Functional innovation through new genes as a general evolutionary process

Published on Nature Genetic at 2025.Feb

Guangji Chen 2025.Apr.25

Article information

nature genetics

Review article

https://doi.org/10.1038/s41588-024-02059-0

Functional innovation through new genes as a general evolutionary process

Received: 12 June 2024

Accepted: 15 December 2024

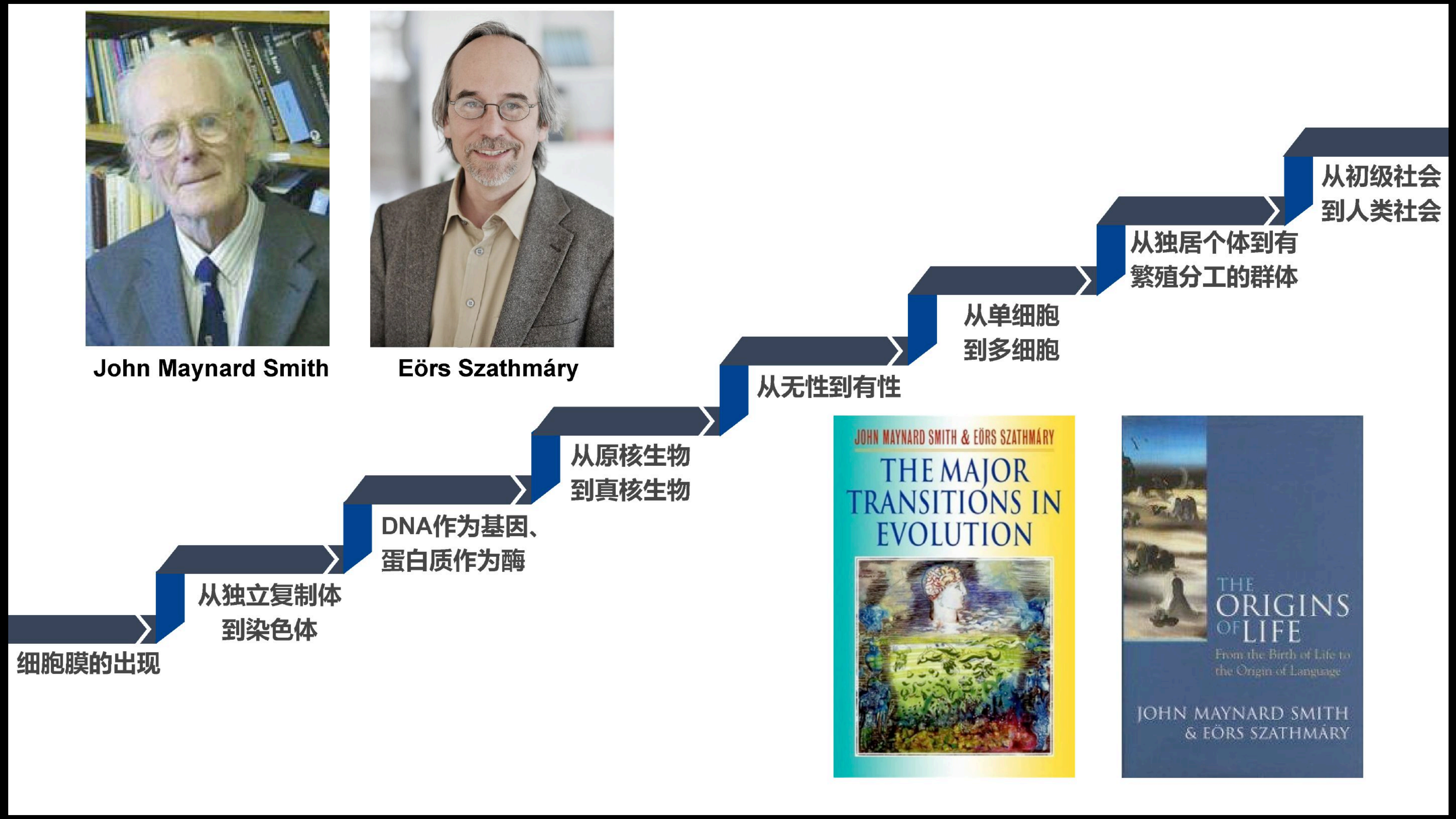
Shengqian Xia[1], Jianhai Chen[1], Deanna Arsala[1], J. J. Emerson[2] & Manyuan Long[1]

Manyuan Long (Left)
龙漫远 (左一)

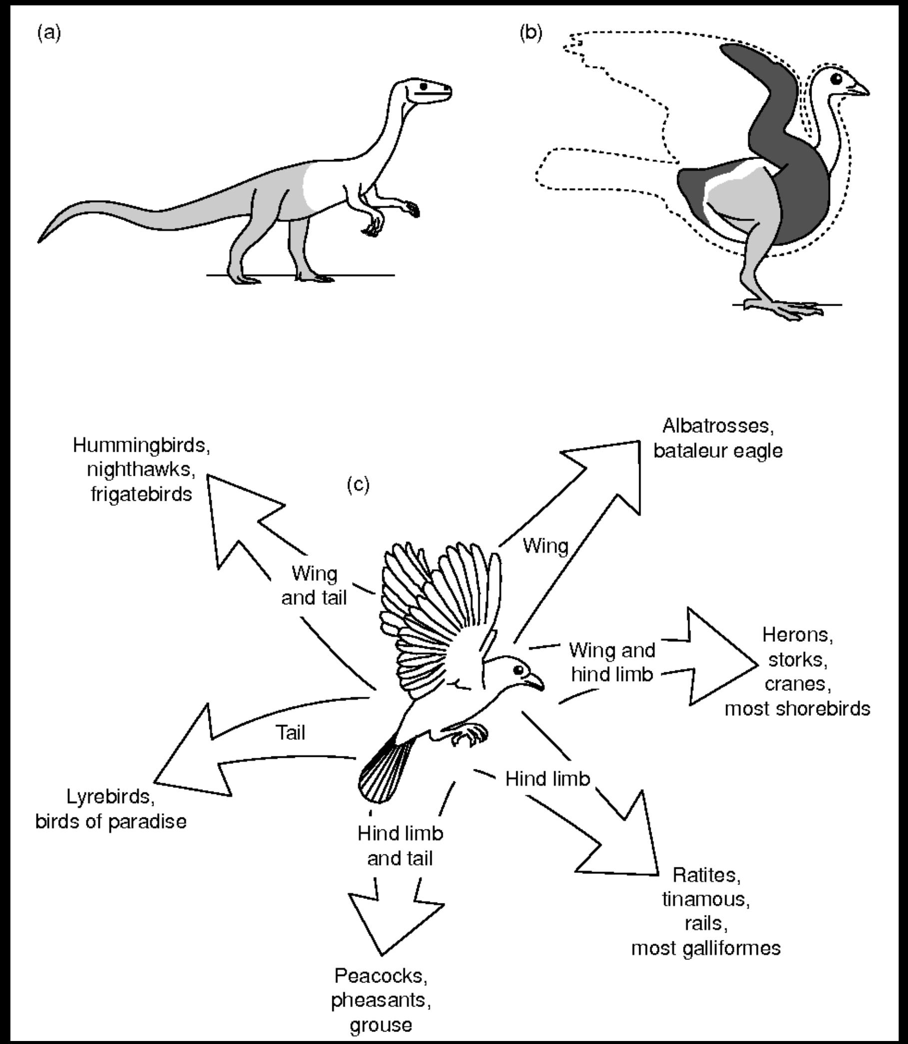
Functional innovation ⟷ **New genes**

Background

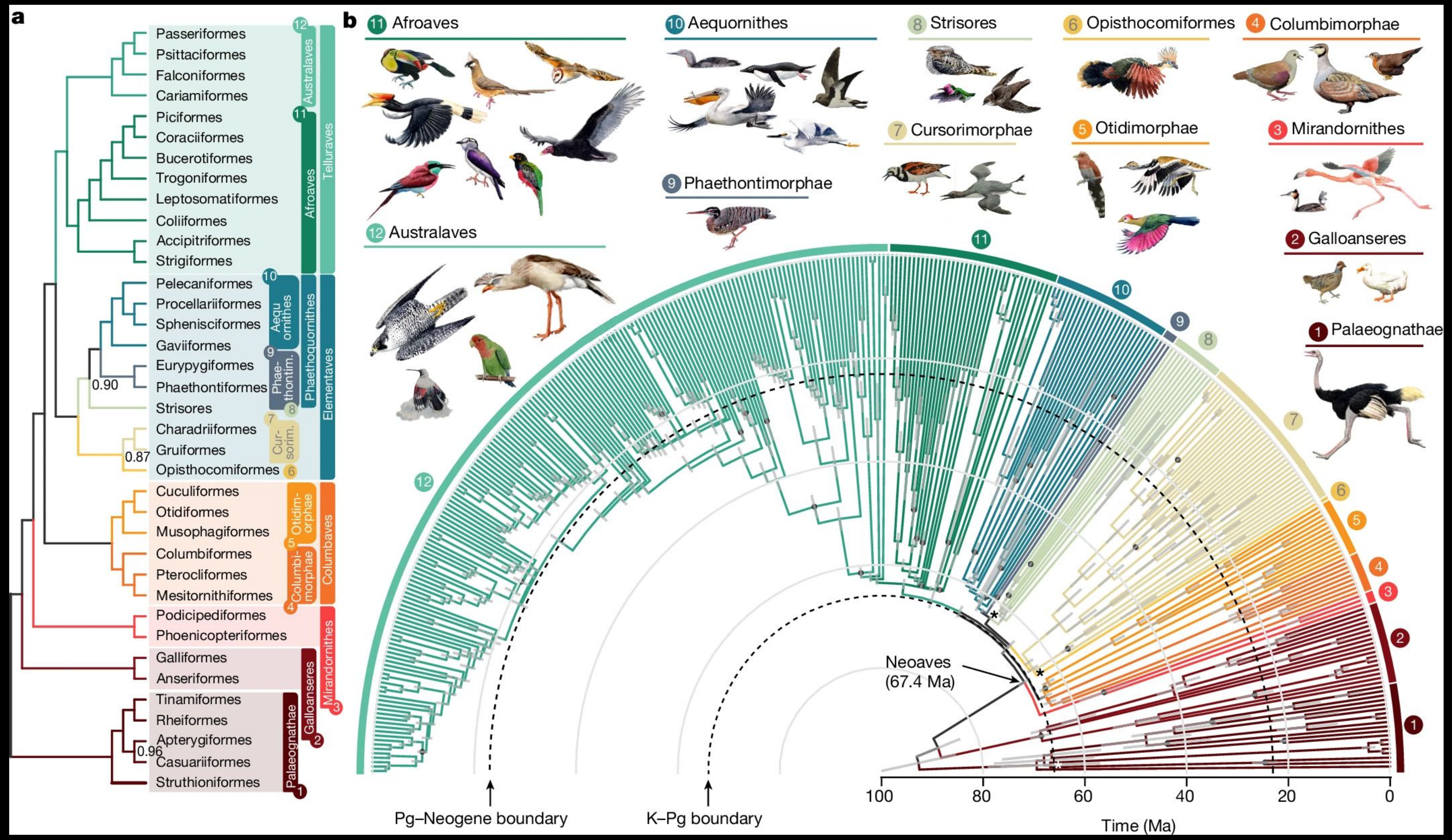
Why do I want to share this article?



8 major transitions in evolution (@*Xiaoxu Xu*)



innovation



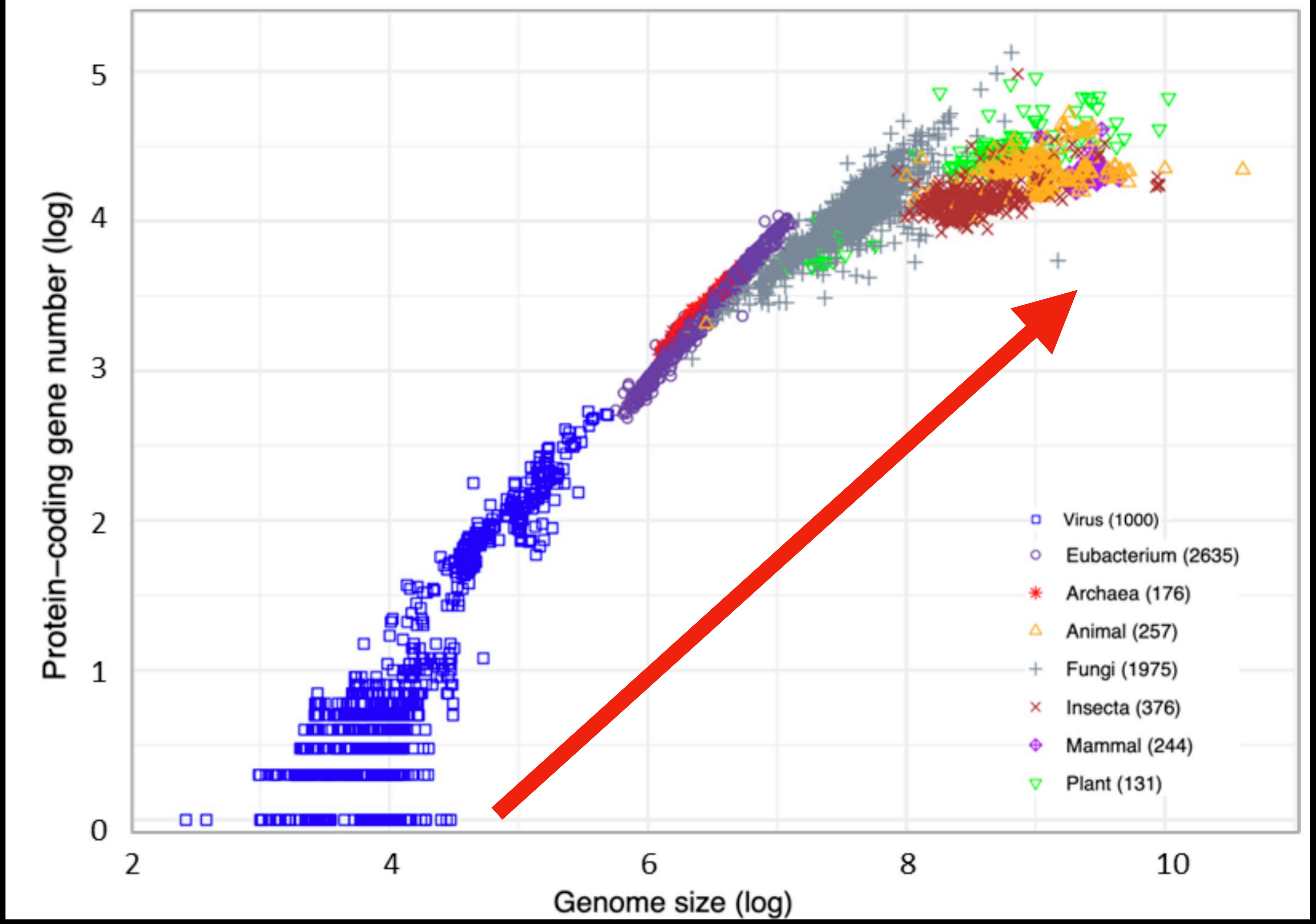
radiation evolution & diversity

Functional innovation

Background

Gene numbers vs. genome sizes

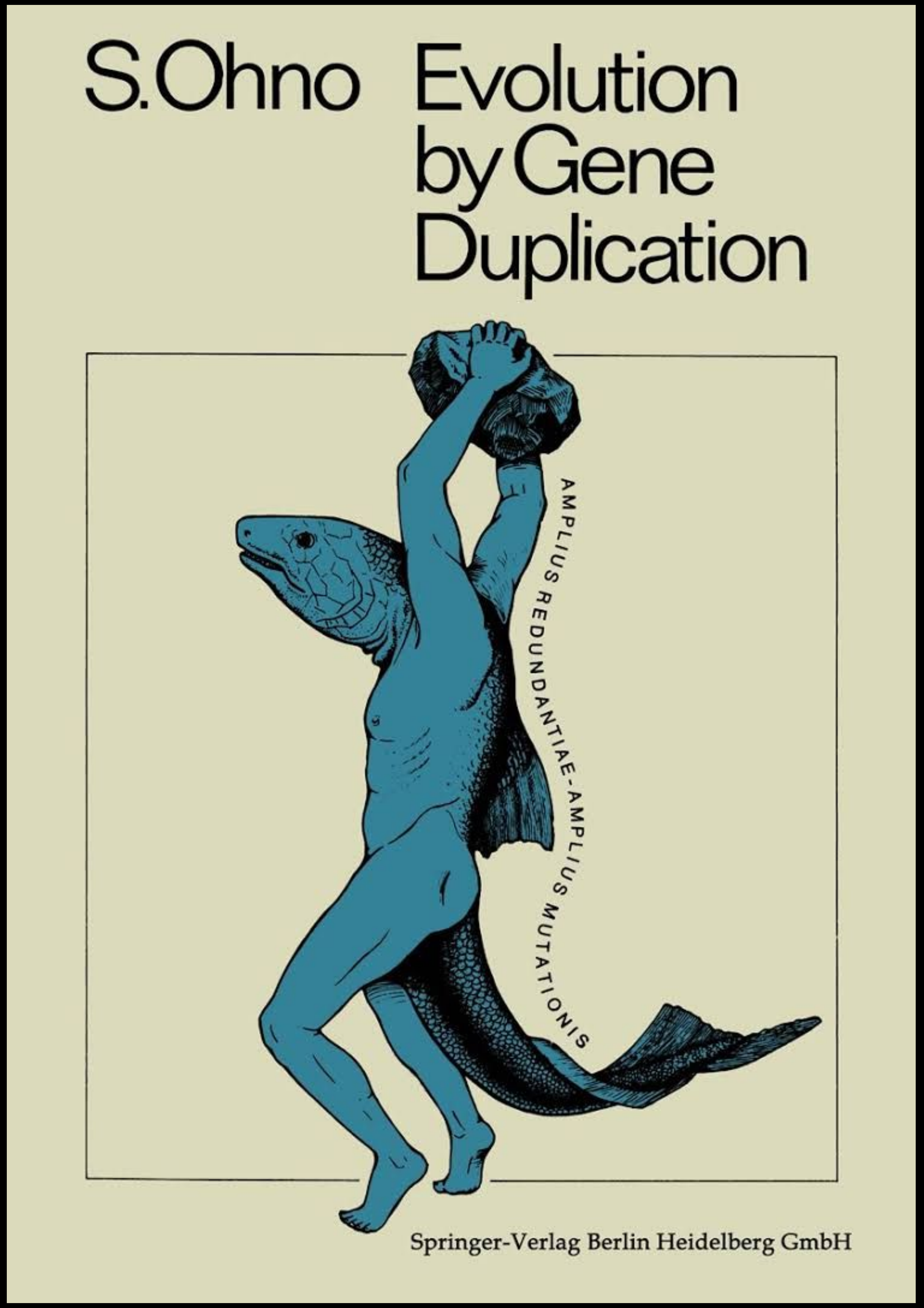
- How do new genes birth?



The distribution of gene numbers and genome sizes across 6,794 extant species

Background

A classic view: Gene duplication

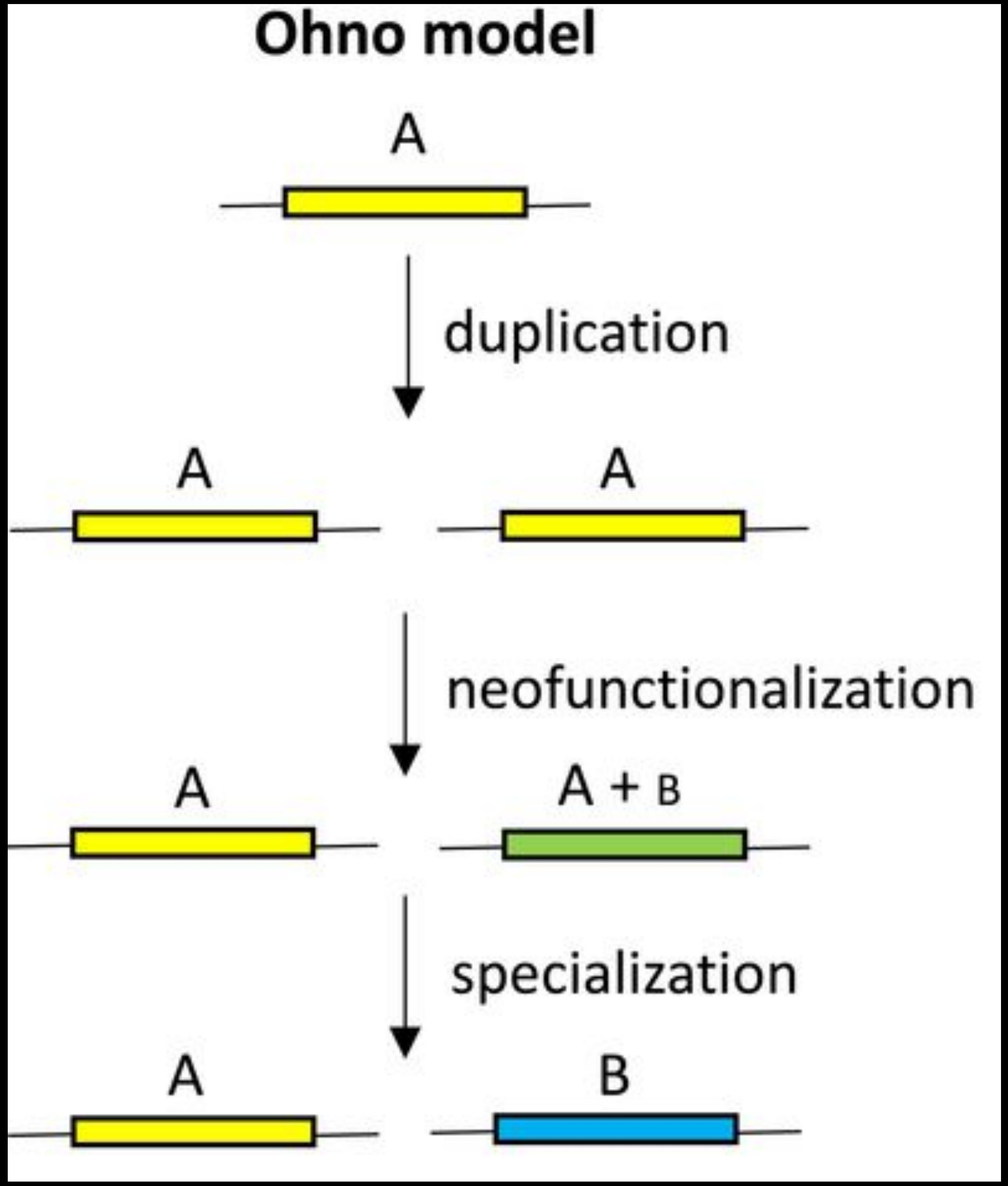


S.Ohno Evolution by Gene Duplication

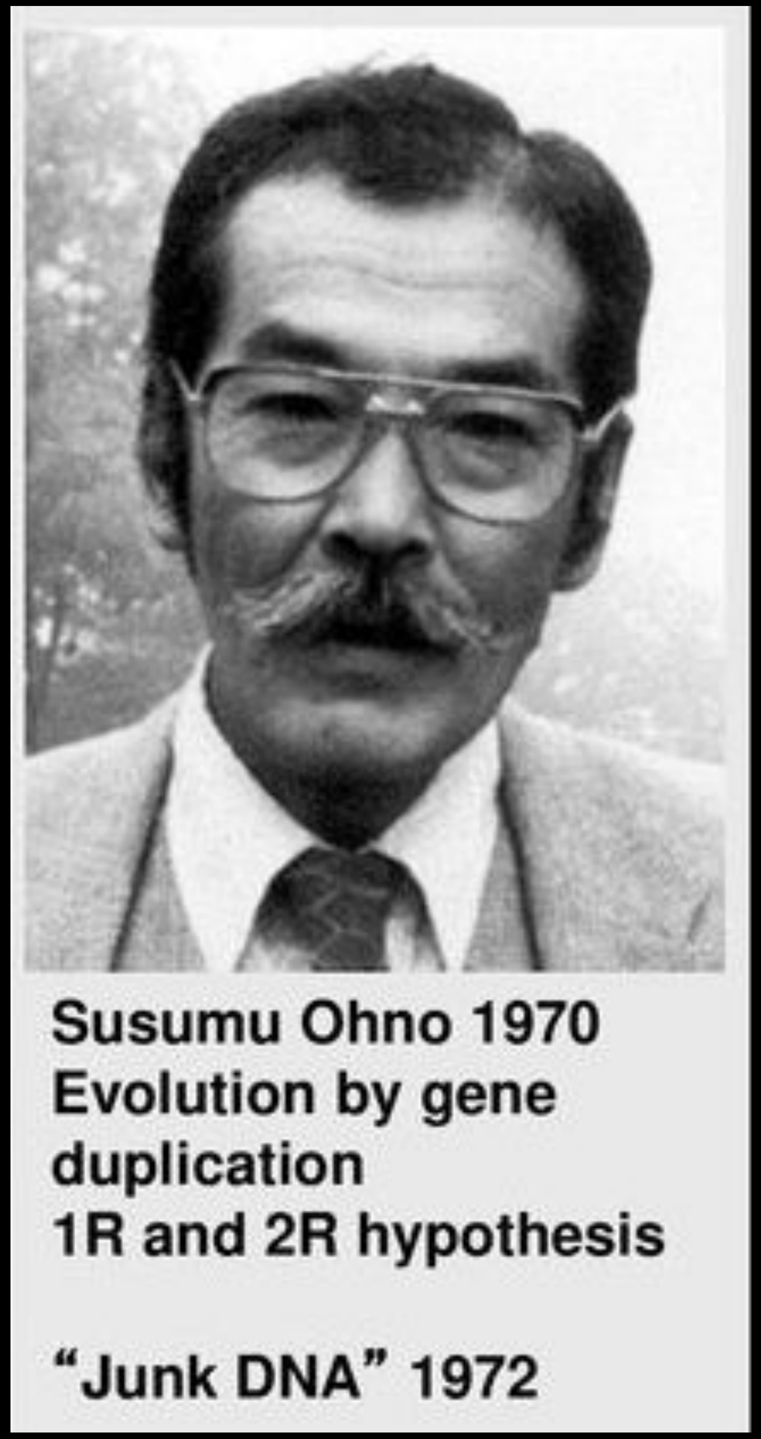
AMPLIUS REDUNDANTIAE-AMPLIUS MUTATIONIS

Springer-Verlag Berlin Heidelberg GmbH

Published in 1970



Gene duplication creates redundancy, the redundant copy accumulates beneficial mutations which provides fuel for innovation



