

Gene duplication and loss

Matthew Hahn
Indiana University

mwh@indiana.edu

How many genes does a human have:

- a. <1
- b. 1-10,000
- c. 10,000-100,000
- d. $>100,000$

How many genes does a human have:

- a. <1
- b. 1-10,000
- c. 10,000-100,000 (~23,000)
- d. $>100,000$

How many genes does a fruitfly have:

- a. <1
- b. 1-10,000
- c. 10,000-100,000
- d. $>100,000$

How many genes does a fruitfly have:

- a. <1
- b. 1-10,000
- c. 10,000-100,000 (~15,000)
- d. >100,000

Gene Number Variation

Corn	40,000 genes
Rice	33,000
Thale cress	28,000
Mouse	23,000
Human	23,000
Worm	19,000
Fruitfly	15,000
Yeast	6,000
<i>M. genitalium</i>	500

Gene duplication and loss result in genome size variation

	<i>S. cerevisiae</i>	<i>C. elegans</i>	<i>D. melanogaster</i>	<i>H. sapiens</i>	<i>A. thaliana</i>
Homeodomain	9	109	148	267	118
Zinc-finger	121	437	357	706	1049
Nuclear receptor	1	183	25	59	4

from Venter et al. (2001)

Similar genomes have similar numbers of genes

Insects: ~15,000
Mammals: ~23,000
Worms: ~19,000
Fungi: 6,000-10,000

Despite this, lots of variation in individual genes.

Gene duplication and loss

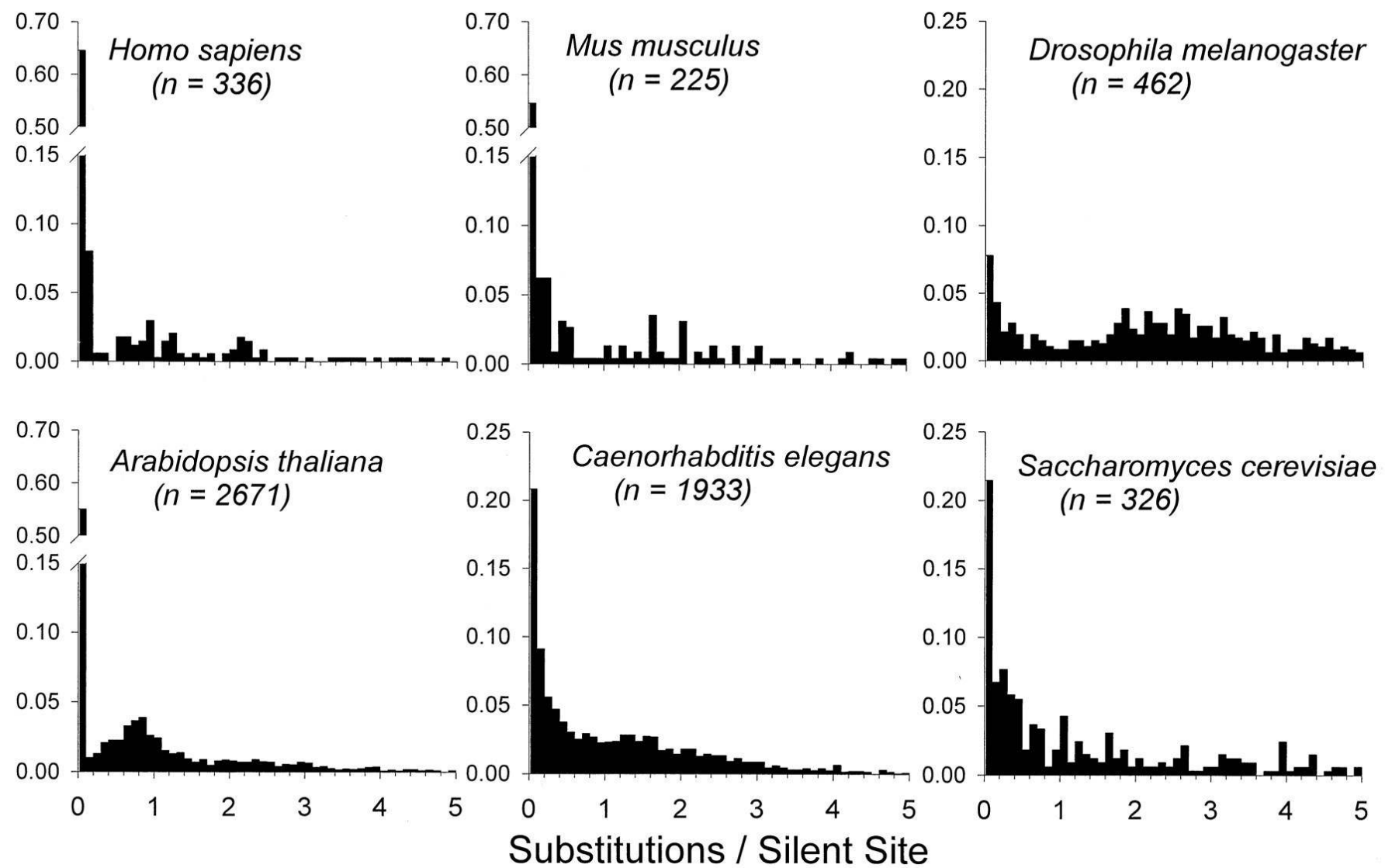
Table 1. Prevalence of gene duplication in all three domains of life^a

	Total number of genes	Number of duplicate genes (% of duplicate genes)	Refs
Bacteria			
<i>Mycoplasma pneumoniae</i>	677	298 (44)	[65]
<i>Helicobacter pylori</i>	1590	266 (17)	[66]
<i>Haemophilus influenzae</i>	1709	284 (17)	[67]
Archaea			
<i>Archaeoglobus fulgidus</i>	2436	719 (30)	[68]
Eukarya			
<i>Saccharomyces cerevisiae</i>	6241	1858 (30)	[67]
<i>Caenorhabditis elegans</i>	18 424	8971 (49)	[67]
<i>Drosophila melanogaster</i>	13 601	5536 (41)	[67]
<i>Arabidopsis thaliana</i>	25 498	16 574 (65)	[69]
<i>Homo sapiens</i>	40 580 ^b	15 343 (38)	[11]

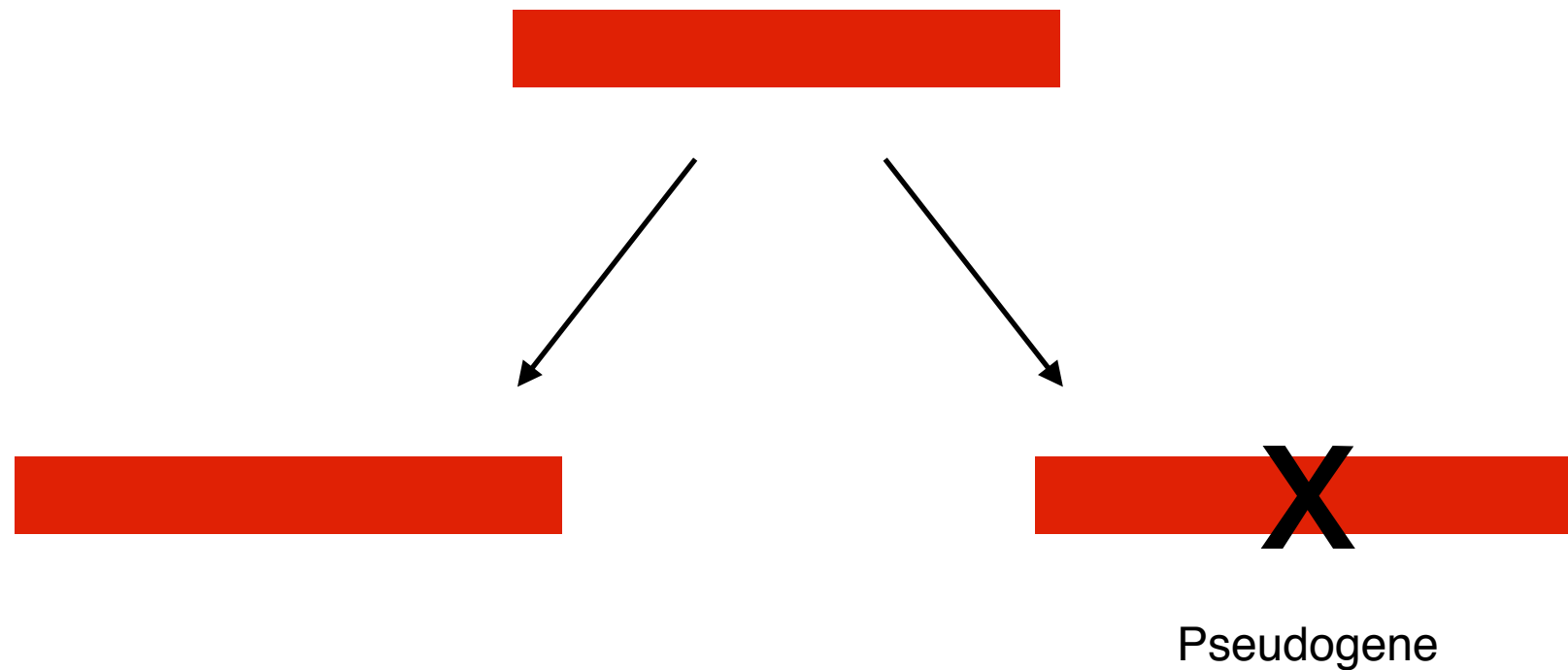
^aUse of different computational methods or criteria results in slightly different estimates of the number of duplicated genes [12].

^bThe most recent estimate is ~30 000 [61].

Gene duplication and loss

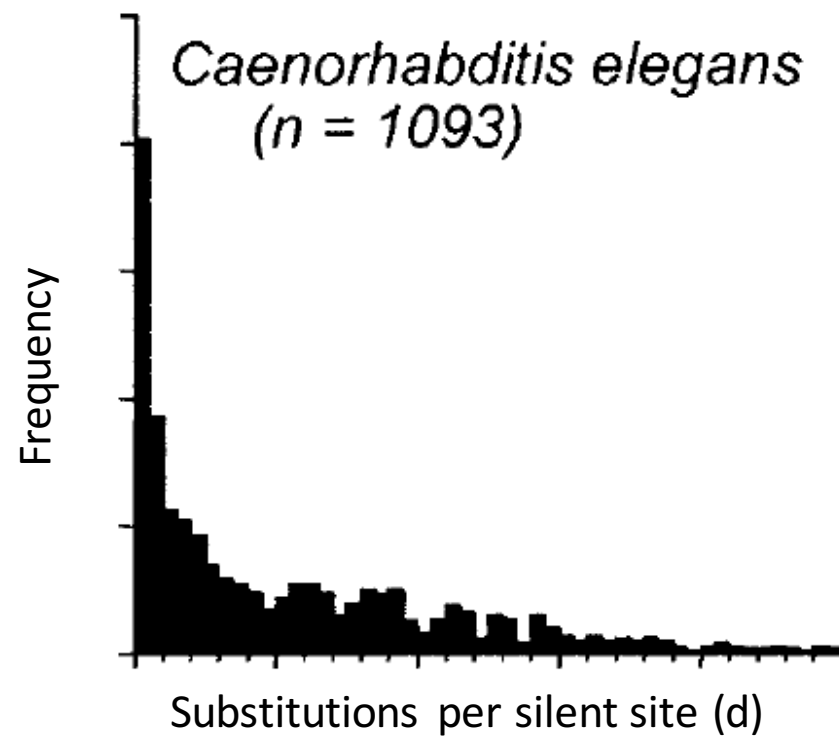


Gene duplication and loss



The most common outcome of duplication is loss

Gene duplication and loss



The most common outcome of duplication is loss

Big questions in gene duplication

“The main interest in duplications lay in ... identical genes which could subsequently mutate separately and diversify their effects.”

--Bridges 1918

Big questions in gene duplication

210

SCIENCE

VOL. 83, No. 2148

SPECIAL ARTICLES

THE BAR "GENE" A DUPLICATION

THE nature of the Bar gene has been the subject of extensive investigation and speculation since February, 1913, when Tice¹ found this reduced-eye mutant as a single male in the progeny of normal-eyed parents. The eye-reduction behaves as a sex-linked dominant, with a locus at 57.0, and has been one of the most important of all the sex-linked characters of *D. melanogaster*. A remarkable peculiarity of the mutant is that occasionally the homozygous stock gives rise to a fly indistinguishable in appearance and genetic behavior from wild-type.² More rarely the stock gives rise to an even more extreme reduction in eye-size, a type which was called Ultra-Bar by Zeleny,³ who found it.

Sturtevant and Morgan⁴ and Sturtevant⁵ found that these two-way changes were the result of a novel type of "unequal" crossing-over, by which the two genes originally present in the two parental chromosomes both emerged in the same chromosome (Bar-double) while the other resultant chromosome was without Bar

The exact point of the insertion is ambiguous, for a reason which will appear below. The normal X in this region (see revised map in Fig. 1) shows in sub-se-

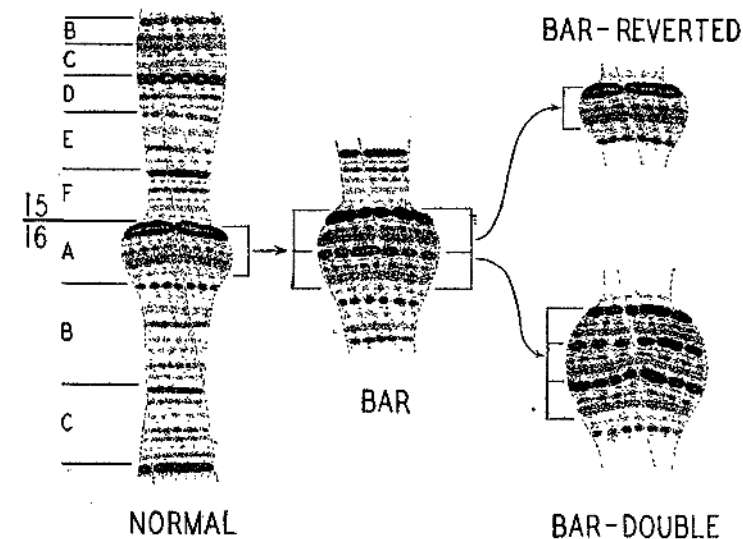


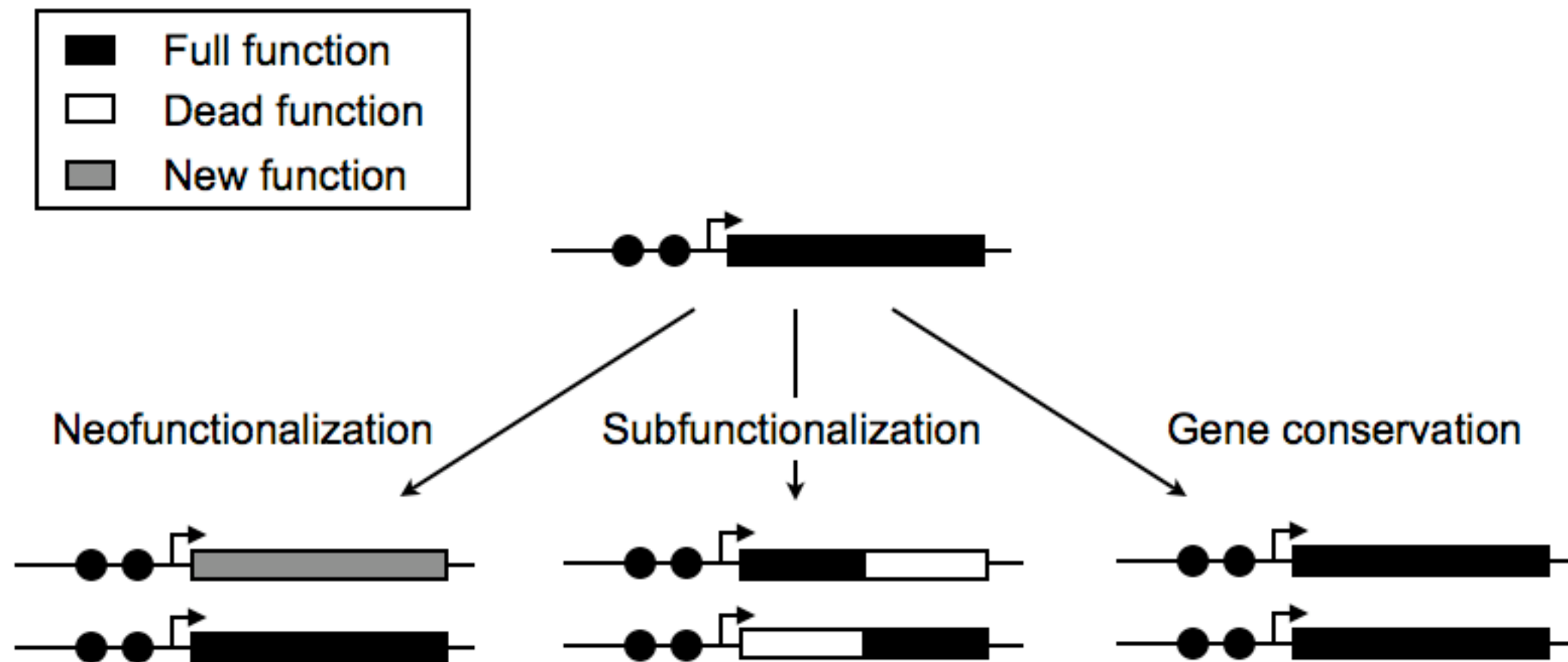
FIG. 1

tion 16A a heavy band, which in well-stretched chro-

Bridges (1935)

