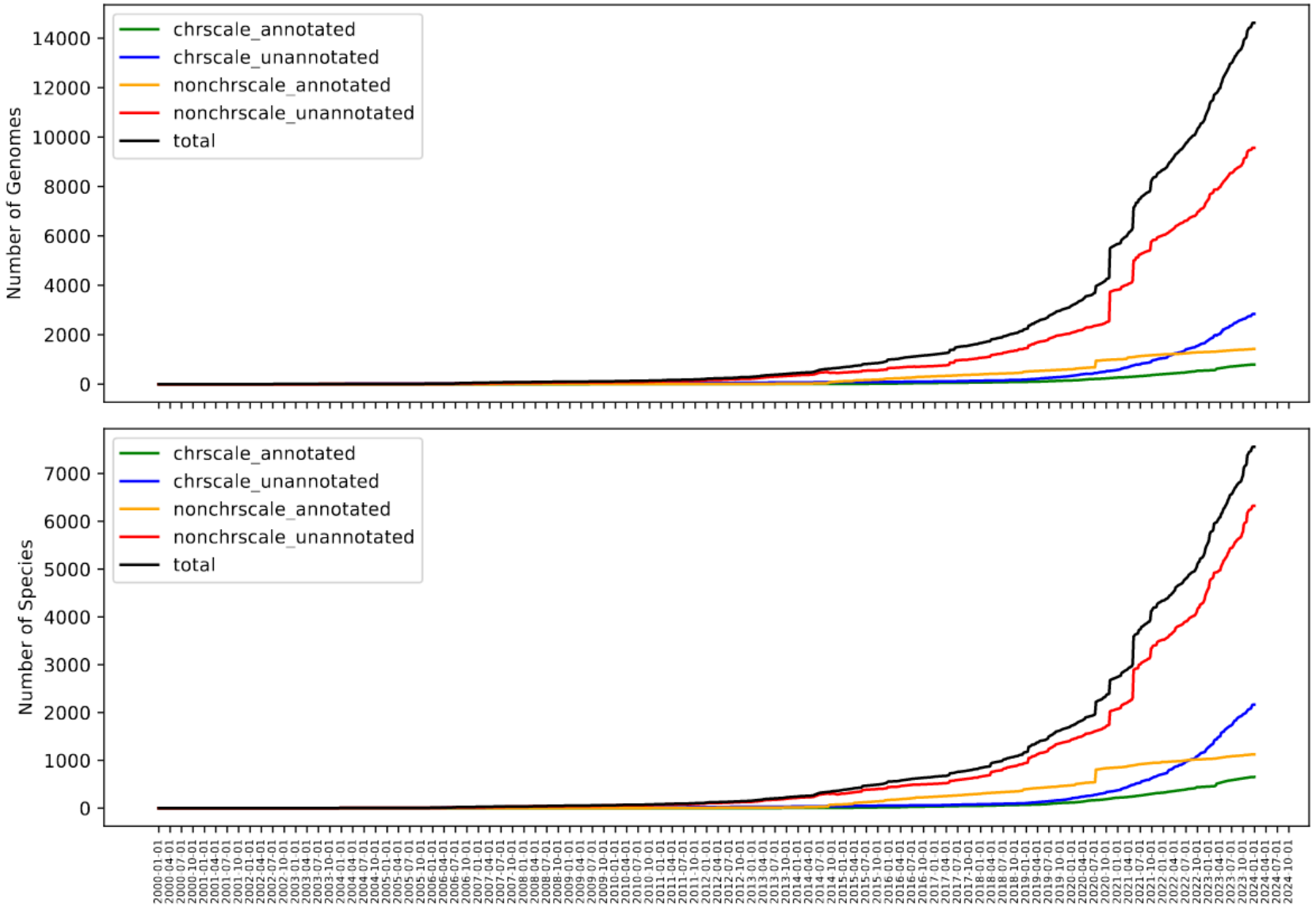
Methods:
Proximity Ligation Approaches



PLANT/ANIMAL

The (recent) avalanche of chromosome-scale genome depositions

NCBI:



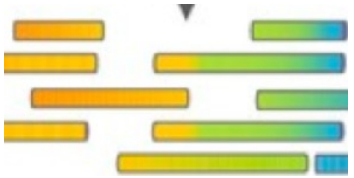
Darrin Schultz

GENOME 10K

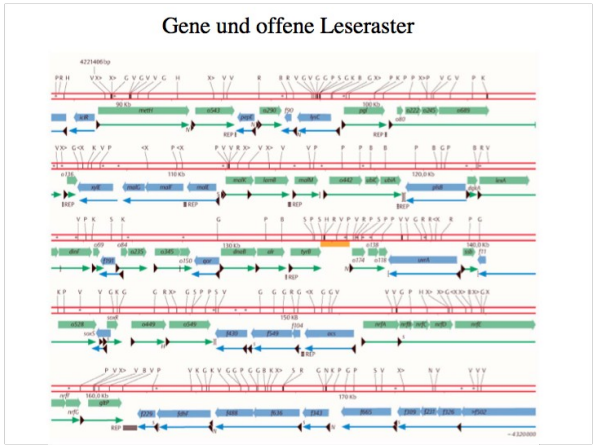
ERGA
EUROPEAN REFERENCE GENOME ATLAS

GORDON AND BETTY MOORE FOUNDATION
wellcome sanger institute
EMBL-EBI
AQUATIC SYMBIOSIS GENOMICS

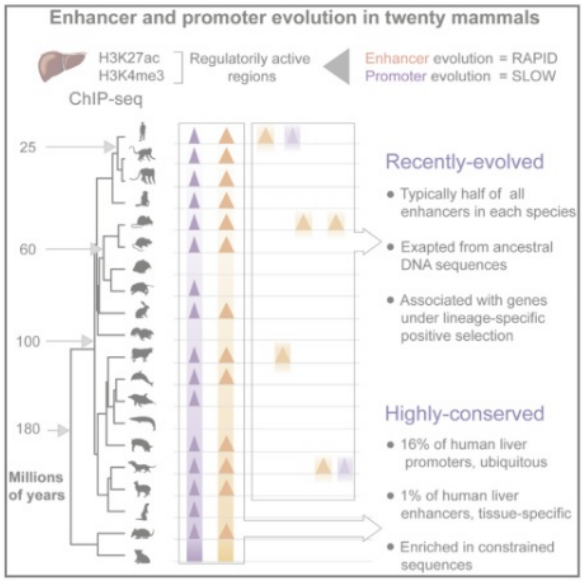
What was a genome in 2010's?



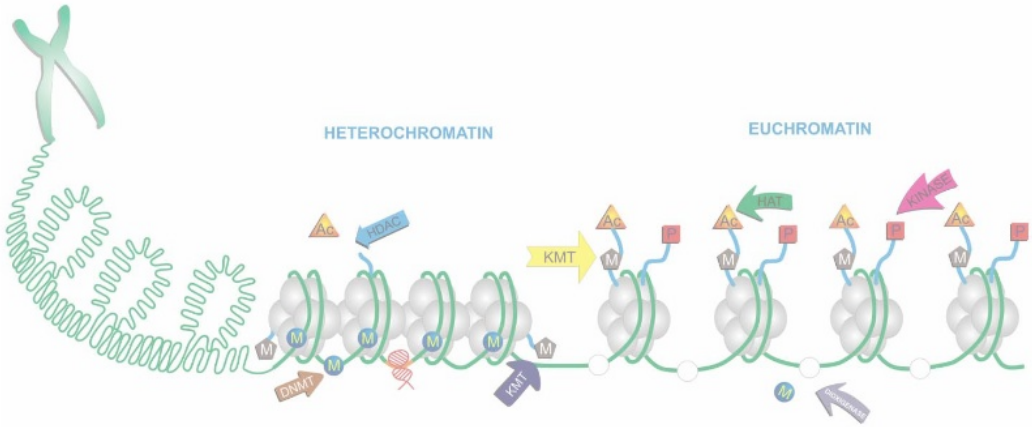
Gene annotation



Regulatory regions



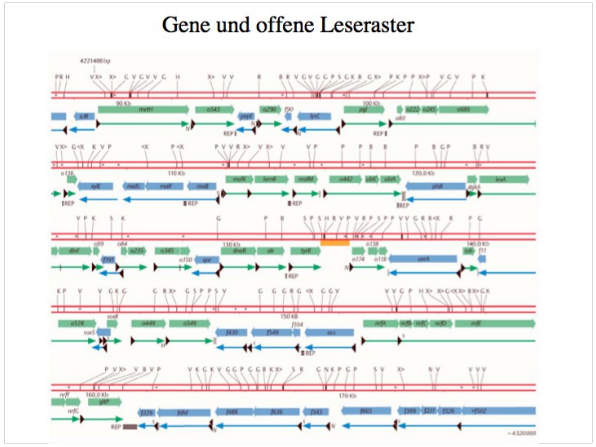
Epigenetic modification



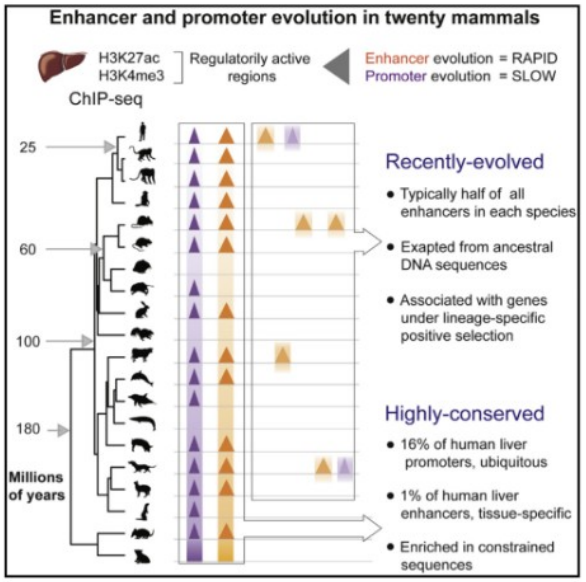
Villar et al, Cell 2015; Villota-Salazar et al, 2016; http://advancedtranscription16.blogspot.co.at/p/overview.html

What is a genome in 2020's?

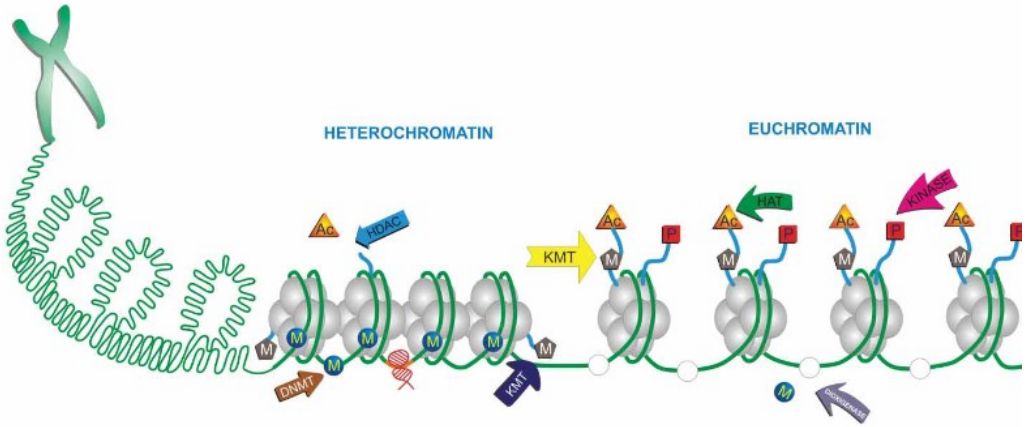
Gene annotation



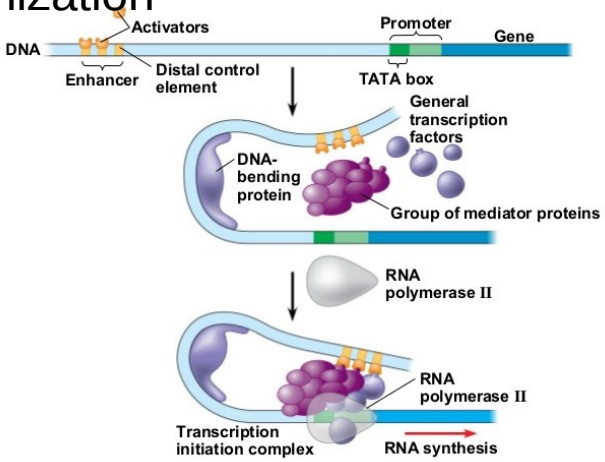
Regulatory regions



Epigenetic modification



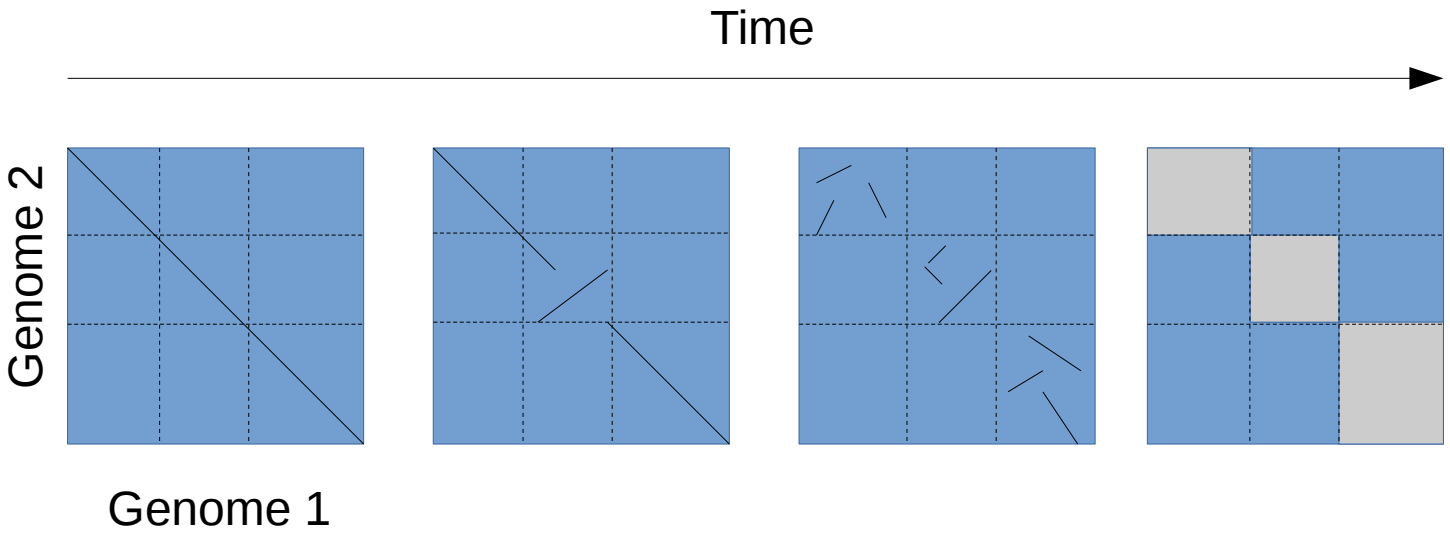
3D organization



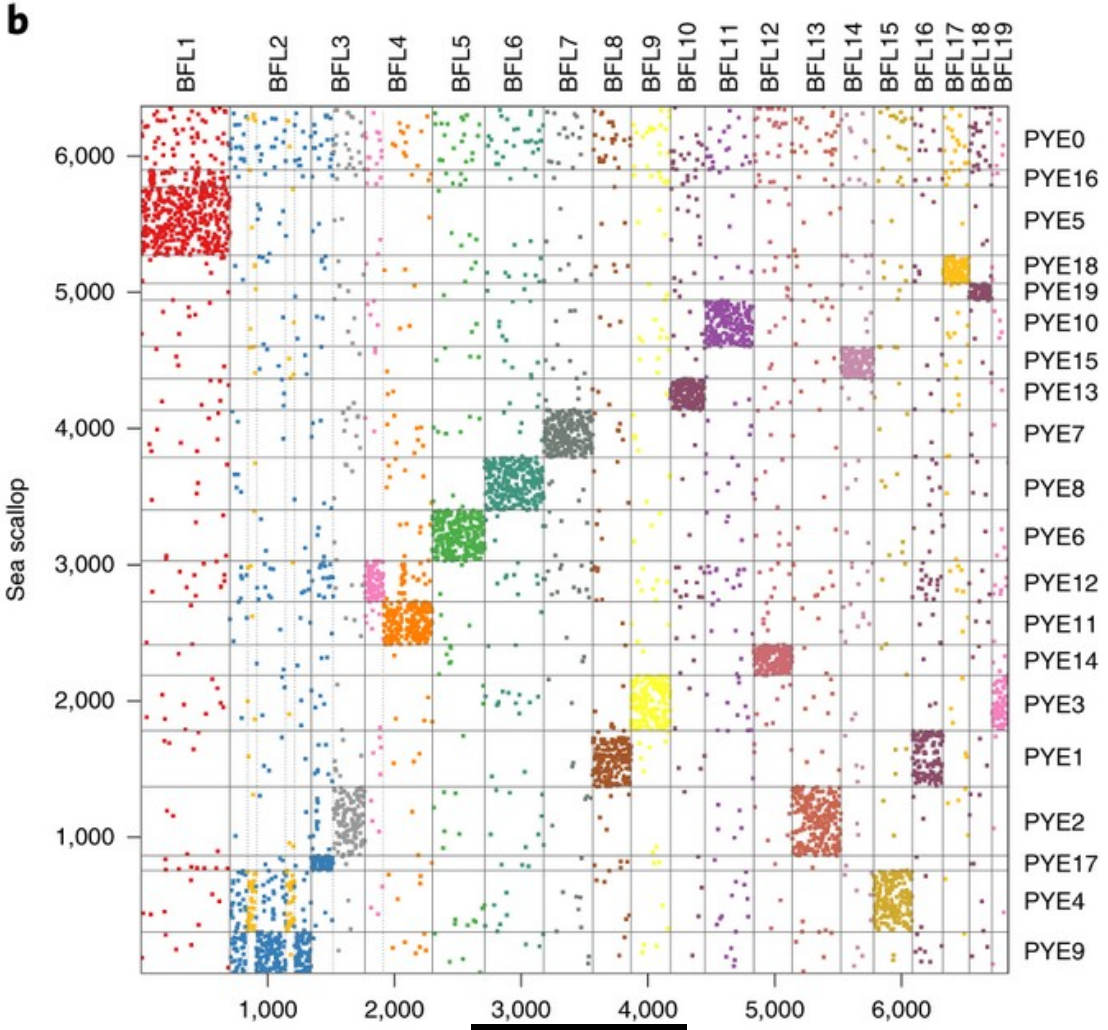
Villar et al, Cell 2015; Villota-Salazar et al, 2016; http://advancedtranscription16.blogspot.co.at/p/overview.html

- Which of the pre-chromosomal assembly age hypotheses can be verified now?

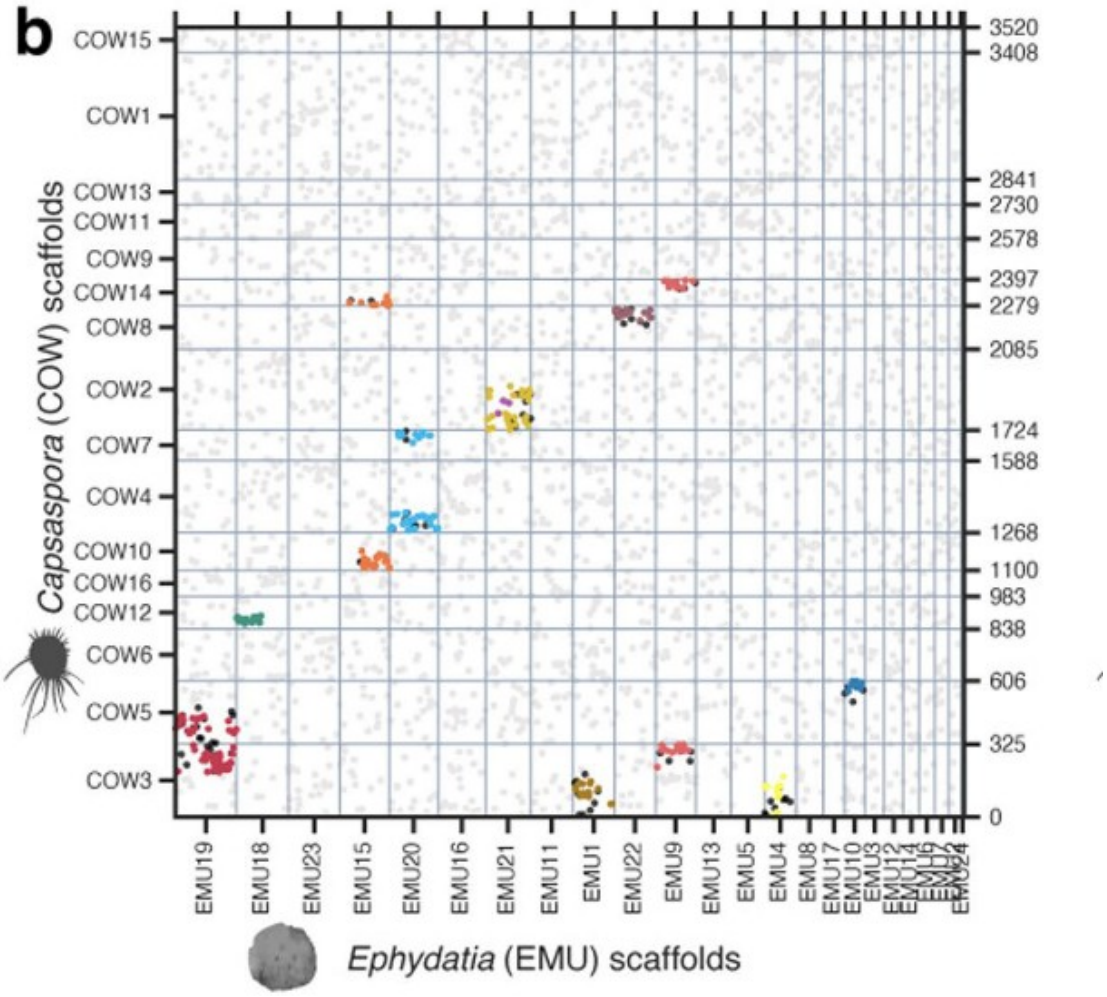
Macro-syntenic patterns in metazoan genomes



