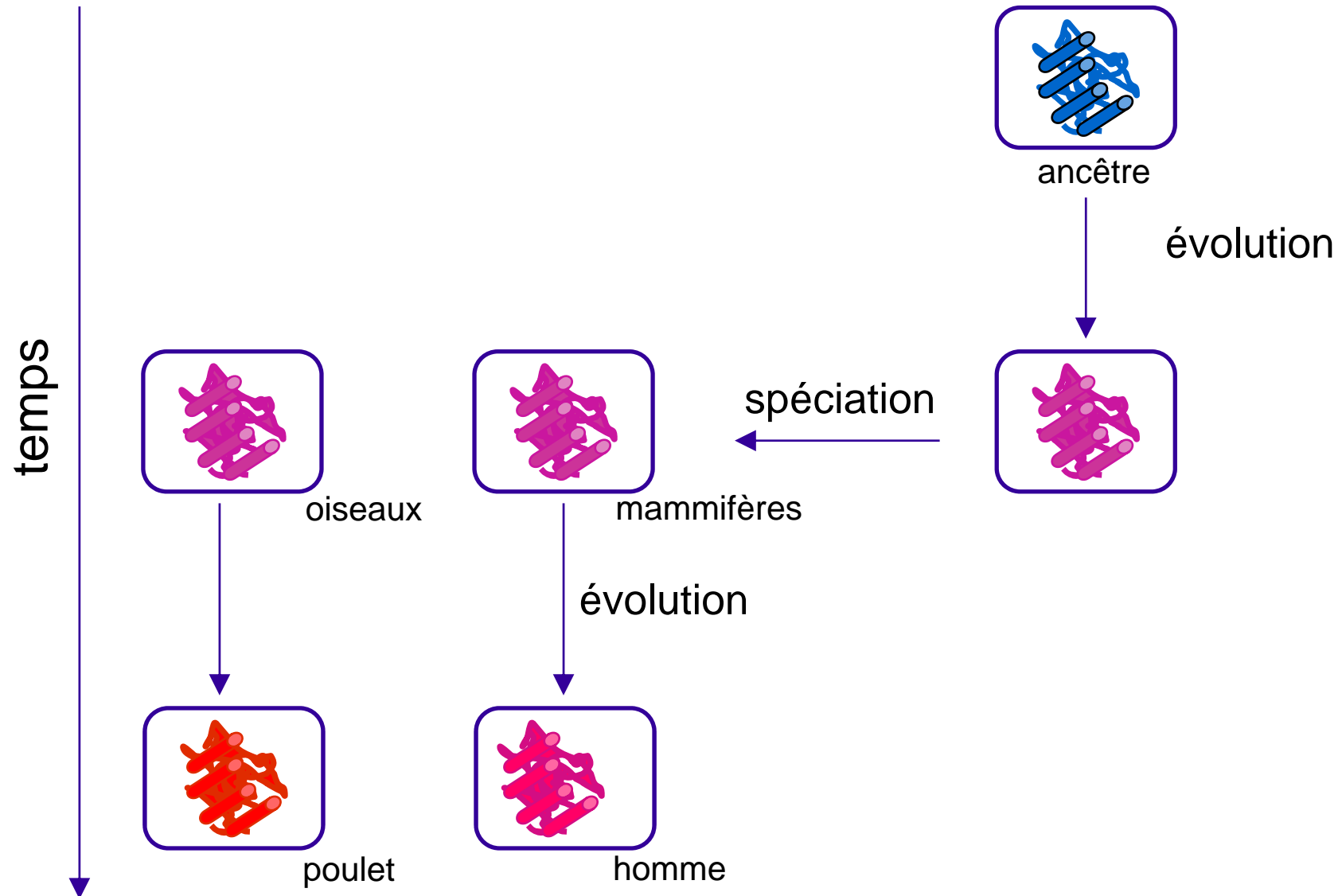


# L'Annotathon

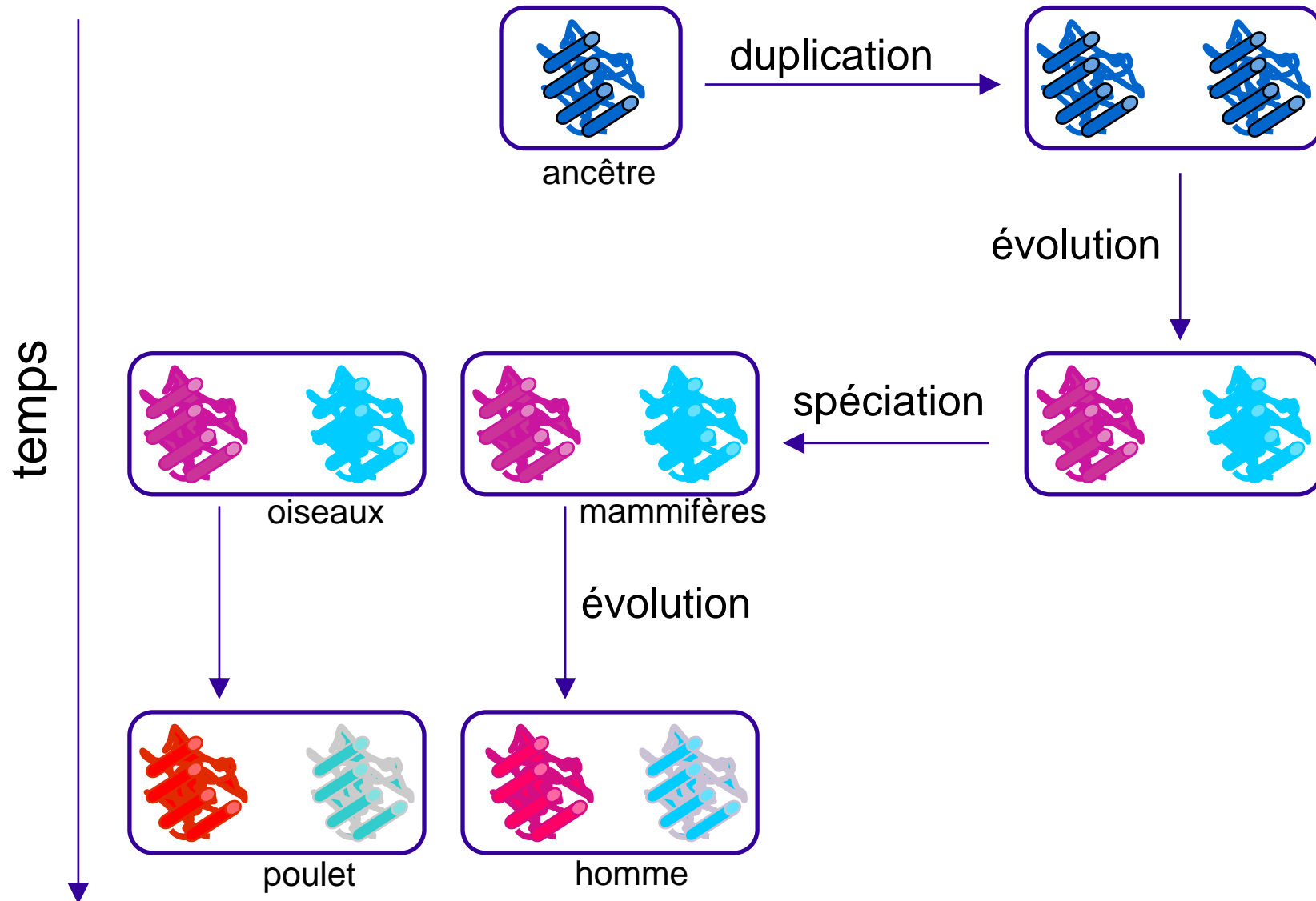
- Annotation collaborative de résultats de séquençage massif
  - Génomes complets
  - Projets de métagénomique
- <http://annotathon.univ-mrs.fr/>