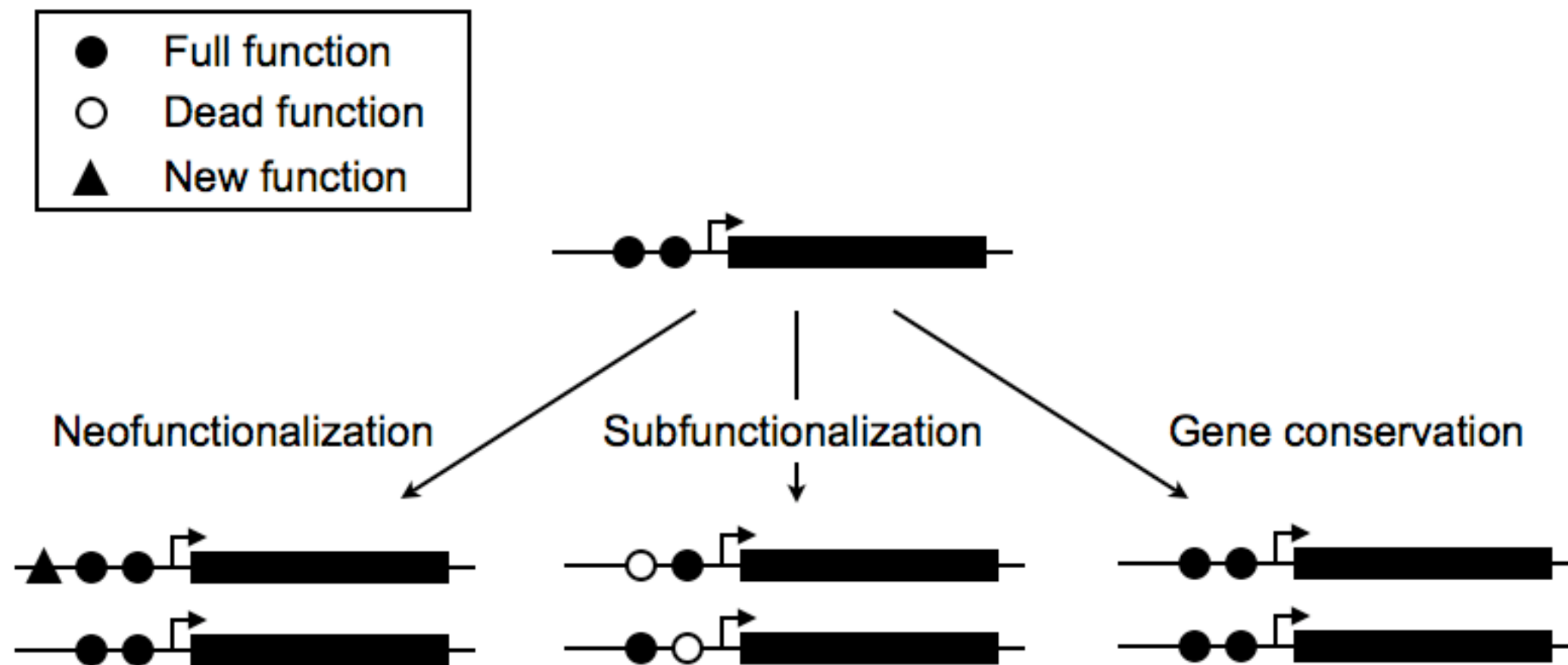
Big questions in gene duplication

Figure 1b

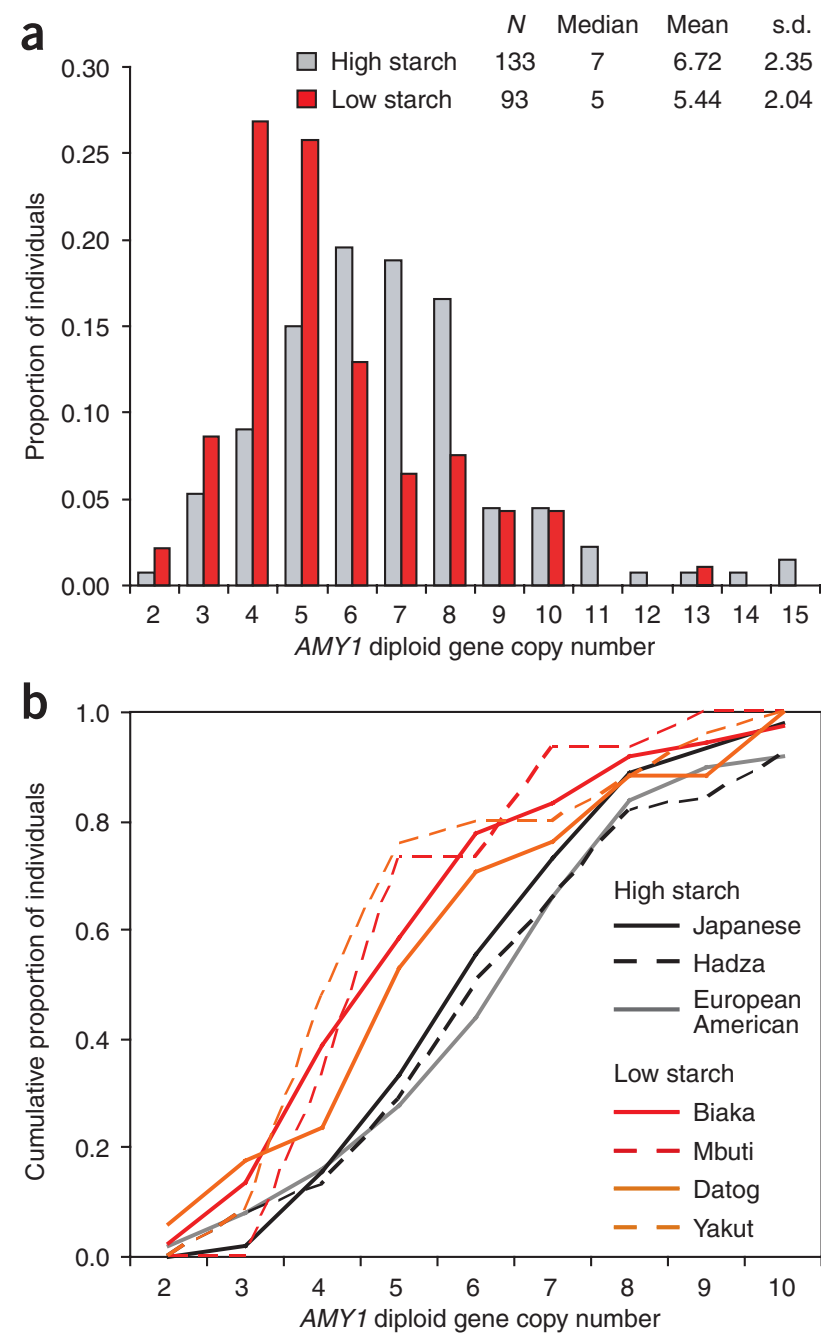


Big questions in gene duplication

Figure 1a



Gene conservation



Perry et al. (2007)

Subfunctionalization

Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 3479–3483, May 1988
Evolution

Gene sharing by δ -crystallin and argininosuccinate lyase

(lens proteins/evolution/gene expression/enzymes/urea cycle)

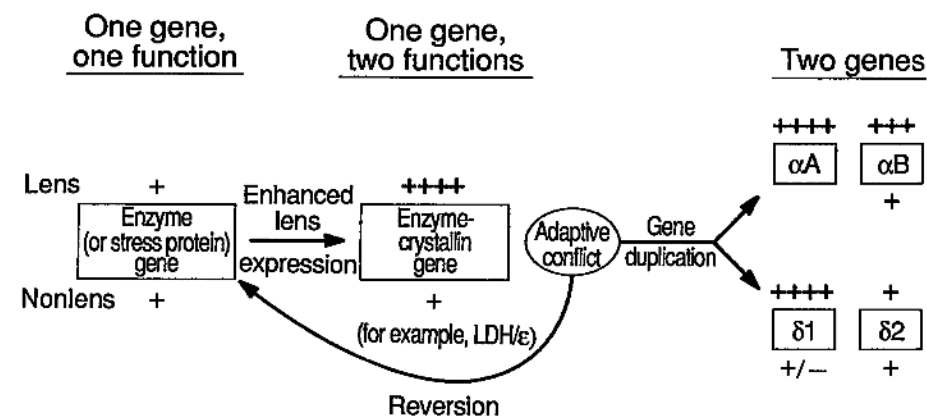
JORAM PIATIGORSKY*, WILLIAM E. O'BRIEN†, BARBARA L. NORMAN*, KAREN KALUMUCK†, GRAEME J. WISTOW*, TERESA BORRAS*, JOHN M. NICKERSON*, AND ERIC F. WAWROUSEK*

*Laboratory of Molecular and Developmental Biology, National Eye Institute, National Institutes of Health, Bethesda, MD 20892; and †Baylor College of Medicine, and Institute for Molecular Genetics, the Howard Hughes Medical Institute, Houston, TX 77030

Communicated by Donald D. Brown, February 1, 1988

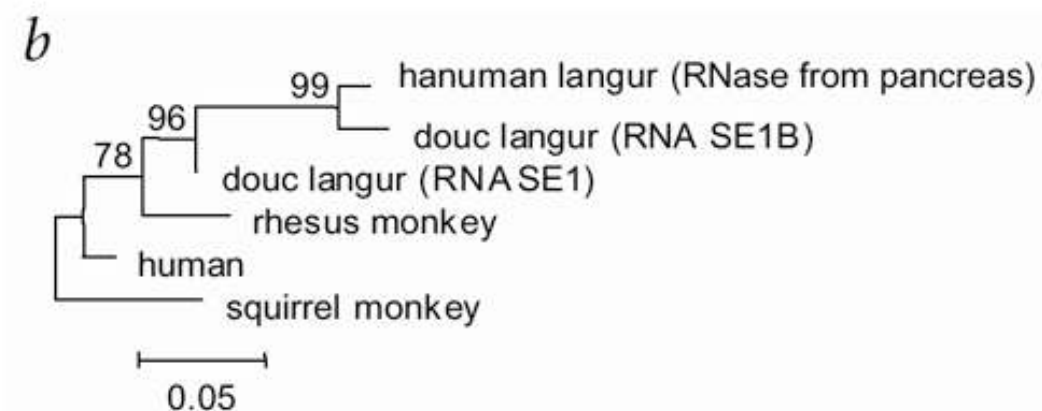
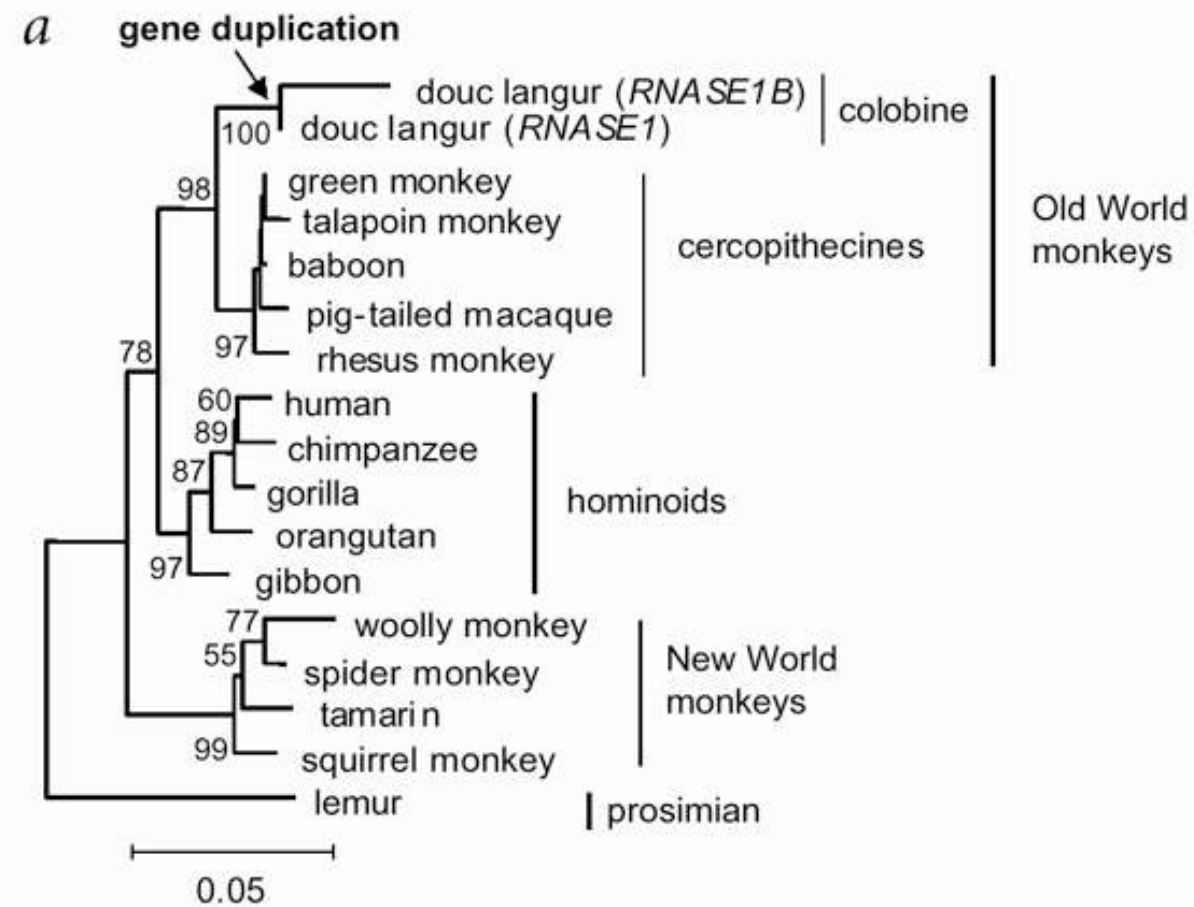
ABSTRACT The lens structural protein δ -crystallin and the metabolic enzyme argininosuccinate lyase (ASL; L-argininosuccinate arginine-lyase, EC 4.3.2.1) have striking sequence similarity. We have demonstrated that duck δ -

Although uricotelic, birds have some activity for ASL as well as for other enzymes of the urea cycle. δ -Crystallin is the dominant crystallin in lenses of birds and reptiles, but it is absent from lenses of mammals (2). Southern blot hybridiza-



Piatigorsky and Wistow (1991)

Neofunctionalization

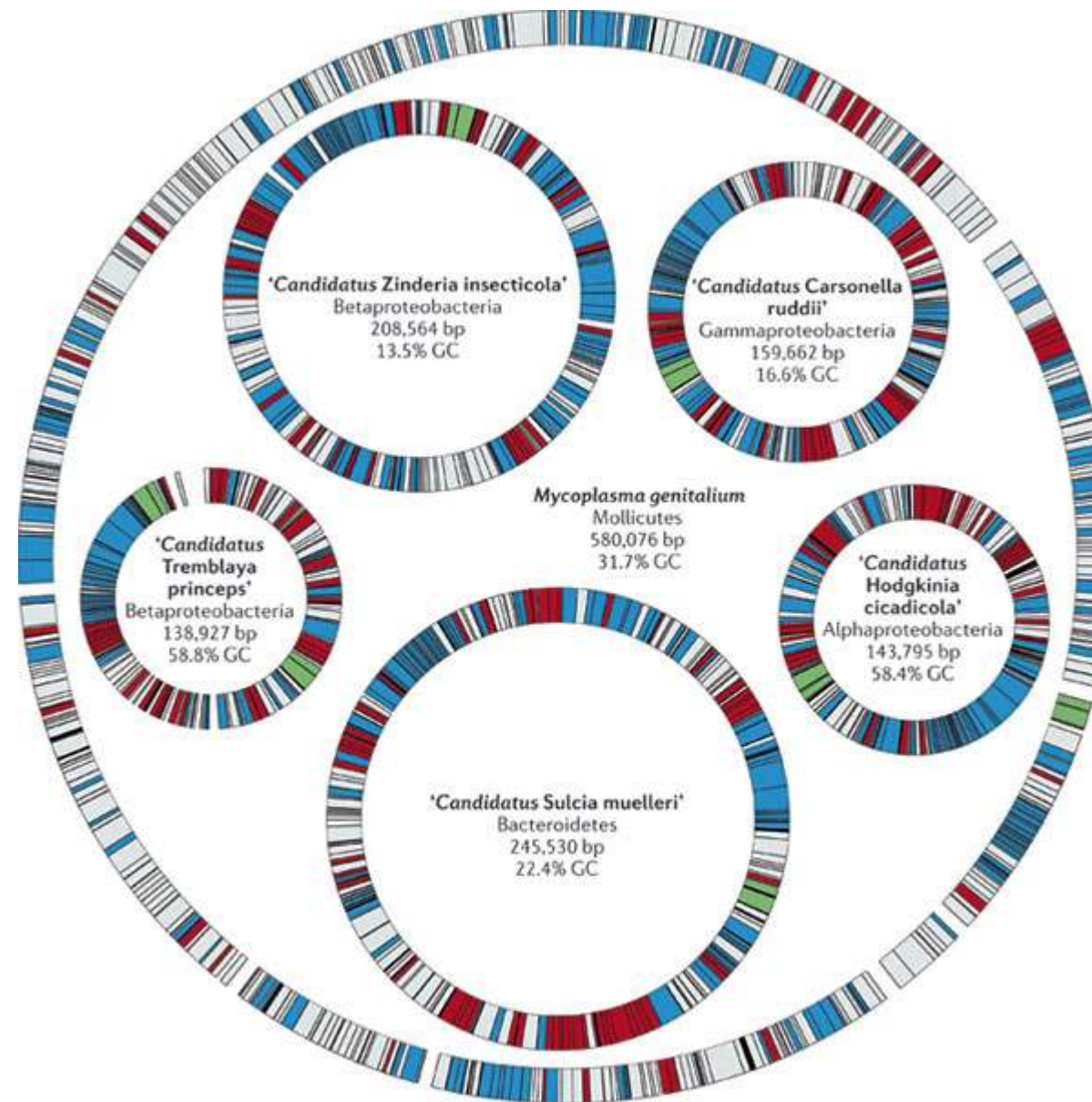


Big questions in gene duplication

Neofunctionalization, subfunctionalization, or conservation?

