A primer for practical phylogenetic data gathering. Uconn EEB3899-007. Spring 2015

Sessions 7 and 8

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

Rafael Medina (rafael.medina.bry@gmail.com)
Yang Liu (yang.liu@uconn.edu)

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:

- 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

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

The concept of molecular clock

Specimen 1

AGGTCAGATTACA

Specimen 2

ATGTCACATTTGA

The concept of molecular clock

Specimen 1

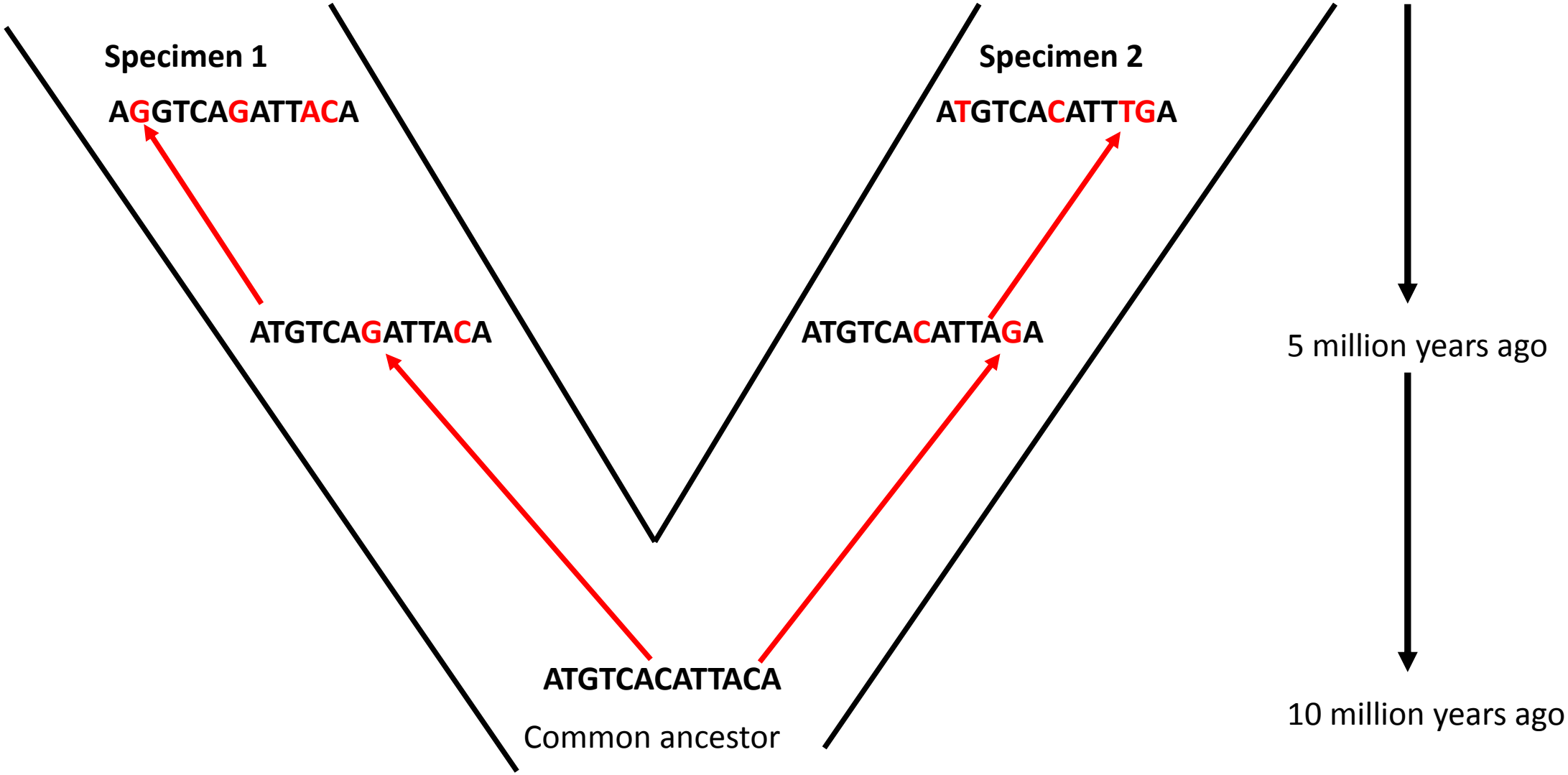
AGGTCAGATTACA

Specimen 2

ATGTCACATTTGA

The concept of molecular clock

The concept of molecular clock



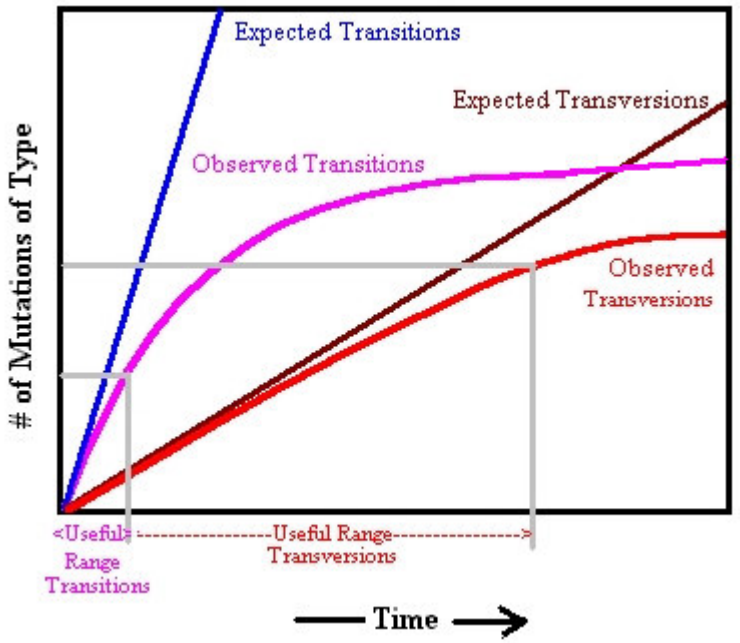
The concept of molecular clock

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

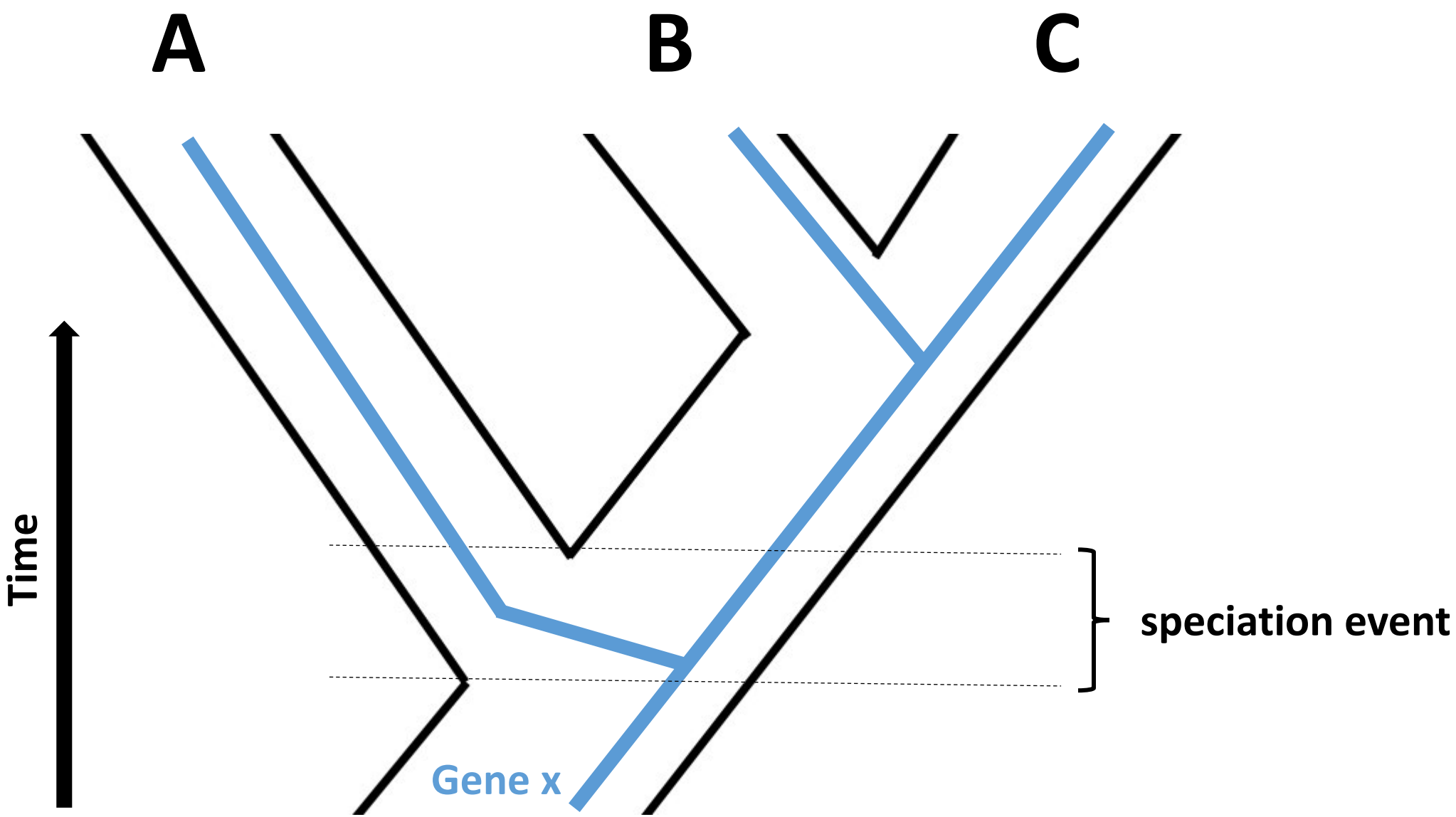
- The rate of change of a sequence may be very variable depending on the kind of molecule, the particular region or the phylogenetic lineage

- Other phenomena such as long branch attraction (due to saturation of very variable sites) have to be taken into account