# Fondements de l'annotation par homologie

# Evolution d'un gène au cours du temps

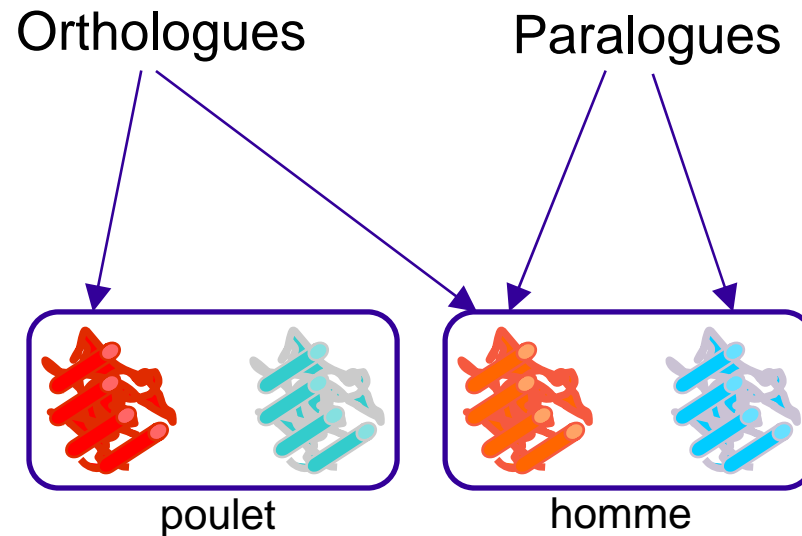


# Apparition de nouveaux gènes par duplication



# Paralogues et orthologues (Fitch, 1970)

- ★ Homologues: gènes provenant d'un ancêtre commun
- ★ Paralogues: gènes homologues issus d'un phénomène de duplication
- ★ Orthologues: gènes homologues issus de la spéciation
- ★ Transfert horizontal: par endosymbiontes, etc. Fitch a aussi introduit "xénologue" pour évoquer ce cas.