Positive selection or genetic drift?

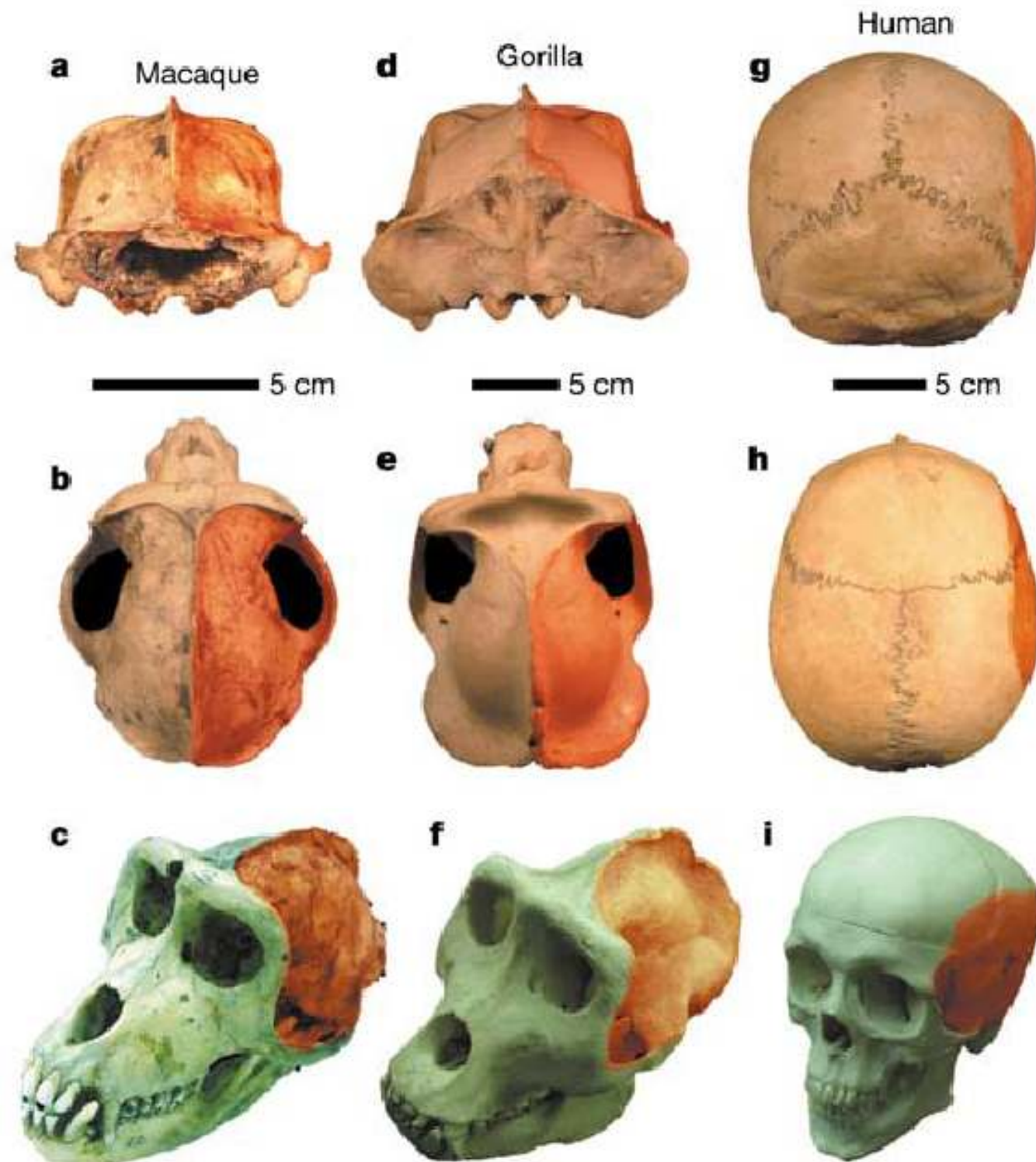
Big questions in gene loss



Nature Reviews | Microbiology

McCutcheon and Moran (2012)

Loss of *Myh16* associated with cranial enlargement



Molecular mechanisms of gene duplication

DNA- or RNA-based

Multiple genes, single genes, partial genes

Molecular mechanisms of gene duplication

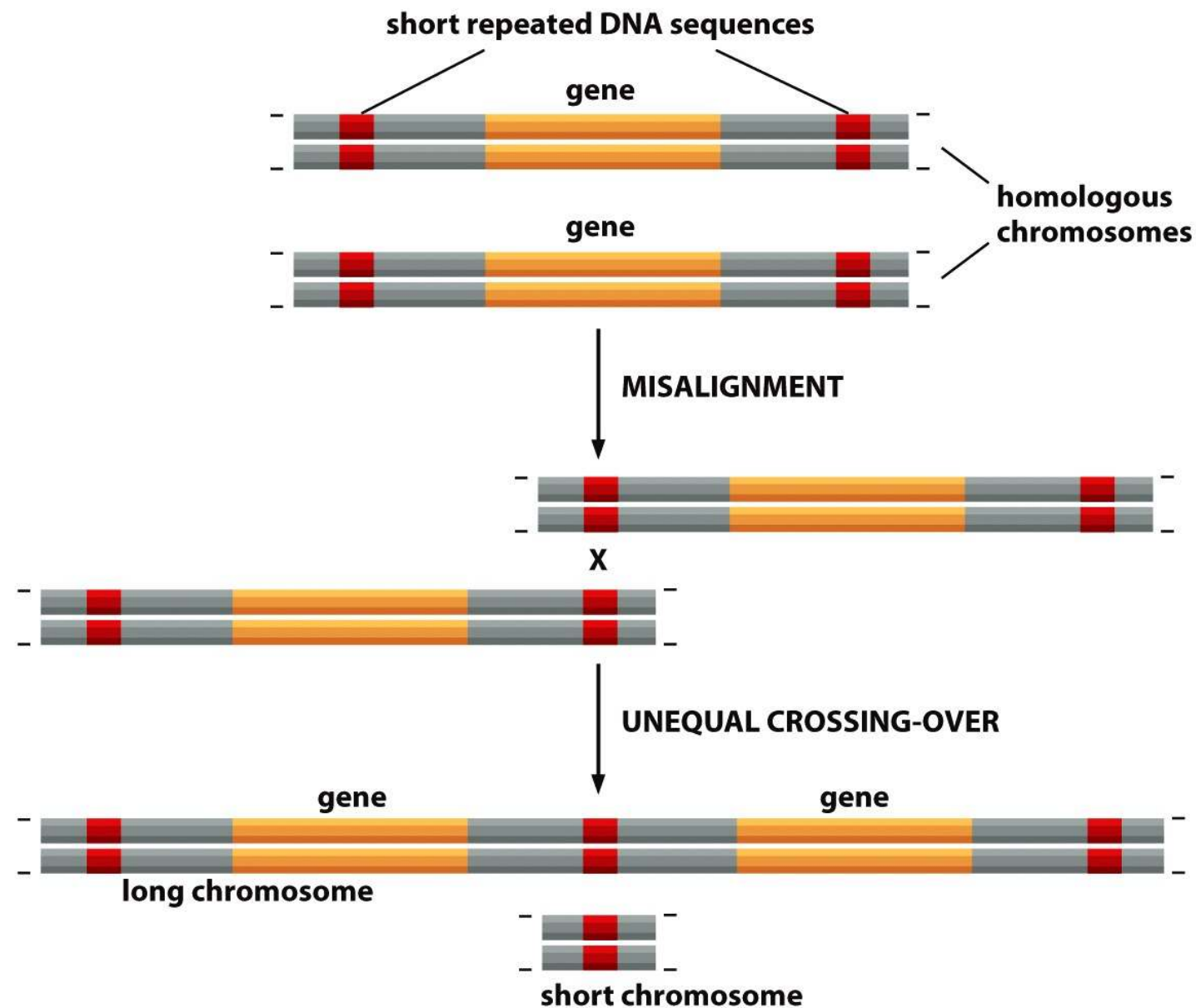


Figure 9-9 Essential Cell Biology 3/e (© Garland Science 2010)

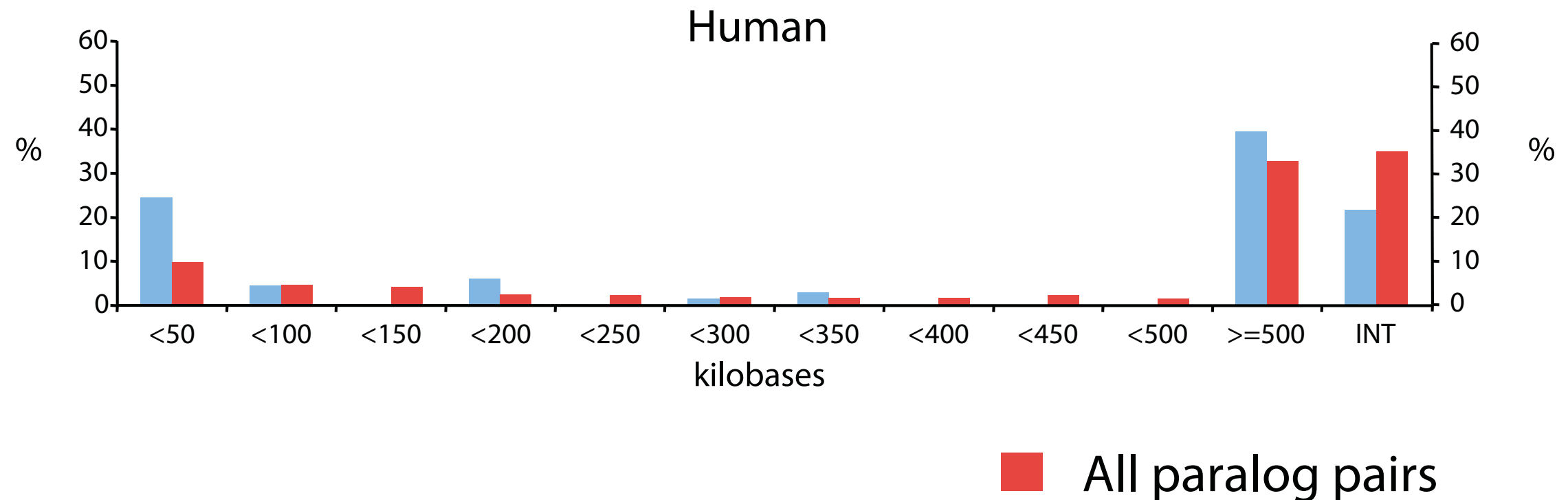
“Unequal crossing-over”

Molecular mechanisms of gene duplication

“Unequal crossing-over”:

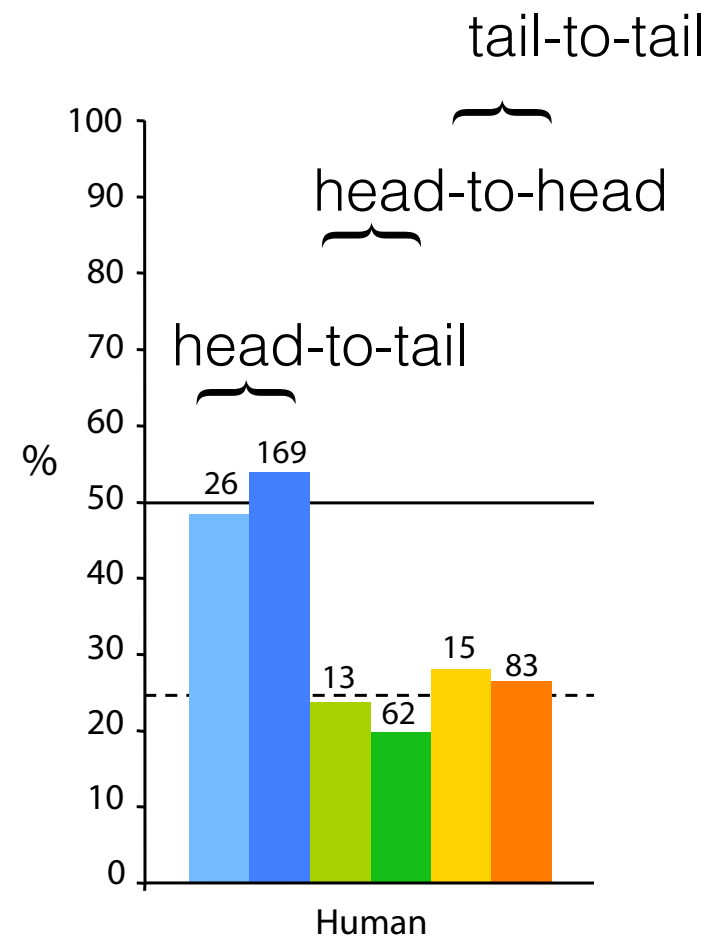
- Requires repeated elements to be present
- Is generally due to NAHR
- Does not always result in tandem duplicates

Molecular mechanisms of gene duplication

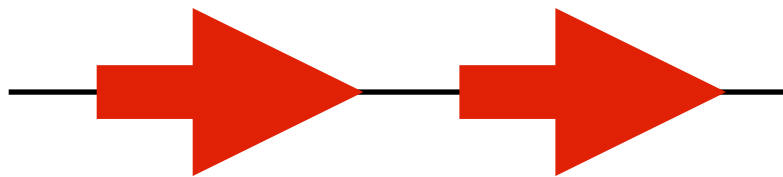


Hypothesis: the more TEs there are in a genome, the farther apart the duplicates are

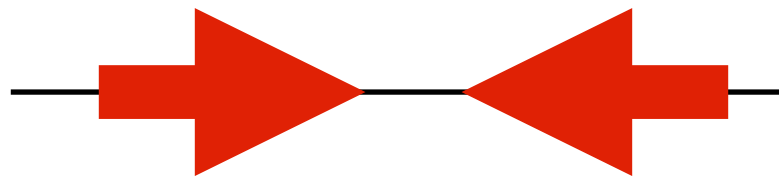
Molecular mechanisms of gene duplication



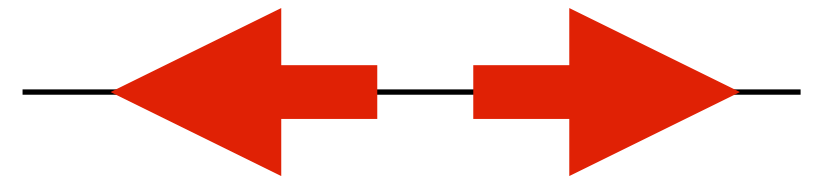
head-to-tail



head-to-head

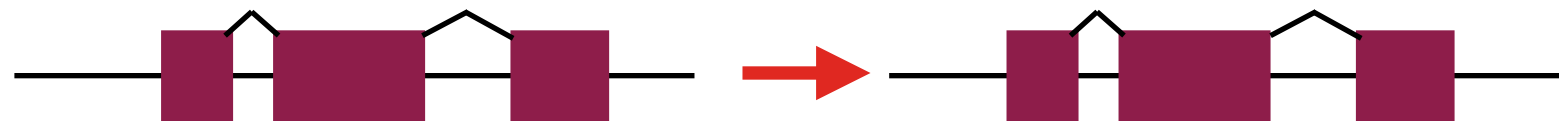


tail-to-tail

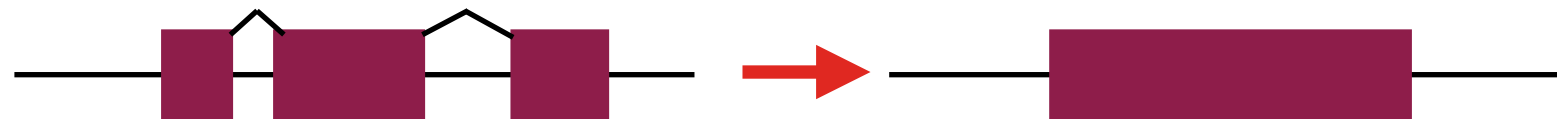


Molecular mechanisms of gene duplication

DNA-based mechanisms



RNA-based mechanisms



“Retrotransposition”

