

# Phyl-ARIANE

## Phylogénomique : Algorithmes et Représentations Intégrés pour l'ANalyse de l'Evolution du vivant

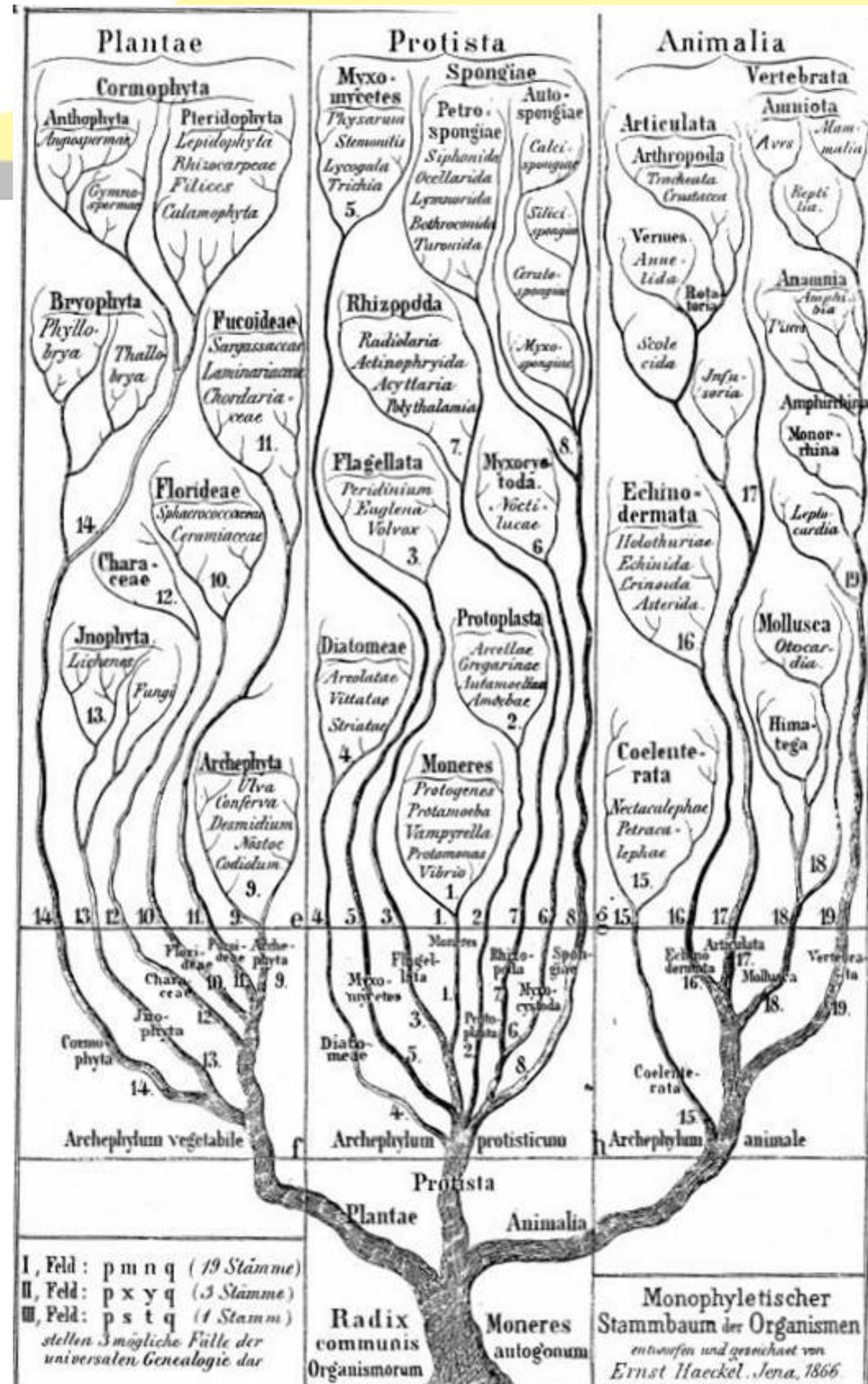
Projet 2008-2001, porté par Vincent Berry (LIRMM, Univ Montpellier 2)

