

**Journée Doctiss 2009 – Montpellier – 09/04/09**

***Propriétés combinatoires des réseaux  
phylogénétiques de niveau k***

**Philippe Gambette**

**Thèse au LIRMM dirigée par V. Berry et C. Paul**



# Plan

---

- **Les réseaux phylogénétiques**
- **Décomposition des réseaux de niveau  $k$**
- **L'encodage par triplets**
- **Réseaux de niveau  $k$  et modèle coalescent**
- **Conclusion**

# Plan

---

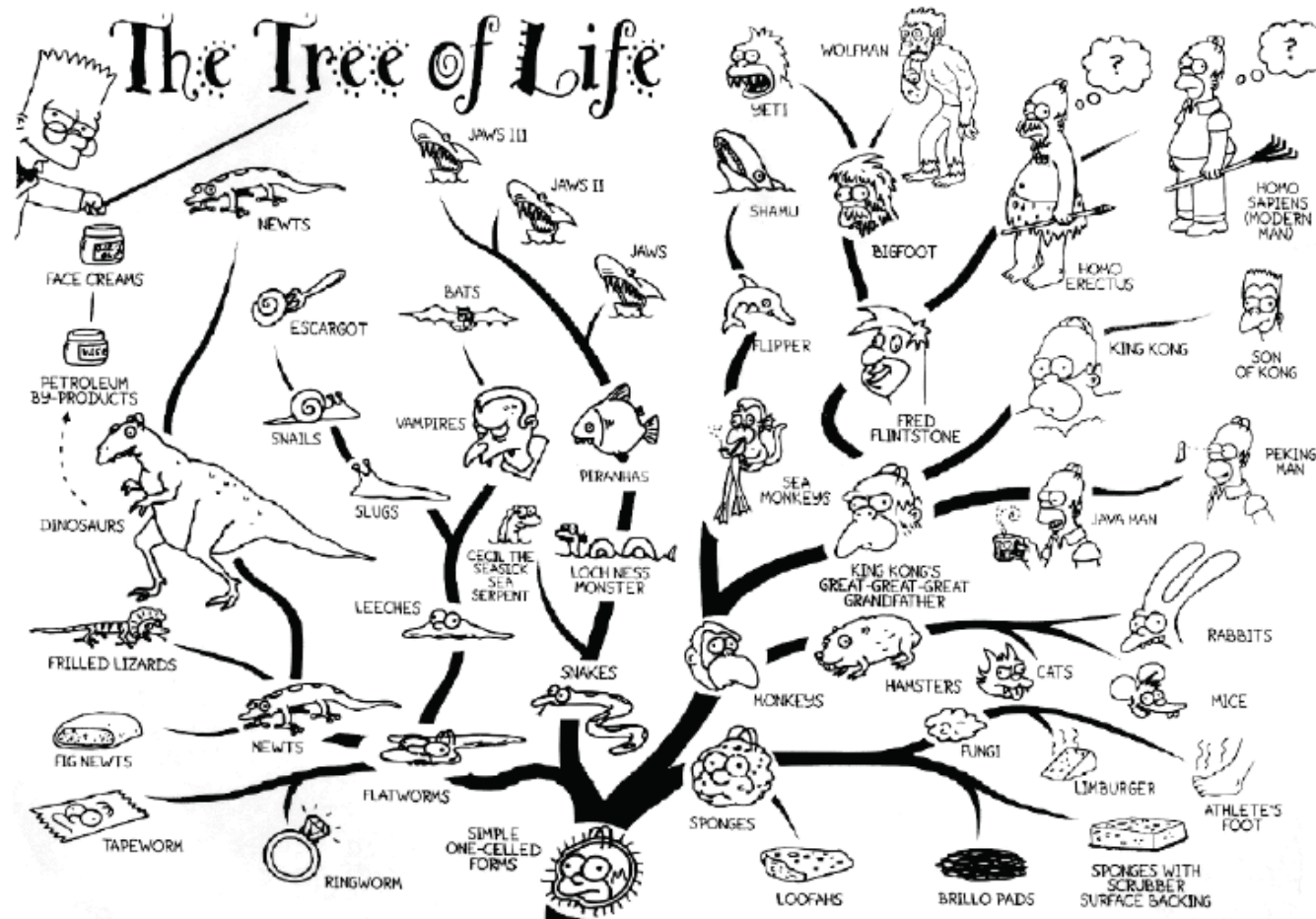
- **Les réseaux phylogénétiques**
- Décomposition des réseaux de niveau  $k$
- L'encodage par triplets
- Réseaux de niveau  $k$  et modèle coalescent
- Conclusion

# Les arbres phylogénétiques

## Arbre phylogénétique



Un arbre phylogénétique est un **arbre** schématique qui montre les relations de parentés entre des entités supposées avoir un ancêtre commun.



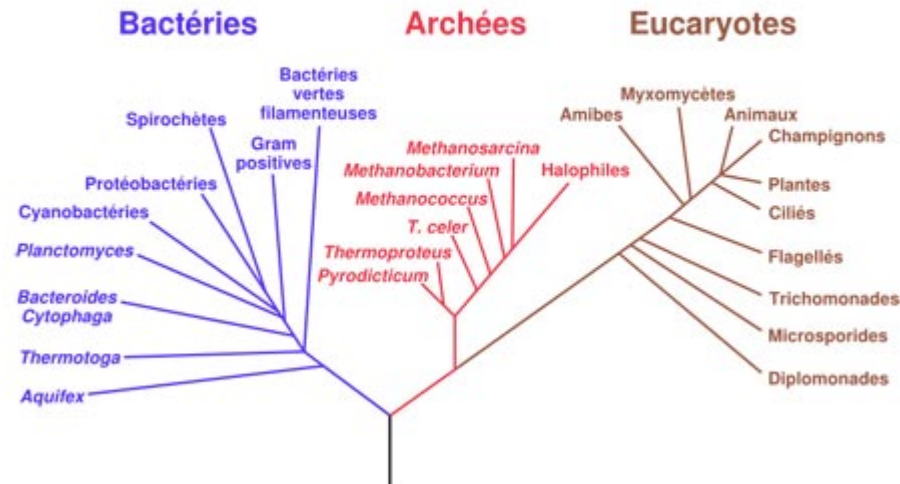
# Les arbres phylogénétiques

## Arbre phylogénétique



Un **arbre phylogénétique** est un **arbre** schématique qui montre les relations de parentés entre des entités supposées avoir un ancêtre commun.

### Arbre phylogénétique de la vie



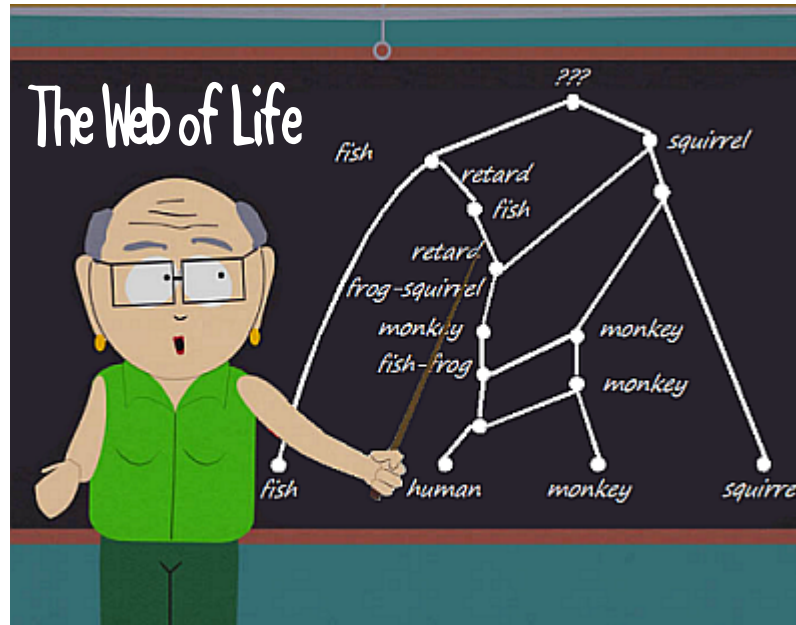
*D'après Woese, Kandler, Wheelis : Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya, Proceedings of the National Academy of Sciences, 87(12), 4576–4579 (1990)*

# Les réseaux phylogénétiques

## Réseau phylogénétique



Un réseau phylogénétique désigne un **graphe** utilisé pour visualiser les relations liées à l'évolution entre des espèces ou des organismes. Il doit être employé quand interviennent des événements d'**hybridations**, de transferts horizontaux de gènes, ou de **recombinaisons génétiques**.