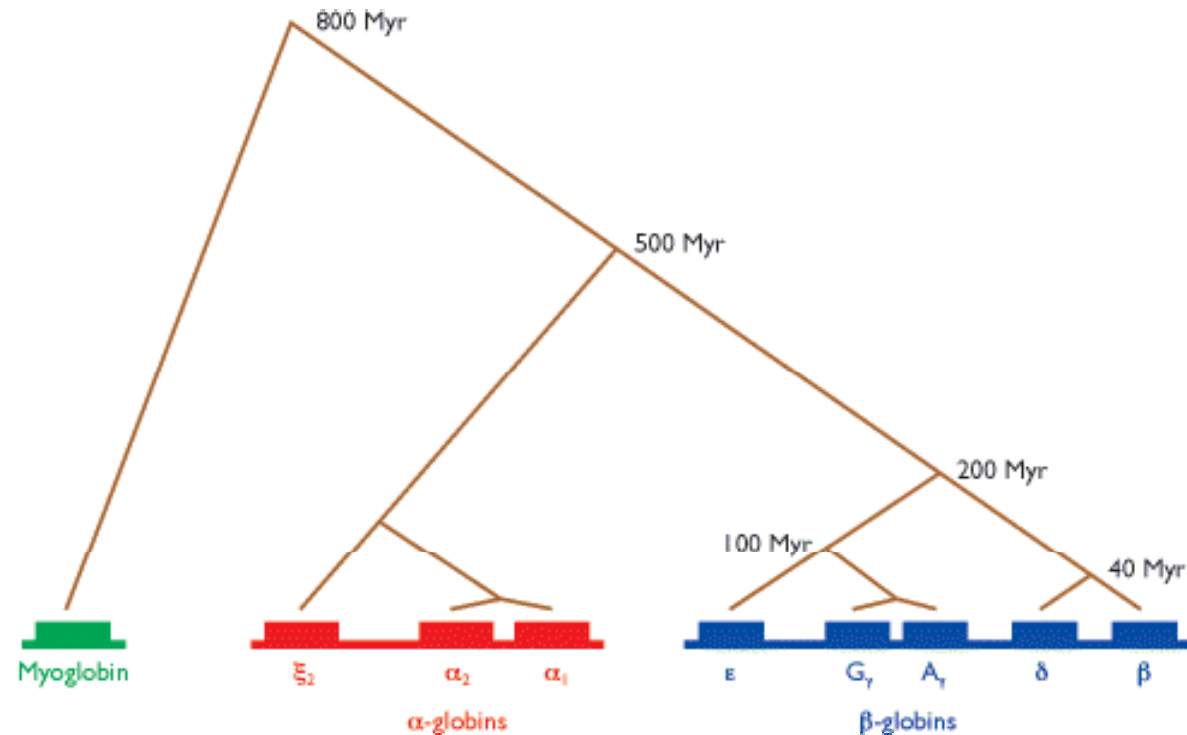


# Fonction et homologie

- ❖ Homologie n'implique pas même fonction: par exemple l'aile de l'oiseau et le bras humain n'ont pas la même fonction
- ❖ Des orthologues rapprochés (p. ex. homme/souris) ont le plus souvent la même fonction dans l'organisme.
- ❖ Des orthologues distants (p. ex. homme/mouche) ont plus rarement le même rôle *phénotypique*, mais peuvent exercer le même rôle dans une *voie* donnée.
- ❖ Les paralogues acquièrent rapidement des fonctions différentes

# Exemple: les gènes de globine humains

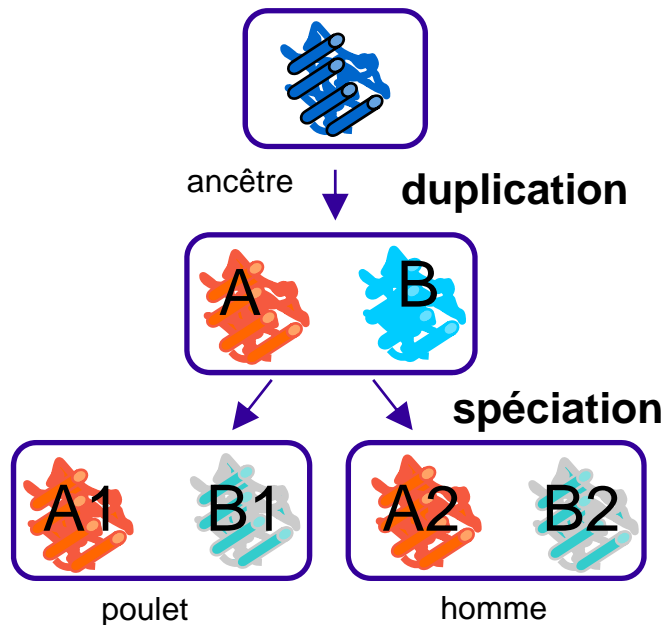
★ Tous paralogues



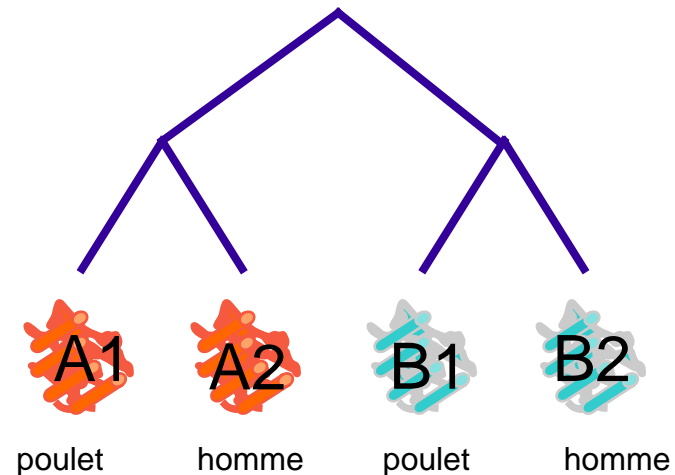
Les gènes se trouvent maintenant sur différents chromosomes: le gène de myoglobine est sur le chromosome 22, les gènes de  $\alpha$ -globines sont sur le chromosome 16 et les gènes de  $\beta$ -globine sont sur le chromosome 11.