Eric Tannier, INRIA, LBBE, Univ Lyon 1

Colloque ANR STIC 2012, Lyon

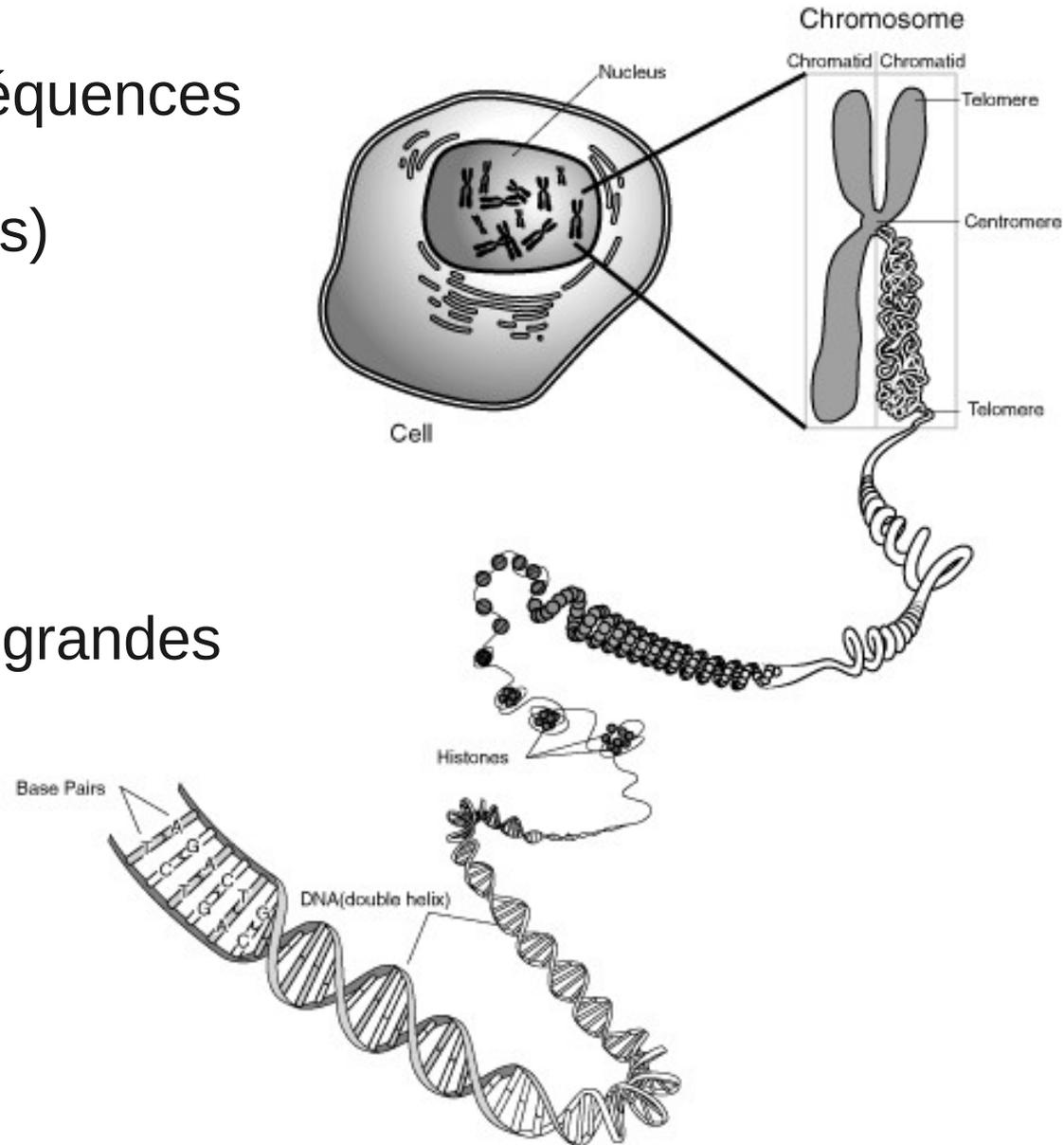


# Phylogénie : histoire des organismes vivants



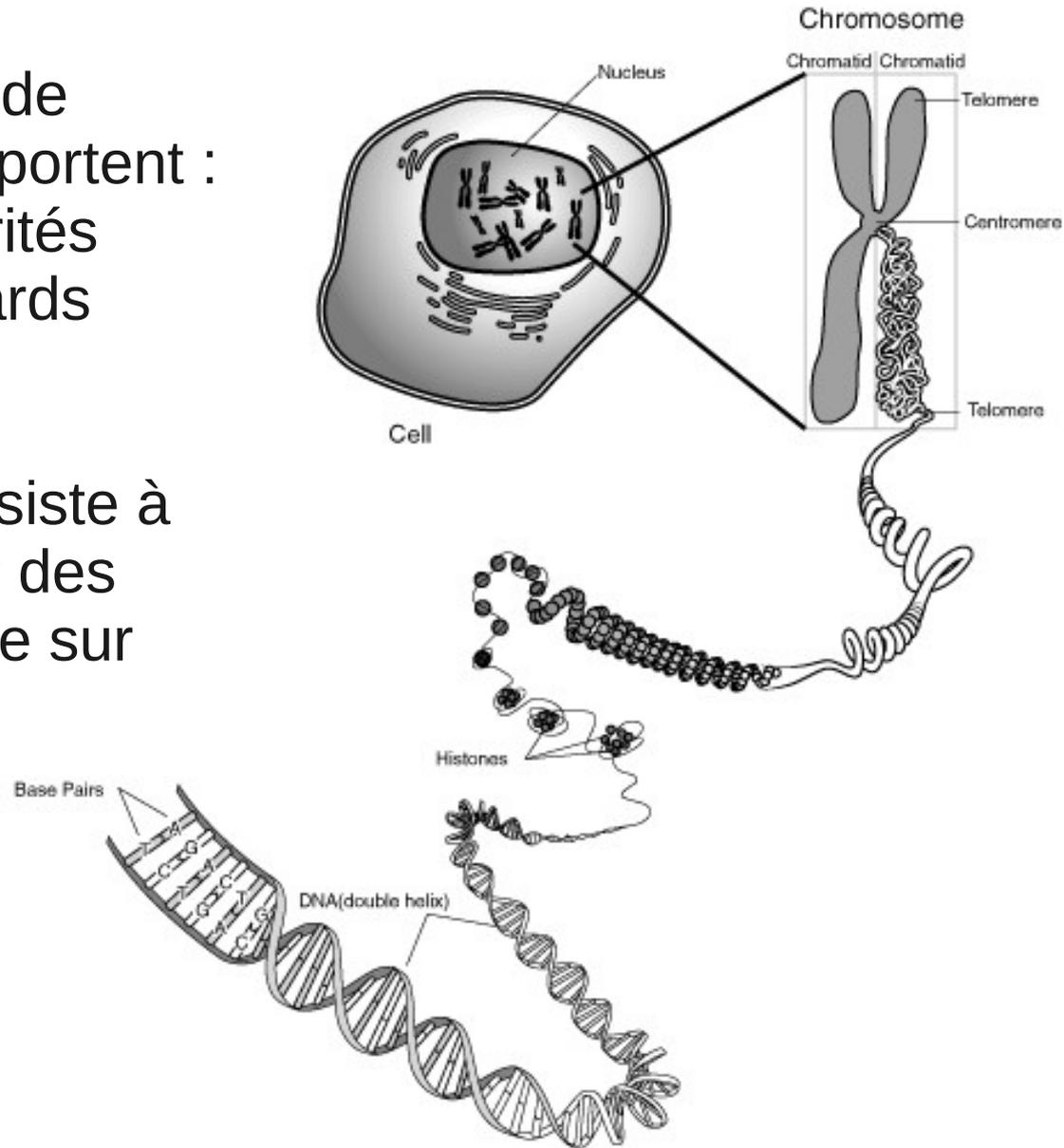
Les génomes sont de longues séquences qui évoluent par réplication et diversification, selon (entre autres)

- des spéciations
- des mutations ponctuelles
- des duplications plus ou moins grandes
- des pertes
- des transferts



Les **génomés** portent les traces de l'histoire des organismes qui les portent : malgré les mutations, des similarités subsistent parfois plusieurs milliards d'années.

