

**Journées annuelles
de la Société Française de Systématique
Angers – 06/10/2009**

Reconstruction combinatoire de réseaux phylogénétiques

Philippe Gambette



Plan

- **Les réseaux phylogénétiques**
- **L'arbre en filigrane**
- **Motivations de l'approche combinatoire**
- **Méthodes de reconstruction combinatoire**
- **Limites des méthodes combinatoires**
- **Perspectives**

Plan

- **Les réseaux phylogénétiques**
- L'arbre en filigrane
- Motivations de l'approche combinatoire
- Méthodes de reconstruction combinatoire
- Limites des méthodes combinatoires
- Perspectives

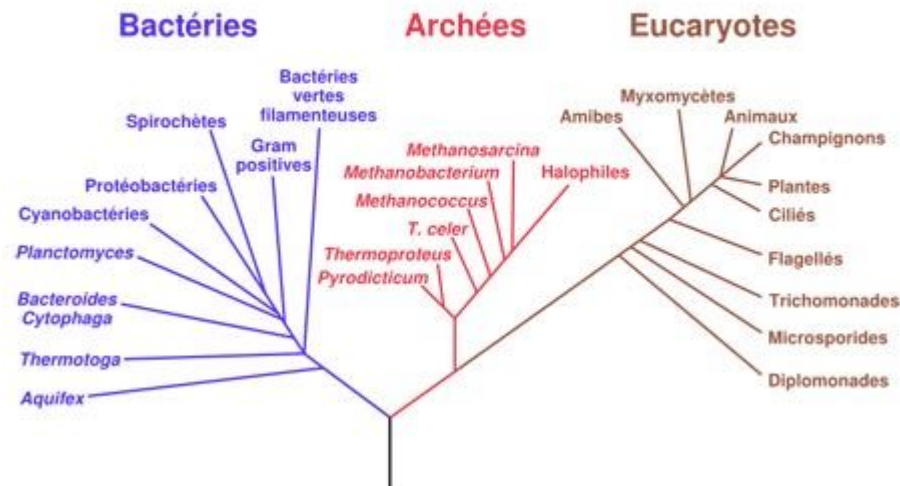
Les arbres phylogénétiques

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



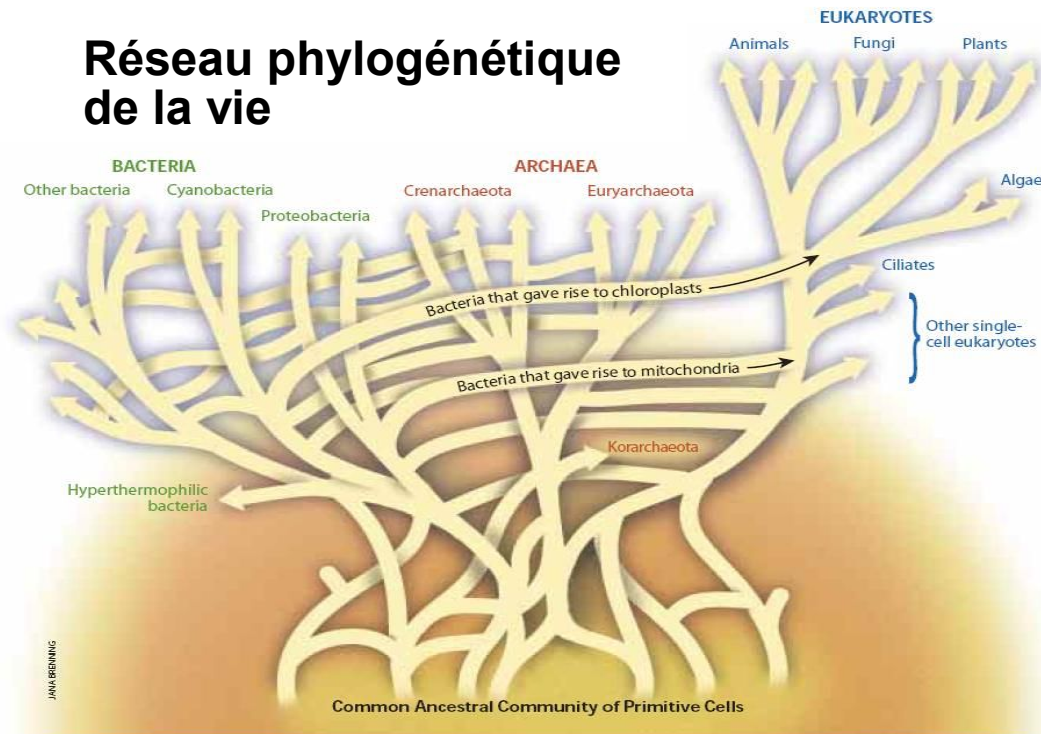
D'après Woese, Kandler, Wheelis : Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya, Proceedings of the National Academy of Sciences, 87(12), 4576–4579 (1990)

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**.



Les réseaux phylogénétiques

Réseau phylogénétique



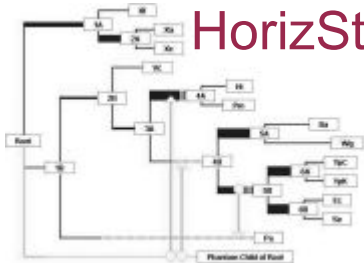
Un réseau phylogénétique désigne un **graphe** utilisé pour visualiser les relations liées à l'évolution entre des espèces ou des organismes. Il doit être employé quand interviennent des événements d'**hybridations**, de transferts horizontaux de gènes, ou de **recombinaisons génétiques**.



