

**Using molecules as data.
Types of homology.
Primary and secondary structure.
The ITS region**



The concept of molecular clock



The idea of using macromolecules as source of data for phylogenetic reconstruction was developed during the 1960's

Organic macromolecules are optimal for this purpose because:

Evolutionary Divergence and Convergence in Proteins

EMILE ZUCKERKANDL

*Laboratoire de Physico-Chimie Colloidale du C. N. R. S.,
Montpellier, France*

AND

LINUS PAULING

*California Institute of Technology,
Pasadena, California*

- Despite performing complex 3-dimensional functions, they can all be reproduced in terms of a sequence (primary structure)
- They are rich in information
- Their changes have a phylogenetic meaning

The concept of molecular clock

Specimen 1

AGGTCAGATTACA

Specimen 2

ATGTCACATTTGA

The concept of molecular clock

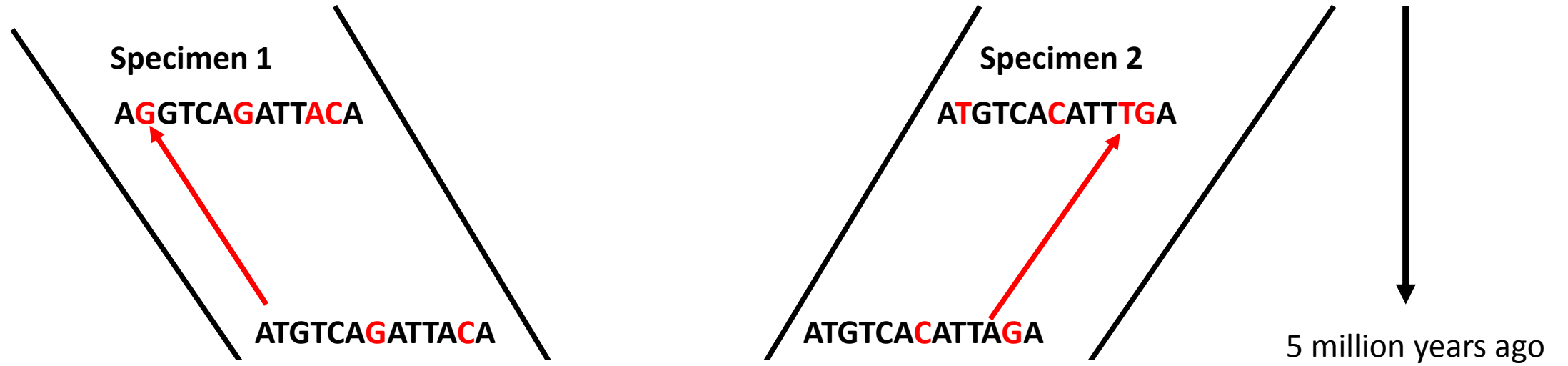
Specimen 1

AGGTCAGATTACA

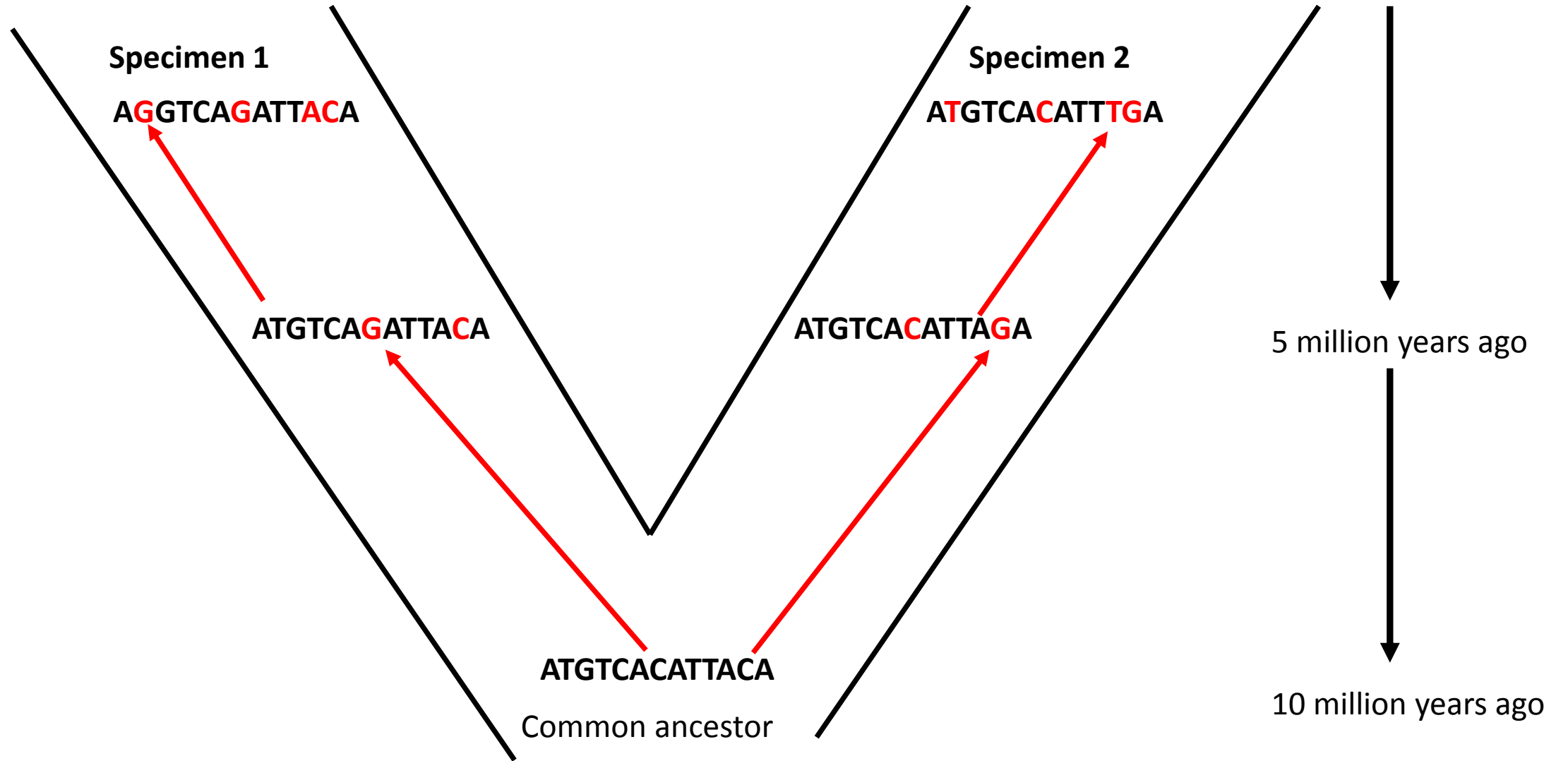
Specimen 2

ATGTCACATTGA

The concept of molecular clock



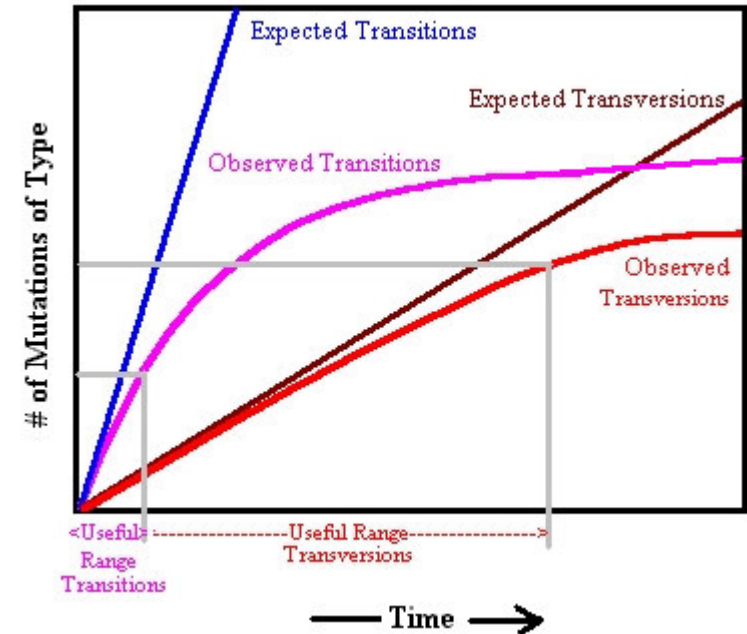
The concept of molecular clock



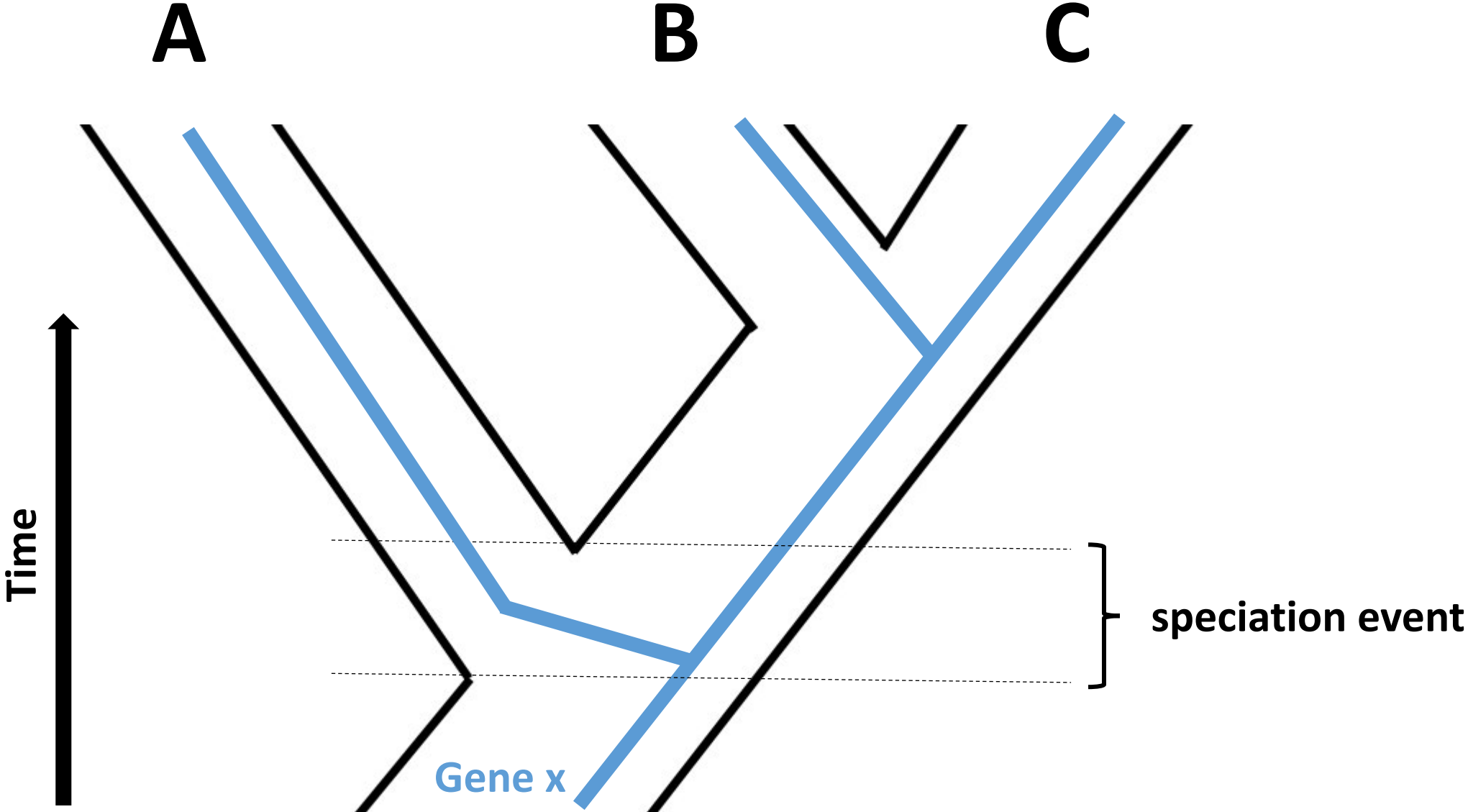
The concept of molecular clock

The basic idea of the molecular clock is central to modern phylogenetic inference, but remember:

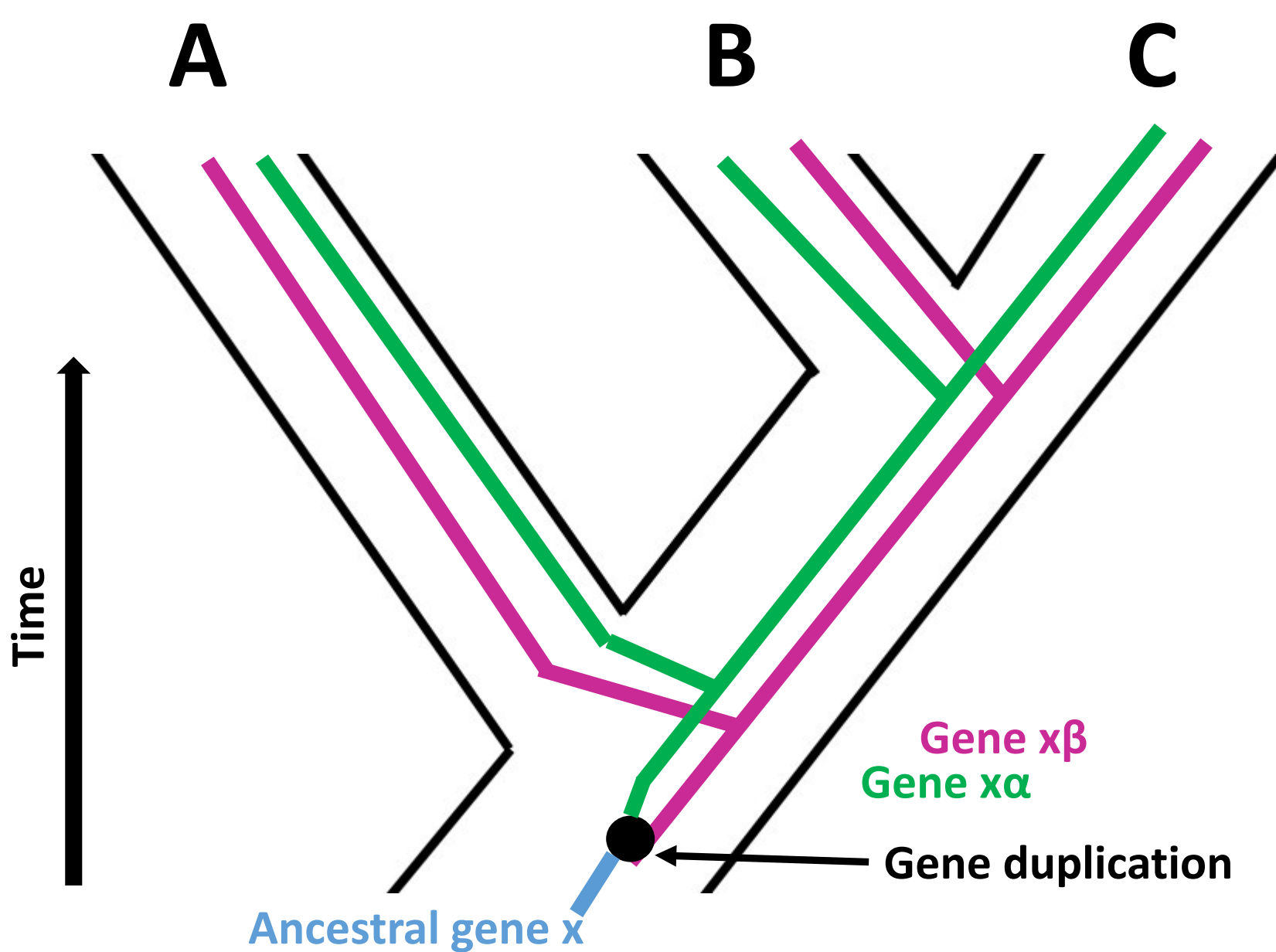
- The rate of change of a sequence may be very variable depending on the kind of molecule, the particular region or the phylogenetic lineage
- Other phenomena such as long branch attraction (due to saturation of very variable sites) have to be taken into account
- The alignment should be based in homology relationships (in particular, **orthology**)



Homology, orthology and paralogy

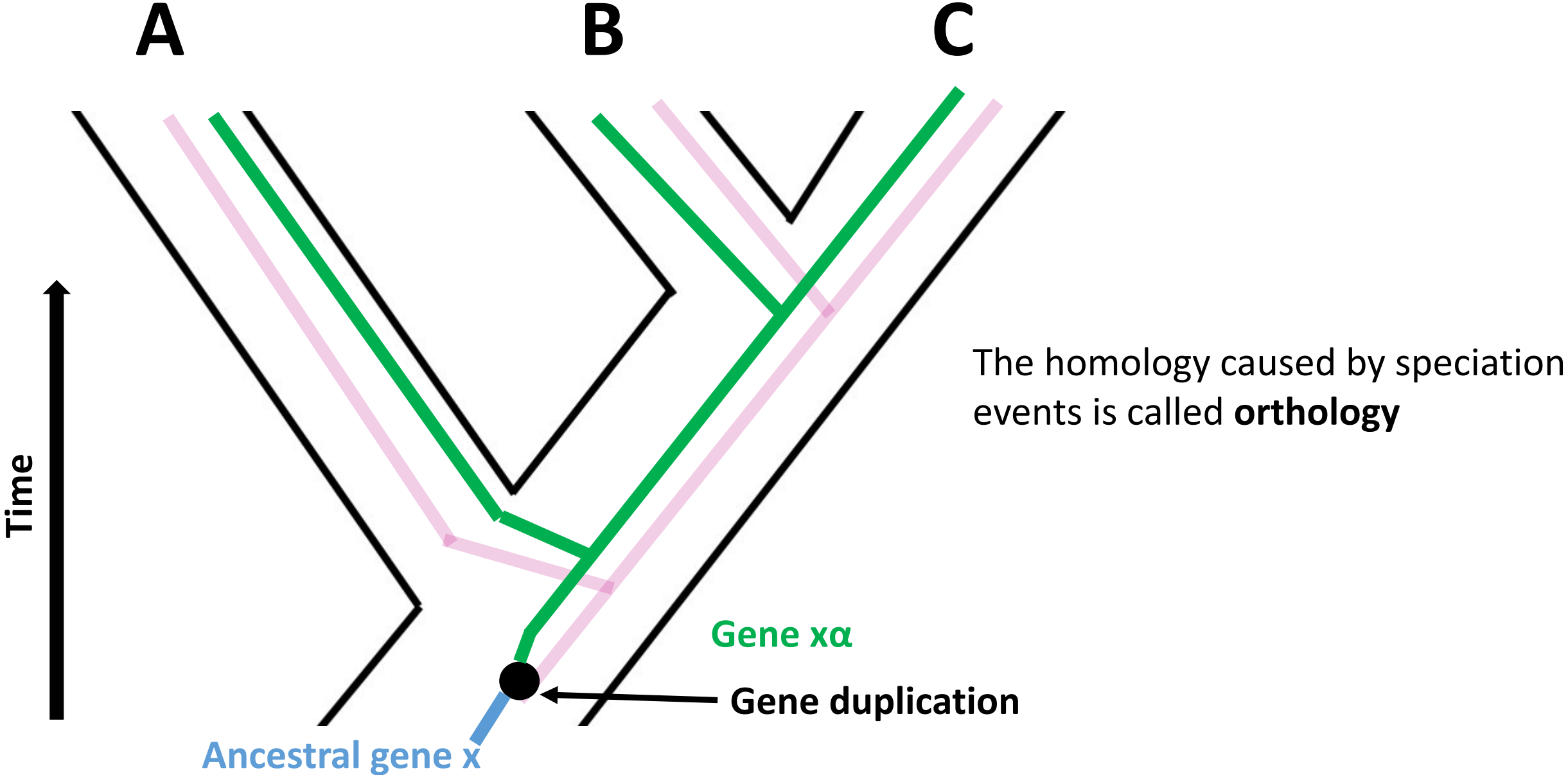


Homology, orthology and paralogy

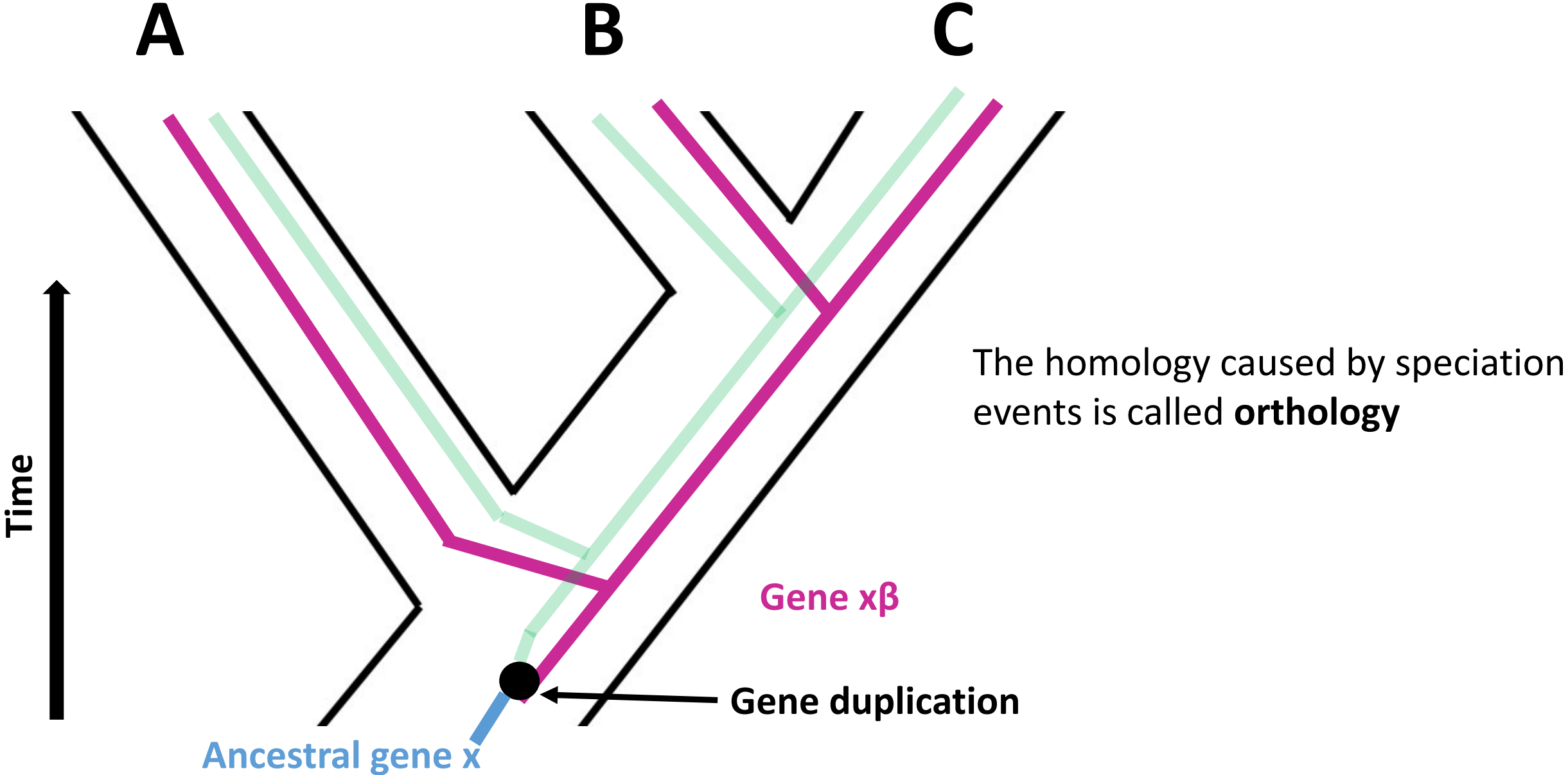


All the resulting sequences of the gen x are **homologous** because they derived from a common ancestor