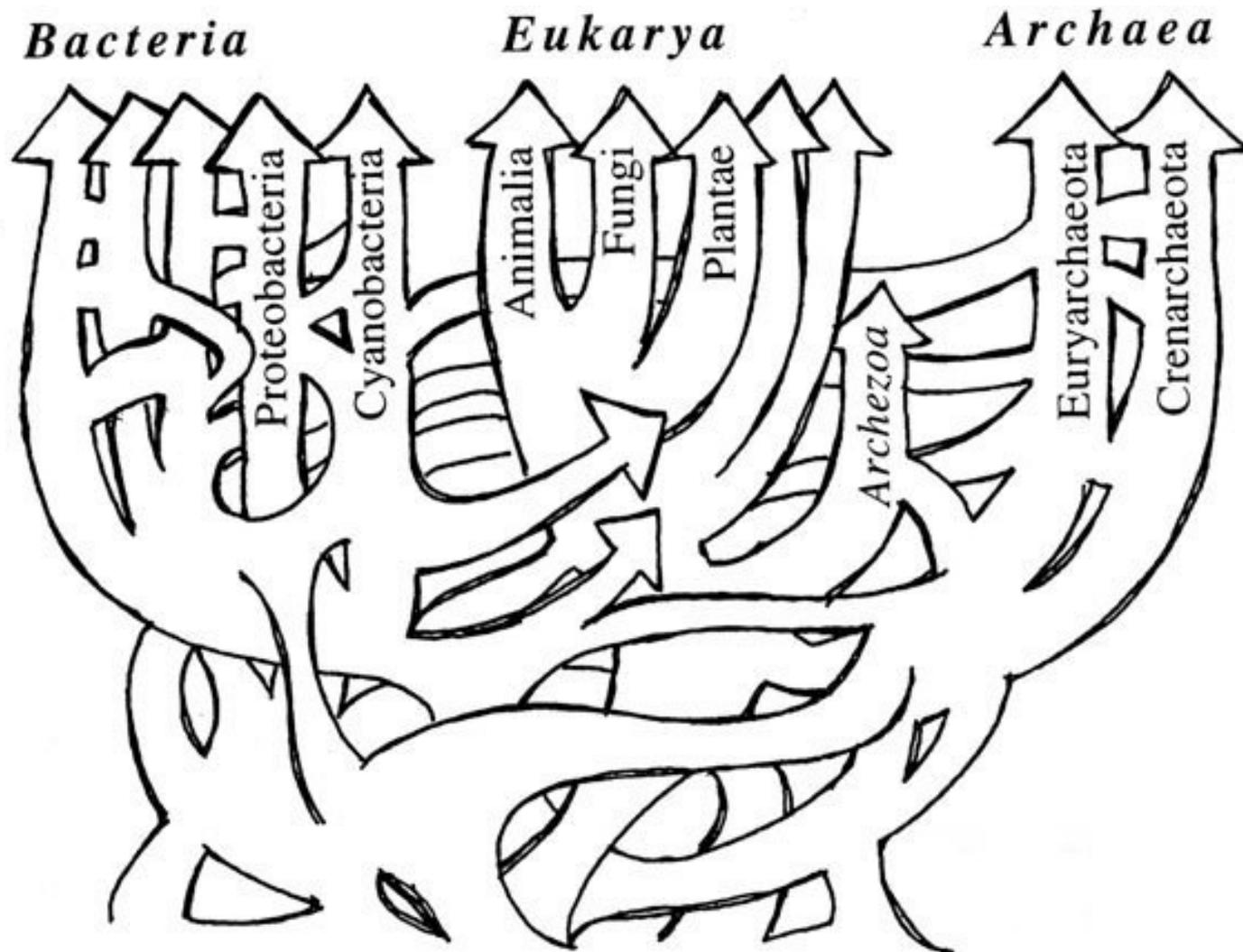
La **phylogénie moléculaire** consiste à retrouver ces traces pour en tirer des informations sur l'histoire de la vie sur Terre.







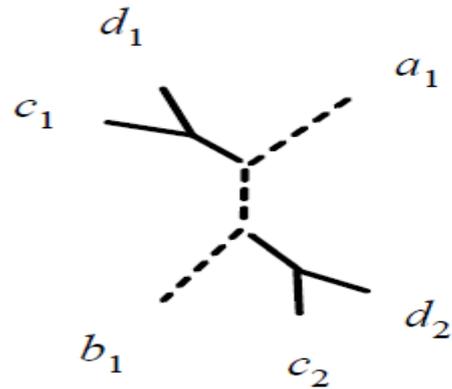
Peut-on même représenter l'histoire de la vie avec un arbre ?  
(Doolittle, *Science*, 1999)



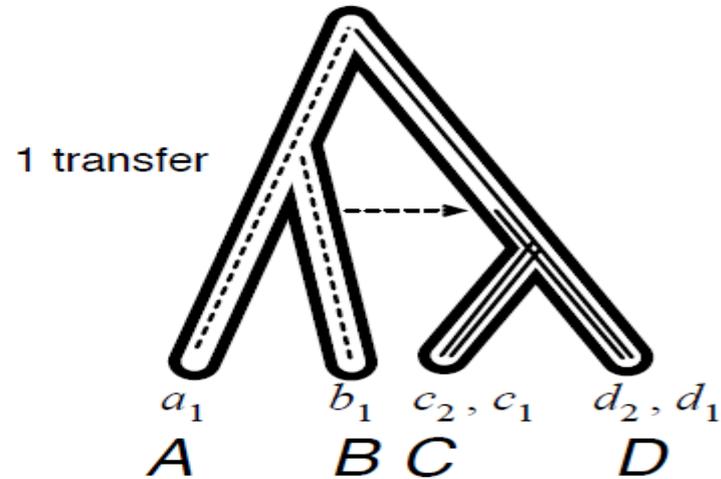
Le projet Phylariane  
décrit l'histoire du vivant avec tous les gènes, en modélisant  
et en tirant parti de leurs histoires complexes  
(origine, duplication, transfert, perte)