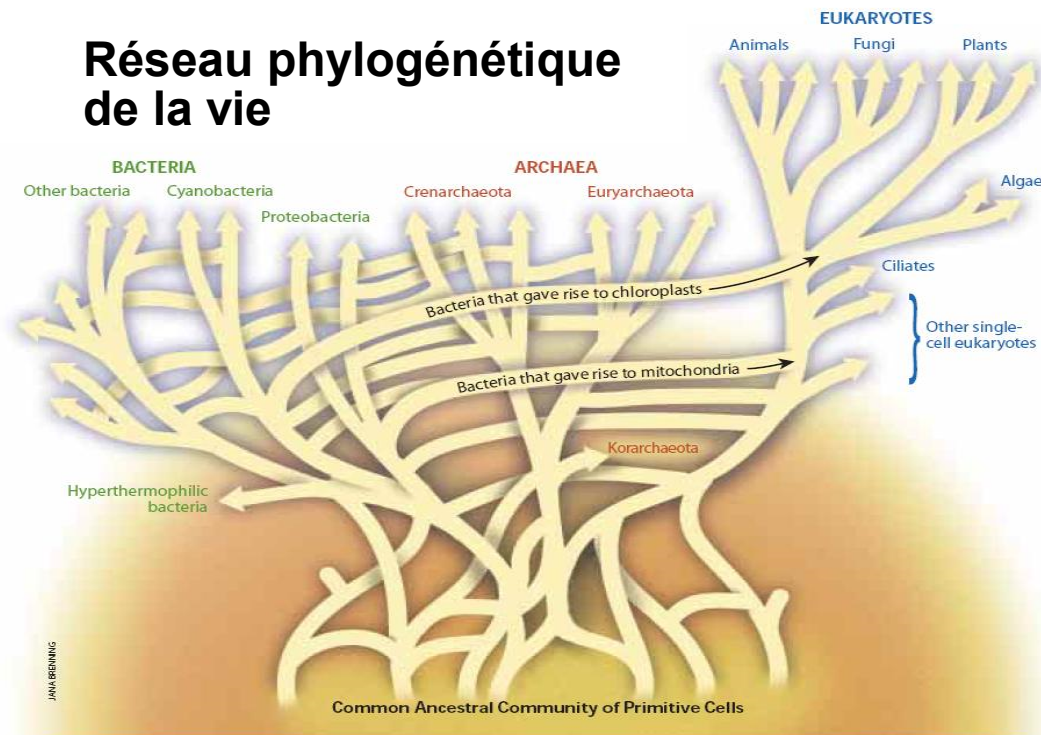
Le "réseau phylogénétique de la vie" d'après Mme Garrison dans South Park S10E12.

# Les réseaux phylogénétiques

## Réseau phylogénétique



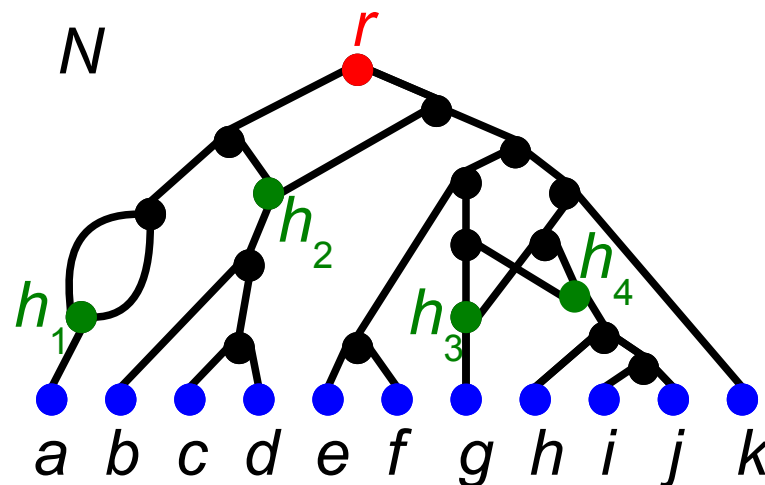
Un réseau phylogénétique désigne un **graphe** utilisé pour visualiser les relations liées à l'évolution entre des espèces ou des organismes. Il doit être employé quand interviennent des événements d'**hybridations**, de transferts horizontaux de gènes, ou de **recombinaisons génétiques**.



# Réseaux phylogénétiques de niveau $k$

Réseau phylogénétique de niveau  $k$  sur un ensemble  $X$  de  $n$  taxons = graphe orienté où :

- un sommet a degré entrant 0 et sortant 2 : la **racine**,
- tous les autres sommets ont :
  - degré entrant 1 et sortant 2: **sommets de spéciation**,
  - degré entrant 2 et sortant  $\leq 1$  : **sommets hybrides**,
  - ou degré entrant 1 et sortant 0 : **feuilles** étiquetées par  $X$ ,
- chaque **blob** a au plus  $k$  sommets hybrides.



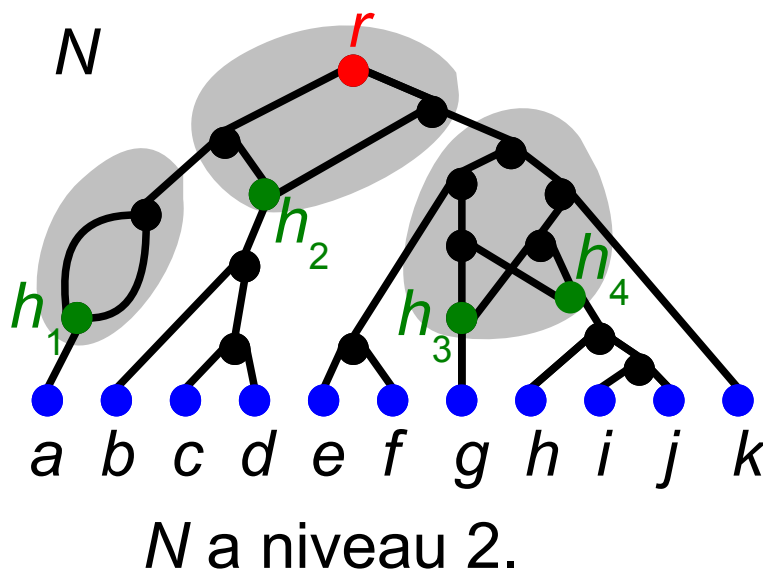
Tous les arcs sont orientés vers le bas



# Réseaux phylogénétiques de niveau $k$

**Réseau phylogénétique de niveau  $k$**  sur un ensemble  $X$  de  $n$  taxons = graphe orienté où :

- un sommet a degré entrant 0 et sortant 2 : la **racine**,
- tous les autres sommets ont :
  - degré entrant 1 et sortant 2: **sommets de spéciation**,
  - degré entrant 2 et sortant  $\leq 1$  : **sommets hybrides**,
  - ou degré entrant 1 et sortant 0 : **feuilles** étiquetées par  $X$ ,
- chaque **blob** a au plus  $k$  sommets hybrides.