réseau de niveau 2

Level-2

diagramme de synthèse

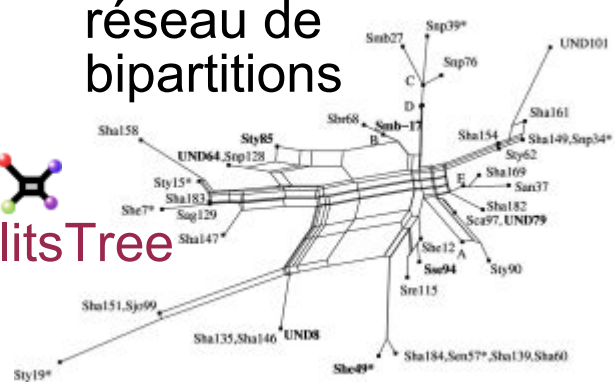


HorizStory

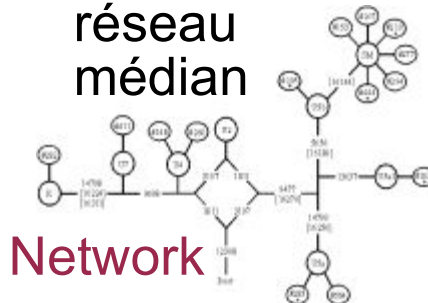
réseau de bipartitions



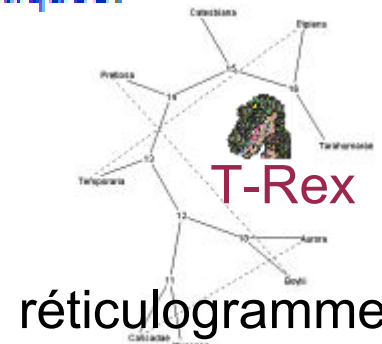
SplitsTree



réseau médian



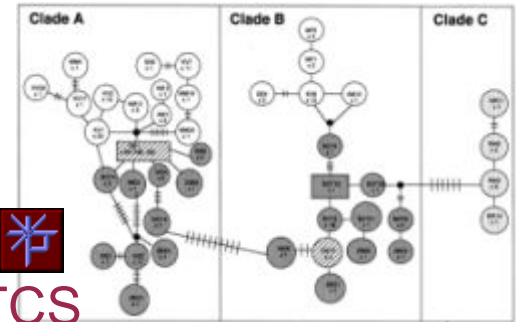
Network



T-Rex

réticulogramme

réseau couvrant minimum



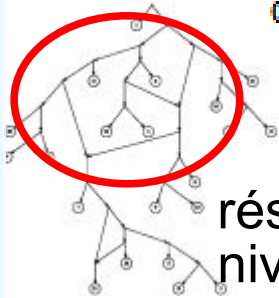
TCS

Les réseaux phylogénétiques

Réseau phylogénétique



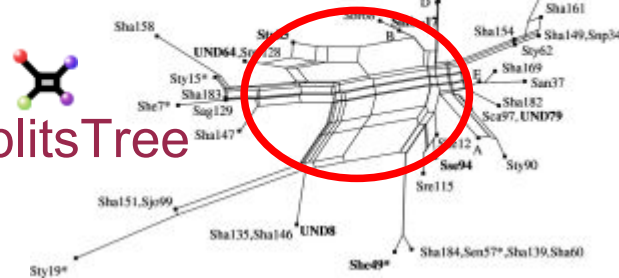
Un réseau phylogénétique désigne un **graphe** utilisé pour visualiser les relations liées à l'évolution entre des espèces ou des organismes. Il doit être employé quand interviennent des événements d'**hybridations**, de transferts horizontaux de gènes, ou de **recombinaisons génétiques**.



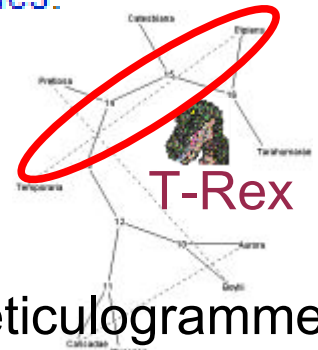
réseau de niveau 2

Level-2

réseau de bipartitions

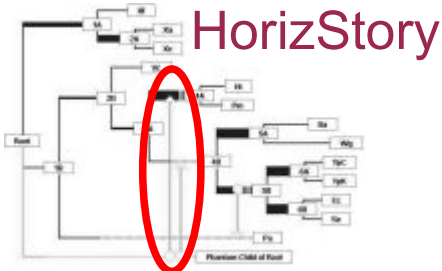


SplitsTree



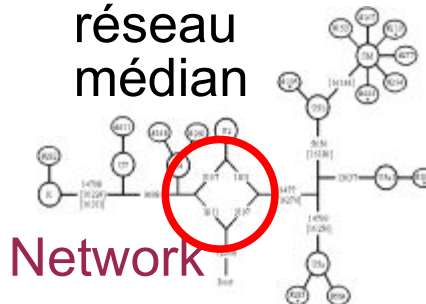
réticulogramme

diagramme de synthèse



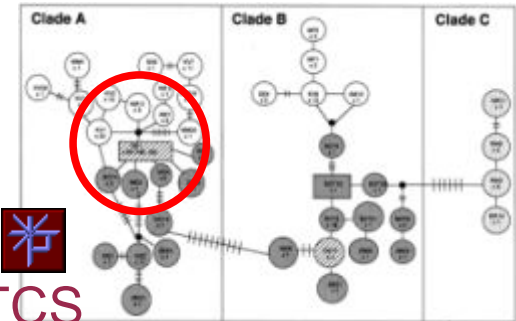
HorizStory

réseau médian



Network

réseau couvrant minimum



TCS

Les réseaux phylogénétiques

Who is Who in Phylogenetic Networks - Articles, Authors & Programs [RSS](#)

Index Browse

[Contribute!](#) [My selection](#)

Search: in [All](#) (word length \geq 3)