Most (invertebrate) animal chromosomes are ‘ancient’ linkage units

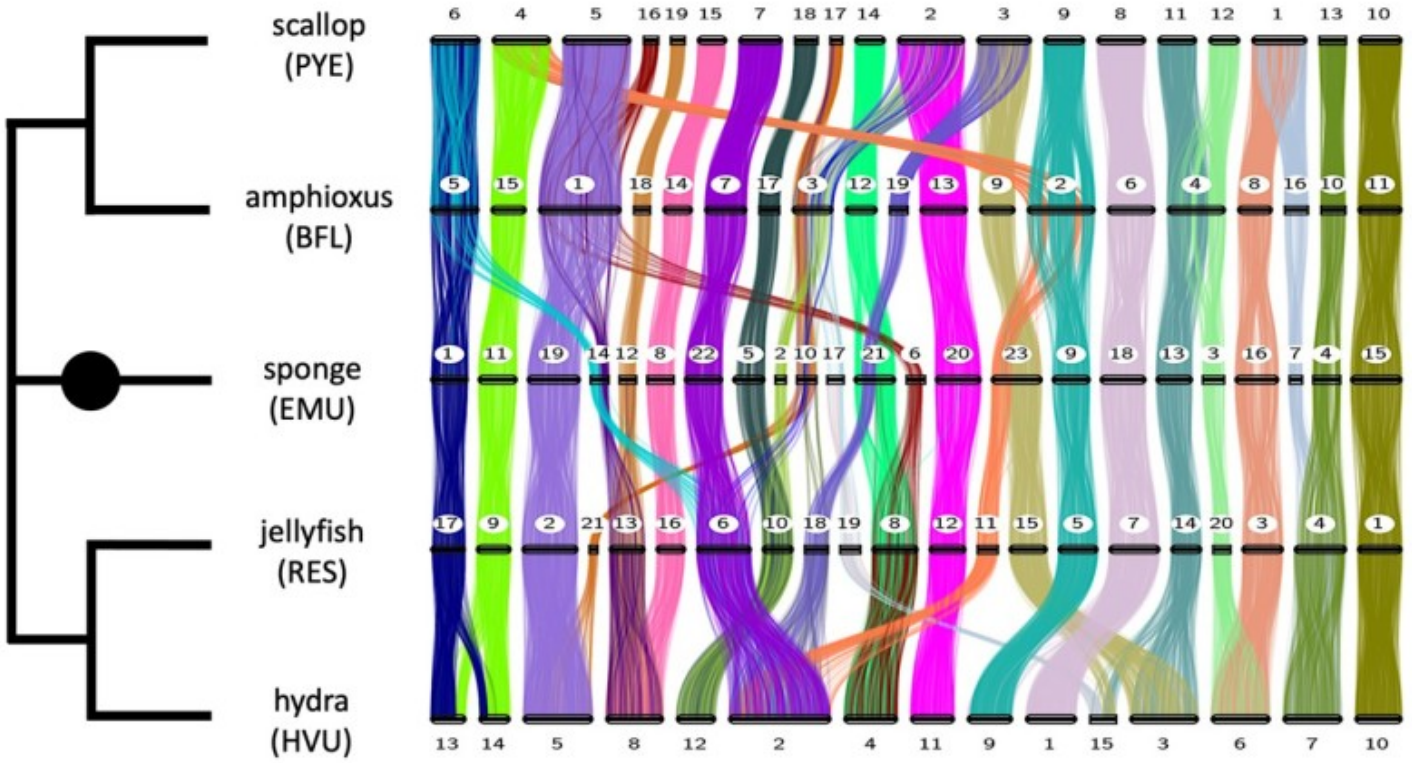


With homologies ranging up to their unicellular relatives



1billion+ years

Schultz et al, Nature 2023



Simakov et al, 2022
Data from:
Wang et al, 2017
Li et al, 2020
Kenny et al, 2020
Simakov et al, 2020

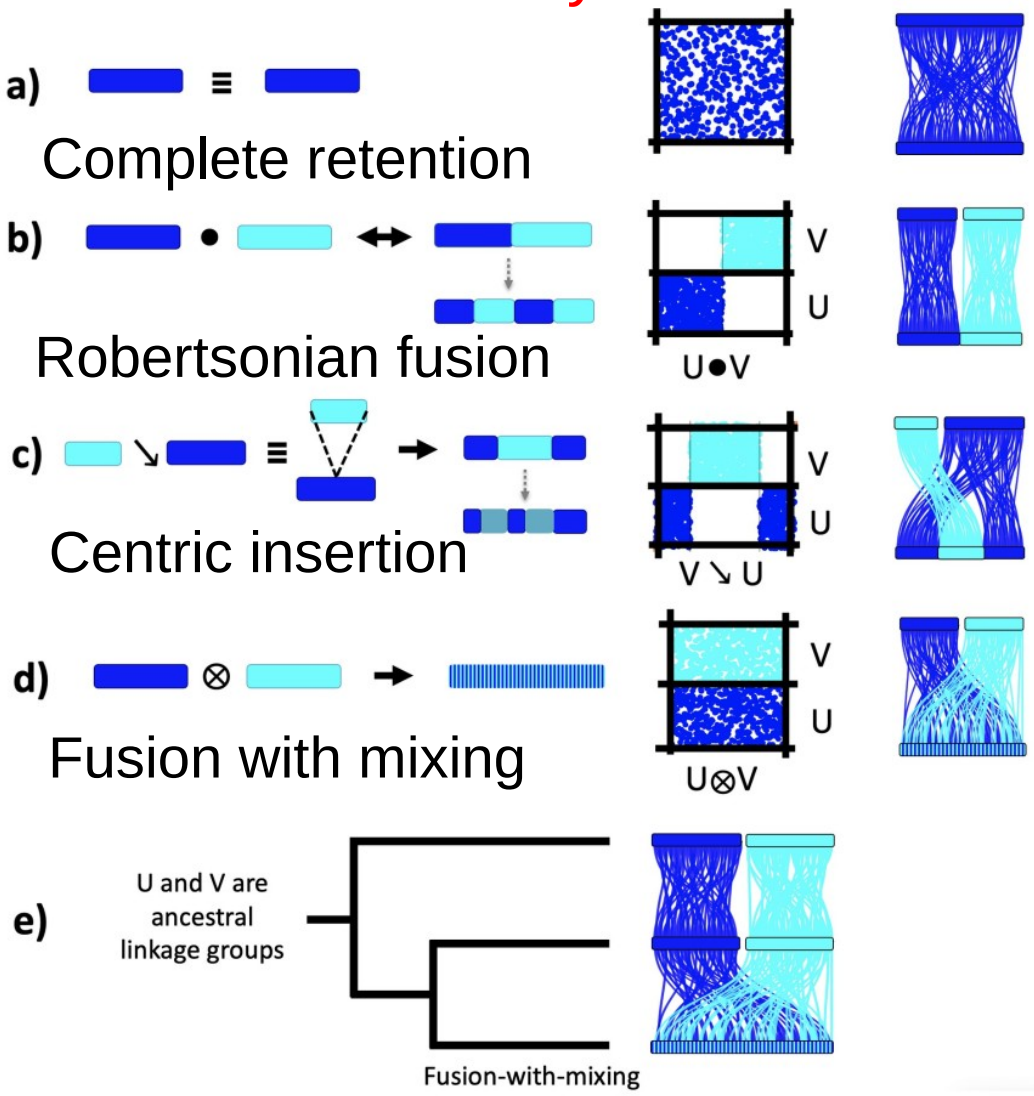
Chromosomes (almost) never break, but they do “add up”
What are the evolutionarily independent chromosomal elements?

Chromosomal elements representation of modern day genomes

Dan Rokhsar

ALG	amphioxus	scallop	bilaterian ancestor (24)	sponge	cnidarian ancestor (21)	jellyfish	hydra	no. of shared orthologs
G	BFL11	PYE10	G	EMU15	G	RES1	HVU10	137
B1	BFL10	PYE13	B1	EMU04	B1⊗B2	RES4	HVU7	100
B2	BFL16	PYE1	B2	EMU07/24				88
M	BFL8	PYE1	M	EMU16	M	RES3	HVU6	102
C2	BFL3L	PYE17	C2	EMU10	C2	RES21	HVU5	20
A1a	BFL1	PYE5	A1a⊗A1b	EMU19	A1a	RES2	HVU5	209
A1b				EMU14R	A1b⊗B3	RES13	HVU8	52
B3	BFL18	PYE19	B3	EMU12				47
P	BFL14	PYE15	P	EMU08	P	RES16	HVU8	76
L	BFL15	PYE4	L	EMU11	L	RES9	HVU14R	72
L_							HVU13R	31
Ea	BFL5	PYE6	Ea⊗Eb	EMU01	Ea	RES17	HVU13L	87
Ea_							HVU14L	26
Eb				EMU14L	Eb⊗F⊗Qb	RES6	HVU2	45
F	BFL7	PYE7	F	EMU22				142
Qb	BFL3R	PYE2	Qb⊗Qa	EMU10				12
Qa				EMU02	Qa⊗J1	RES10	HVU12	31
J1	BFL17	PYE18	J1	EMU05				55
R	--	PYE12	R	EMU17	R⊗Qc	RES19	HVU15L	32
Qc	BFL3R	PYE2	Qc⊗Qd					13
Qd				EMU10	Qd⊗O2	RES18	HVU2	21
O2	BFL19	PYE3	O2	EMU21mid				46
N	BFL12	PYE14	N	EMU21end	A2⊗N	RES8	HVU4	106
A2	BFL1	PYE16	A2	EMU06				41
H	BFL13	PYE2	H	EMU20	H	RES12	HVU11	136
J2	BFL2eve	PYE4	J2	EMU23L	J2	RES11	HVU2	64
C1	BFL2odd	PYE9	C1	EMU09	C1	RES5	HVU9	142
D	BFL6	PYE8	D	EMU18	D	RES7	HVU1	158
D_							HVU9t	14
K	BFL9	PYE3	K	EMU23R	K	RES15	HVU3	100
K_							HVU15R	17
I	BFL4R	PYE11	I	EMU13	I	RES14	HVU3	79
I_							HVU15R	14
O1	BFL4L	PYE12	O1	EMU03	O1	RES20	HVU6	54

Fundamental processes in chromosomal evolution and their macrosyntenic outcomes

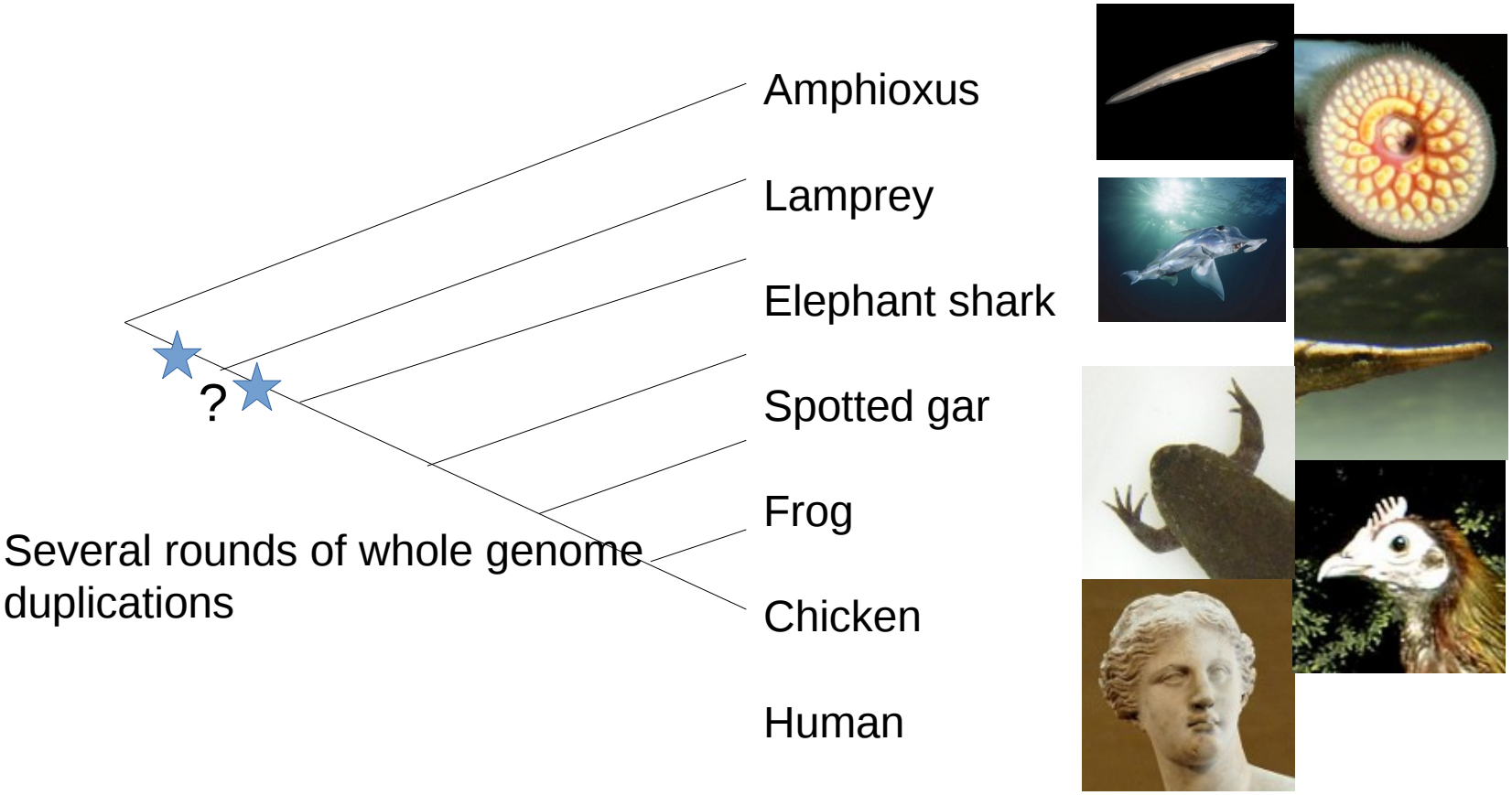


Algebra representation of

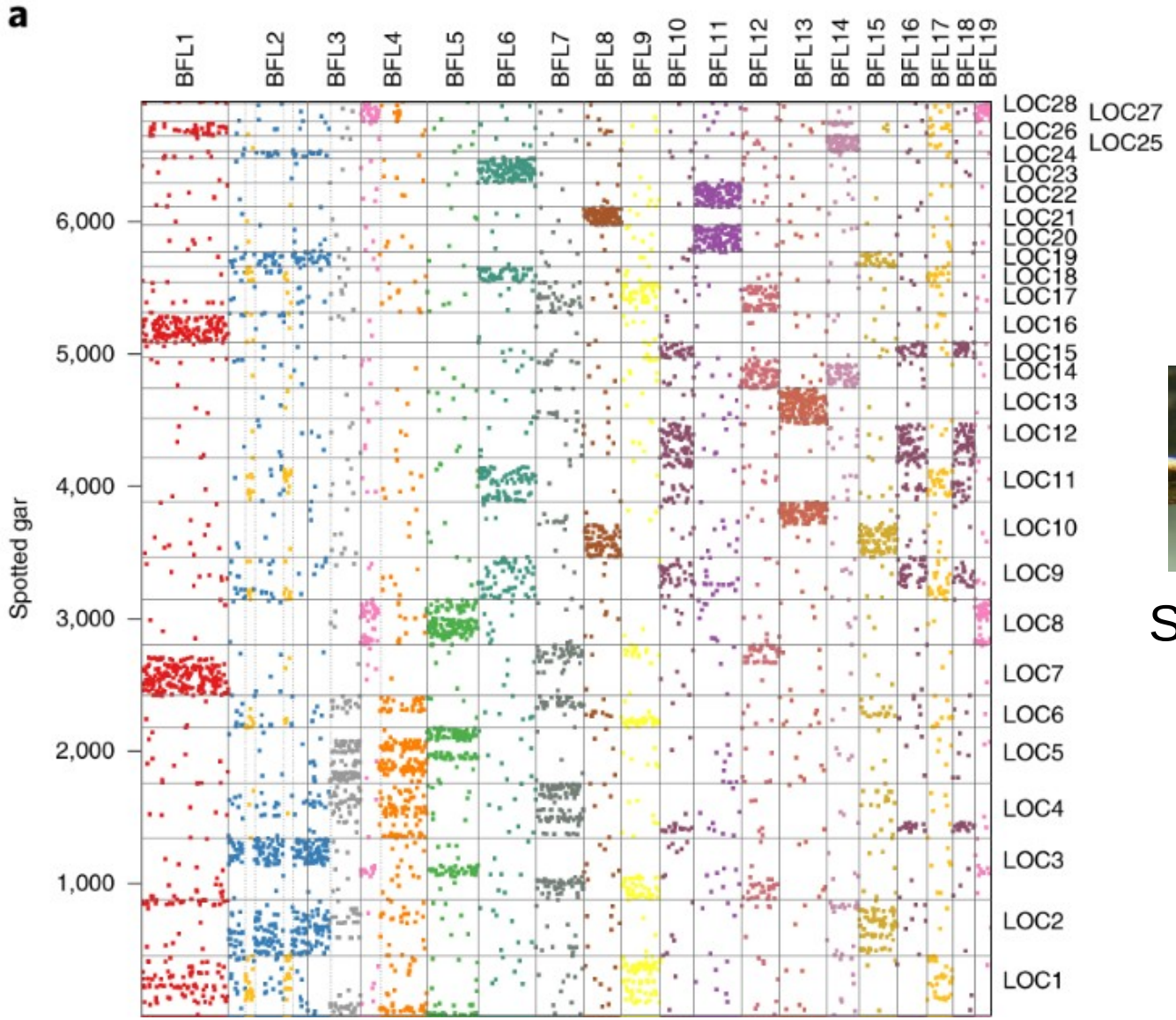
$$\mathrm{BFL1}={}^{41}A2 \otimes {}^{258}A1$$
$$\mathrm{BFL2}={}^{142}C1 \searrow {}^{66}J2$$
$$\mathrm{BFL3}={}^{18}C2 \bullet {}^{78}Q$$
$$\mathrm{BFL4}={}^{90}I \bullet {}^{55}O1$$
$$\mathrm{BFL5}={}^{162}E$$
$$\mathrm{BFL6}={}^{172}D$$
$$\mathrm{BFL7}={}^{145}F$$
$$\mathrm{BFL8}={}^{102}M$$
$$\mathrm{BFL9}={}^{119}K$$
$$\mathrm{BFL10}={}^{95}B1$$
$$\mathrm{BFL11}={}^{138}G$$
$$\mathrm{BFL12}={}^{107}N$$
$$\mathrm{BFL13}={}^{135}H$$
$$\mathrm{BFL14}={}^{78}P$$
$$\mathrm{BFL15}={}^{104}L$$
$$\mathrm{BFL16}={}^{86}B2$$
$$\mathrm{BFL17}={}^{54}J1$$
$$\mathrm{BFL18}={}^{46}B3$$
$$\mathrm{BFL19}={}^{46}O2$$

ALG	amphioxus	scallop	bilaterian ancestor (24)	sponge	cnidarian ancestor (21)	jellyfish	hydra
G	BFL11	PYE10	G	EMU15	G	RES1	HVU10
B1	BFL10	PYE13	B1	EMU04	B1⊗B2	RES4	HVU7
B2	BFL16	PYE1	B2	EMU07/24			
M	BFL8	PYE1	M	EMU16	M	RES3	HVU6
C2	BFL3L	PYE17	C2	EMU10	C2	RES21	HVU5
A1a	BFL1	PYE5	A1a⊗A1b	EMU19	A1a	RES2	HVU5
A1b				EMU14R	A1b⊗B3	RES13	HVU8
B3	BFL18	PYE19	B3	EMU12			
P	BFL14	PYE15	P	EMU08	P	RES16	HVU8
L	BFL15	PYE4	L	EMU11	L	RES9	HVU14R
L_							HVU13R
Ea	BFL5	PYE6	Ea⊗Eb	EMU01	Ea	RES17	HVU13L
Ea_							HVU14L
Eb				EMU14L	Eb⊗F⊗Qb	RES6	HVU2
F	BFL7	PYE7	F	EMU22			
Qb	BFL3R	PYE2	Qb⊗Qa	EMU10			
Qa				EMU02	Qa⊗J1	RES10	HVU12
J1	BFL17	PYE18	J1	EMU05			
R	--	PYE12	R	EMU17	R⊗Qc	RES19	HVU15L
Qc	BFL3R	PYE2	Qc⊗Qd				
Qd				EMU10	Qd⊗O2	RES18	HVU2
O2	BFL19	PYE3	O2	EMU21mid			
N	BFL12	PYE14	N	EMU21end	A2⊗N	RES8	HVU4
A2	BFL1	PYE16	A2	EMU06			
H	BFL13	PYE2	H	EMU20	H	RES12	HVU11
J2	BFL2eve	PYE4	J2	EMU23L	J2	RES11	HVU2
C1	BFL2odd	PYE9	C1	EMU09	C1	RES5	HVU9
D	BFL6	PYE8	D	EMU18	D	RES7	HVU1
D_							HVU9t
K	BFL9	PYE3	K	EMU23R	K	RES15	HVU3
K_							HVU15R
I	BFL4R	PYE11	I	EMU13	I	RES14	HVU3
I_							HVU15R
O1	BFL4L	PYE12	O1	EMU03	O1	RES20	HVU6

Modification of the ancestral(-ish) karyotype



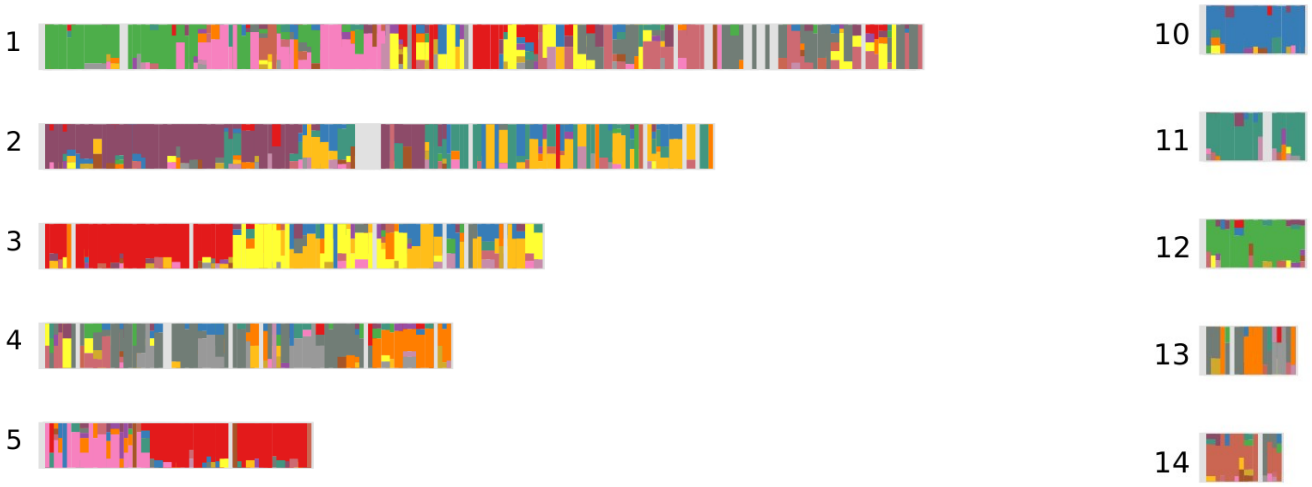
Ancestral linkage groups were largely retained (and multiplied) in vertebrates, however...