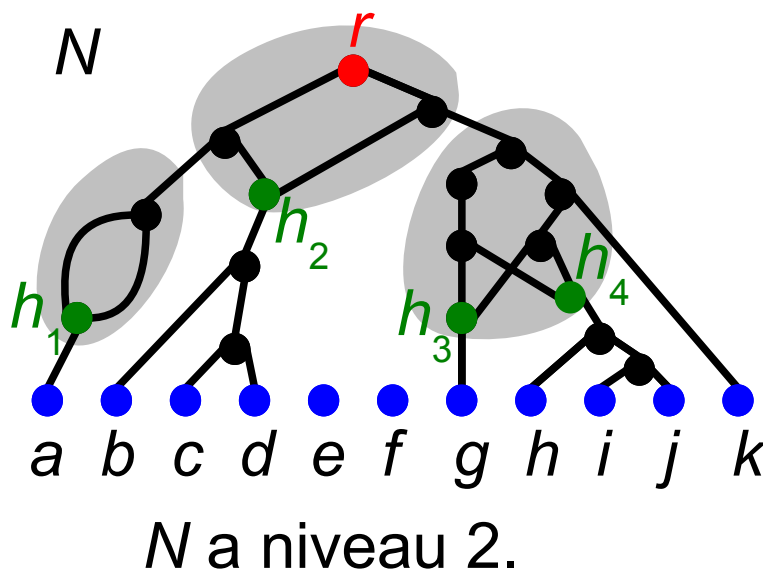


Un **blob** est une composante **biconnexe** maximale du graphe non orienté sous-jacent, c'est à dire un sous-graphe maximal non déconnecté par la suppression d'un sommet quelconque.

# Réseaux phylogénétiques de niveau $k$

**Réseau phylogénétique de niveau  $k$**  sur un ensemble  $X$  de  $n$  taxons = graphe orienté où :

- un sommet a degré entrant 0 et sortant 2 : la **racine**,
- tous les autres sommets ont :
  - degré entrant 1 et sortant 2: **sommets de spéciation**,
  - degré entrant 2 et sortant  $\leq 1$  : **sommets hybrides**,
  - ou degré entrant 1 et sortant 0 : **feuilles** étiquetées par  $X$ ,
- chaque **blob** a au plus  $k$  sommets hybrides.

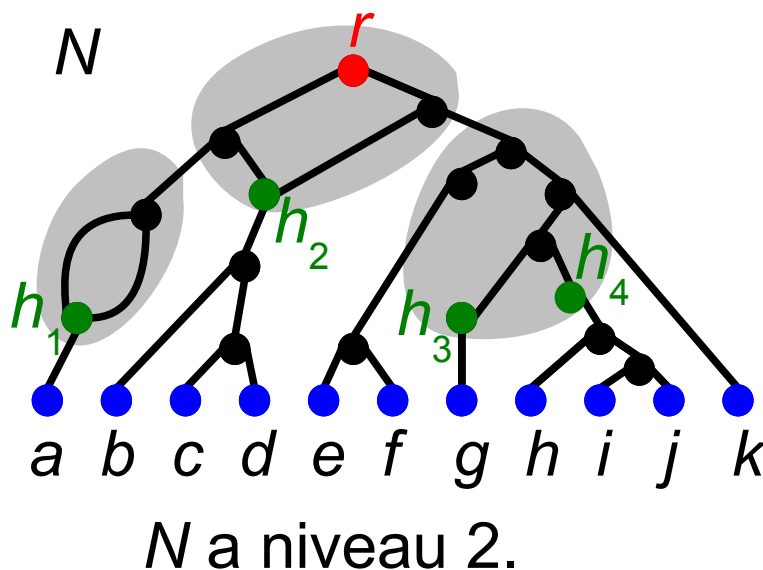


Un **blob** est une composante **biconnexe** maximale du graphe non orienté sous-jacent, c'est à dire un sous-graphe maximal non déconnecté par la suppression d'un sommet quelconque.

# Réseaux phylogénétiques de niveau $k$

**Réseau phylogénétique de niveau  $k$**  sur un ensemble  $X$  de  $n$  taxons = graphe orienté où :

- un sommet a degré entrant 0 et sortant 2 : la **racine**,
- tous les autres sommets ont :
  - degré entrant 1 et sortant 2: **sommets de spéciation**,
  - degré entrant 2 et sortant  $\leq 1$  : **sommets hybrides**,
  - ou degré entrant 1 et sortant 0 : **feuilles** étiquetées par  $X$ ,
- chaque **blob** a au plus  $k$  sommets hybrides.

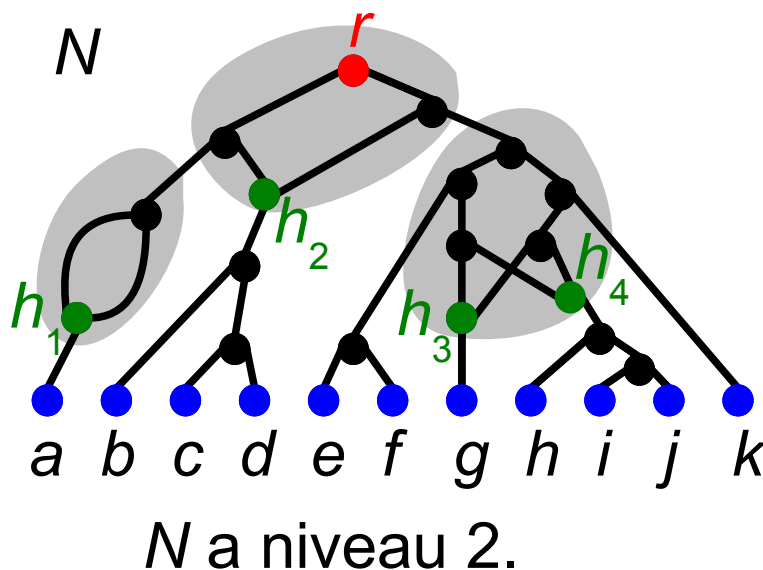


Un **blob** est une composante **biconnexe** maximale du graphe non orienté sous-jacent, c'est à dire un sous-graphe maximal non déconnecté par la suppression d'un sommet quelconque.

# Réseaux phylogénétiques de niveau $k$

**Réseau phylogénétique de niveau  $k$**  sur un ensemble  $X$  de  $n$  taxons = graphe orienté où :

- un sommet a degré entrant 0 et sortant 2 : la **racine**,
- tous les autres sommets ont :
  - degré entrant 1 et sortant 2: **sommets de spéciation**,
  - degré entrant 2 et sortant  $\leq 1$  : **sommets hybrides**,
  - ou degré entrant 1 et sortant 0 : **feuilles** étiquetées par  $X$ ,
- chaque **blob** a au plus  $k$  sommets hybrides.



Un **blob** est une composante **biconnexe** maximale du graphe non orienté sous-jacent, c'est à dire un sous-graphe maximal non déconnecté par la suppression d'un sommet quelconque.