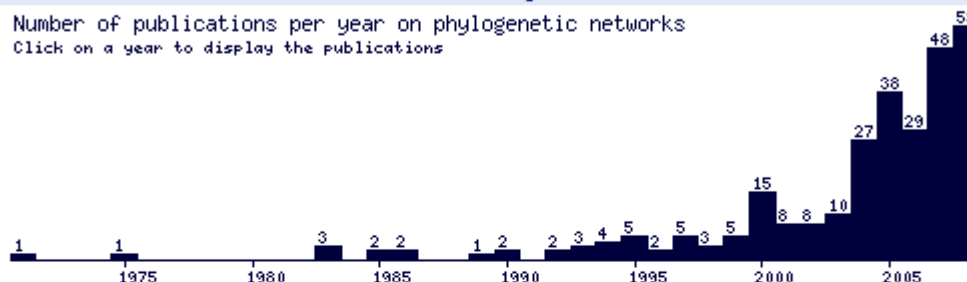
[Login](#)

[Publications - Index](#) ([All 277 publications](#))

Selection by: [Year](#) | [Category](#) | [Keyword](#) | [Author](#)

Selection by Year

Number of publications per year on phylogenetic networks
Click on a year to display the publications



Selection by Category

[Article \(Journal\)](#) (159) [InProceedings](#) (77) [InBook](#) (13)
[PhdThesis](#) (11) [Misc](#) (17) [Programs](#) (37)

Selection by Keyword


[abstract-network](#)(21) [approximation](#)(4) [APX-hard](#)(1) [ARG](#)(5) [block-realization](#)(1) [bootstrap](#)(1) [bound](#)(3) [branch-and-bound](#)(1) [cactus-graph](#)(1) [characterization](#)(3) [clustering](#)(2) [coalescent](#)(5) [consensus](#)(8) [consistency](#)(2) [construction](#)(2) [distance-between-networks](#)(19) [diversity](#)(1) [enumeration](#)(1) [evaluation](#)(22) [explicit-network](#)(23) [FPT](#)(6) [from-clusters](#)(3) [from-distances](#)(18) [from-multilabeled-tree](#)(2) [from-network](#)(5) [from-quartets](#)(5) [from-rooted-trees](#)(26) [from-sequences](#)(17) [from-splits](#)(10) [from-trees](#)(7) [from-triplets](#)(10) [from-unrooted-trees](#)(7) [galled-network](#)(2) [galled-tree](#)(26) [generation](#)(5) [haplotype-network](#)(2) [haplotyping](#)(1) [heuristic](#)(4) [HMM](#)(1) [hybridization](#)(16) [labeling](#)(3) [lateral-gene-transfer](#)(13) [level-f-phylogenetic-network](#)(7) [likelihood](#)(7) [lineage-sorting](#)(1) [MASN](#)(4) [median-network](#)(14) [MedianJoining](#)(2) [minimum-number](#)(8) [minimum-spanning-network](#)(2) [mu-distance](#)(1) [NeighborNet](#)(9) [nested-network](#)(2) [netting](#)(3) [normal-network](#)(1) [NP-complete](#)(13) [optimal-realization](#)(2) [parsimony](#)(10) [perfect](#)(5) [phylogenetic-network](#)(151) [phylogeny](#)(155) [polynomial](#)(26) [Program-Arlequin](#)(5) [Program-Beagle](#)(3) [Program-Bio-PhyloNetwork](#)(4) [Program-CombineTrees](#)(2) [Program-Dendroscope](#)(5) [Program-EEEP](#)(2) [Program-GalledTree](#)(1) [Program-HapBound](#)(1) [Program-HorizStory](#)(2) [Program-HybridNumber](#)(1) [Program-LatTrans](#)(3) [Program-Level2](#)(1) [Program-Marlon](#)(1) [Program-](#)


Les réseaux phylogénétiques




Who is Who in Phylogenetic Networks - Articles, Authors & Programs [RSS](#)

Index **Browse** [Contribute!](#) [My selection](#)

Search: in **All** (word length \geq 3) [Login](#)





Publications of Year << 2008 >> 

<< [Article \(Journal\)](#) >> 

1    





[Gabriel Cardona](#), [Mercè Llabrés](#), [Francesc Rosselló](#) and [Gabriel Valiente](#). [Metrics for phylogenetic networks II: Nodal and triplets metrics](#). 2008. [Comment] [BIBTeX](#)

Keywords: [distance between networks](#), [phylogenetic network](#), [phylogeny](#). **Note:** [Submitted](#). [Annote]

2    

[Cuong Than](#), [Derek Ruths](#) and [Luay Nakhleh](#). [PhyloNet: A Software Package for Analyzing and Reconstructing Reticulate Evolutionary Relationships](#). In *BMC Bioinformatics*, Vol. 9(322), 2008. [Comment] [BIBTeX](#)

Keywords: [Program PhyloNet](#), [software](#). **Note:** <http://dx.doi.org/10.1186/1471-2105-9-322>. [Annote]

3     

[Iyad A. Kanj](#), [Luay Nakhleh](#), [Cuong Than](#) and [Ge Xia](#). [Seeing the Trees and Their Branches in the Network is Hard](#). In *TCS*, Vol. 401:153-164, 2008. [Comment] [BIBTeX](#)

Keywords: [evaluation](#), [from network](#), [from rooted trees](#), [NP-complete](#), [phylogenetic network](#), [phylogeny](#). **Note:** <http://www.cs.rice.edu/~nakhleh/Papers/tcs08.pdf>. [Annote]

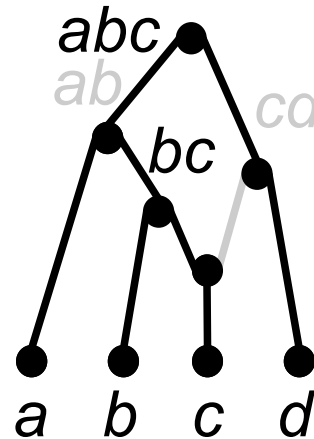
Plan

- Les réseaux phylogénétiques
- **L'arbre en filigrane**
- Motivations de l'approche combinatoire
- Méthodes de reconstruction combinatoire
- Limites des méthodes combinatoires
- Perspectives

L'arbre en filigrane

Modèle de **transmission arborée** de gènes
(gène transmis intégralement)