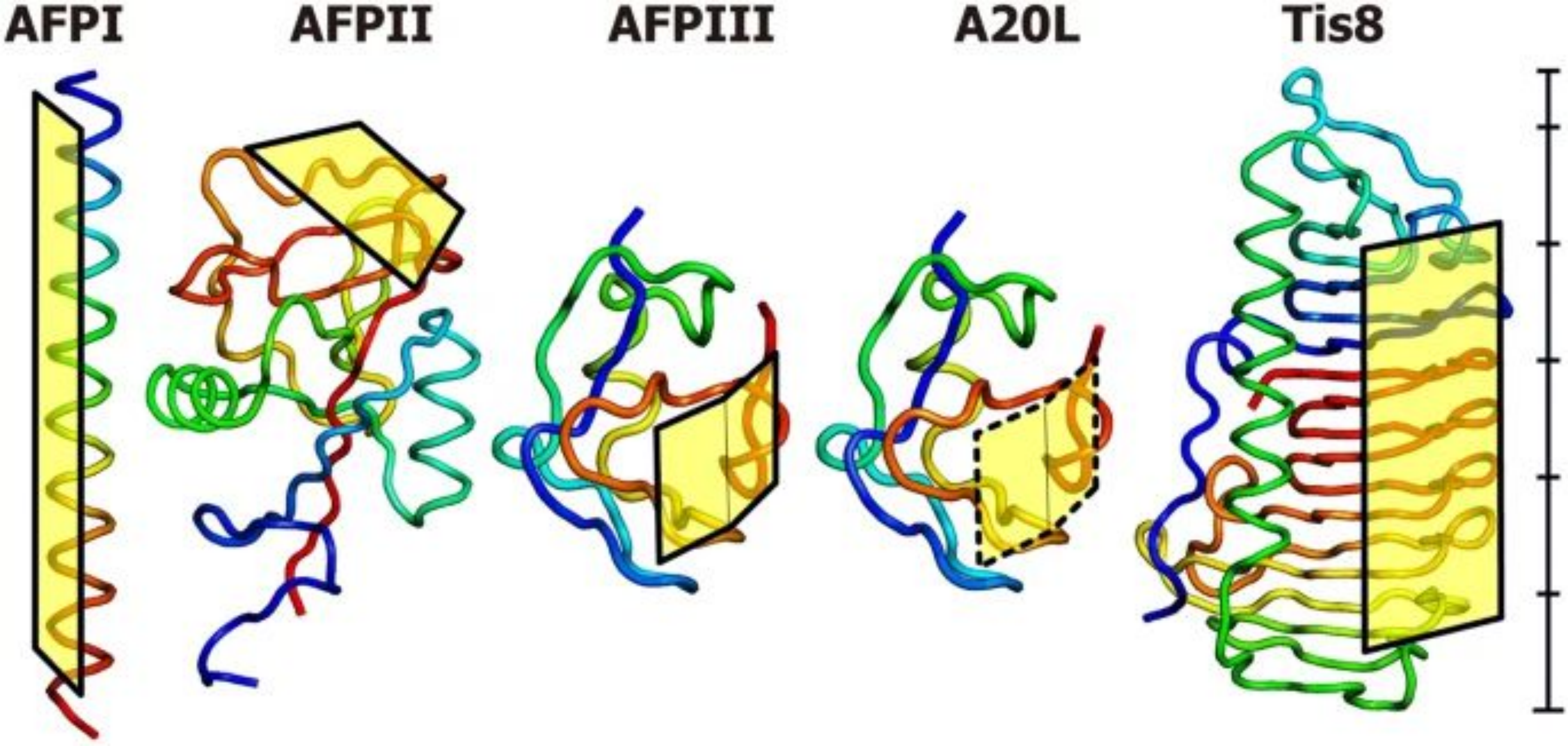
Susumu Ohno 1970
Evolution by gene duplication
1R and 2R hypothesis

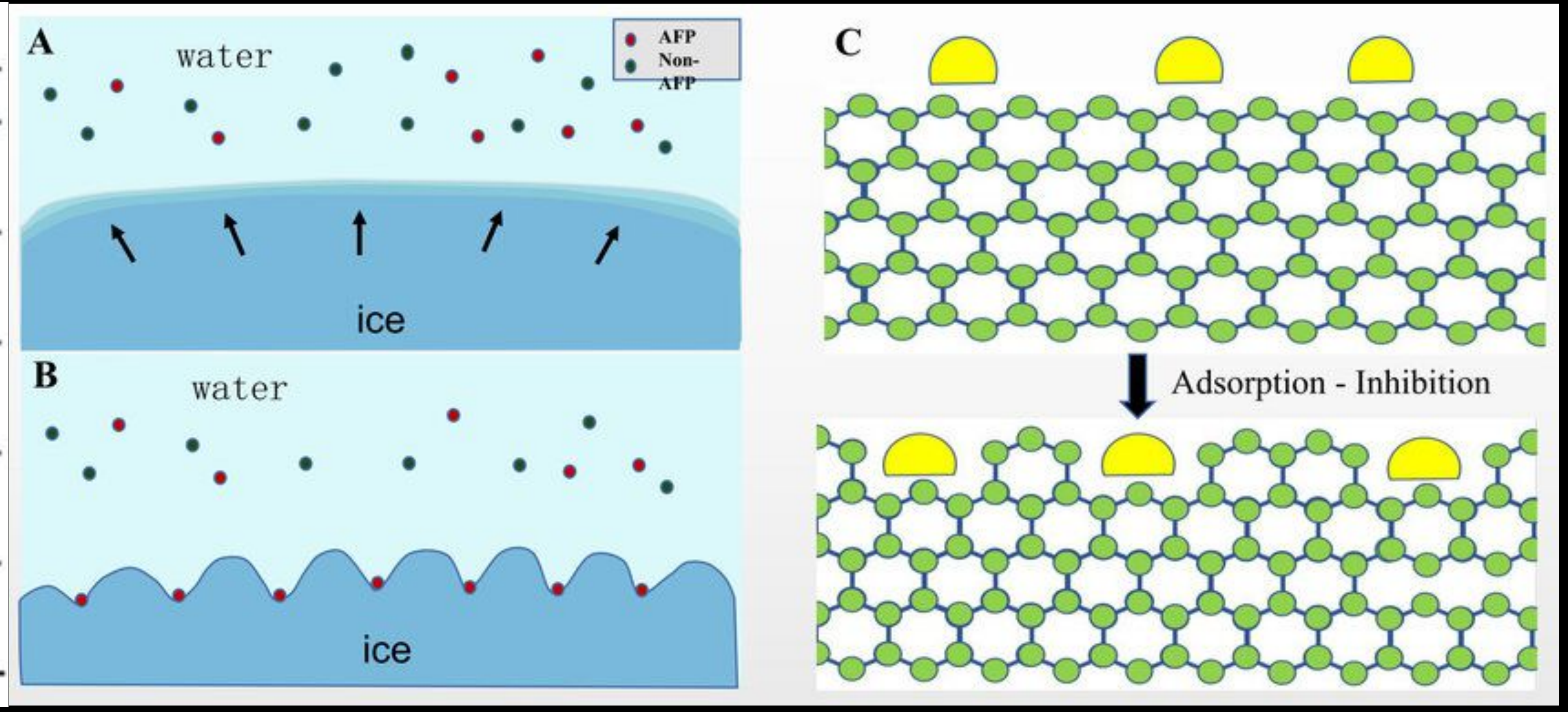
“Junk DNA” 1972

大野乾

Examples of Functional innovation via New gene

1. Antifreeze Proteins (AFPs) 抗冻蛋白





Principle:

- AFPs specifically bind to ice crystal surfaces
- Stops or slows further ice crystal growth
- Prevents ice recrystallization
- Protects cells and tissues from freezing damage

0 °C to 11 °C
even -2 °C

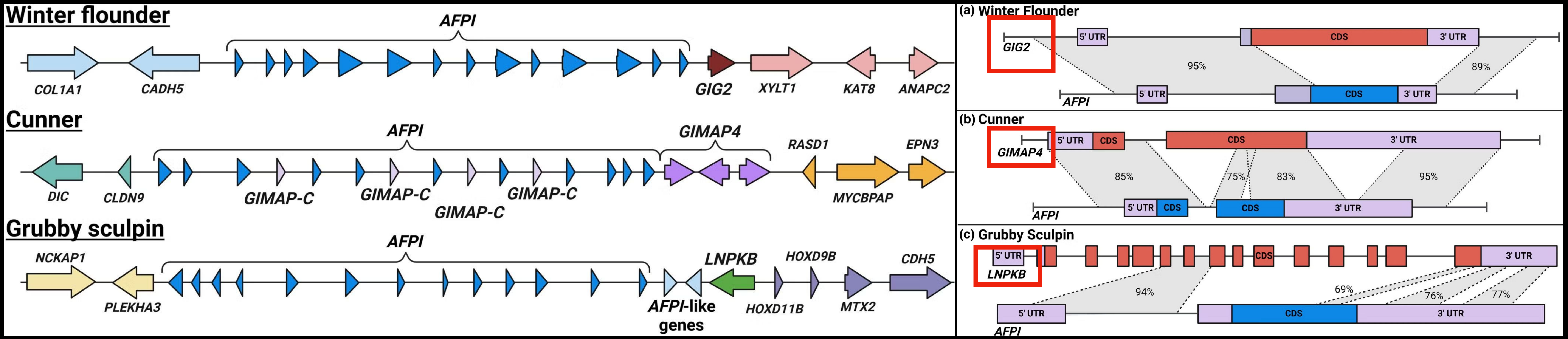
Winter flounder

0 °C to 25 °C

Cuuner

0 °C to 21°C

Grubby sculpin

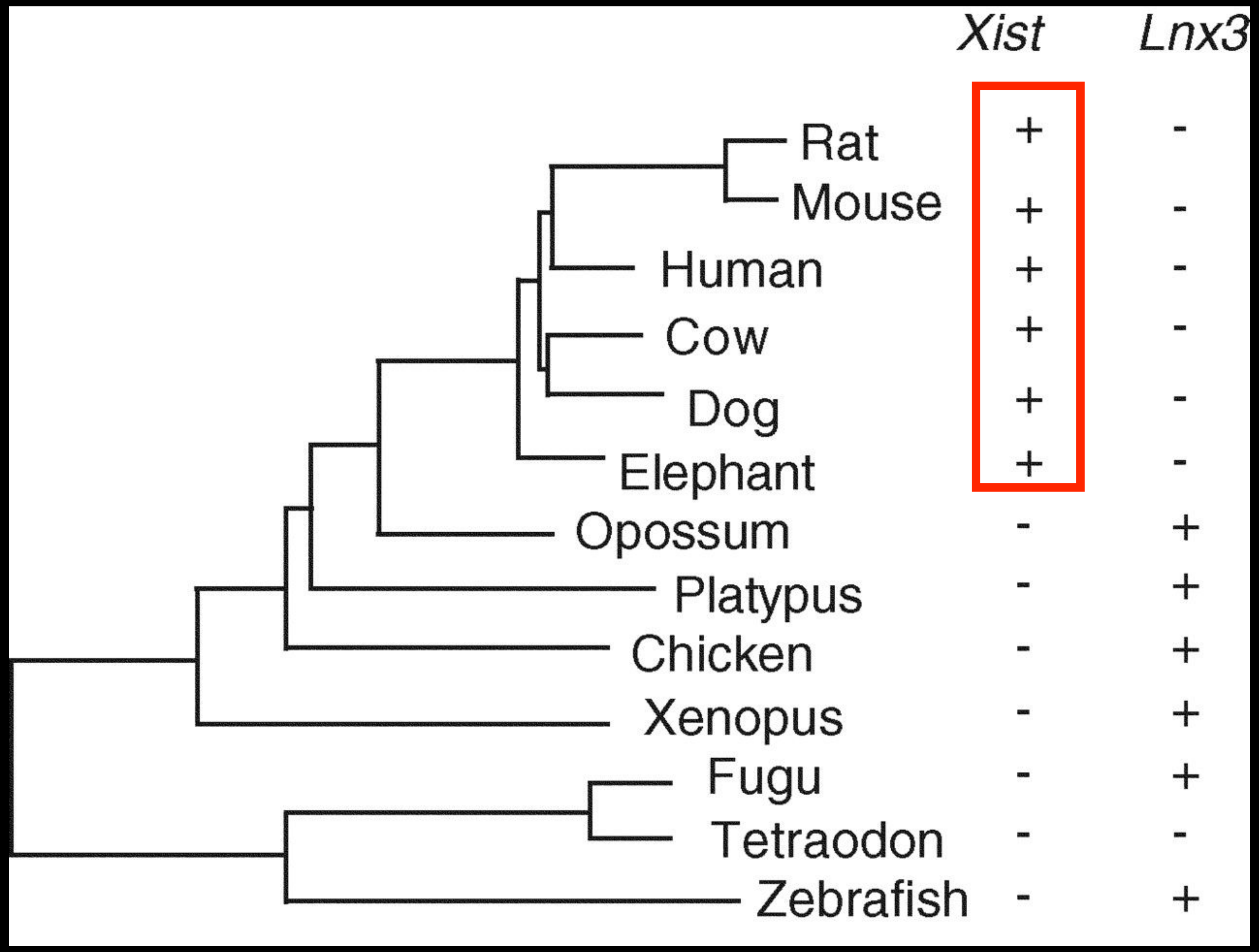


Different way of new genes *AFPI* brith for "*Antifreeze*"

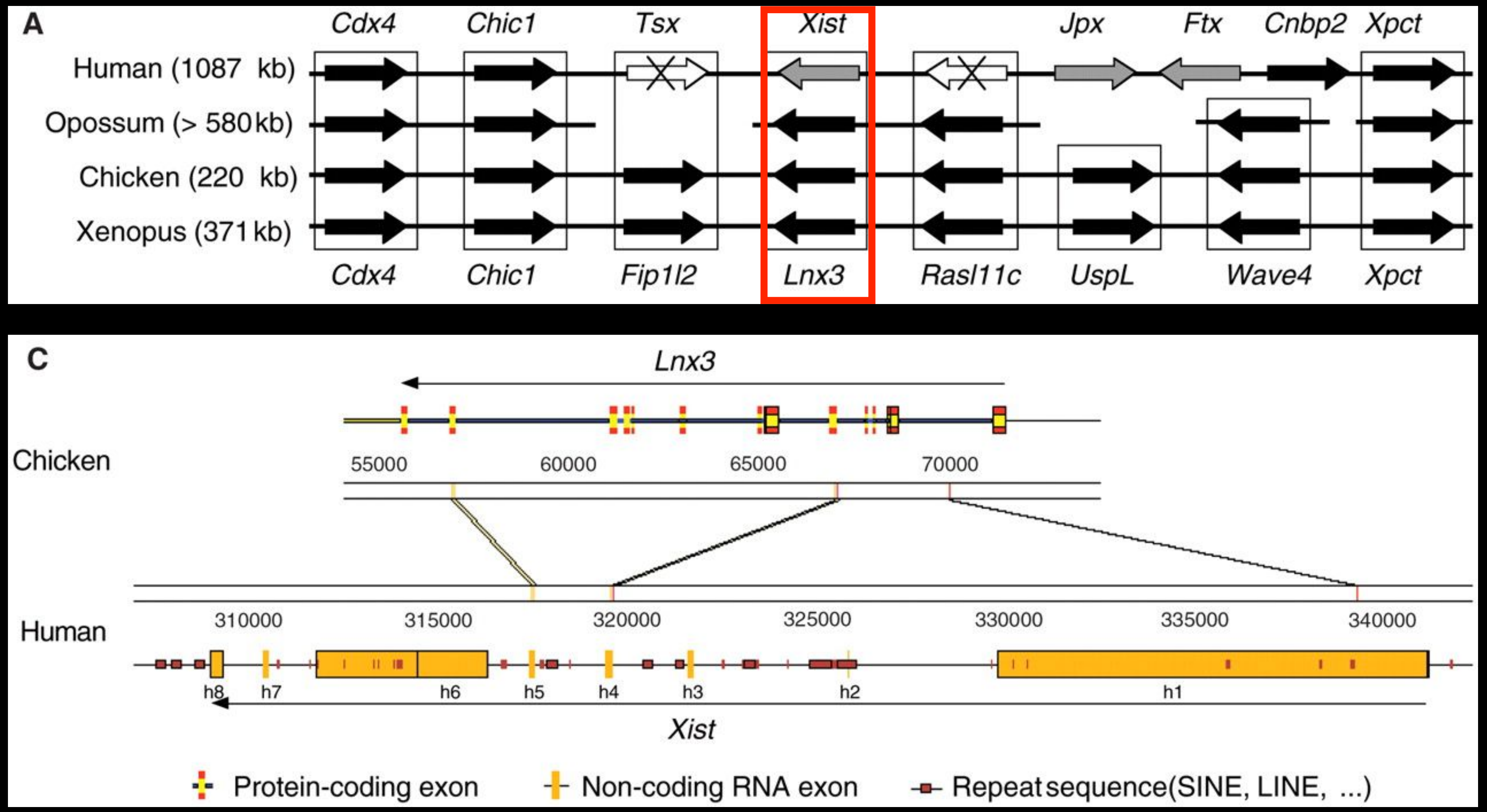
Rahman, A.T. et al. 2019.
Rives, N. et al. 2024.

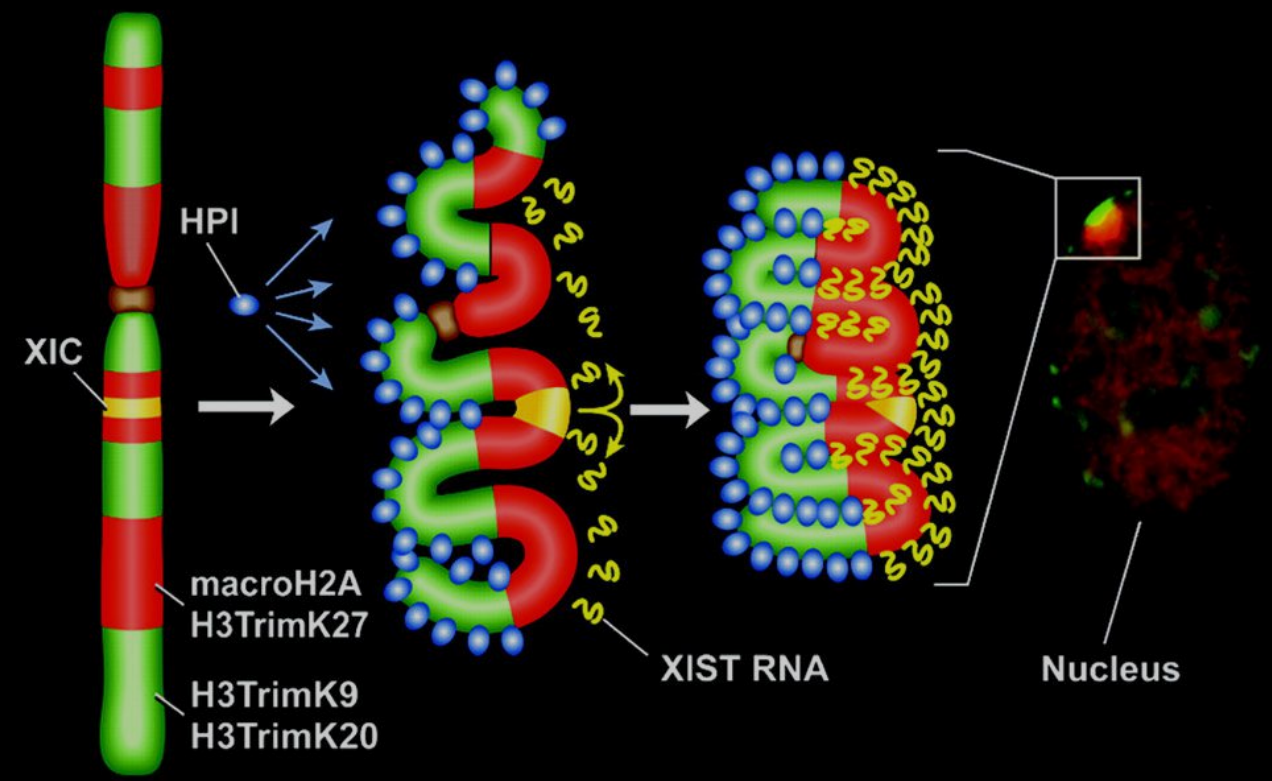
Examples of Functional innovation via New gene

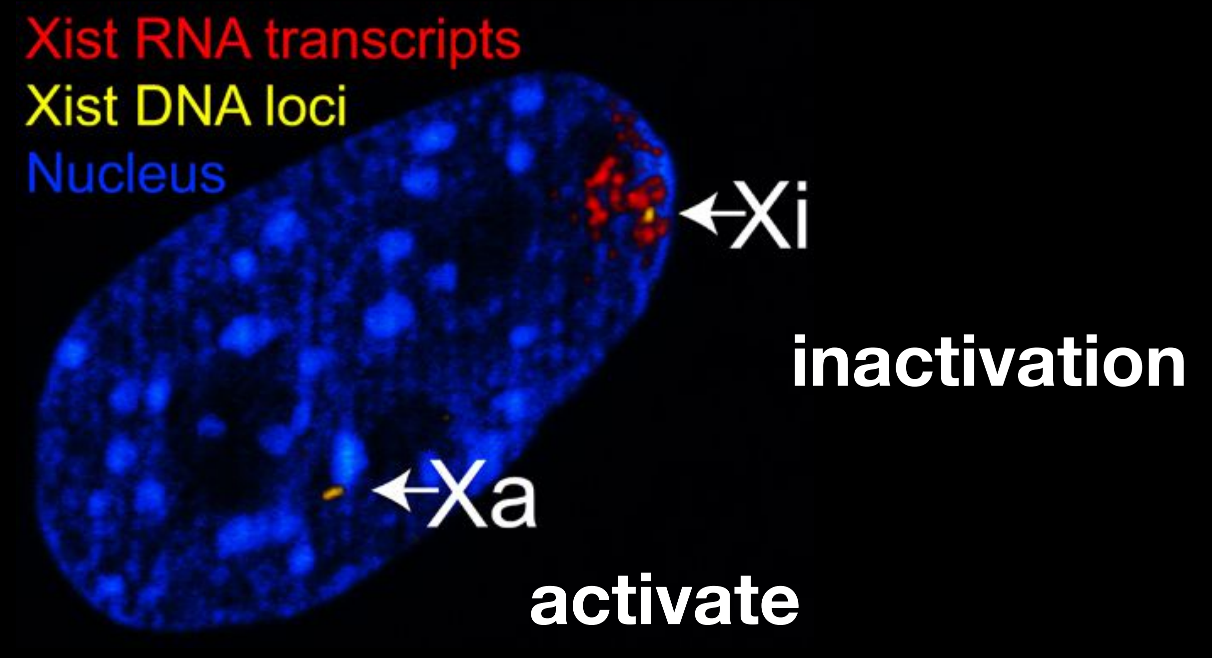
2. X-inactive specific transcript (*Xist*) in Eutherians[真兽类]