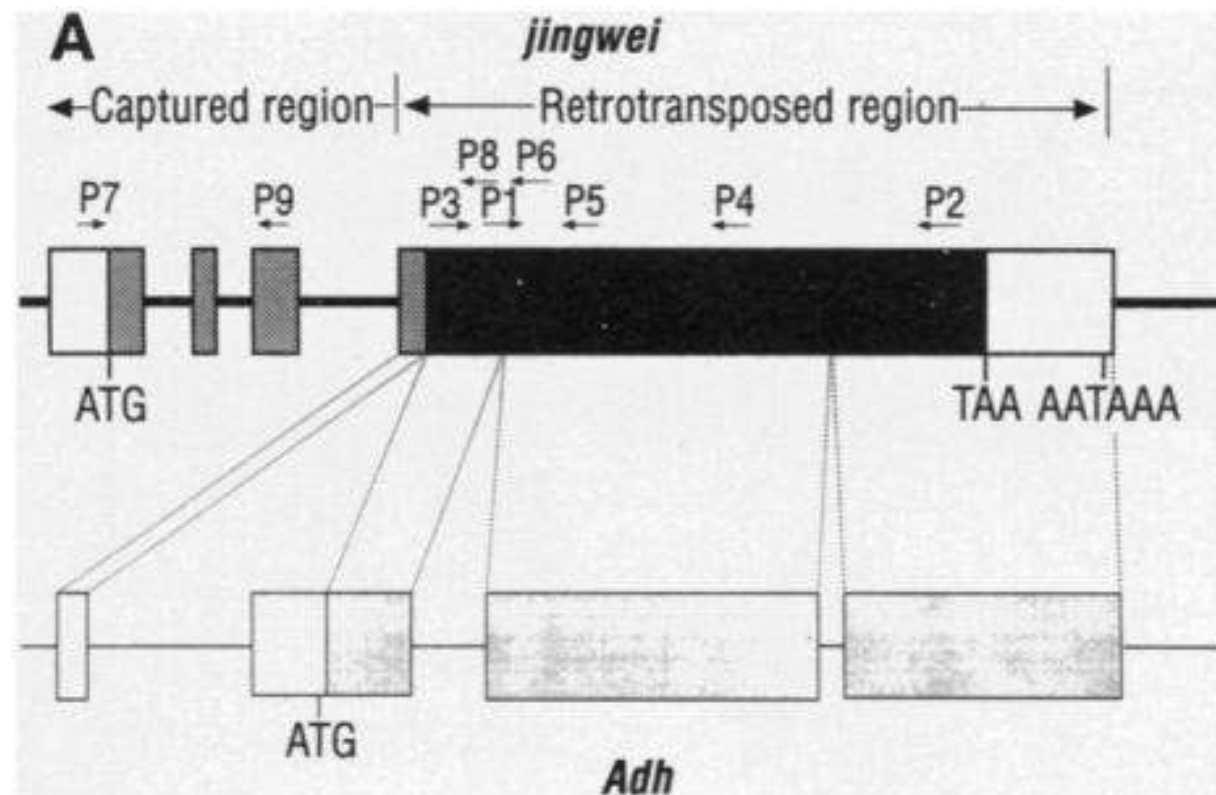
Molecular mechanisms of gene duplication

Retrotransposition:

- Results in a daughter copy without introns
- Brings along (almost) no flanking sequence
- Can only copy one gene at a time

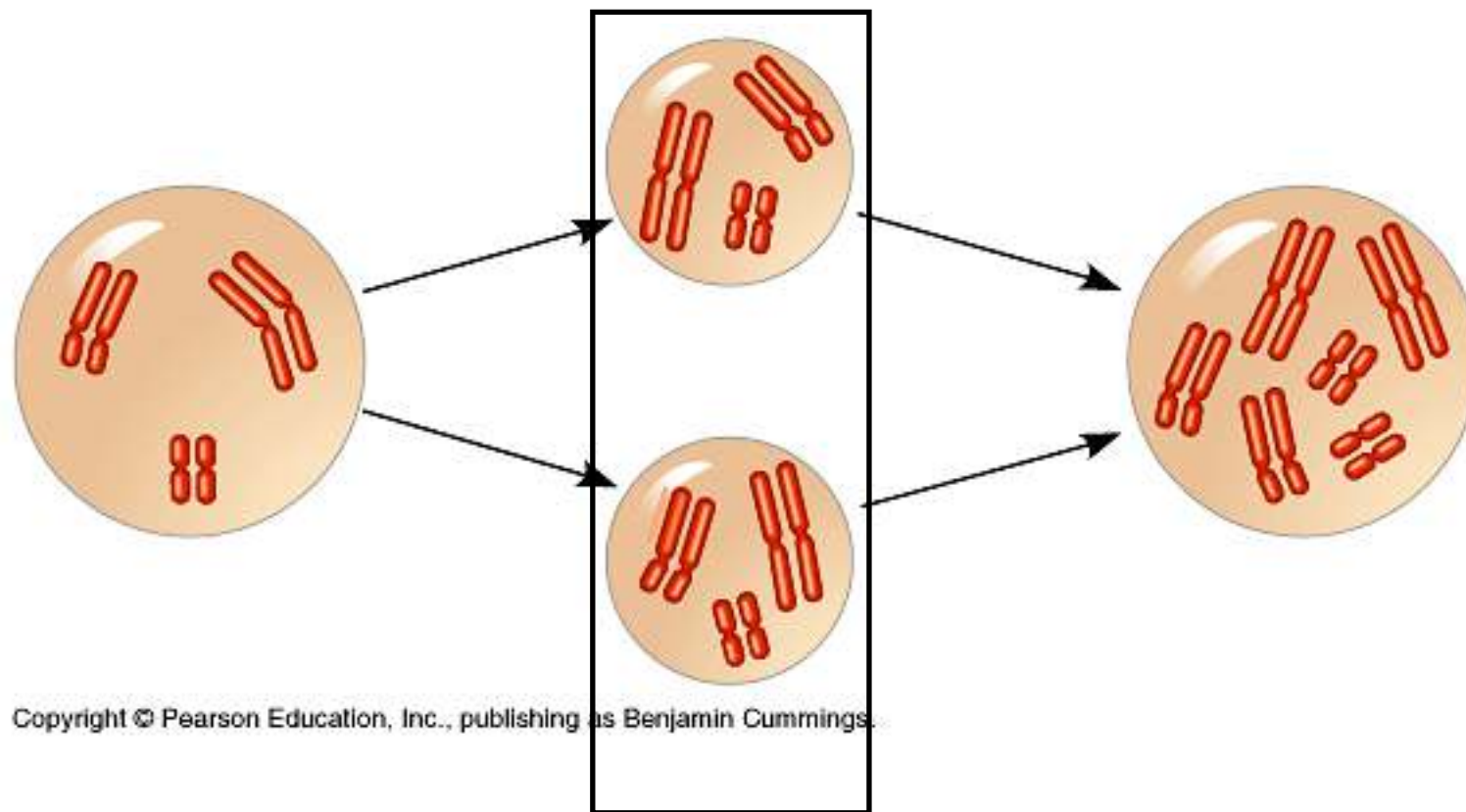
Molecular mechanisms of gene duplication

Weird hybrid case: “chimeric” gene duplicates



Molecular mechanisms of gene duplication

Polyploidy



Unreduced gametes

Molecular mechanisms of gene duplication

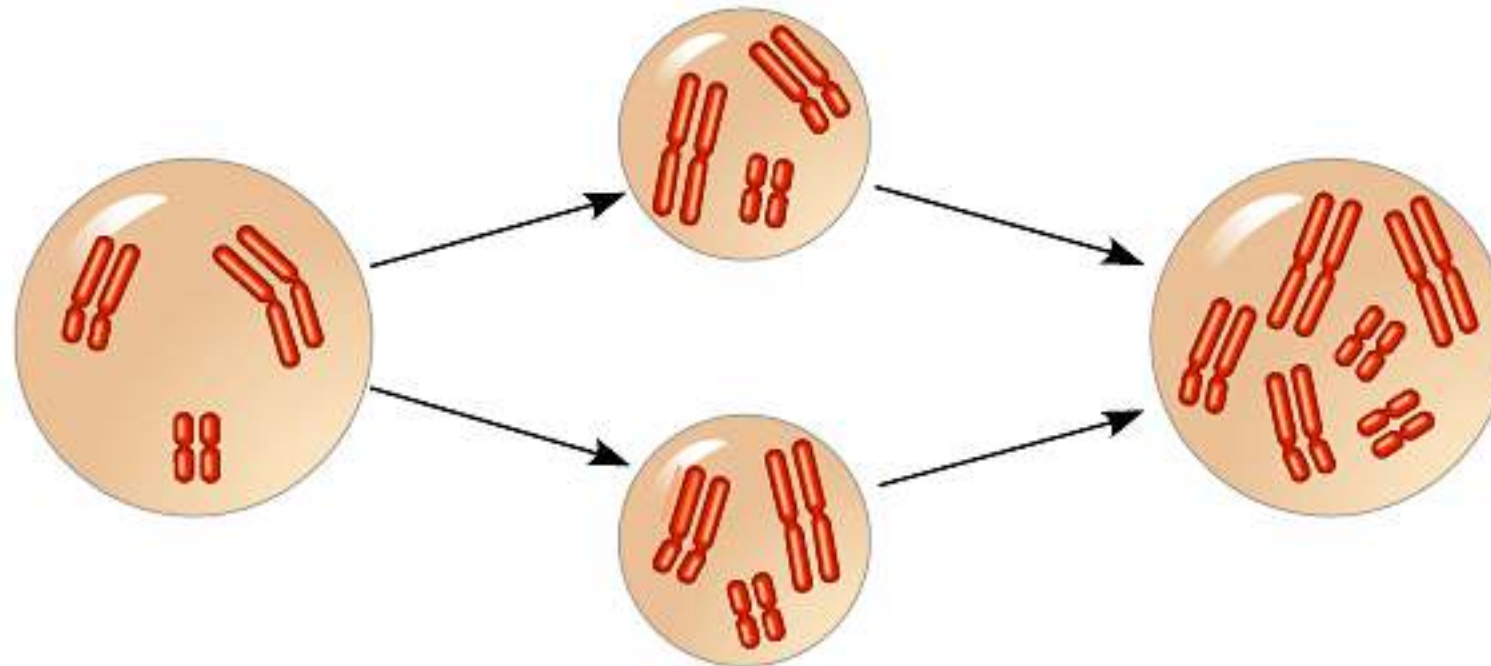
Polyploidy:

- Doubles the entire content of the genome
- Most genes subsequently return to single-copy
- Two types (at approximately equal frequency in nature):
Autopolyploidy and allopolyploidy

Molecular mechanisms of gene duplication

Autopolyploidy:

Doubling the number of chromosomes, where both parents are from the same species (or same individual)

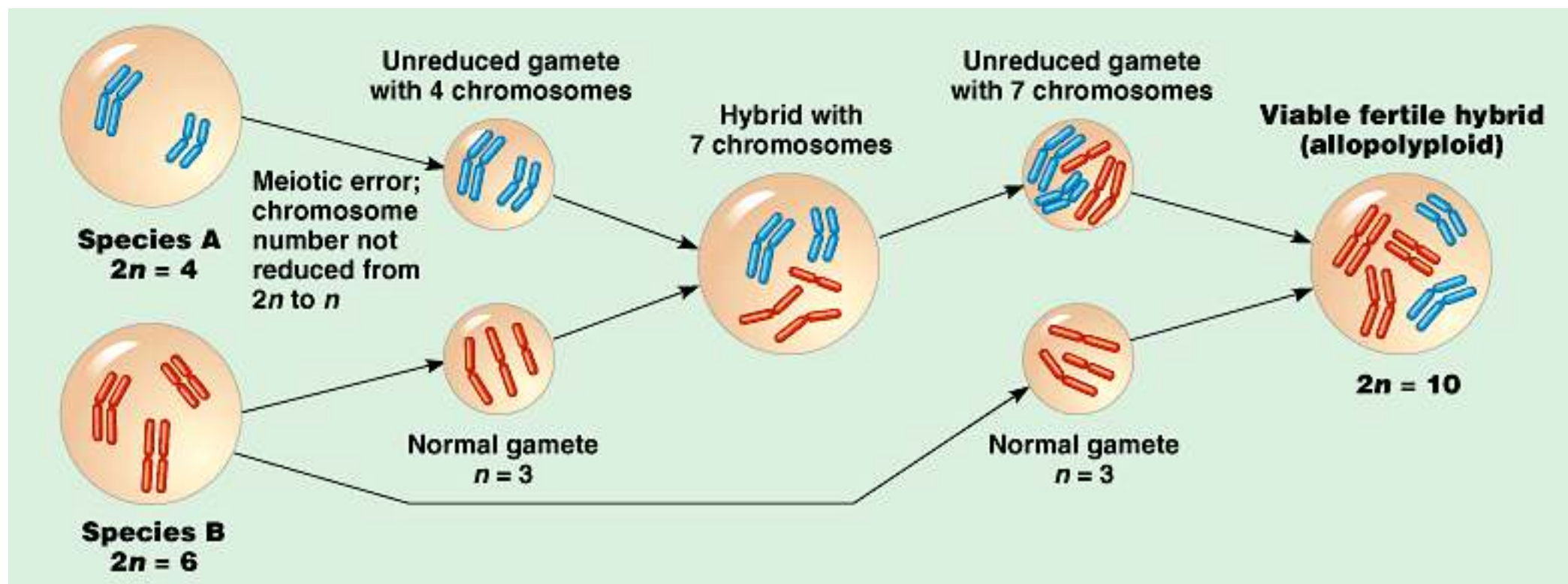


Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

Molecular mechanisms of gene duplication

Allopolyploidy:

Doubling the number of chromosomes, where the parents are from *different* species



Molecular mechanisms of gene loss

Nonsense mutation

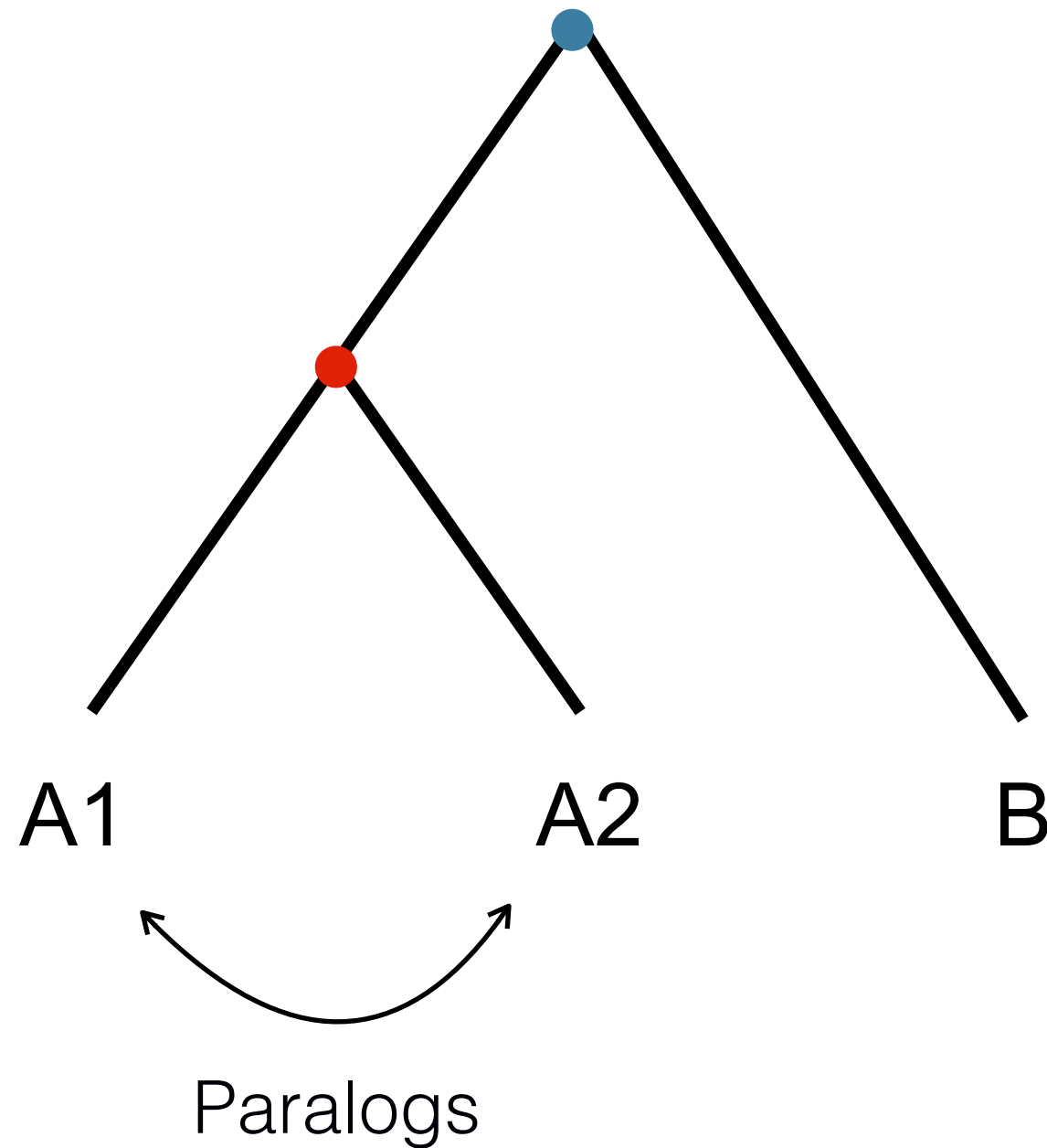
Frame shifting indel

Complete deletion (often due to NAHR)