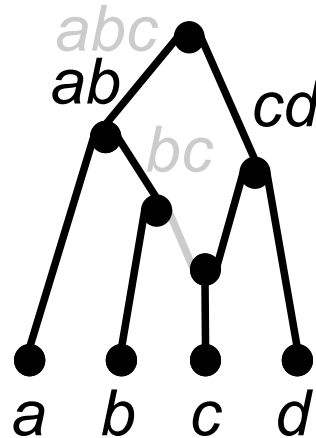
Clades “souples”



L'arbre en filigrane

Modèle de **transmission arborée** de gènes
(gène transmis intégralement)

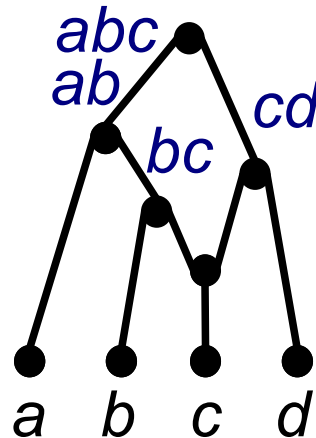
Clades “souples”



L'arbre en filigrane

Modèle de **transmission arborée** de gènes
(gène transmis intégralement)

Clades “souples”

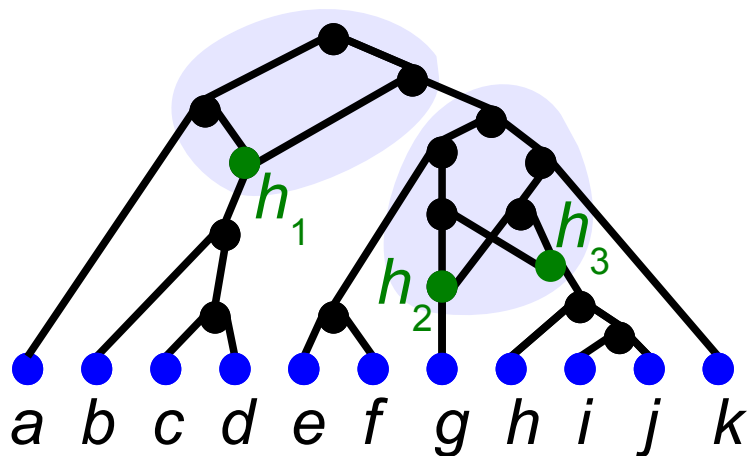


L'ensemble $C(N)$ de **tous les clusters** **souplement compatibles** avec N peut être de taille **exponentielle**.
Test de compatibilité souple **NP-complet**

(Kanj, Nakhleh, Than, Xia, TCS, 2008)

L'arbre en filigrane : réseaux restreints

Algorithmes rapides pour des réseaux à **structure proche d'un arbre**.

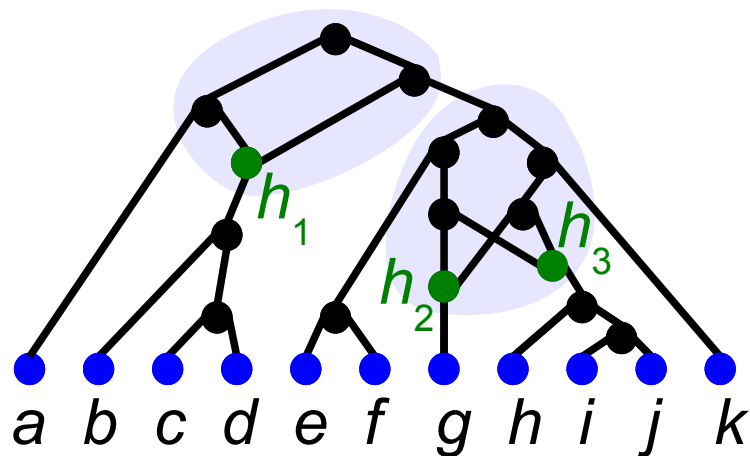


réseau de niveau 2.

niveau = nombre maximum de noeuds de réticulation par partie non arborée.

L'arbre en filigrane : réseaux restreints

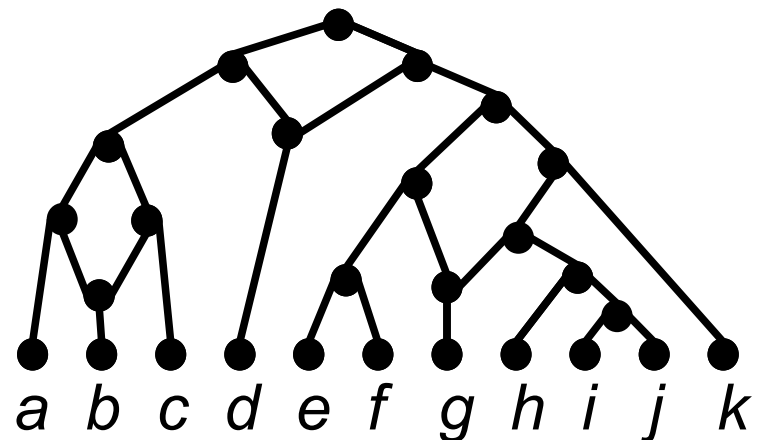
Algorithmes rapides pour des réseaux à **structure proche d'un arbre**.



réseau de niveau 2.

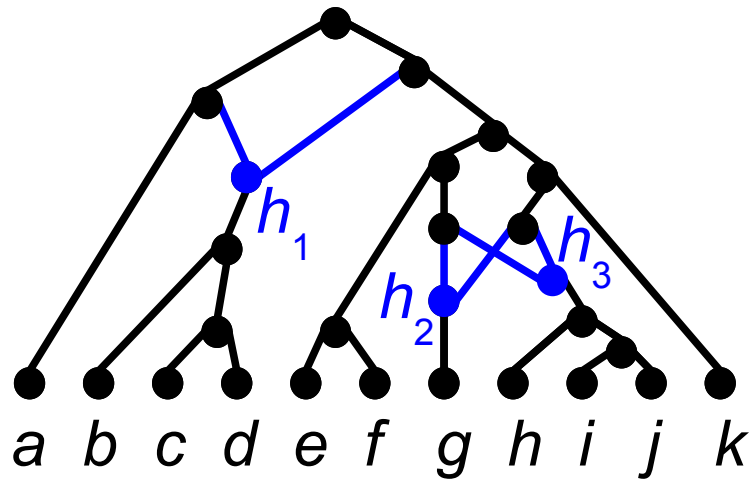
niveau = nombre maximum de noeuds de réticulation par partie non arborée.

réseau de niveau 1
("galled tree")



L'arbre en filigrane : réseaux restreints

Algorithmes rapides pour des réseaux à **structure proche d'un arbre**.