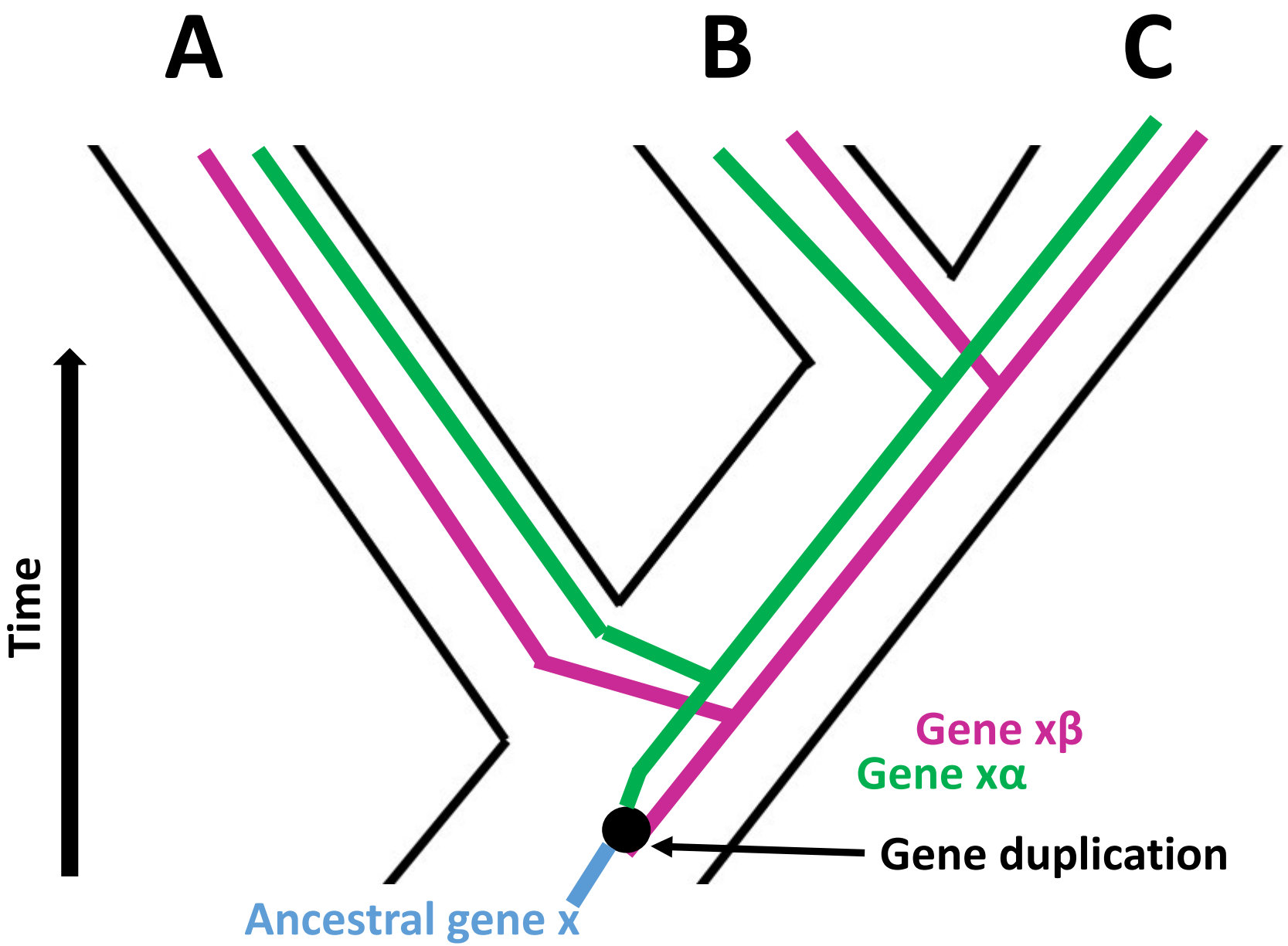
- The alignment should be based in homology relationships (in particular, **orthology**)



Homology, orthology and paralogy

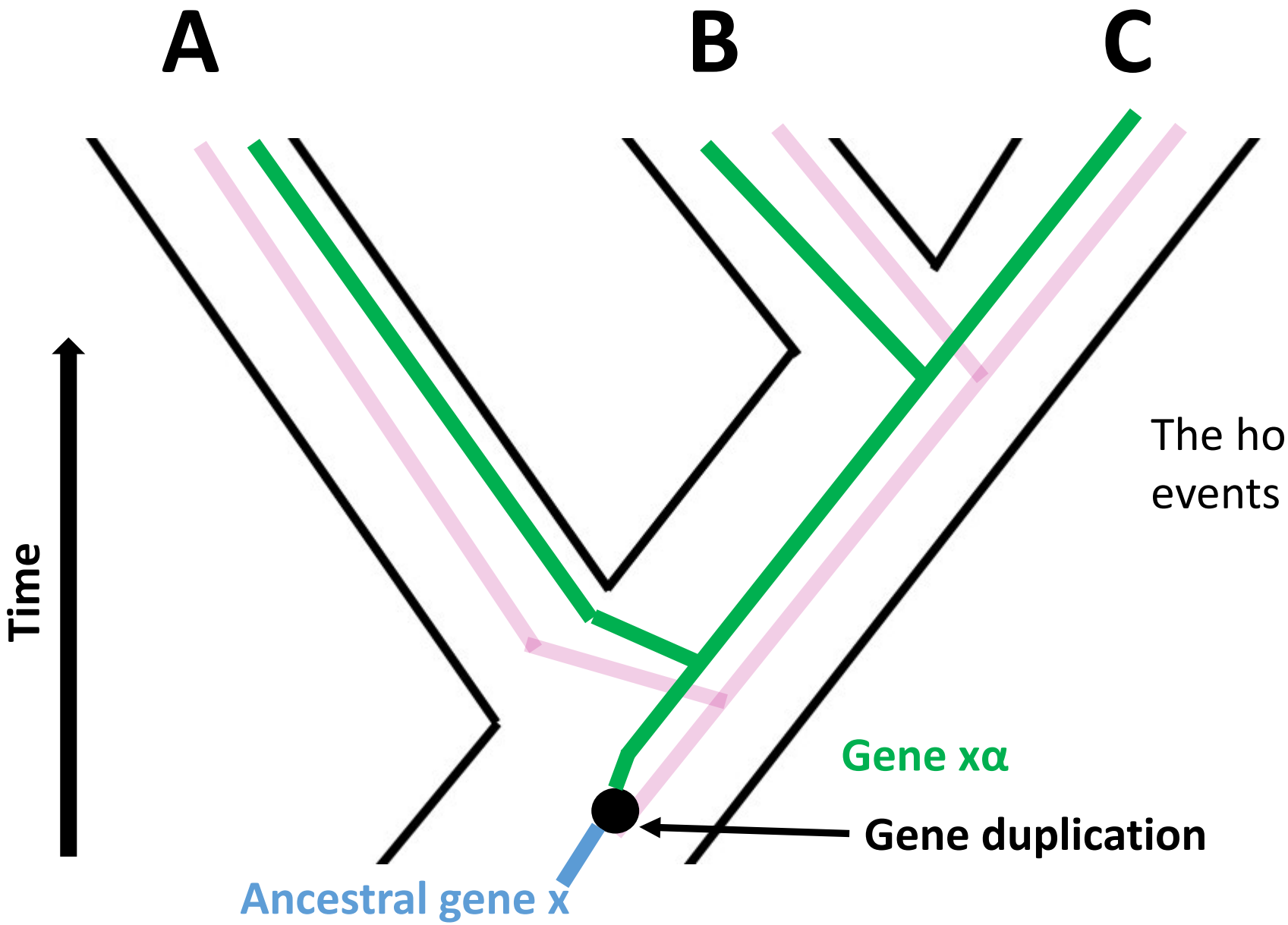


Homology, orthology and paralogy



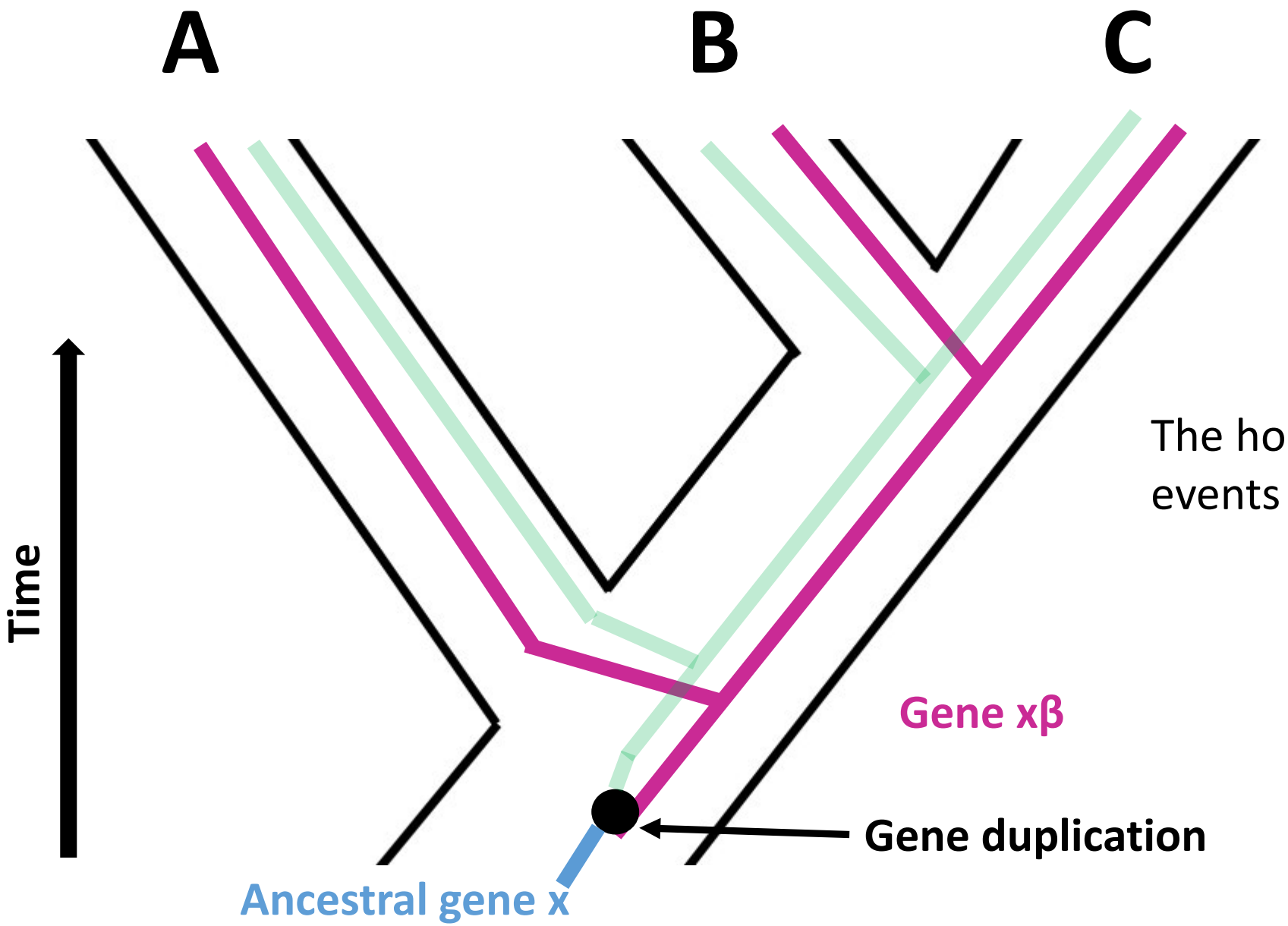
All the resulting sequences of the gen x are **homologous** because they derived from a common ancestor