	Xist	*Lnx3*
Rat	+	-
Mouse	+	-
Human	+	-
Cow	+	-
Dog	+	-
Elephant	+	-
Opossum	-	+
Platypus	-	+
Chicken	-	+
Xenopus	-	+
Fugu	-	+
Tetraodon	-	-
Zebrafish	-	+







Mechanism:

- **Pseudogenization [假基因化]**

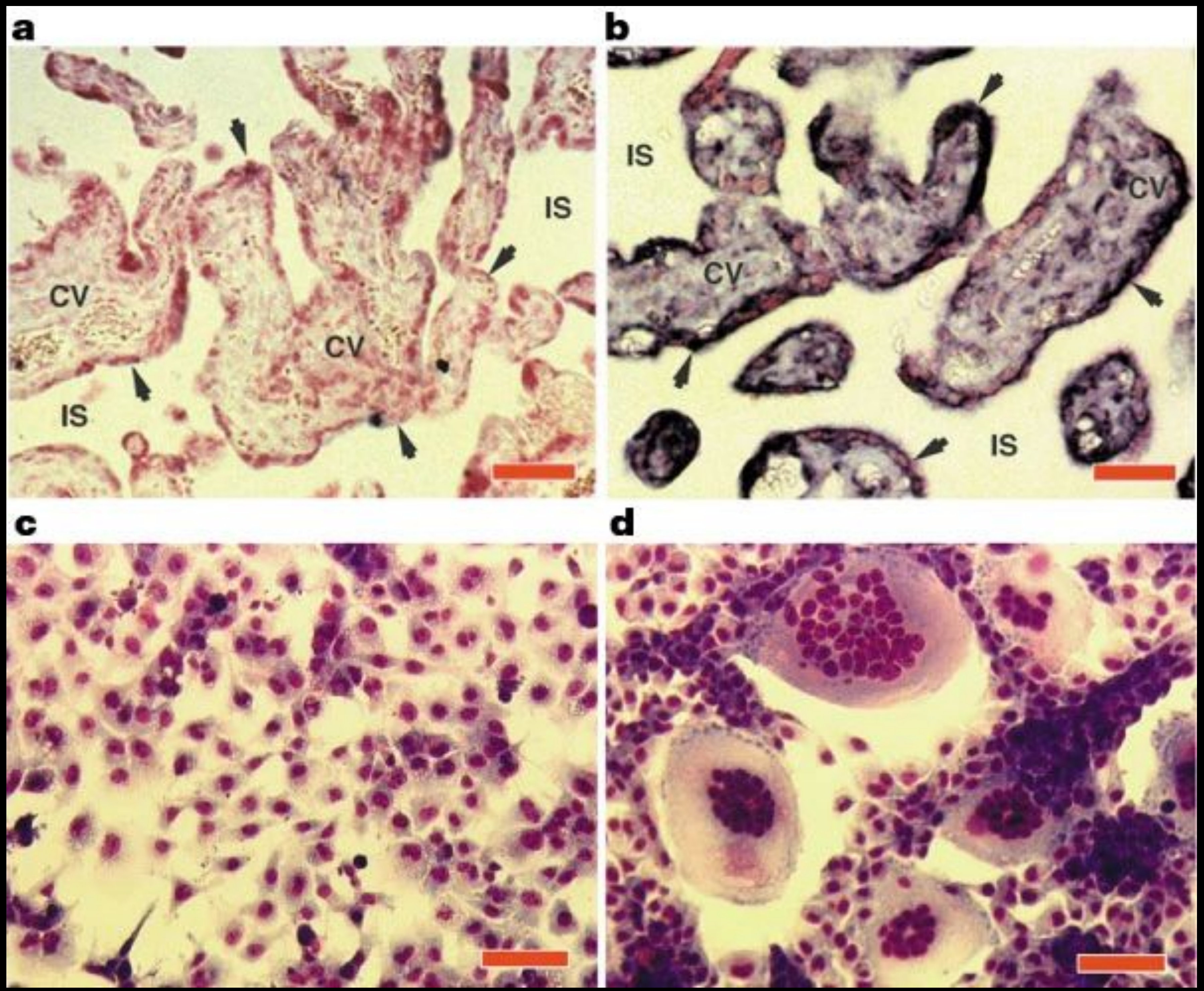
Function:

- **X-chromosome inactivation**

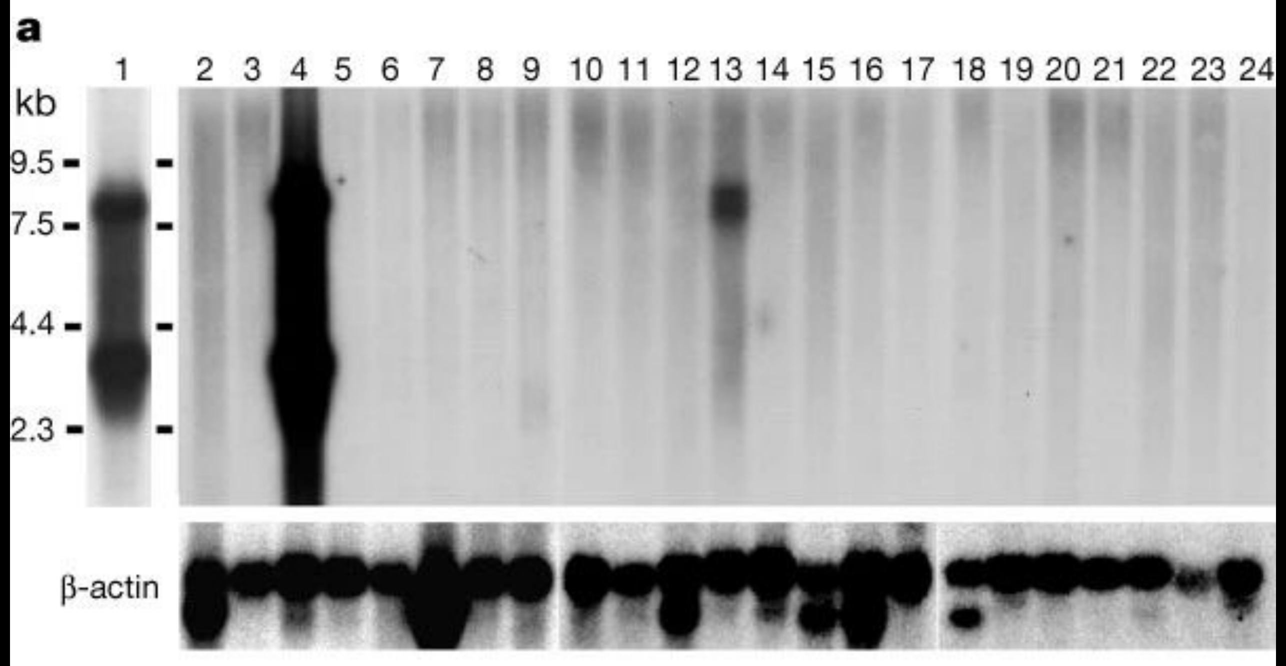
Laurent D. et al. 2006.

Examples of Functional innovation via New gene

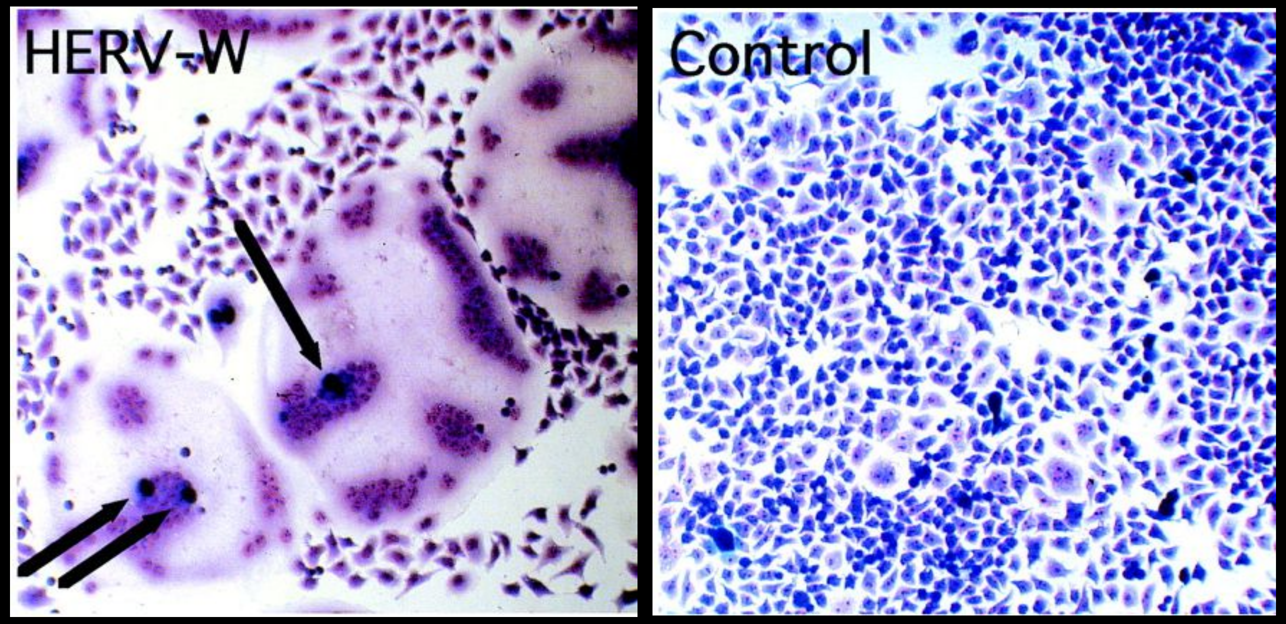
3. *Syncytin 1* (*ERVW1*) in *Homo sapiens*



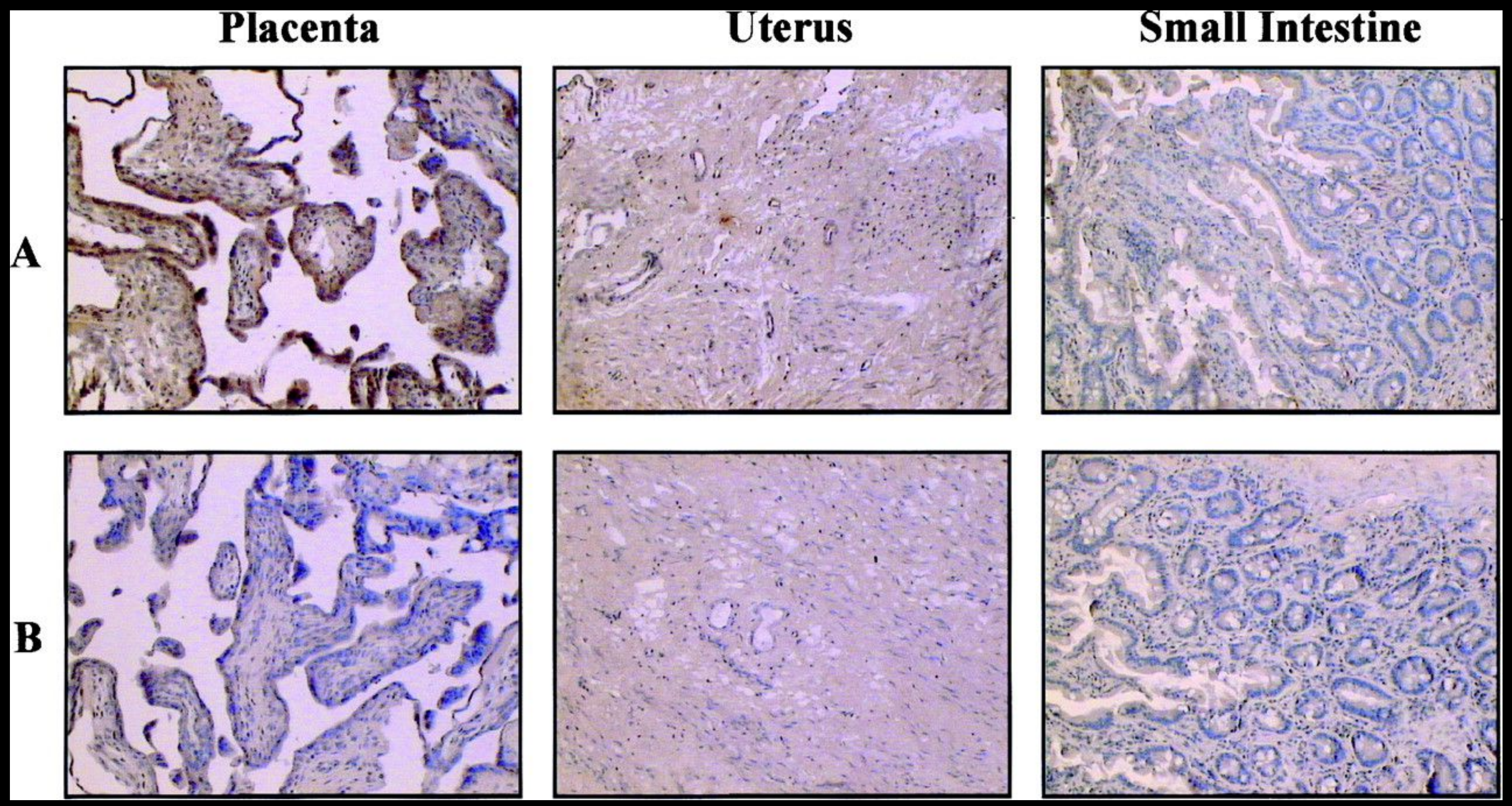
a-**c** negative control
d *synctin 1* in COS cells



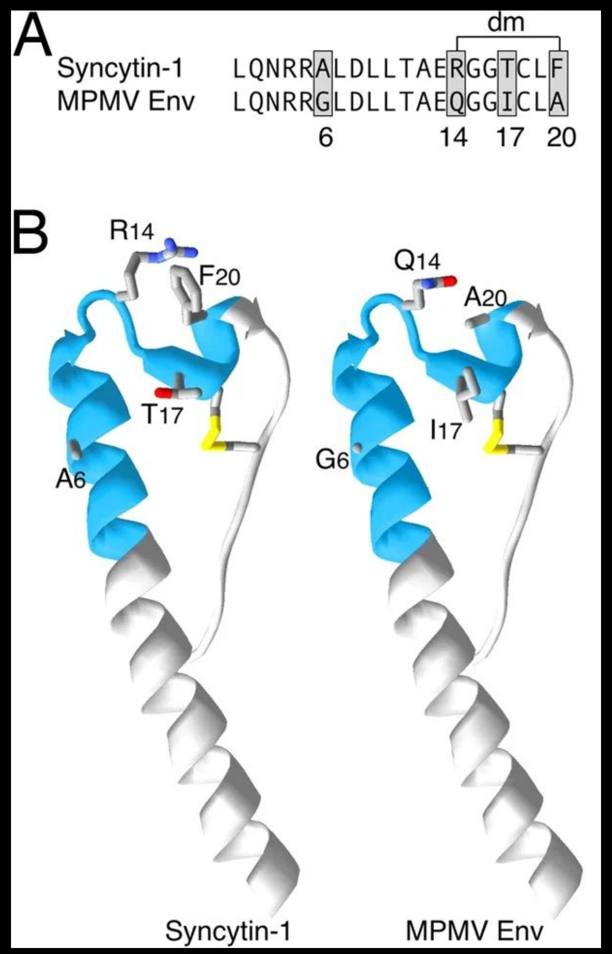
Synctin 1 expression in placenta (4).



Formation of syncytia by *HERV-W* envelope glycoprotein.

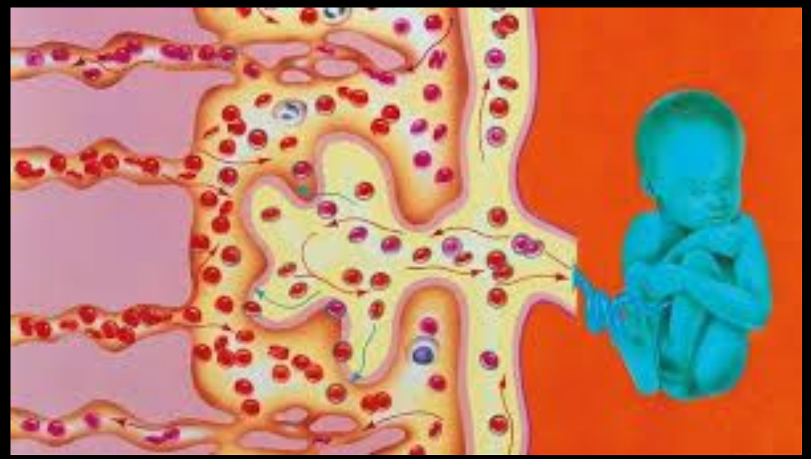


In situ detection of *HERV-W* envelope glycoprotein in tissue sections



synctin 1

MPMV Env



合胞素

Mechanism:

- ***Retroviral* genes**

Function:

- **Human placental morphogenesis**

Mi, S. et al. 2000.
Blond, J. L. et al. 2000.

Outline

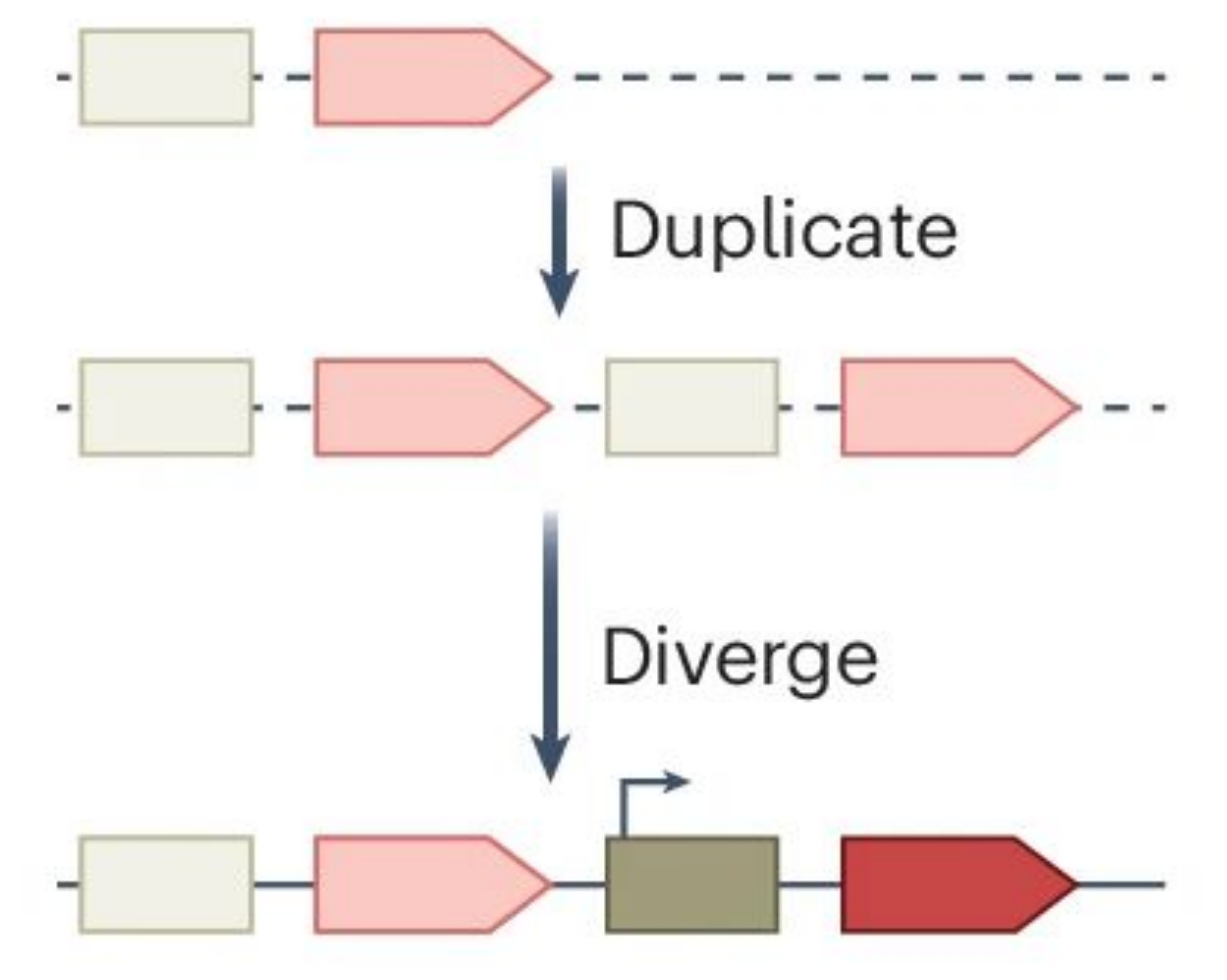
1. Molecular mechanisms to generate new genes
    - *De novo* gene generation

2. How did new genes evolve?
    - Evolutionary forces acting on new genes
    - Rate of new gene evolution
    - Patterns of new gene evolution

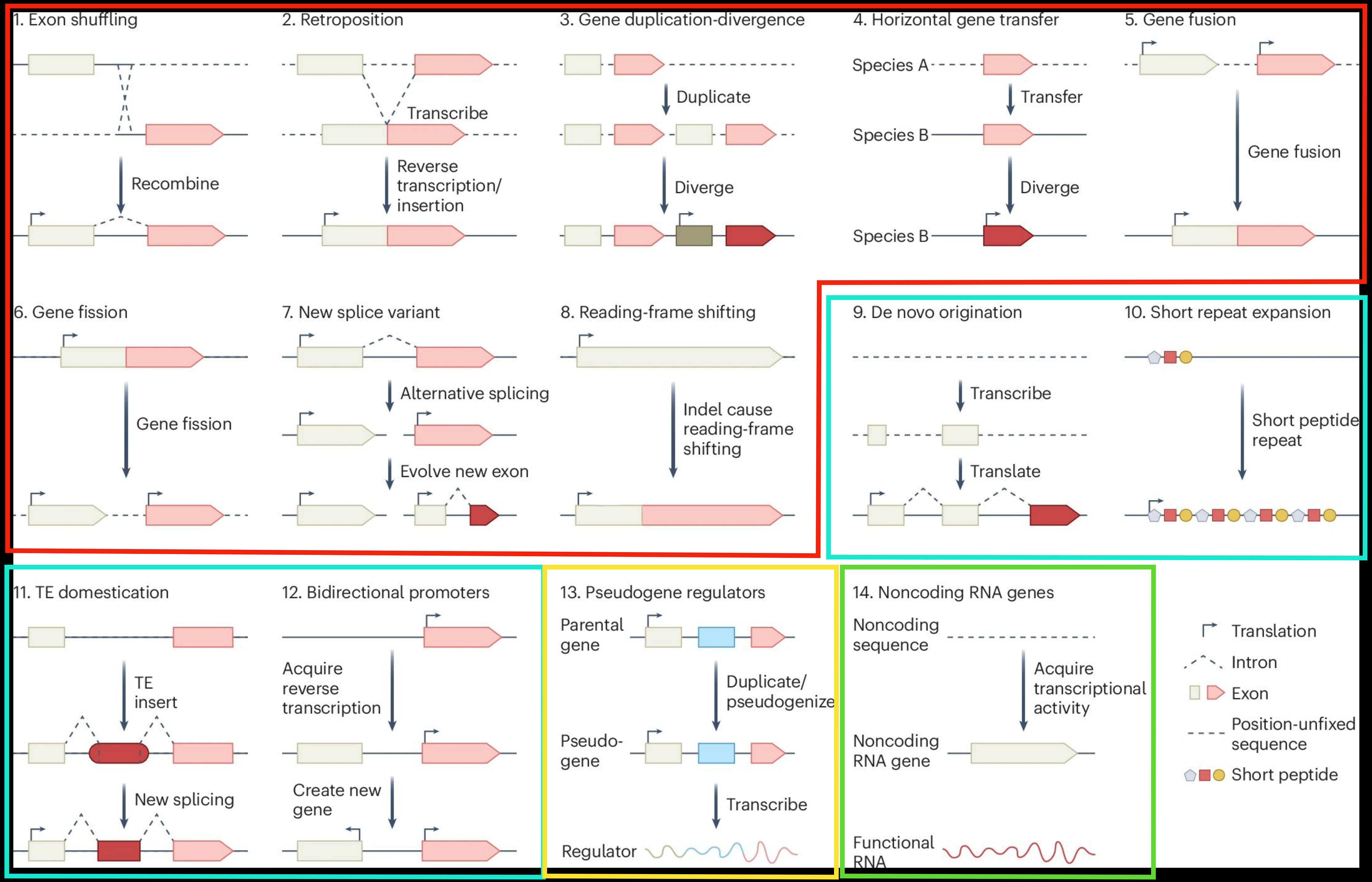
3. How to date gene ages?
    - How can we use the age of genes?