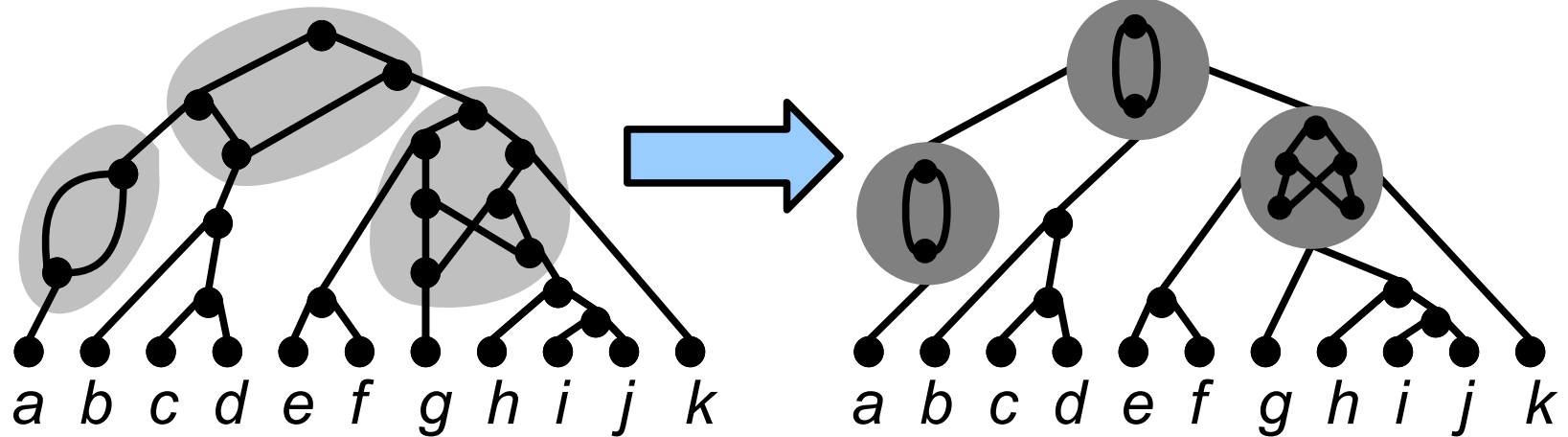
# Plan

---

- Les réseaux phylogénétiques
- **Décomposition des réseaux de niveau  $k$**
- L'encodage par triplets
- Réseaux de niveau  $k$  et modèle coalescent
- Conclusion

# Décomposition des réseaux de niveau $k$

réseau de niveau  $k$  = arbre de générateurs de niveau  $k$



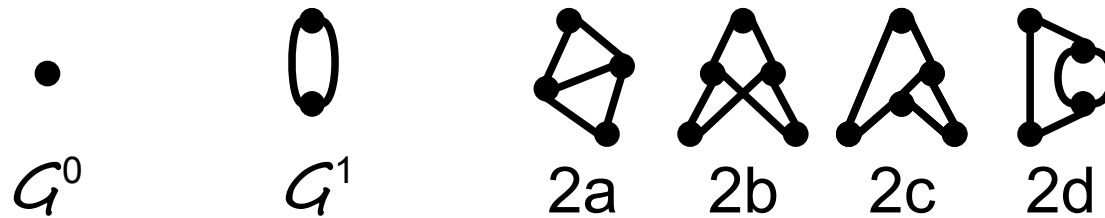
$N$ , un réseau de niveau  $k$ .

$N$  décomposé comme  
**arbre de générateurs.**

*Générateurs introduits par van Iersel & al (Recomb 2008)  
pour une classe restreinte de réseaux de niveau  $k$ .*

# Générateurs de niveau $k$

Un **générateur de niveau  $k$**  est un réseau de niveau  $k$  biconnexe.



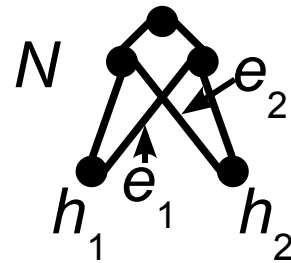
Les **côtés** d'un générateur sont :

- ses arcs
- ses sommets hybrides de degré sortant 0

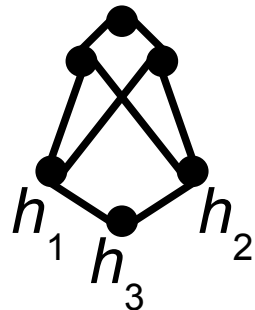
$\mathbf{S}_k$  est l'ensemble des générateurs de niveau au plus  $k$ .

# Construction des générateurs

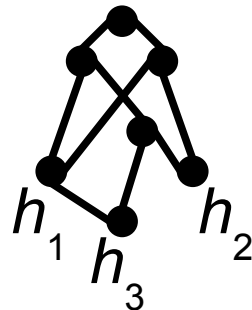
Règles de construction des générateurs de niveau  $k+1$  à partir des générateurs de niveau  $k$  :



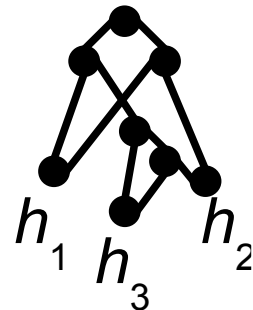
Règle  $R_1$  :



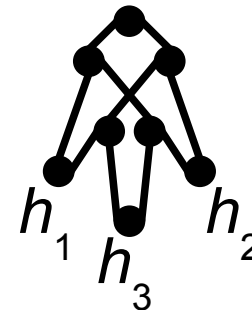
$R_1(N, h_1, h_2)$



$R_1(N, h_1, e_2)$



$R_1(N, e_2, e_2)$

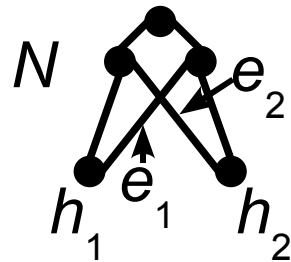


$R_1(N, e_1, e_2)$

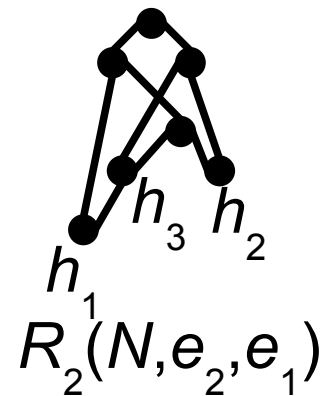
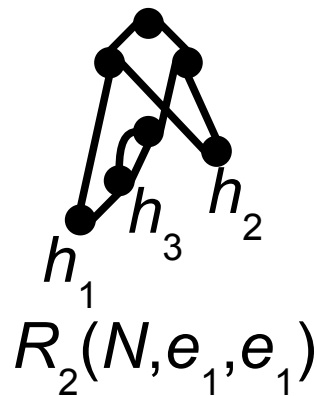
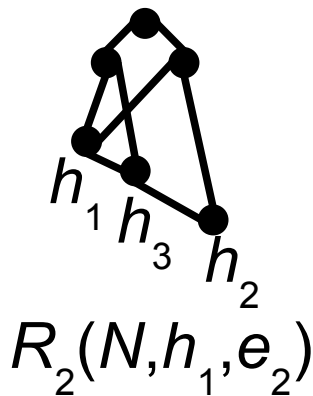


# Construction des générateurs

Règles de construction des générateurs de niveau  $k+1$  à partir des générateurs de niveau  $k$  :



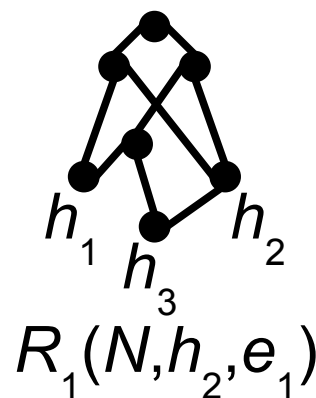
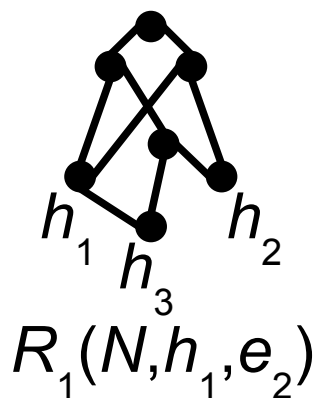
Règle  $R_2$  :



# Nombre de générateurs

## **Problème !**

Certains des générateurs de niveau  $k+1$  ainsi obtenus depuis les générateurs de niveau  $k$  sont isomorphes !



Greetings from [The On-Line Encyclopedia of Integer Sequences!](http://www.oeis.org/)

[Hints](#)

Search: 1, 4, 65, 1993

I am sorry, but the terms do not match anything in the table.