Homology, orthology and paralogy



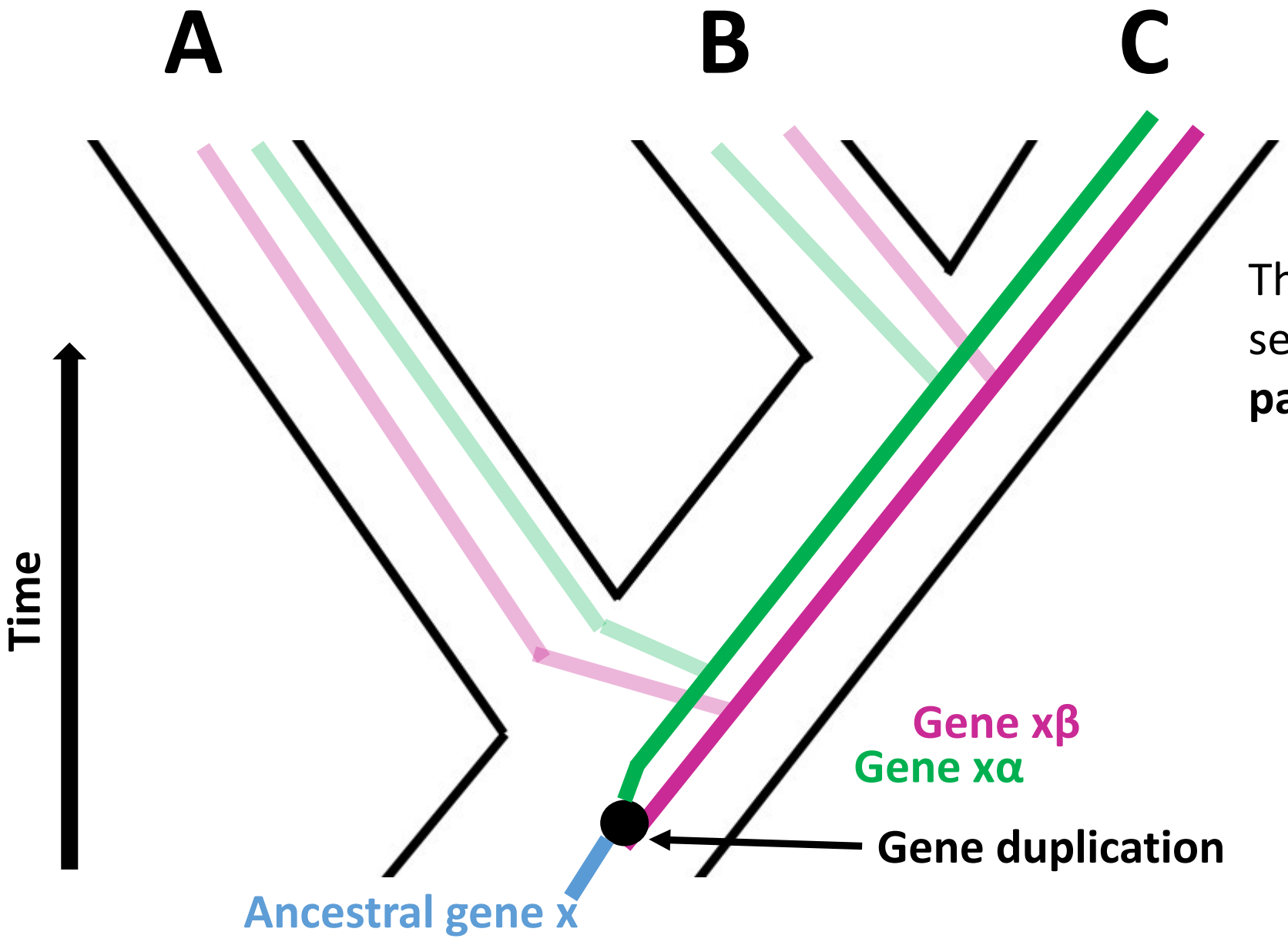
The homology caused by speciation events is called **orthology**

Homology, orthology and paralogy



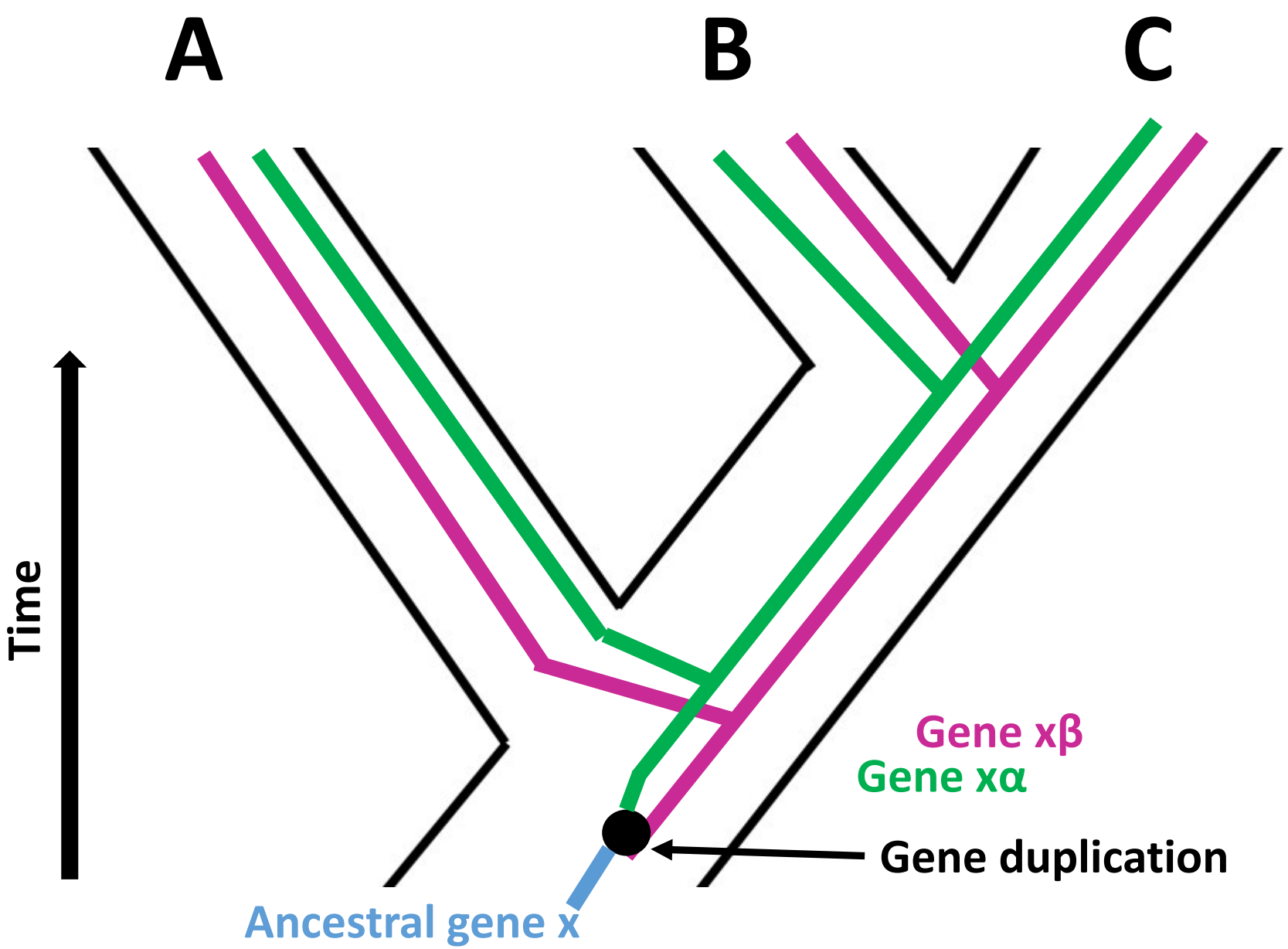
The homology caused by speciation events is called **orthology**

Homology, orthology and paralogy



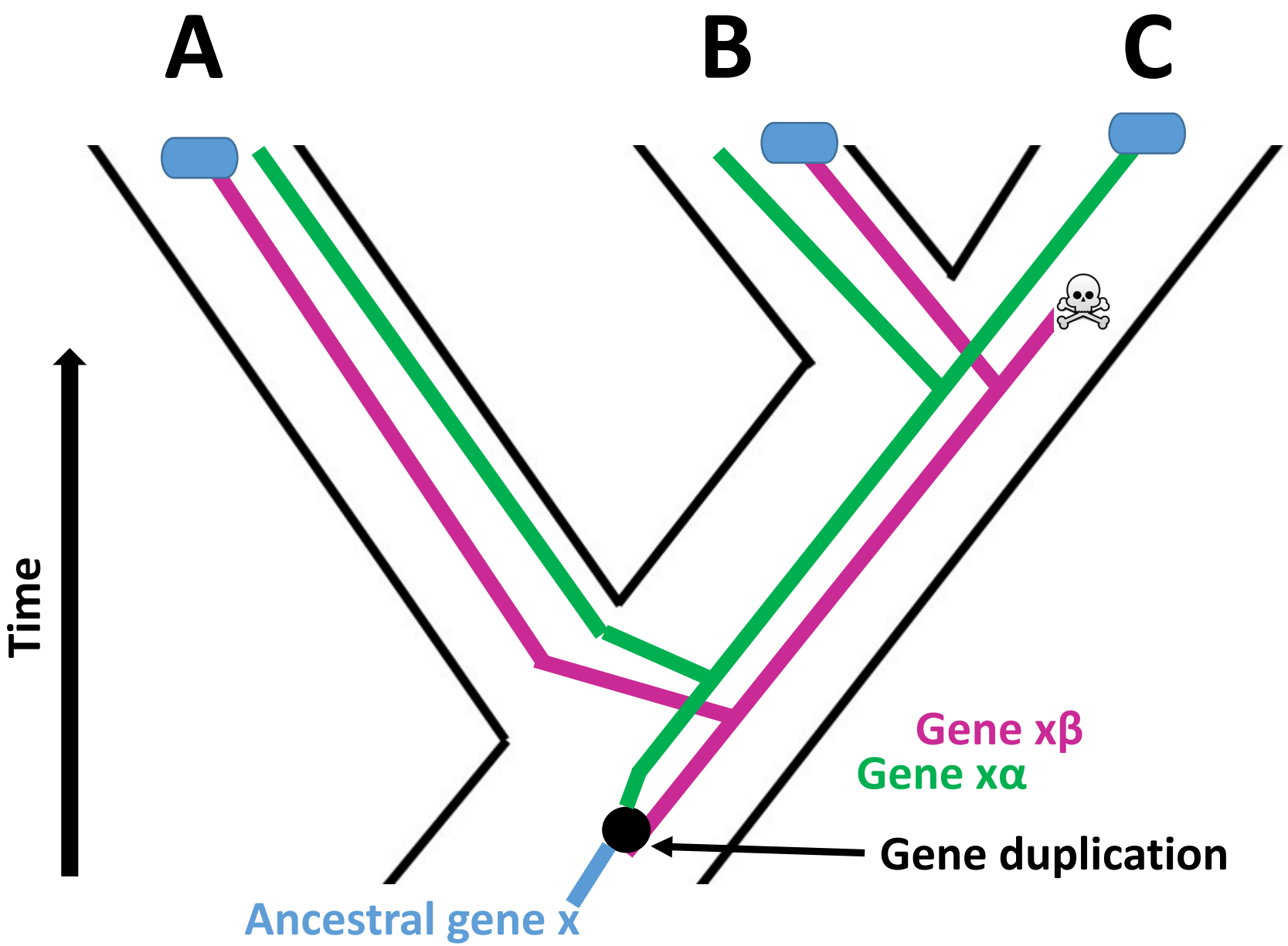
The homology caused by a sequence duplication is called **paralogy**

Homology, orthology and paralogy

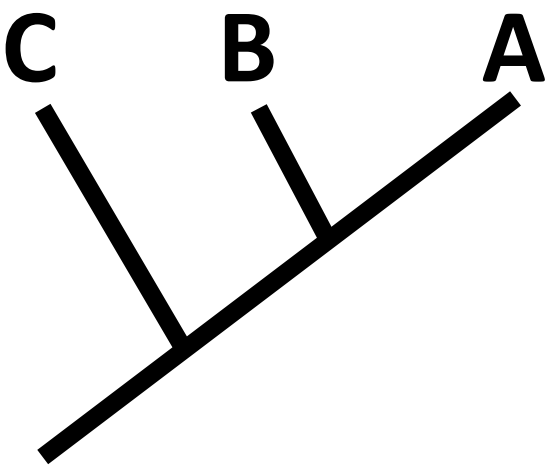


If our objective is phylogenetic reconstruction we should only align orthologous sequences, otherwise our result would be misleading