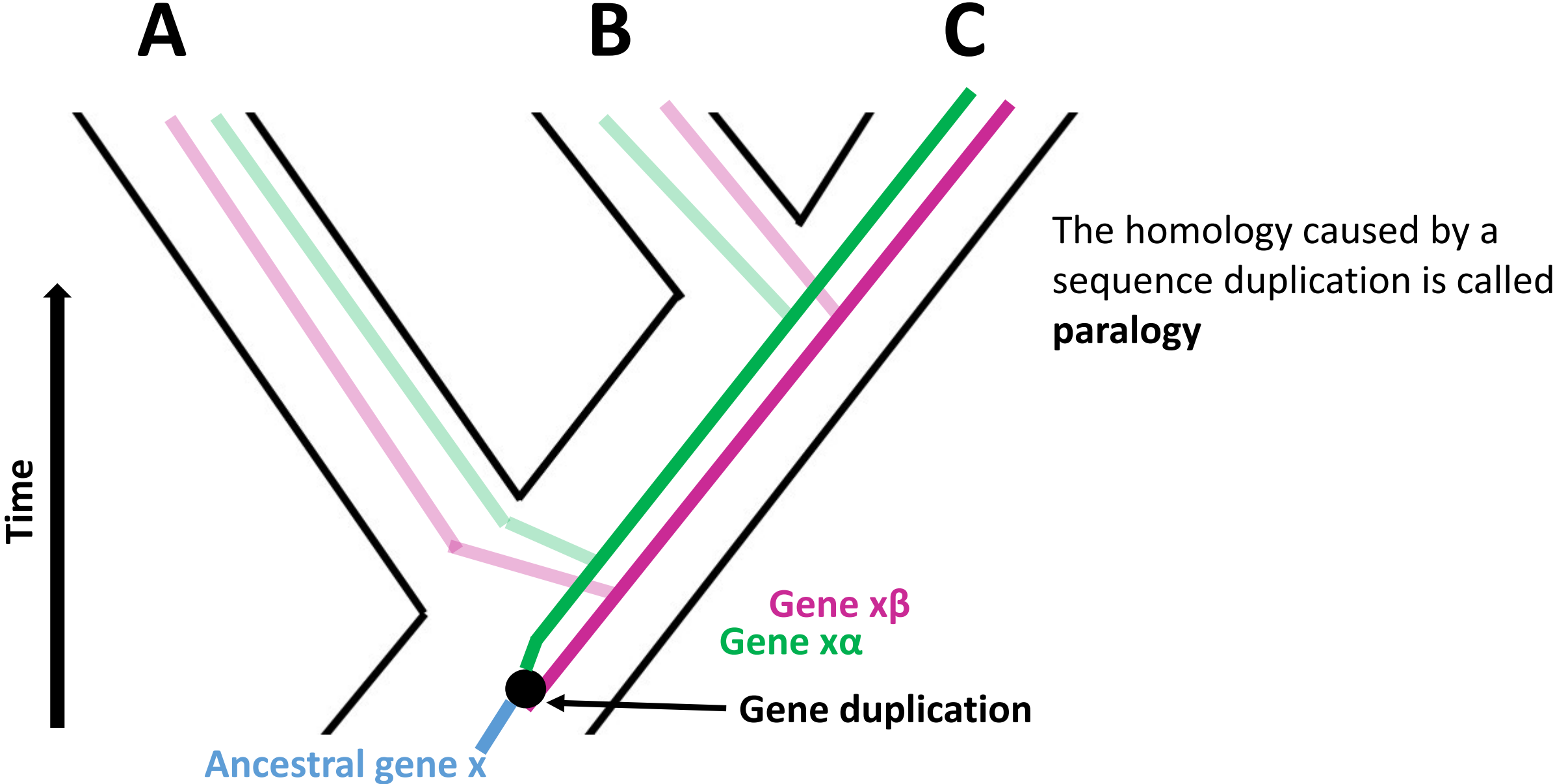
Homology, orthology and paralogy



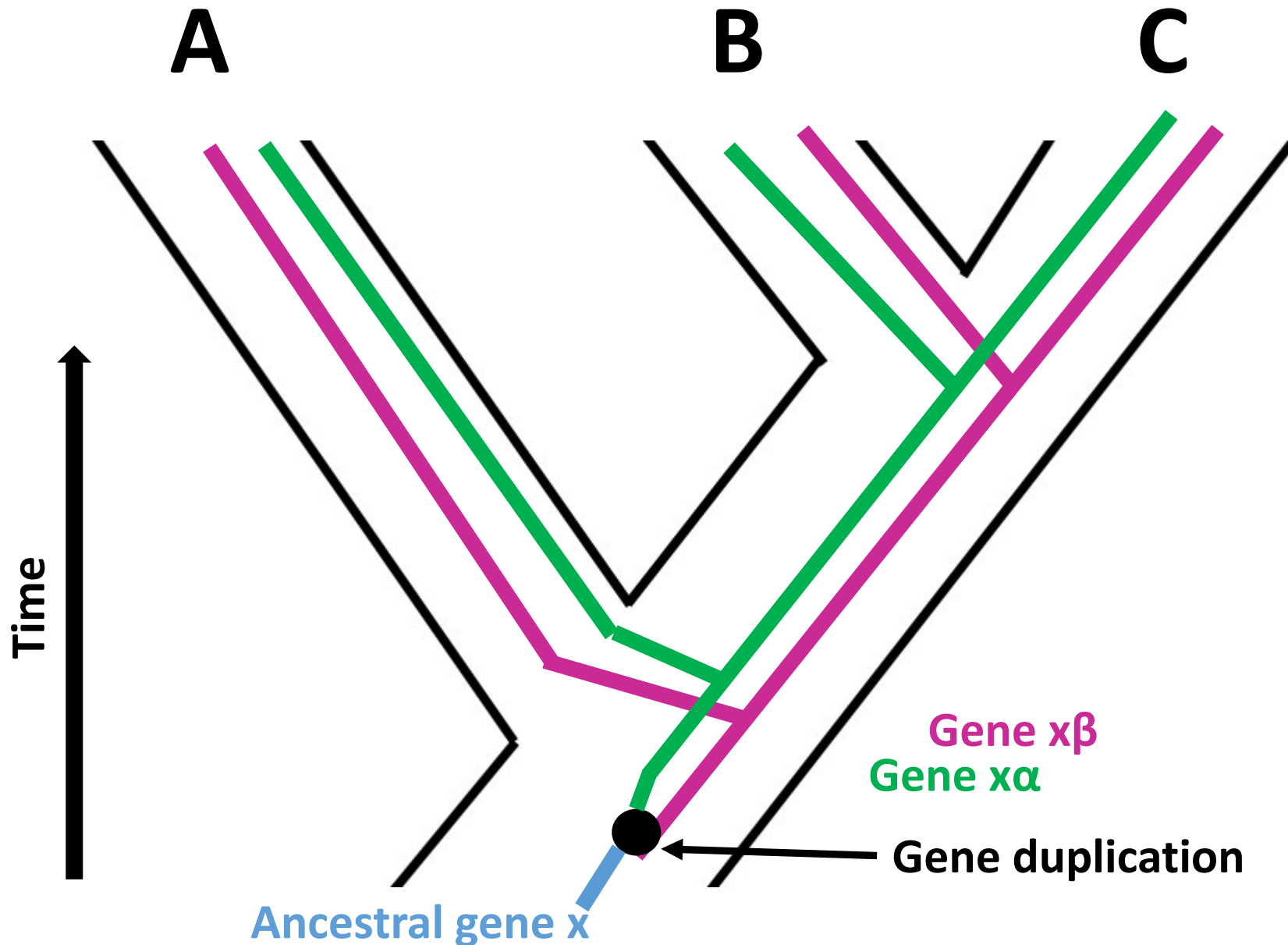
Homology, orthology and paralogy



Homology, orthology and paralogy

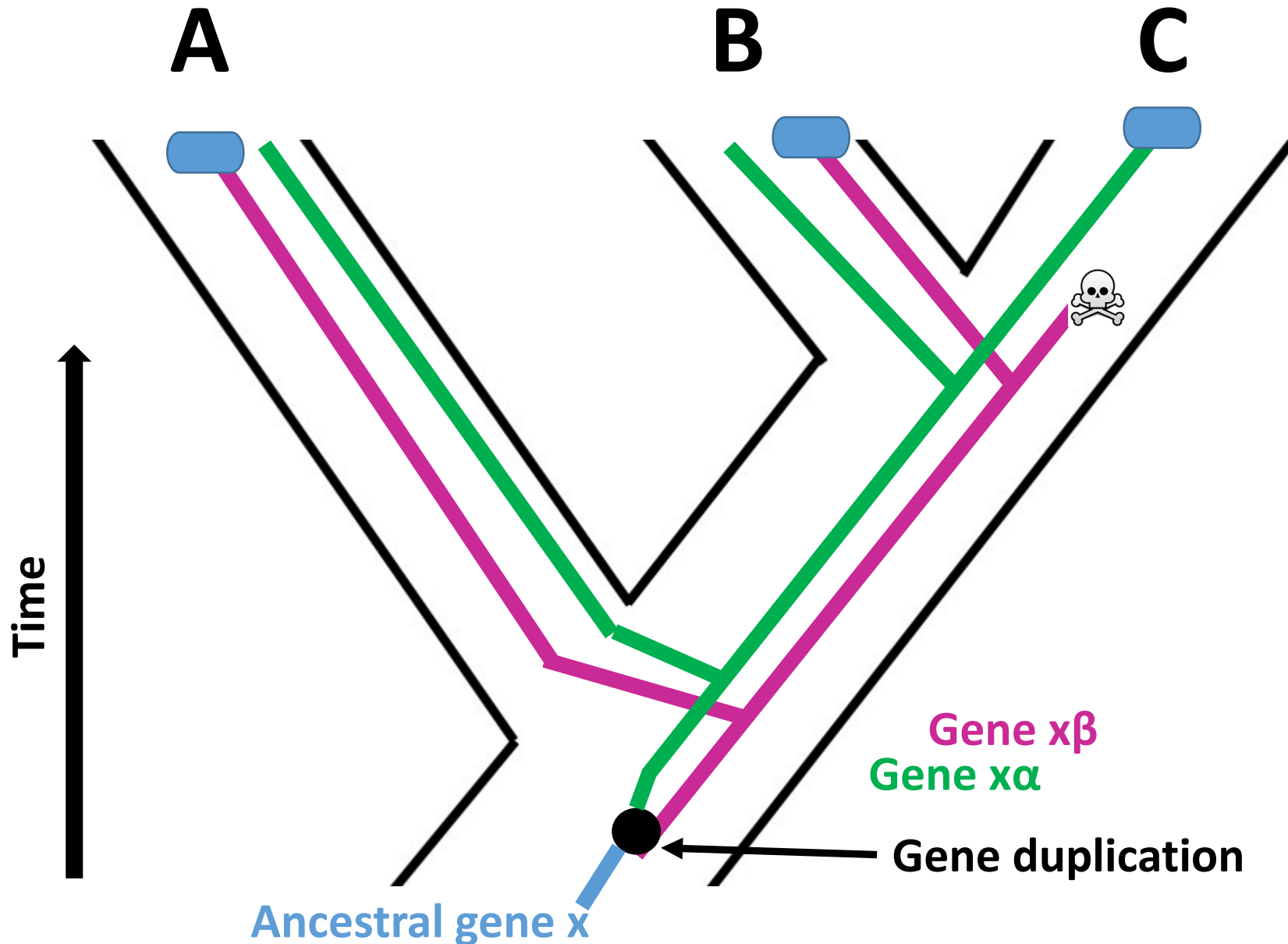


Homology, orthology and paralogy

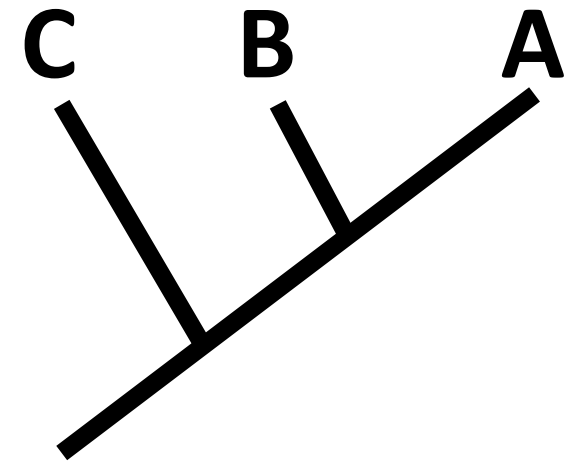


If our objective is phylogenetic reconstruction we should only align orthologous sequences, otherwise our result would be misleading