**A** gene phylogeny

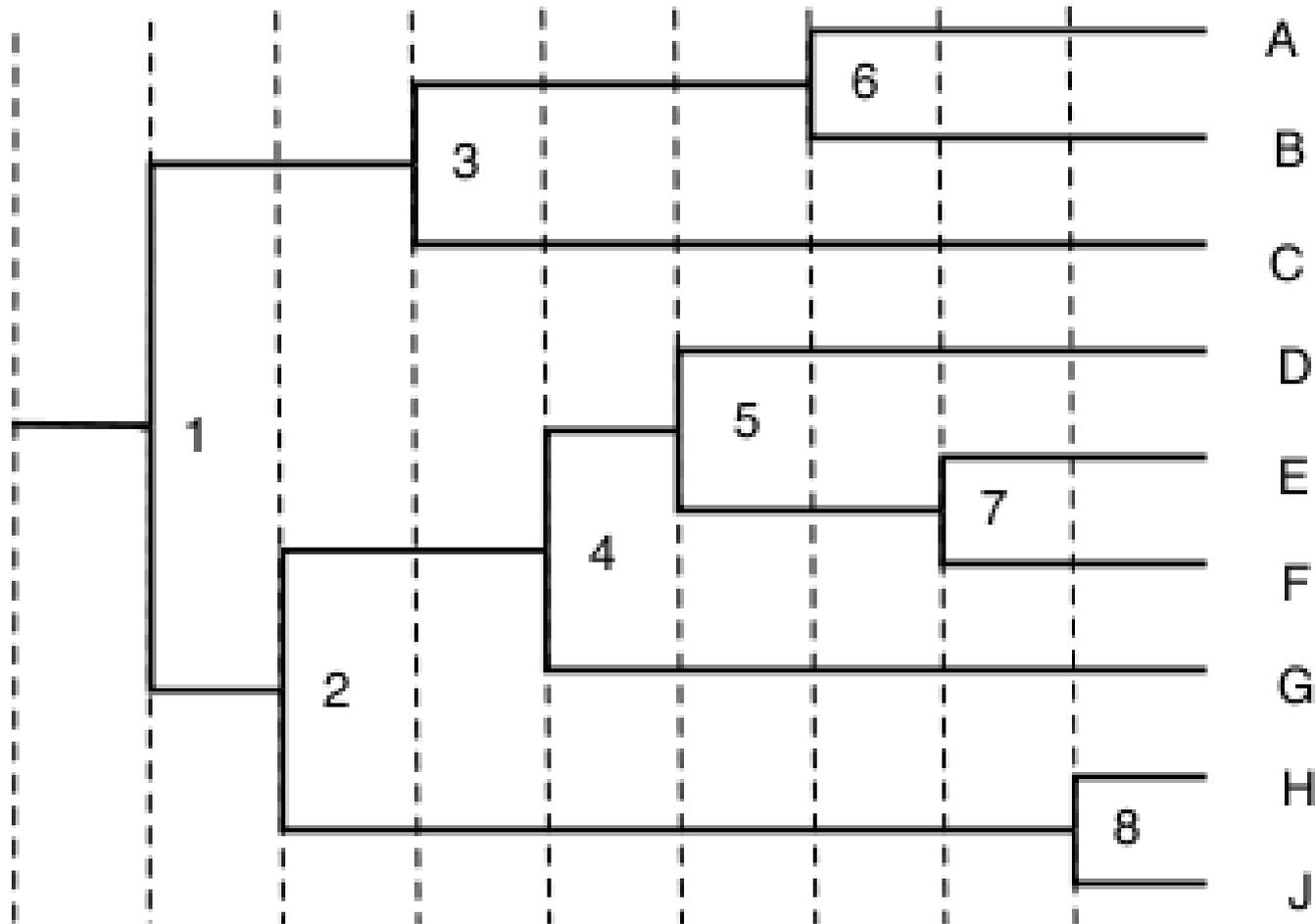


**B** reconciliation



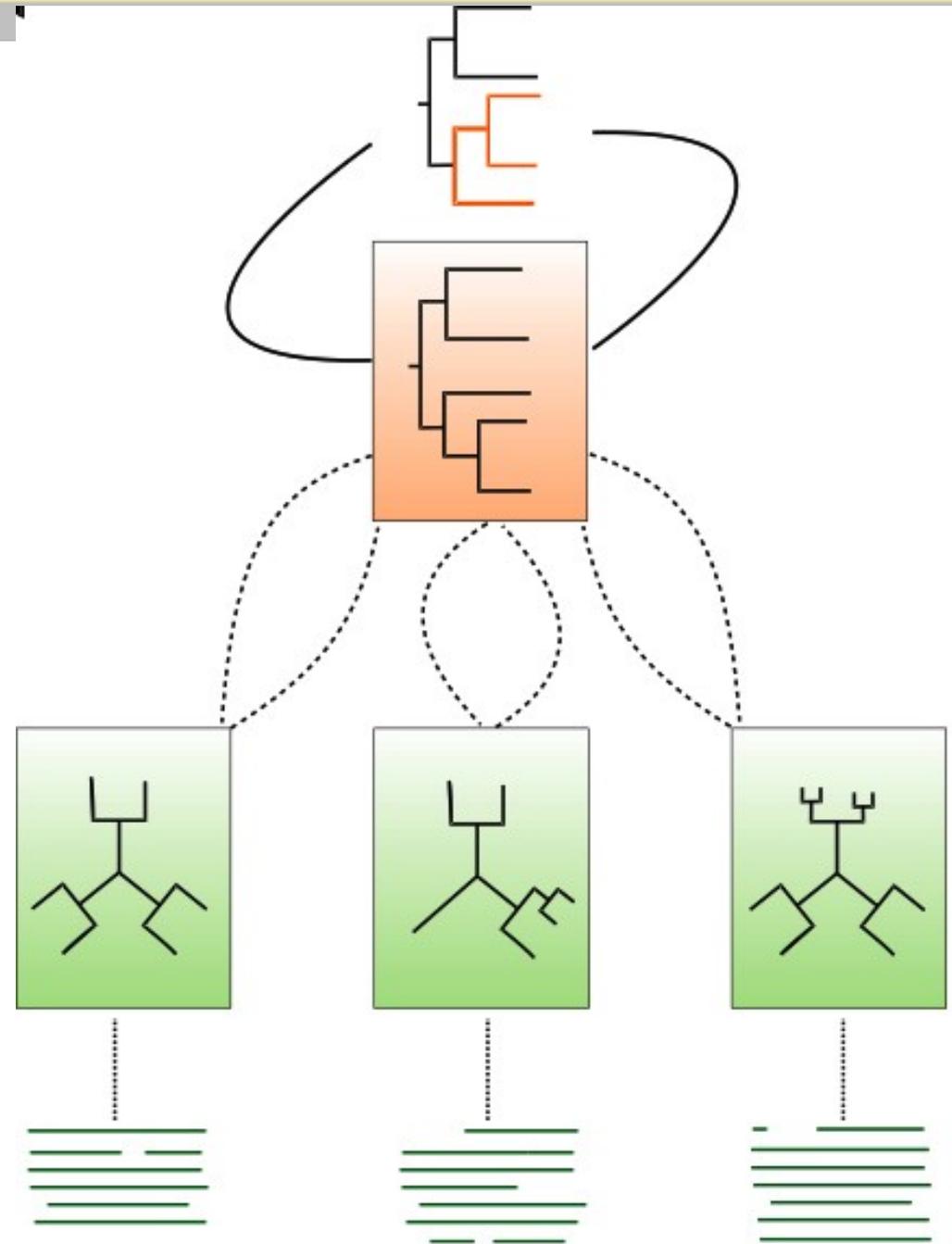
Plongement des arbres dans l'histoire des espèces : réconciliation  
 Nouvel algorithme de complexité  $O(n^2)$ ,  
 (Doyon et al, RECOMB Comparative Genomics, 2010)

Un arbre d'espèce contient la chronologie des spéciations, car les transferts contiennent une information sur la contemporanéité des espèces