Genealogical relationships among genes

Paralogs: genes (loci) whose most recent common ancestor is a duplication node ●

Genealogical relationships among genes

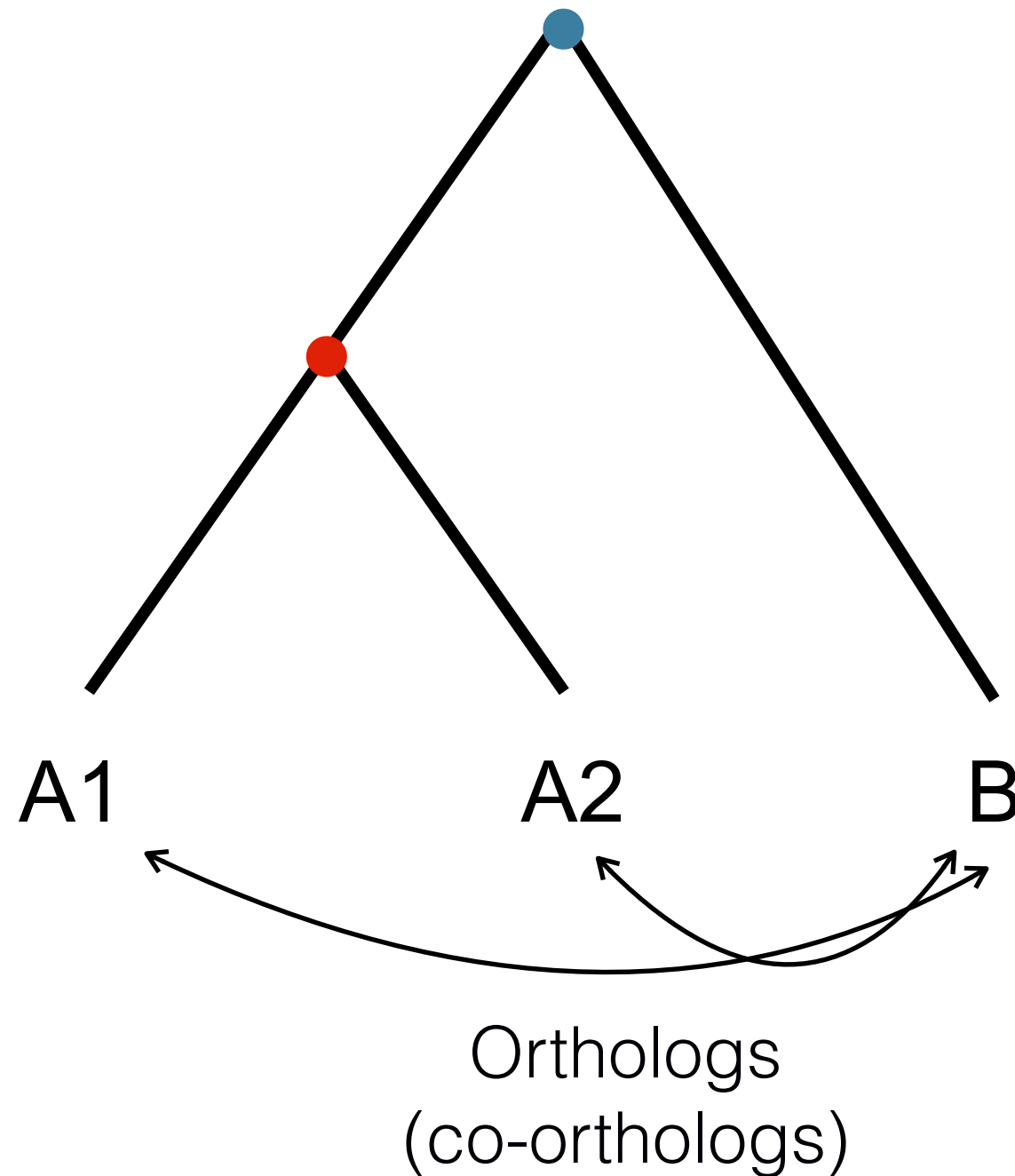


Genealogical relationships among genes

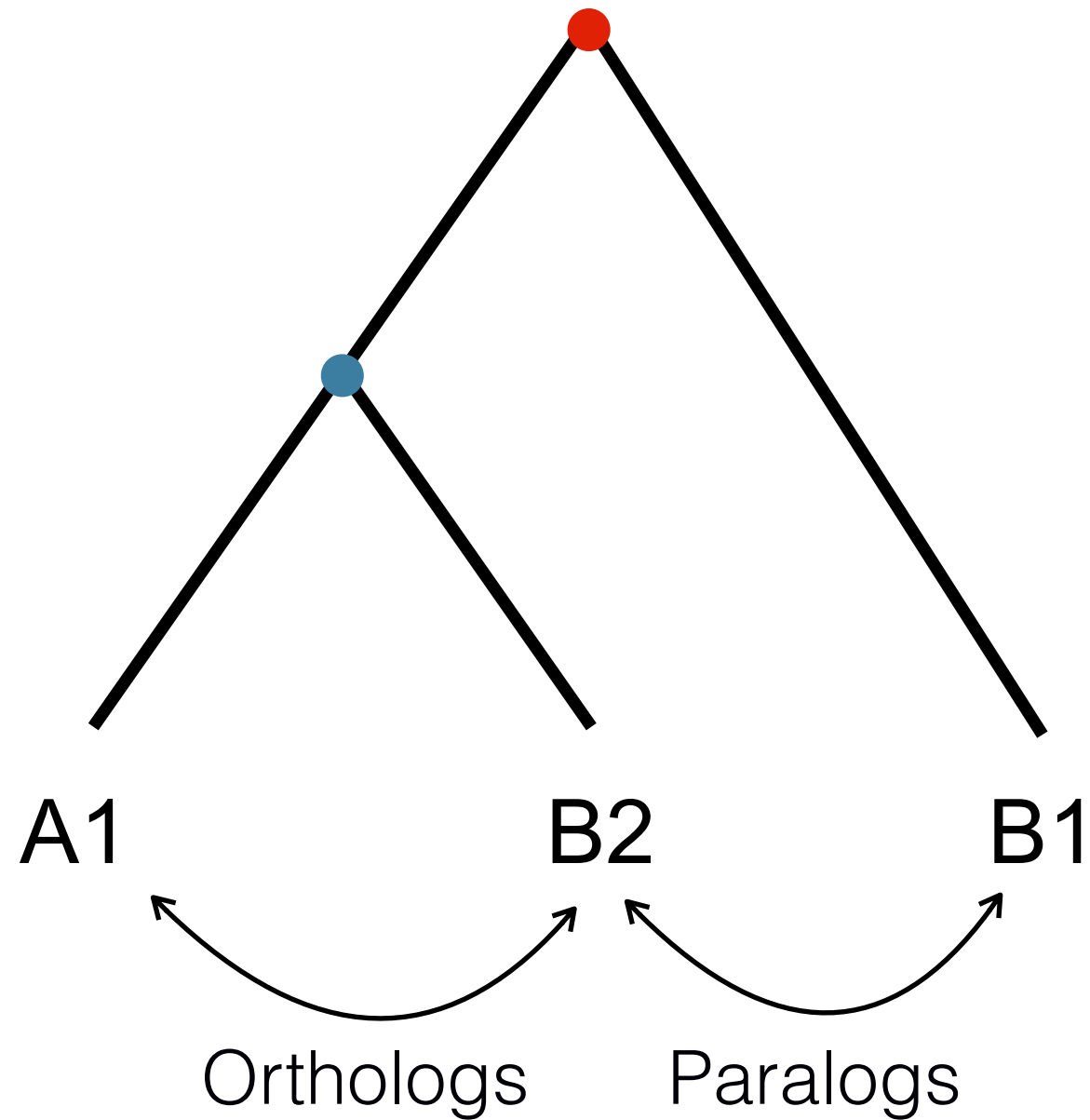
Paralogs: genes (loci) whose most recent common ancestor is a duplication node ●