Homology, orthology and paralogy

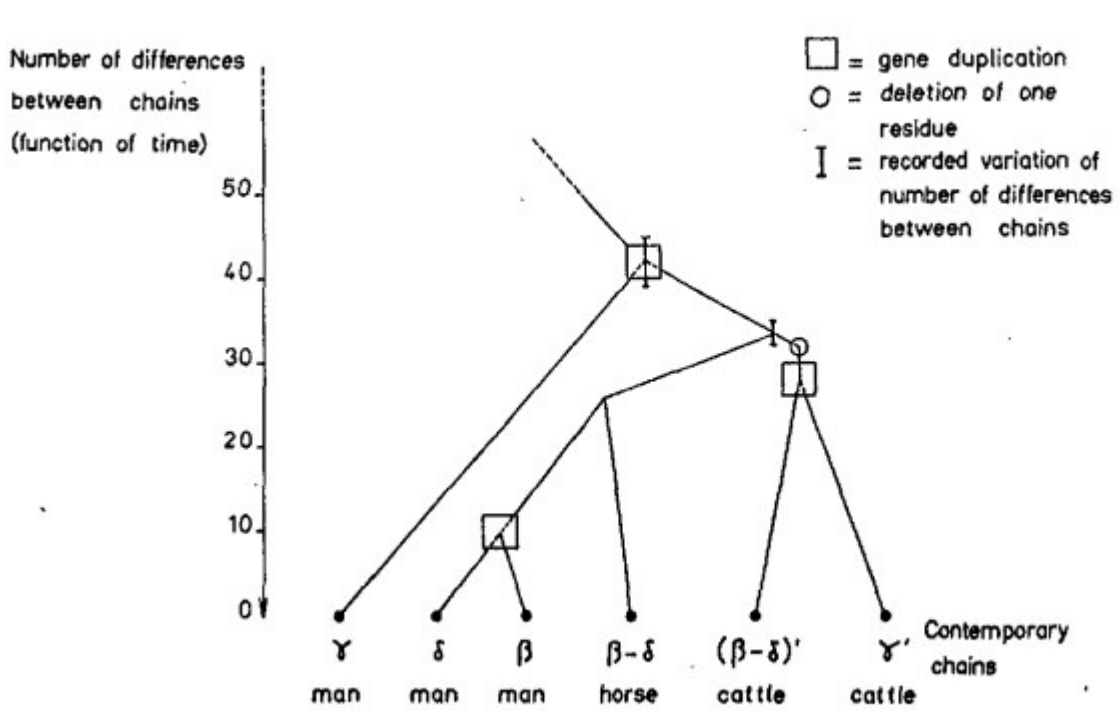


If our objective is phylogenetic reconstruction we should only align orthologous sequences, otherwise our result would be misleading

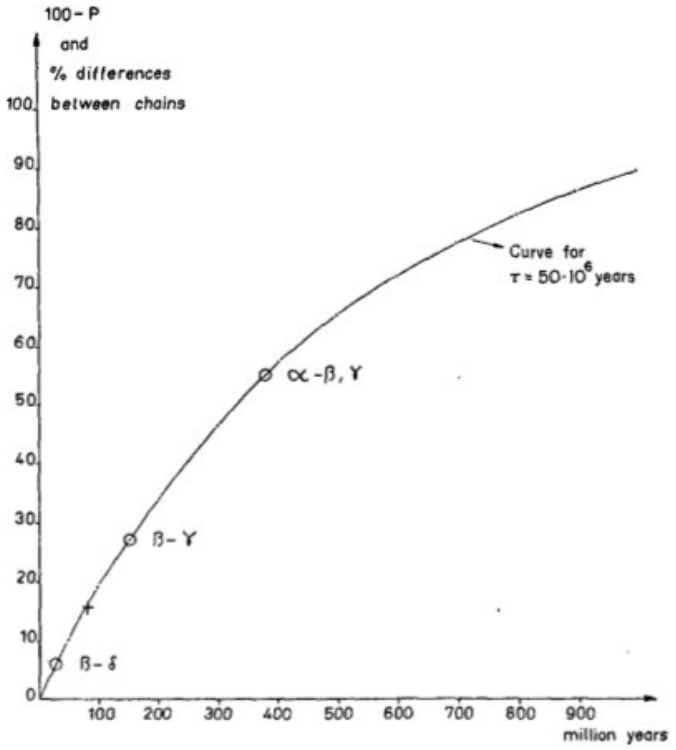


Homology, orthology and paralogy

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

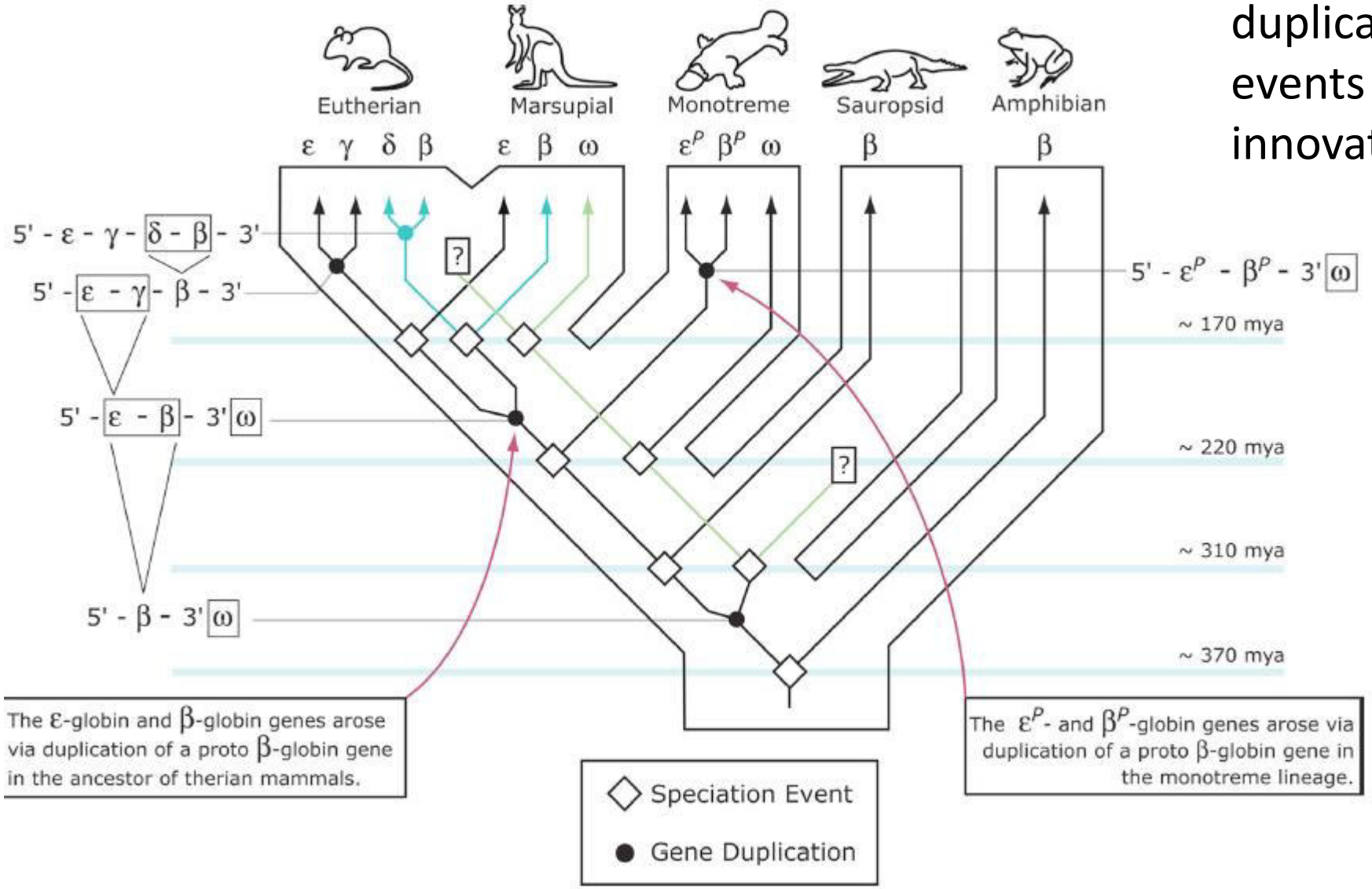


4. Probable evolutionary relationship of some mammalian hemoglobin chains.



Homology, orthology and paralogy

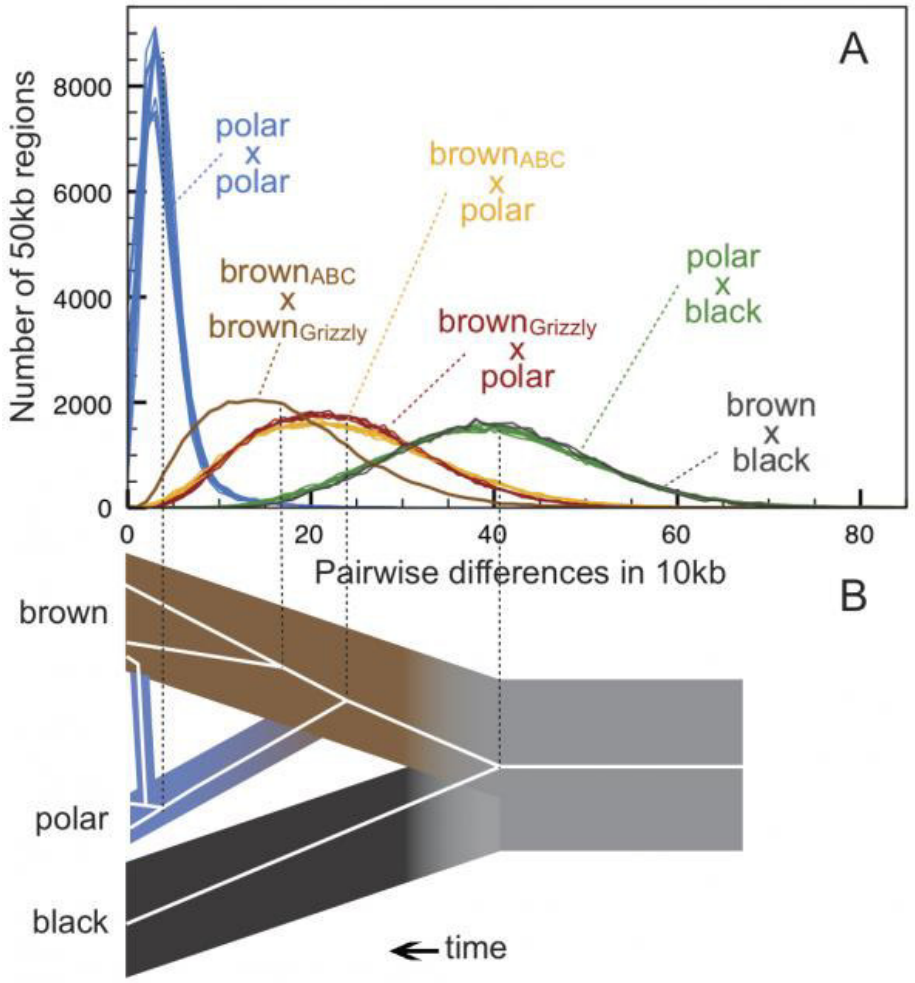
Gene (and genome) duplications have been key events for evolutionary innovation