# Nombre de générateurs

## ***Borne inférieure :***

Il y a un nombre **exponentiel** de générateurs !

## ***Idée :***

Coder les nombres entre 0 et  $2^{k-1}-1$  par des générateurs de niveau  $k$ .

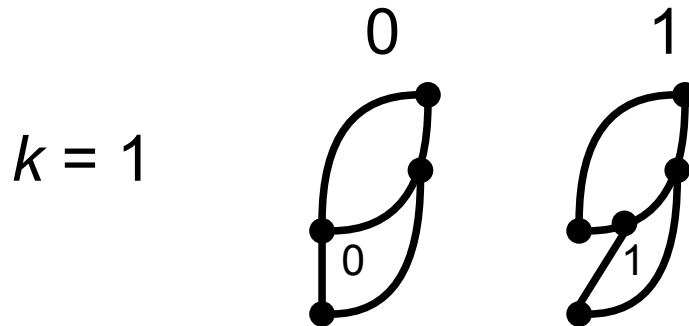
# Nombre de générateurs

## ***Borne inférieure :***

Il y a un nombre **exponentiel** de générateurs !

## ***Idée :***

Coder les nombres entre 0 et  $2^{k-1}-1$  par des générateurs de niveau  $k$ .



# Nombre de générateurs

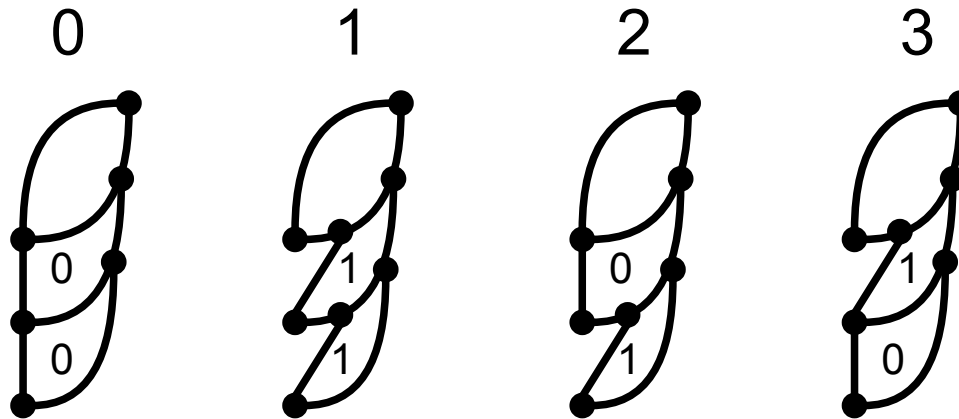
## **Borne inférieure :**

Il y a un nombre **exponentiel** de générateurs !

## **Idée :**

Coder les nombres entre 0 et  $2^{k-1}-1$  par des générateurs de niveau  $k$ .

$k = 2$



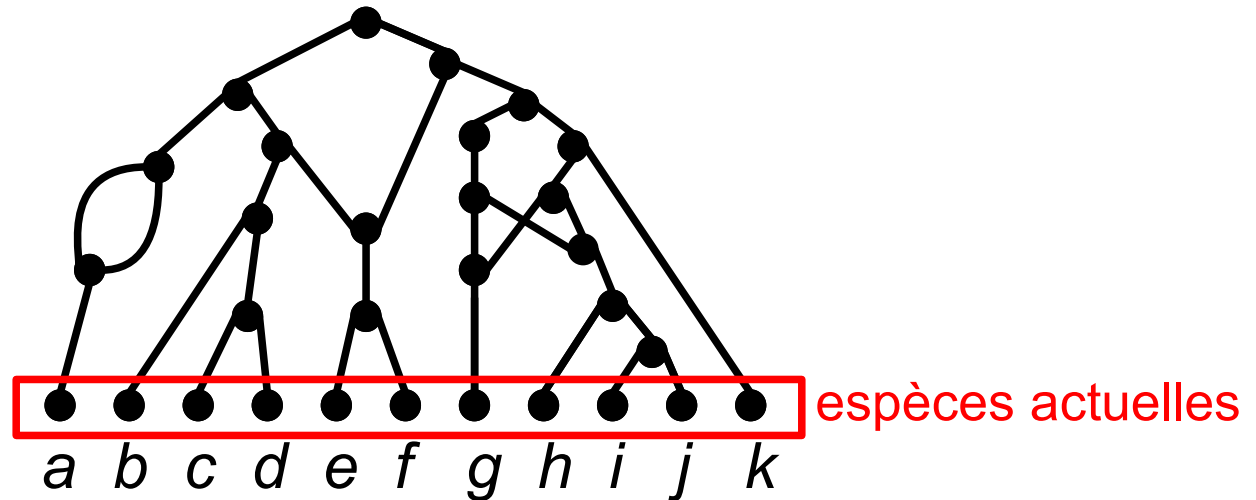
# Plan

---

- Les réseaux phylogénétiques
- Décomposition des réseaux de niveau  $k$
- **L'encodage par triplets**
- Réseaux de niveau  $k$  et modèle coalescent
- Conclusion

# Reconstruction d'un réseau de niveau $k$

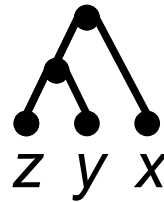
On cherche à reconstruire un réseau de niveau minimum



qui explique des données **actuellement disponibles**  
(à propos des feuilles uniquement) :  
séquences ADN, distances entre les séquences, triplets...

# Reconstruction depuis les triplets

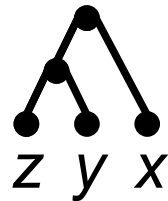
Un **triplet**  $x|yz$  est un arbre phylogénétique enraciné sur 3 taxons  $\{x,y,z\}$  tel que  $x$ , et le père de  $y$  et  $z$ , sont des fils de la racine.





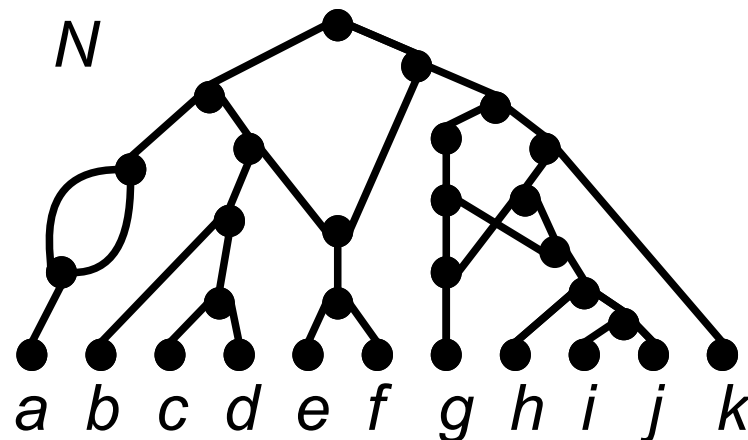
# Reconstruction depuis les triplets

Un **triplet**  $x|yz$  est un arbre phylogénétique enraciné sur 3 taxons  $\{x,y,z\}$  tel que  $x$ , et le père de  $y$  et  $z$ , sont des fils de la racine.



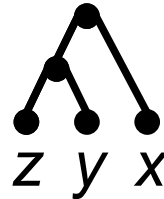
$x|yz$  est **compatible** avec un réseau phylogénétique  $N$  si:

- $N$  contient deux noeuds  $u$  et  $v$
- et des chemins intérieurement disjoints deux à deux :
  - de  $u$  à  $y$ ,
  - de  $u$  à  $z$ ,
  - de  $v$  à  $u$ ,
  - et de  $v$  à  $x$ .