Molecular mechanisms to generate new genes

Classical model: gene duplication-divergence



Molecular mechanisms to generate new genes



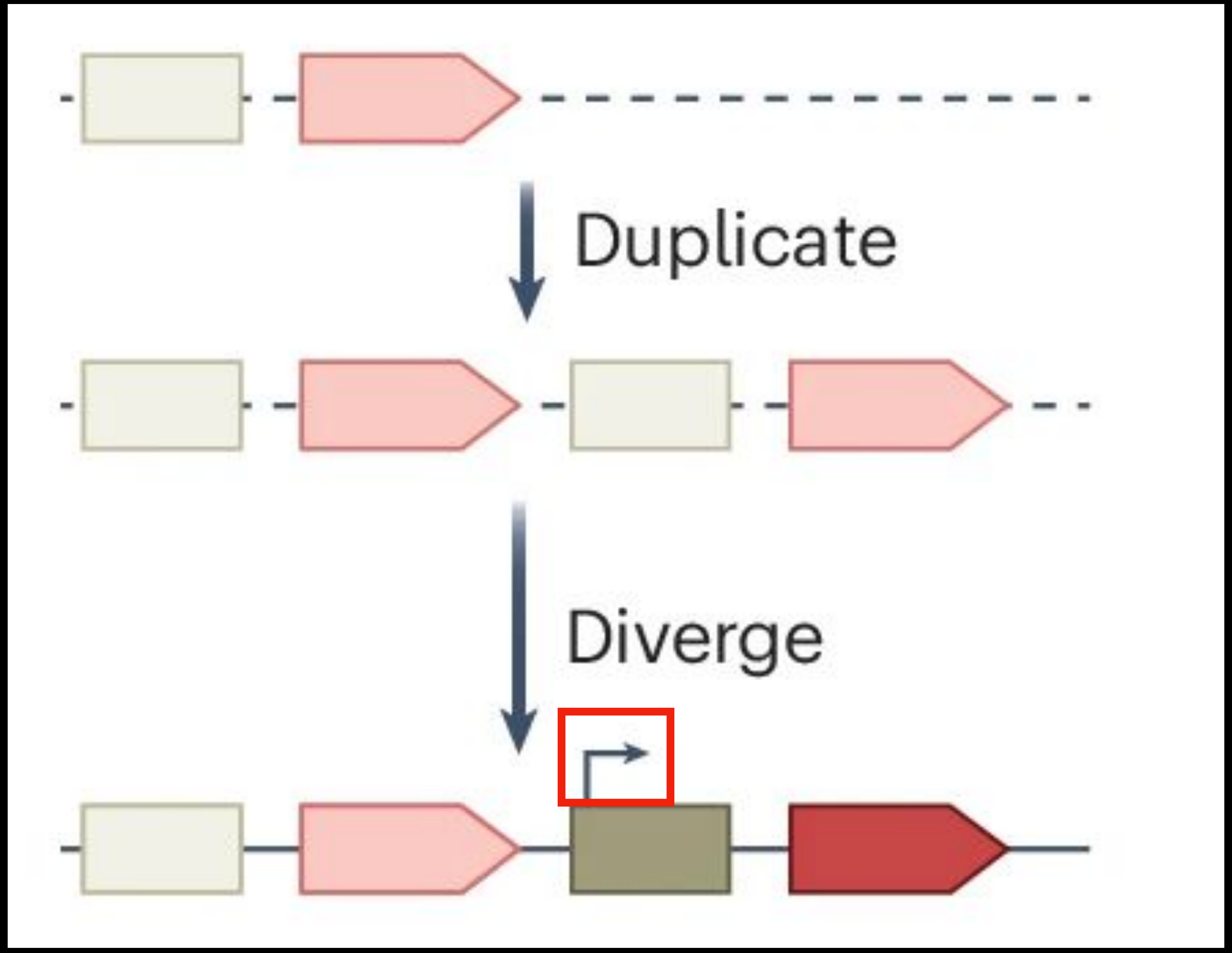
Based on the <u>products of ancestors</u> and <u>products of new genes</u>

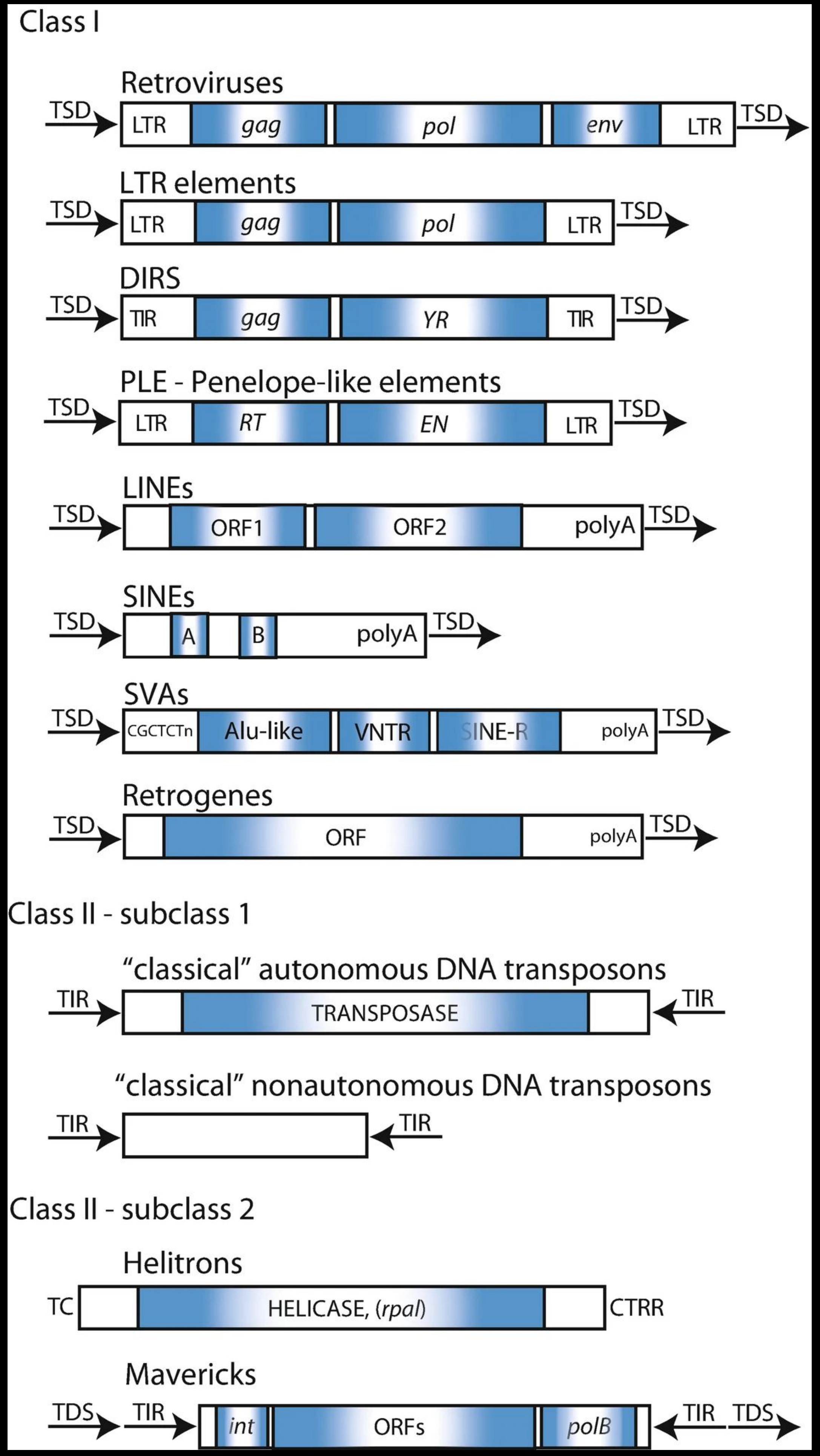
Classified into 4 types:

- Protein to Protein
  1. exon shuffling, 2. retroposition,
  3. gene duplication-divergence,
  4. Horizontal gene transfer, 5. gene fusion, 6. gene fission, 7. new isoform divergence, 8. reading frameshift
- Noncoding to Protein
  9. de novo genes, 10. short repeat expansion, 11. TE domestication, 12. bidirectional promoter
- Protein to Noncoding
  13. pseudogene regulators (*Xist*)
- Noncoding to Noncoding
  14. lncRNA genes

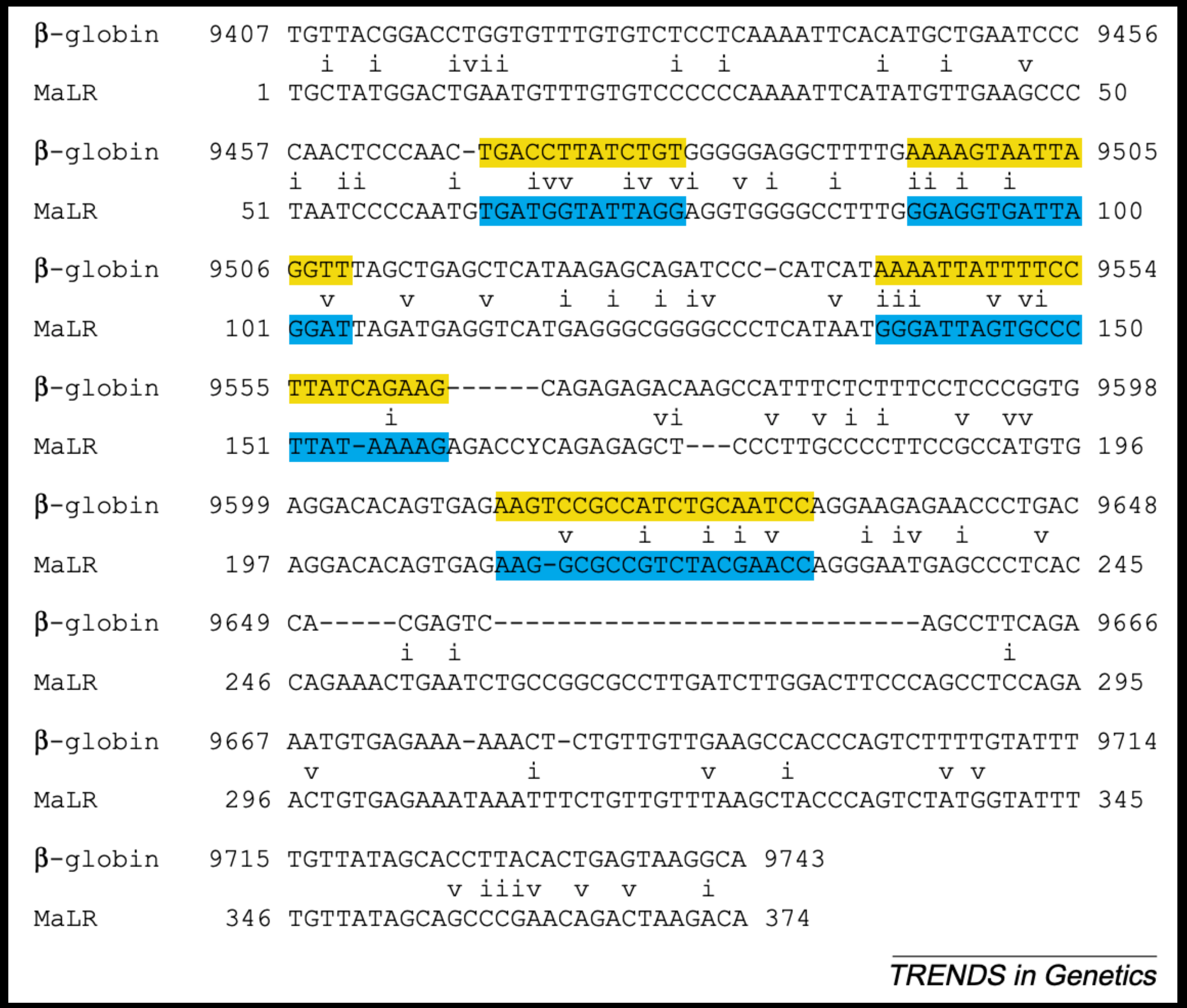
Molecular mechanisms to generate new genes

The generation of new regulatory elements in new genes





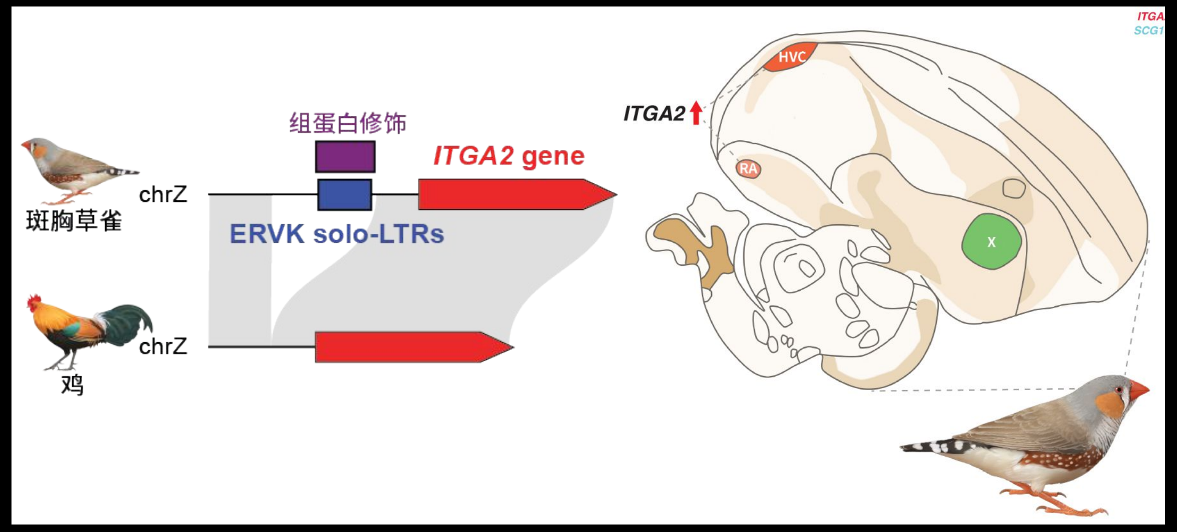
Structures of eukaryotic mobile elements



```
β-globin  9407 TGTTACGGACCTGGTGTTTGTGTCTCCTCAAAATTCACATGCTGAATCCC 9456
                 i  i    ivii          i  i         i   i    v
MaLR         1 TGCTATGGACTGAATGTTTGTGTCCCCCCAAAATTCATATGTTGAAGCCC 50

β-globin  9457 CAACTCCCAAC-TGACCTTATCTGTGGGGGAGGCTTTTGAAAAGTAATTA 9505
               i  ii     i    ivv   iv vi  v i   i    ii i  i
MaLR        51 TAATCCCCAATGTGATGGTATTAGGAGGTGGGGCCTTTGGGAGGTGATTA 100

β-globin  9506 GGTTTAGCTGAGCTCATAAGAGCAGATCCC-CATCATAAAATTATTTTCC 9554
                 v    v    v    i  i  i iv       v  iii    v vi
MaLR       101 GGATTAGATGAGGTCATGAGGGCGGGGCCCTCATAATGGGATTAGTGCCC 150

β-globin  9555 TTATCAGAAG------CAGAGAGACAAGCCATTTCTCTTTCCTCCCGGTG 9598
                     i                vi     v  v i i    v  vv
MaLR       151 TTAT-AAAAGAGACCYCAGAGAGCT---CCCTTGCCCCTTCCGCCATGTG 196

β-globin  9599 AGGACACAGTGAGAAGTCCGCCATCTGCAATCCAGGAAGAGAACCCTGAC 9648
                                v    i   i i v     i iv  i    v
MaLR       197 AGGACACAGTGAGAAG-GCGCCGTCTACGAACCAGGGAATGAGCCCTCAC 245

β-globin  9649 CA-----CGAGTC---------------------------AGCCTTCAGA 9666
                      i  i                                  i
MaLR       246 CAGAAACTGAATCTGCCGGCGCCTTGATCTTGGACTTCCCAGCCTCCAGA 295

β-globin  9667 AATGTGAGAAA-AAACT-CTGTTGTTGAAGCCACCCAGTCTTTTGTATTT 9714
                v             i          v    i         v v
MaLR       296 ACTGTGAGAAATAAATTTCTGTTGTTTAAGCTACCCAGTCTATGGTATTT 345

β-globin  9715 TGTTATAGCACCTTACACTGAGTAAGGCA 9743
                         v iiiv  v  v    i
MaLR       346 TGTTATAGCAGCCCGAACAGACTAAGACA 374
```

TRENDS in Genetics

Experimentally characterized *cis*-regulatory sequences from the β-globin locus are highlighted in yellow and the homologous nucleotides from the MaLR (LTR) sequence are highlighted in blue.



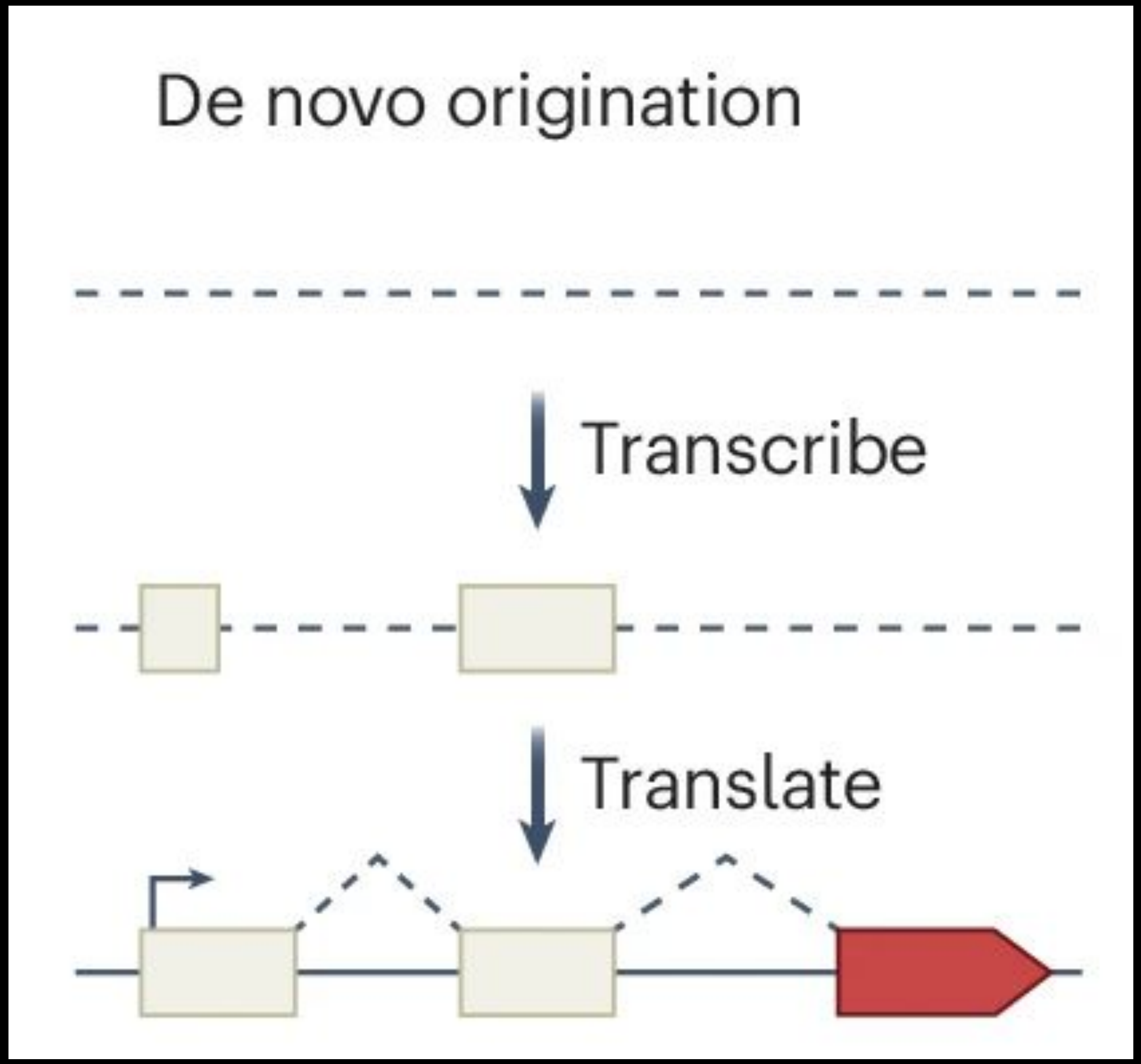
Sourced from

- recruited existing regulatory elements
- transposable elements

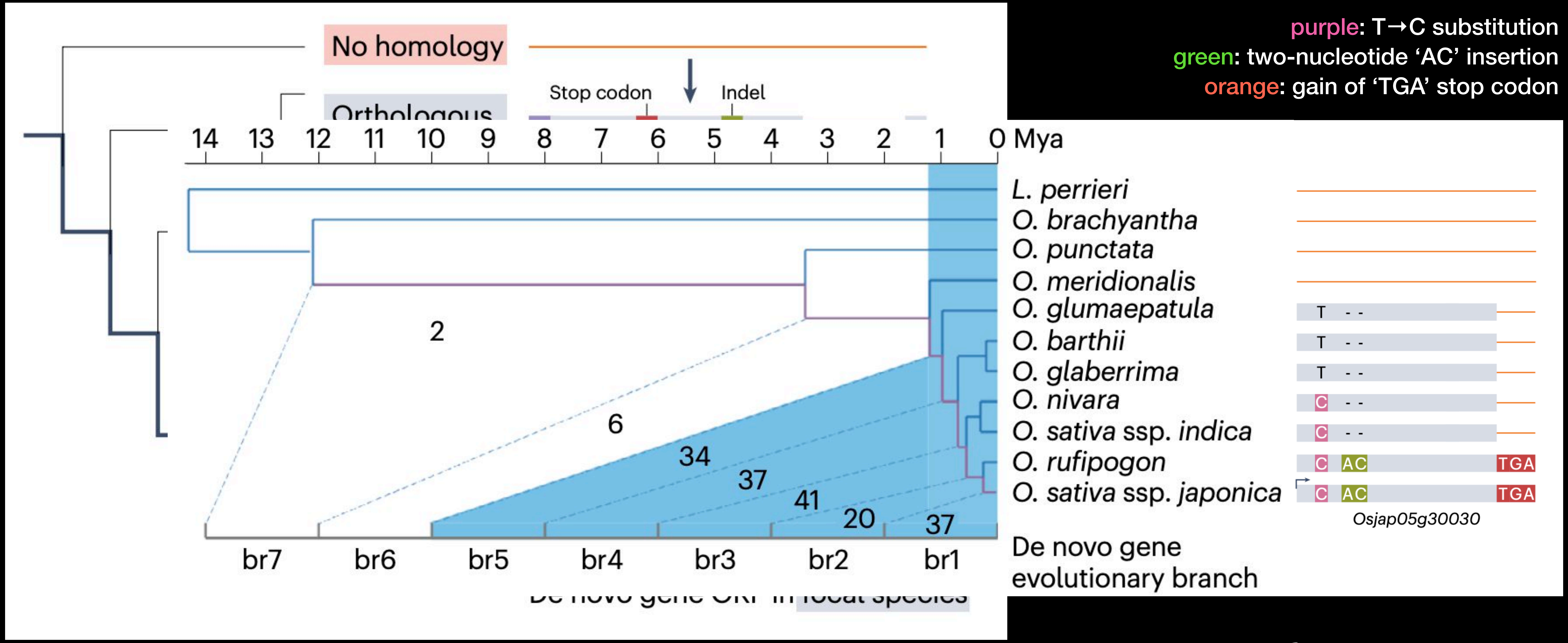
Makałowski, W. et al. 2019.
Jordan, I. K. et al. 2003.
Chen, G. et al. 2024.