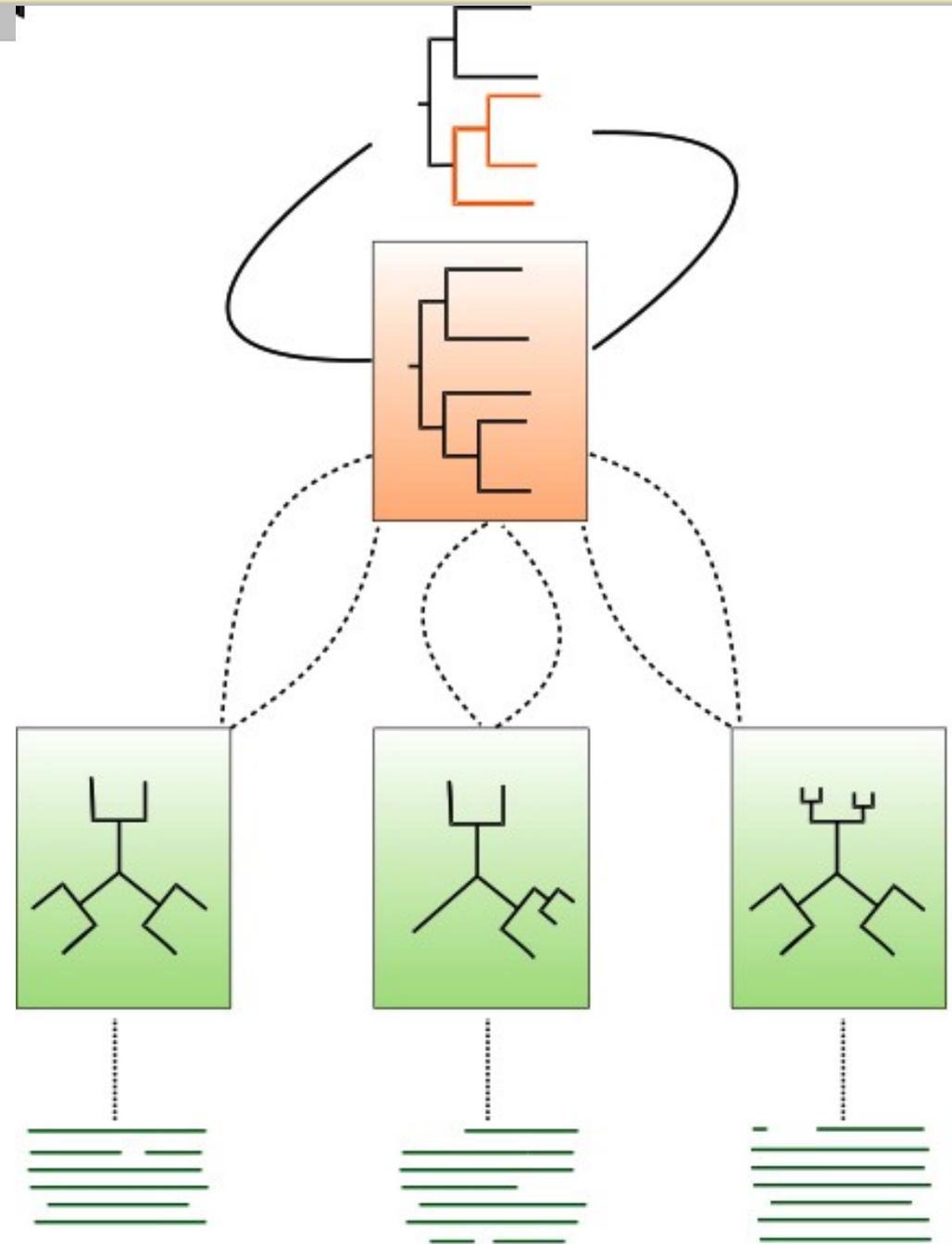
Optimisation simultanée :  
On cherche la combinaison  
- arbre d'espèces daté  
- arbres de gènes  
qui ait la meilleure vraisemblance

Utilisation de clusters locaux ou  
nationaux (Jade)



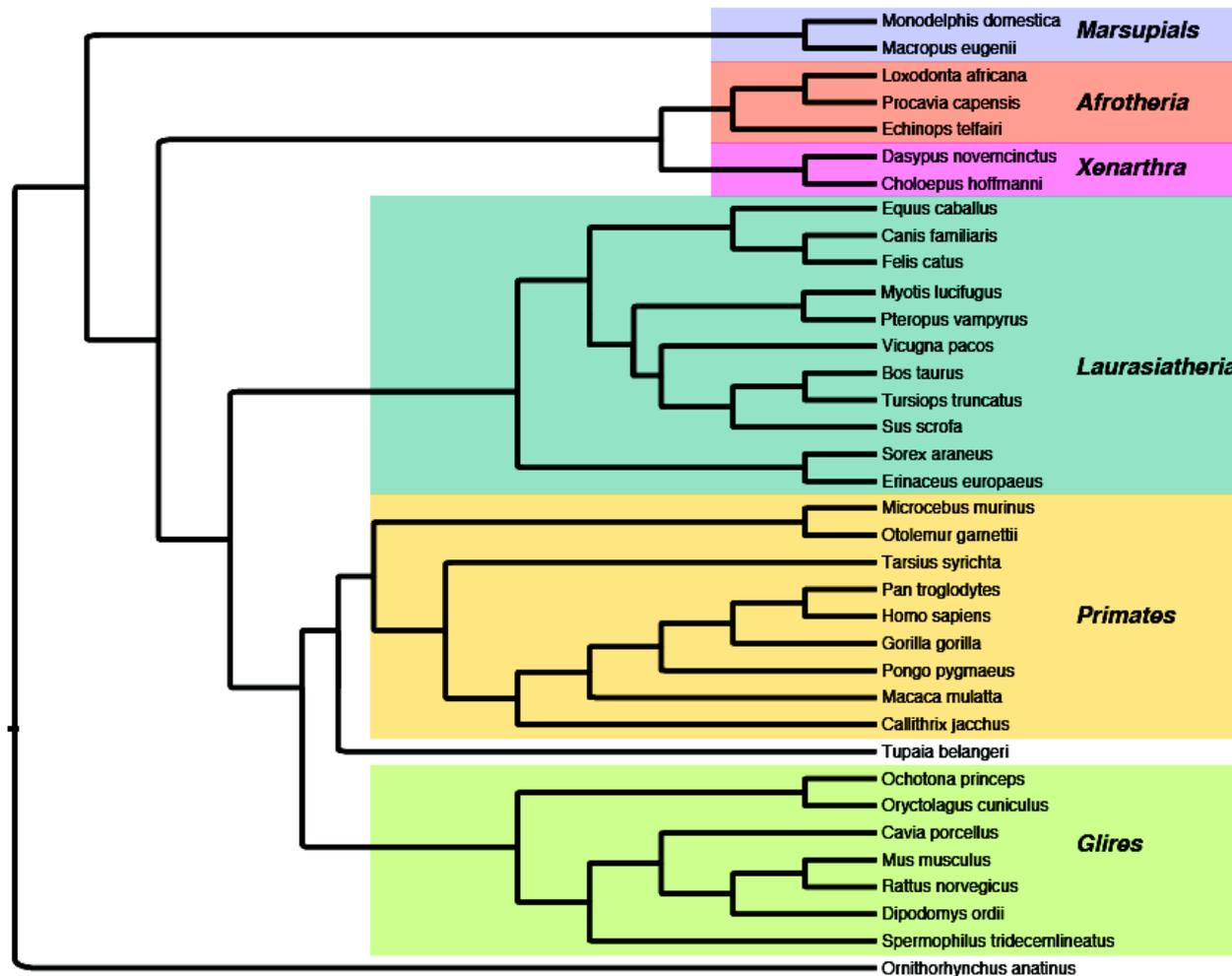
Prise en compte des erreurs et de l'incertitude liées aux données :