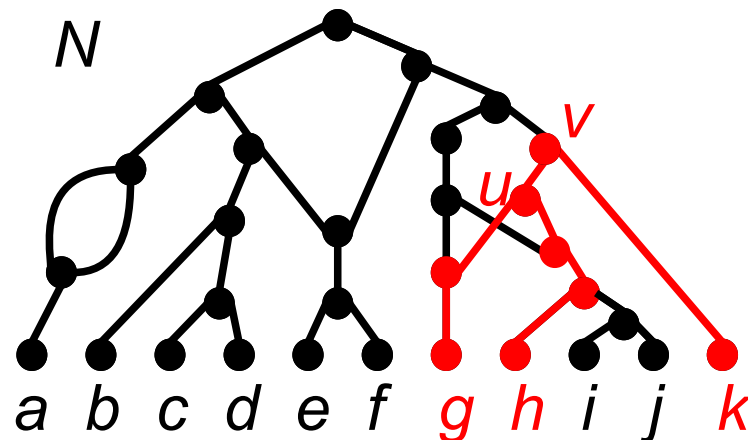
# Reconstruction depuis les triplets

Un **triplet**  $x|yz$  est un arbre phylogénétique enraciné sur 3 taxons  $\{x,y,z\}$  tel que  $x$ , et le père de  $y$  et  $z$ , sont des fils de la racine.



$x|yz$  est **compatible** avec un réseau phylogénétique  $N$  si:

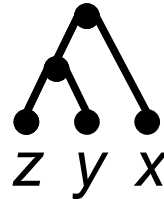
- $N$  contient deux noeuds  $u$  et  $v$
- et des chemins intérieurement disjoints deux à deux :
  - de  $u$  à  $y$ ,
  - de  $u$  à  $z$ ,
  - de  $v$  à  $u$ ,
  - et de  $v$  à  $x$ .



$k|gh$  **compatible** avec  $N$ .

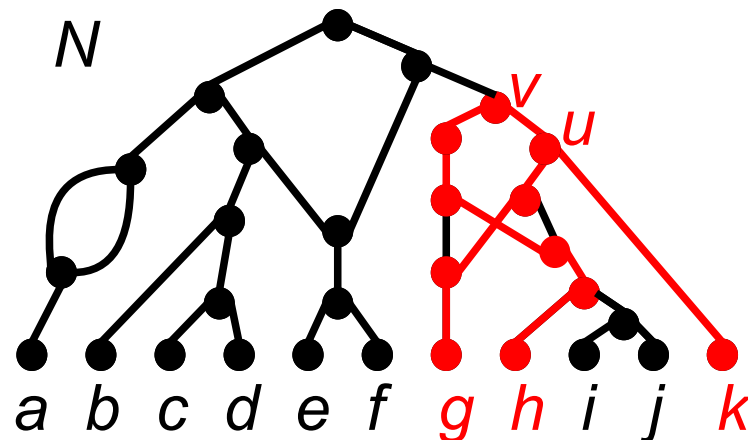
# Reconstruction depuis les triplets

Un **triplet**  $x|yz$  est un arbre phylogénétique enraciné sur 3 taxons  $\{x,y,z\}$  tel que  $x$ , et le père de  $y$  et  $z$ , sont des fils de la racine.



$x|yz$  est **compatible** avec un réseau phylogénétique  $N$  si:

- $N$  contient deux noeuds  $u$  et  $v$
- et des chemins intérieurement disjoints deux à deux :
  - de  $u$  à  $y$ ,
  - de  $u$  à  $z$ ,
  - de  $v$  à  $u$ ,
  - et de  $v$  à  $x$ .



$h|gk$  compatible avec  $N$ .



# Reconstruction depuis les triplets

Algorithmes polynomiaux de reconstruction de réseaux :

- de niveau 1

*(Jansson & Sung, COCOON 2004,  
Jansson, Nguyen & Sung, SODA 2005)*

- de niveau 2

*(Iersel et al, RECOMB 2008)*

- de niveau  $>2$

*(To & Habib, CPM 2009)*

compatibles avec un ensemble de triplets.

# Reconstruction depuis les triplets

Algorithmes polynomiaux de reconstruction de réseaux :

- de niveau 1

*(Jansson & Sung, COCOON 2004,  
Jansson, Nguyen & Sung, SODA 2005)*

- de niveau 2

*(Iersel et al, RECOMB 2008)*

- de niveau  $>2$

*(To & Habib, CPM 2009)*

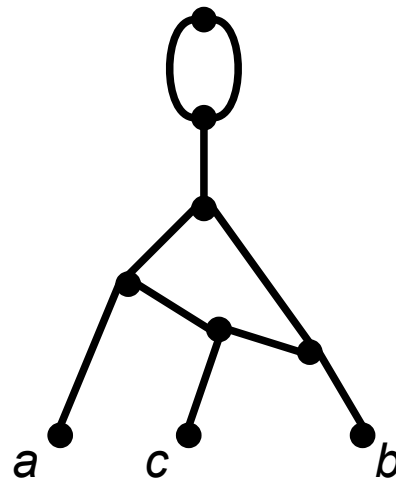
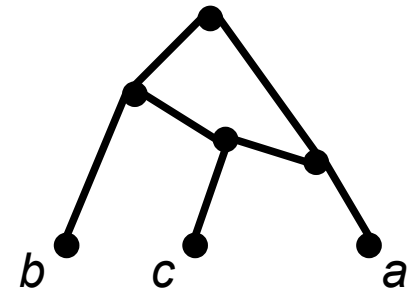
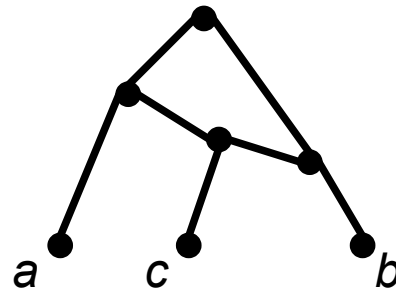
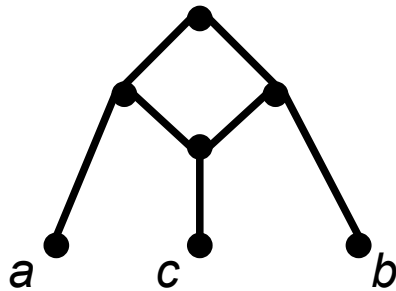
compatibles avec un ensemble de triplets

**MAIS**

**l'ensemble des triplets représente-t-il vraiment le réseau ?**

# Encodage par triplets au niveau 1

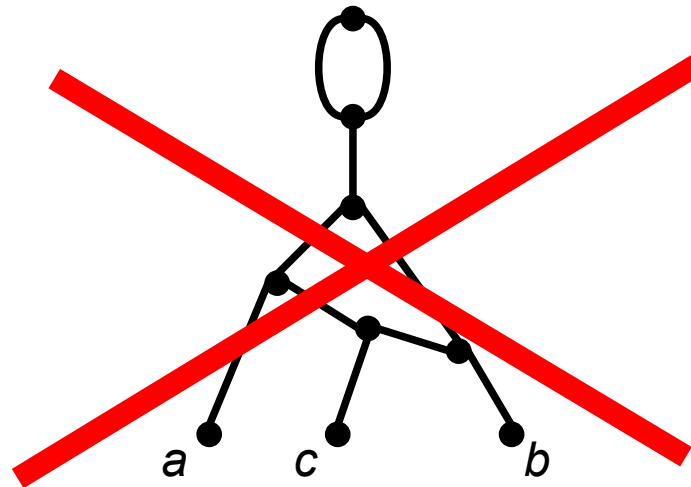
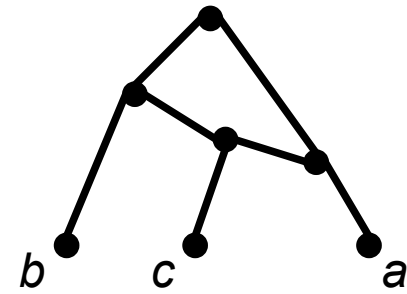
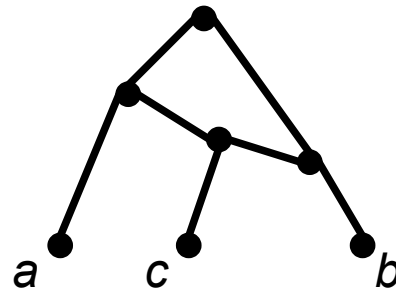
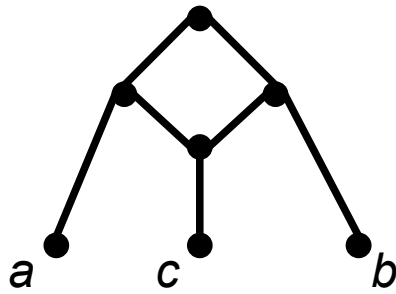
Combien de réseaux de niveau 1 sont compatibles avec exactement  $\{a|bc, b|ac, c|ab\}$  ?