Orthologs: genes (loci) whose most recent common ancestor is a speciation node ●

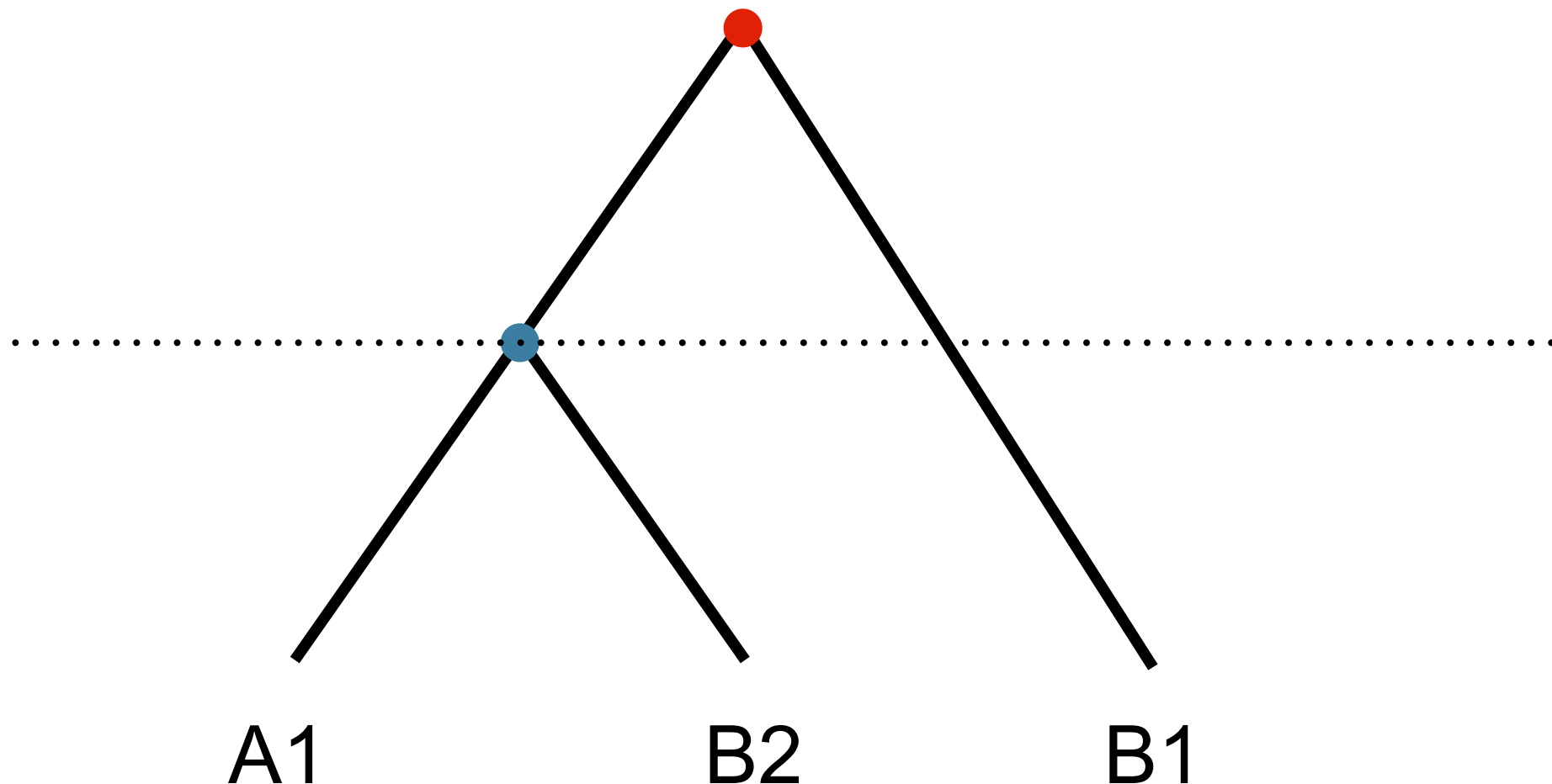
Genealogical relationships among genes



Genealogical relationships among genes

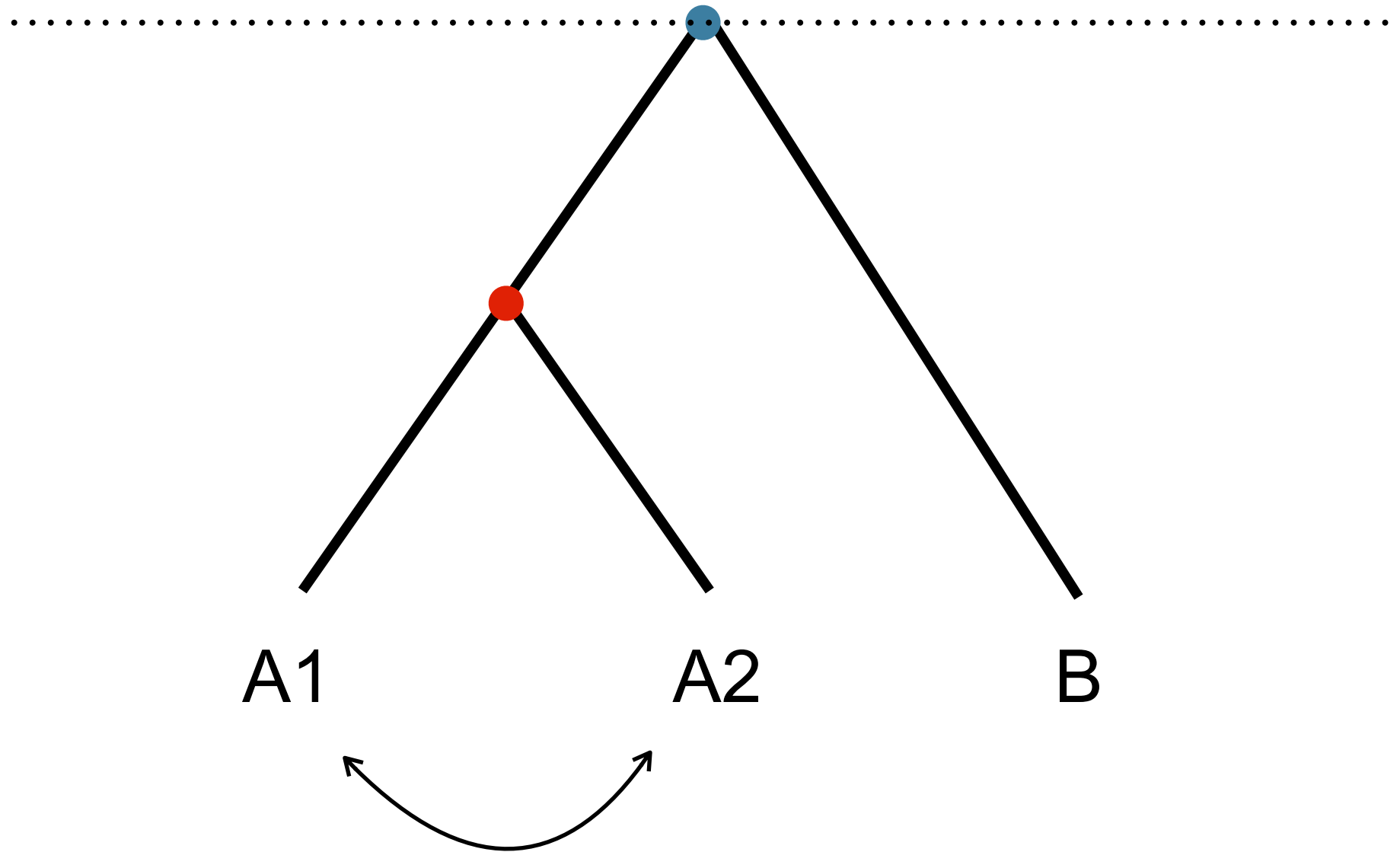


Genealogical relationships among genes



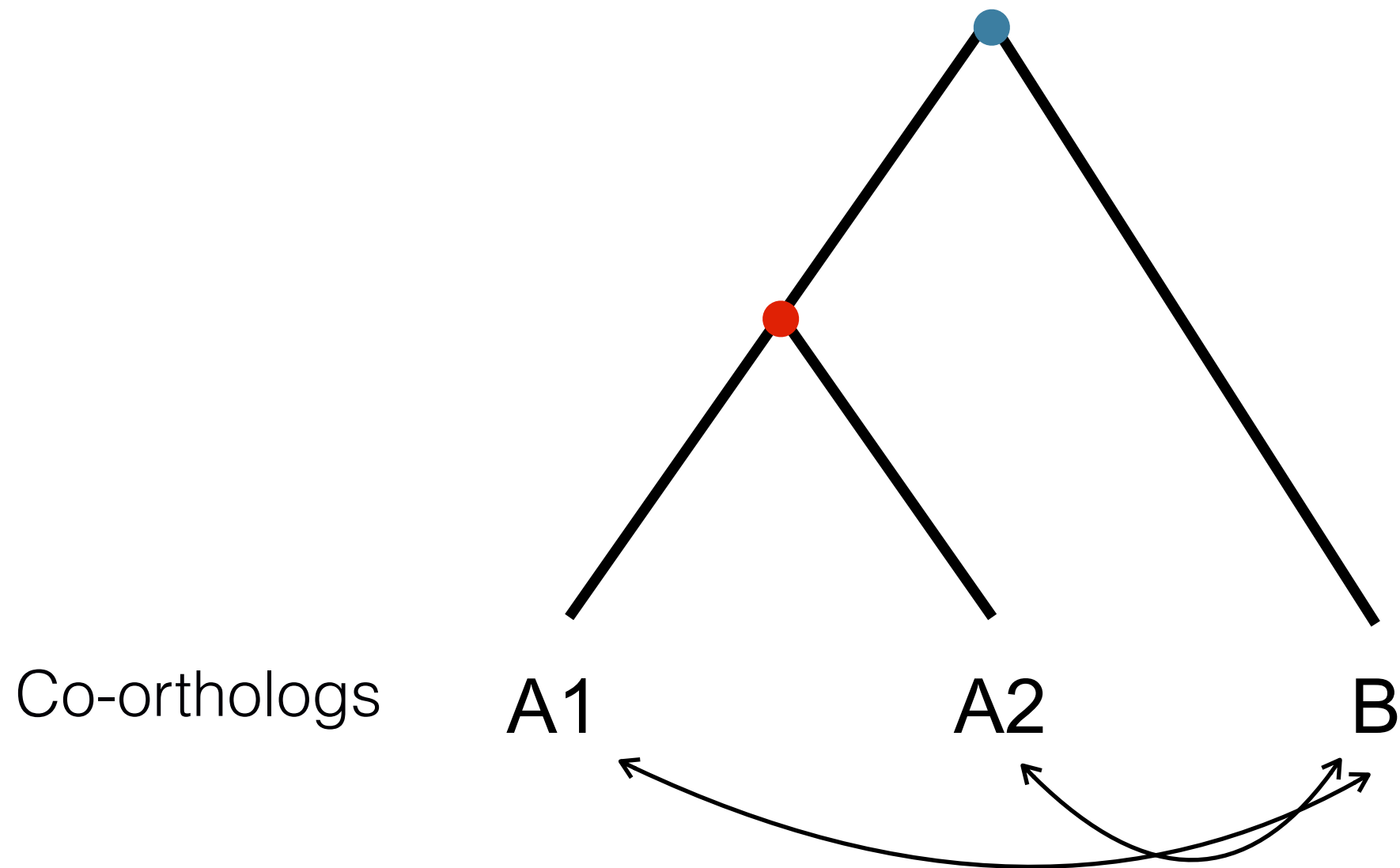
Paralogs
(out-paralogs wrt the A-B speciation event)

Genealogical relationships among genes

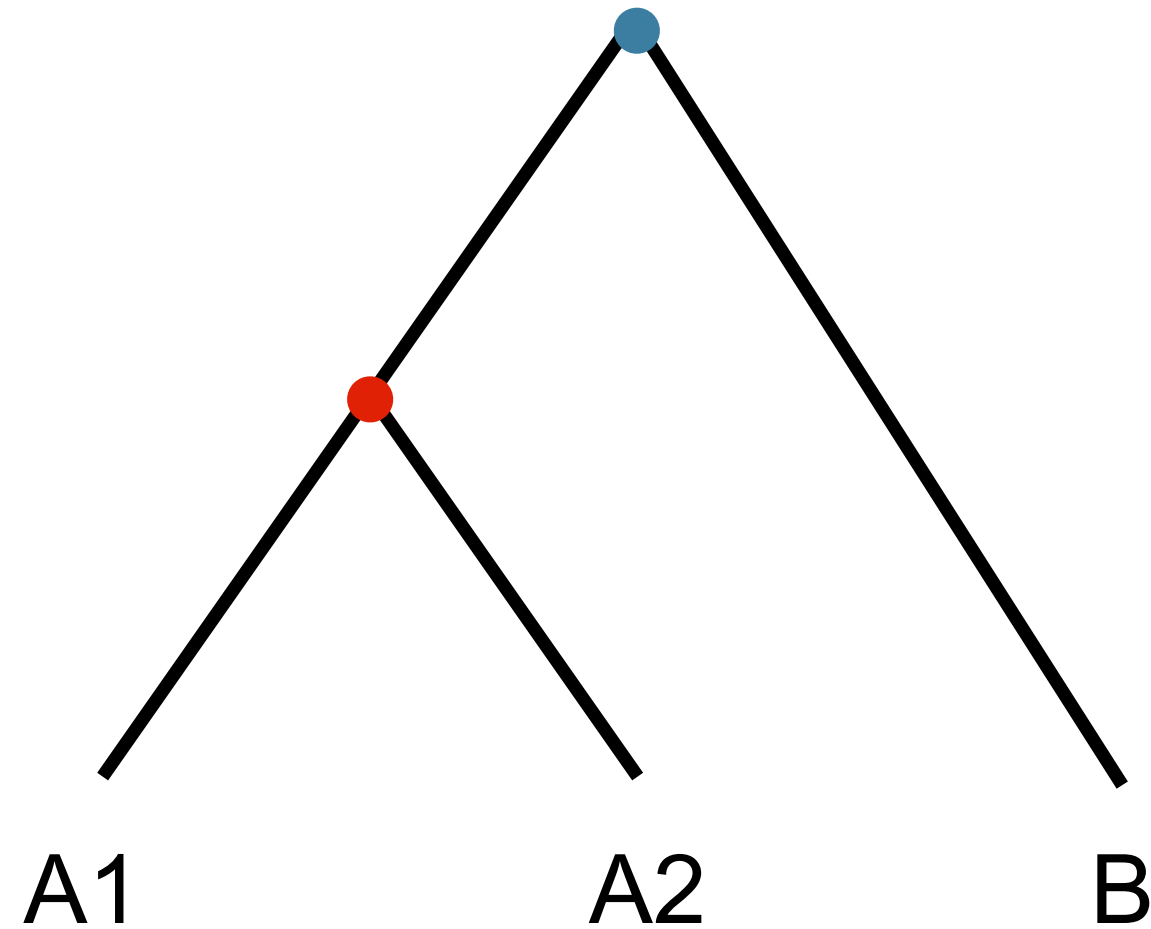


In-paralogs wrt the A-B speciation event

More genealogical relationships among genes



More genealogical relationships among genes



“Positional orthologs”
(Dewey 2011)



16

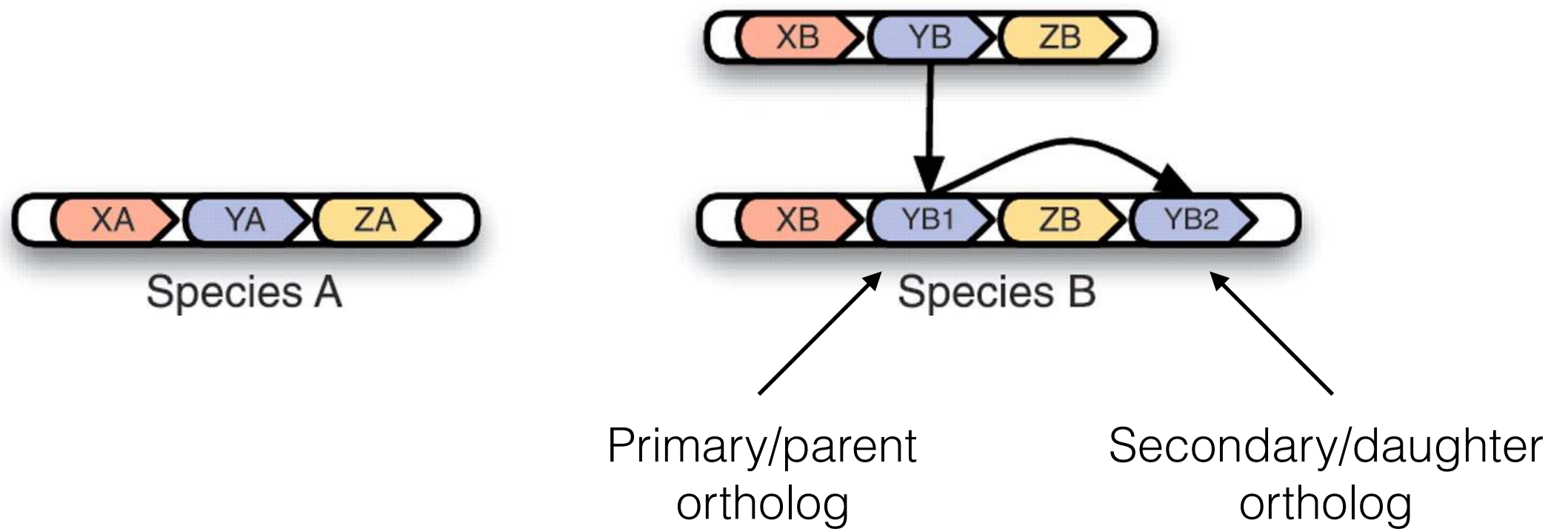


x



x

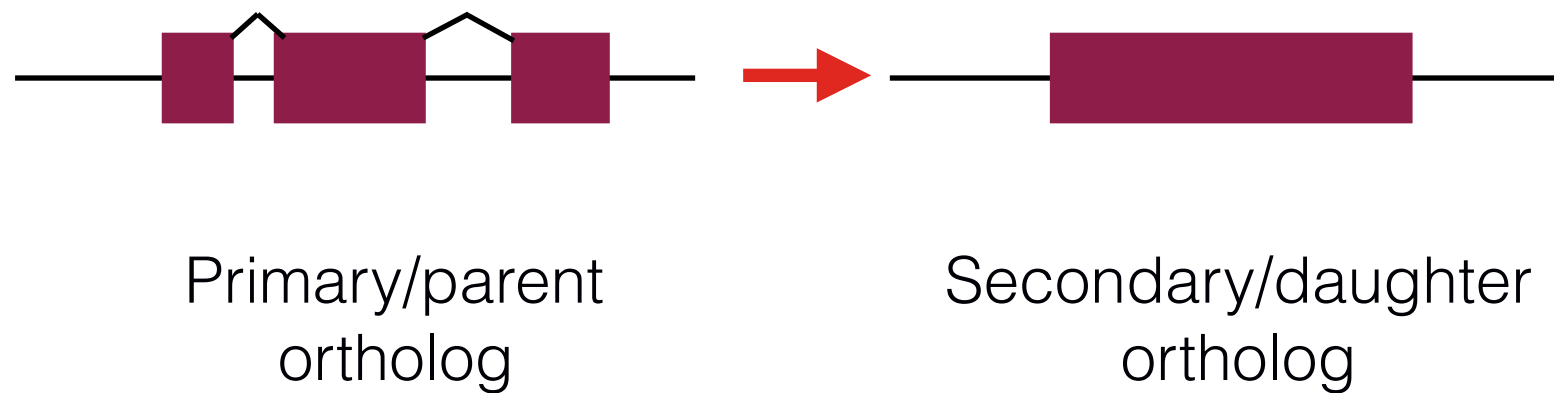
More genealogical relationships among genes



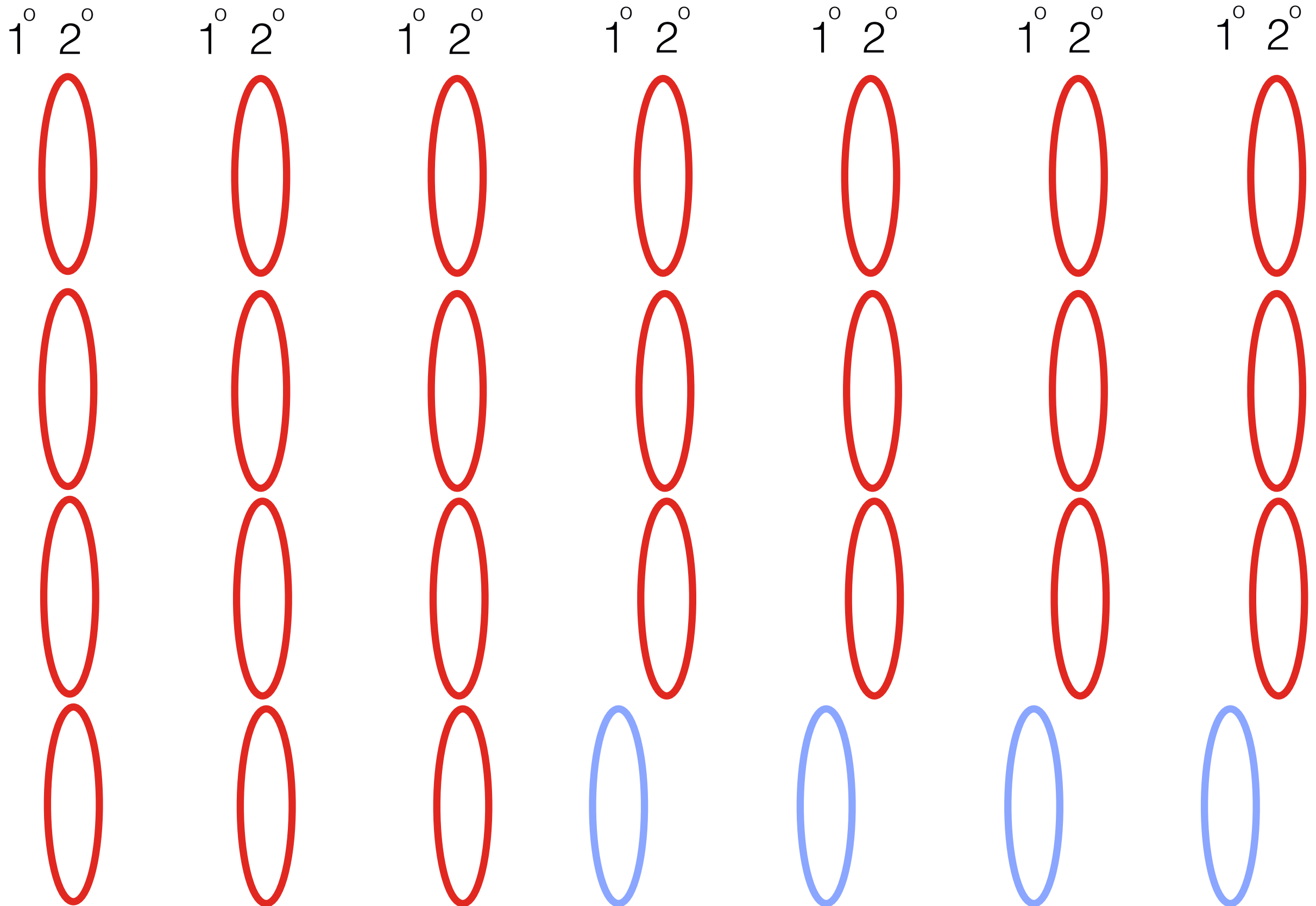
“Positional orthologs”
(Dewey 2011)