- support statistique des branches
- amélioration des arbres en fonction de la phylogénie des espèces

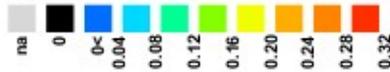


# Résultats

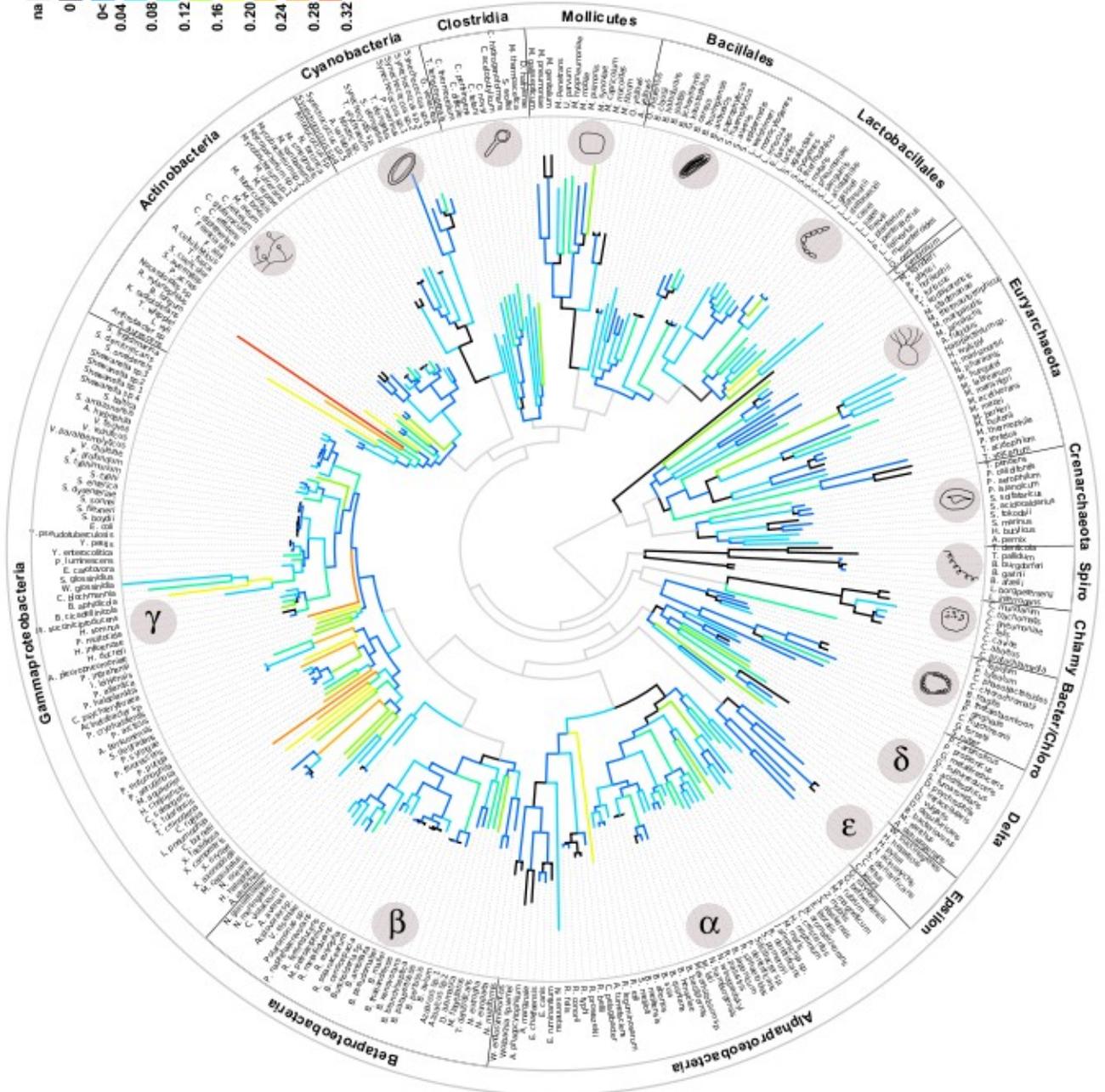
## Une phylogénie des mammifères avec 6000 gènes Boussau et al, en préparation