Homology, orthology and paralogy

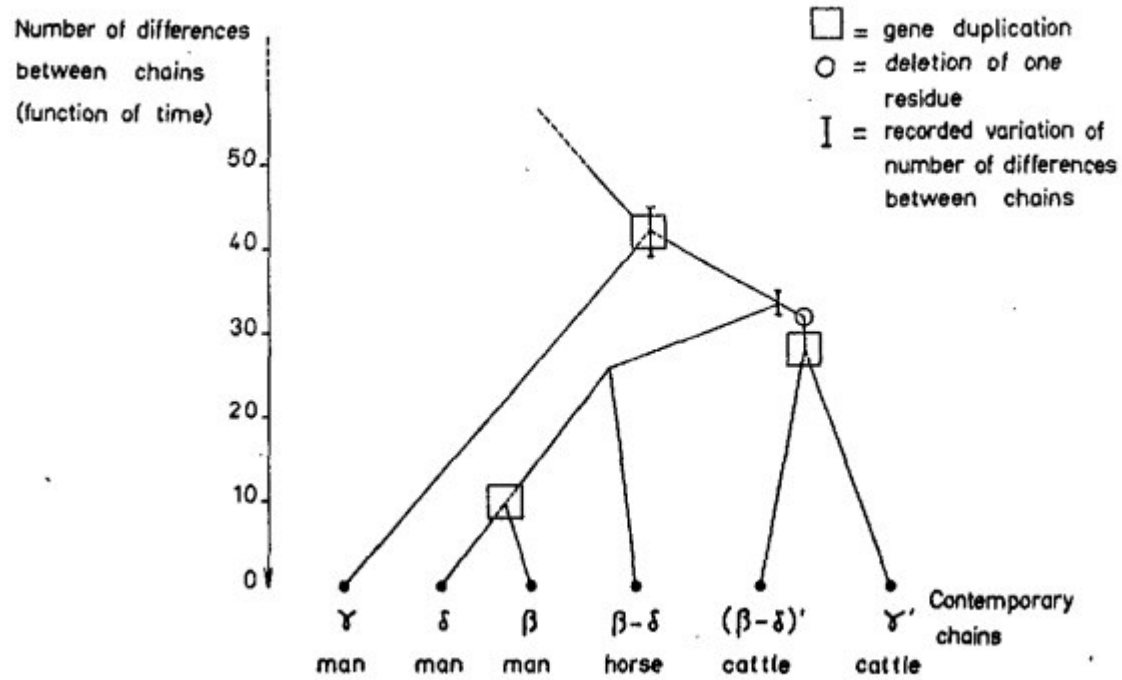


If our objective is phylogenetic reconstruction we should only align orthologous sequences, otherwise our result would be misleading