# Arbres avec paralogues et orthologues

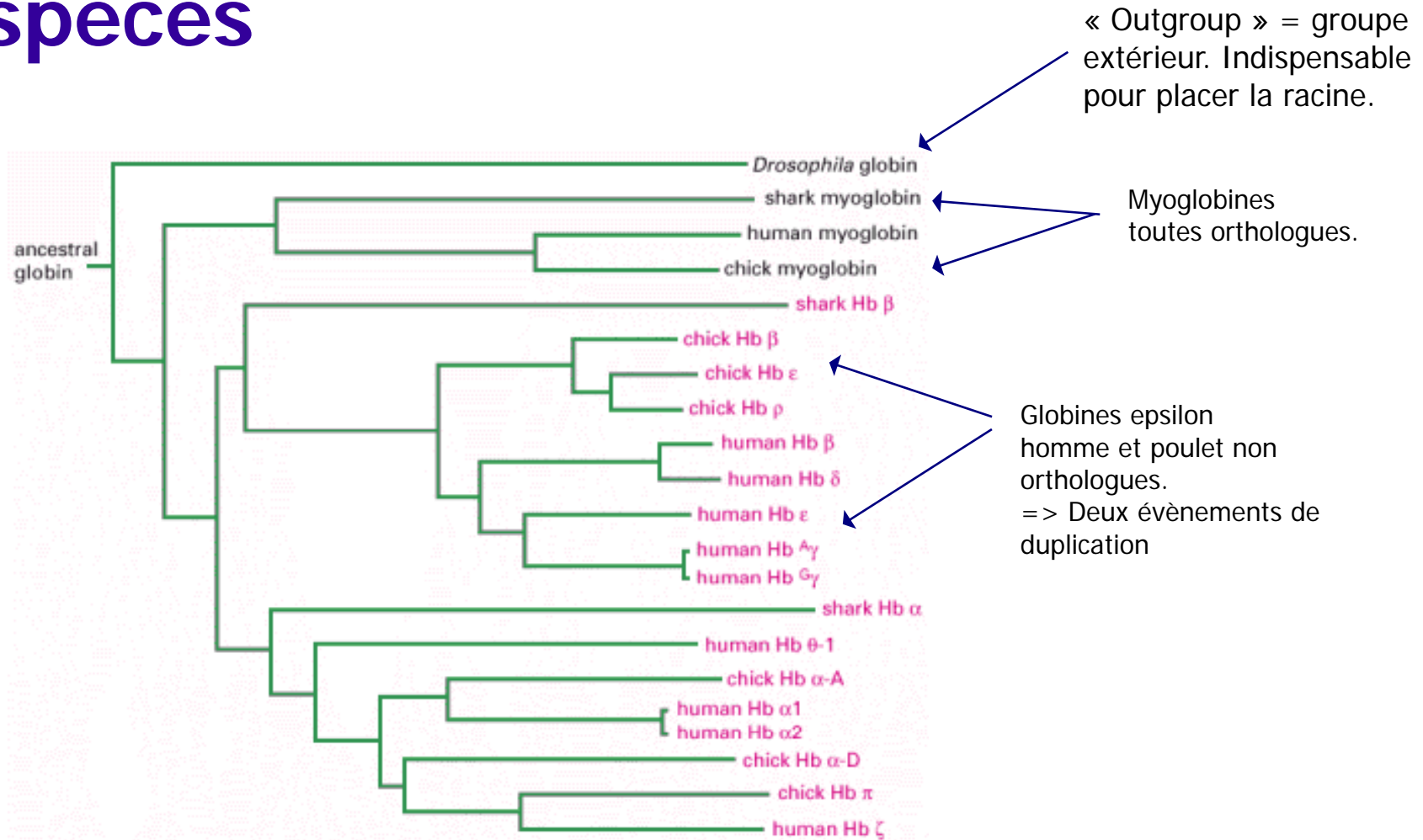
★ Admettons le schéma évolutif suivant (à gauche) ayant produit deux gènes paralogues présents chez tous les vertébrés.



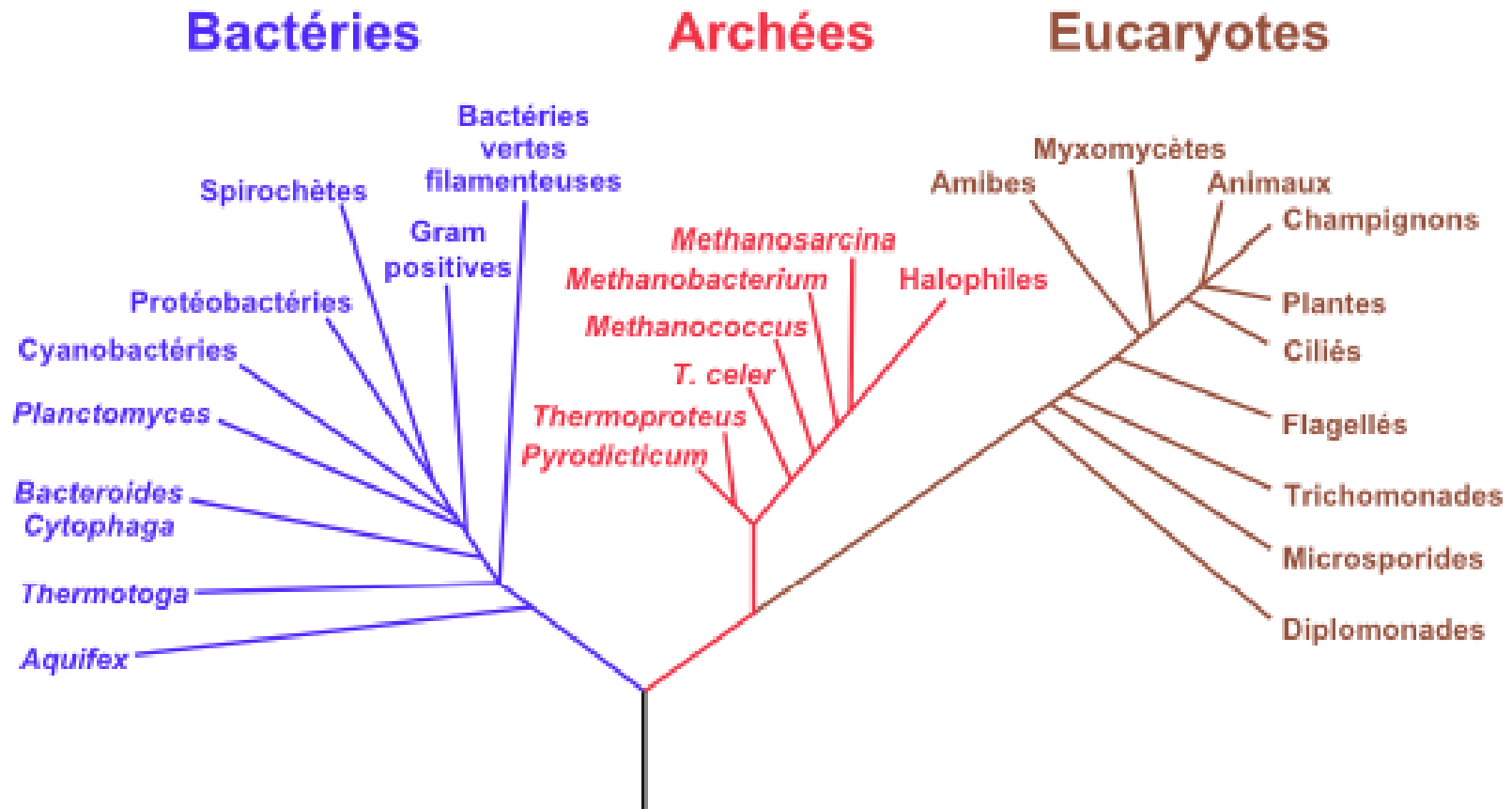
★ Etant donné que la duplication (ayant produit les paralogues) a eu lieu AVANT la spéciation (ayant produit les orthologues), les orthologues devraient être plus proches entre eux que les paralogues. L'arbre devrait donc ressembler à ceci:

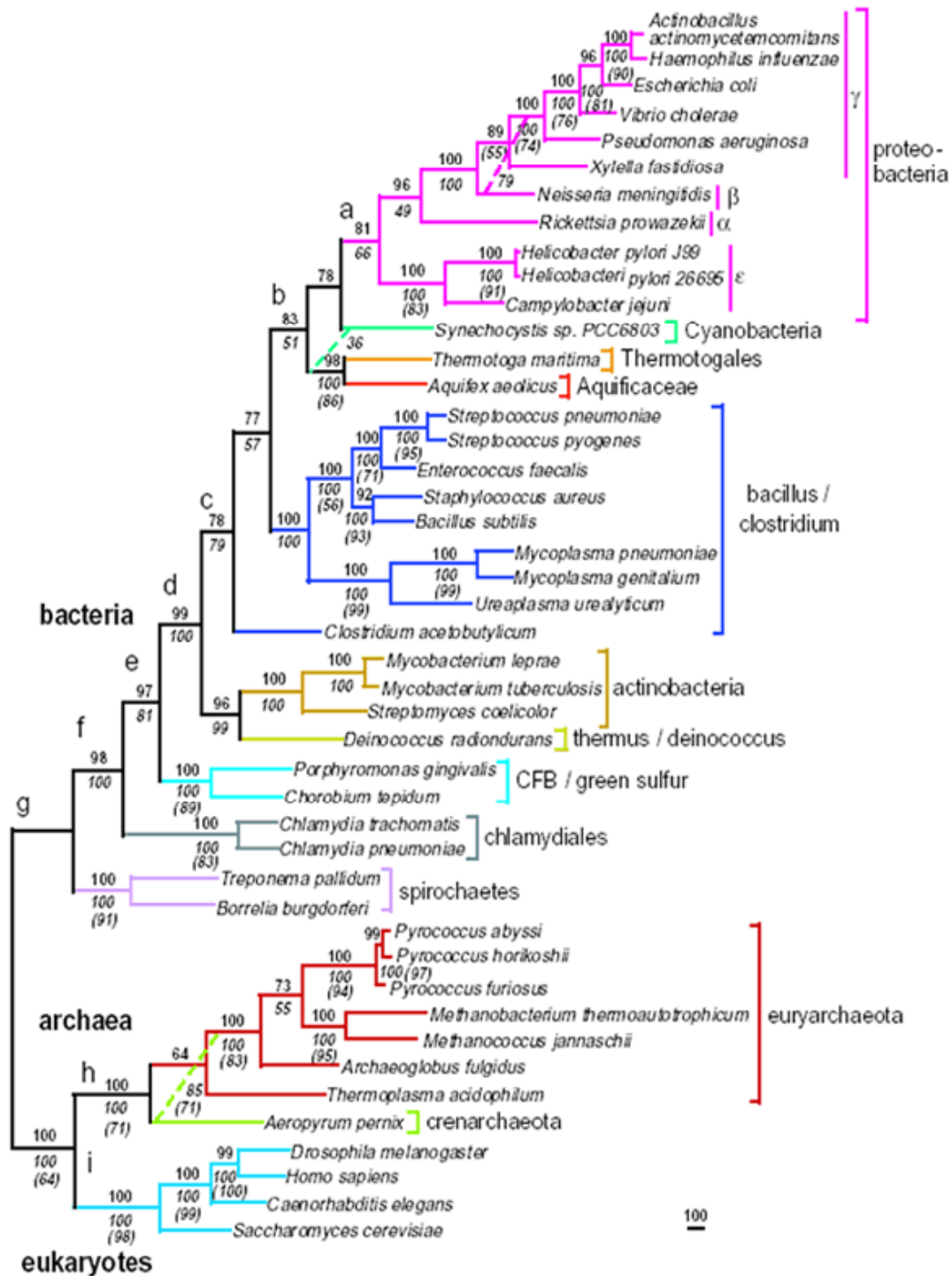


# Les gènes de globine chez # espèces

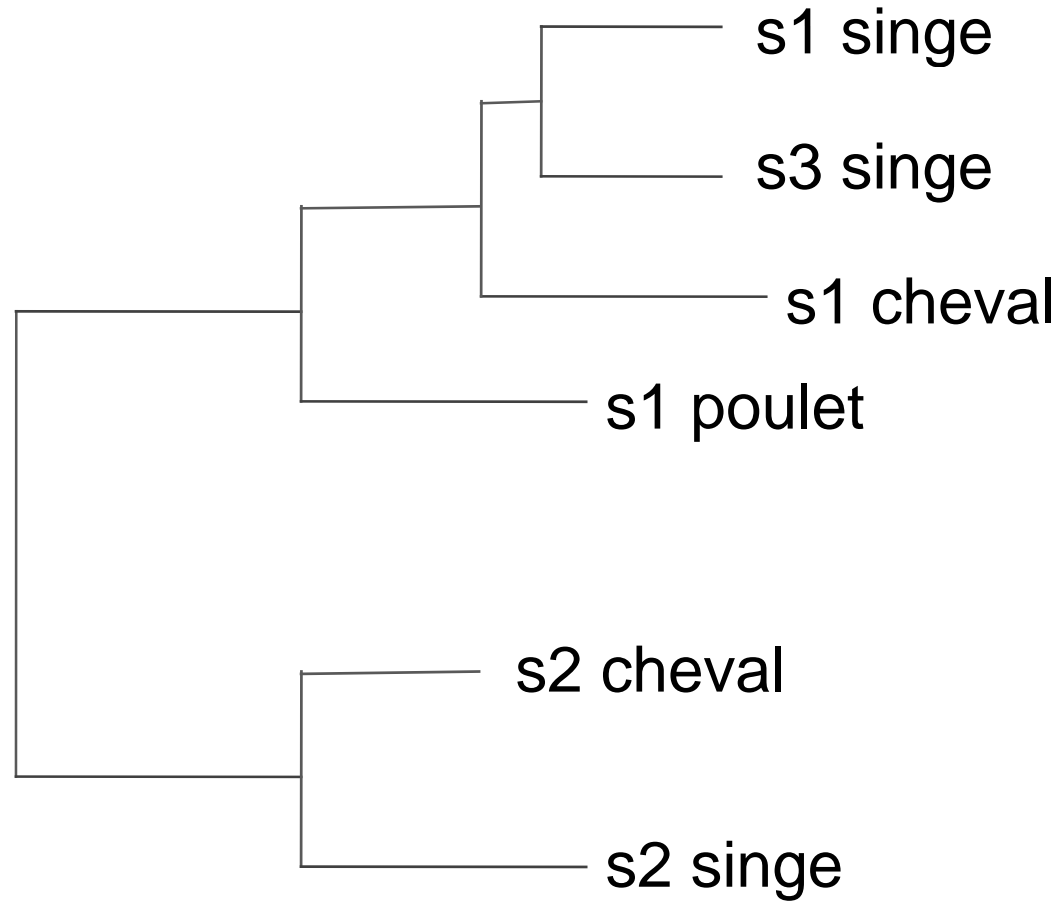


# Importance de la phylogénie

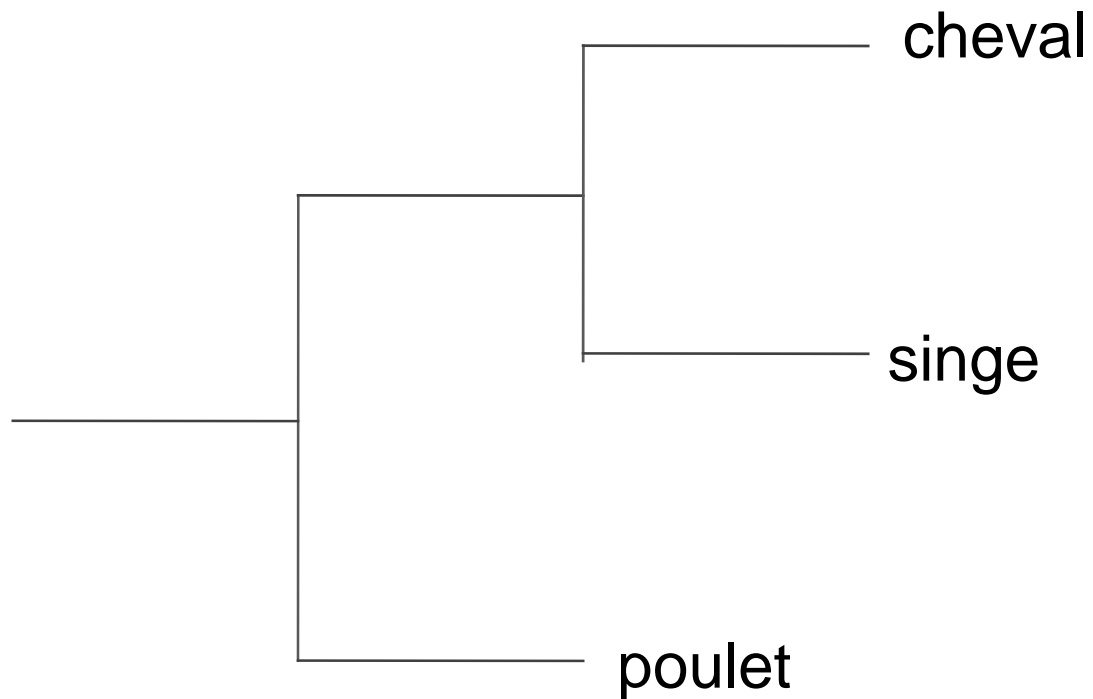




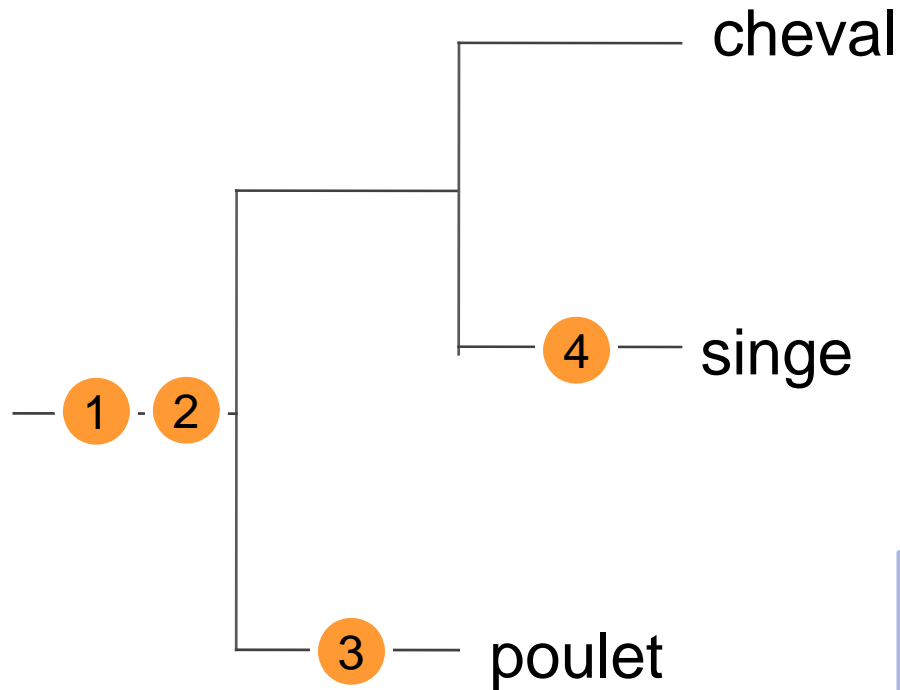
# Exercice



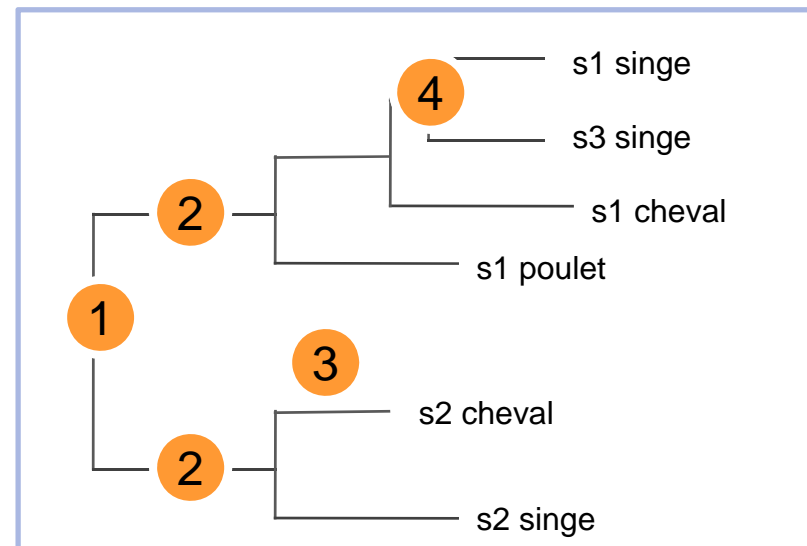
# D'abord faire l'arbre des espèces



# Puis placer les évènements



- 1 Apparition du gène ancestral P+C+S
- 2 Duplication chez ancêtre mamm+oiseaux
- 3 Perte chez le poulet
- 4 Duplication « récente » chez le singe



# 2004: Environmental Genome Shotgun Sequencing of the Sargasso Sea

- ✦ 1Gbase
- ✦ 1800 espèces bactériennes
- ✦ 148 nouvelles espèces
- ✦ 1,2 millions de nouveaux gènes

# 2007



PLOS BIOLOGY

a peer-reviewed open-access journal published by the Public Library of Science

MY PROFILE

Search PLoS

Home Browse Articles About For Readers For Authors and Reviewers

RESEARCH ARTICLE

OPEN ACCESS

## The *Sorcerer II* Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific

Douglas B. Rusch<sup>1\*</sup>, Aaron L. Halpern<sup>1</sup>, Granger Sutton<sup>1</sup>, Karla