Proportion families with LGT



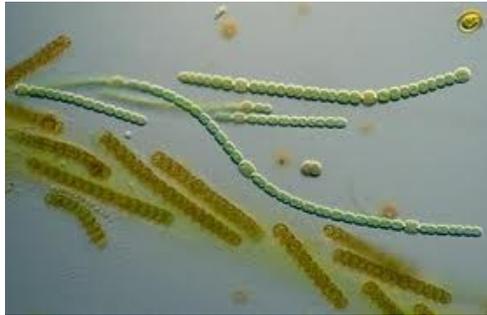
0.3 substitutions/site



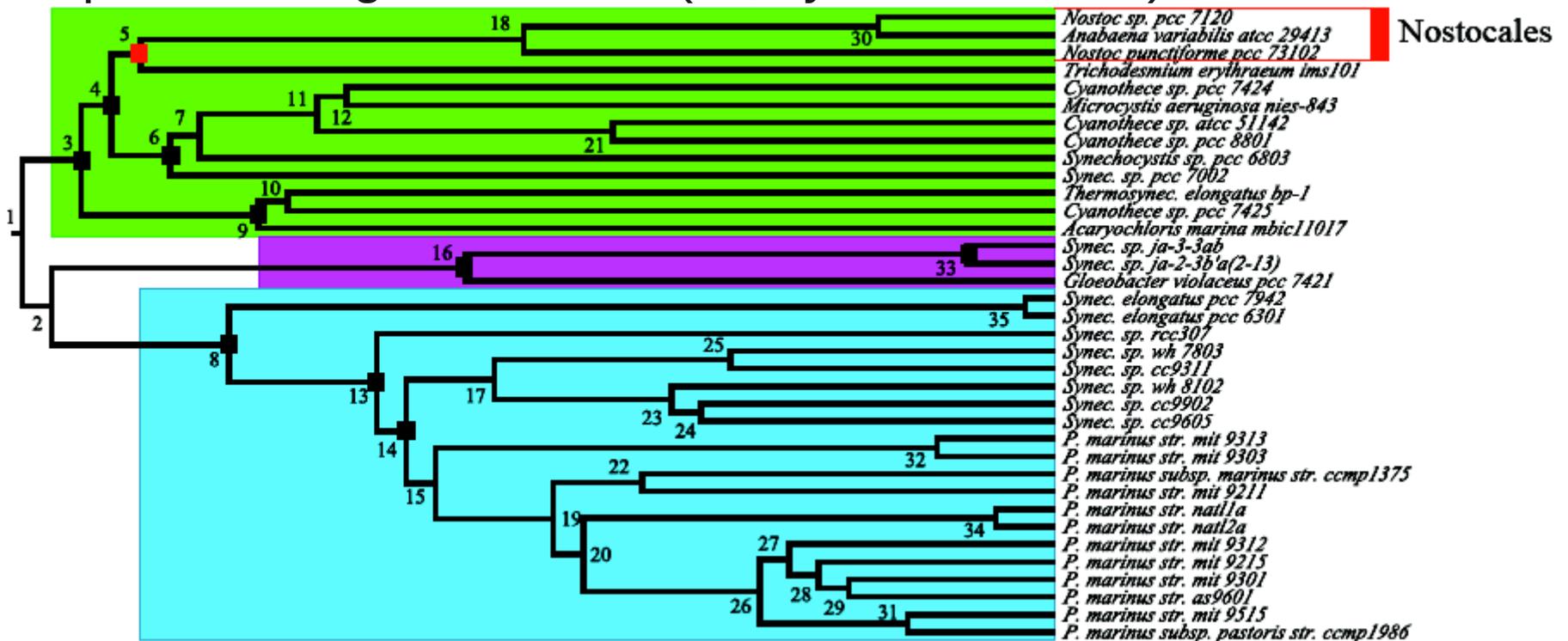
Non seulement les transferts ne brouillent pas le signal phylogénétique, mais ils le renforcent