De novo gene generation

Stepwise *de novo* gene origination



De novo gene origination



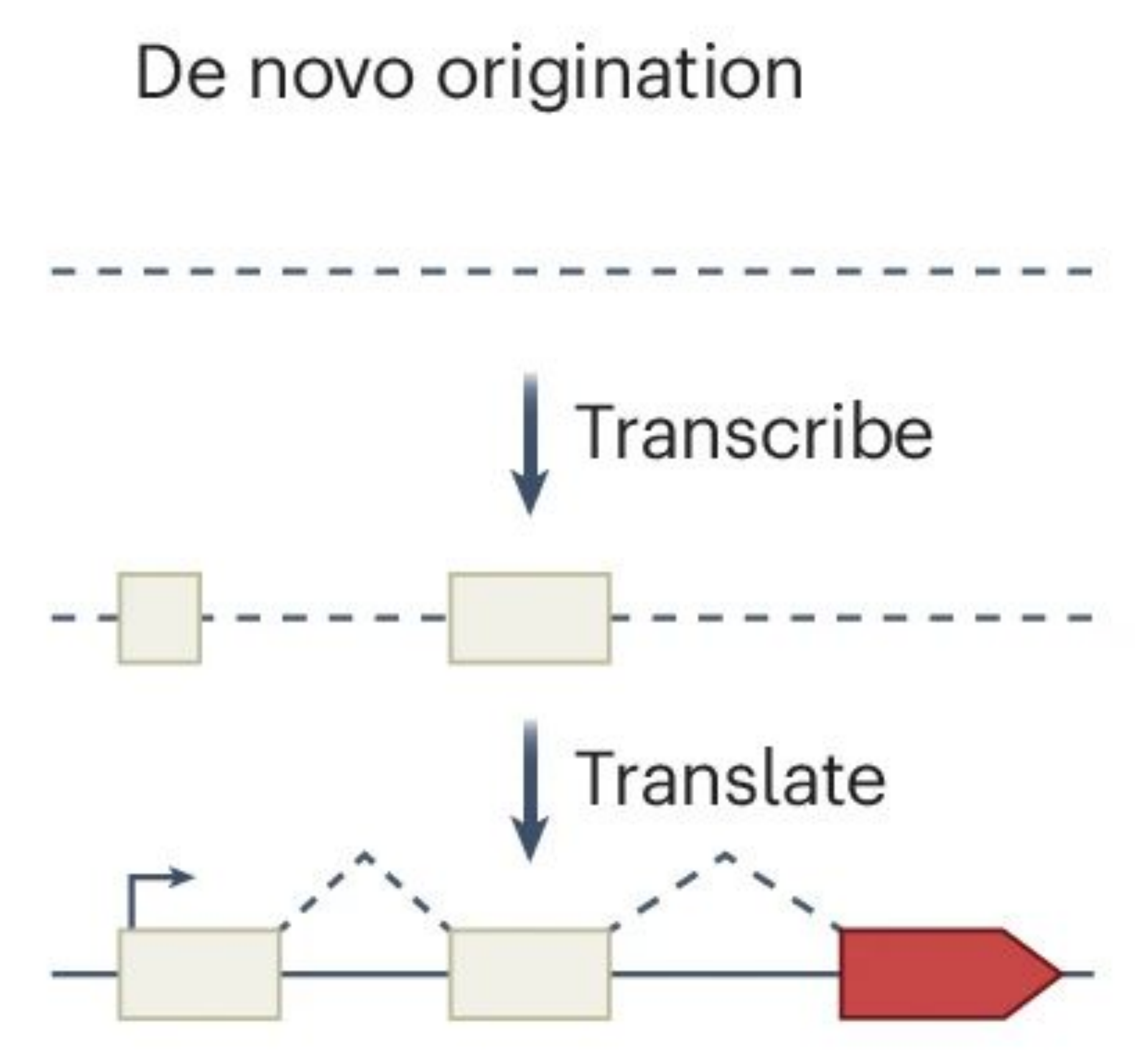
purple: T→C substitution
green: two-nucleotide 'AC' insertion
orange: gain of 'TGA' stop codon

Stepwise *de novo* gene origination

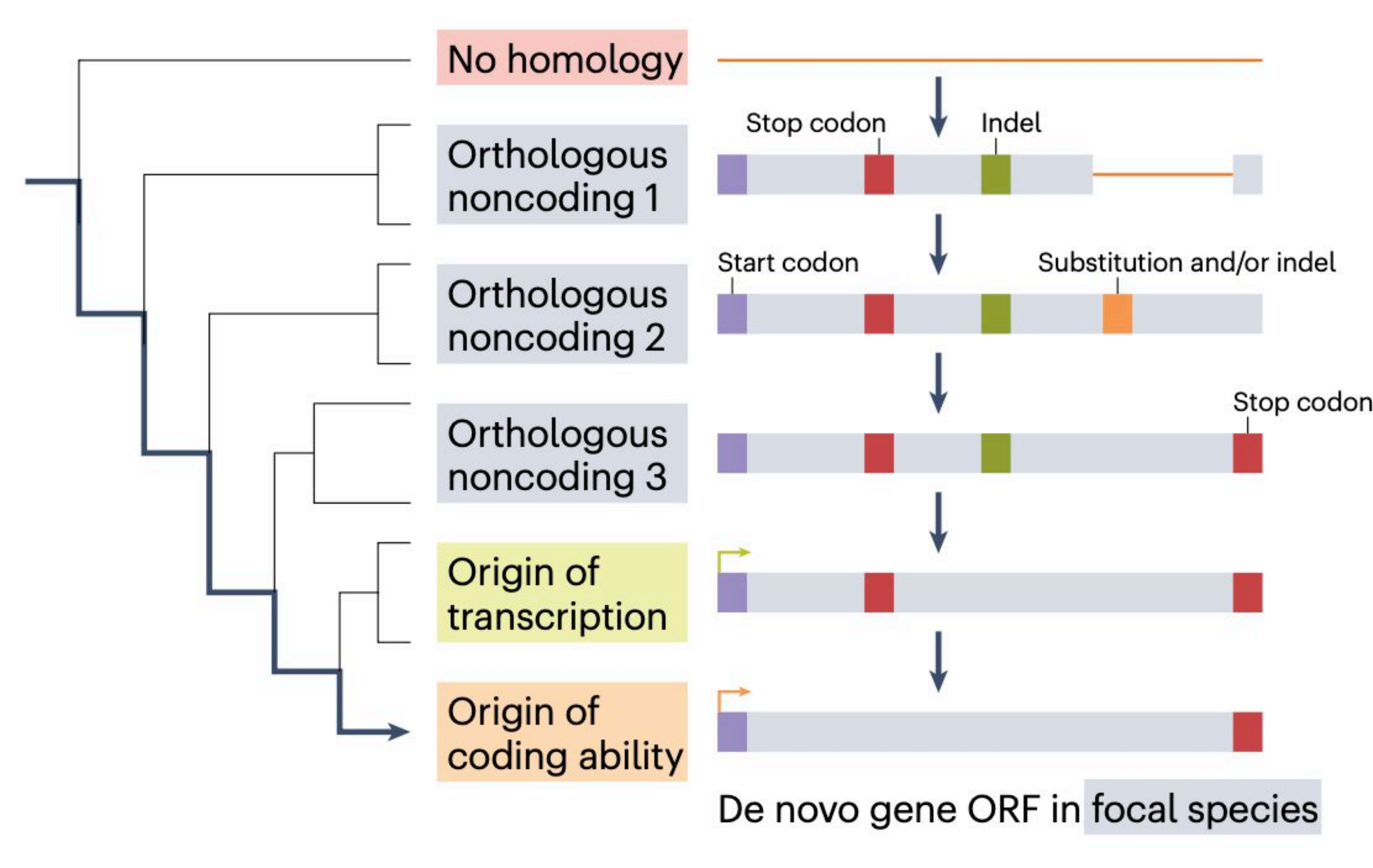
Oryza [稻属]

De novo gene generation

Stepwise *de novo* gene origination



De novo gene origination



Stepwise *de novo* gene origination

De novo gene generation

De novo genes vs. orphan genes

Orphan genes are genes that lack a detectable homologue outside of a given species or lineage.

	***de novo* genes**	**orphan genes**
Focus	**Origin mechanism** novel genes arising from previously **non-coding** DNA	**Phylogenetic uniqueness or isolation** genes without detectable homologs in other species
Emphasis	Functional origin	Evolutionary isolation
Species-specific?	Not necessarily	Yes
Homology	Usually **Yes**	Usually no

rapid divergence

alternative reading frames

gene transfer from rapidly evolved viruses or bacteria

gene loss

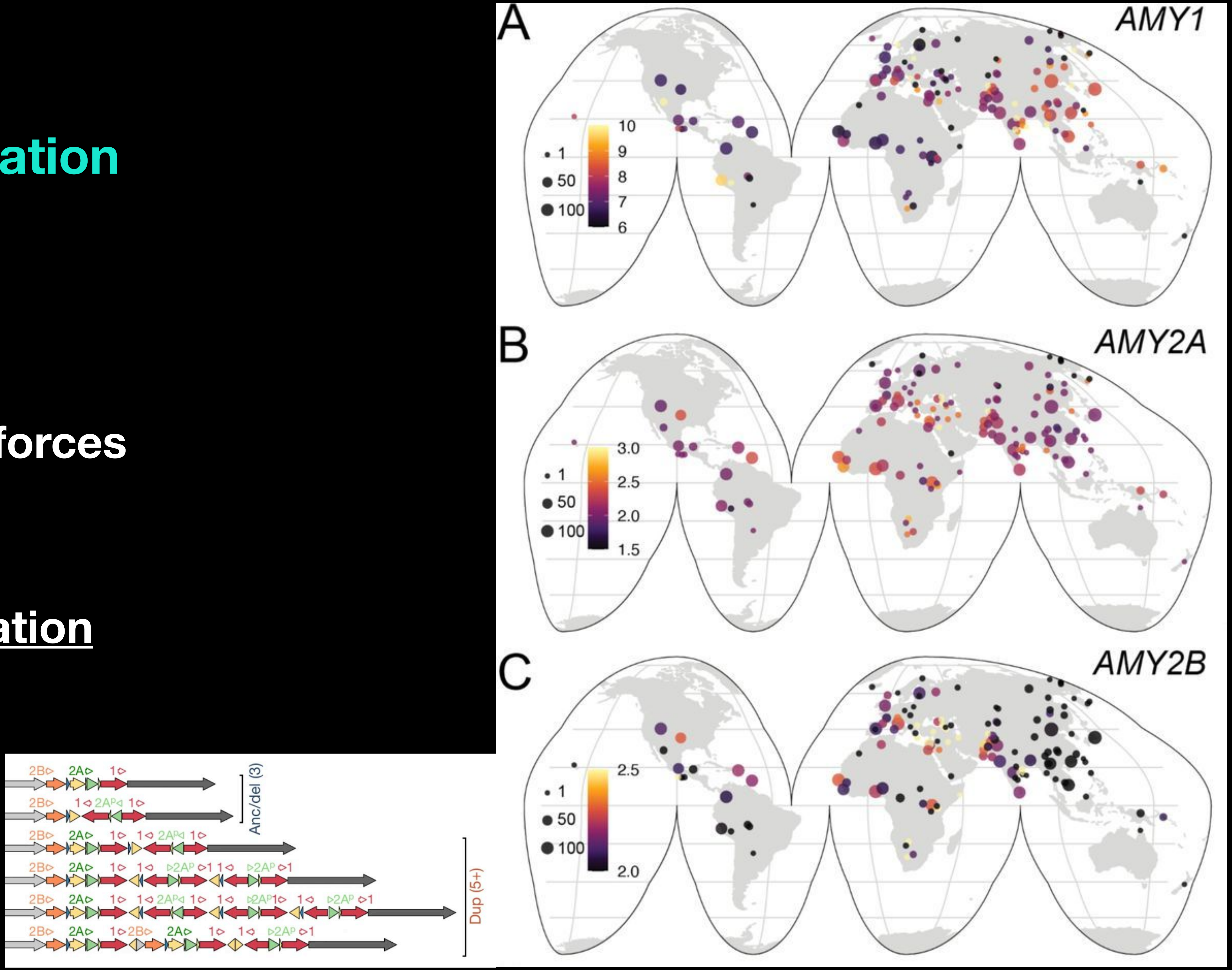
Q: How did new genes evolve?

Evolutionary forces acting on new genes

- **Natural selection for functional innovation**
- **Sexually antagonistic selection**

- **Approaches to detect these evolutionary forces**
  - **Positive selection**
  - **Copy number variants (CNVs) in population genomics**
    - *AMY* gene in human population

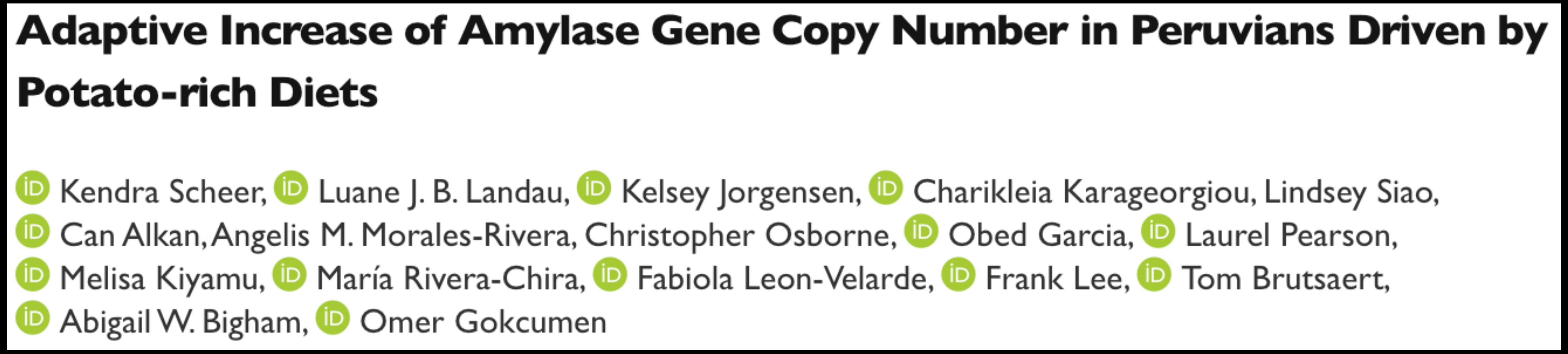
Amylase: 淀粉酶



Bolognini, D. et al. 2024.

Examples of Copy number variants (CNVs) in population genomics

Amylase: 淀粉酶



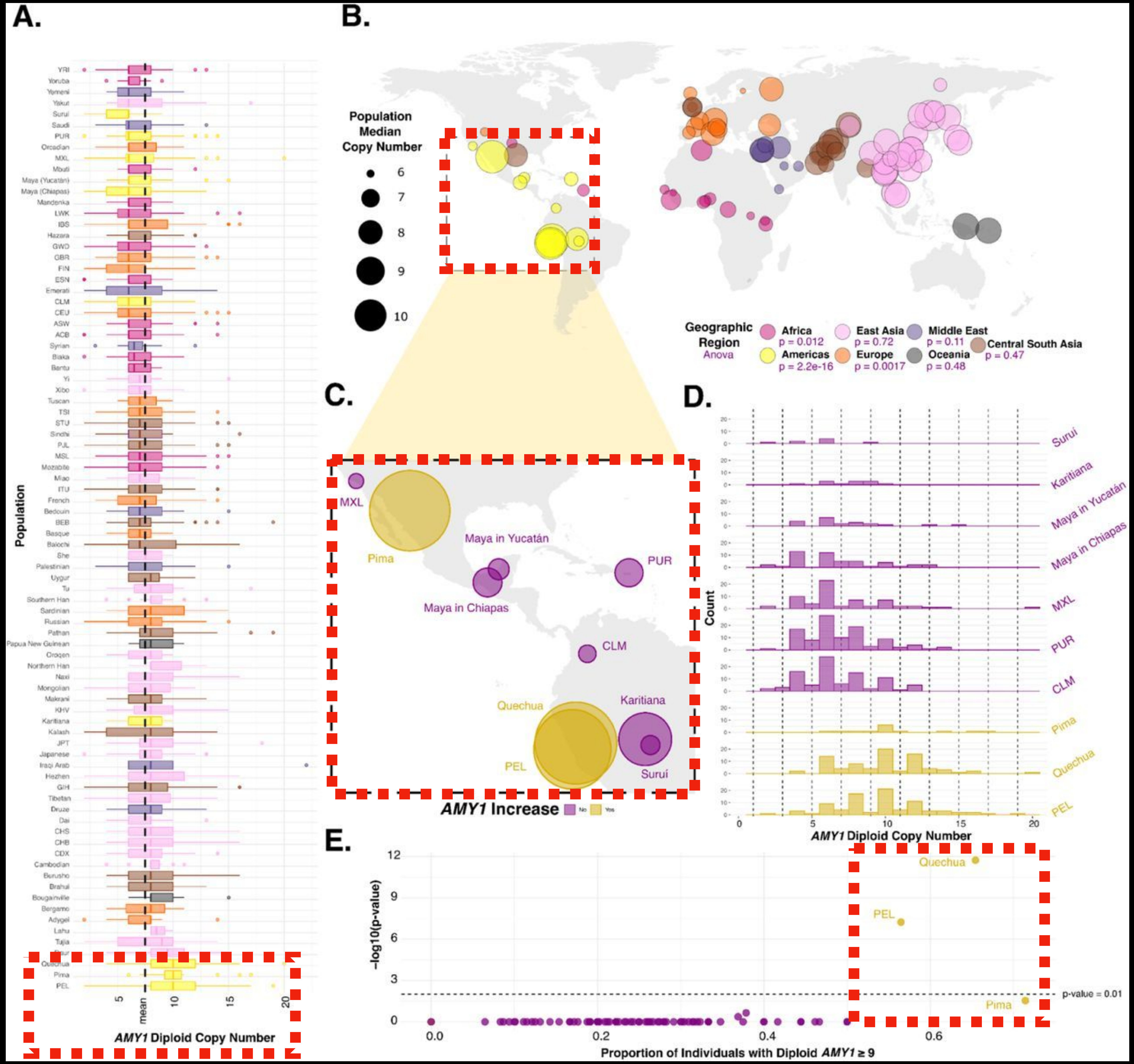
Adaptive Increase of Amylase Gene Copy Number in Peruvians Driven by Potato-rich Diets

Kendra Scheer, Luane J. B. Landau, Kelsey Jorgensen, Charikleia Karageorgiou, Lindsey Siao, Can Alkan, Angelis M. Morales-Rivera, Christopher Osborne, Obed Garcia, Laurel Pearson, Melisa Kiyamu, María Rivera-Chira, Fabiola Leon-Velarde, Frank Lee, Tom Brutsaert, Abigail W. Bigham, Omer Gokcumen

***AMY1* gene**

Potato-rich Diets

Peru 秘鲁
Peruvians 秘鲁

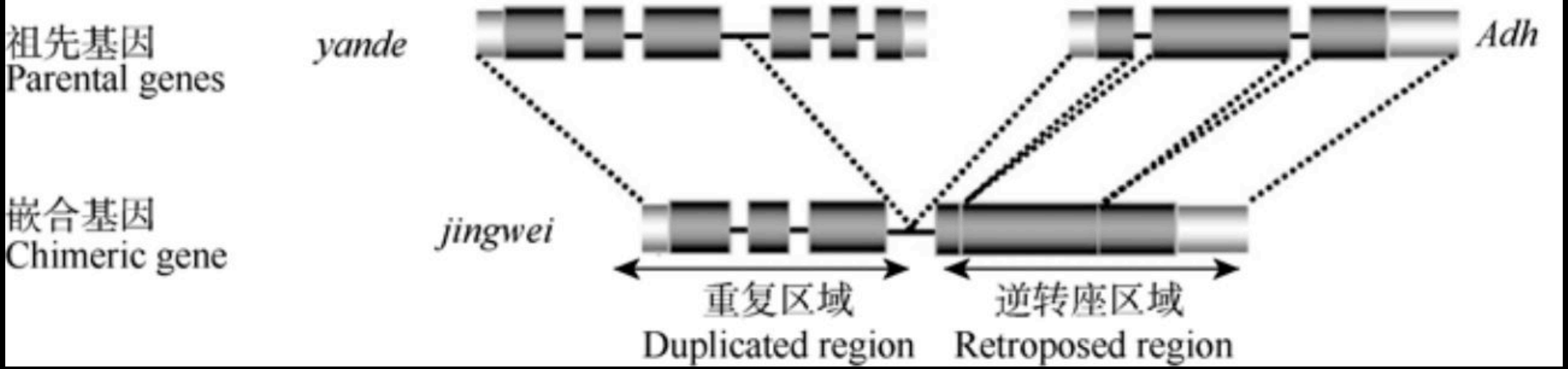


The potato is native to the Andes Mountains in South America!

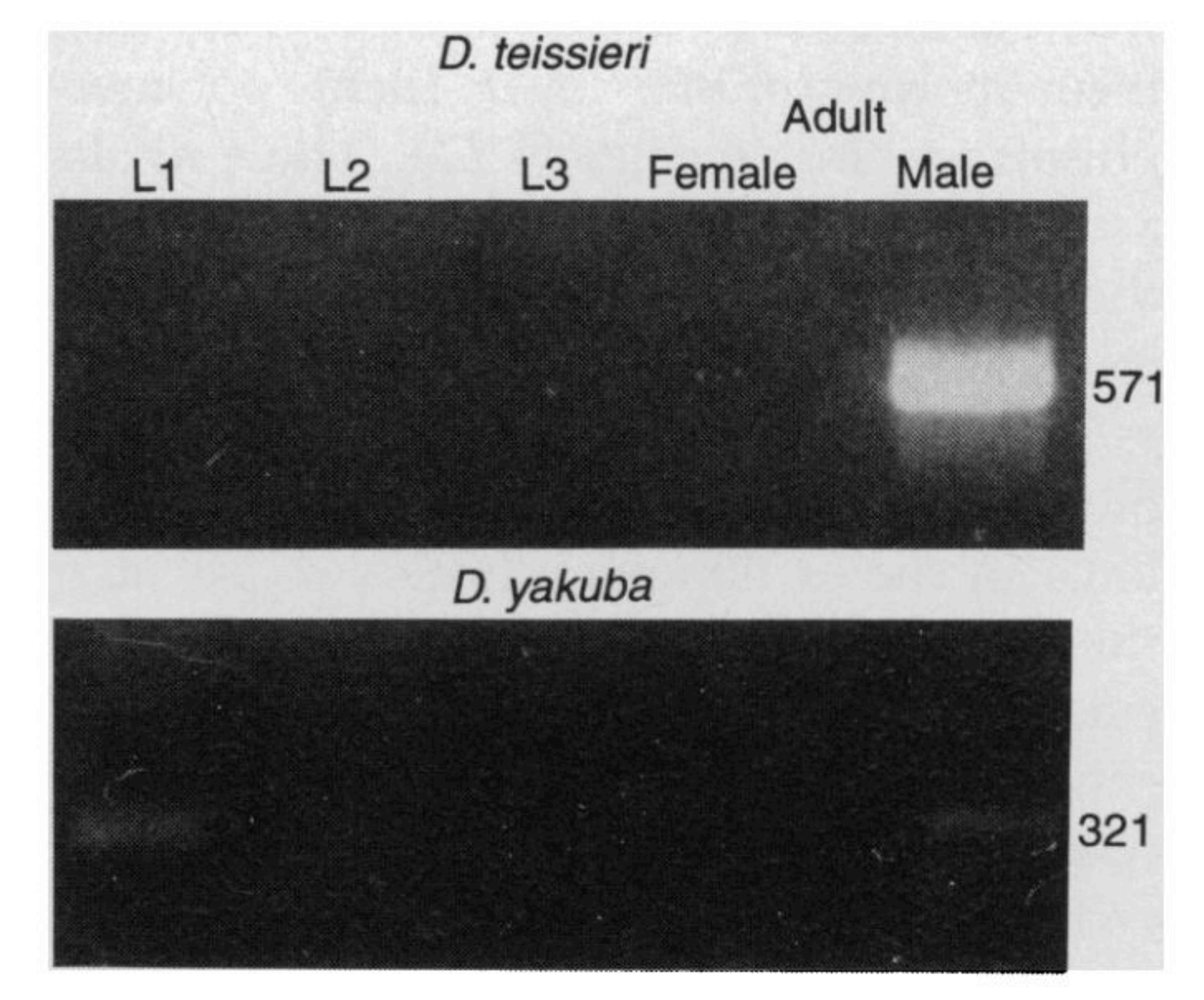
Kendra S. et al. 2025.