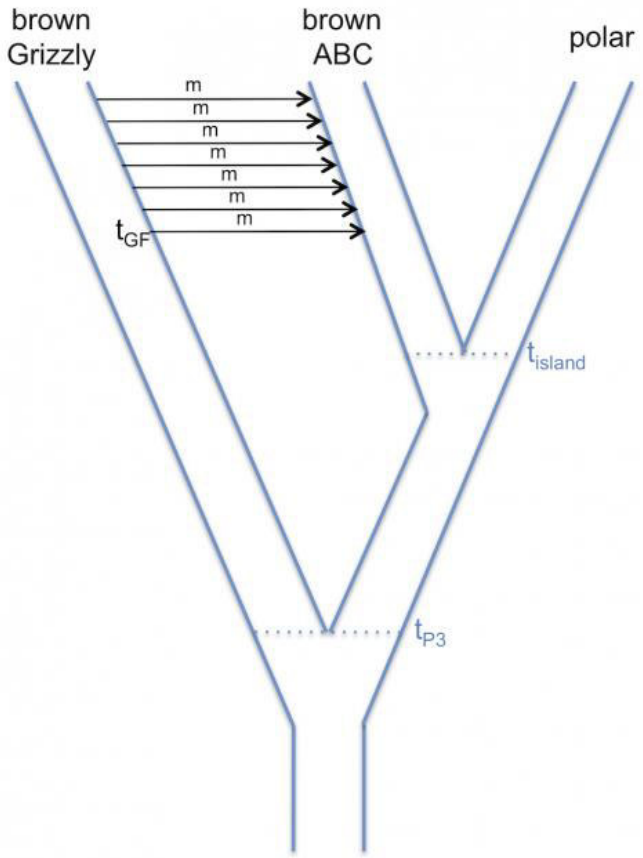


Opazo et al. 2008. Proc Natl Acad Sci USA 105

Non tree-like evolution

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

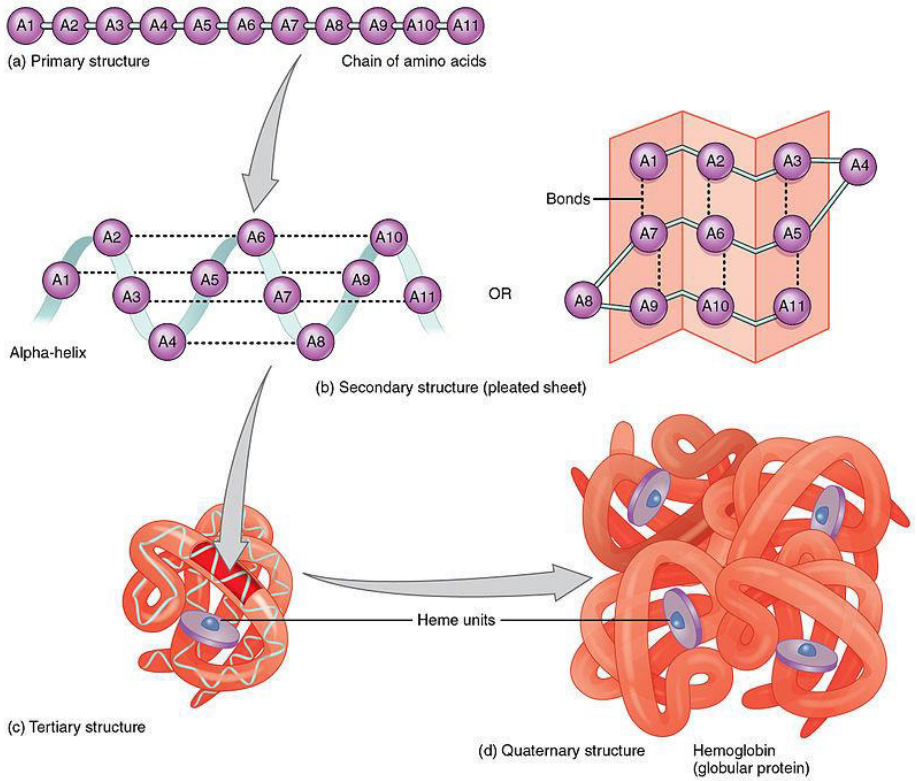




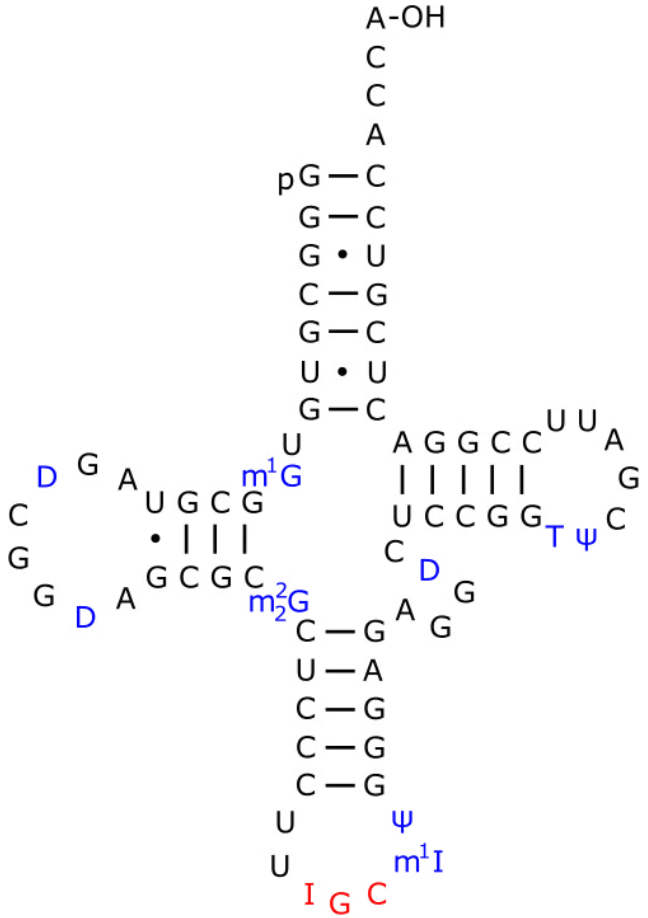
Cahill et al. 2013. PLoS Genetics9 e1003345