Arbre de la vie (prokaryotique) avec des milliers de gènes

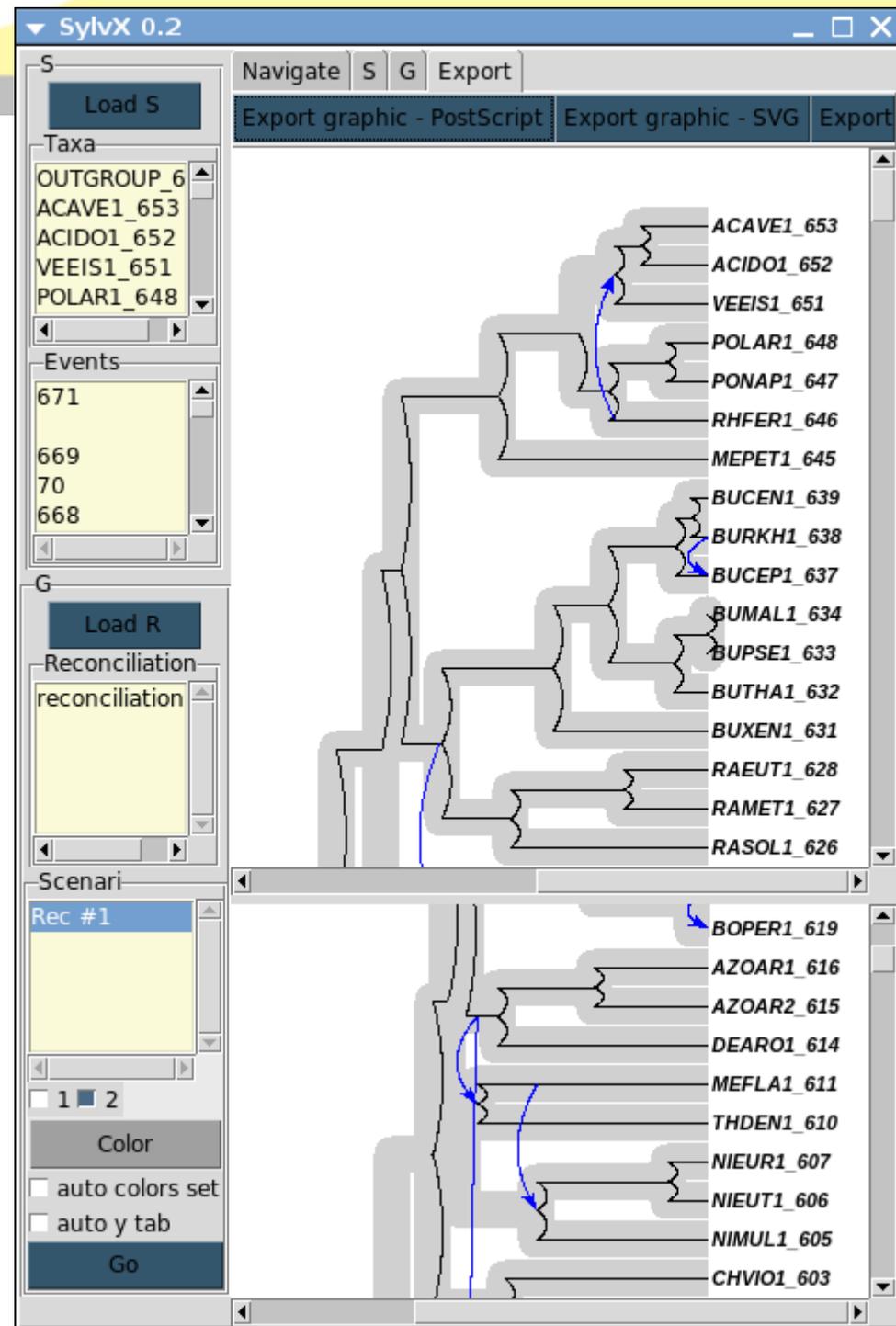
Abby et al, *PNAS*, en révision



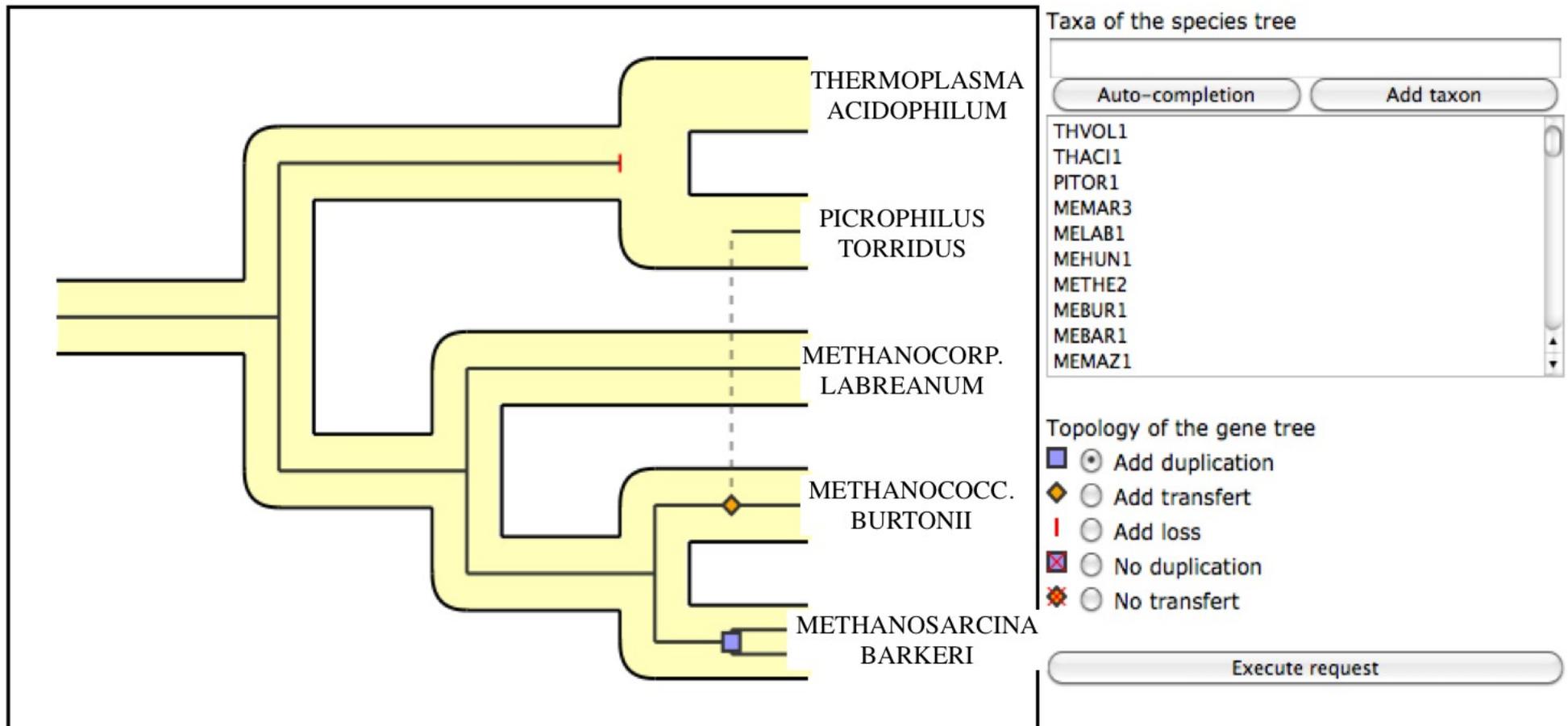
On peut dater des évènements de diversification des espèces avec les transferts de gènes (Szollosi et al, soumis), et comparer au registre fossile (ici, cyanobactéries)



Un logiciel de  
visualisation des  
réconciliations



Une banque de données de transferts de gènes  
Interrogeable depuis le web



The screenshot displays a phylogenetic tree with several taxa and associated gene transfer events. The taxa listed are:

- THERMOPLASMA ACIDOPHILUM
- PICROPHILUS TORRIDUS
- METHANOCORP. LABREANUM
- METHANOCOCC. BURTONII
- METHANOSARCINA BARKERI

Gene transfer events are indicated by symbols on the branches:

- A red vertical line on the branch leading to THERMOPLASMA ACIDOPHILUM indicates an "Add loss" event.
- A yellow diamond on the branch leading to METHANOCOCC. BURTONII indicates an "Add transfert" event.
- A blue square on the branch leading to METHANOSARCINA BARKERI indicates an "Add duplication" event.

The control panel on the right includes:

- Taxa of the species tree:** A search box with "Auto-completion" and "Add taxon" buttons. A list of taxa is shown: THVOL1, THAC11, PITOR1, MEMAR3, MELAB1, MEHUN1, METHE2, MEBUR1, MEBAR1, MEMAZ1.
- Topology of the gene tree:** A legend with radio buttons for:
  - Add duplication (blue square)
  - Add transfert (yellow diamond)
  - Add loss (red line)
  - No duplication (blue square)
  - No transfert (yellow diamond)
- Execute request:** A button at the bottom.

# Publications

## **Deux publications impliquant les trois laboratoires partenaires**

- J-P Doyon, V Ranwez, V Daubin and V Berry. Models, algorithms, and programs for phylogeny reconciliation. Briefings in Bioinformatics (2011) 12(5):392-400.
- J-P Doyon, C Scornavacca, KY Gorbunov, G Szöllösi, V Ranwez and V Berry. An efficient algorithm for gene/species trees parsimonious reconciliation with losses, duplications and transfers. RECOMB Comparative Genomics, september 2010.

## **En tout, une vingtaine d'articles publiés reliés au projet**

Résultats algorithmiques sur les arbres et les réseaux

Bases de données

Outils d'interrogation et visualisation

Résultats biologiques

Revue générale sur un thème

Article de vulgarisation

Ouvrages

## Perspectives

- Intégrer plus d'évènement évolutifs (remaniements chromosomiques)
- Intégrer les dépendances fonctionnelles entre gènes (interactions, régulations, expressions)
- Intégrer des données environnementales et des traits d'histoire de vie

Projet Ancestrome, financé par le CGI, appel Bioinformatique



AGENCE NATIONALE DE LA RECHERCHE  
ANR

ANR-08-EMER-011-01

Merci

<http://www2.lirmm.fr/phylariane/>