Spotted gar

Simakov *et al*, NEE 2020

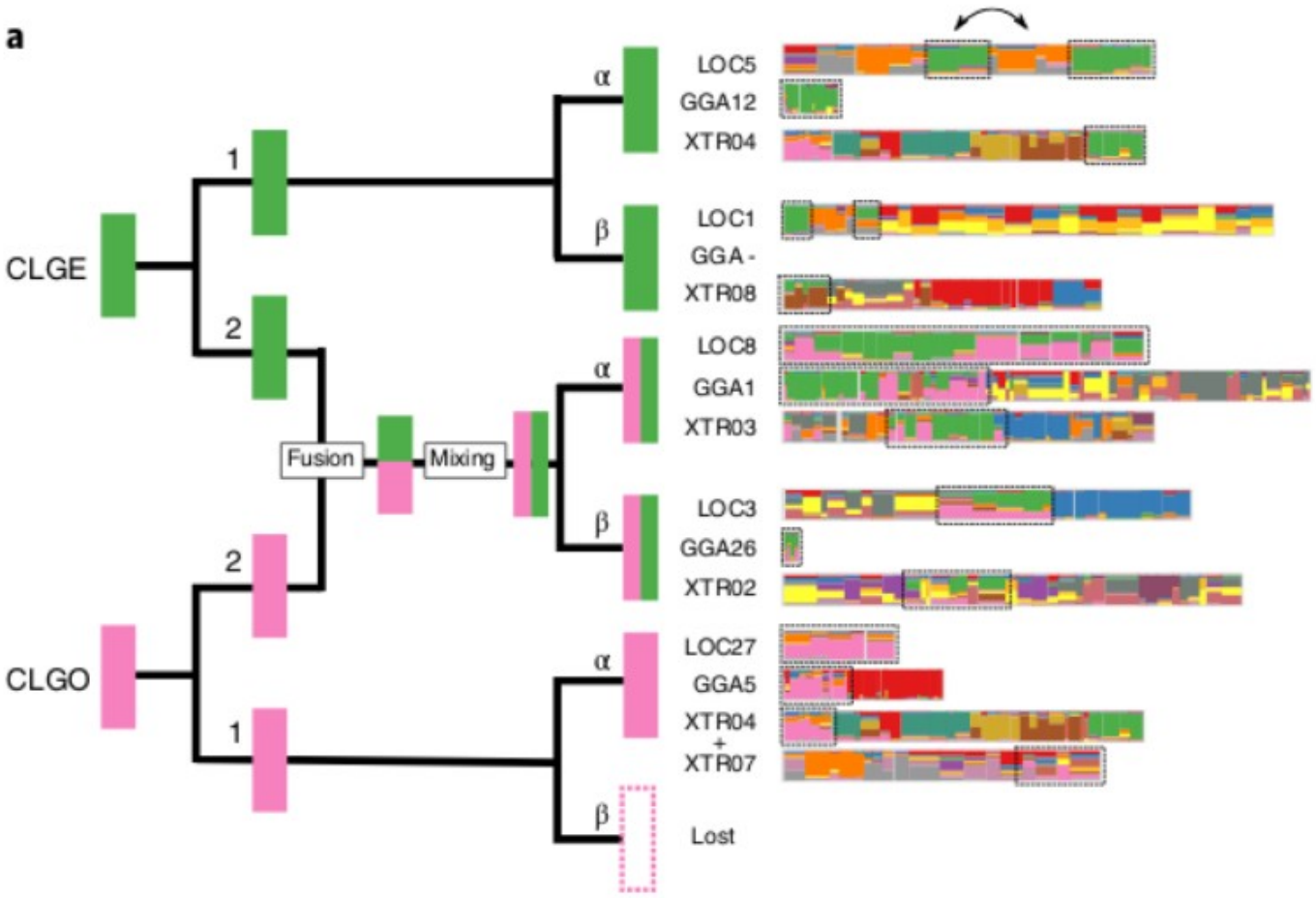
… many underwent fusions (with and without mixing).



Unfused chromosomes: mostly micro-chromosomes

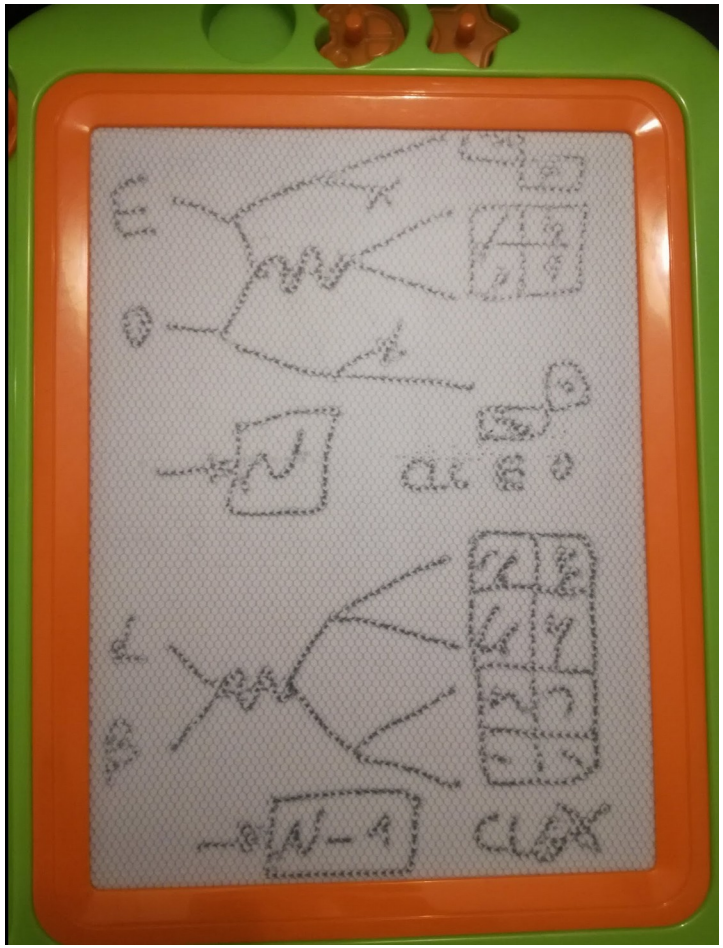
Chicken-specific fusions (sharp syntenic boundaries)
and old vertebrate fusions (mixed) – mostly on macro-chromosomes

FWM events specific for the second round of vertebrate whole genome duplications



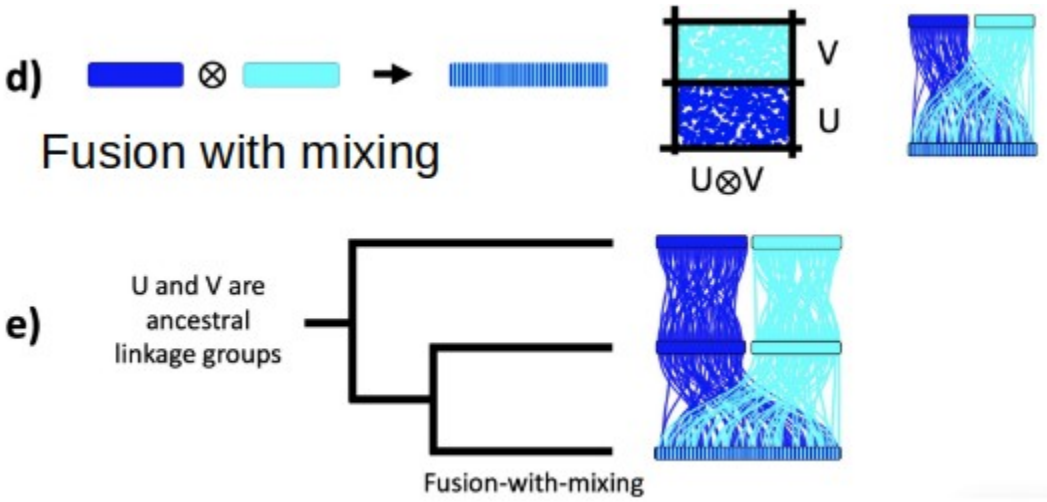
(CLG = Chordate ancestral linkage group
for more details, see Simakov et al, NEE 2020, also Sacerdot et al, Genome Biology 2018)

Worst parenting...

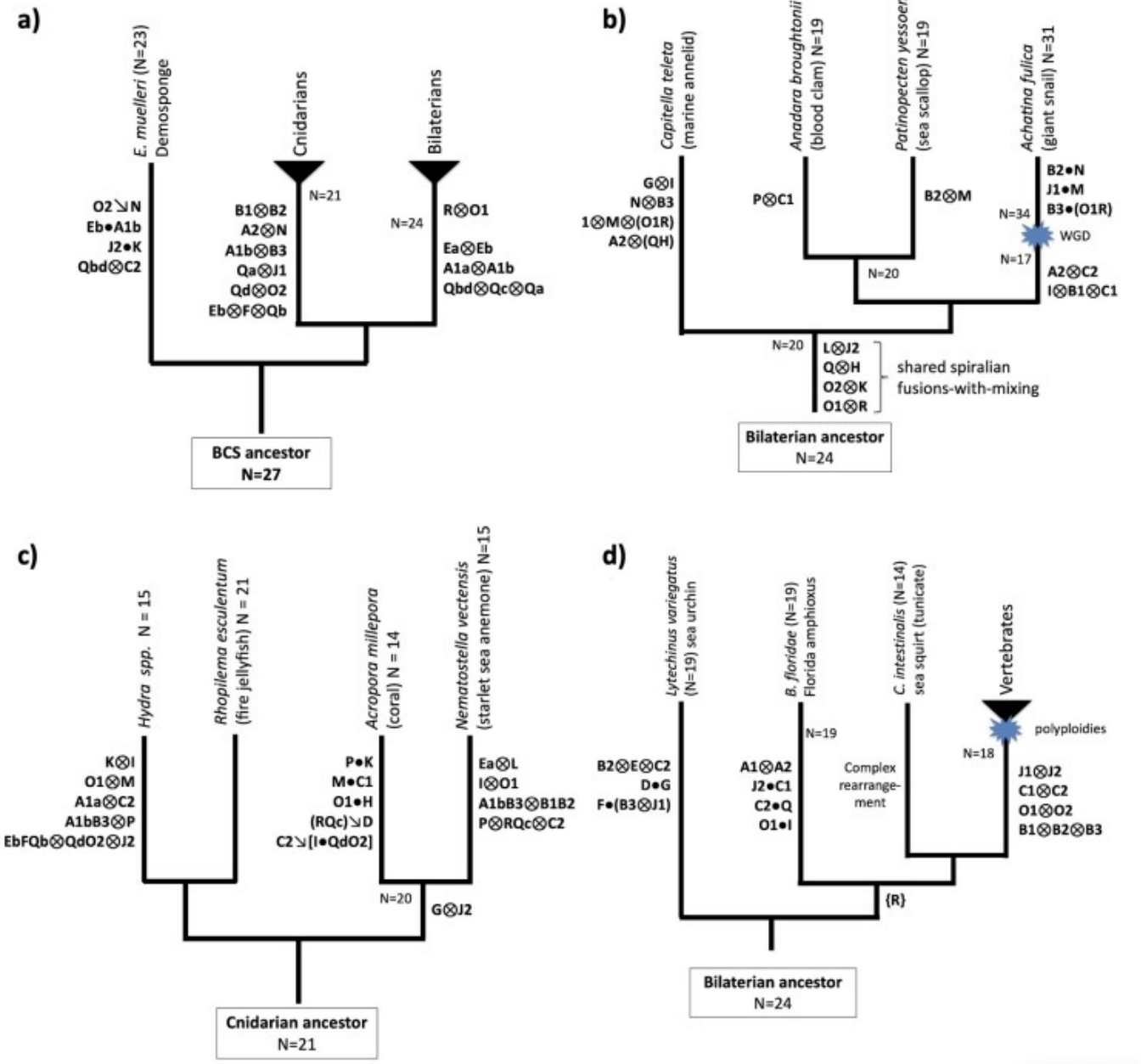


Jan 1st 2020

Irreversible algebraic processes in animal chromosome evolution



Fusion-with-mixing (FWM) – a synapomorphic character in animal chromosomal evolution



Article

Ancient gene linkages support ctenophores as sister to other animals

https://doi.org/10.1038/s41586-023-05936-6

Received: 30 April 2022

Accepted: 9 March 2023

Published online: 17 May 2023

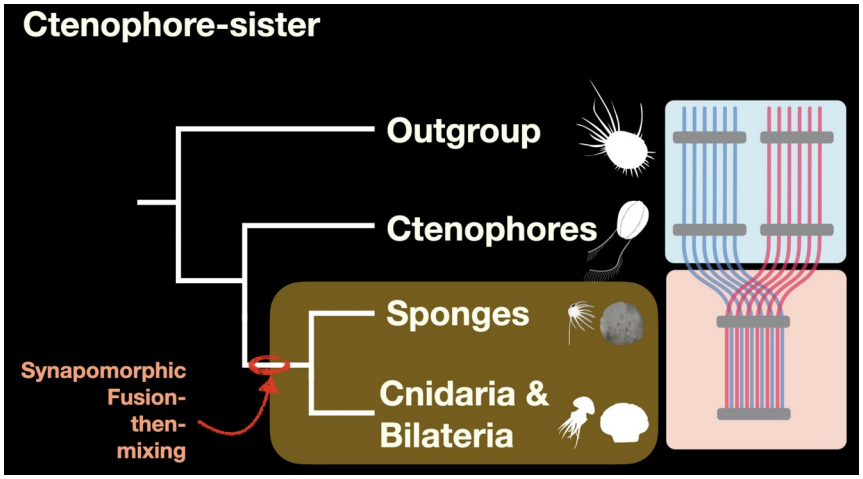
Open access

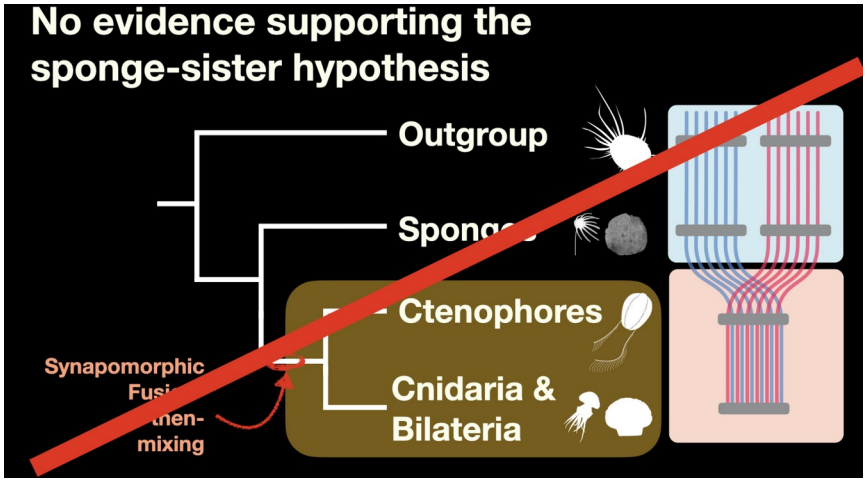
Check for updates

Darrin T. Schultz[1,2,3]✉, Steven H. D. Haddock[2,4], Jessen V. Bredeson[5], Richard E. Green[3], Oleg Simakov[1]✉ & Daniel S. Rokhsar[5,6,7]✉

A central question in evolutionary biology is whether sponges or ctenophores (comb jellies) are the sister group to all other animals. These alternative phylogenetic hypotheses imply different scenarios for the evolution of complex neural systems and other animal-specific traits[1–6]. Conventional phylogenetic approaches based on morphological characters and increasingly extensive gene sequence collections have not been able to definitively answer this question[7–11]. Here we develop chromosome-scale gene linkage, also known as synteny, as a phylogenetic character for resolving this question[12]. We report new chromosome-scale genomes for a ctenophore and two

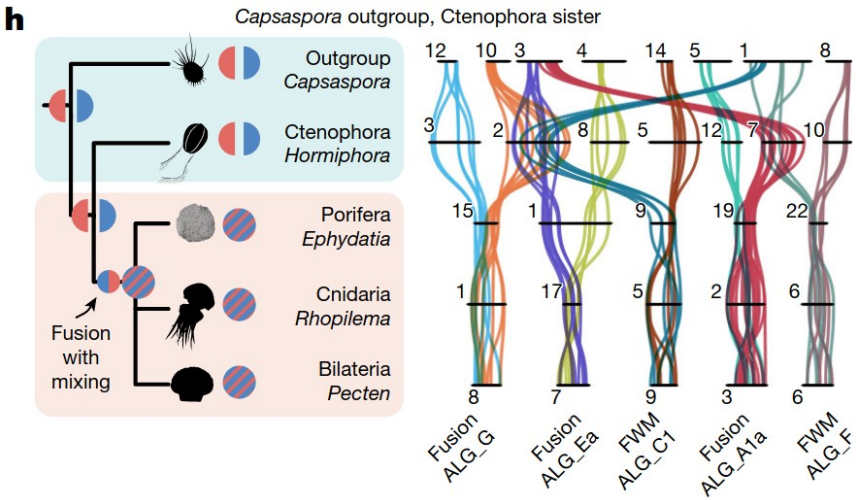
Darrin Schultz

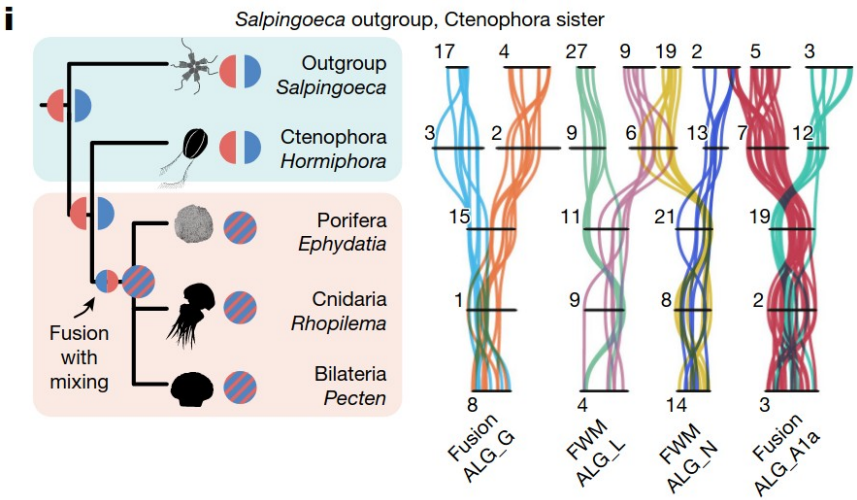




Schultz et al, Nature 2023

Several fusion and fusion-with-mixing characters support “ctenophore sister”





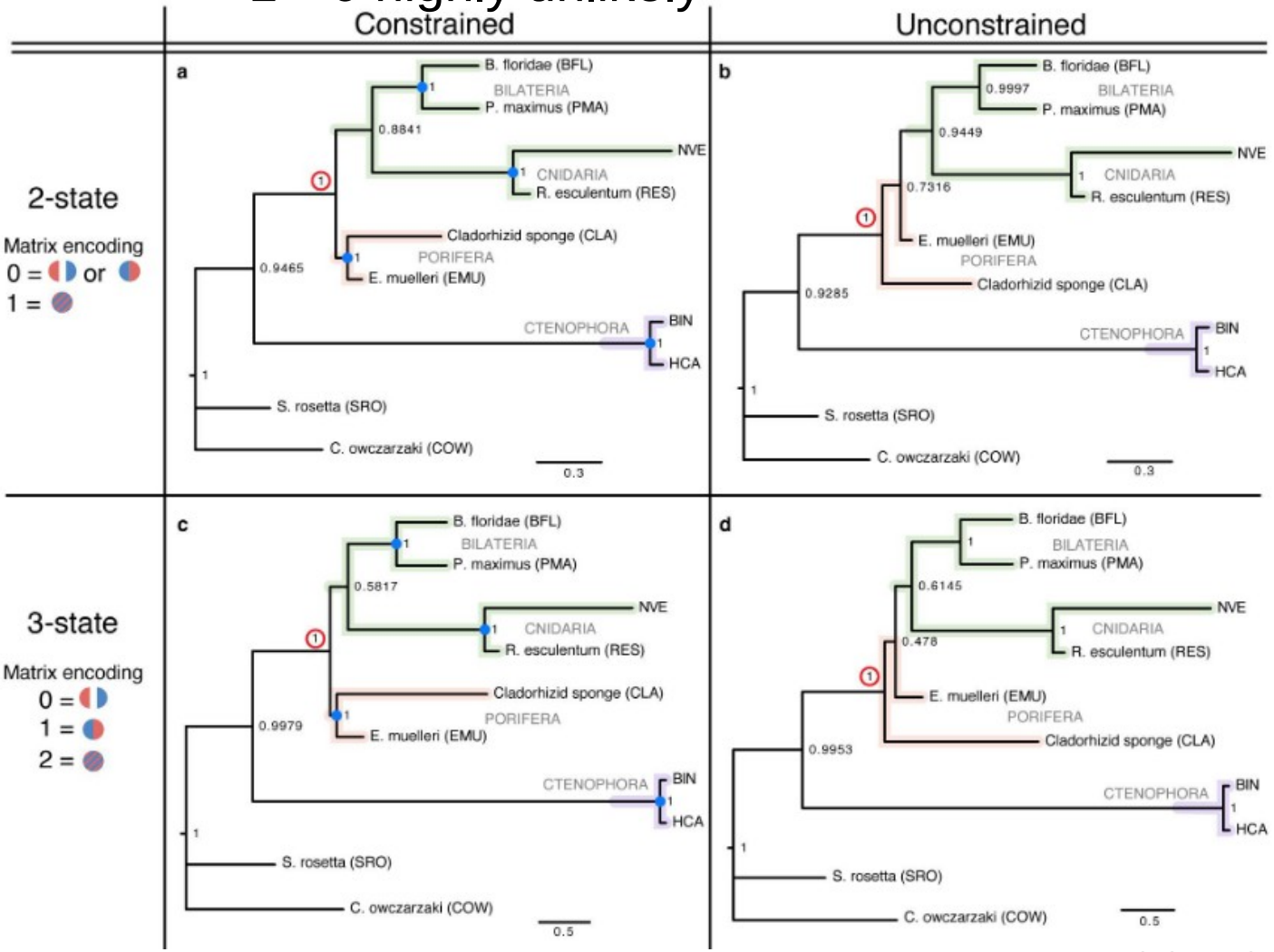
Schultz et al, Nature 2023

But also important to take into account convergent FWM events

Transition rates:

0 → 1 (observing two chromosomes fused by chance)

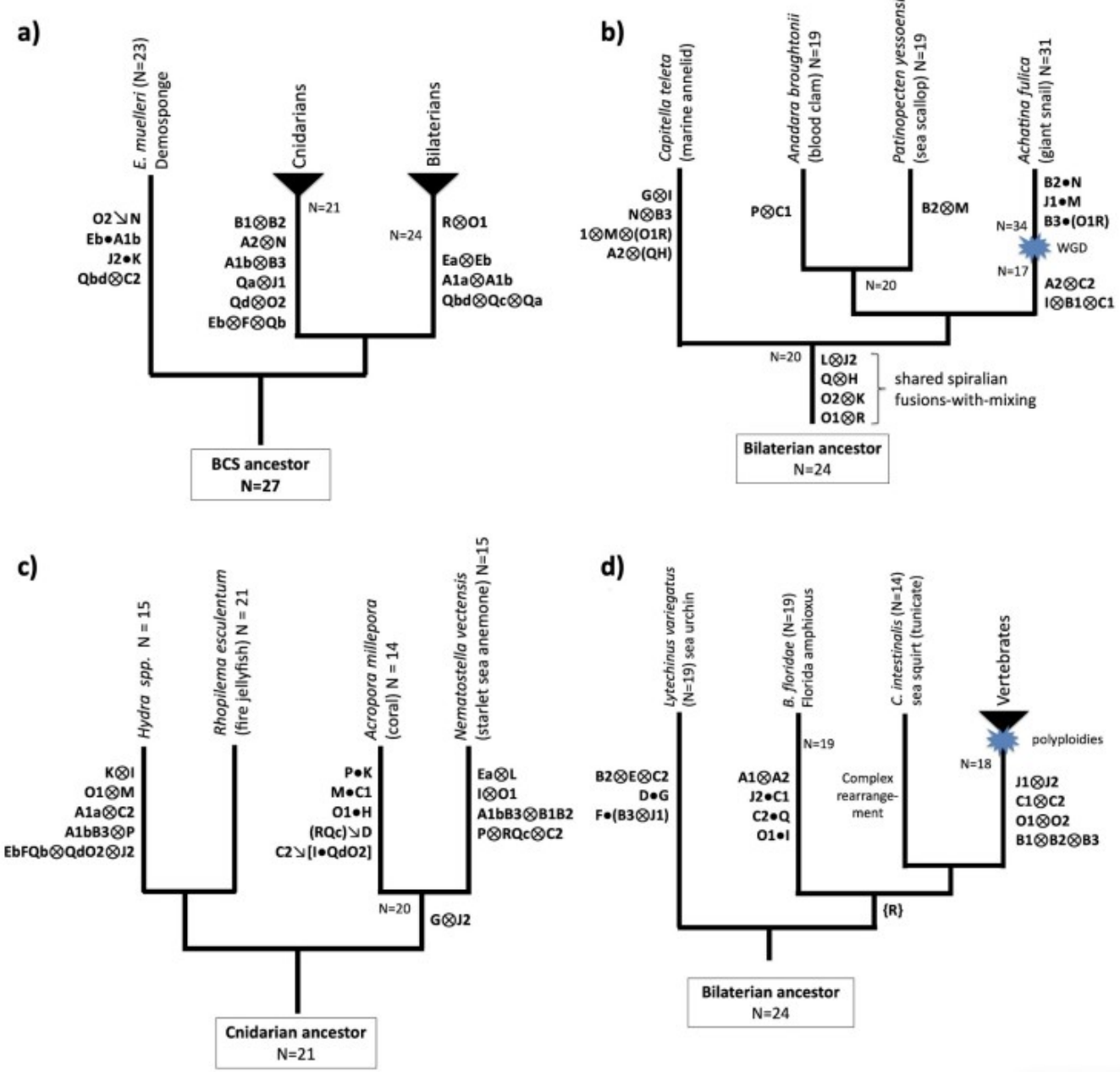
1 → 0 highly unlikely



Schultz et al, Nature 2023

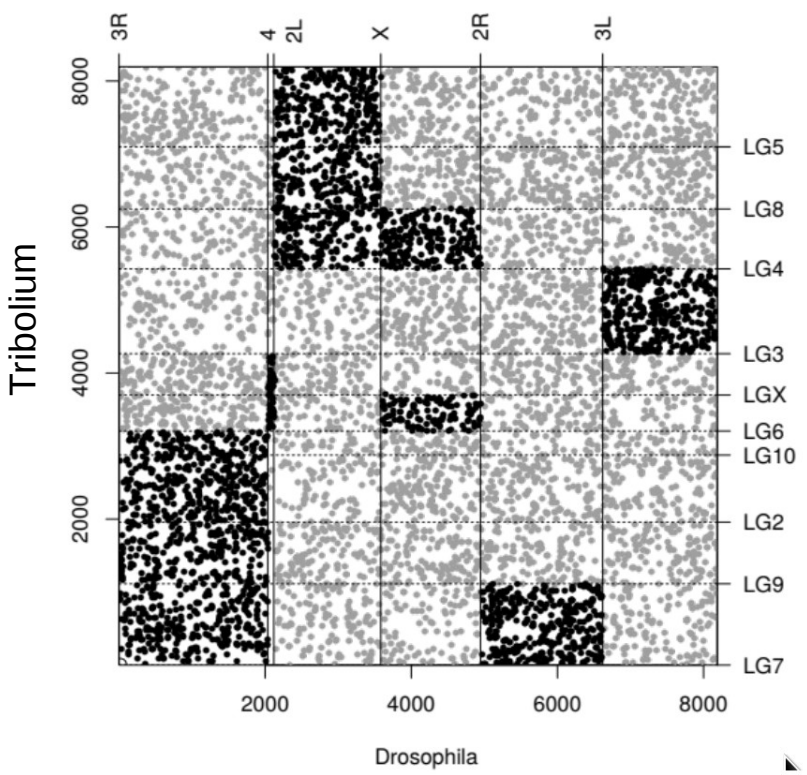
Surprise

Exercise! What clade (bonus for species guess) does each dotplot correspond to?

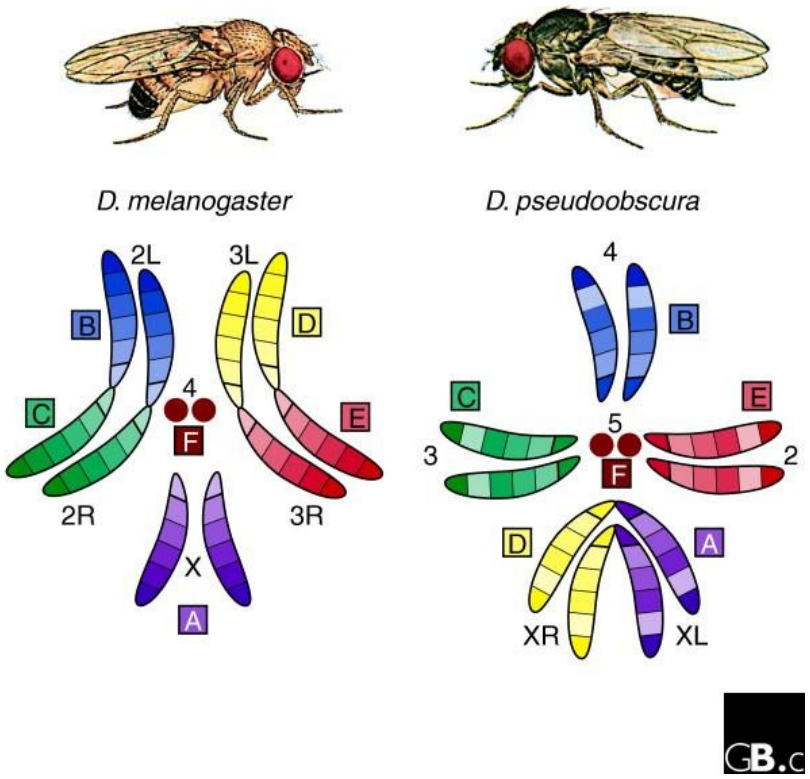


Are the chromosomal “units” the same in all animals?

Loss of the ancestral chromosomal building scheme and the emergence of novel chromosomal elements



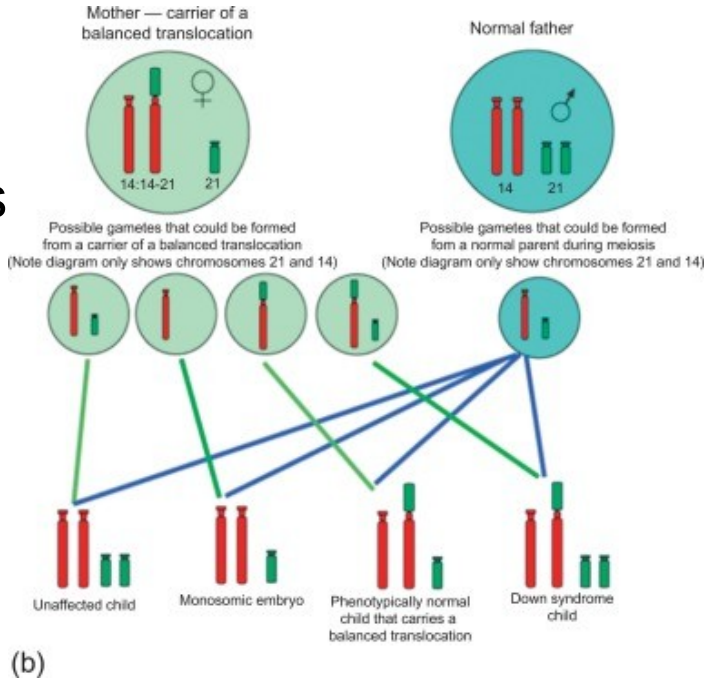
Simakov, 2022



Kulathinal, Hartl, Genome Biol. 2005

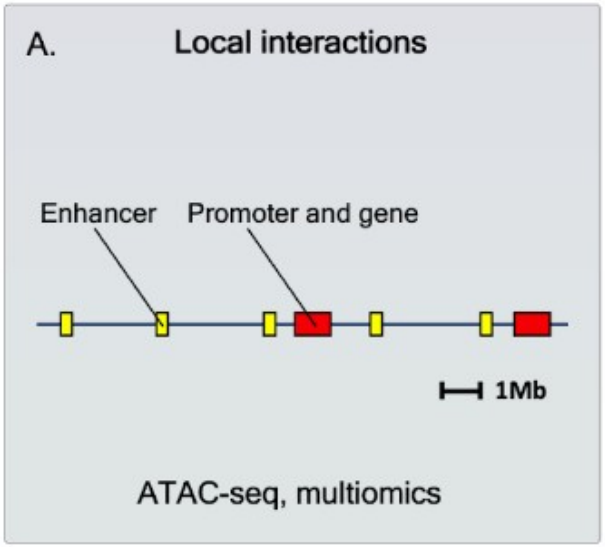
Why are chromosomes preserved?
Why do they break?
Is there any impact?

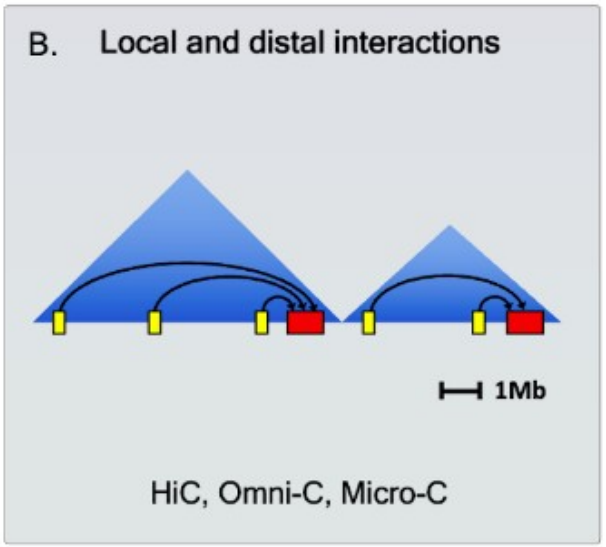
Meiotic constraints

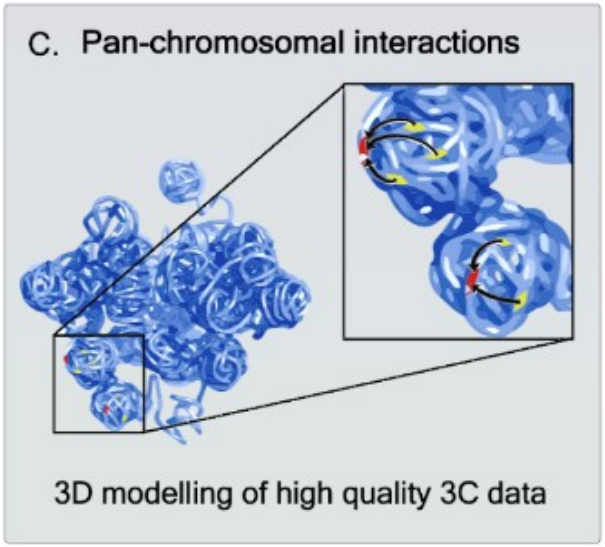


Seidman et al, 2016

Pan-chromosomal regulatory constraints







Rogers et al 2023

NEWS

Scientists declare that octopuses are basically aliens

BY MEREDITH PLACKO 08.14.2015 :: 5:01PM EDT MPLACKO

646 SHARES

THE UNEXPLAINED

Scientists conclude Octopus DNA is out of this world

By Ancient Code

Octopuses 'are aliens', scientists decide after DNA study

Rob Waugh Wednesday 12 Aug 2015 2:02 pm

Shocking Claim: Scientists Think Octopuses Might Be Aliens After Studying Their DNA

According to a group of scientists, the lovable octopus, nature's aquatic contortionists, might actually be aliens not of this world. Researchers who mapped out the genetic code of the octopus found it to be so strange that it could actually be an extraterrestrial.

It's the first genome sequencing ever conducted on a cephalopod and it showed remarkable complexity with some 33,000 protein-coding genes identified - more than is found in a human.

Octopus genetic code is so strange it could be an ALIEN, according to scientists

24,217 views 131 6 SHARE

Unique mode of cephalopod genome evolution

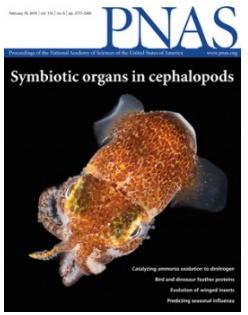
“Vertebrates without WGD”

Convergent expansion of C2H2/PCDH? Yes

Extensive RNAediting? Yes

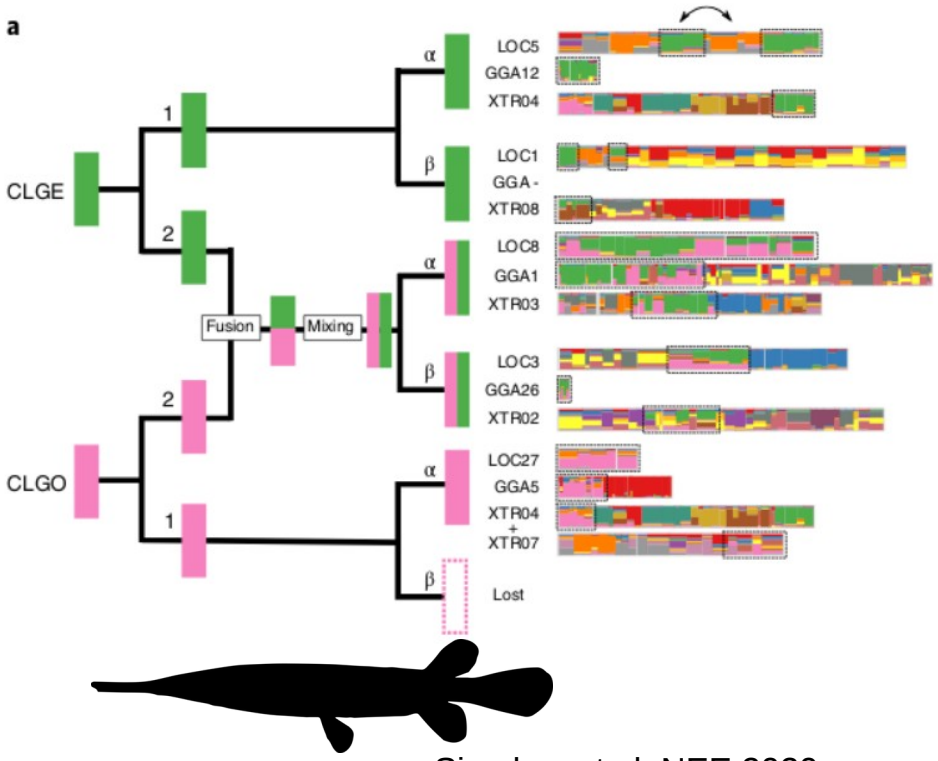
Whole genome duplication? No

Whole genome rearrangements? Yes

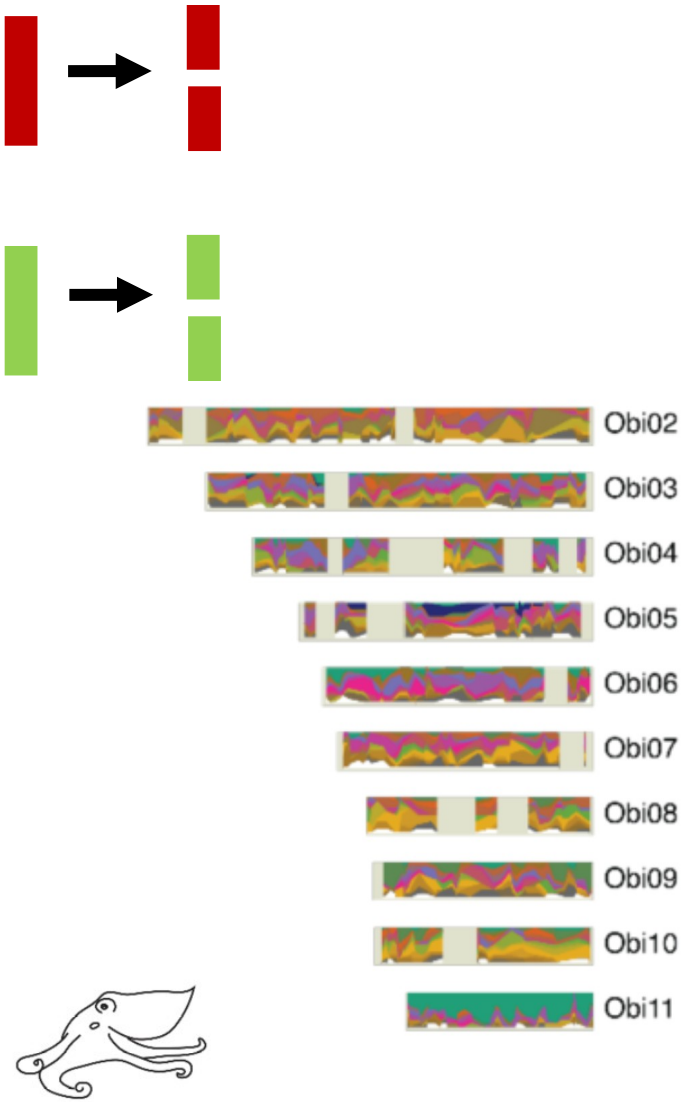


Albertin *et al*, Nature 2015
Belcaid *et al*, PNAS 2019
Albertin et al, Nat Comm 2022
Schmidbaur et al, Nat Comm 2022
Albertin and Simakov, Ann Rev Anim Bio 2020
Ritschard et al, Bioessays 2019