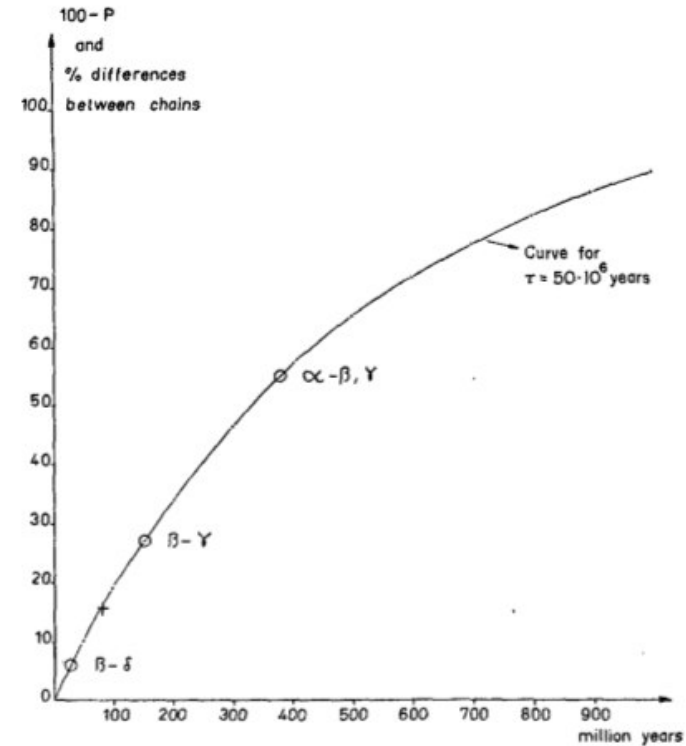


Homology, orthology and paralogy

Both orthology and paralogy were present in the study of Zuckerkandl and Pauling

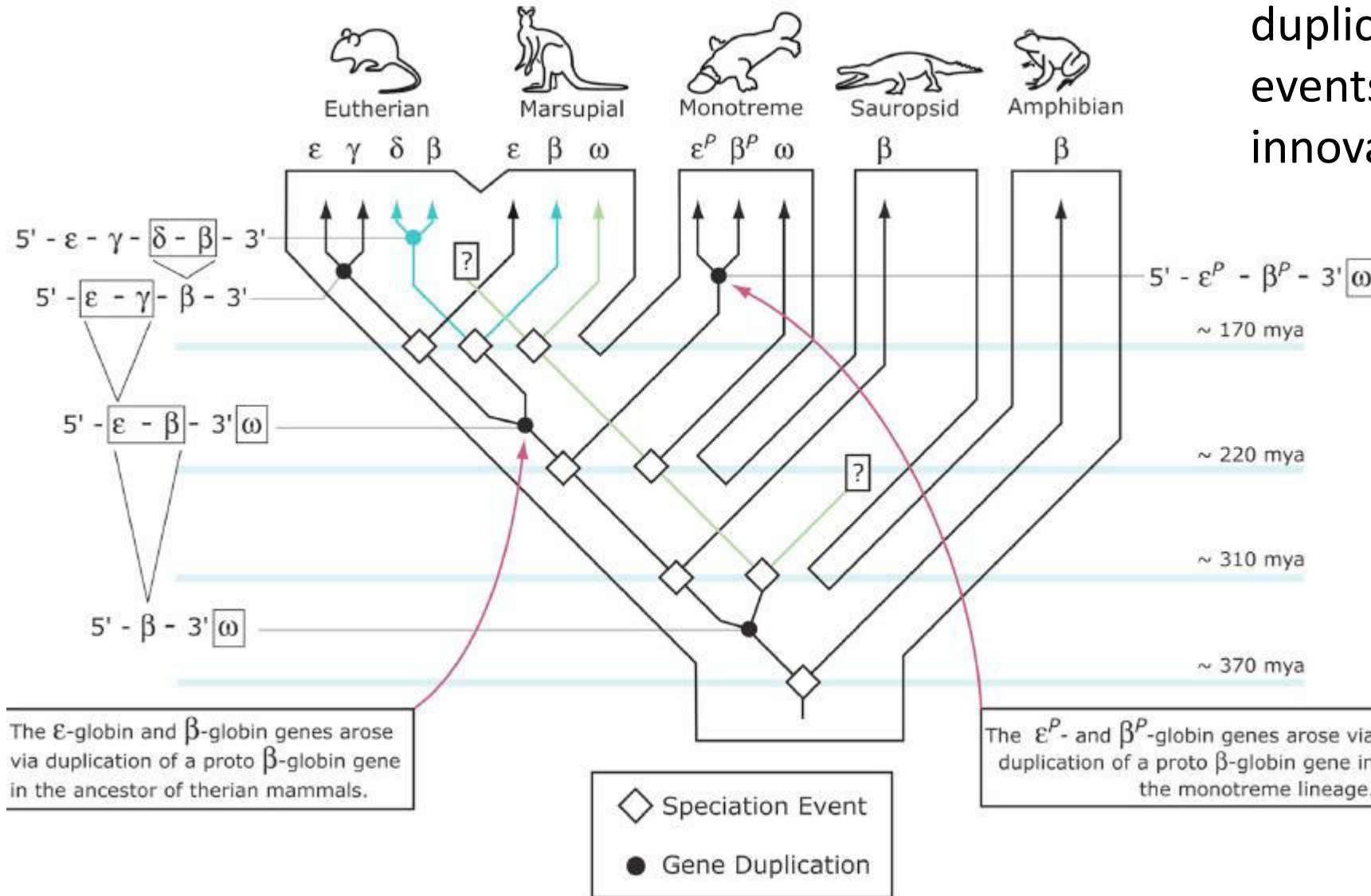


4. Probable evolutionary relationship of some mammalian hemoglobin chains.



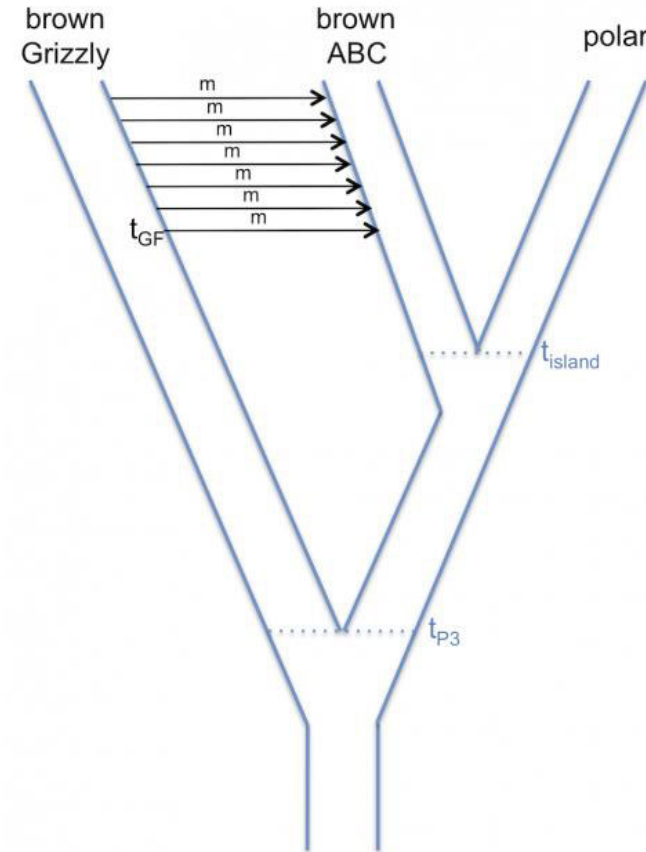
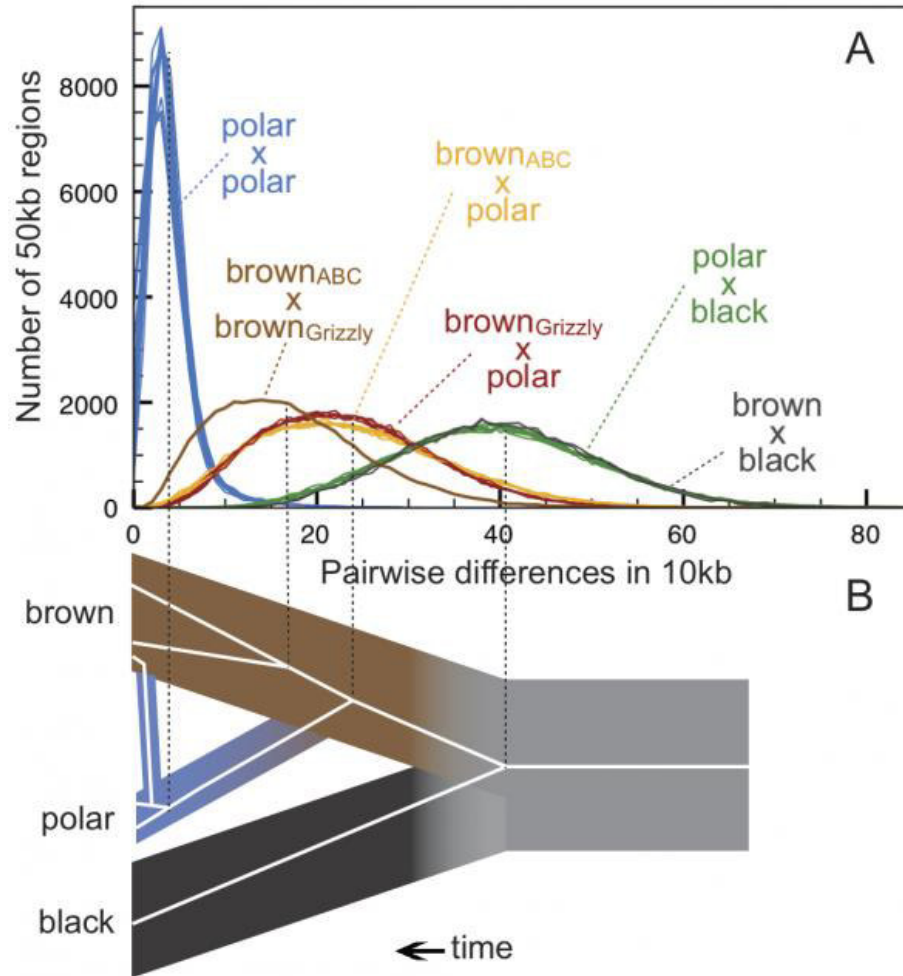
Homology, orthology and paralogy

Gene (and genome) duplications have been key events for evolutionary innovation



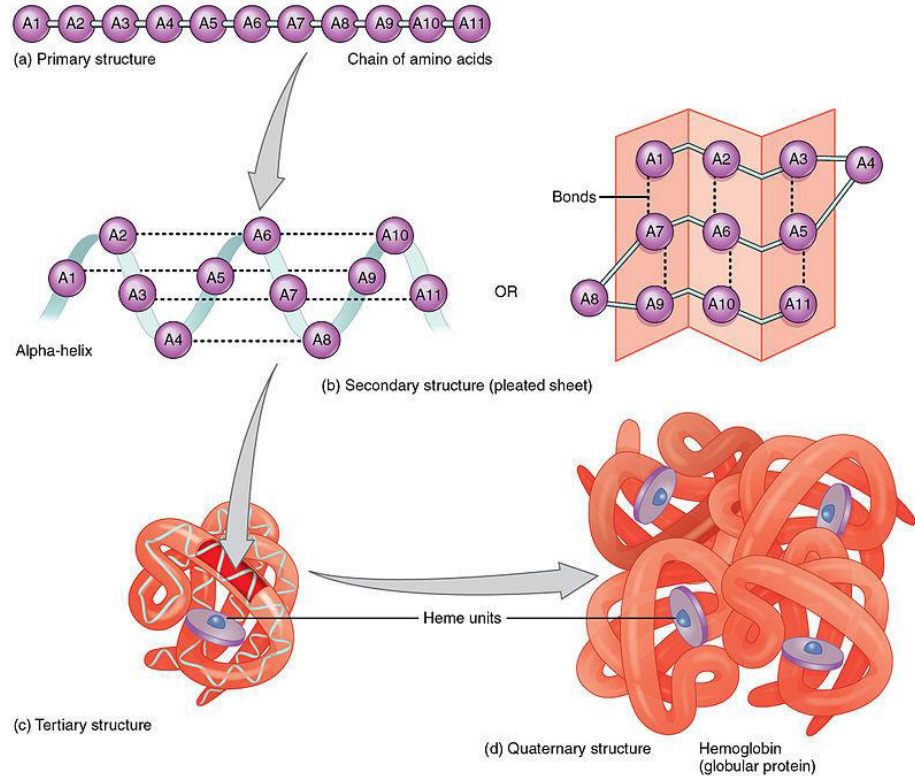
Non tree-like evolution

Example: Introgression of brown bear genes in the polar bears of the ABC islands

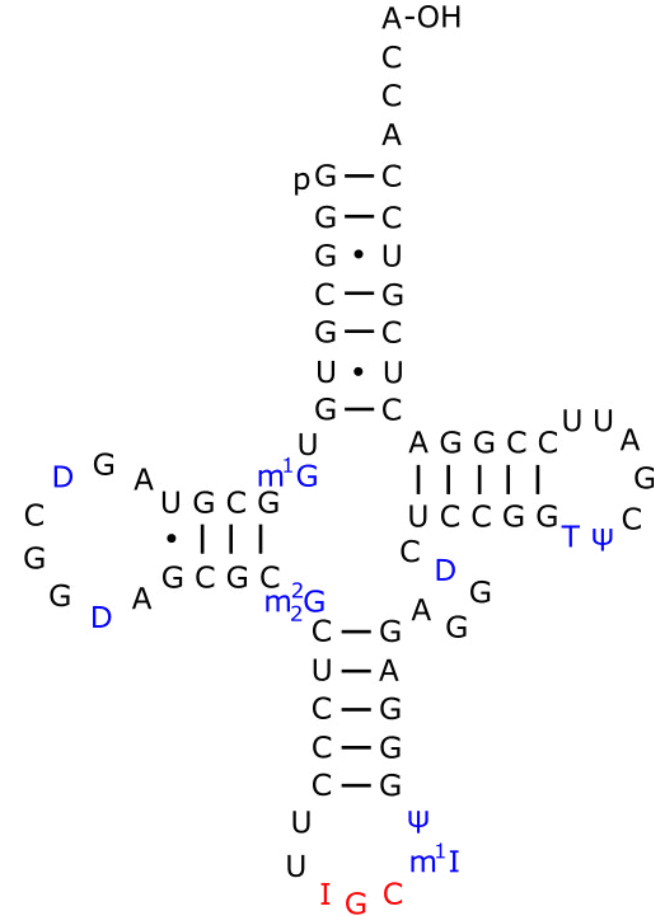


Primary and secondary structure

The function of macromolecules is explained in terms of their structure (particularly for proteins and RNA)

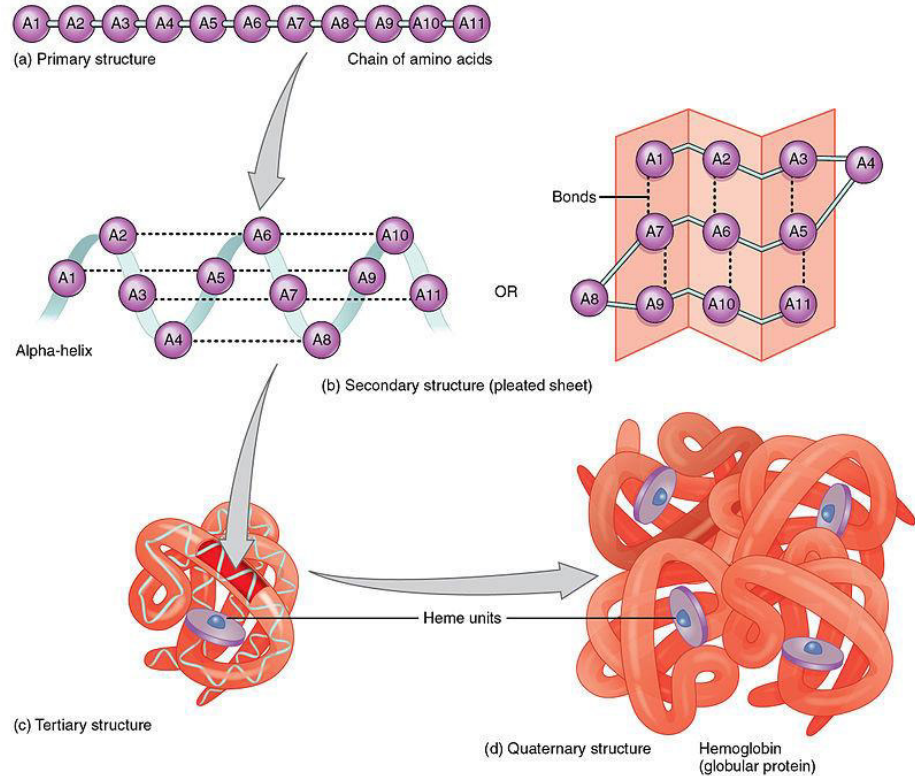


Compensatory changes

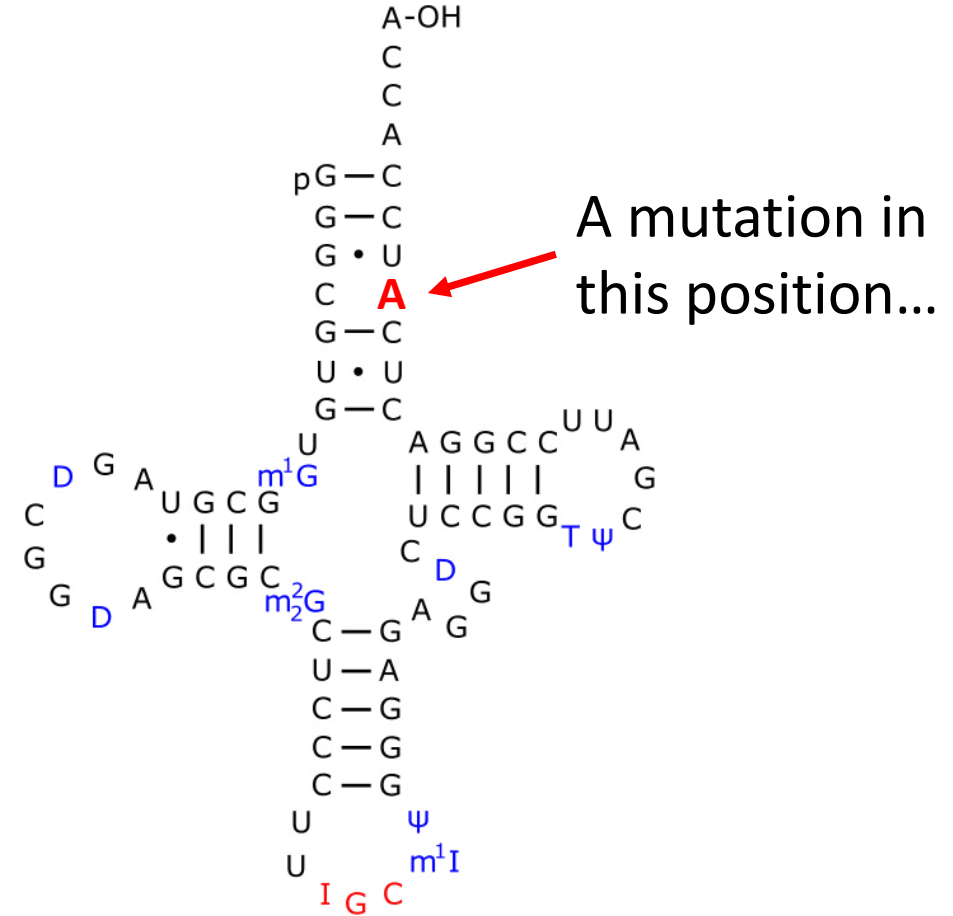


Primary and secondary structure

The function of macromolecules is explained in terms of their structure (particularly for proteins and RNA)