New genes with detectable molecular functions and fertility effects

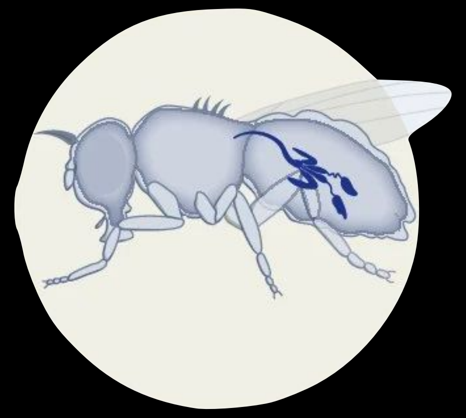
jgw gene



The *Adh* gene inserted into *yande* gene after reverse transcription, and the exons of the two genes are transcribed together to produce a chimeric new gene (*jingwei*).



jingwei gene *(jgw)* show specific high expression in Adult Male of *D. teissieri.*



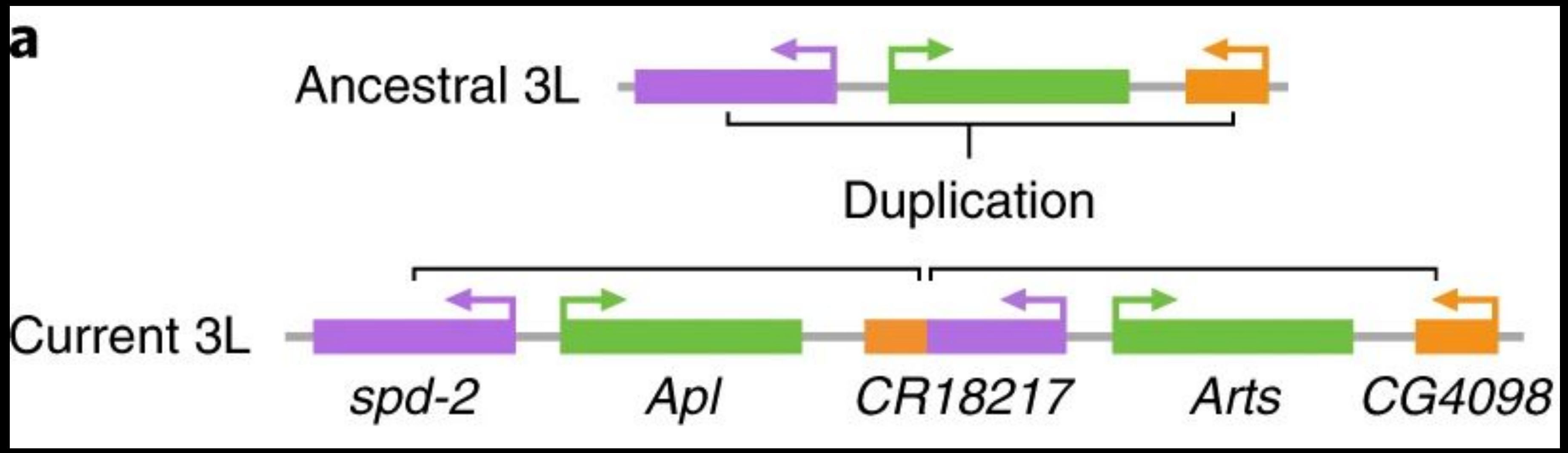
Drosophila

***jingwei* (*jgw*) encodes a novel alcohol dehydrogenase-like enzyme[脱氢酶], likely contributing to *Drosophila*'s metabolic adaptation and possibly reproductive fitness.**

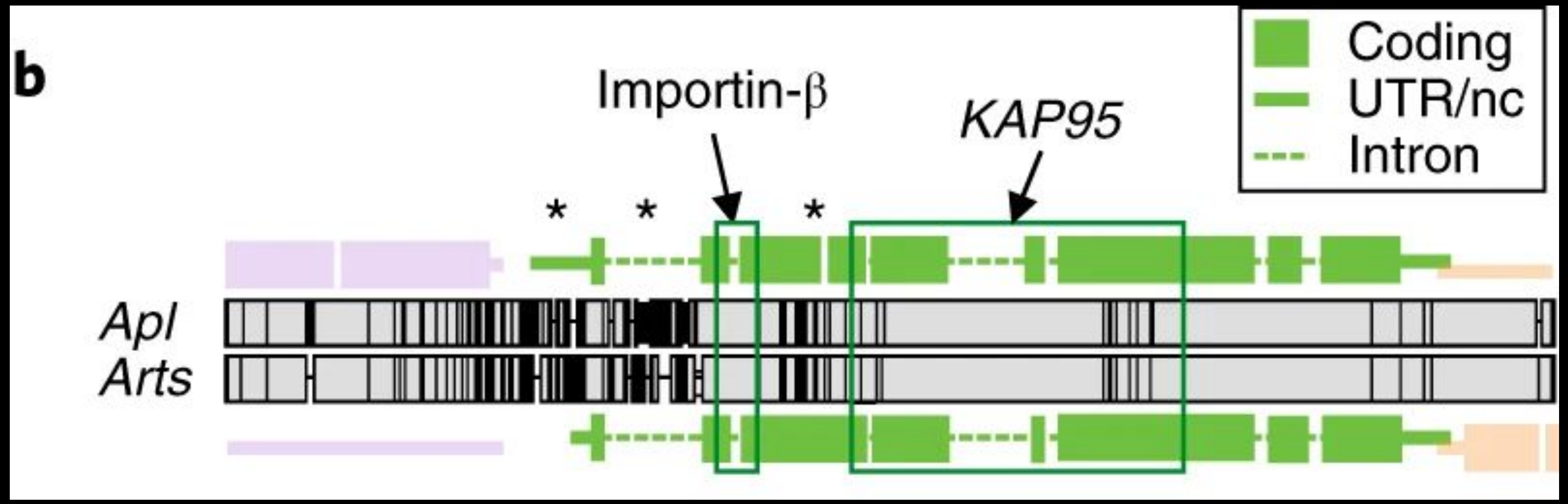
Long, M. et al. 1993.
Zhan, Z. et al. 2011.

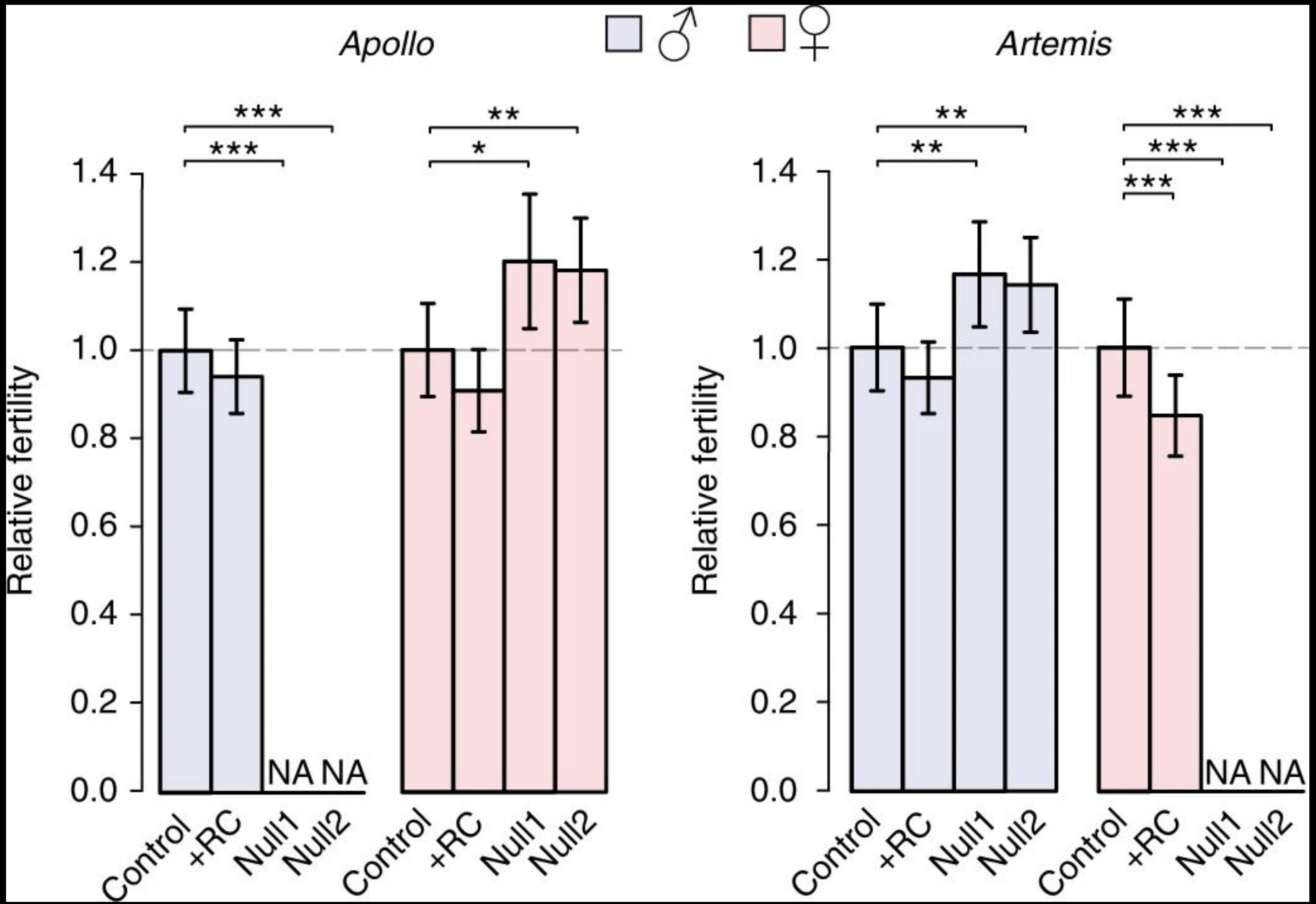
New genes with detectable molecular functions and fertility effects

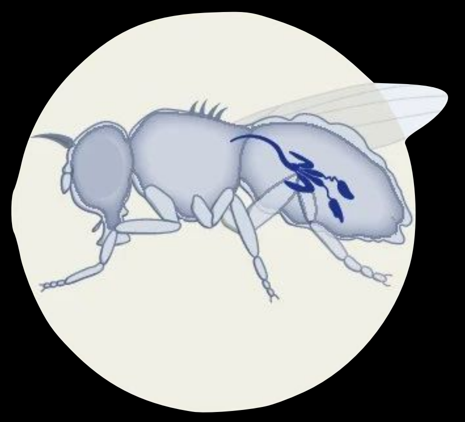
Apl & Arts genes



originated through a duplication event on chromosome 3L (paralogous relationship)







Drosophila

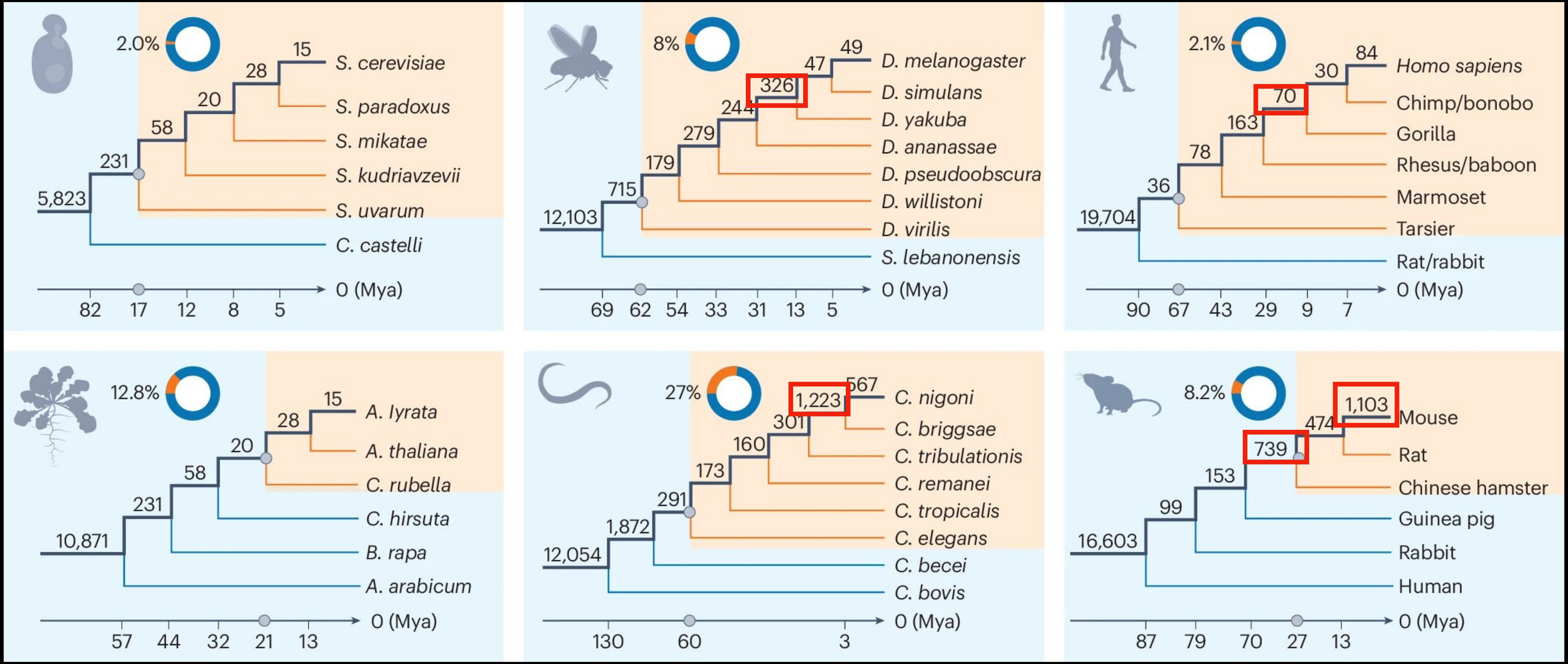
Apl & Arts
- distinct exon–intron organizations
- unique coding regions

- Null means the CRISPR/Cas9 knockouts.
- Raw counts of offspring are scaled to the mean counts from control lines.
  - Knocking out *Apollo* significantly reduces male fertility.
  - Knocking out *Artemis* significantly affects female fertility.

Functional essentiality and fertility effects

VanKuren, N. W. et al. 2018.

Rate of new gene evolution



Overview of new gene origination in humans and several other species

1. New genes have continuously emerged on a long time scale, often throughout the entire course of evolutionary history.
2. Recent lineages tend to acquire new genes more rapidly than ancient lineages.
    - 25$g\ M^{-1}$ compared to 4–7 $g\ M^{-1}$ in human, 34 $g\ M^{-1}$ compared to 5–11 $g\ M^{-1}$ in *Arabidopsis.*

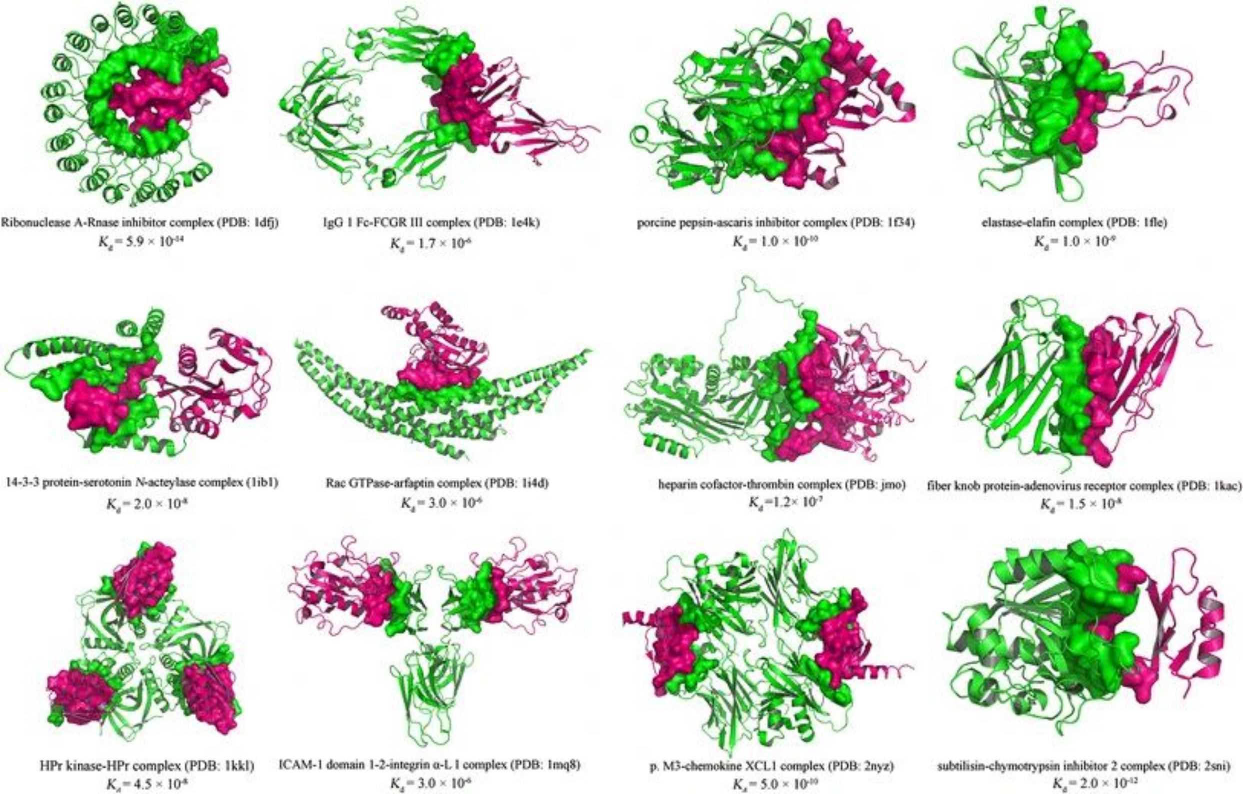
Patterns of new gene evolution

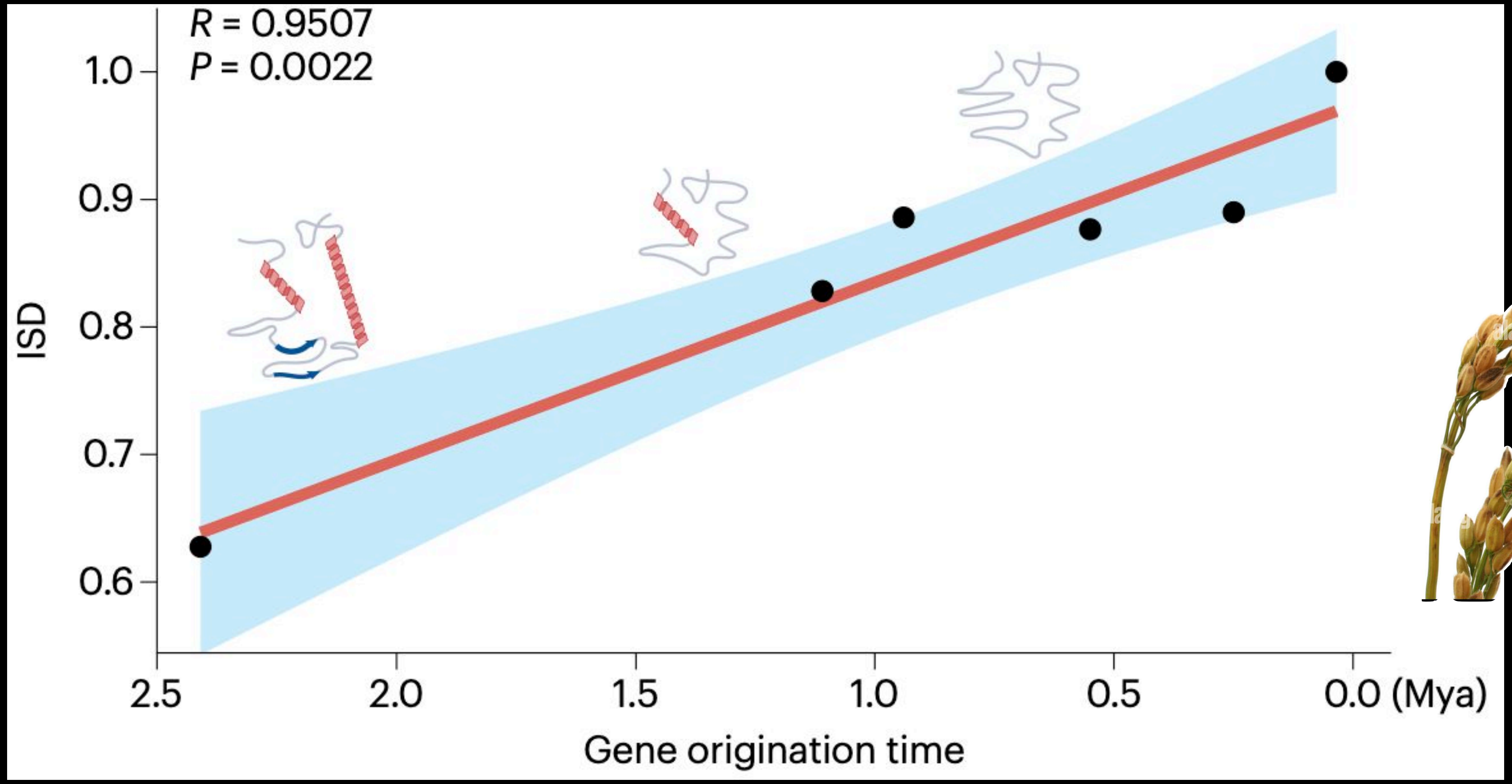
1. Structure

Debated

How does the structure of *de novo* proteins evolves?

more complex?





Tertiary structure of *de novo* proteins in *Oryza*.

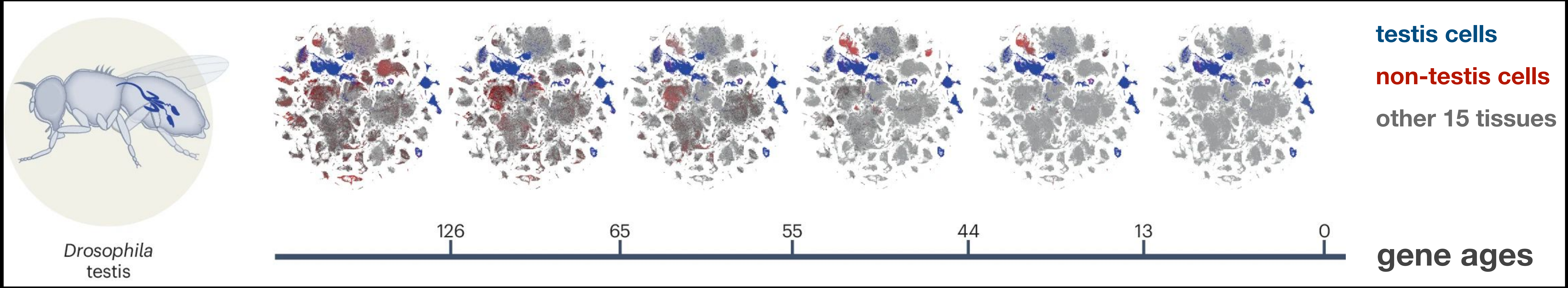
Oryza [稻属]

ISD: intrinsic structural disorder [内在结构紊乱]

- *Oryza de novo* genes showed a steady evolution of structure that is accompanied by increasing complexity in a relatively short time of 2.4 million years.

Patterns of new gene evolution

2. Expression patterns of new genes



‘Out of testis’ expression pattern of new genes at the single-cell level in *Drosophila*

As genes become older, they gradually accumulate expression in non-testis tissues.

‘**out of testis**’ pattern

majority of new genes exhibit testis expression

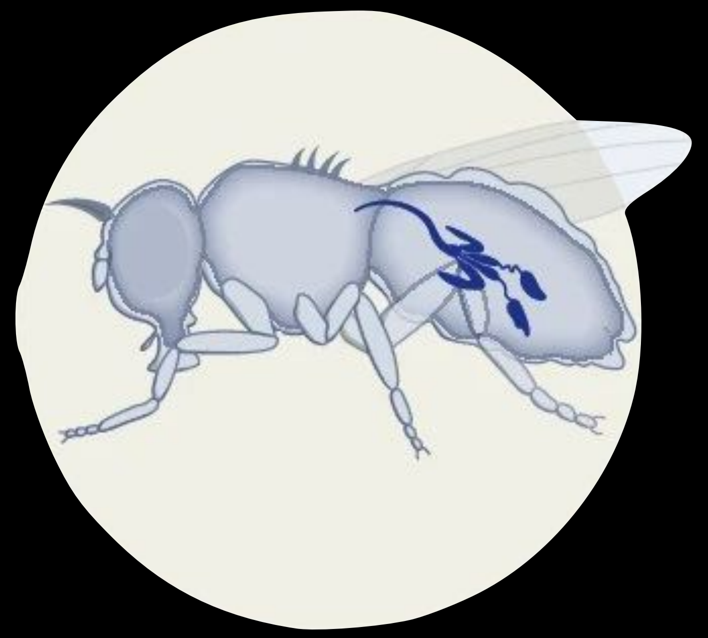
Patterns of new gene evolution

2. Expression patterns of new genes

As genes become older, they gradually accumulate expression in non-testis tissues.

‘**out of testis**’ pattern

majority of new genes exhibit testis expression



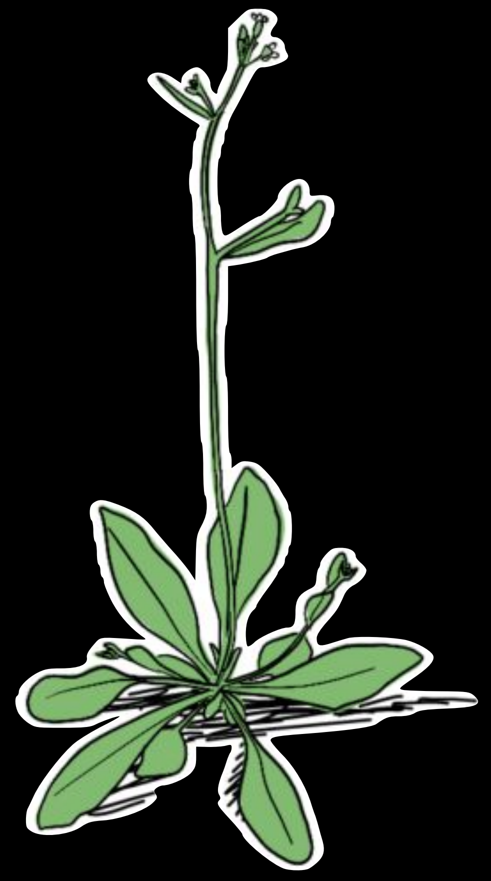
Drosophila

Human

‘**out of pollen** [花粉]’ hypothesis

more highly expressed in anthers[囊] and other male flower tissues

Oryza [稻属]



Arabidopsis [拟南芥]

Q: How to date gene ages?