B. Heidelberg<sup>1,2</sup>, Shannon Williamson<sup>1</sup>, Shibu Yooseph<sup>1</sup>, Dongying Wu<sup>1,3</sup>, Jonathan A. Eisen<sup>1,3</sup>, Jeff M. Hoffman<sup>1</sup>, Karin Remington<sup>1,4</sup>, Karen Beeson<sup>1</sup>, Bao Tran<sup>1</sup>, Hamilton Smith<sup>1</sup>, Holly Baden-Tillson<sup>1</sup>, Clare Stewart<sup>1</sup>, Joyce Thorpe<sup>1</sup>, Jason Freeman<sup>1</sup>, Cynthia Andrews-Pfannkoch<sup>1</sup>, Joseph E. Venter<sup>1</sup>, Kelvin Li<sup>1</sup>, Saul Kravitz<sup>1</sup>, John F. Heidelberg<sup>1,2</sup>, Terry Utterback<sup>1</sup>, Yu-Hui Rogers<sup>1</sup>, Luisa I. Falcón<sup>5</sup>, Valeria Souza<sup>5</sup>, Germán Bonilla-Rosso<sup>5</sup>, Luis E. Eguarte<sup>5</sup>, David M. Karl<sup>6</sup>, Shubha Sathyendranath<sup>7</sup>, Trevor Platt<sup>7</sup>, Eldredge Bermingham<sup>8</sup>, Victor Gallardo<sup>9</sup>, Giselle Tamayo-Castillo<sup>10</sup>, Michael R. Ferrari<sup>11</sup>, Robert L. Strausberg<sup>1</sup>, Kenneth Nealson<sup>1,12</sup>, Robert Friedman<sup>1</sup>, Marvin Frazier<sup>1</sup>, J. Craig Venter<sup>1</sup>

**1** J. Craig Venter Institute, Rockville, Maryland, United States of America, **2** Department of Biological Sciences, University of Southern California, Avalon, California, United States of America, **3** Genome Center, University of California Davis, Davis, California, United States of America, **4** Your Genome, Your World, Rockville, Maryland, United States of America, **5** Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico, **6** Department of Oceanography, University of Hawaii, Honolulu, Hawaii, United States of America, **7** Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada, **8** Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama, **9** Departamento de Oceanografía, Universidad de Concepción, Concepción, Chile, **10** Escuela de Química, Universidad de Costa Rica, San Pedro, Costa Rica, **11** Department of Environmental Sciences, Rutgers University, New Brunswick, New Jersey, United States of America, **12** Department of Earth Sciences, University of Southern California, Los Angeles, California, United States of America