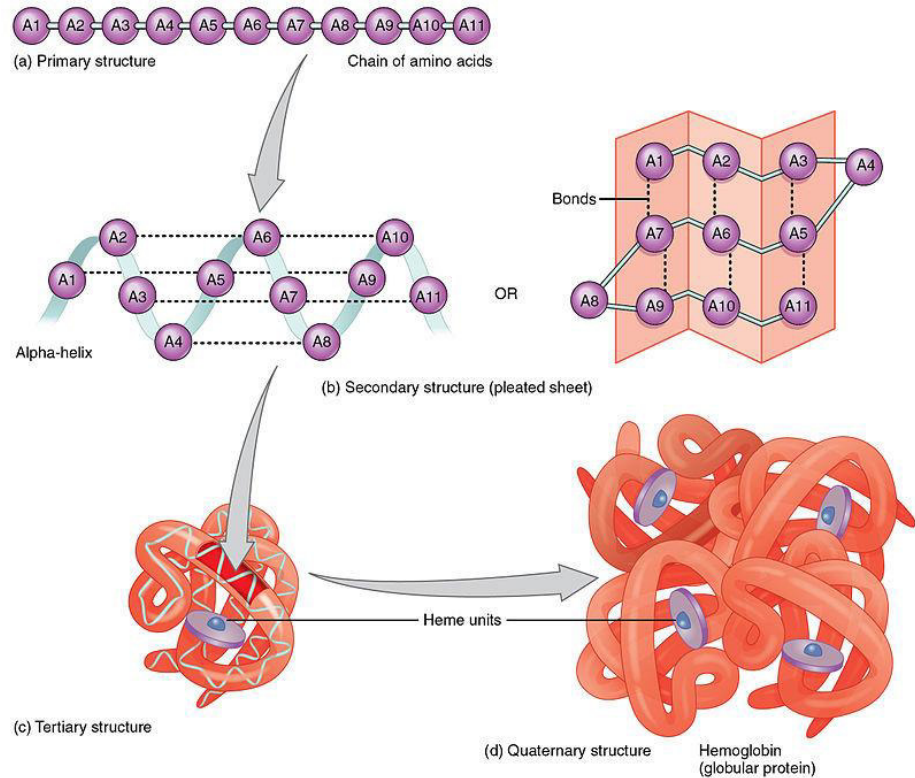


Compensatory changes



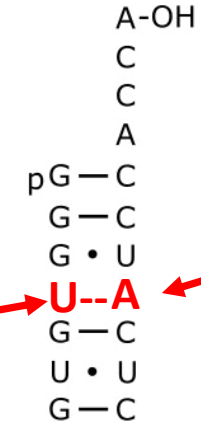
Primary and secondary structure

The function of macromolecules is explained in terms of their structure (particularly for proteins and RNA)

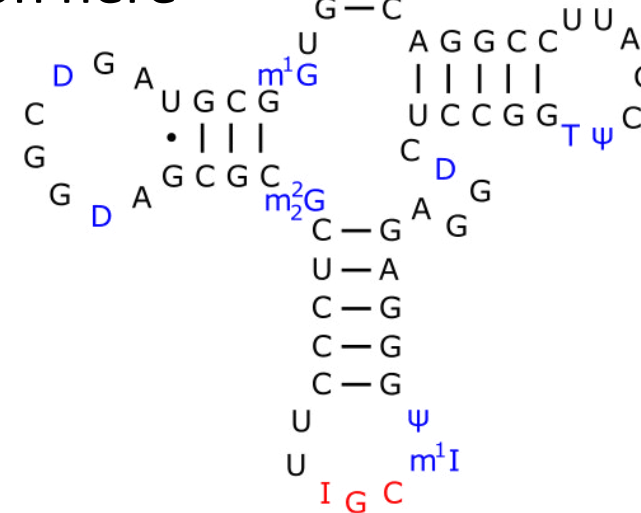


Compensatory changes

...may favor a mutation here

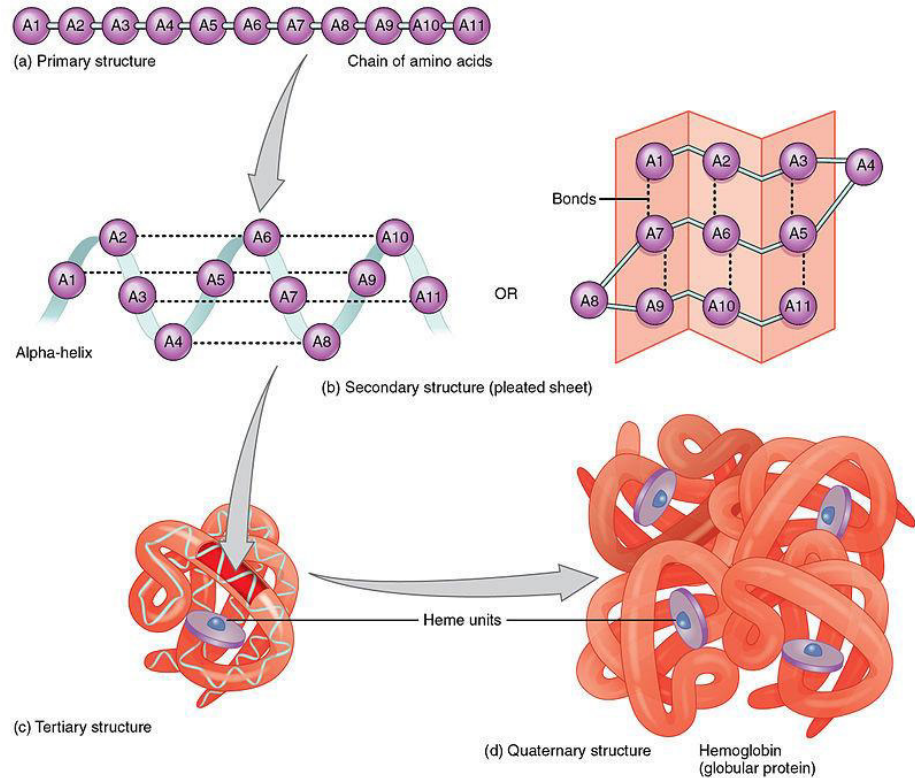


A mutation in this position...



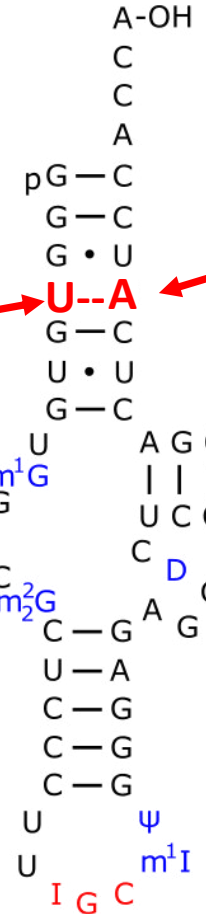
Primary and secondary structure

The function of macromolecules is explained in terms of their structure (particularly for proteins and RNA)

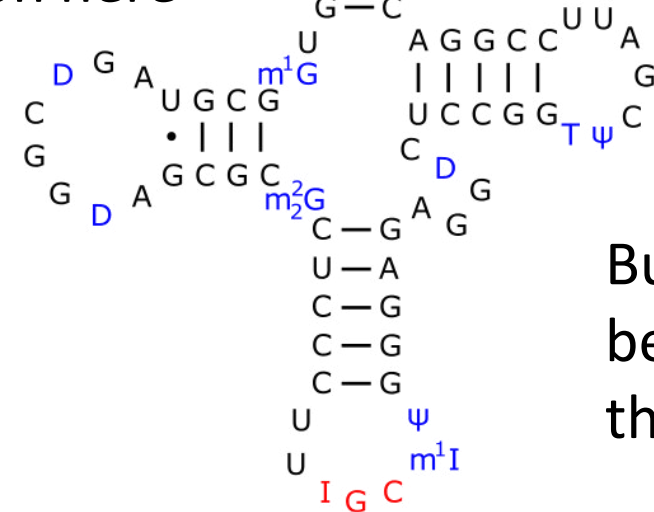


Compensatory changes

...may favor a mutation here



A mutation in this position



But it will not be obvious in the alignment

Primary and secondary structure


RNAfold (RNA structure prediction) <http://http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>

RNAfold WebServer

1 Enter Input Parameters 2 View Results

[\[Home\]](#) [New job](#) [\[Help\]](#)

The **RNAfold web server** will predict secondary structures of single stranded RNA or DNA sequences. Current limits are 7,500 nt for partition function calculations and 10,000 nt for minimum free energy only predictions.





Simply paste or upload your sequence below and click *Proceed*. To get more information on the meaning of the options click the  symbols. You can test the server using [this sample sequence](#).

Paste or type your **sequence** here: [\[clear\]](#)

[Show constraint folding](#)




Or upload a file in FASTA format: No file chosen

Fold algorithms and basic options

- minimum free energy (MFE) and partition function 
- minimum free energy (MFE) only 
- no GU pairs at the end of helices 
- avoid isolated base pairs 

[Show advanced options](#)

Output options

- interactive RNA secondary structure plot 
- RNA secondary structure plots with reliability annotation (Partition function folding only) 
- Mountain plot 

Notification via e-mail upon completion of the job (optional):

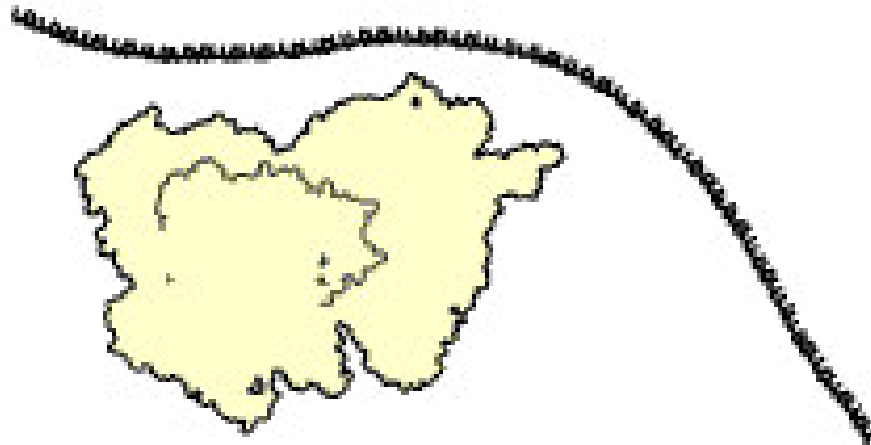
Ribosome

— A molecular machine, serves for biological protein synthesis (translation).