Primary and secondary structure

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

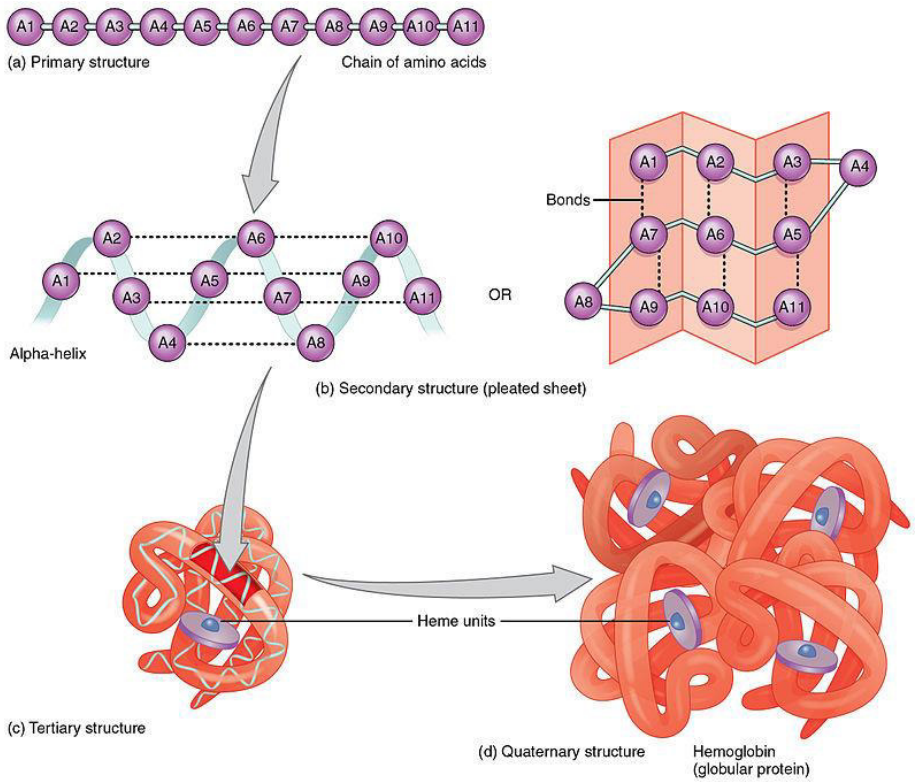


Compensatory changes

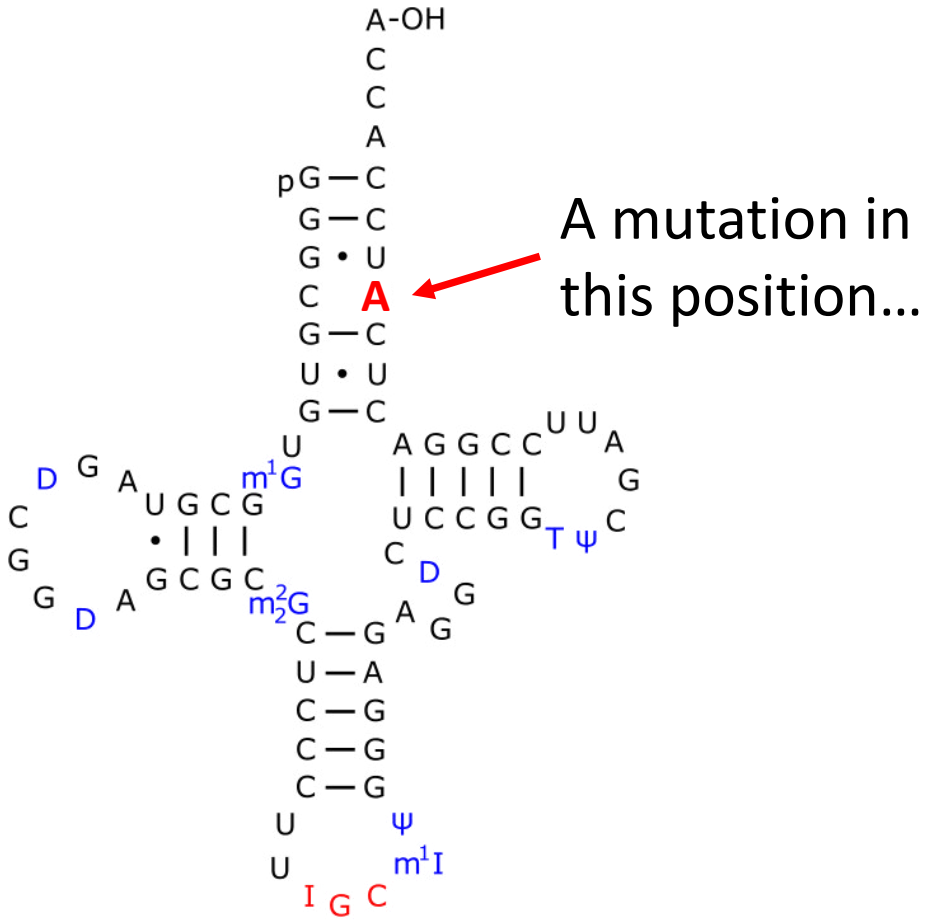


Primary and secondary structure

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

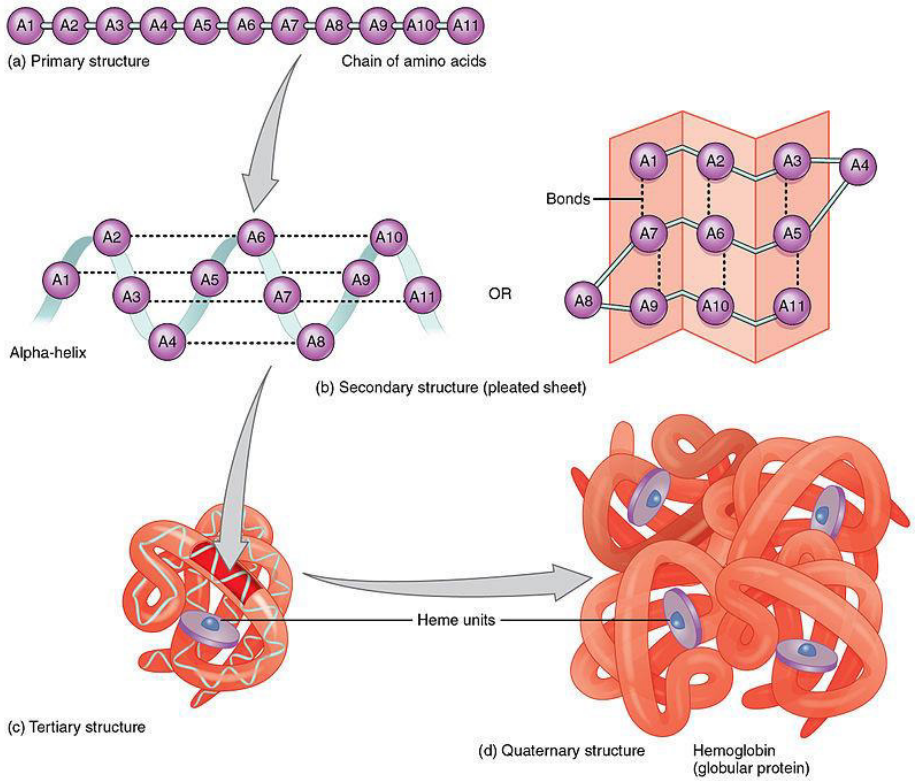


Compensatory changes

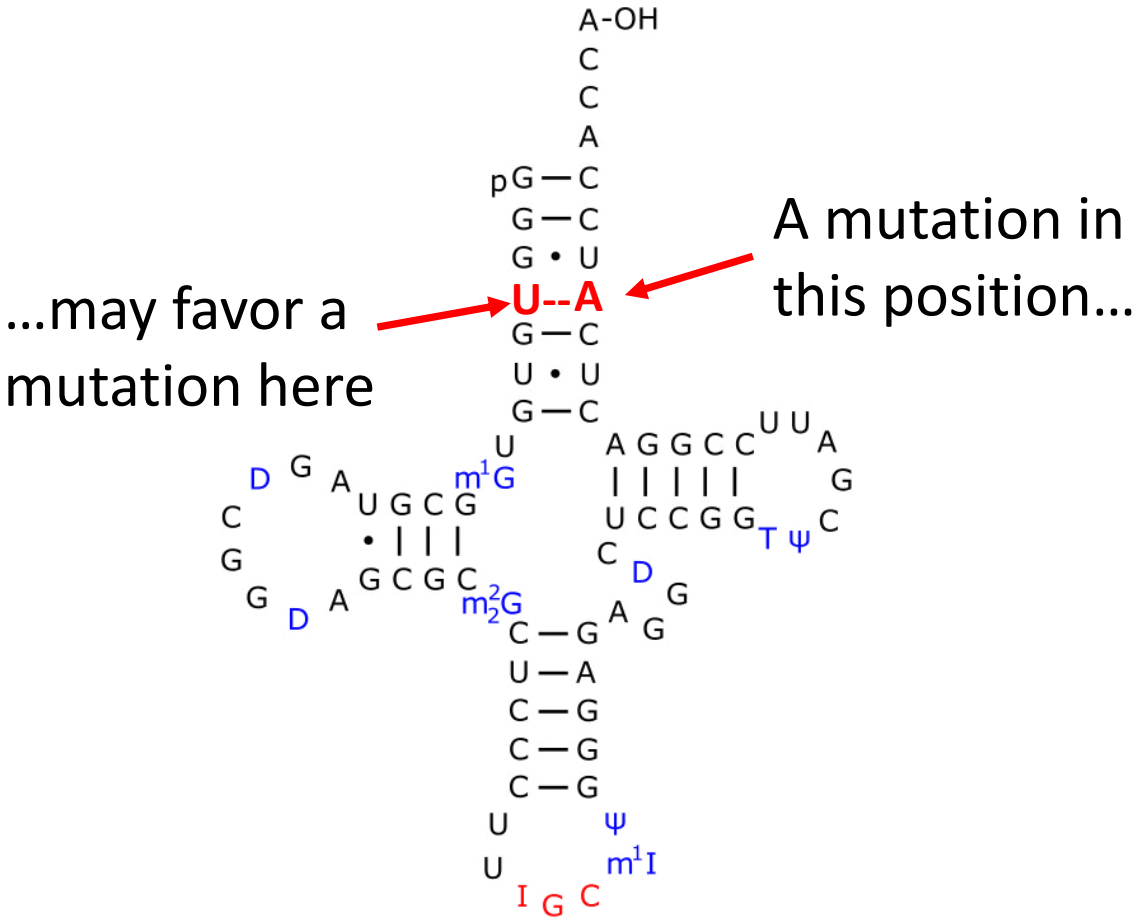


Primary and secondary structure

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



Compensatory changes

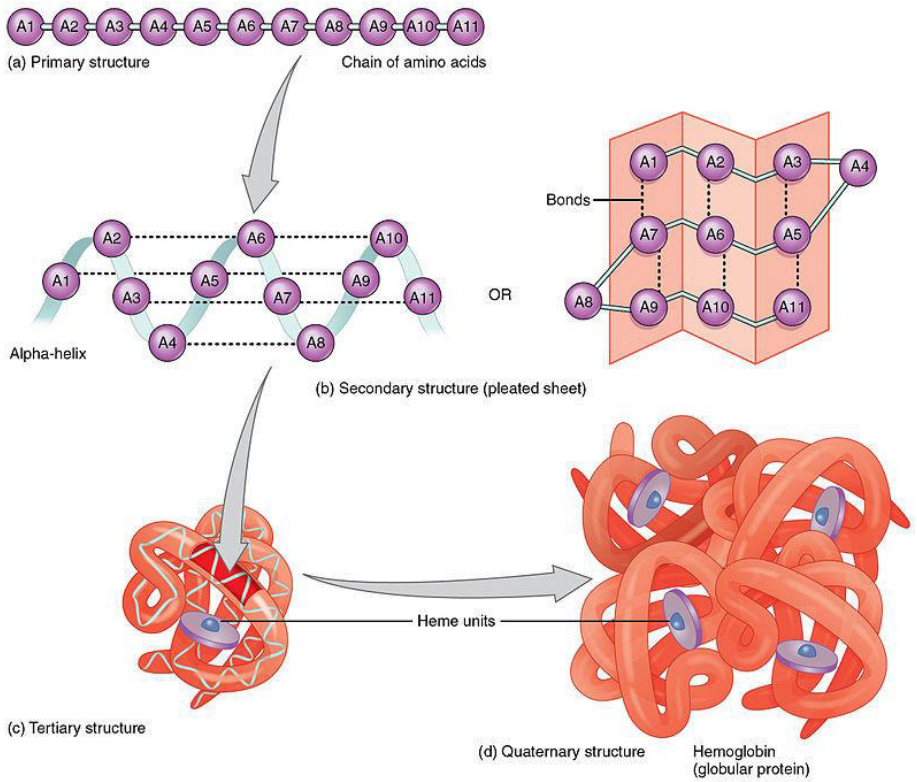


…may favor a mutation here

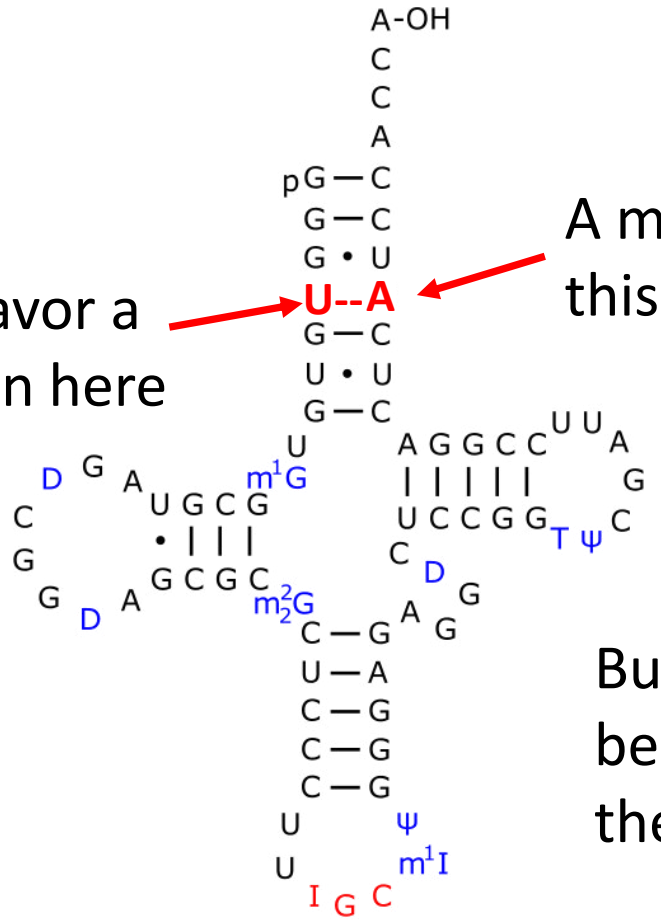
A mutation in this position…

Primary and secondary structure

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



Compensatory changes



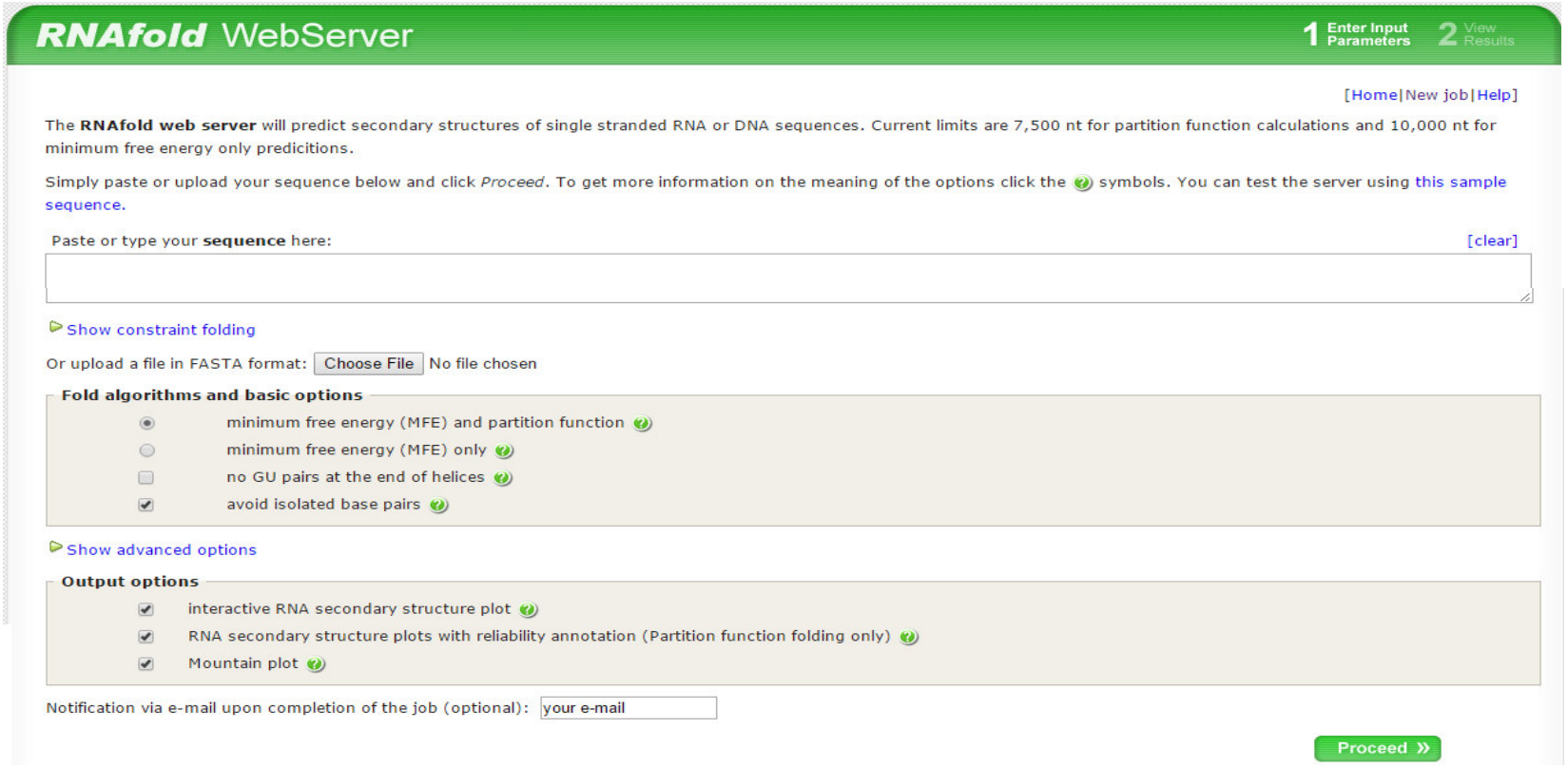
…may favor a mutation here

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