More genealogical relationships among genes

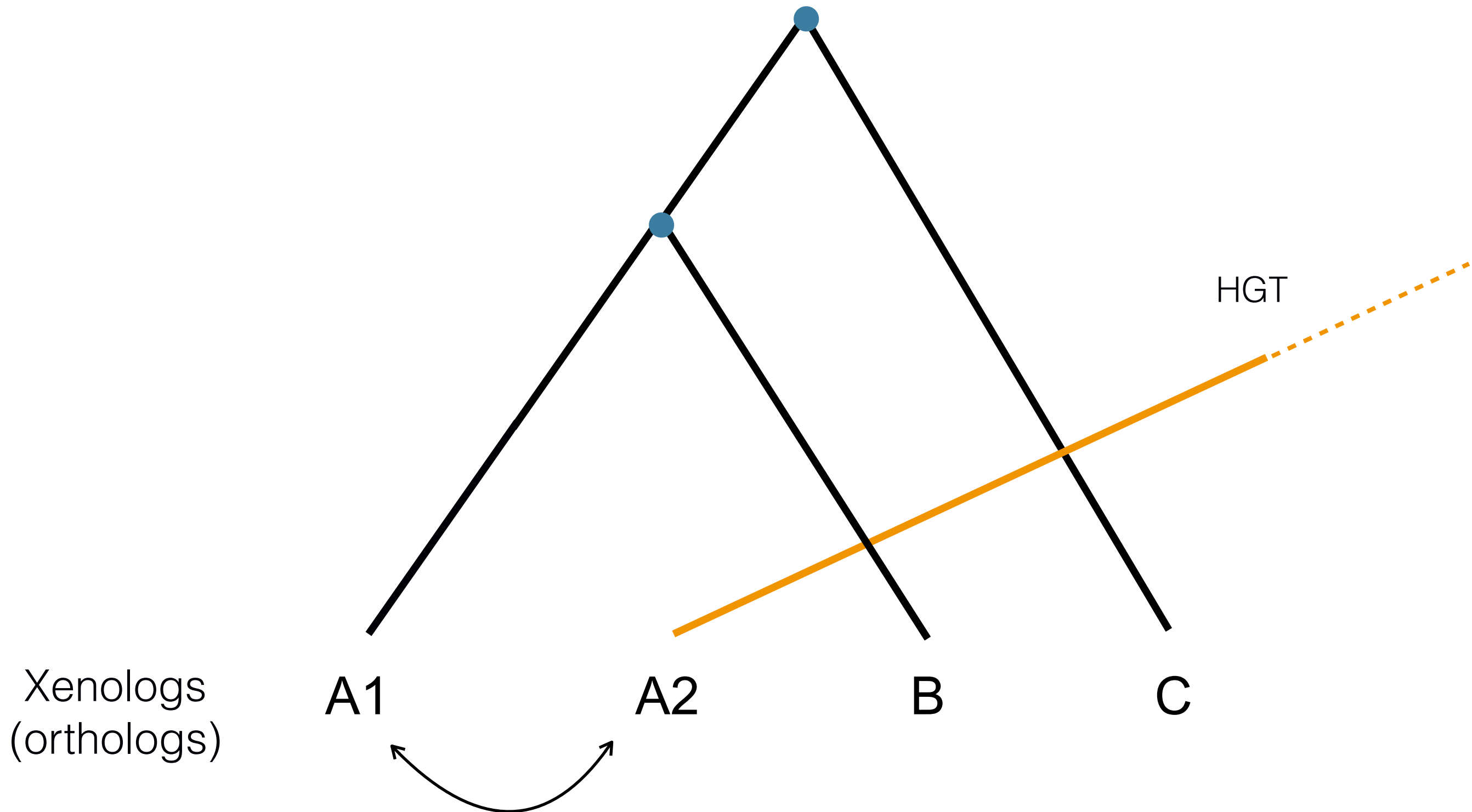
Retrotransposition



Polarized duplicates and adaptation



More genealogical relationships among genes



More genealogical relationships among genes

 OPEN ACCESS  PEER-REVIEWED

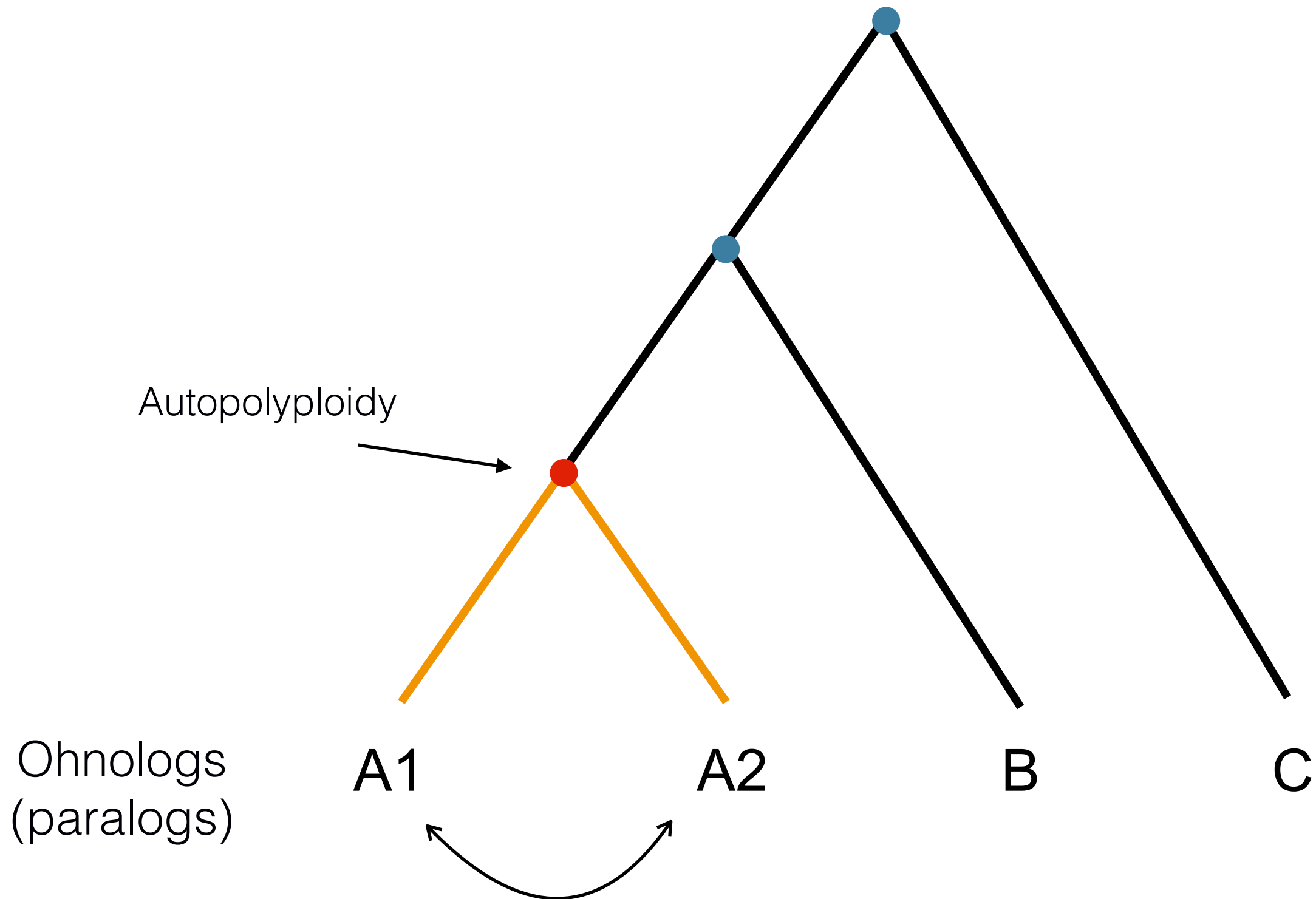
RESEARCH ARTICLE

Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes

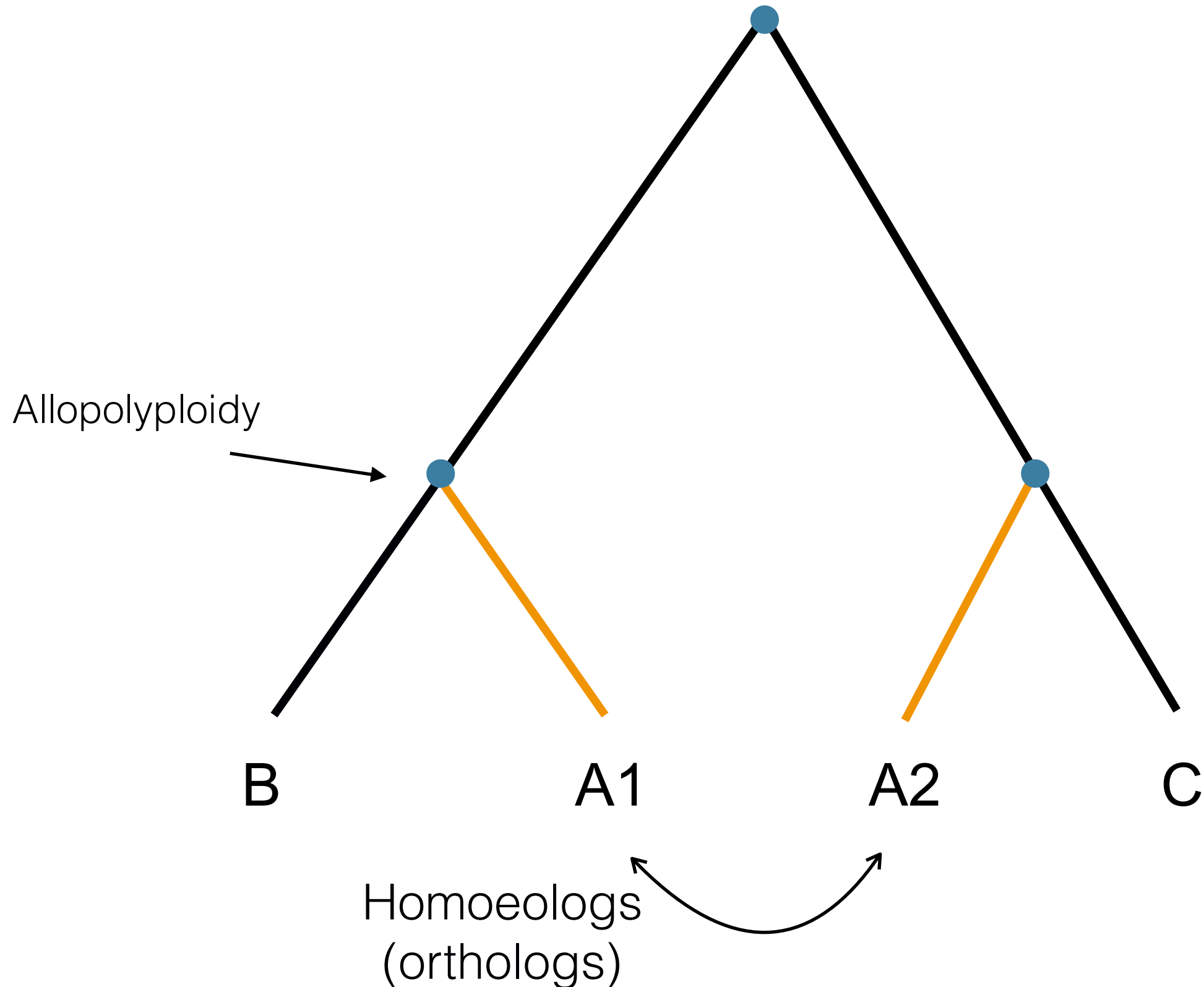
Todd J. Treangen , Eduardo P. C. Rocha

Published: January 27, 2011 • <http://dx.doi.org/10.1371/journal.pgen.1001284>

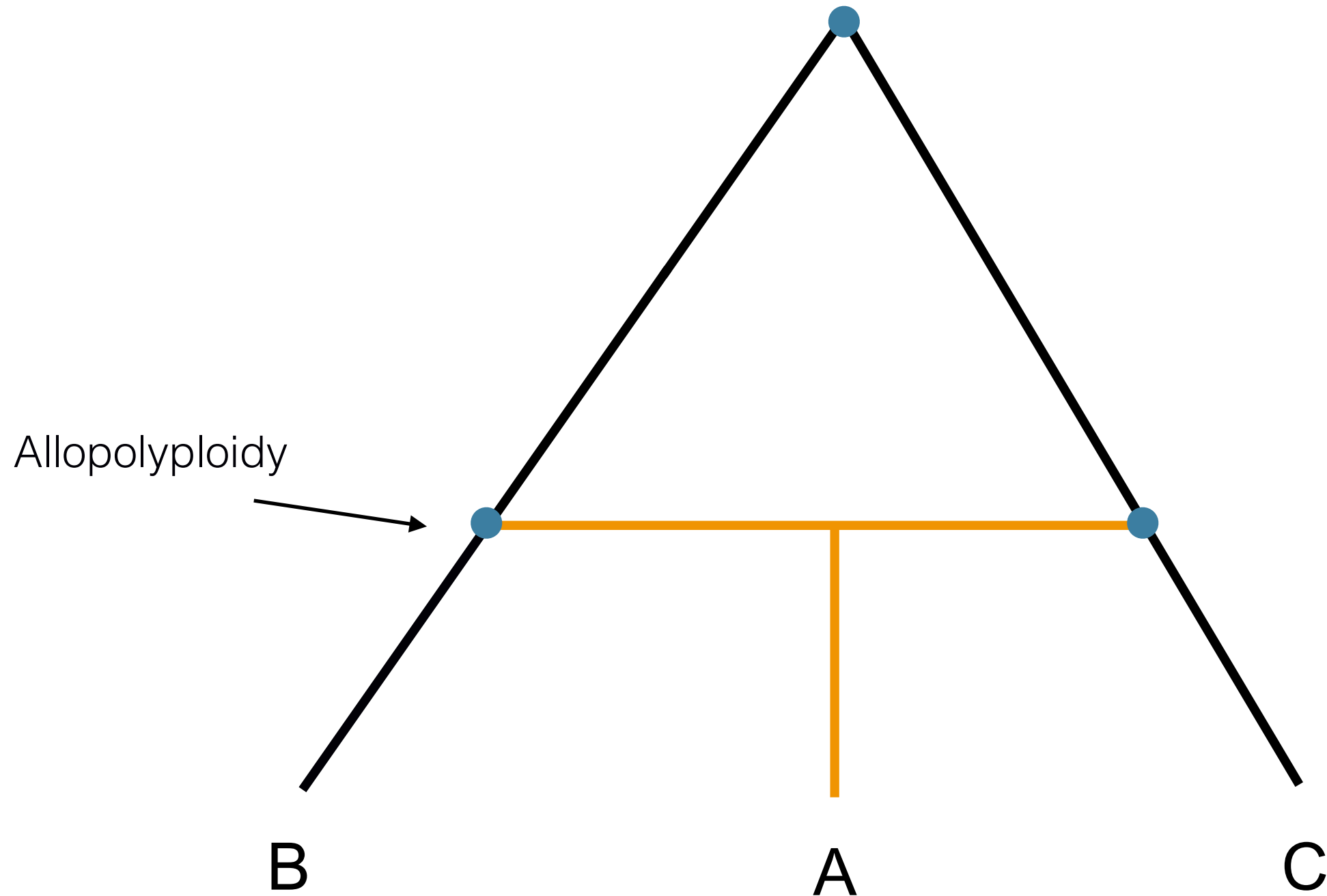
More genealogical relationships among genes



More genealogical relationships among genes



More genealogical relationships among genes



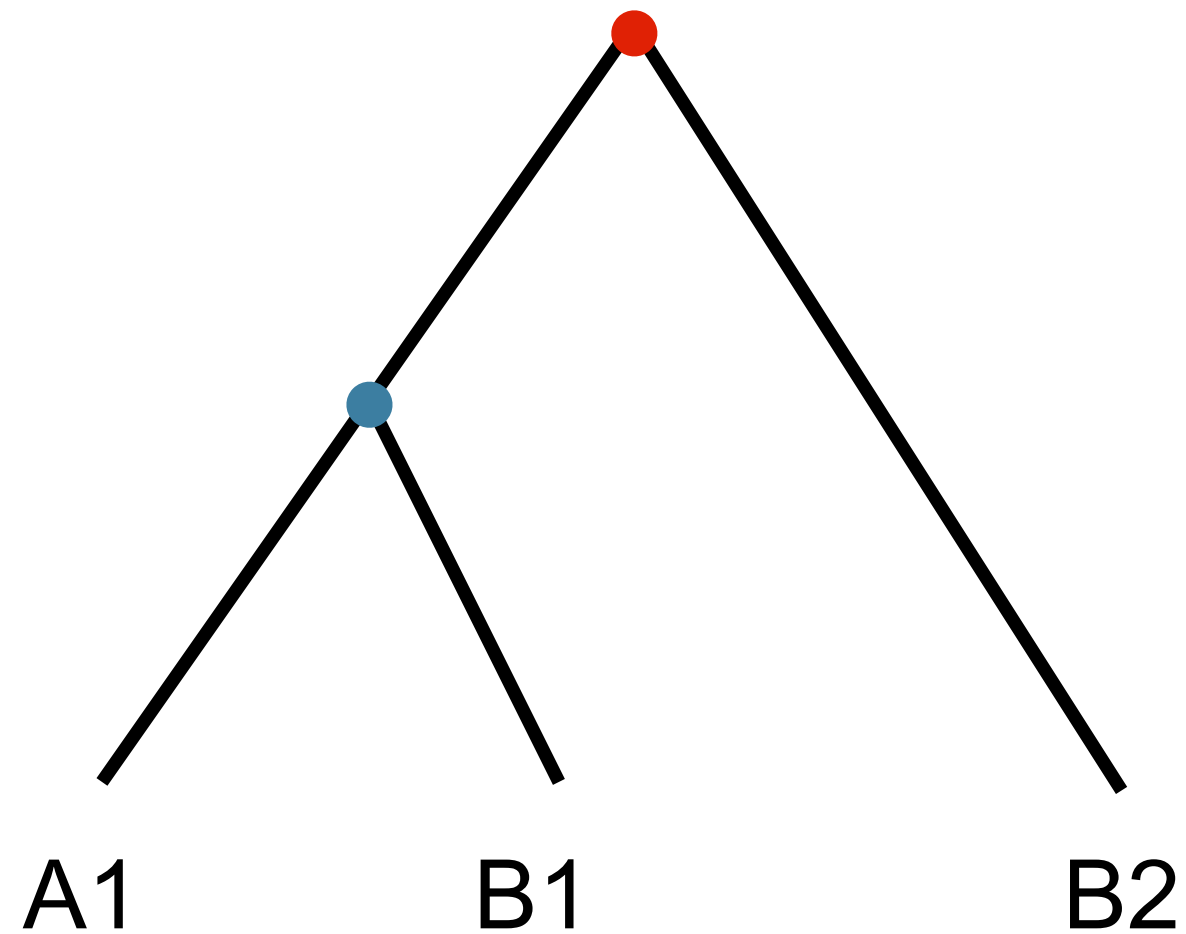
More genealogical relationships among genes