Fusion-with-mixing after duplication or large-scale rearrangements



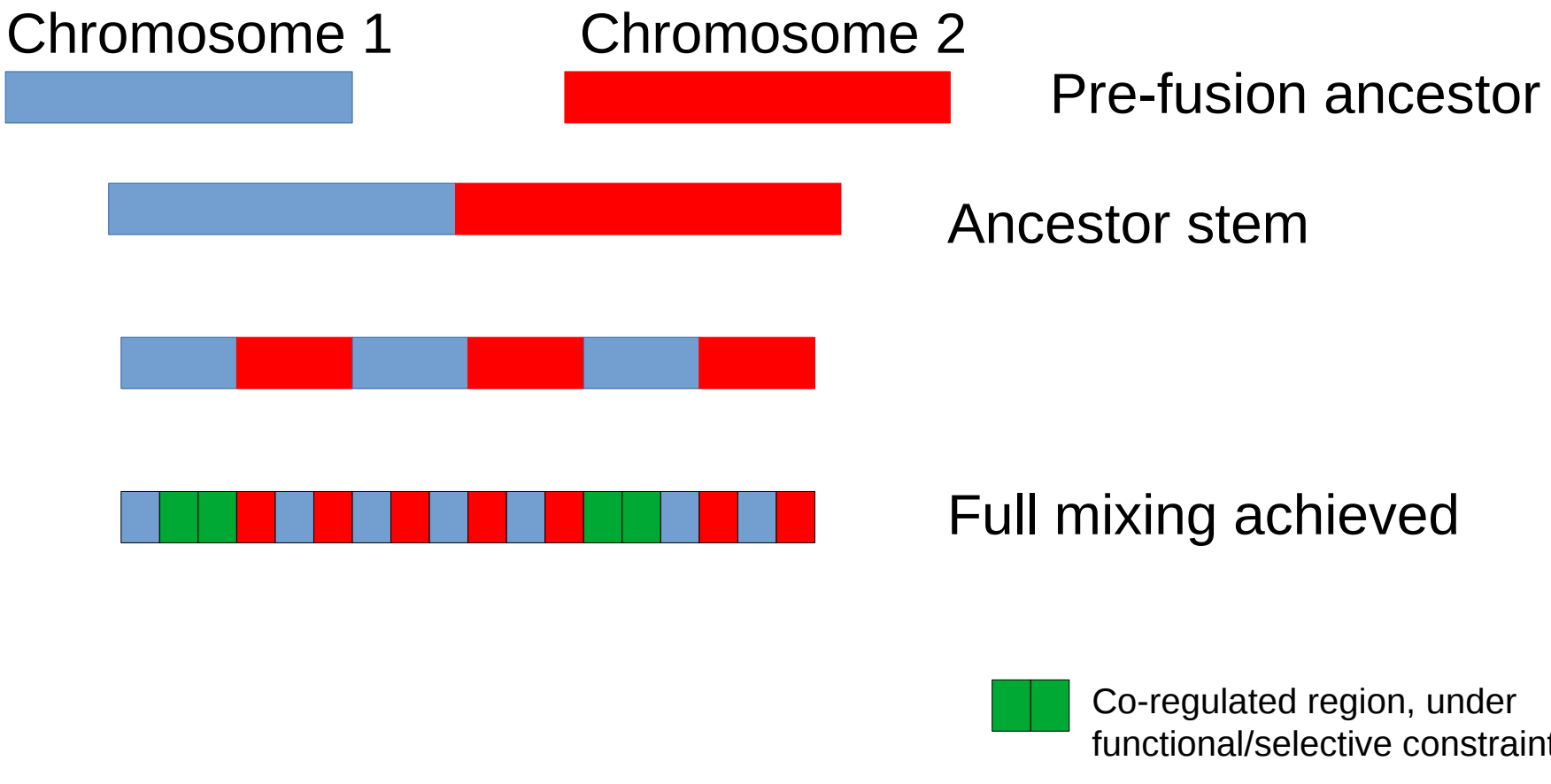
Simakov et al, NEE 2020



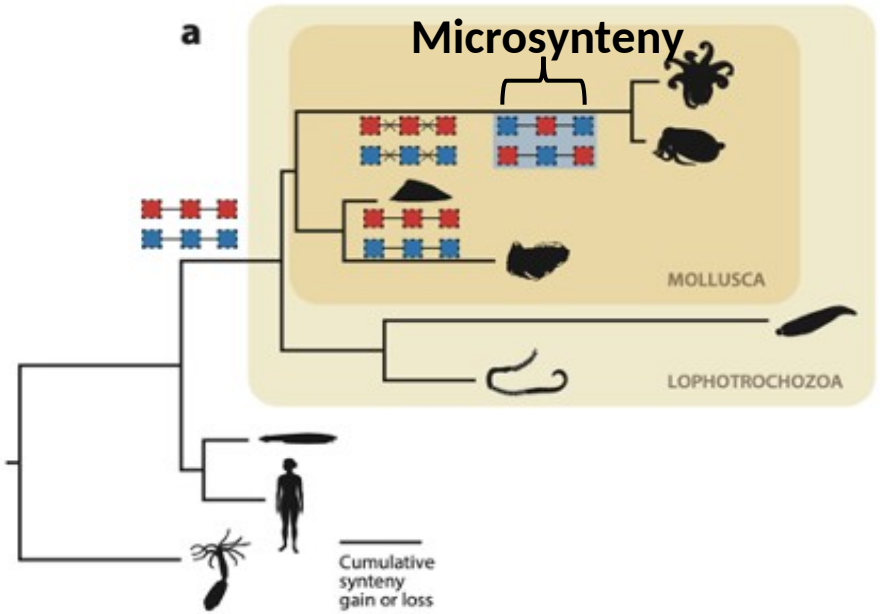
Albertin et al, Nat Comm 2022

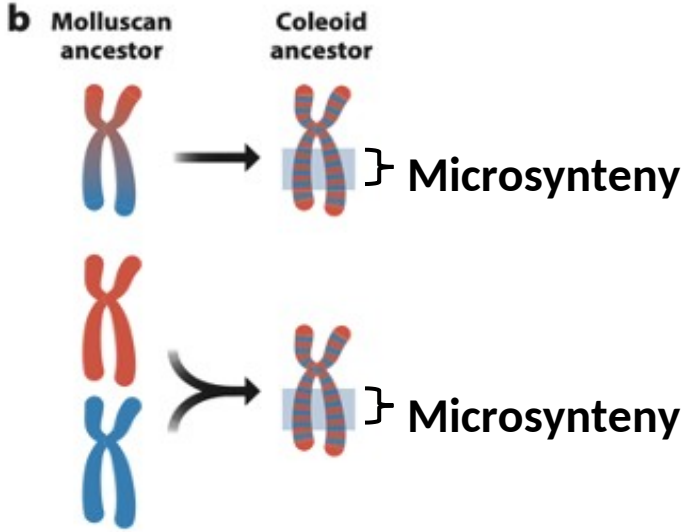
What is going on after such fusions?

Fusion-with-mixing –
a single tectonic event with long-lasting consequences

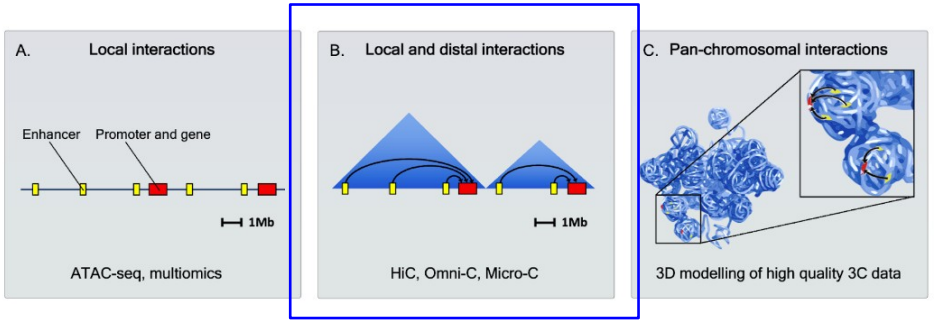


Hundreds of novel gene neighborhoods (micro-syntenies) in the coleoid ancestor



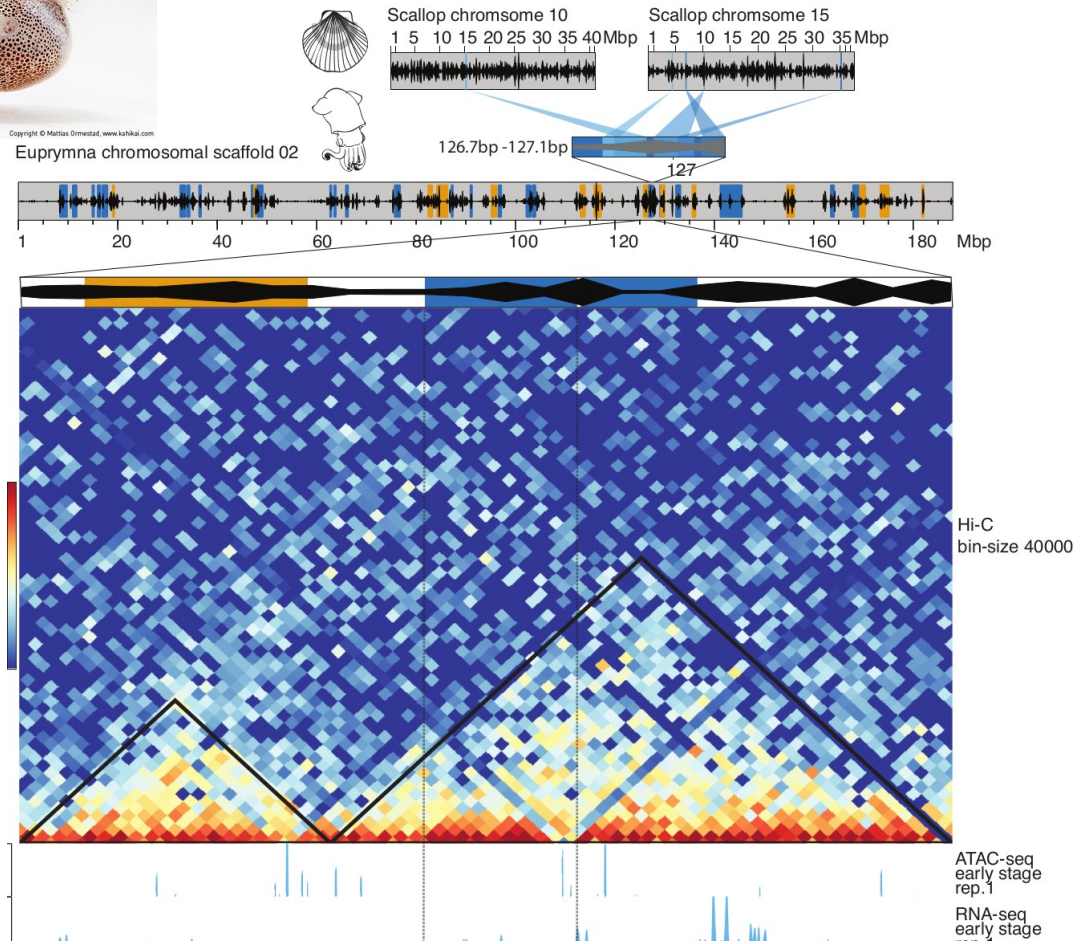


Albertin and Simakov, Ann Rev 2020
Belcaid et al, PNAS 2019
Schmidbaur et al, Nat Comm 2022
Robert et al, BMC Genomics 2022
Robert et al, Bioinformatics 2022
Zimmermann et al, NEE 2019

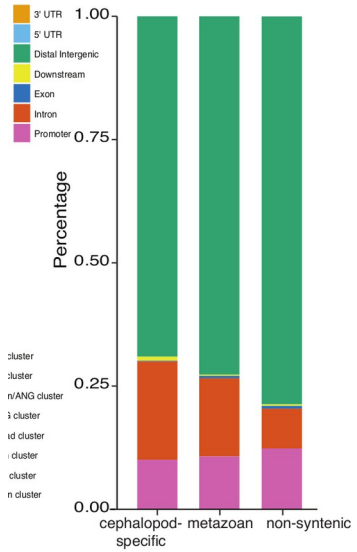


Rogers et al, 2023

Microsyntenies Associated with Cephalopod Innovations (MACIs) form separate topological units



ATAC-peaks in introns

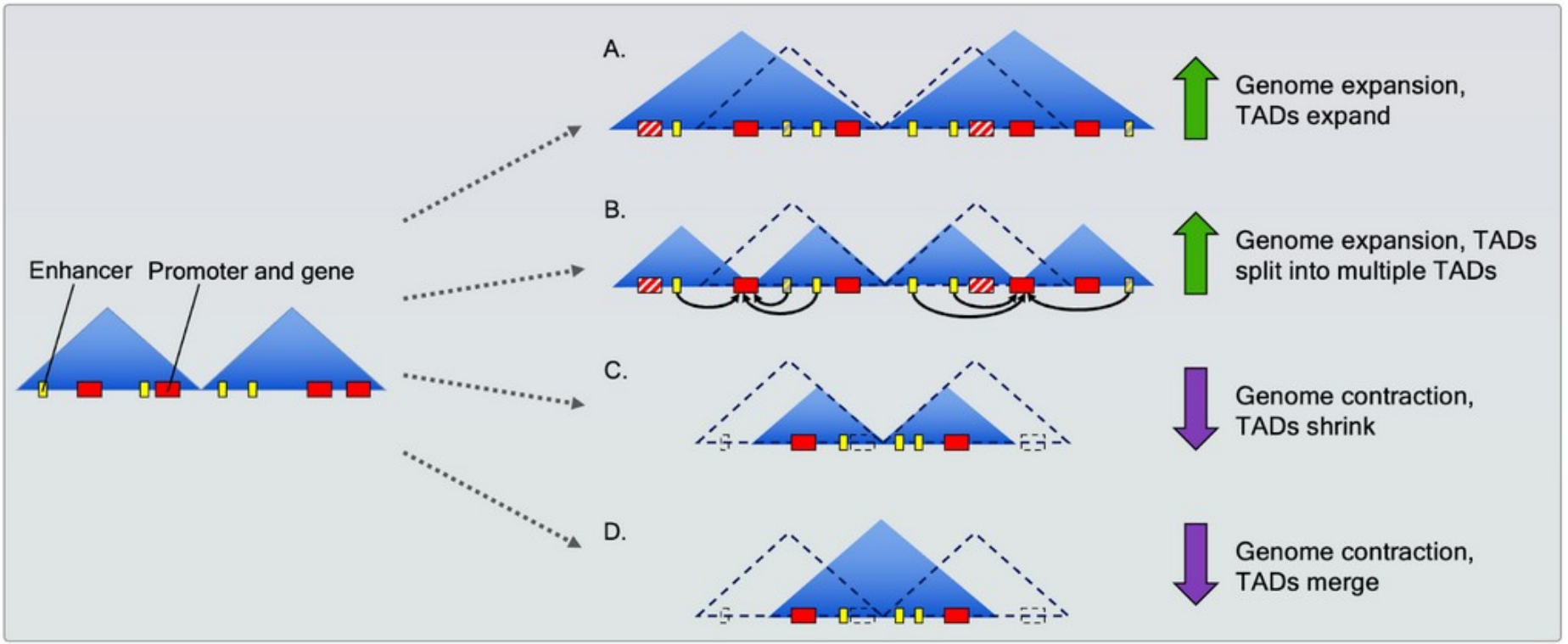


- novel microsyntenic regions show different co-expression and topological properties
- putative enhancers in intronic regions

Schmidbaur *et al*, Nat Comm 2022
Rouressol *et al*, iScience 2023
Rogers *et al*, in preparation

Thea Rogers

Impact of topology in changing genomes

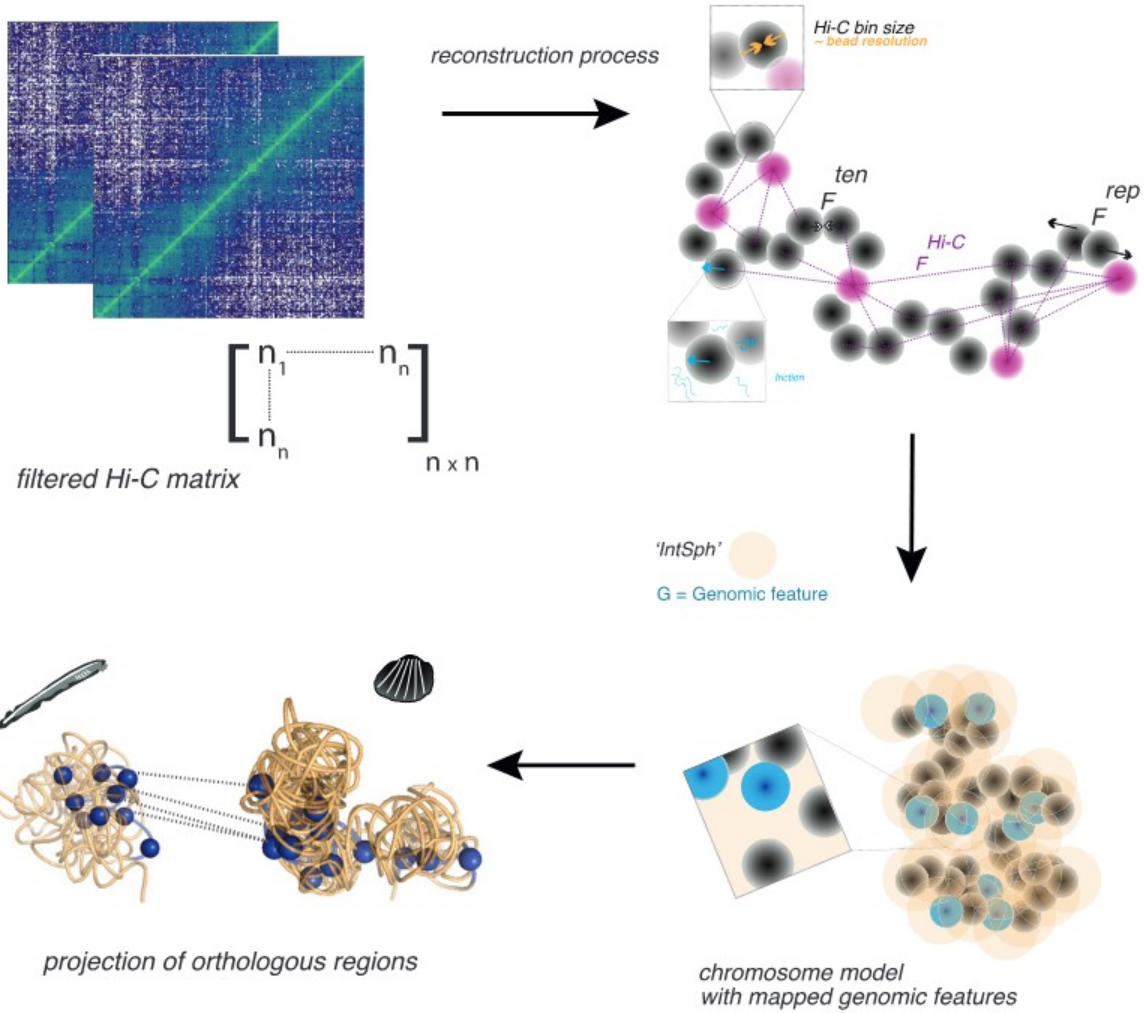


Rogers et al, in press

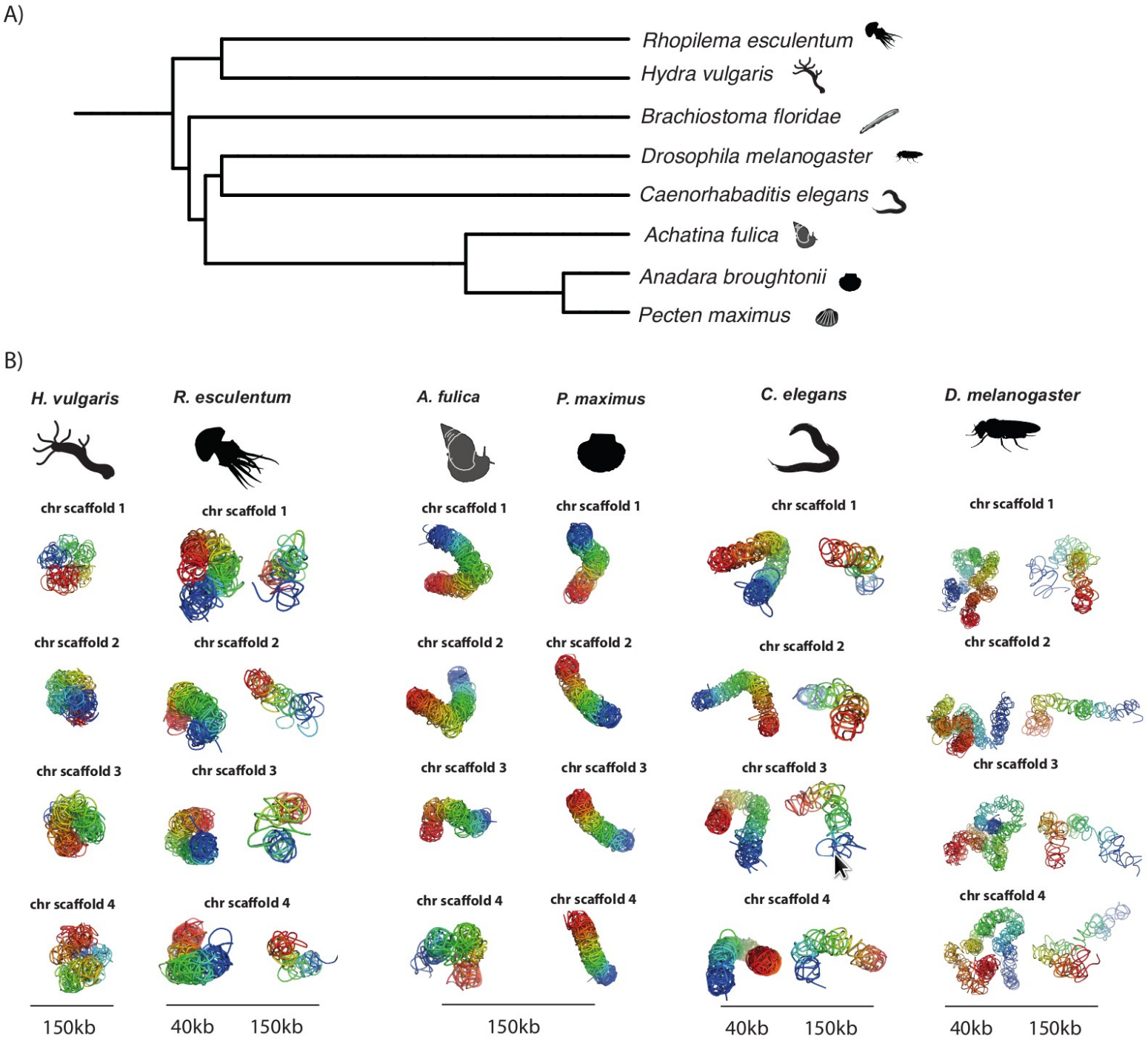
Thea Rogers

Squid genomes are on average 2x larger than the octopus genomes

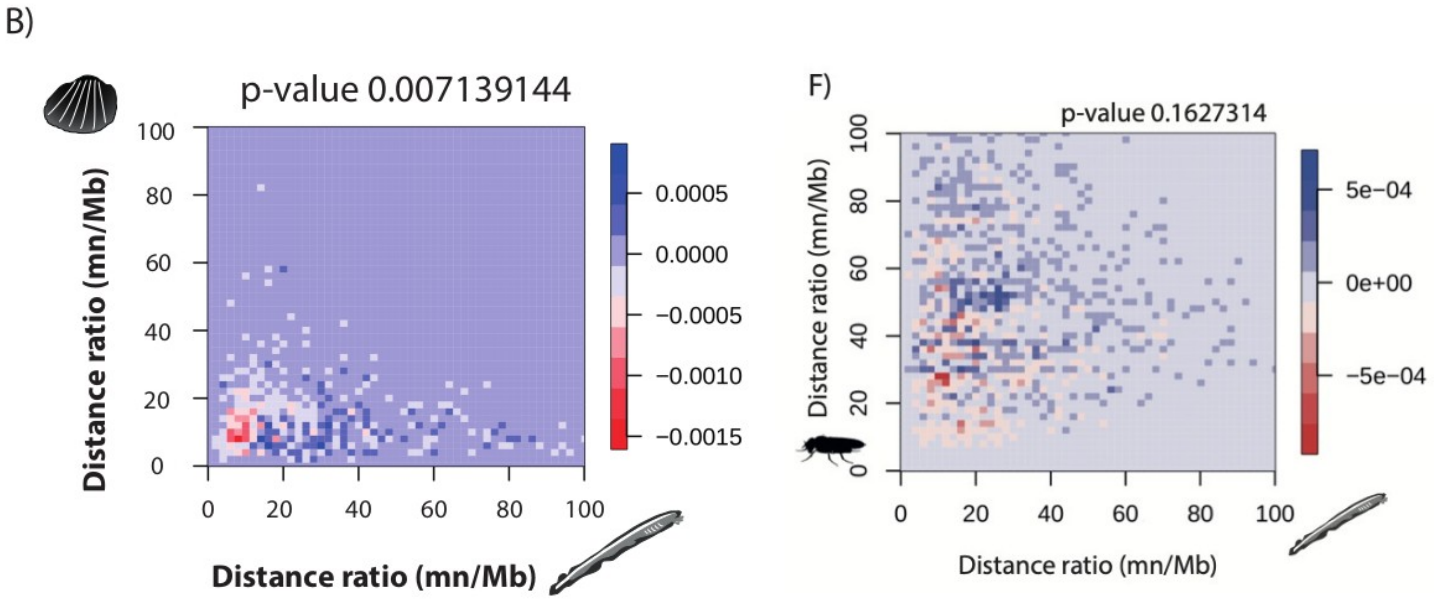
Test for evolutionarily conserved pan-chromosomal interactions



Clarence et al, iScience 2023

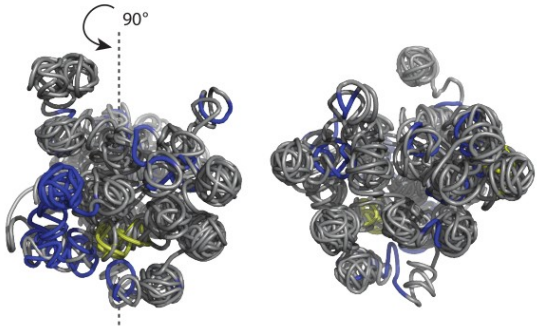

A)
Rhopilema esculentum
Hydra vulgaris
Brachiostoma floridae
Drosophila melanogaster
Caenorhabaditis elegans
Achatina fulica
Anadara broughtonii
Pecten maximus
B)
H. vulgaris
R. esculentum
A. fulica
P. maximus
C. elegans
D. melanogaster
chr scaffold 1
chr scaffold 2
chr scaffold 3
chr scaffold 4
150kb
40kb
150kb
150kb
40kb
150kb
40kb
150kb

Synteny beyond local gene clusters?