How to date gene ages?

- Genomic screen of a gene for paralogous sequences
  - previously existing genetic material

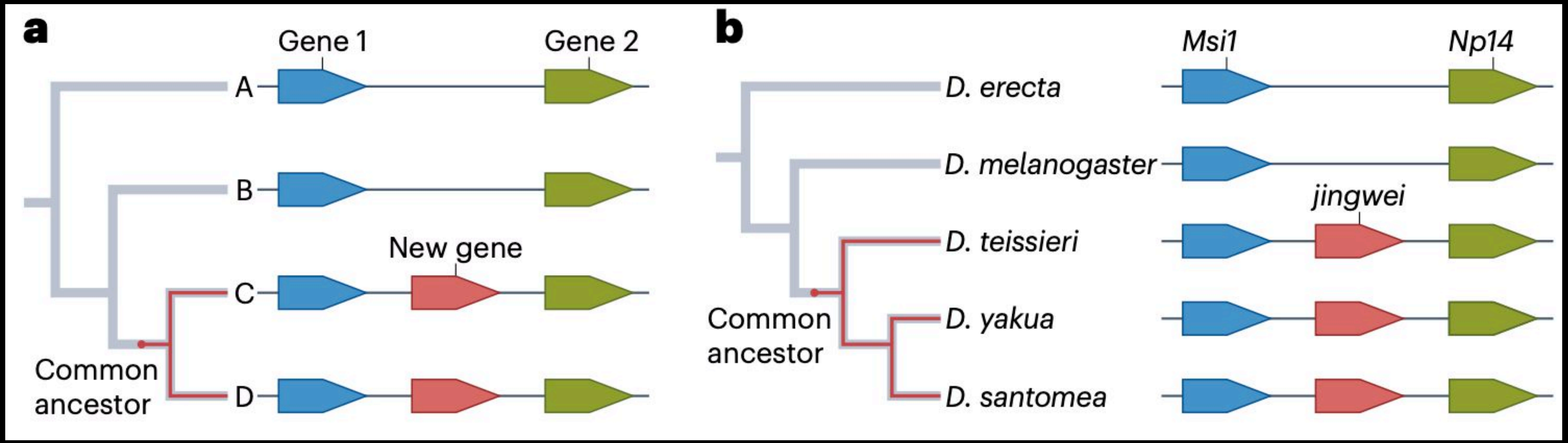
    gene duplication, TEs or genic recombination
  - previously non-existing material

    potential *de novo* gene

GenTree
The Time Tree of Genes Along the Evolution History

Human Data **Fly Data** **Mouse Data**
Rat Data **Opossum Data** **Chicken Data**

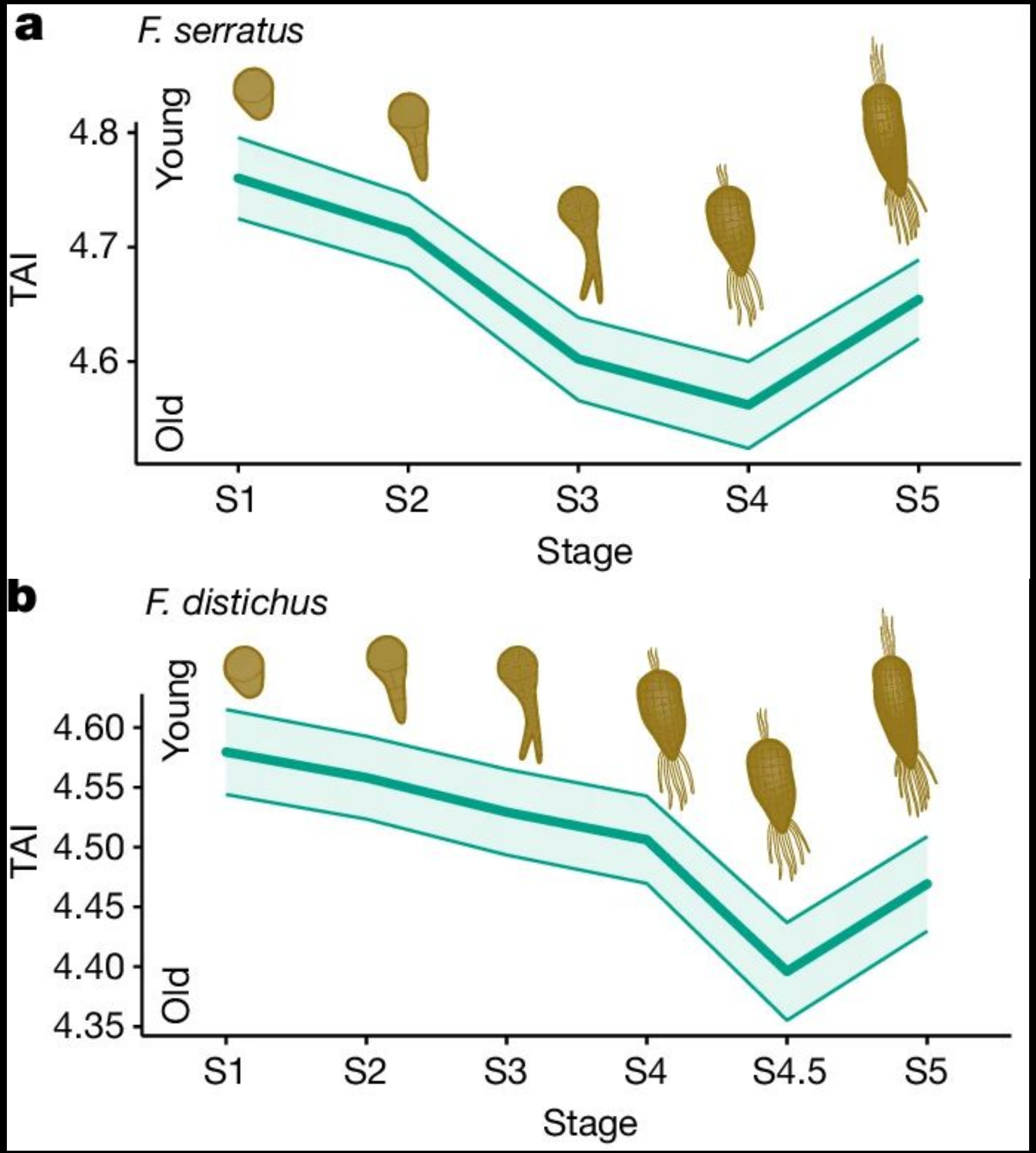
- Phylostratigraphy
  - Phylo-: 系统发育
  - stratigraphy: 地层学



The ***orthologous relationship*** across species groups is determined by a syntenic map surrounding the gene of interest, which is created by a high-quality reciprocal genome alignment.

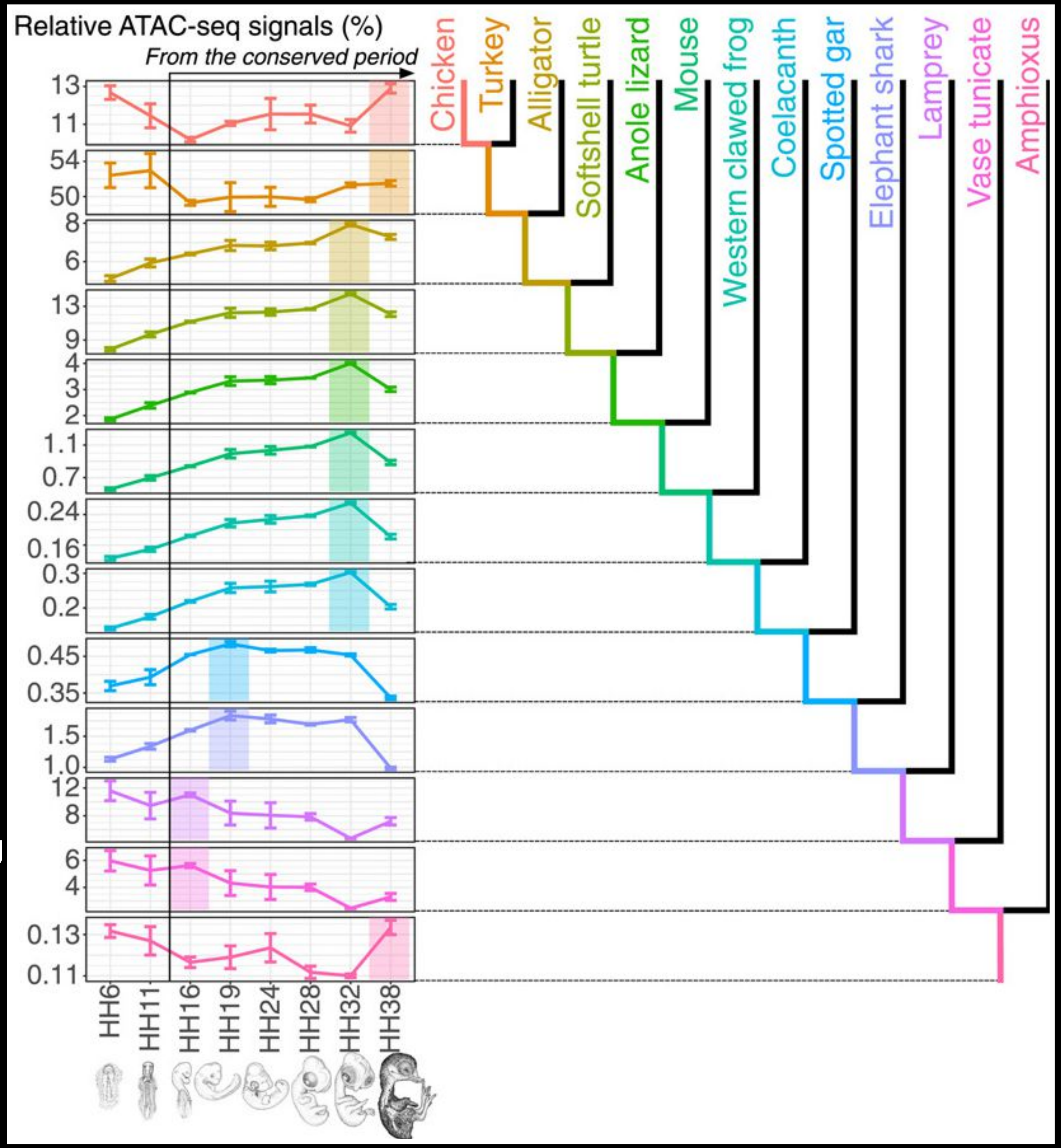
Shao Y. et al. 2019.

How can we use the age of genes?



TAI: transcriptome age indices

- Old genes show conserved expression in the middle stages of development.



young

age

old

early

late

Stages

Vertical axis represents:

- the percentage (%) of relative chromatin accessibilities of the genomic regions categorized by the evolutionary ages of their sequences

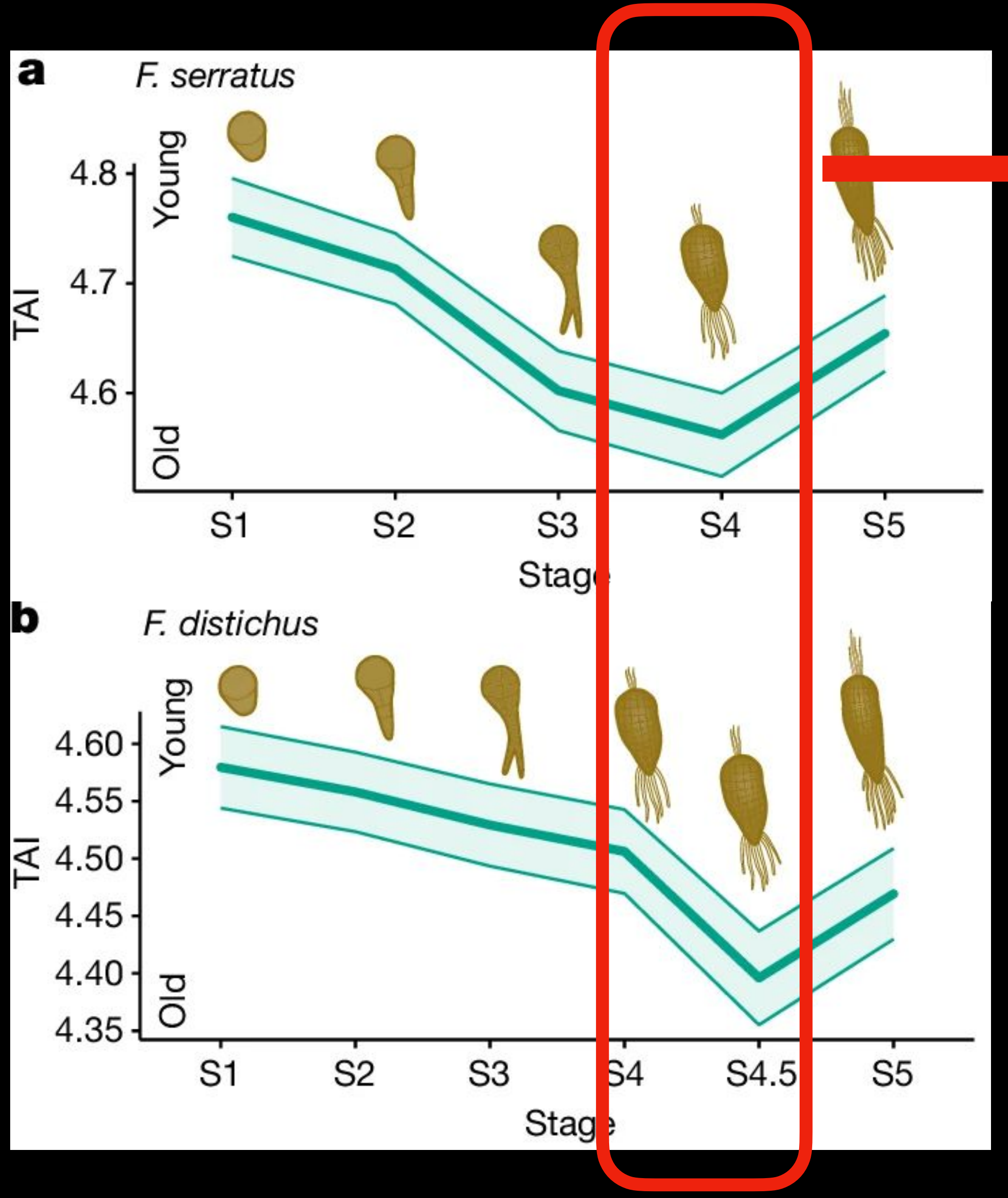
In brief:

when the new genes expression

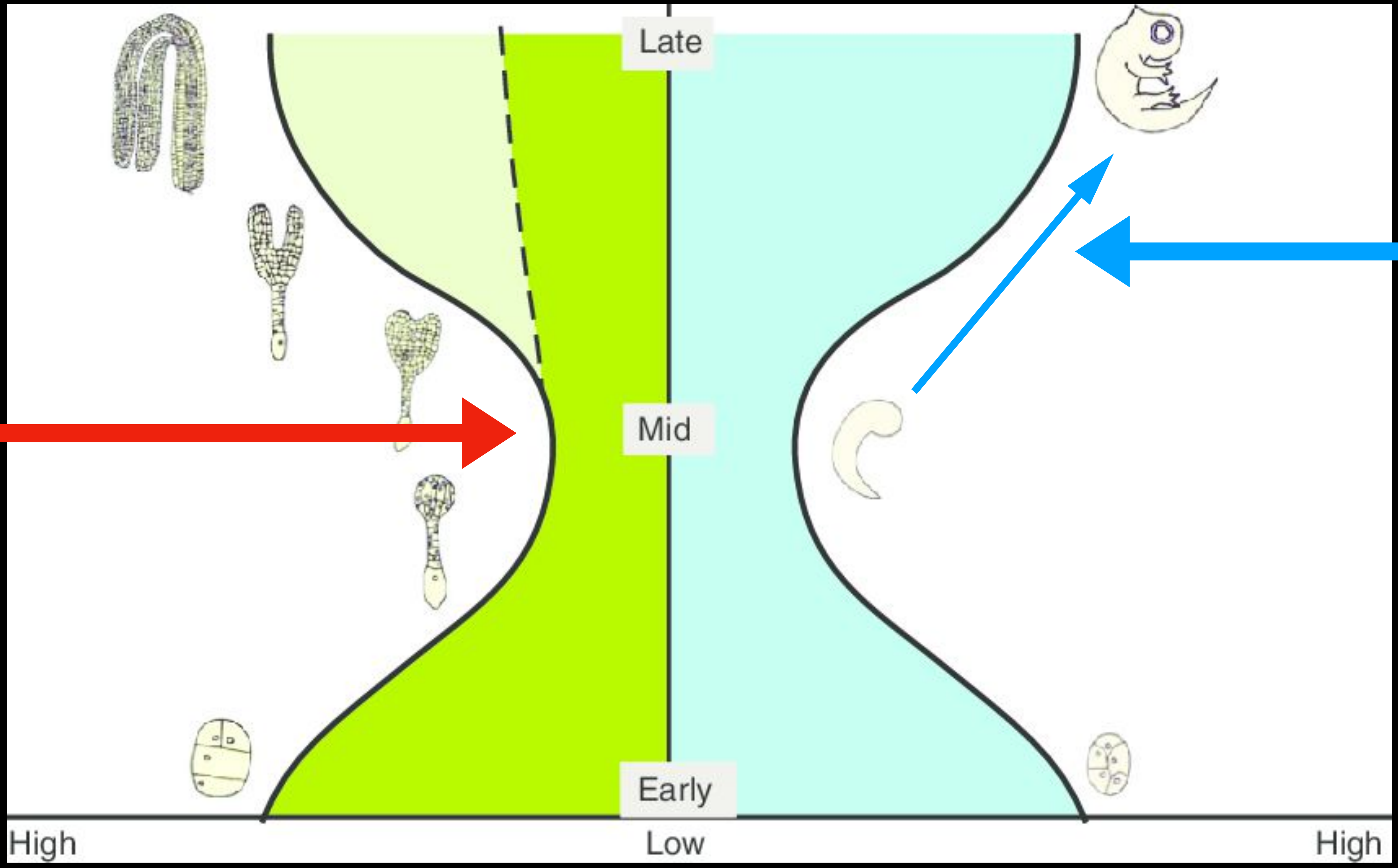
- Young genes show high expression in the later stages of (embryonic) development.

Uesaka, M. et al. 2021.
Lotharukpong, J. S. et al. 2024.

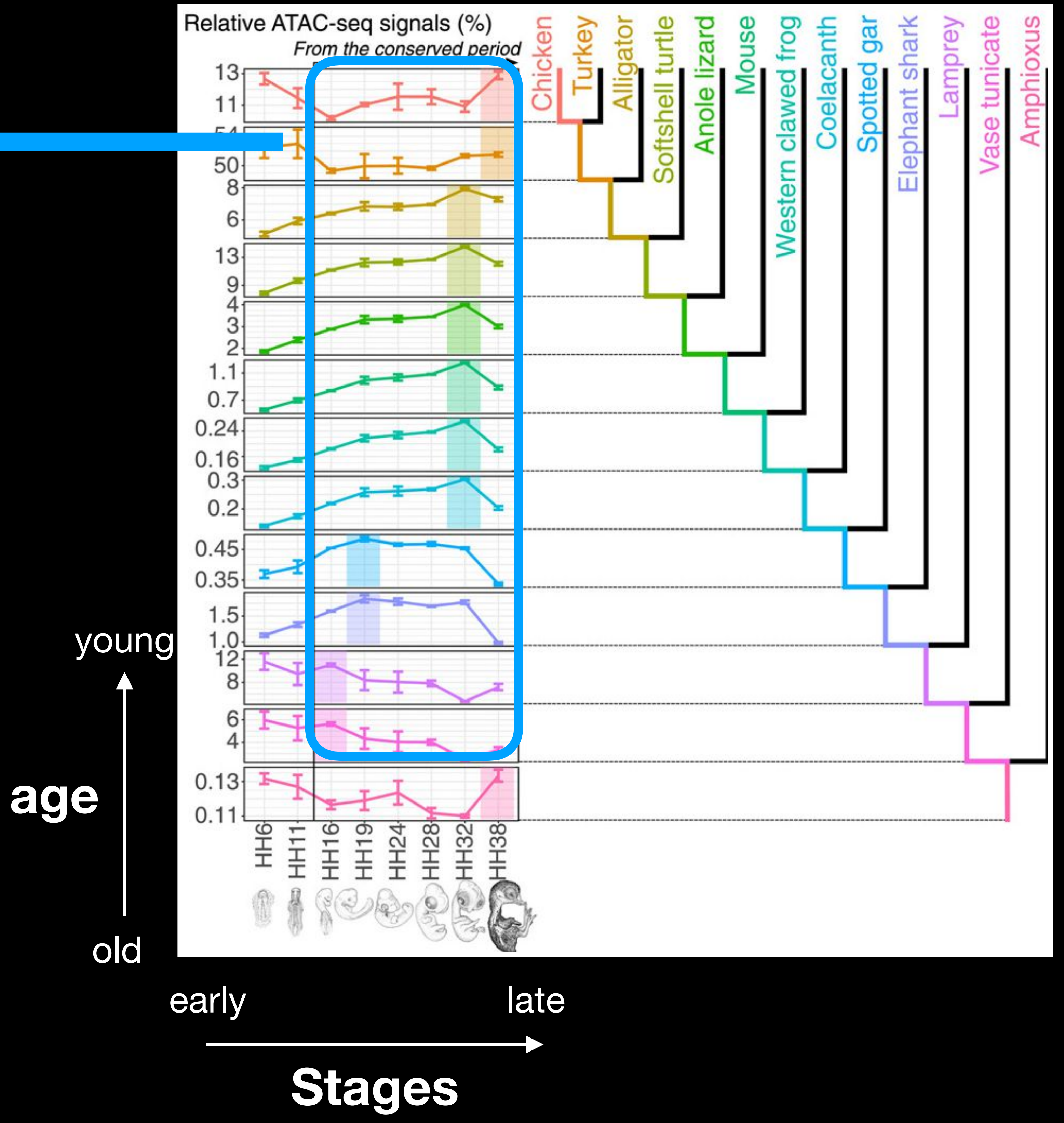
How can we use the age of genes?



TAI: transcriptome age indices



Hourglass model



- Old genes show conserved expression in the middle stages of development.
- Young genes show high expression in the later stages of (embryonic) development.

Uesaka, M. et al. 2021.
Lotharukpong, J. S. et al. 2024.