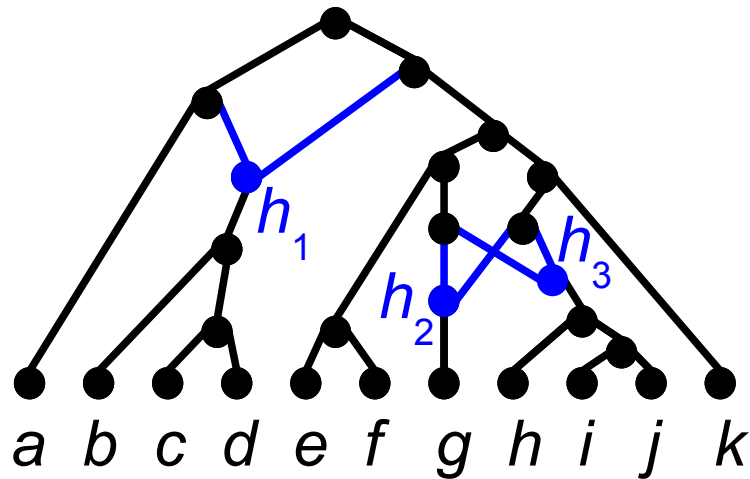


réseau à une couche de réticulation.

réseau à **une couche de réticulation** (“*galled network*”) : la suppression d'un noeud de réticulation déconnecte le réseau.

L'arbre en filigrane : réseaux restreints

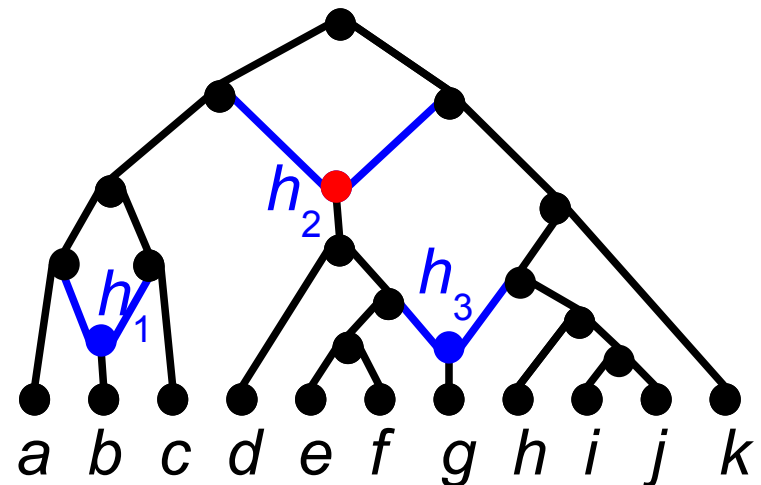
Algorithmes rapides pour des réseaux à **structure proche d'un arbre**.



réseau à une couche de réticulation.

réseau à **une couche de réticulation** (“*galled network*”) : la suppression d'un noeud de réticulation déconnecte le réseau.

réseau à deux couches de réticulation.



Plan

- Les réseaux phylogénétiques
- L'arbre en filigrane
- **Motivations de l'approche combinatoire**
- Méthodes de reconstruction combinatoire
- Limites des méthodes combinatoires
- Perspectives

Reconstruction de réseaux

{séquences de gènes}



méthodes de distance

Bandelt & Dress 1992 - Legendre & Makarenkov 2000 - Bryant & Moulton 2002

méthodes de parcimonie

Hein 1990 - Kececioglu & Gusfield 1994 - Jin, Nakhleh, Snir, Tuller 2009

méthodes de vraisemblance

Snir & Tuller 2009 - Jin, Nakhleh, Snir, Tuller 2009 - Velasco & Sober 2009

réseau *N*

Reconstruction de réseaux

**Problème : méthodes généralement lentes,
explosion du nombre de séquences.**

{séquences de gènes}

méthodes de distance

*Bandelt & Dress 1992 - Legendre &
Makarenkov 2000 - Bryant & Moulton 2002*

méthodes de parcimonie

*Hein 1990 - Kececioglu & Gusfield 1994 -
Jin, Nakhleh, Snir, Tuller 2009*

méthodes de vraisemblance

*Snir & Tuller 2009 - Jin, Nakhleh, Snir,
Tuller 2009 - Velasco & Sober 2009*



réseau N

Reconstruction combinatoire de réseaux

{séquences de gènes}



*Reconstruction d'un arbre
par ensemble de gènes
homologues*

phylome = {arbres}



*Réconciliation ou
consensus d'arbres*

super-réseau N

Reconstruction combinatoire de réseaux

{séquences de gènes}



*Reconstruction d'un arbre
par ensemble de gènes
homologues*

phylome = {arbres}



*Réconciliation ou
consensus d'arbres*

super-réseau N

**Problème : le consensus d'arbres est un
problème NP-complet pour 2 arbres**

Triplets/quadruplets, splits/clades

Problème :

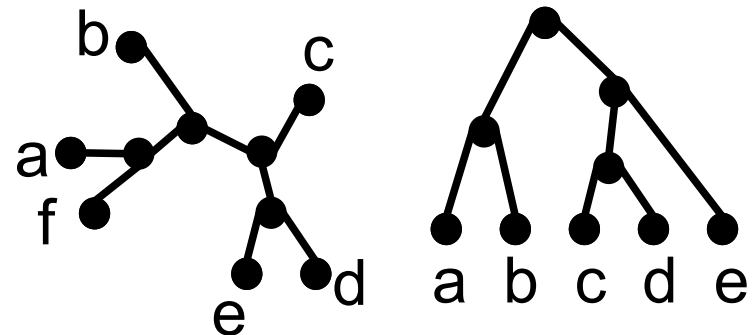
Reconstruire le **super-réseau** d'un ensemble d'arbres est **difficile**.