Enrichment of long-range interaction distance ratio among orthologous genes

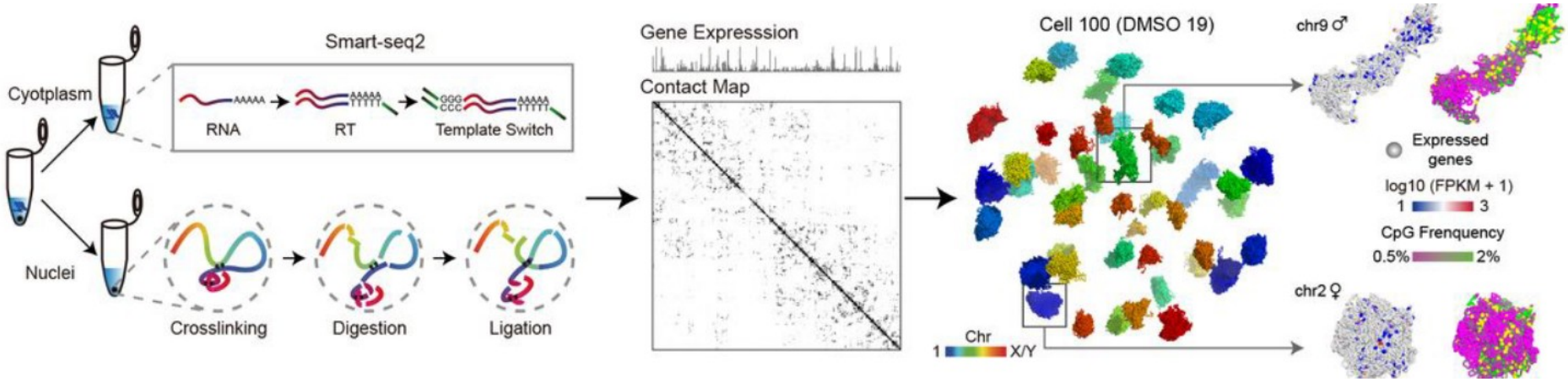
Clarence et al, 2023



Preferential localization of micro-syntenic clusters?

Schmidbaur et al, 2022

Bottleneck: HiC at single cell resolution



Tan lab at Stanford:
https://3dgeno.me/publications

Recap: Current status and challenges

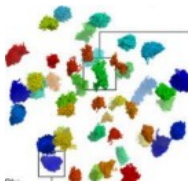
- Chromosomal-scale genomes and extensive species sampling is almost a norm

- Big challenge to identify regulatory sequences and EP interactions, test using knock-out etc

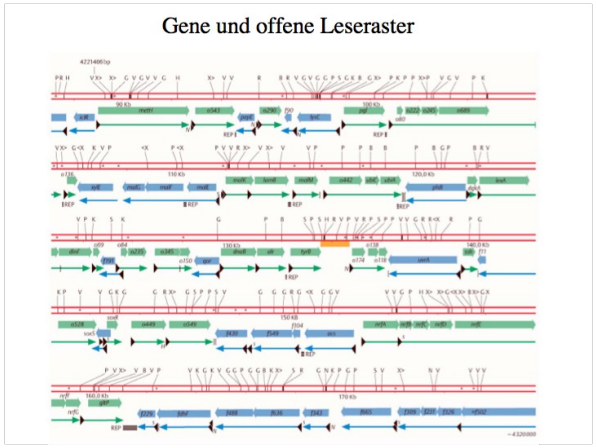
- Evidence for the role of topology at local (sub-chromosomal) scales

- “It’s complicated” at the whole chromosome scale

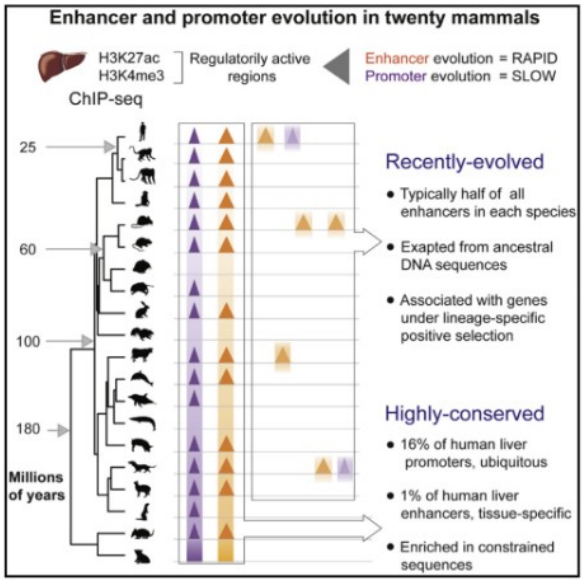
What will genome be like in 2030's/aka next year?



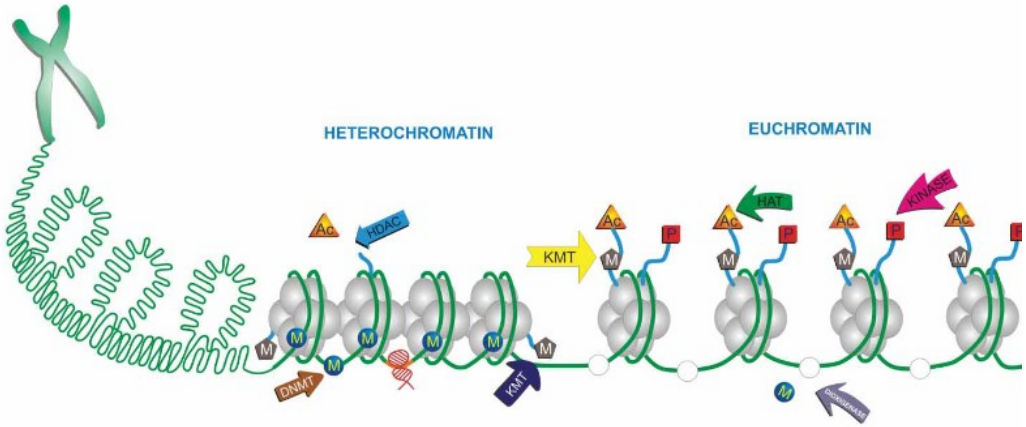
Gene annotation



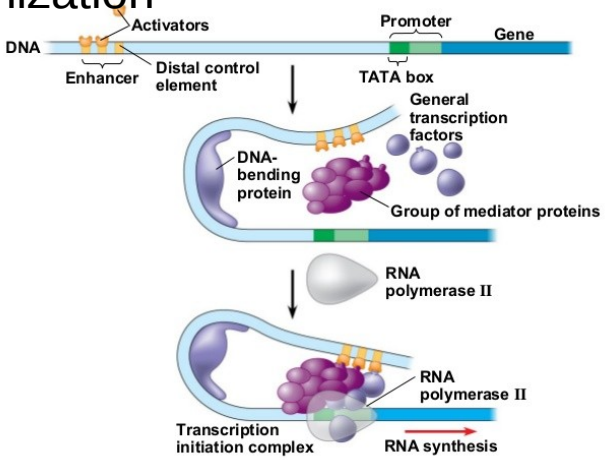
Regulatory regions



Epigenetic modification



3D organization



Villar et al, Cell 2015; Villota-Salazar et al, 2016; http://advancedtranscription16.blogspot.co.at/p/overview.html

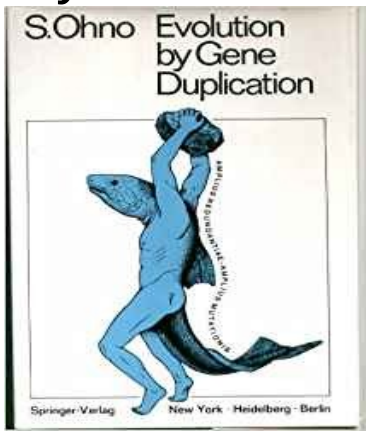
III. Future theoretical genomics, predictions, and applications

“Evolutionary topology of genomes, form and function”

We can reconstruct the genomes of the first animals

Over 600 million years ago

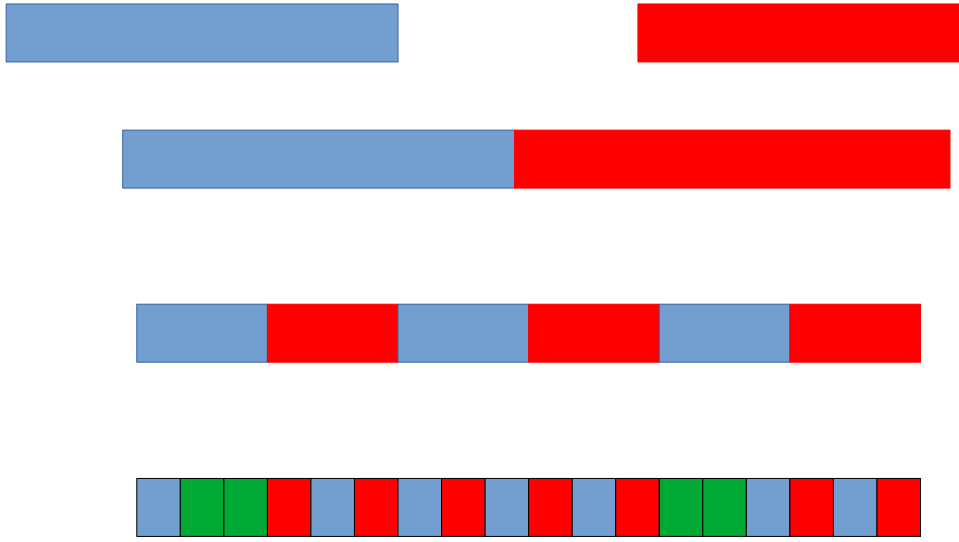
… but how do clade-specific (i.e. macro-) evolutionary “trajectories” contribute to evolutionary novelty?

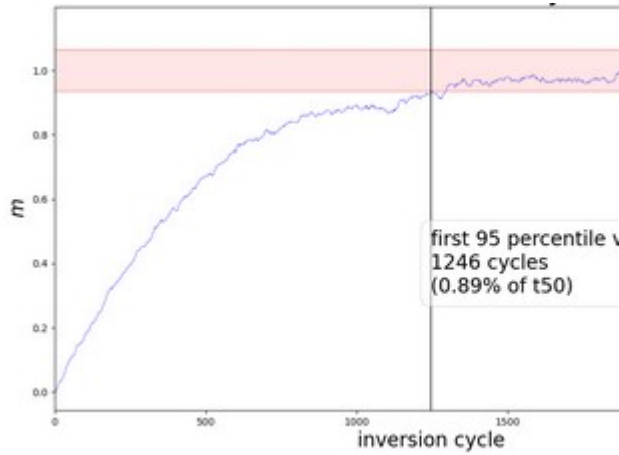


Fusion-with-mixing –
a single tectonic event with long-lasting consequences

Chromosome 1

Chromosome 2





Full mixing achieved

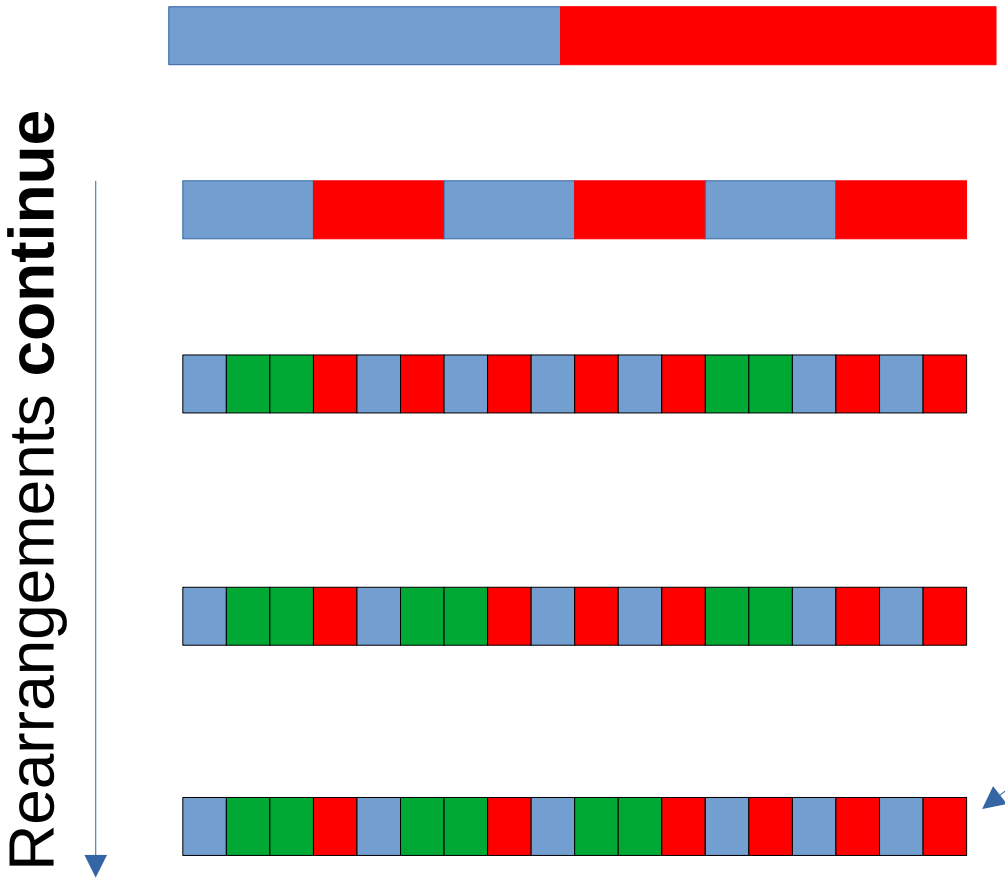
Co-regulated region, under functional/selective constraint

Arno Bluemel Darrin Sch

Time to explore half of all possible neighborhoods (τ_{50}) – macro-evolutionary time-scale

Chromosome 1

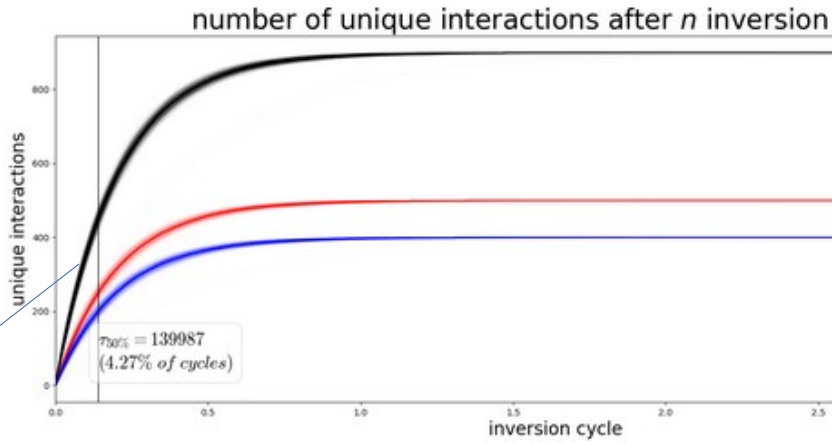
Chromosome 2



Rearrangements **continue**

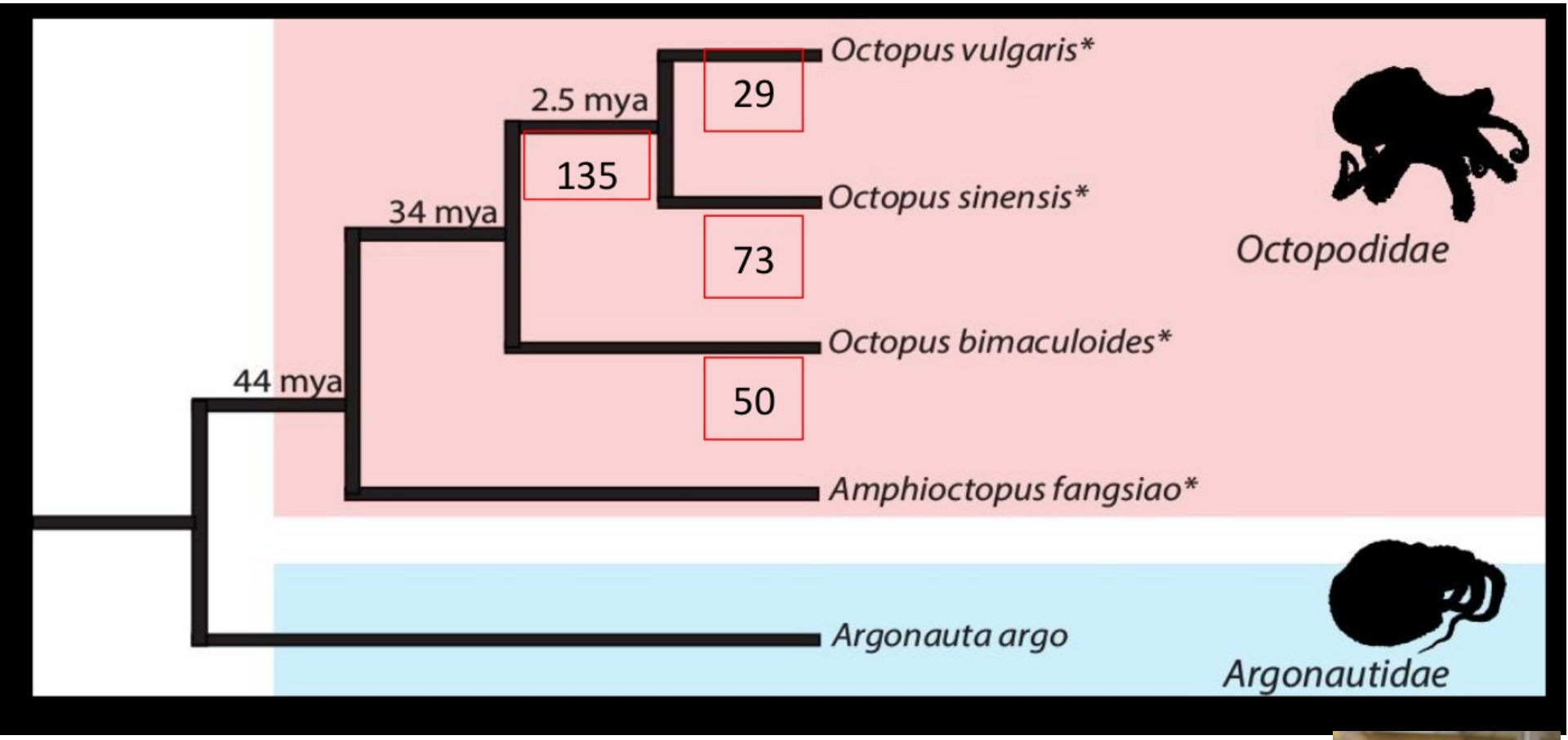
50% of all possible combinations explored

$\tau_e << \tau_{50}$



Arno Bluemel Darrin Sch

τ_{50} for an “average” chromosome



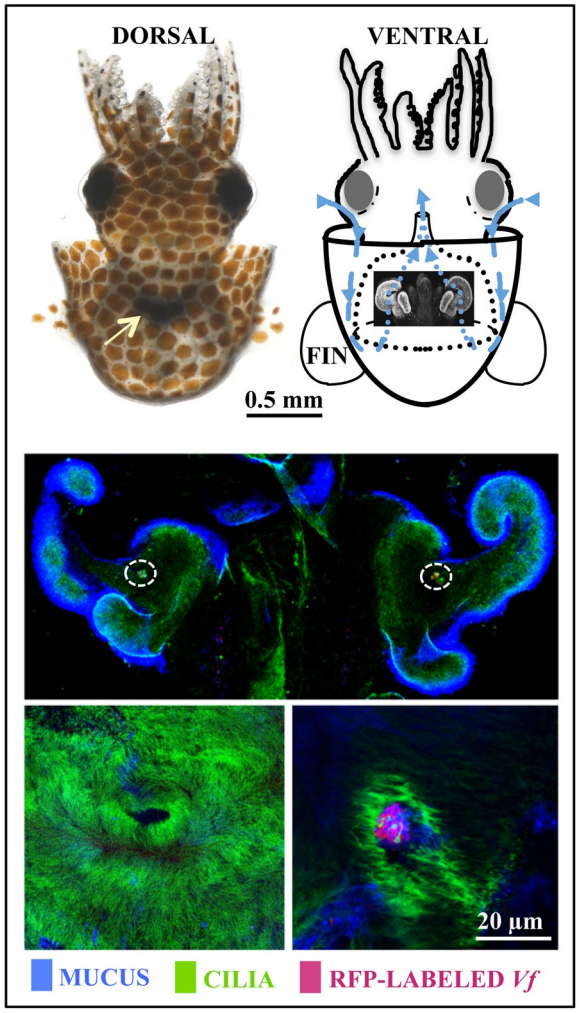
→ breaks occur at highly insulated regions (TAD boundaries)
→ many neighborhoods still to be explored after 600 million years!