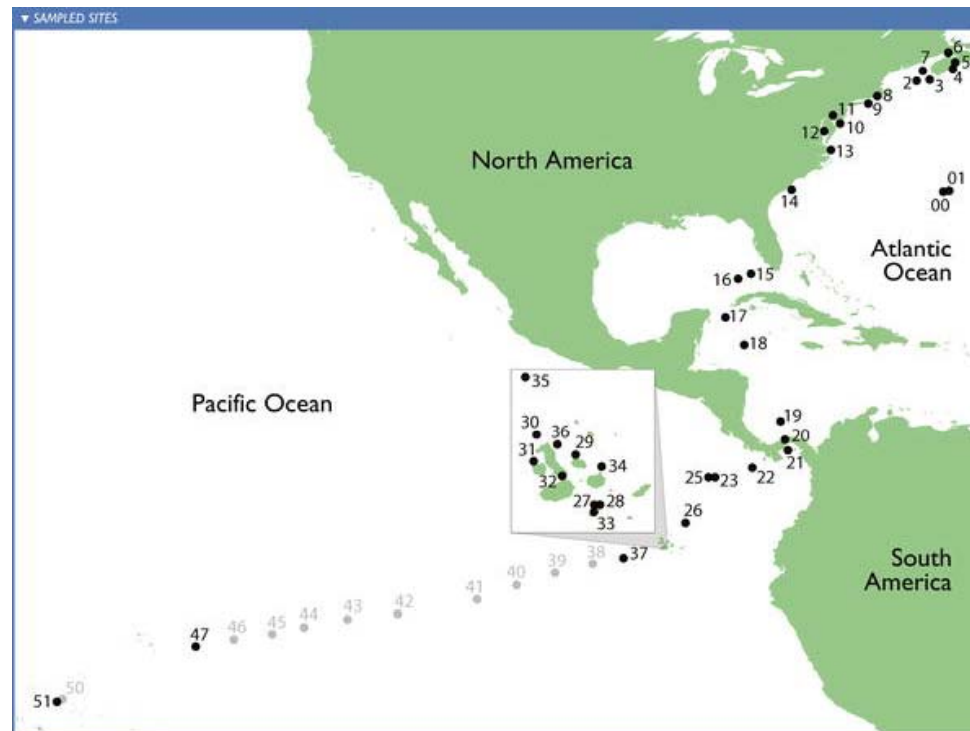


This article is part of the Oceanic Metagenomics collection.

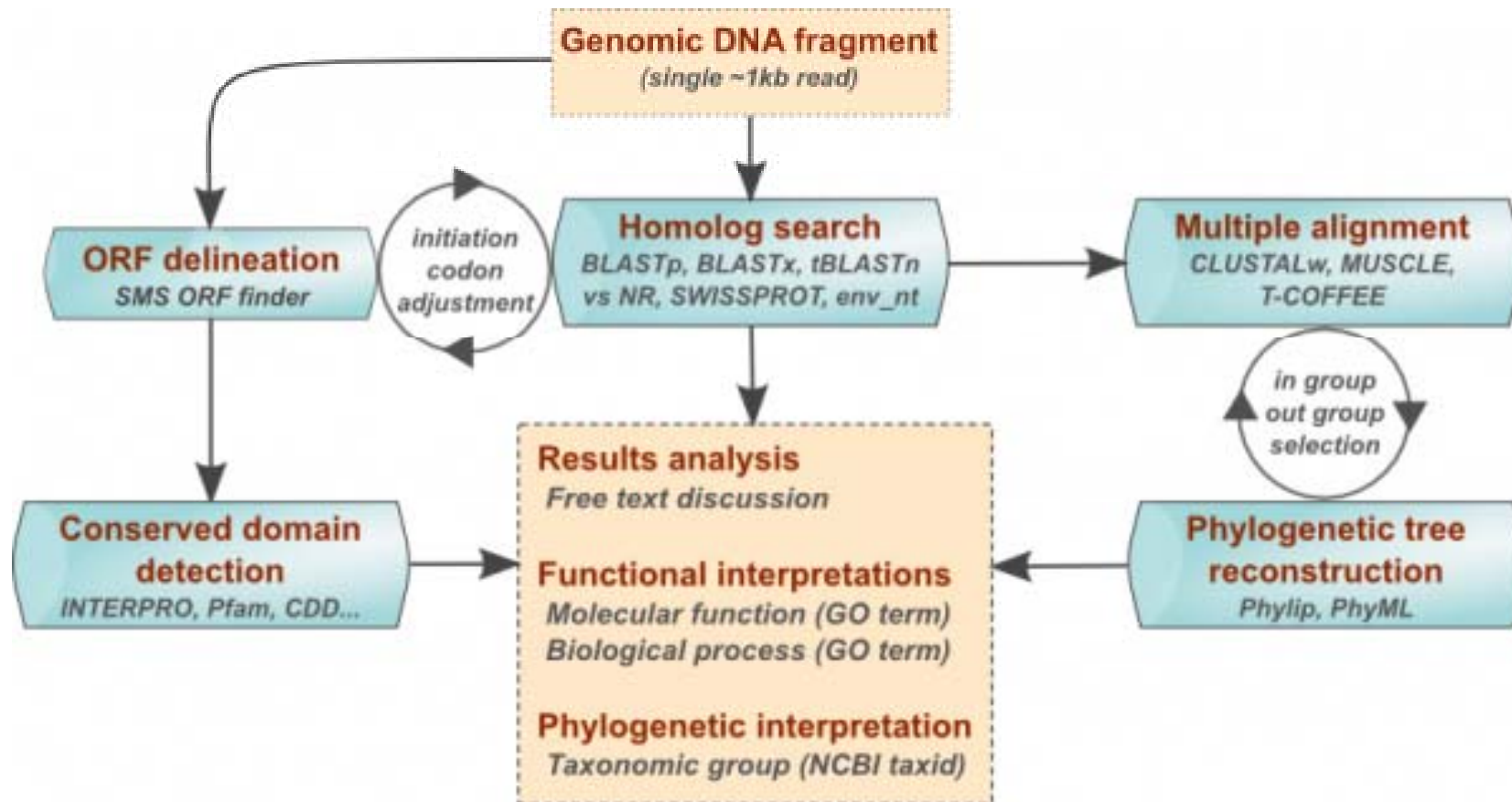
The world's oceans contain a complex mixture of micro-organisms that are for the most part, uncharacterized both genetically and biochemically. We report here a metagenomic study of the marine planktonic microbiota in which surface (mostly marine) water samples were analyzed as part of the *Sorcerer II* Global Ocean Sampling expedition. These samples, collected across a

# 2007: Global Ocean Sampling (GOS)

- ◆ 41 échantillons sur un trajet de 8000km
  - ★ 7.7 M de séquences
  - ★ 6,25 Gb



# Annotation workflow



# Wiki export

- ✦ Les meilleures annotations sont exportées vers le Wiki « Metagenes »

Indexé par Google

Interrogeable par auteur,  
lieu, fonction...