Idée :

reconstruire un réseau contenant tous les :

triplets
quadruplets
clades
splits

des arbres en entrée ?



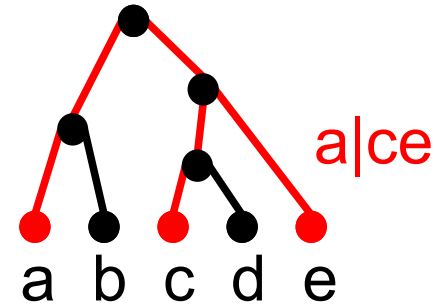
Motivations algorithmiques !

Triplets/quadruplets, splits/clades

Idée :

reconstituer un réseau contenant tous les :

triplets



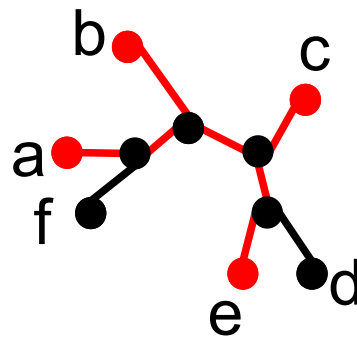
des arbres en entrée ?

Triplets/quadruplets, splits/clades

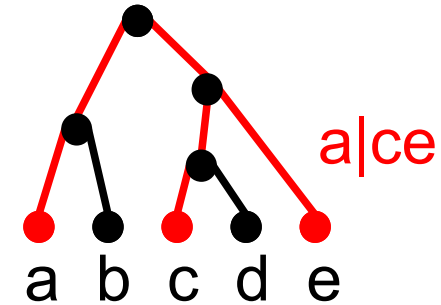
Idée :

reconstituer un réseau contenant tous les :

ab|ce



triplets



a|ce

quadruplets

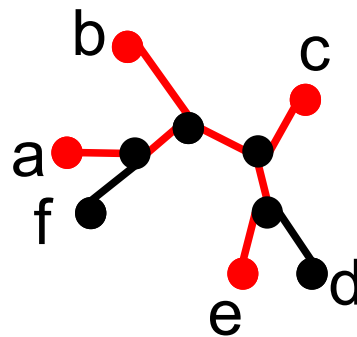
des arbres en entrée ?

Triplets/quadruplets, splits/clades

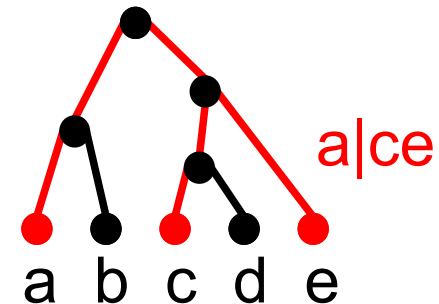
Idée :

reconstituer un réseau contenant tous les :

ab|ce

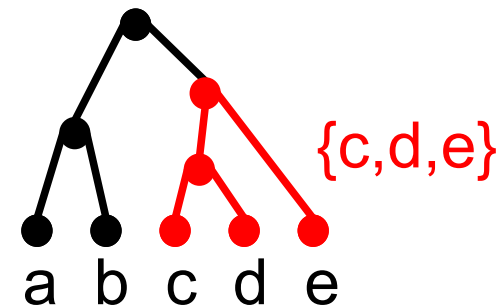


triplets



quadruplets

clades



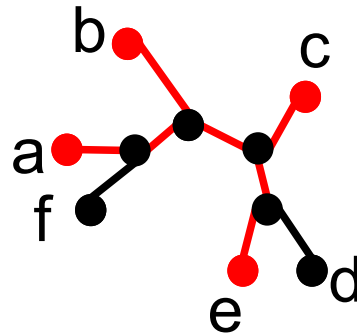
des arbres en entrée ?

Triplets/quadruplets, splits/clades

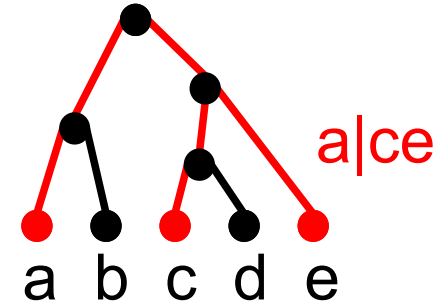
Idée :

reconstituer un réseau contenant tous les :

$ab|ce$

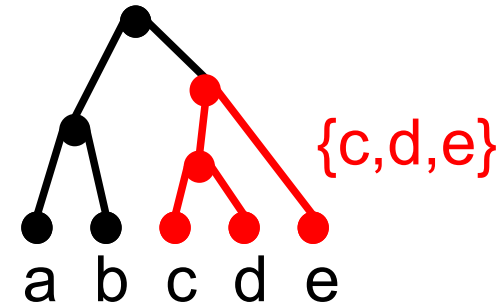


triplets

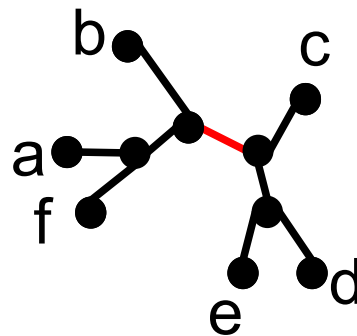


quadruplets

clades



$\{a,b,f\}$
 $\{c,d,e\}$



splits

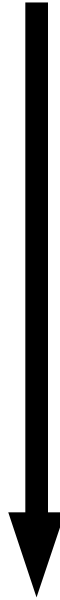
des arbres en entrée ?