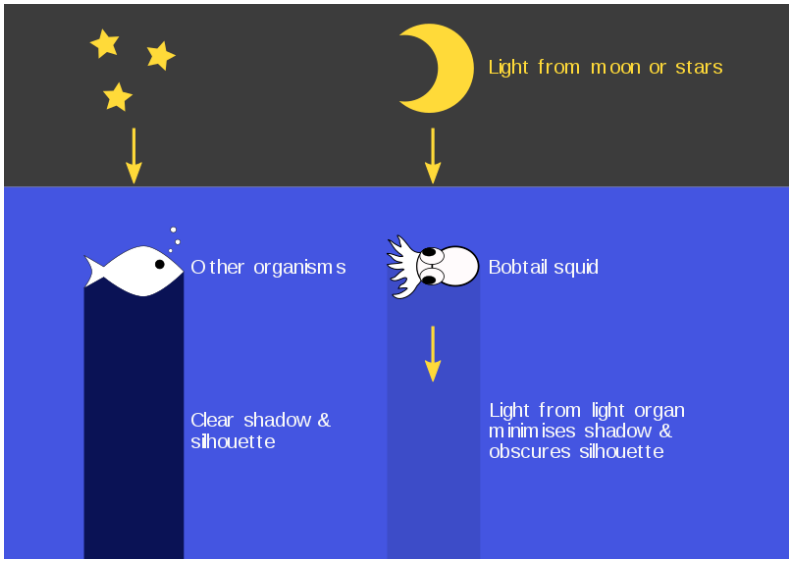
Dalila Destanc

Hypothesis: (singular) ancient genomic events* continue to shape clade-specific evolution

* irreversible

Light Organ (LO)





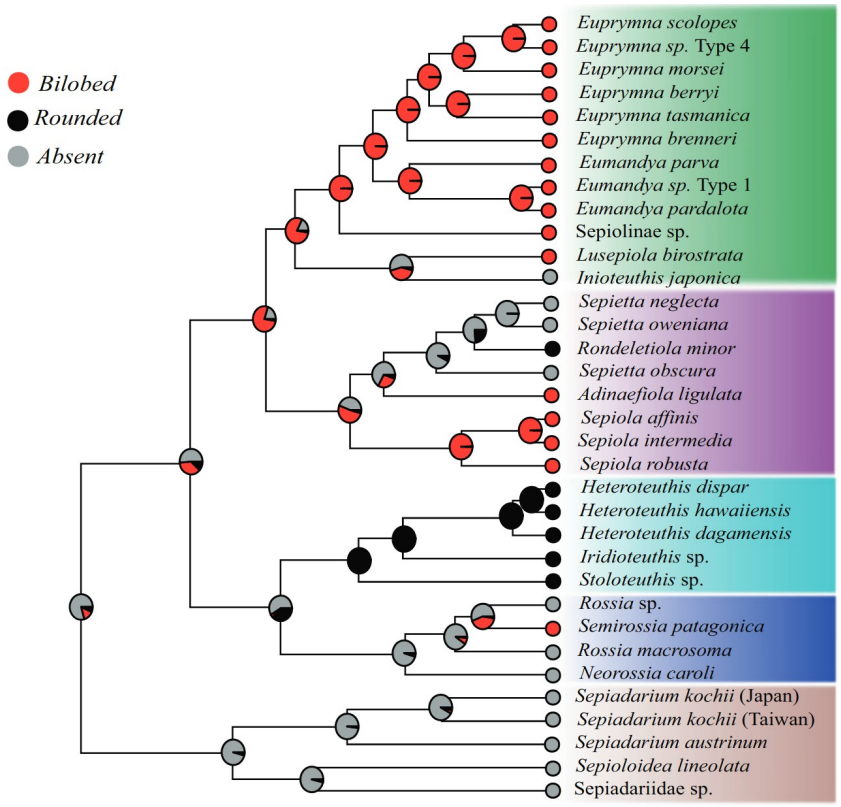
http://manabu-biology.com

McFall-Ngai, 2014

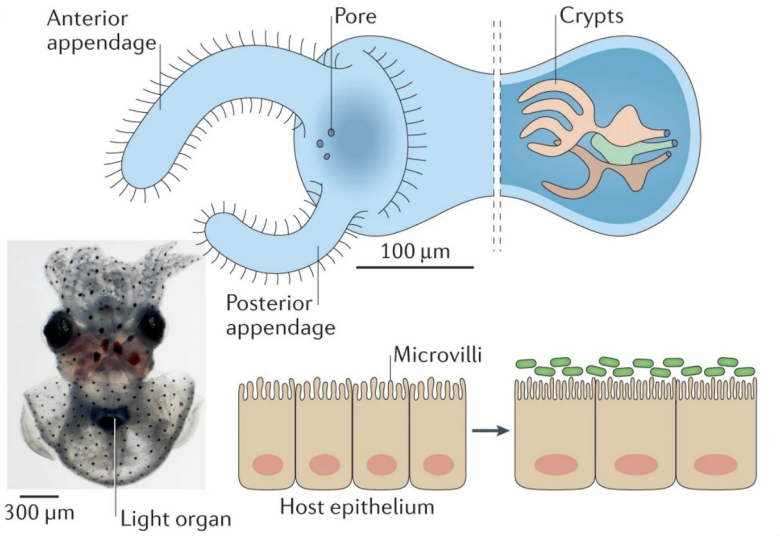
Apical epithelial surface

Symbiosis – light production (not nutritionally obligate, but key ecologically)

Origin of novel organ (LO) in bobtail squids



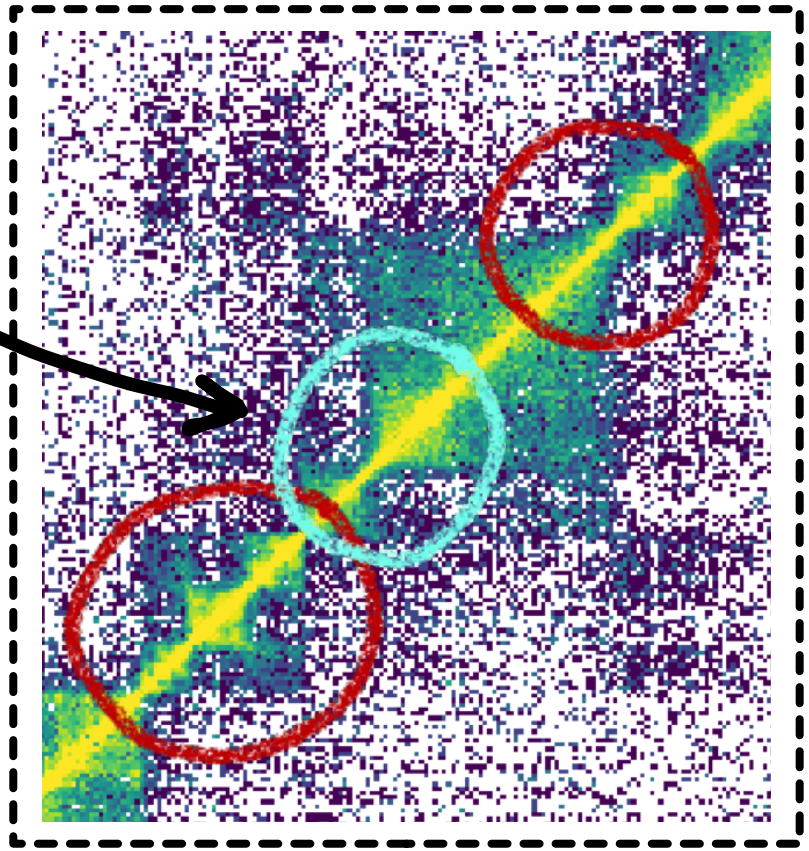
Sanchez, et al, 2021

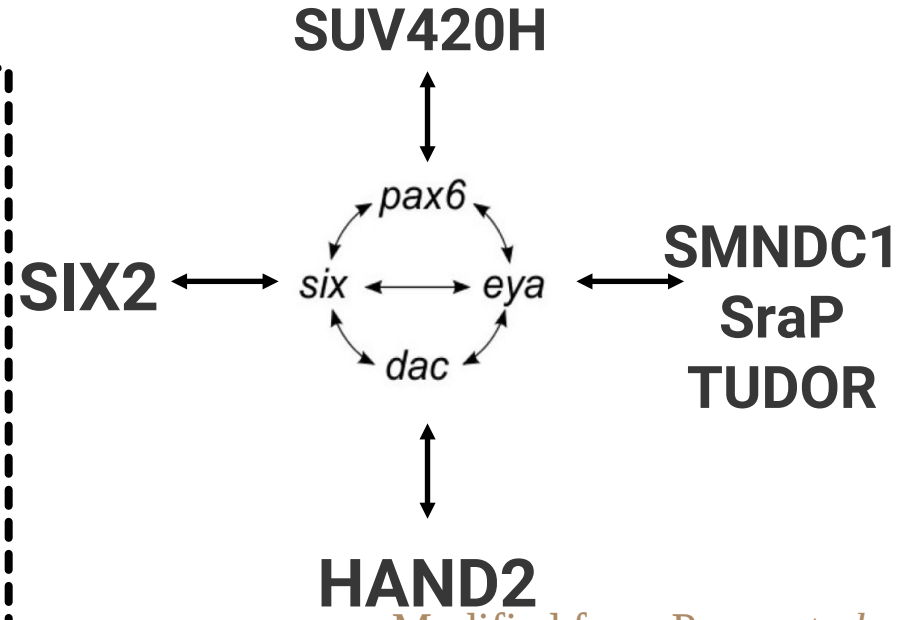


Nyholm, McFall-Ngai, 2021

Time-scale ~several 100 my

Novel neighborhoods associated with LO development



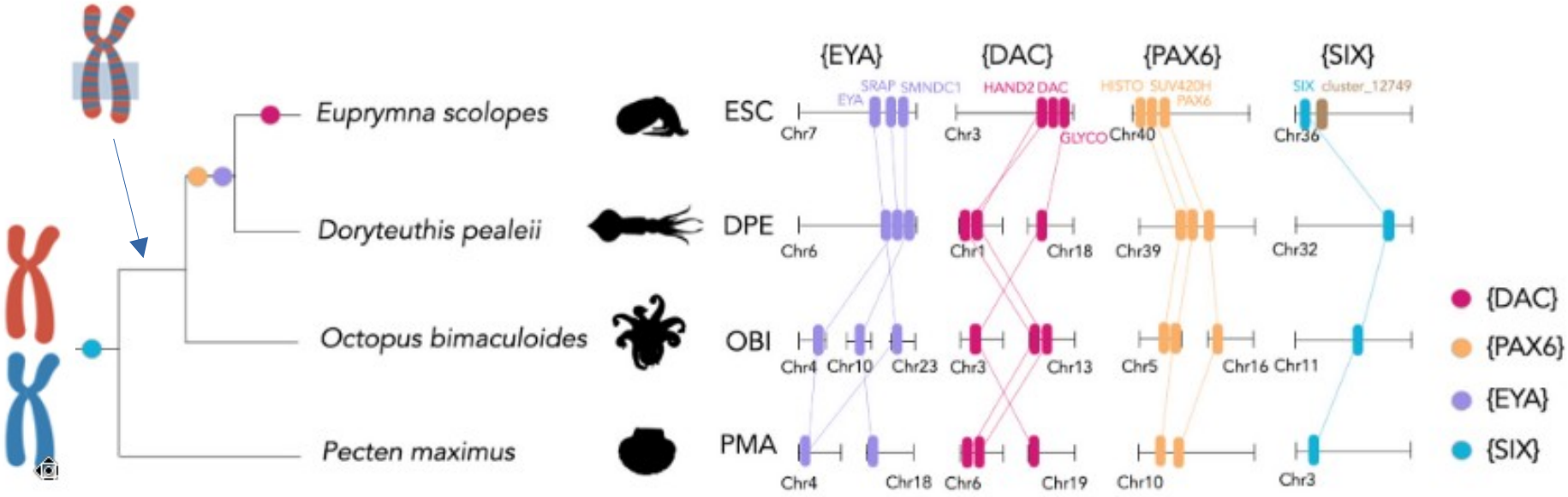


Modified from Peyer *et al.*, 201

Lisa Rouressol
U Vienna/Caltech

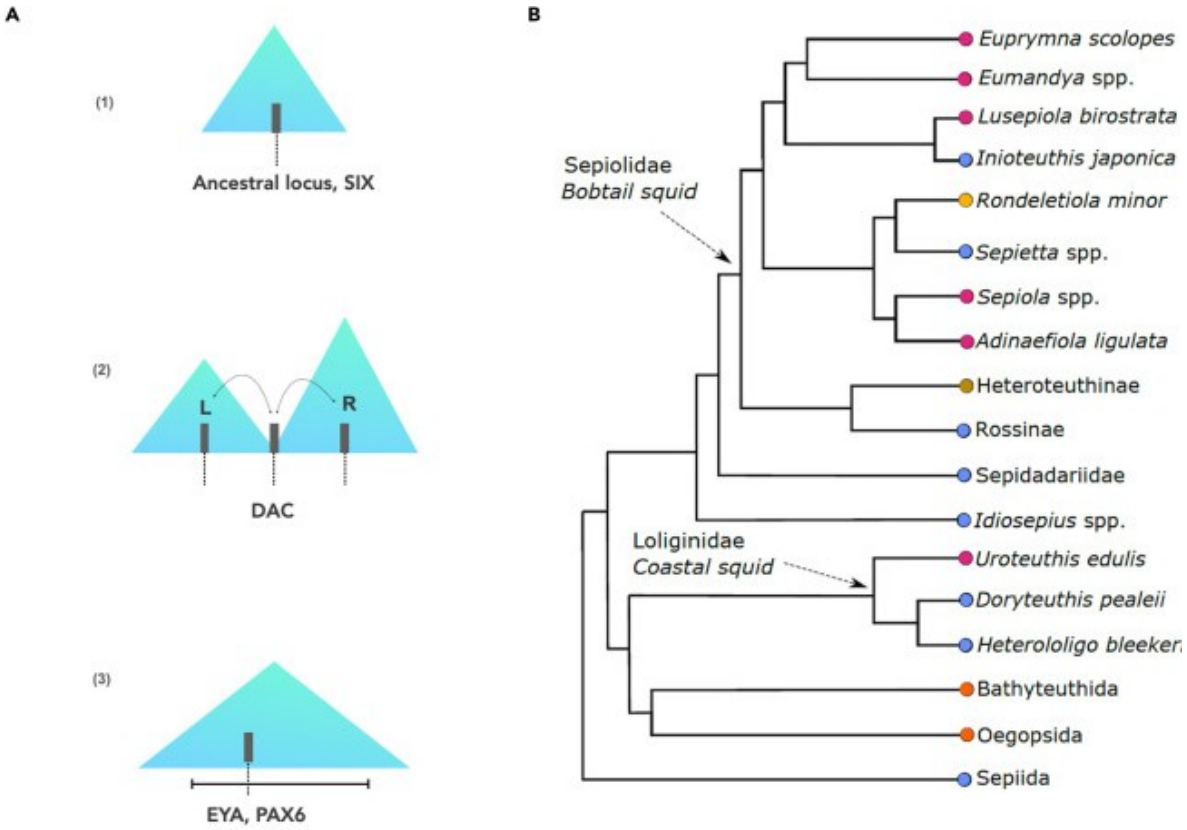
Margaret
McFall-Ngai
Caltech

Continuous emergence of new interactions



Rouressol et al, iScience 2023

Continuous emergence of new interactions



Rouressol et al, iScience 2023

CephASG hub

Margaret McFall-Ngai, U Hawaii

AQUATIC SYMBIOSIS GENOMICS

Spencer Nyholm, U Connecticut

Elizabeth Heath-Heckman, Michigan State University

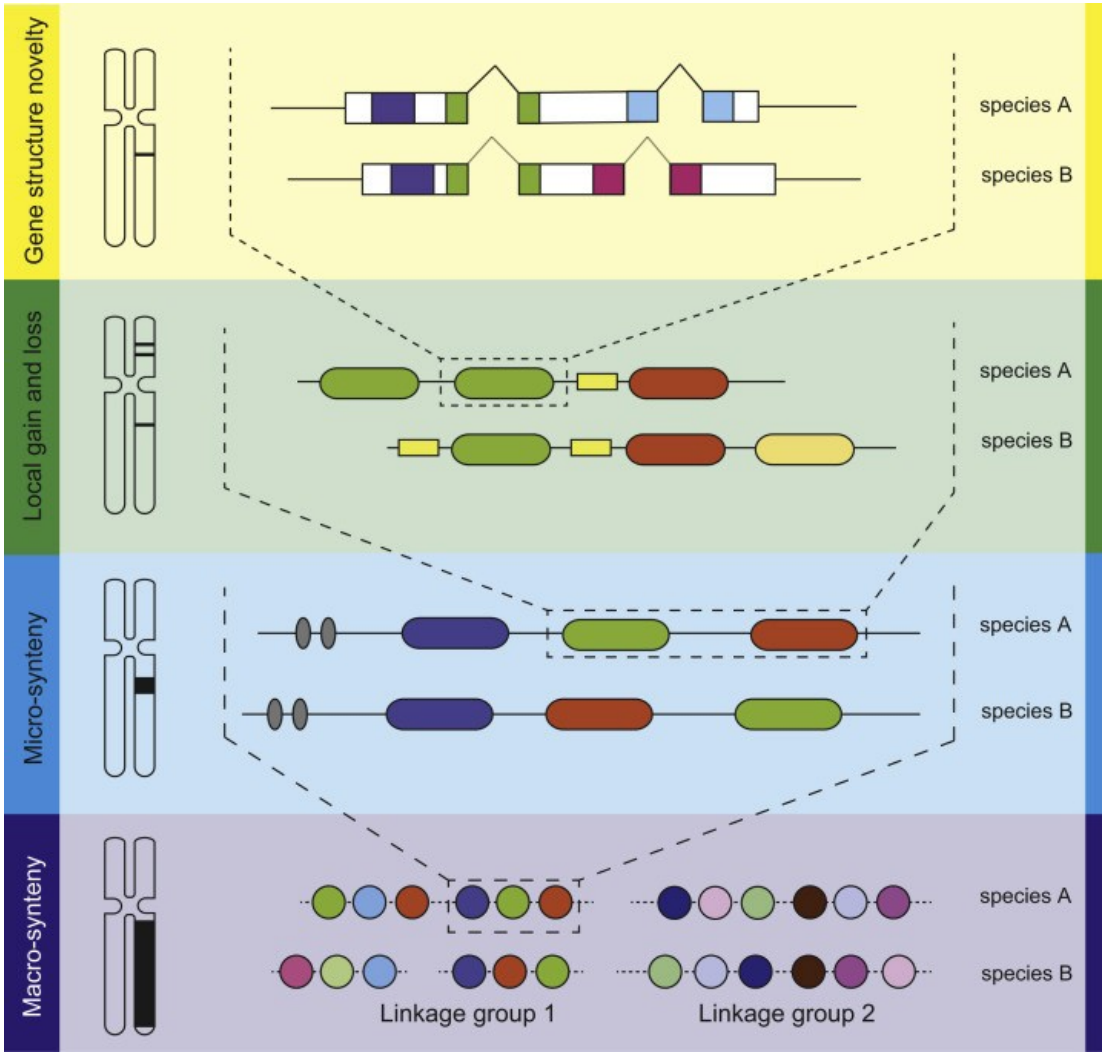
Raphael Lami, Sorbonne University and Marine station of Banyuls

Michele Nishiguchi, UC Merced

Gustavo Sanchez, U Hiroshima

Oleg Simakov, U Vienna

Beyond chromosomal “algebra”: combine different level of orthologies and capture their evolutionary interaction properties



Simakov and Kawashima, 2016

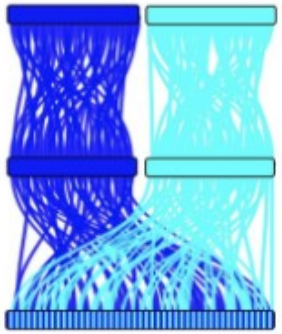
Generalization of FWM

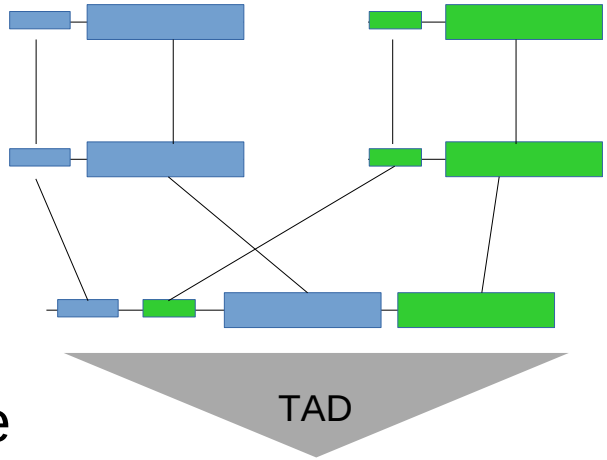
Chromosomes

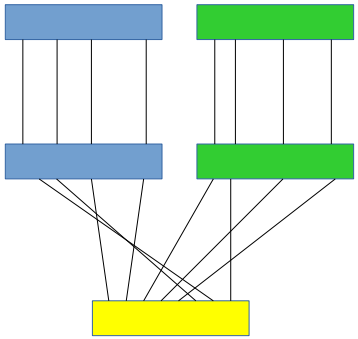
Gene regulation

Protein “domains”

Homology







“New” chromosome

“New” domain

“New” linkage (micro-synteny)

Constraints:

meiotic

regulatory linkage

folding

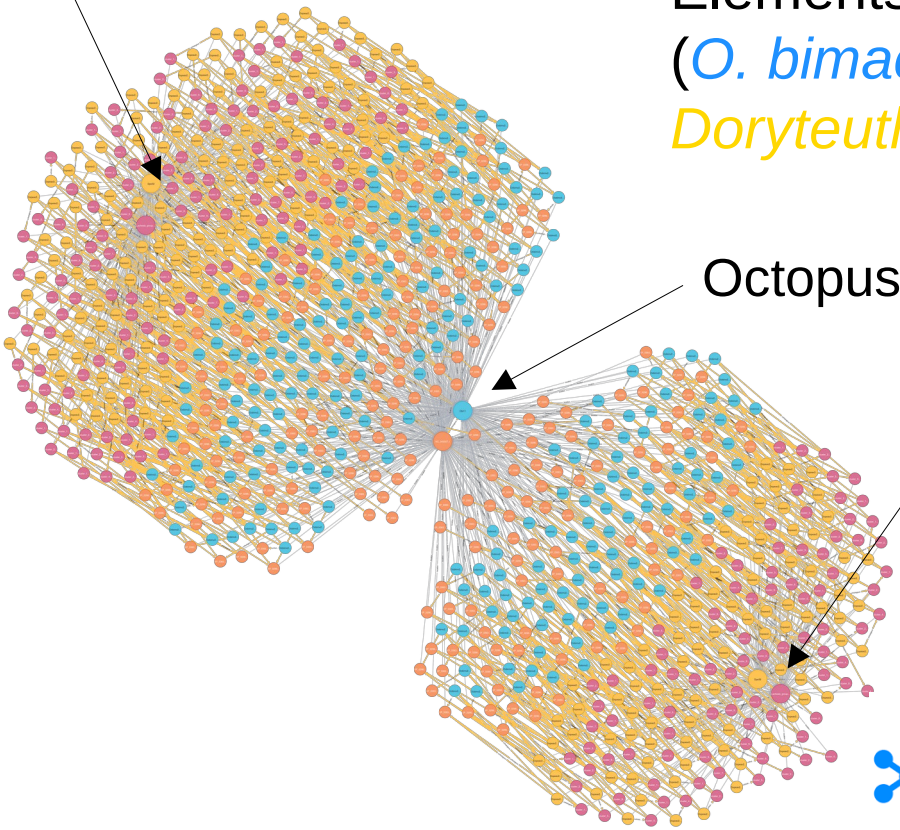
How to visualize orthologies at multiple levels?