$\{a|bc, b|ac, c|ab\}$  n'est pas **encodé** par un réseau de niveau 1.

# Encodage par triplets au niveau 1

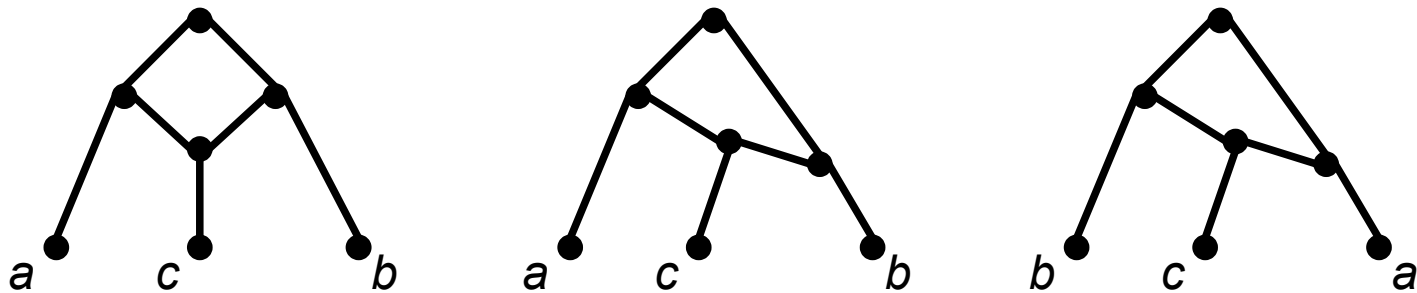
Combien de réseaux de niveau 1 **minimaux** sont compatibles avec exactement  $\{a|bc, b|ac, c|ab\}$  ?





# Encodage par triplets au niveau 1

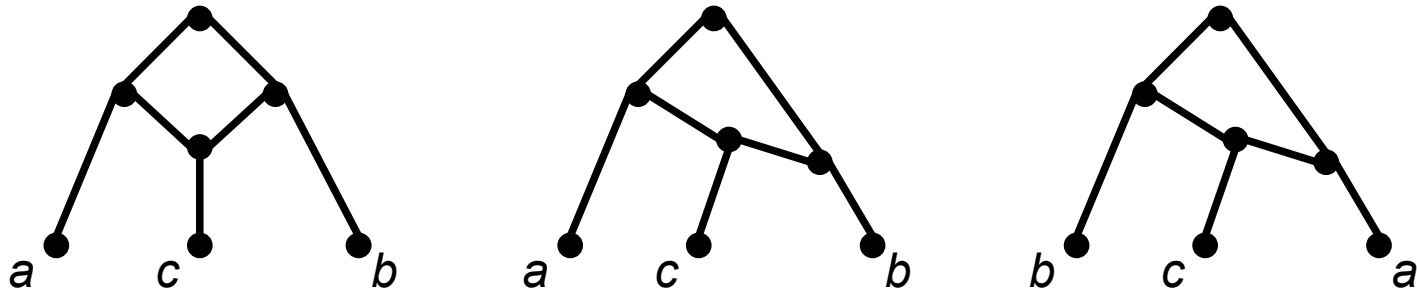
Combien de réseaux de niveau 1 minimaux sont compatibles avec exactement  $\{a|bc, b|ac, c|ab\}$  ?



$\{a|bc, b|ac, c|ab\}$  n'est pas **encodé**  
par un réseau de niveau 1.

# Encodage par triplets au niveau 1

Combien de réseaux de niveau 1 minimaux sont compatibles avec exactement  $\{a|bc, b|ac, c|ab\}$  ?

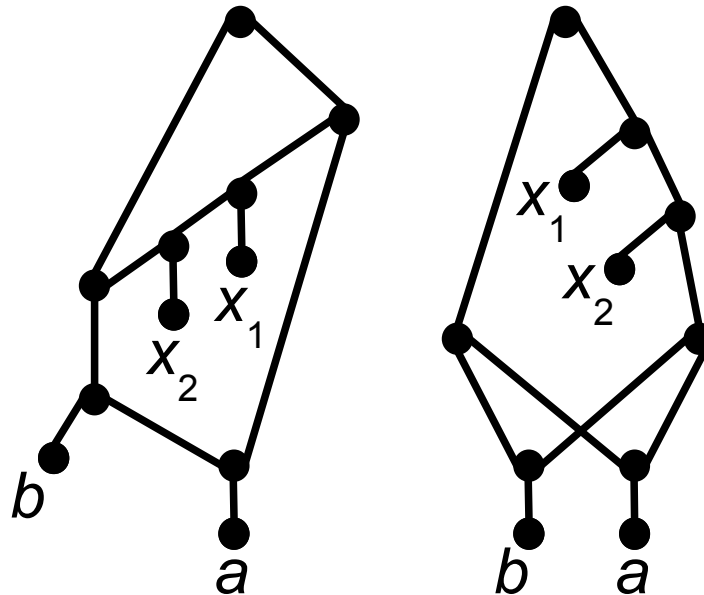


Ces trois réseaux ne sont pas **encodés** par leur ensemble de triplets  $\{a|bc, b|ac, c|ab\}$ .

Un réseau de niveau 1 n'est pas encodé par ses triplets ssi il contient un blob à 4 sommets

# Encodage par triplets au niveau 2

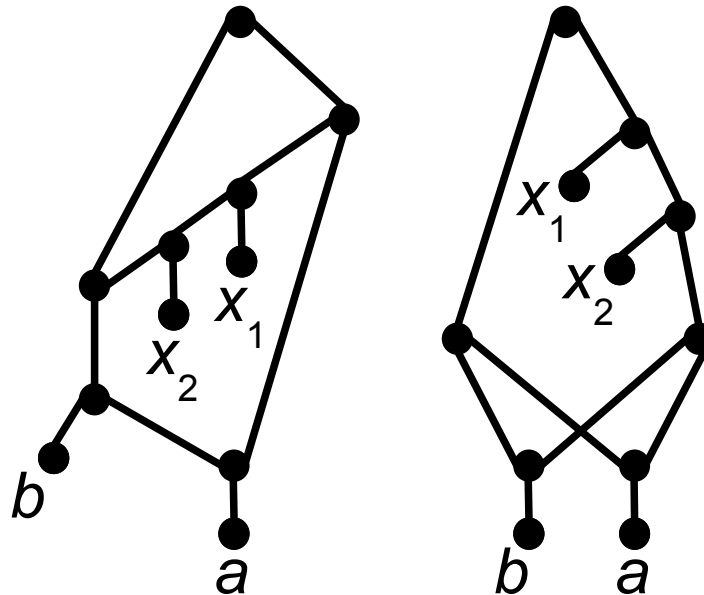
Au niveau 2, pas de caractérisation de l'encodage.



Ces deux réseaux de niveau 2 ne sont pas **encodés** par leur ensemble de triplets.