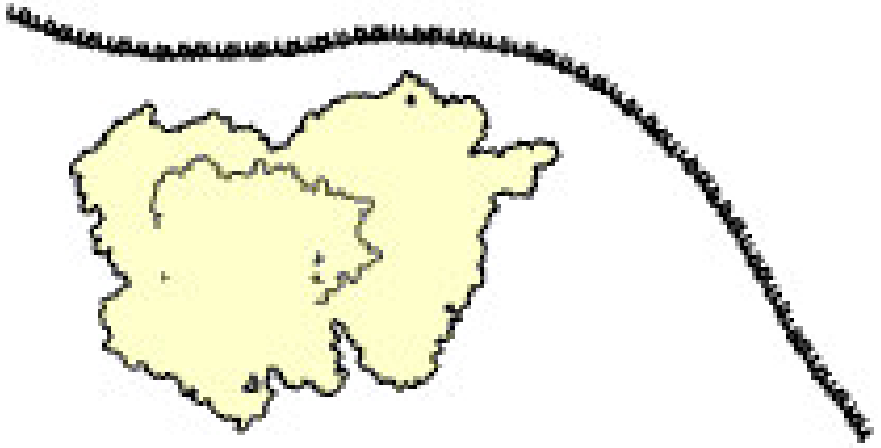
Ribosome

—— 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



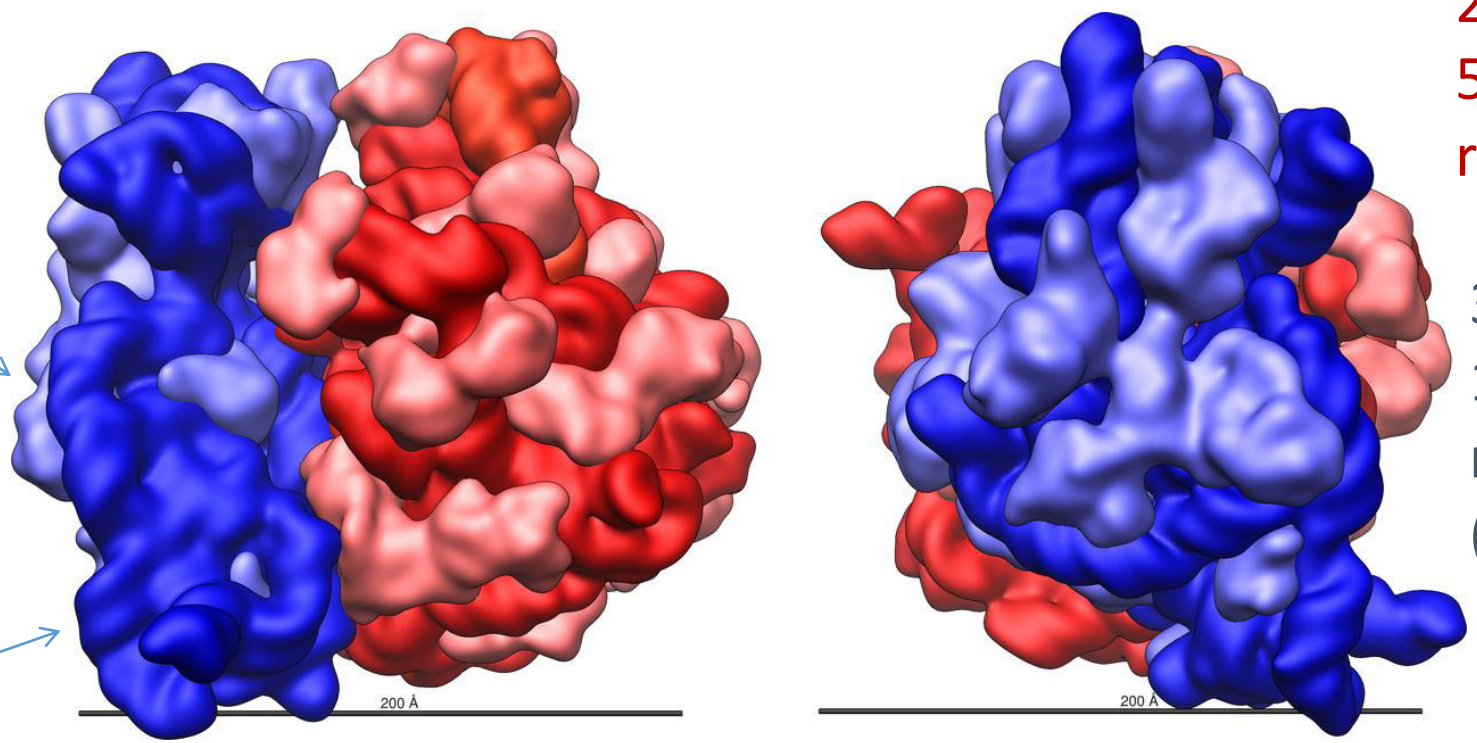
By Bensaccount

Structure of the 70S *E.coli* ribosome

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

ribosomal proteins

rRNA



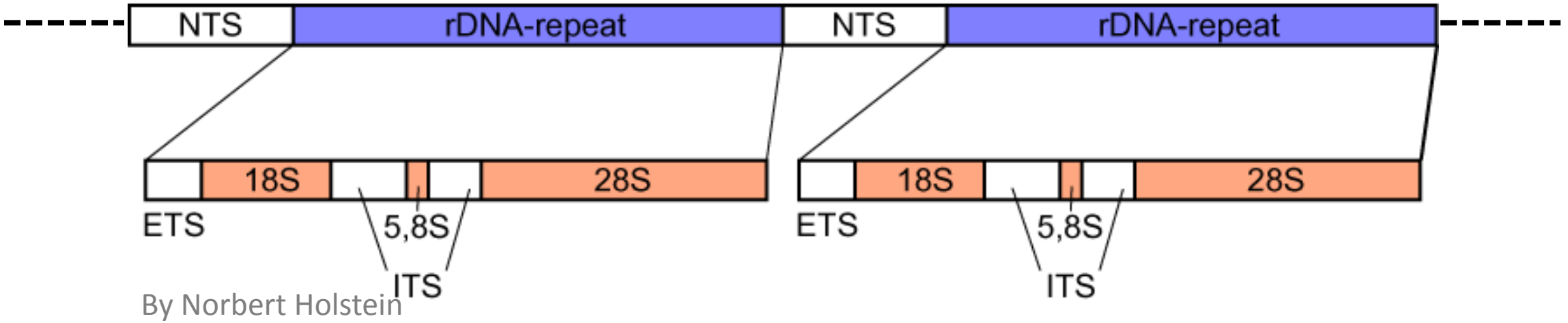
By Vossman

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



By Norbert Holstein

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

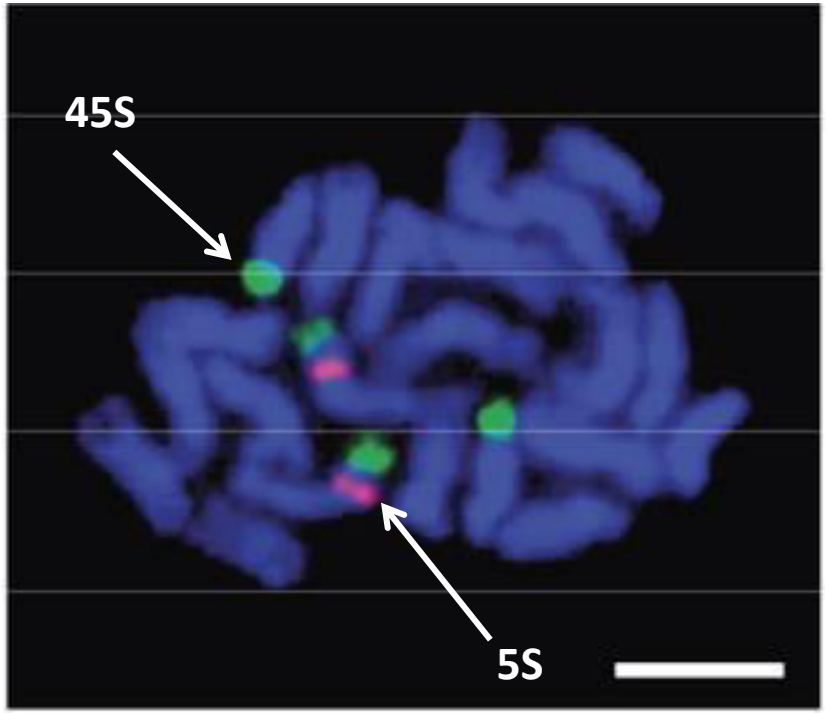
Human genome: ribosomal RNA gene clusters

- 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

Guo et al. 2012

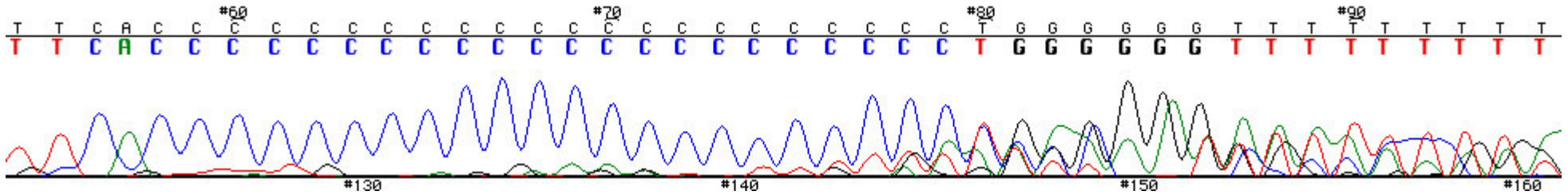
Concerted evolution

- 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

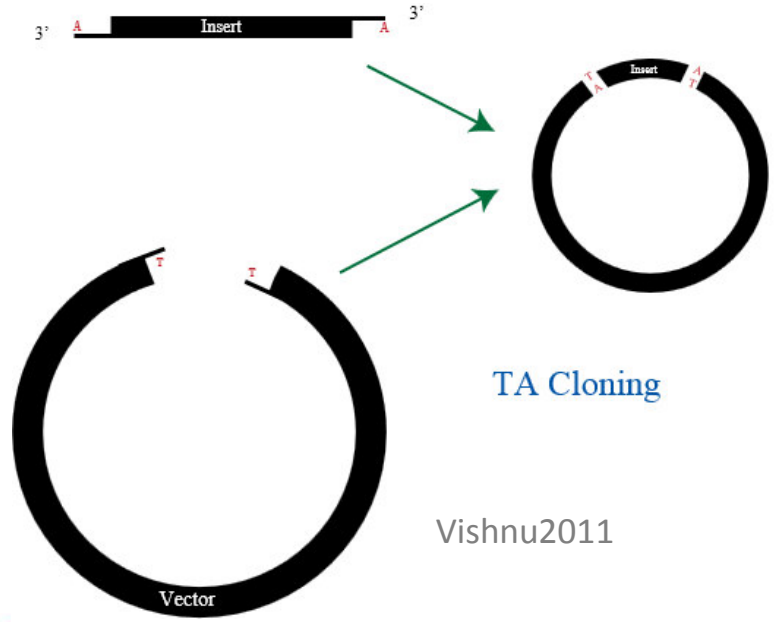
- Concerted evolution can explain the similarity among multicopy rDNA genes