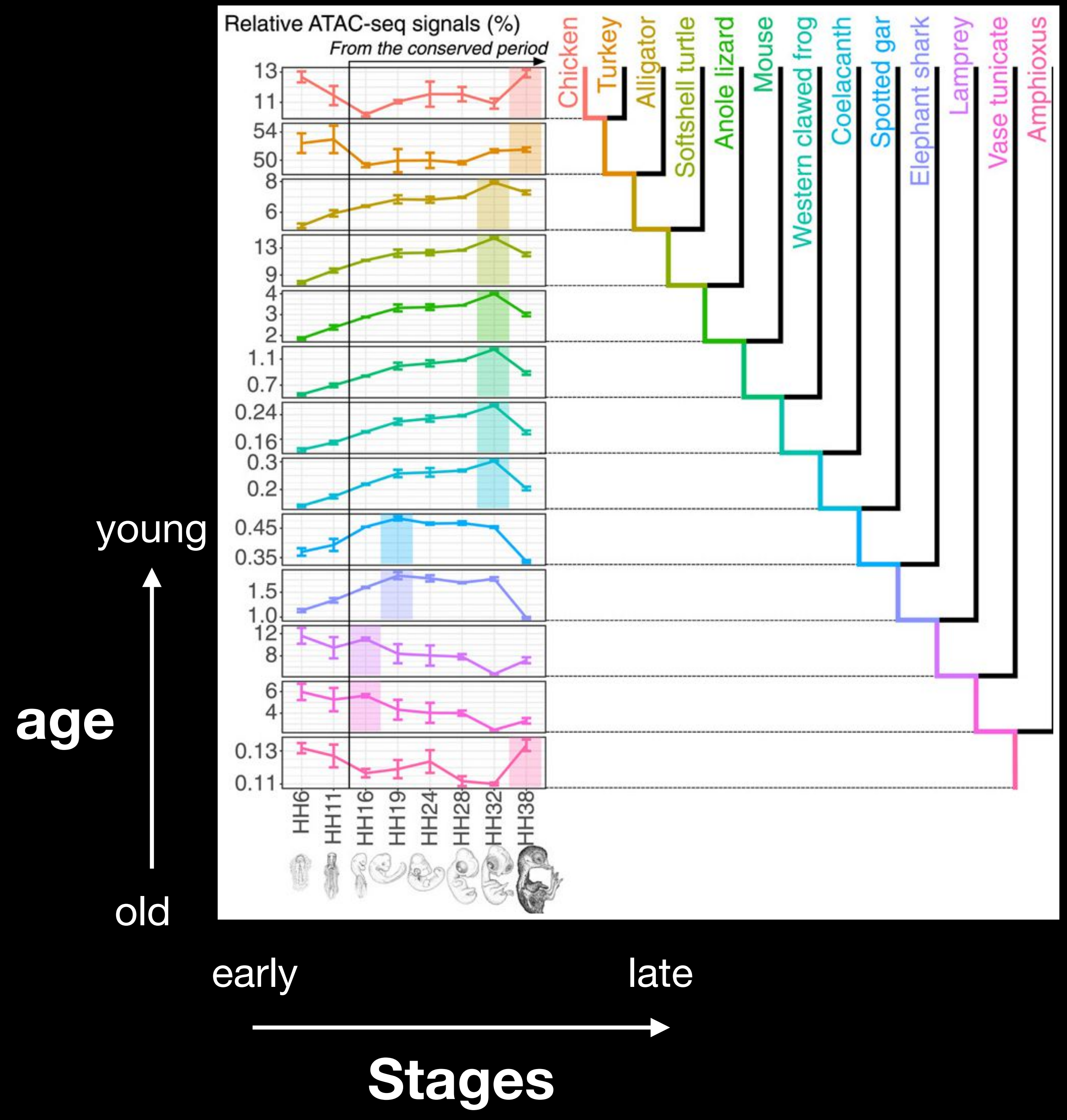
How can we use the age of genes?

What am I curious about ……

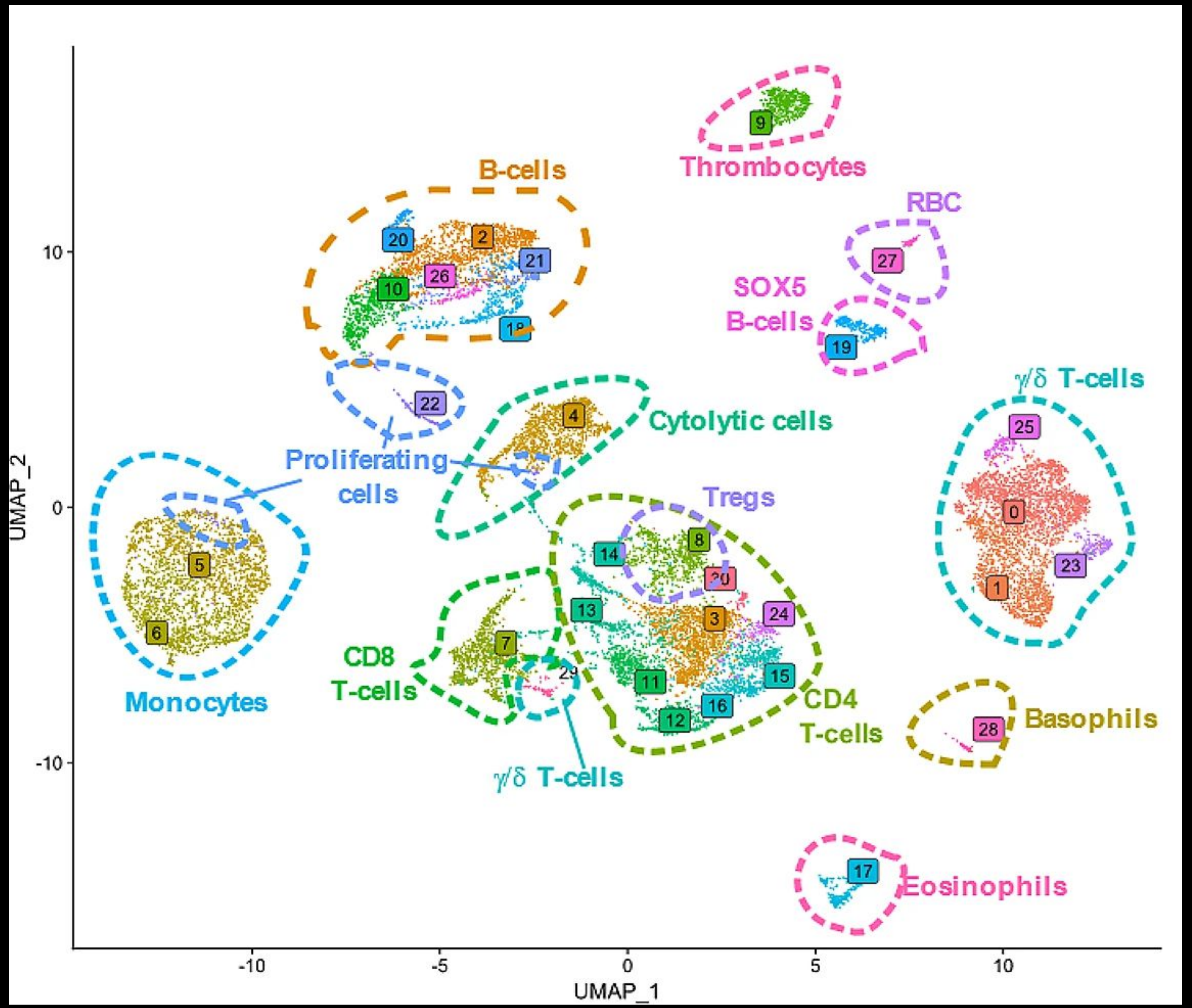
GenTree
The Time Tree of Genes Along the Evolution History

Human Data **Fly Data** **Mouse Data**

Rat Data **Opossum Data** **Chicken Data**



The **ages** of the **key genes** (or marker genes) affecting **late development and morphological diversity** (wing shape, leg or tail, to the extreme, the sex organs)?



Uesaka, M. et al. 2021.
Li, S. et al. 2019.

Thanks for your attention!

Q & A

Guangji Chen 2025.Apr.25

Due to time limitations,
a lot has not been shared yet...

Outline

1. Molecular mechanisms to generate new genes
    - *De novo* gene generation
2. How did new genes evolve?
    - Evolutionary forces acting on new genes
    - Rate of new gene evolution
    - Patterns of new gene evolution
3. How to date gene ages?
    - How can we use the age of genes?

- Molecular mechanisms to generate new genes
    - Diverse molecular mechanisms for forming new gene structures
    - The generation of new regulatory elements in new genes
- De novo gene generation
    - De novo genes versus orphan genes
    - The prevalence of de novo genes
- Rate of new gene evolution
    - Methods to date gene ages
- New genes with detectable molecular functions and fertility effects
    - New genes with the evolution of functional essentiality
    - New genes integrate into and change existing gene **networks**
- Patterns of new gene evolution
    - The rapid evolution of the nascent structure of de novo protein
    - Increasing structural complexity of de novo genes in evolution
    - Expression and interaction patterns of new genes
- Evolutionary forces acting on new genes
    - Functional and phenotypic evidence for adaptation and sexual conflicts
    - Sequence evolution and molecular population genetics
    - Molecular population genomics
- The **application** of new genes in agriculture and medicine
- Conclusions and perspectives

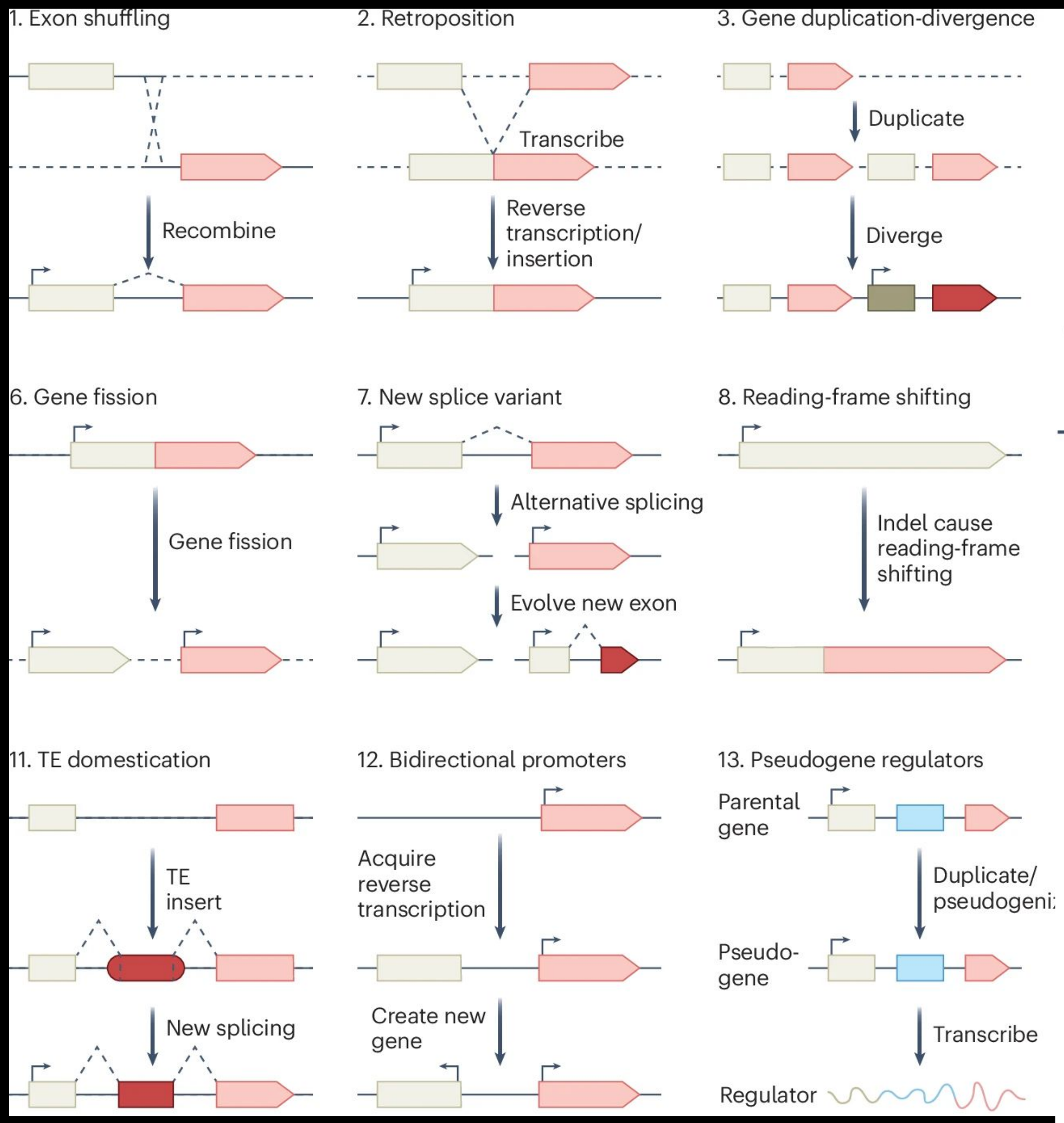


Figure 1

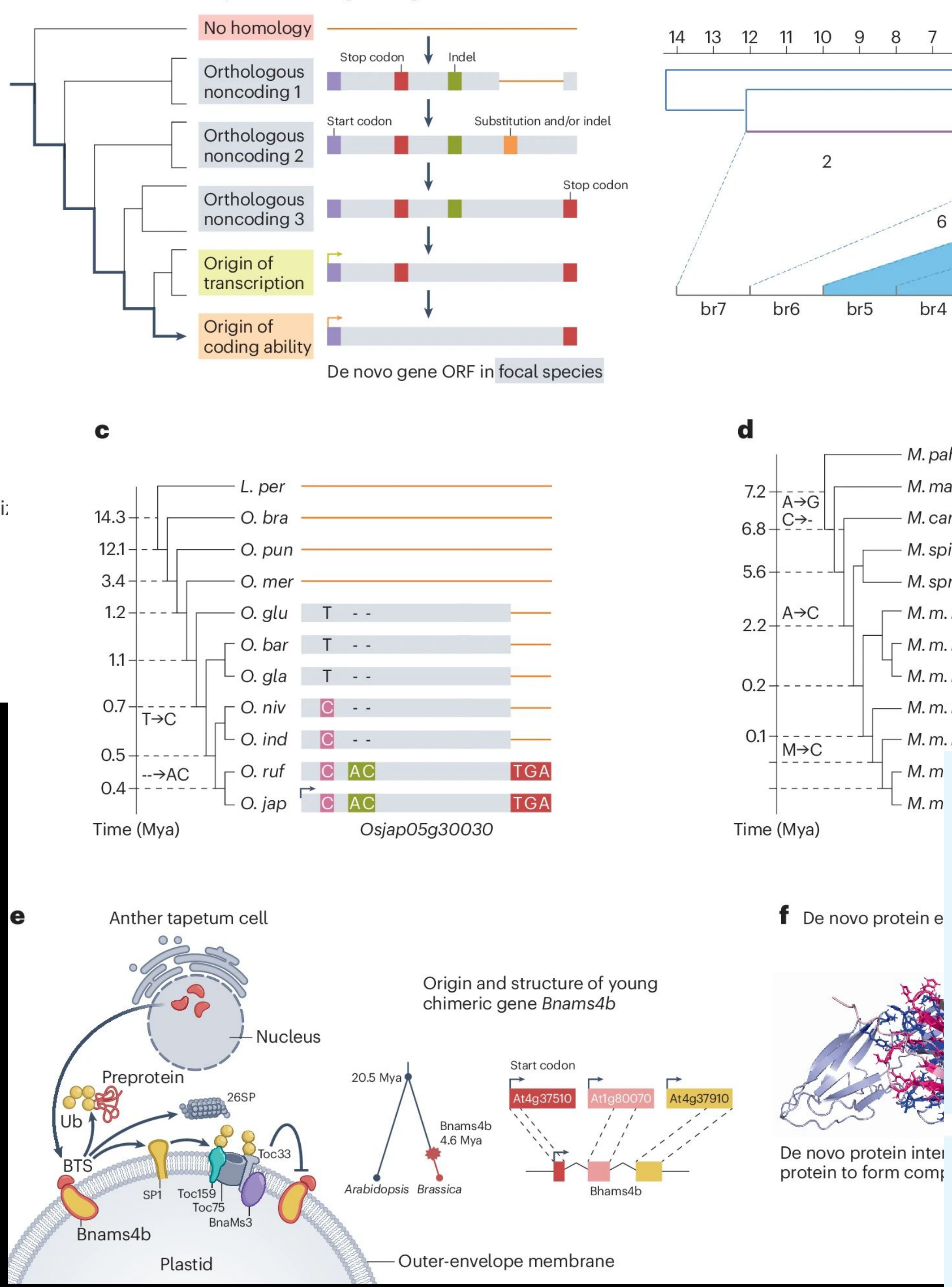


Figure 2

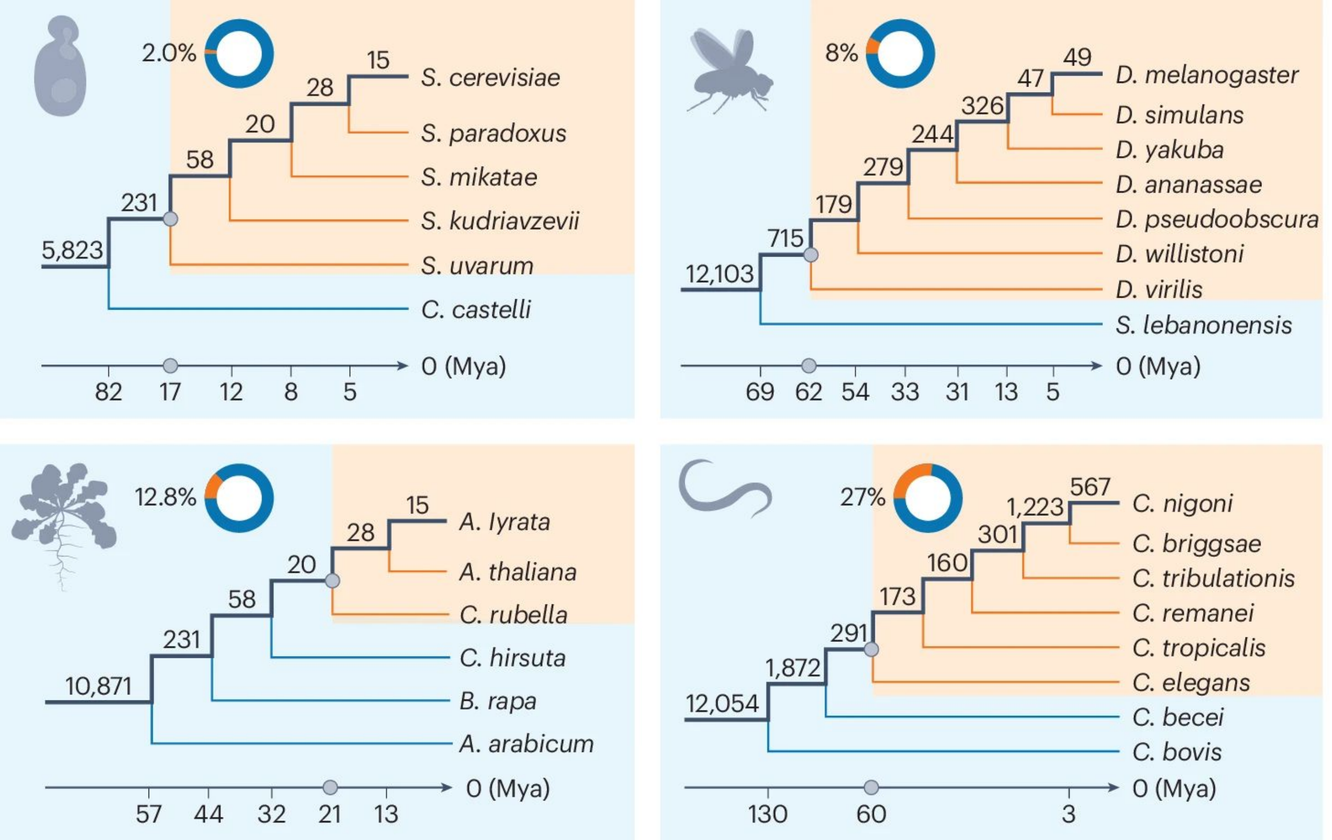


Figure 3

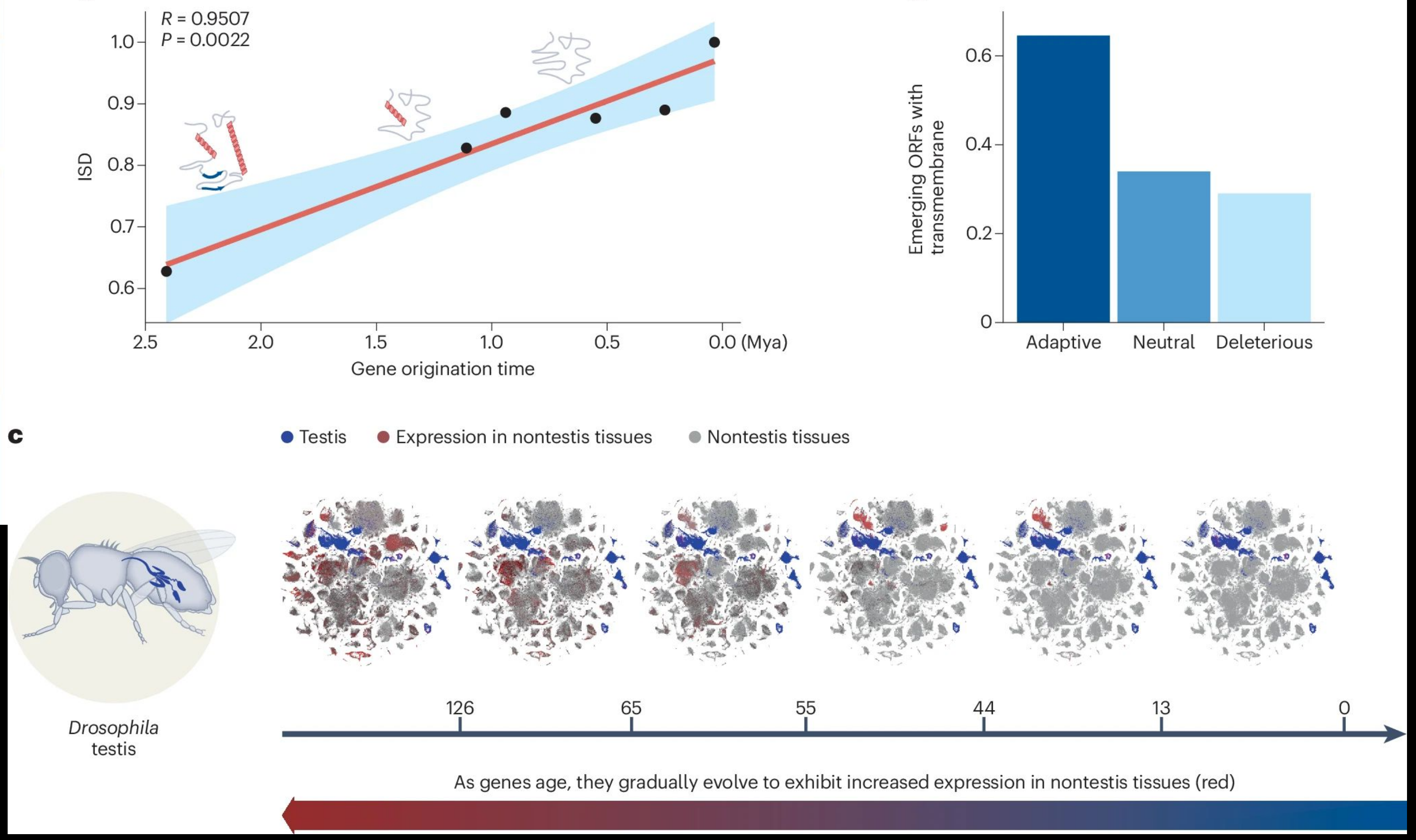
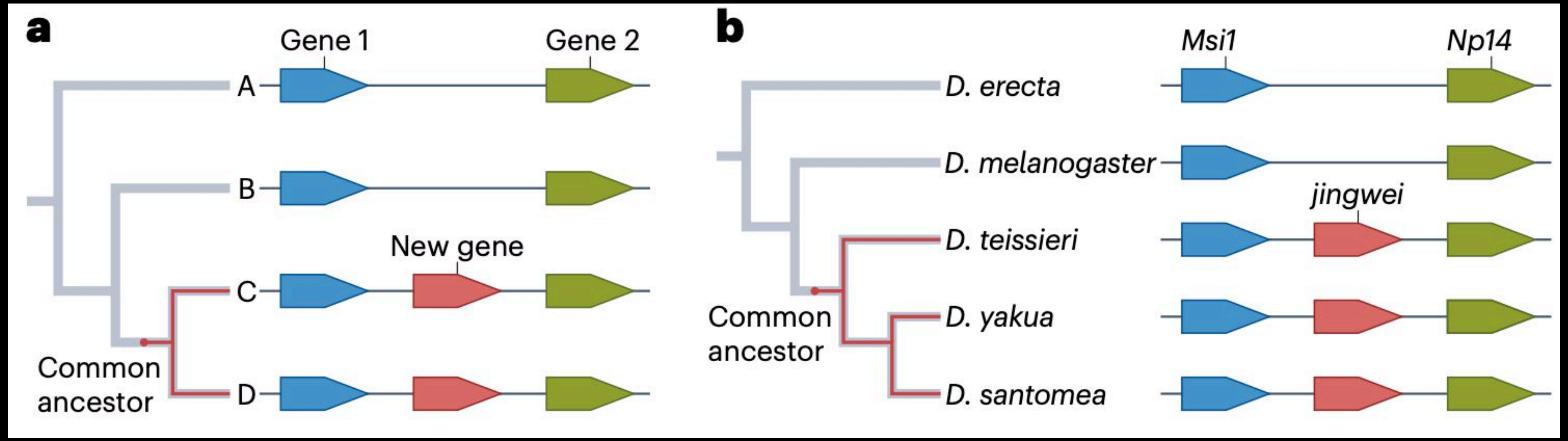


Figure 4



Box 1

Thanks again for your attention!

Q & A

Guangji Chen 2025.Apr.25