# Encodage par triplets au niveau 2

Au niveau 2, pas de caractérisation de l'encodage.



Ces deux réseaux de niveau 2 ne sont pas **encodés** par leur ensemble de triplets.

Deux histoires évolutives **très différentes**,  
mais le **même ensemble de triplets**.

# Plan

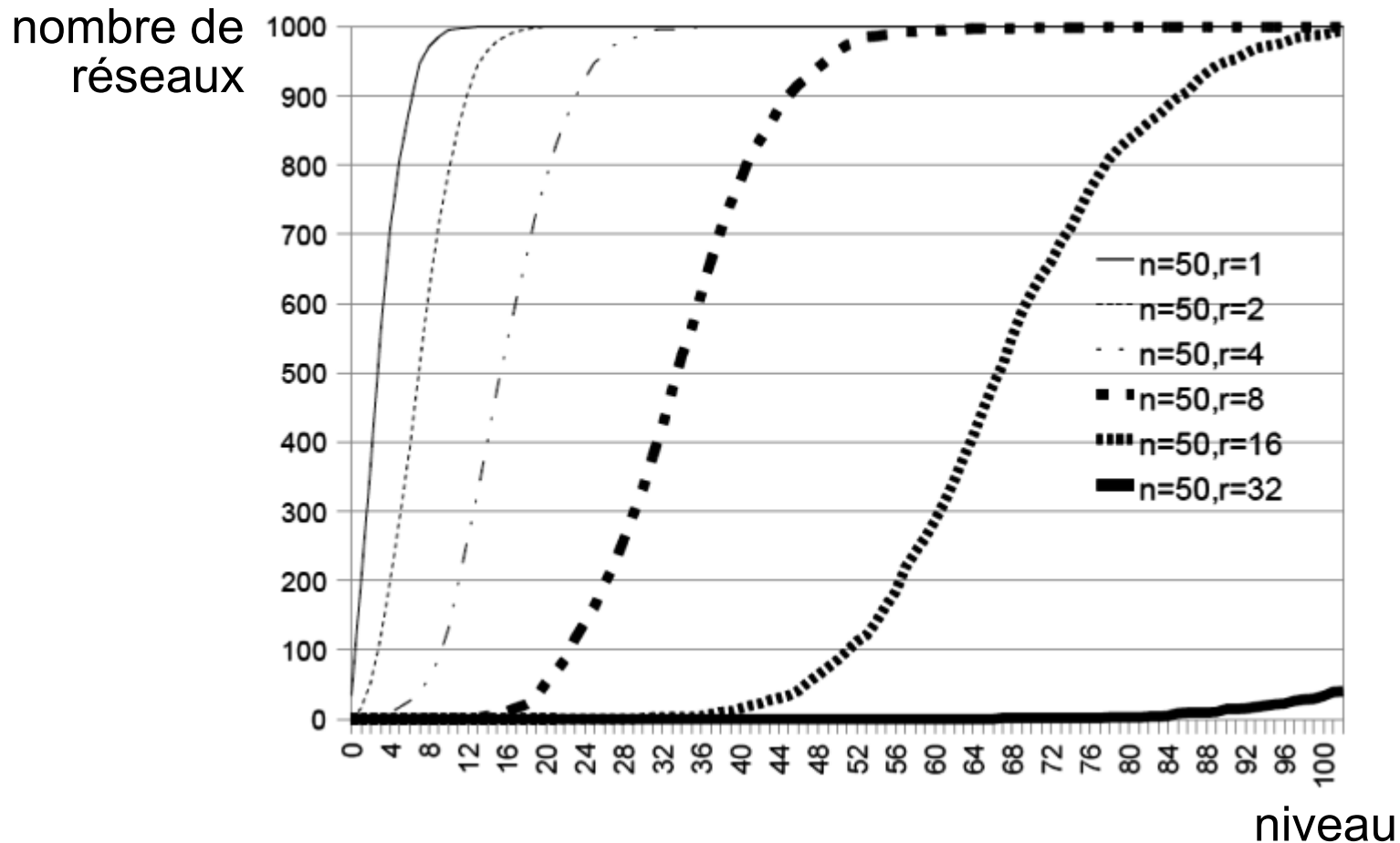
---

- Les réseaux phylogénétiques
- Décomposition des réseaux de niveau  $k$
- L'encodage par triplets
- **Réseaux de niveau  $k$  et modèle coalescent**
- Conclusion

# Réseaux de niveau $k$ et modèle coalescent

On simule l'évolution par le modèle coalescent.

Quel est le **niveau des réseaux obtenus** ?



# Plan

---

- Les réseaux phylogénétiques
- Décomposition des réseaux de niveau  $k$
- L'encodage par triplets
- Réseaux de niveau  $k$  et modèle coalescent
- **Conclusion**

# Réseaux de niveau $k$ ? Prudence !

Les réseaux de niveau  $k$  ont une jolie structure arborée, mais :

- grande complexité à l'intérieur des blobs
- ne sont pas toujours **encodés** par leur ensemble de triplets
- ont un **niveau élevé** dans les cas où le **modèle coalescent avec recombinaison** s'applique.

Perspective :

trouver un **paramètre alternatif** pour exprimer la complexité des réseaux phylogénétiques ?



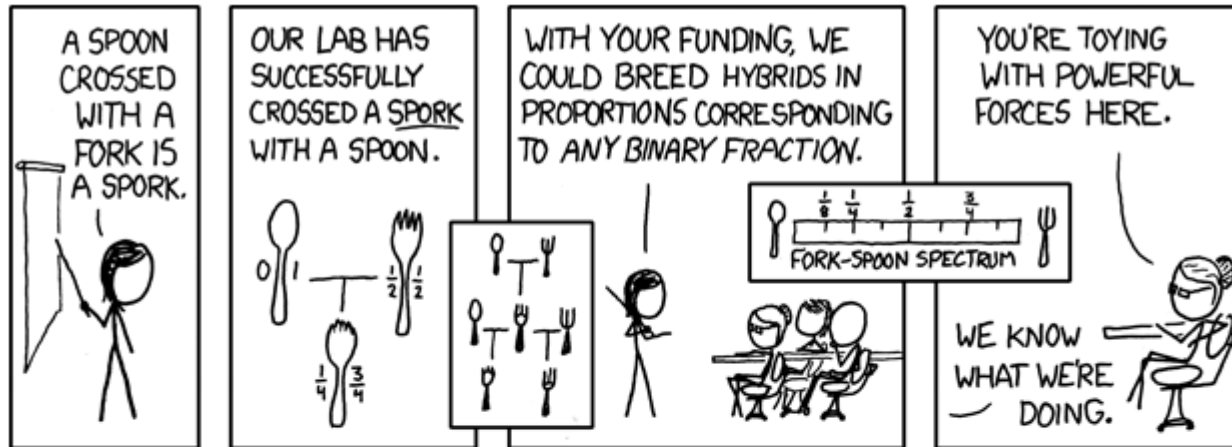
# Des questions ?

---

Merci pour votre attention !

# Des questions ?

*Une dernière incitation à la prudence en cas d'hybridation...*



- UNE CUILLER CROISÉE AVEC UNE FOURCHETTE EST UNE CUIRCHETTE.

NOTRE LABO A RÉUSSI À CROISER UNE CUIRCHETTE AVEC UNE CUILLER.

AVEC VOTRE FINANCEMENT, NOUS POURRIONS CRÉER DES HYBRIDES DANS DES PROPORTIONS CORRESPONDANT À N'IMPORTE QUEL NOMBRE DIADIQUE.

- VOUS MANIPULEZ DES FORCES TRÈS PUISSANTES.

- NOUS SAVONS CE QUE NOUS FAISONS.



DEUX SEMAINES PLUS TARD...

Randall Munroe - Xkcd... en français - <http://xkcd.free.fr>

Merci pour votre attention !

<http://www.lirmm.fr/~gambette>