	Pairs of genes found in the same species	Pairs of genes found in different species
Genes that originated by a speciation event	Homoeologs	Orthologs
Genes that originated by a duplication event	Whole genome duplication: Ohnologs	Paralogs
	Small scale duplication: Paralogs	

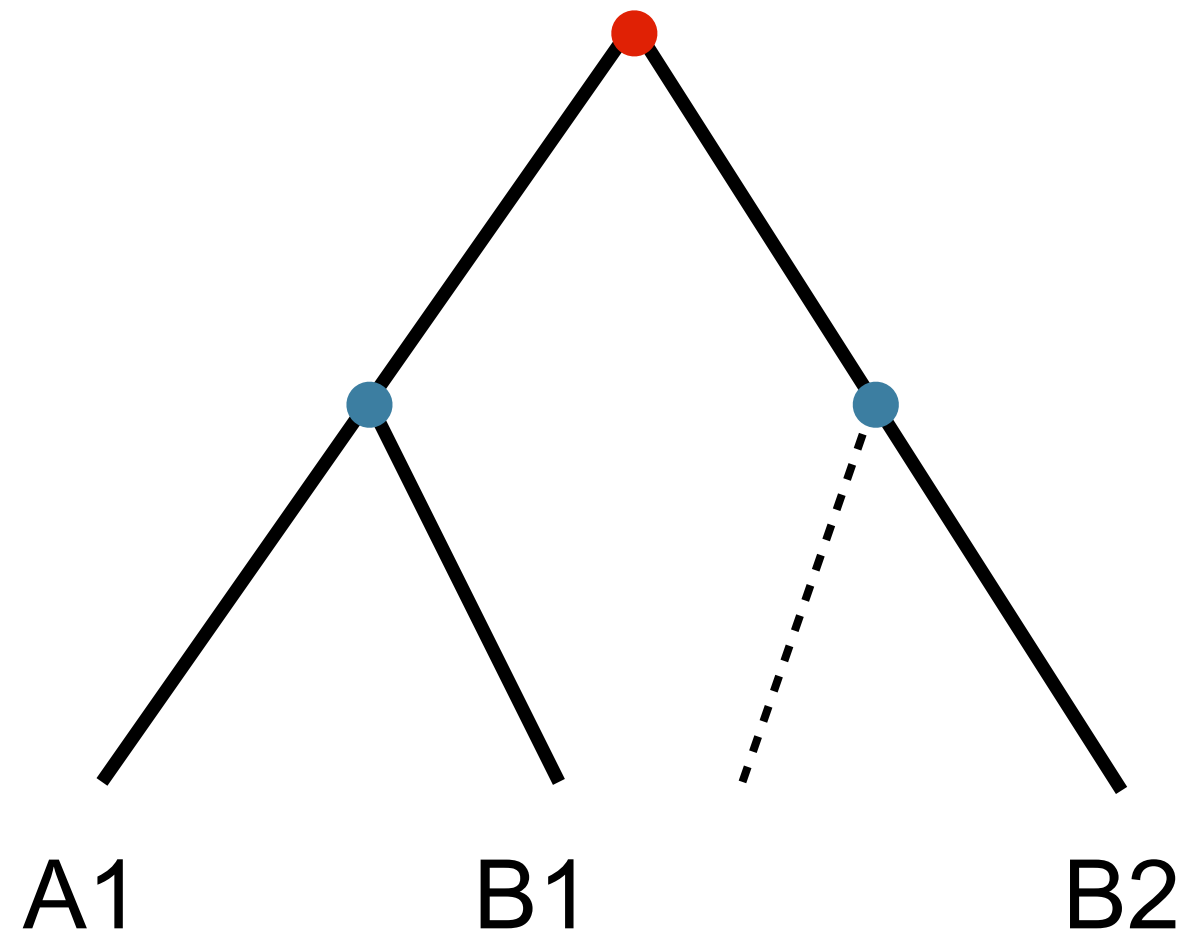
Trends in Plant Science

Glover et al. (2016)

More genealogical relationships among genes

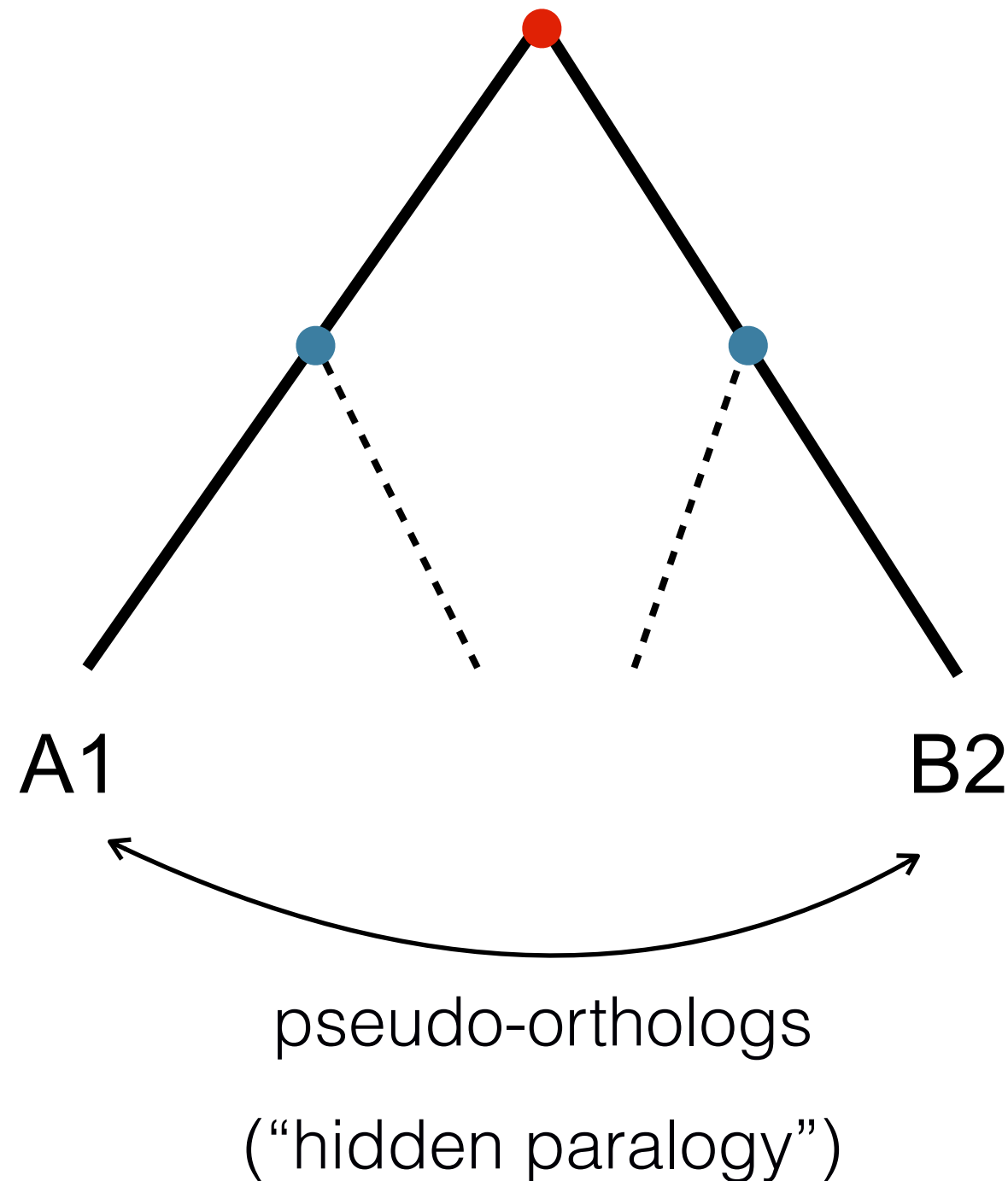


More genealogical relationships among genes

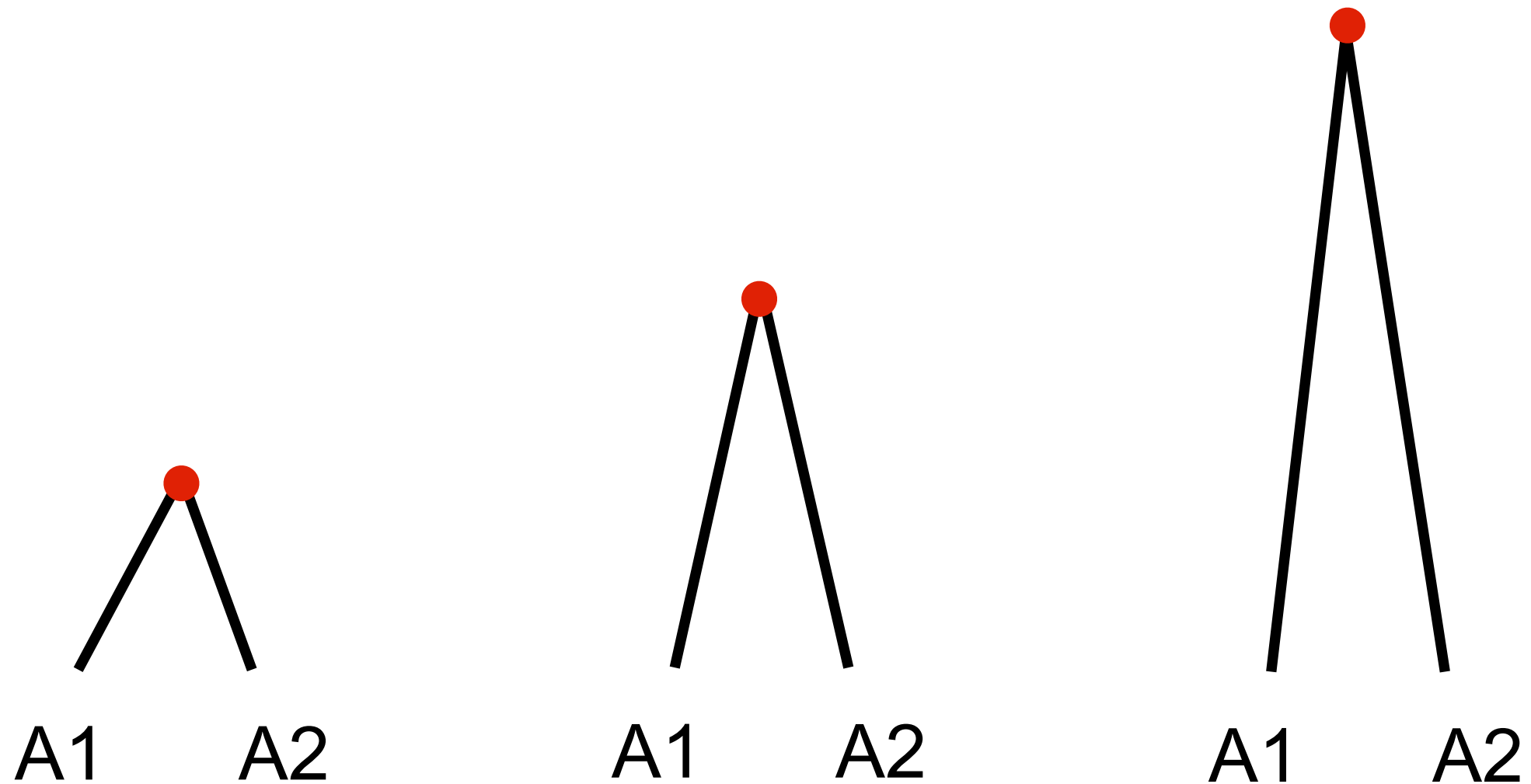


“Reconciled” gene tree

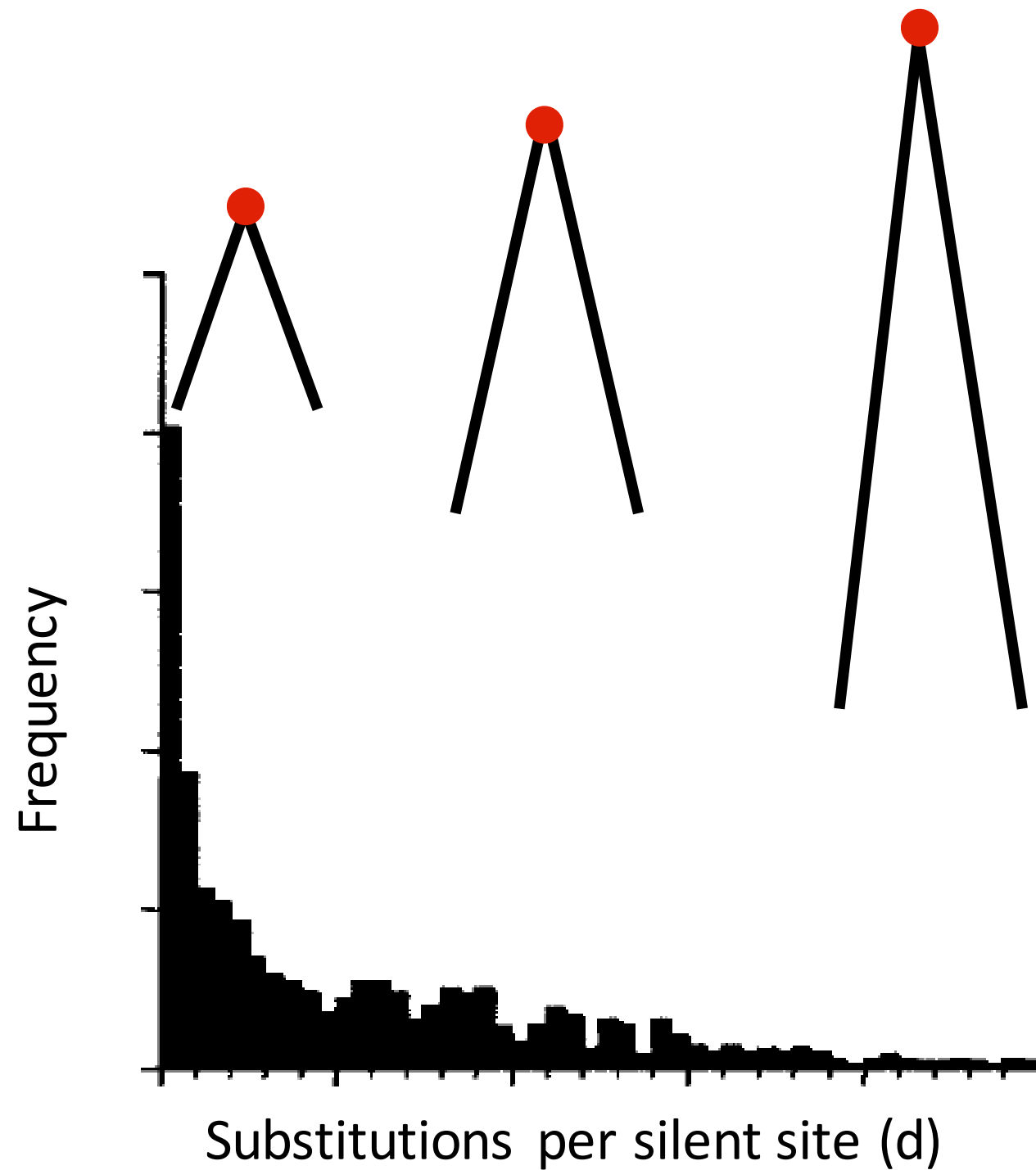
More genealogical relationships among genes



Divergence between paralogs



Divergence between paralogs



Divergence between paralogs

How divergent are new paralogs?