Two subunit:

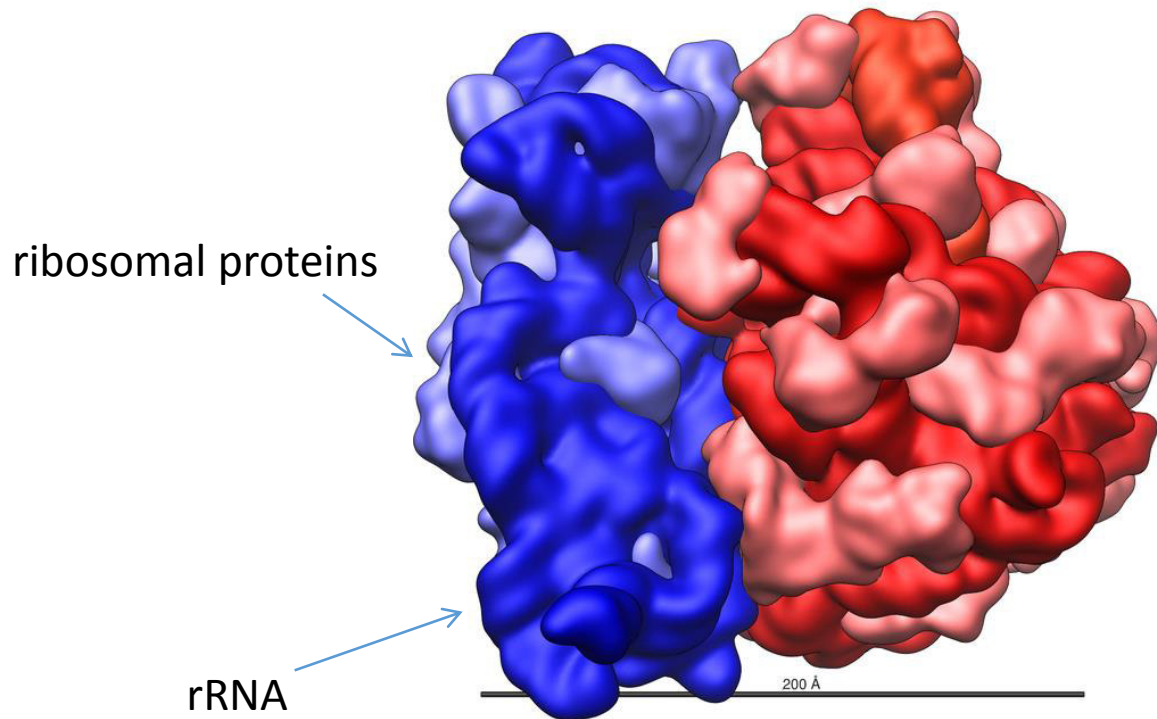
Large: binds tRNA and joins amino acids to form the protein

Small: binds and reads the mRNA

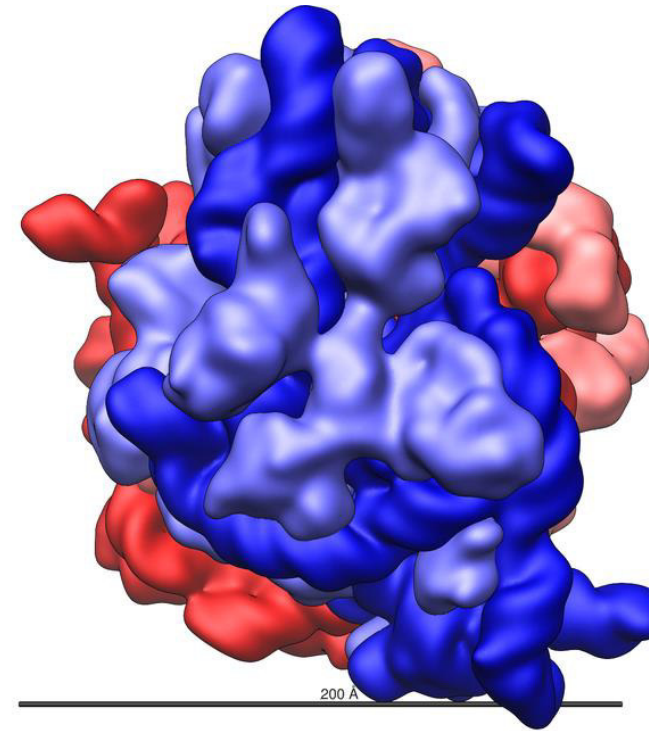


Structure of the 70S *E.coli* ribosome

With large 50S ribosomal subunit (red); small 30S ribosomal subunit (blue)



By Vossman



50S:

23S rRNA

5S rRNA

ribosomal proteins (pink)

30S:

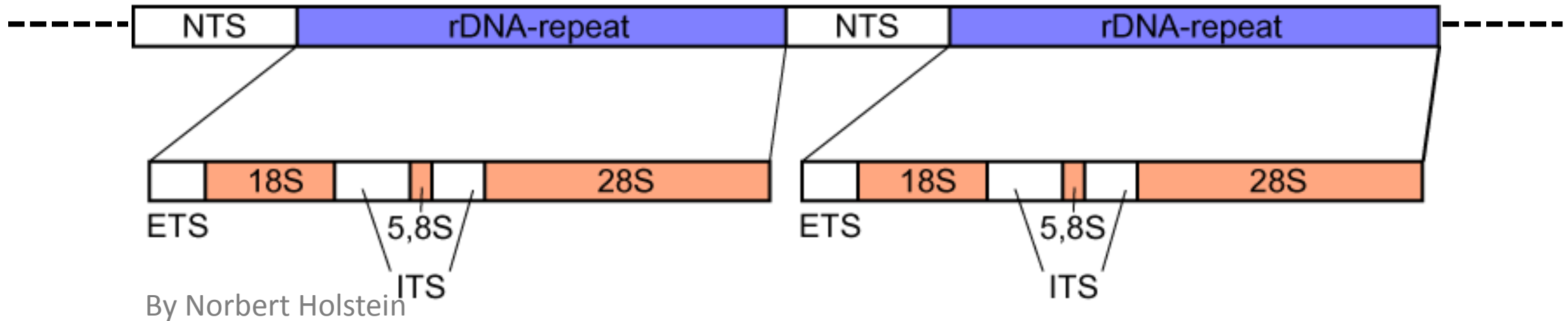
16S rRNA (dark blue)

ribosomal proteins

(light blue)

Ribosomal DNA

— a DNA sequence that codes for ribosomal RNA



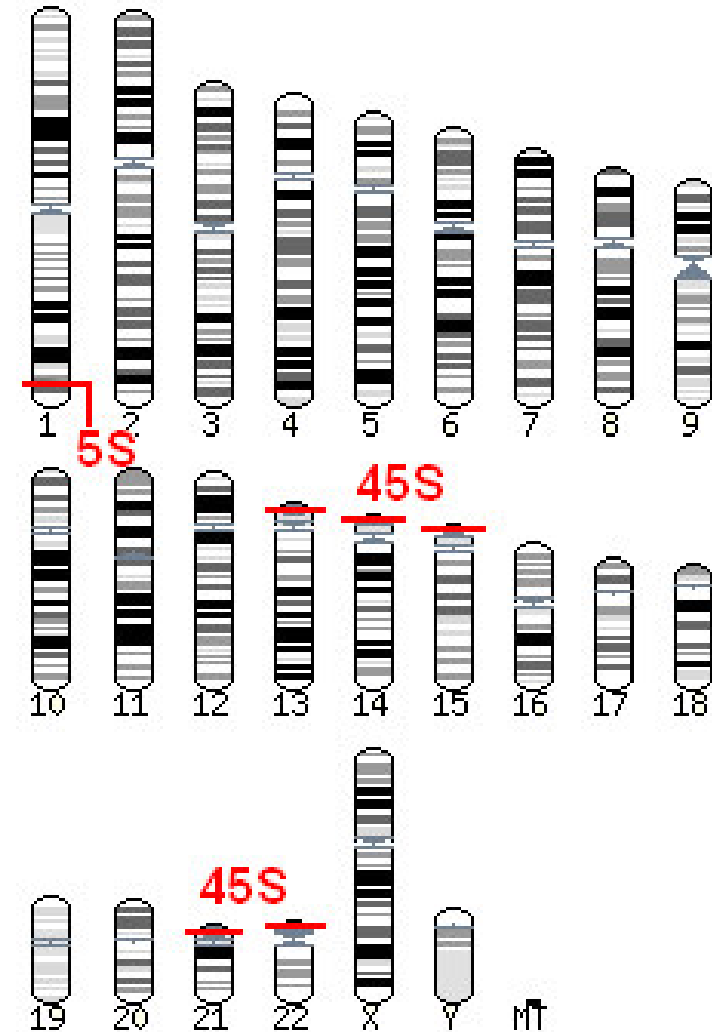
Gene cluster of 18S, 5.8S, and 28S
NTS, non-transcribed spacer
ETS, external transcribed spacer
ITS, **internal transcribed spacers 1 and 2**

Human genome:
5 chromosomes with the repeat unit
chromosomes 13, 14, 15, 21 and 22

Human genome: ribosomal RNA gene clusters

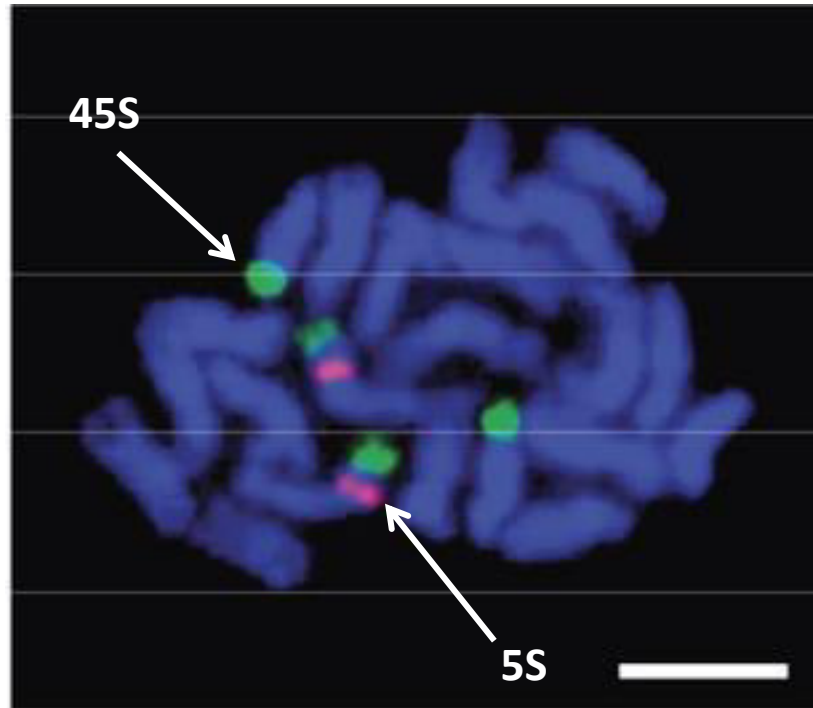
- One cluster of 5S RNA genes: chromosome 1
- Five clusters of 45S ribosomal RNA genes: chromosomes 13, 14, 15, 21 and 22
- 5S: 100 copies
- 45S: 300 copies

Caburet et al. 2005; Stults et al. 2008



Laurence A. Moran

Citrullus lanatus chromosomes

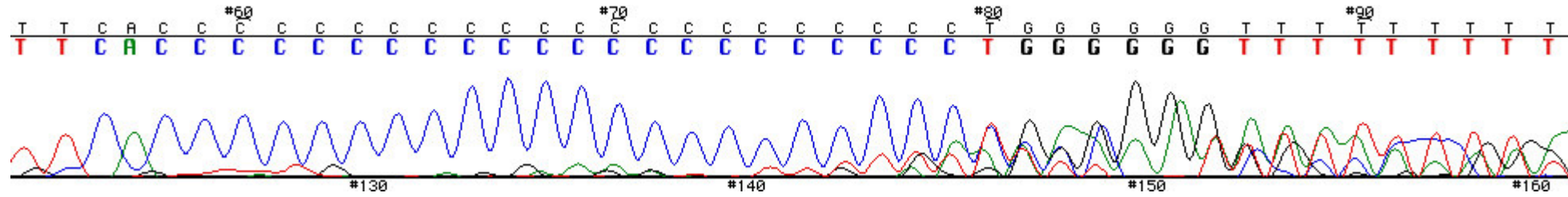


FISH (Fluorescence in situ hybridization)
using rDNA (green, 45S; pink, 5S) probes