Triplets/quadruplets, splits/clades

Idée :

modifier le type de données à traiter

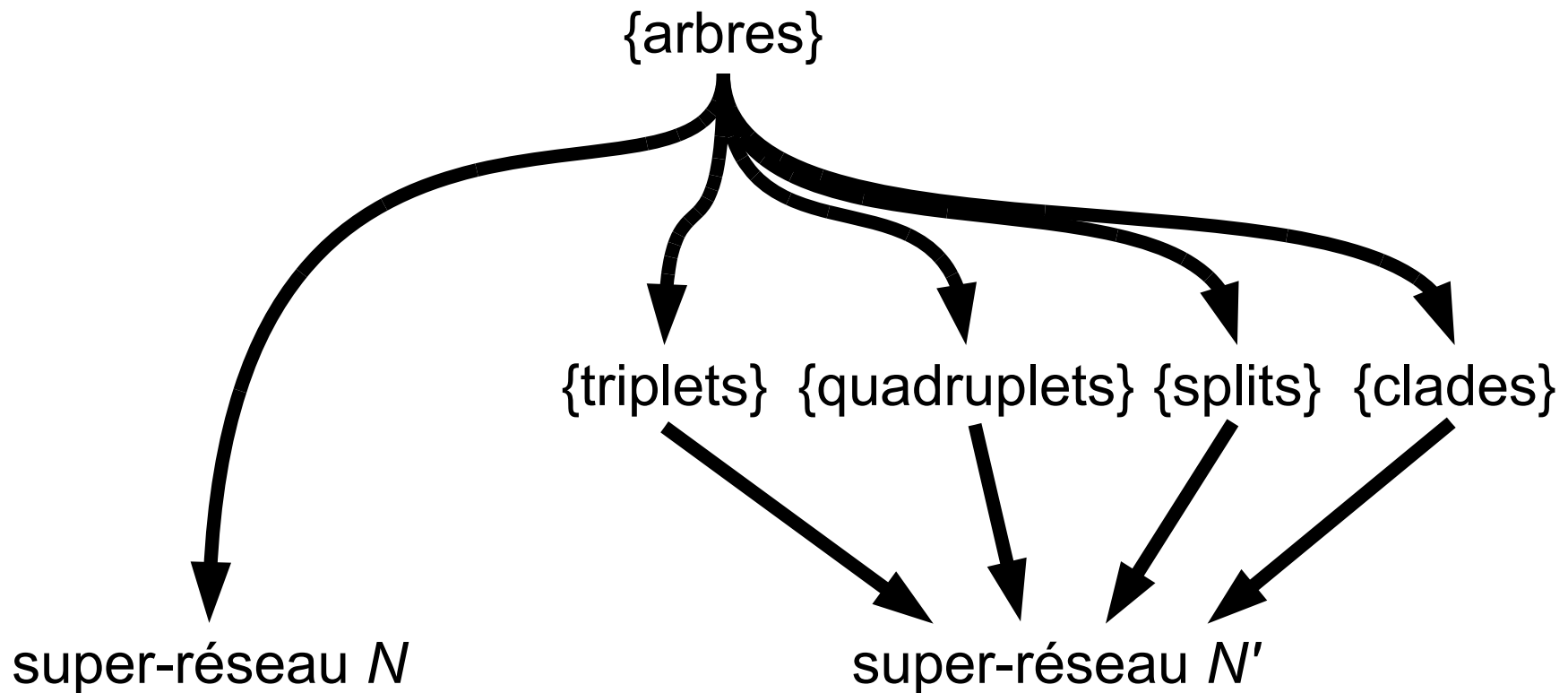
{arbres}



super-réseau N

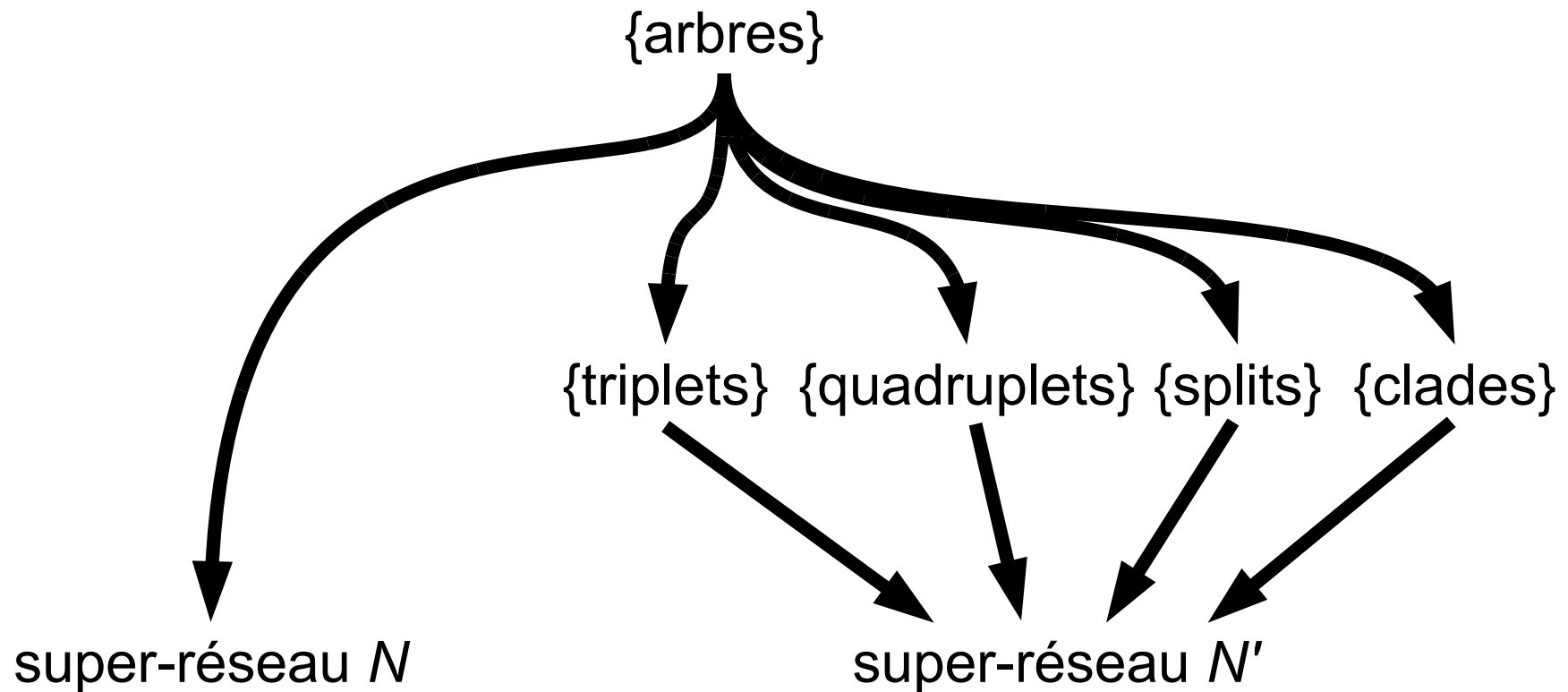
Triplets/quadruplets, splits/clades

Idée :
modifier le type de données à traiter



Triplets/quadruplets, splits/clades

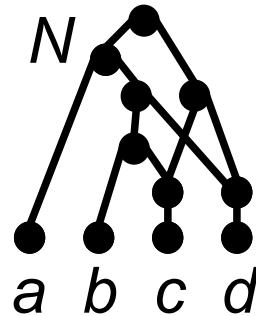
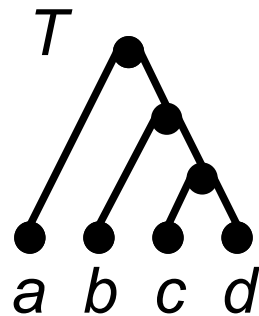
Idée :
modifier le type de données à traiter