- INCOMPLETE concerted evolution results in polymorphisms among copies of rDNA sequences

Consequence of incomplete concerted evolution

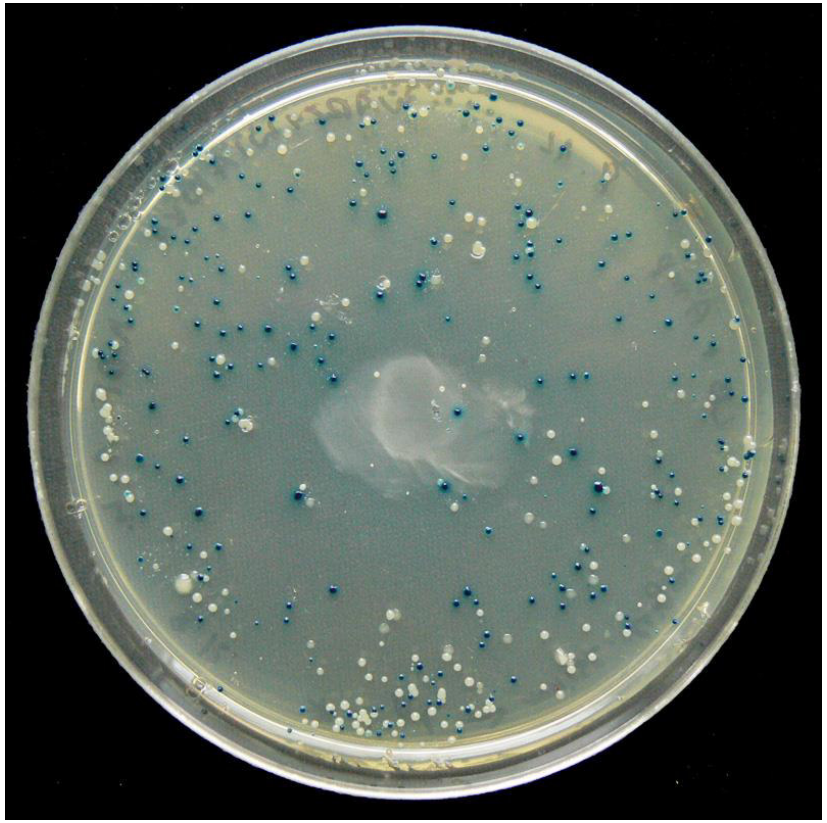


How to resolve it? –cloning the PCR product



TA Cloning

Vishnu2011



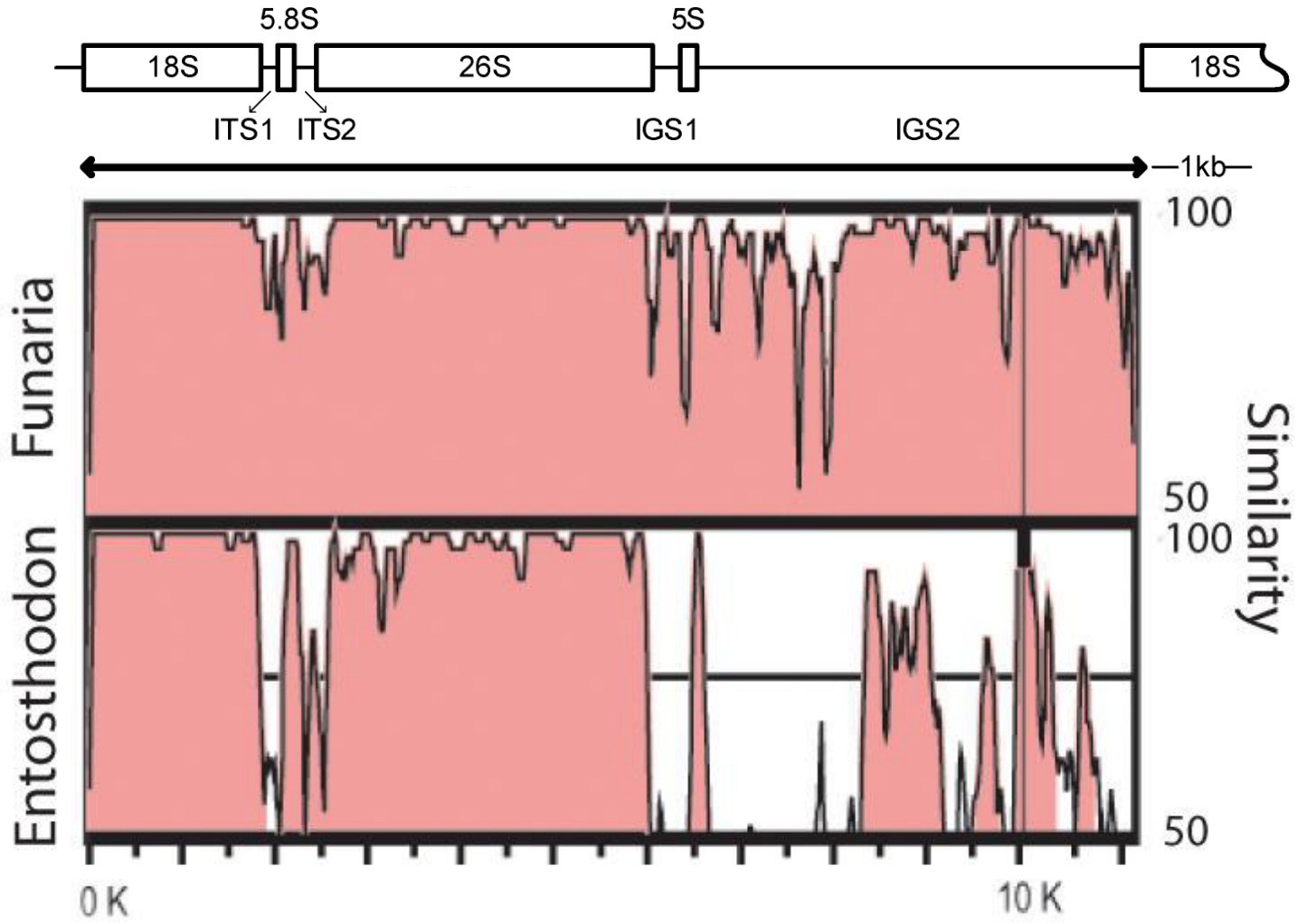
By Stefan Walkowski

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 orgnisms

Nuclear ribosomal RNA repeat unit in mosses



Liu et al. 2013

Proceedings of the National Academy of Sciences
Vol. 65, No. 3, pp. 609–616, March 1970

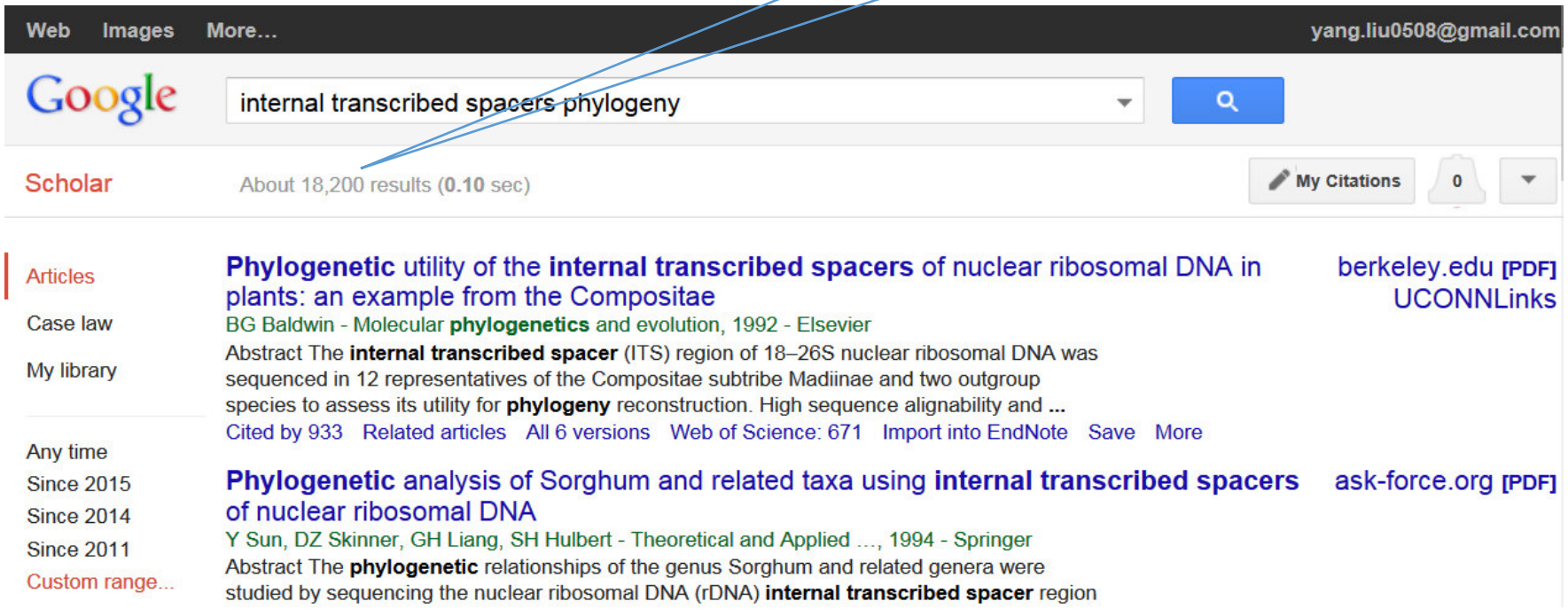
Evolution of the Transcription Unit of Ribosomal RNA*

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



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

[a]National Center for Biotec and Microbiology), Agricul [d]Centraalbureau voor Schi

Edited* by Daniel H. Janze

OPEN ACCESS Freely available online

PLoS one

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

Hui Yao[1☯], Jingyuan Song[1☯], Chang Liu[1☯], Kun Luo[1,2], Jianping Han[1], Ying Li[1], Xiaohui Pang[1], Hongxi Xu[4], Yingjie Zhu[3*], Peigen Xiao[1], Shilin Chen[1*]

1 Institute of Medicinal P Pharmacy, Hubei Univers Chengdu, Sichuan, Peopl China

OPEN ACCESS Freely available online

PLoS one

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

Shilin Chen[1*], Hui Yao[1], Jianping Han[1], Chang Liu[2], Jingyuan Song[1*], Linchun Shi[1], Yingjie Zhu[1], Xinye Ma[1], Ting Gao[1], Xiaohui Pang[1], Kun Luo[3], Ying Li[1], Xiwen Li[1], Xiaocheng Jia[1], Yulin Lin[1], Christine Leon[4]

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