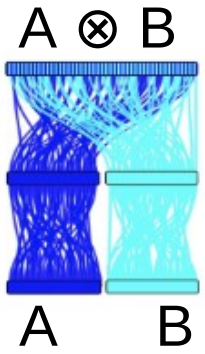
Graph database representation
for orthologous genes and non-coding
Elements for two octopus vs two squids
(*O. bimaculoides*, *O. sinensis*,
Doryteuthis pealeii, *Euprymna scolopes*)

Squid chromosomes (A)

Octopus chromosomes (= A ⊗ B)

Squid chromosomes (B)

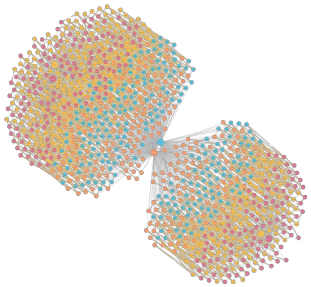




{A} = *Euprymna* LG35; *Doryteuthis* Dpe32; *Octopus* Obi/Osi11
{B} = *Euprymna* LG31; *Doryteuthis* Dpe36; *Octopus* Obi/Osi11

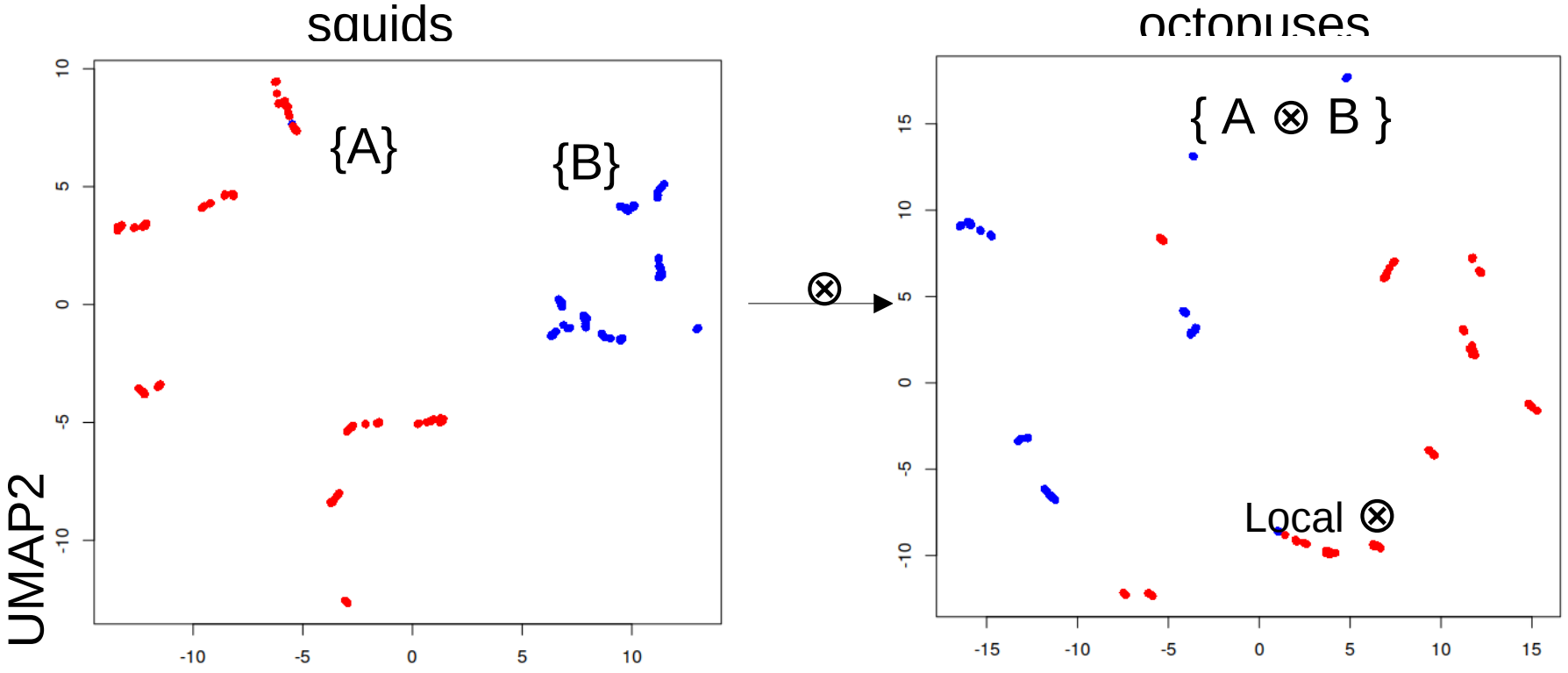
Fatih Sarigöl

Manifold representation of distances – Evolutionary Topology



Averaged distances

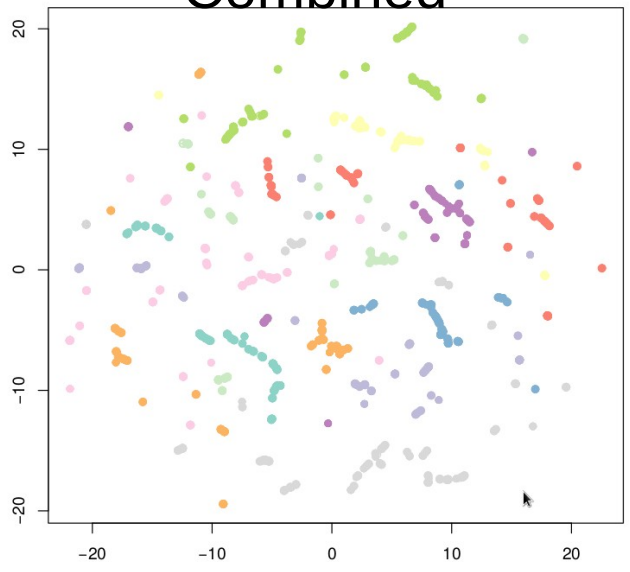
UMAP



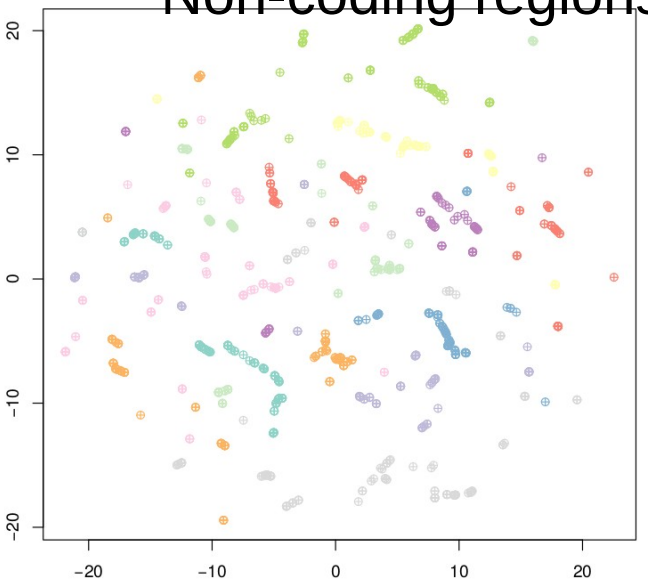
{A} = *Euprymna* LG35; *Doryteuthis* Dpe32; *Octopus* Obi/Osi11
{B} = *Euprymna* LG31; *Doryteuthis* Dpe36; *Octopus* Obi/Osi11

Multi-scale orthologies and paths through manifolds

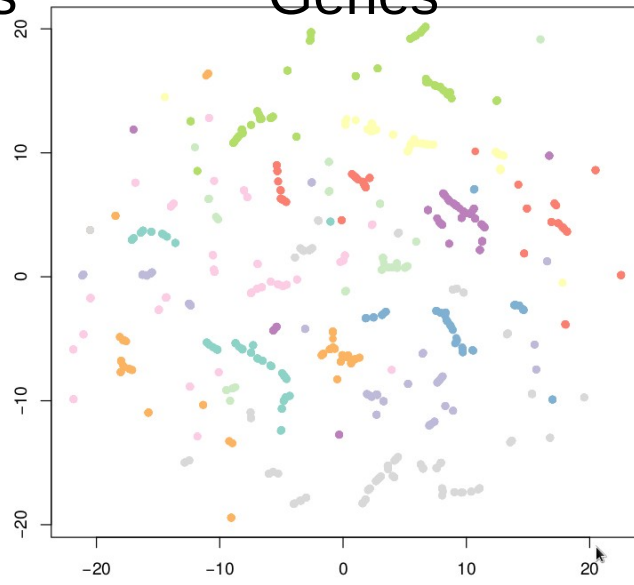
Combined



Non-coding regions

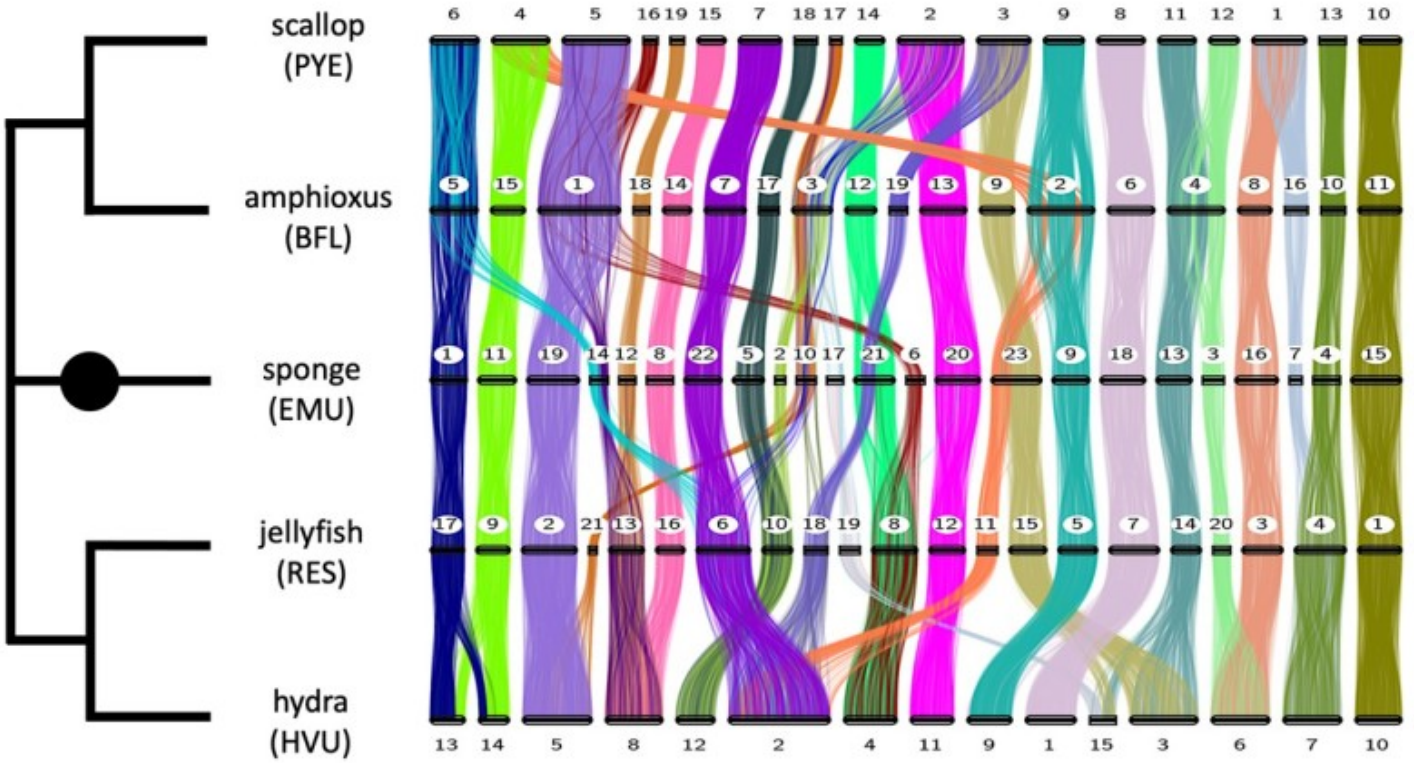


Genes

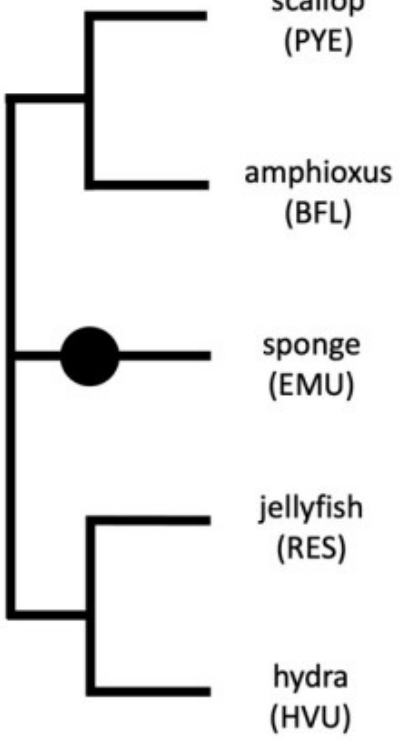


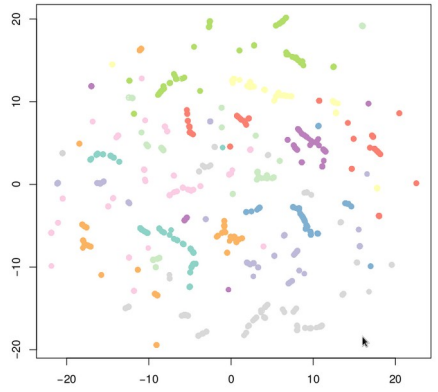
~10 squid chromosomes

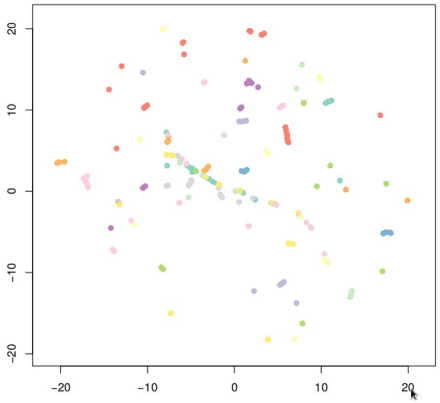
Towards multi-scale orthologies and evolutionary topology



Towards multi-scale orthologies and evolutionary topology







→ identify ⊗ at sub-chromosomal level, including non-coding regions (novel synapomorphic characters)

→ comparative evolutionary topology (detect orthology presence for faster evolving characters provided same topological context)

→ predict possible future interactions (paths through manifold space)

→ non-chromosomal context (FWM with transposon insertions)

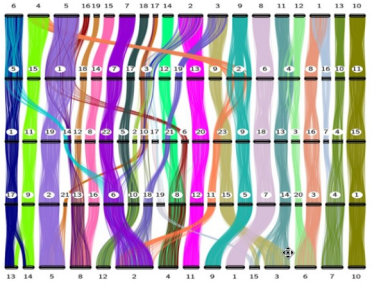
Schultz et al, in preparation
Kon et al, in preparation

Recap: evolutionary topology of animal genomes?

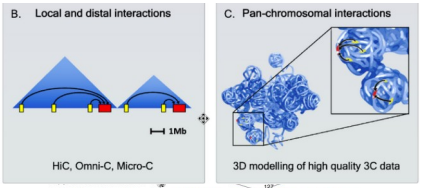
How to quantify (macro-)evolutionary trajectories of genomes?

Do different trajectories contribute to diverging (regulatory) modes of genome evolution?

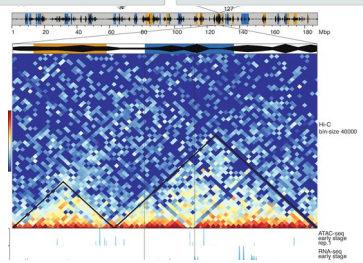
- *Genome "architecture" in evolution (synteny et al)*
  Syntenic elements are highly preserved (including some sub-chromosomal linkages)

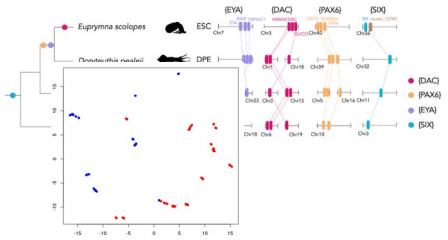


- *Regulatory function vs meiotic constraints*
  Selective pressure to maintain (mostly local) chromosomal interactions



- *"Transitions" in genome organization onto novel regulation*
  Genome-wide rearrangements create new regulatory neighborhoods to be explored





- *Evolutionary topology: emergence of novel (regulatory) gene neighborhoods*

Thank you for your attention!

Graduate students:
Lisa Rouressol
Koto Kon

Postdocs:
Thea Rogers
Darrin Schultz
Tetsuo Kon

Technician:
Fatih Sarigoel

MSc students:
Laura Johanesova
David Stohlmann
George Efthymiadis

BSc students:
Arno Bluemel
YehorTertyshnyk

Admin: Nina Znidaric

Alumni: Nicola Wong, Hannah Schmidbaur, Nicolas Robert, Christina Holzinger, Elena Ritschard, Dalila Destanovic, Gozde Yalcin

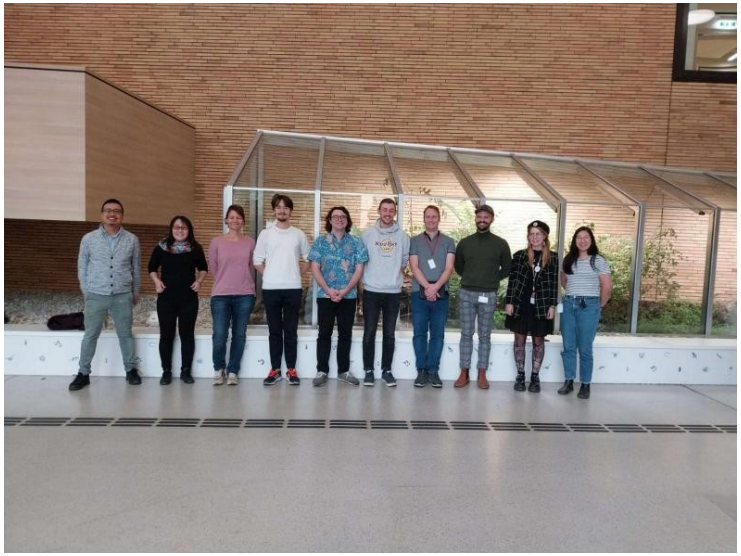
molecular evolution & development

TIERGARTEN SCHÖNBRUNN

FWF ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

Marine Biological Laboratory
THE UNIVERSITY OF CHICAGO

erc



(soon™ looking for postdocs)

I was thinking to talk along the lines of a historical-ish perspective on the (somewhat) early days of animal comparative genomics, the questions and early predictions (ancestral complexity of animal genomes, synteny retention etc), and their validation after the transition to the chromosomal-scale world.
This can be followed by new approaches and discussion if there are any new technical bottlenecks for comparative field beyond the current continuous haplotype-resolved chromosome-scale sequencing efforts (maybe along the lines of 3D genomics, single-cell/tissue resolved genomics etc).
I could end up by presenting some of the work we have been up to in the recent years and where we think it leads.

One potentially interesting and technical aspect is data exploration across different homologies in the genome (chromosome, genes, CNEs etc), for that we have been