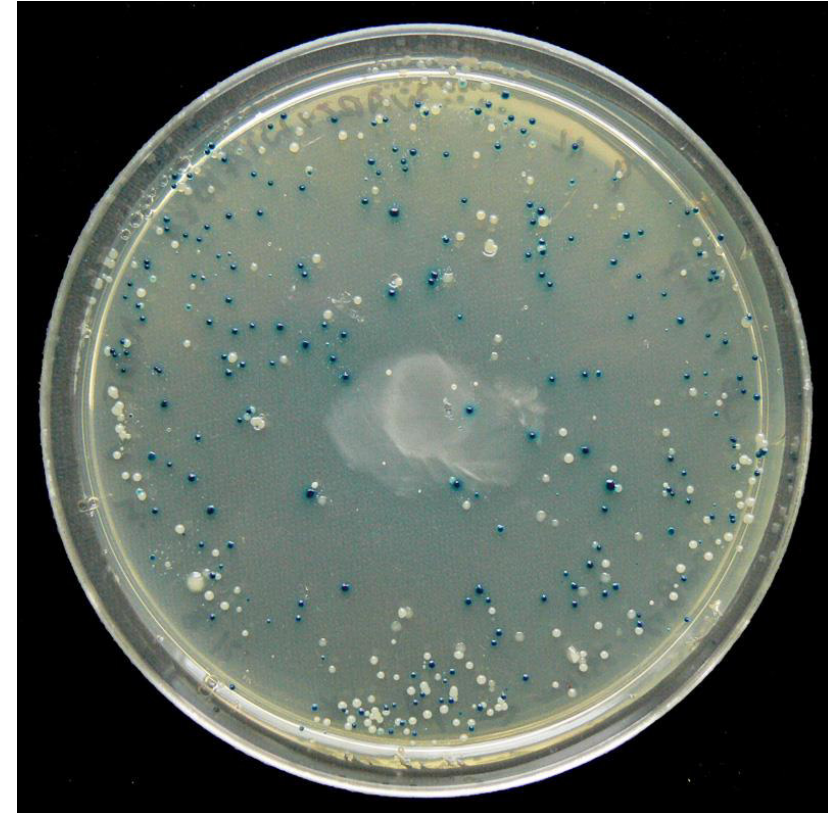
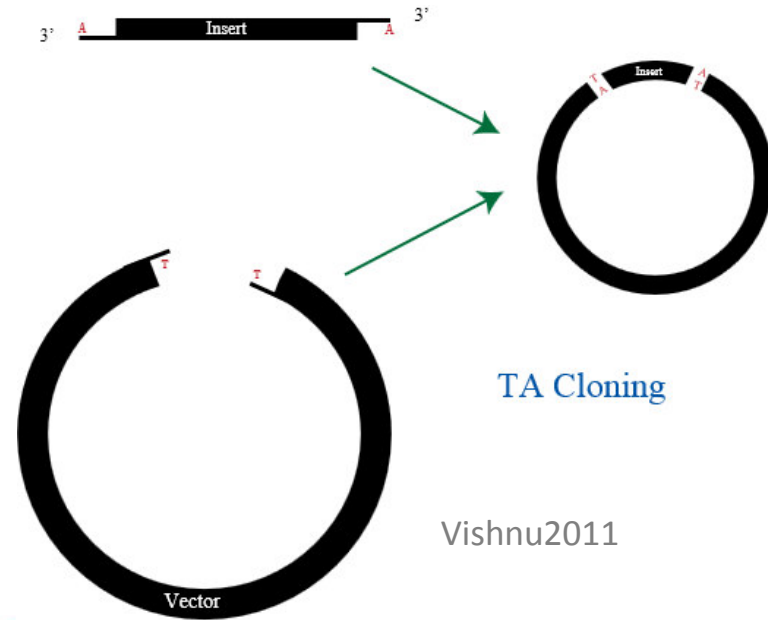
Concerted evolution

- Concerted evolution: a process in which related genes within a species undergo genetic exchange, causing their sequence evolution to be concerted over some period of time
- Concerted evolution can explain the similarity among multicopy rDNA genes
- INCOMPLETE concerted evolution results in polymorphisms among copies of rDNA sequences

Consequence of incomplete concerted evolution



How to resolve it? –cloning the PCR product



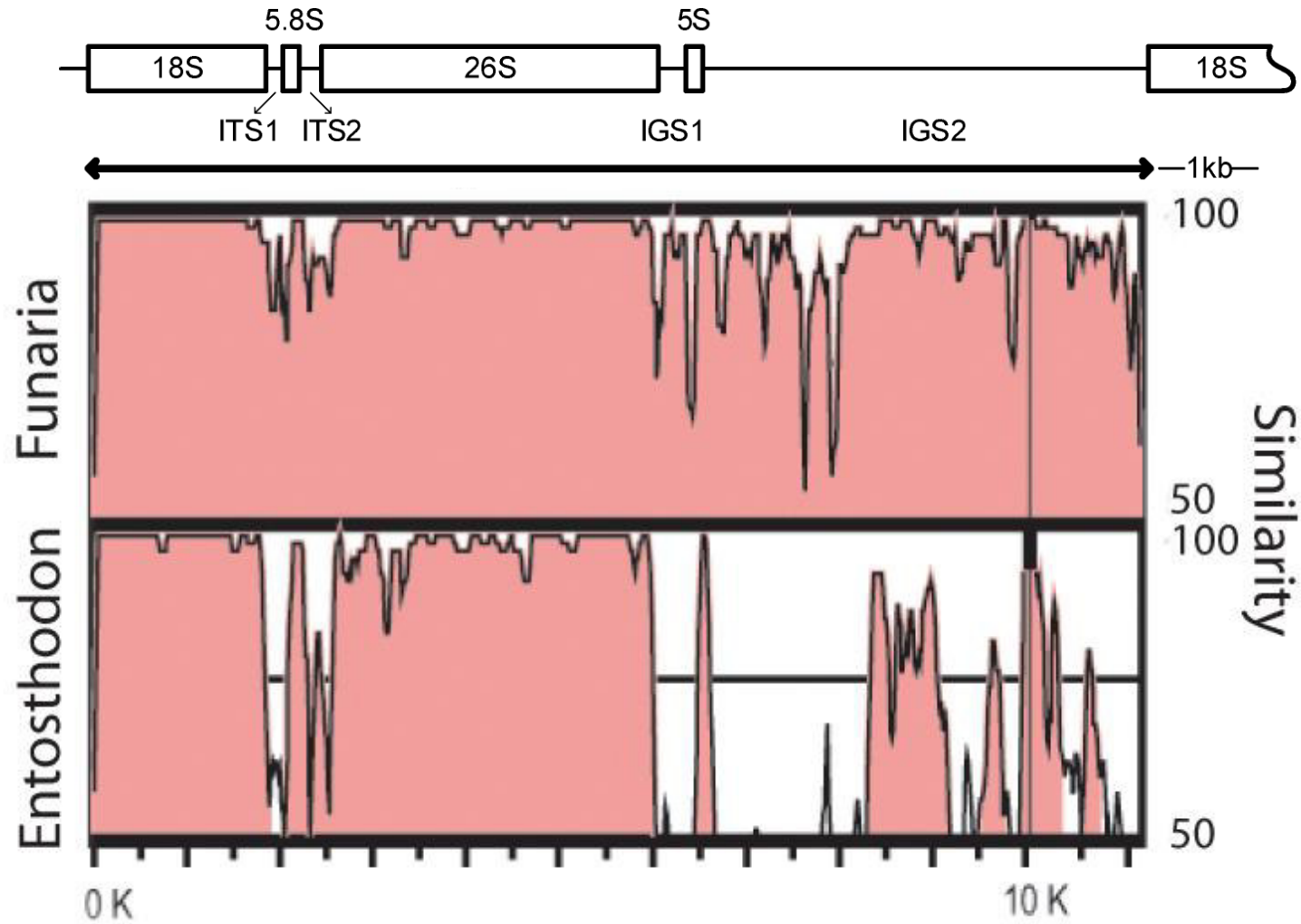
By Stefan Walkowski

An LB agar plate showing the result of a blue white screen

ITS sequences are widely used for phylogenetic studies and DNA barcoding, because:

- As 18S and 26S genes are so conserved, it is easy to design universal primers
- ITS has multiple copies, and is easy to be amplified
- The locus evolves fast, so that it is variable and provides much phylogenetic information, potentially useful in resolving closely related organisms

Nuclear ribosomal RNA repeat unit in mosses



Evolution of the Transcription Unit of Ribosomal RNA*

Robert P. Perry, Tsai-Ying Cheng, Jerome J. Freed,
Jay R. Greenberg, Dawn E. Kelley, and Kenneth D. Tartof

THE INSTITUTE FOR CANCER RESEARCH, PHILADELPHIA, PENNSYLVANIA

Communicated by Thomas F. Anderson, October 15, 1969

18, 200 papers

The screenshot shows a Google Scholar search interface. The search bar contains the text "internal transcribed spacers phylogeny". Below the search bar, it indicates "About 18,200 results (0.10 sec)". The search results are listed in a table-like format with a left-hand navigation menu. The first result is titled "Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae" by BG Baldwin, published in 1992. The second result is titled "Phylogenetic analysis of Sorghum and related taxa using internal transcribed spacers of nuclear ribosomal DNA" by Y Sun, DZ Skinner, GH Liang, and SH Hulbert, published in 1994. The left-hand menu includes options for "Articles", "Case law", "My library", "Any time", "Since 2015", "Since 2014", "Since 2011", and "Custom range...".

Articles	Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae	berkeley.edu [PDF] UCONNLinks
Case law	BG Baldwin - <i>Molecular phylogenetics and evolution</i> , 1992 - Elsevier	
My library	Abstract The internal transcribed spacer (ITS) region of 18–26S nuclear ribosomal DNA was sequenced in 12 representatives of the Compositae subtribe Madiinae and two outgroup species to assess its utility for phylogeny reconstruction. High sequence alignability and ...	
Any time	Cited by 933 Related articles All 6 versions Web of Science: 671 Import into EndNote Save More	
Since 2015	Phylogenetic analysis of Sorghum and related taxa using internal transcribed spacers of nuclear ribosomal DNA	ask-force.org [PDF]
Since 2014	Y Sun, DZ Skinner, GH Liang, SH Hulbert - <i>Theoretical and Applied ...</i> , 1994 - Springer	
Since 2011	Abstract The phylogenetic relationships of the genus <i>Sorghum</i> and related genera were studied by sequencing the nuclear ribosomal DNA (rDNA) internal transcribed spacer region	
Custom range...		

Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*

Conrad L. Schoch^{a,1},
Wen Chen^b, and Fun

[OPEN ACCESS](#) Freely available online



^aNational Center for Biotechnology and Microbiology, Agricultural University of Amsterdam
^dCentraalbureau voor Schiedplantsoorten

Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals

Edited* by Daniel H. Janze

Hui Yao¹*, Jingyuan Song¹*, Chang Liu¹*, Kun Luo^{1,2}, Jianping Han¹, Ying Li¹, Xiaohui Pang¹, Hongxi Xu⁴, Yingjie Zhu^{3*}, Peigen Xiao¹, Shilin Chen^{1*}

¹ Institute of Medicinal Plant Development, Hubei University of Chinese Medicine, Chengde, Sichuan, People's Republic of China

[OPEN ACCESS](#) Freely available online



Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species

Shilin Chen^{1*}, Hui Yao¹, Jianping Han¹, Chang Liu², Jingyuan Song^{1*}, Linchun Shi¹, Yingjie Zhu¹, Xinye Ma¹, Ting Gao¹, Xiaohui Pang¹, Kun Luo³, Ying Li¹, Xiwen Li¹, Xiaocheng Jia¹, Yulin Lin¹, Christine Leon⁴

¹ Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, People's Republic of China, ² Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, People's Republic of China, ³ Department of Pharmacy, Hubei University of Chinese Medicine, Wuhan, Hubei, People's Republic of China, ⁴ Royal Botanic Gardens, Kew, Richmond, United Kingdom