$N' = N ?$

Triplets/quadruplets, splits/clades

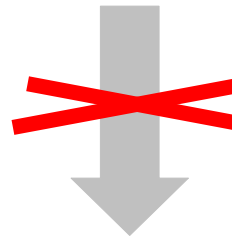
Un réseau compatible avec l'ensemble de **tous les triplets** d'un arbre T n'est pas forcément compatible avec T .



compatible avec
 $\{a|bc, a|bd, a|cd, b|cd\}$
mais pas avec T

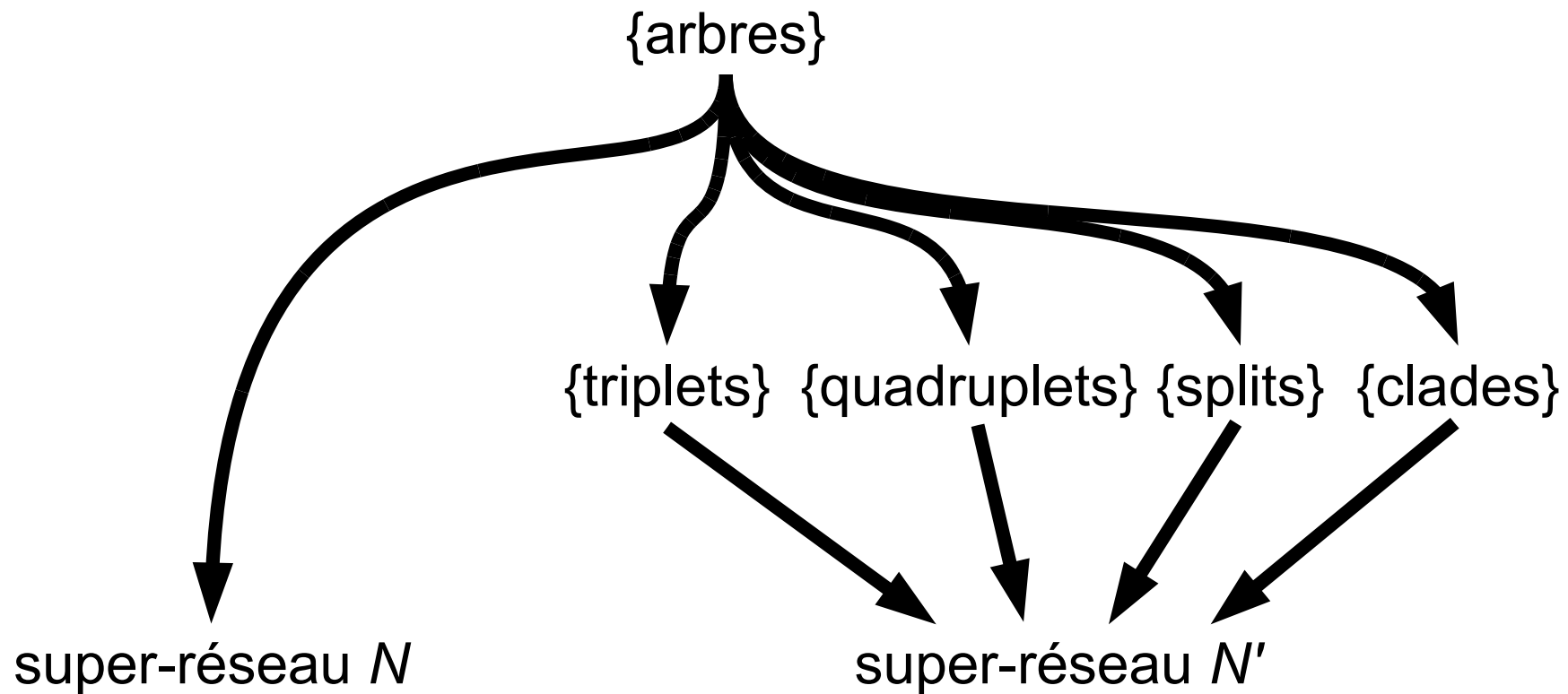
compatible avec
 $\{abcd, bcd, cd, a, b, c\}$
mais pas avec T

compatible avec les clades d'un arbre T



compatible avec T .

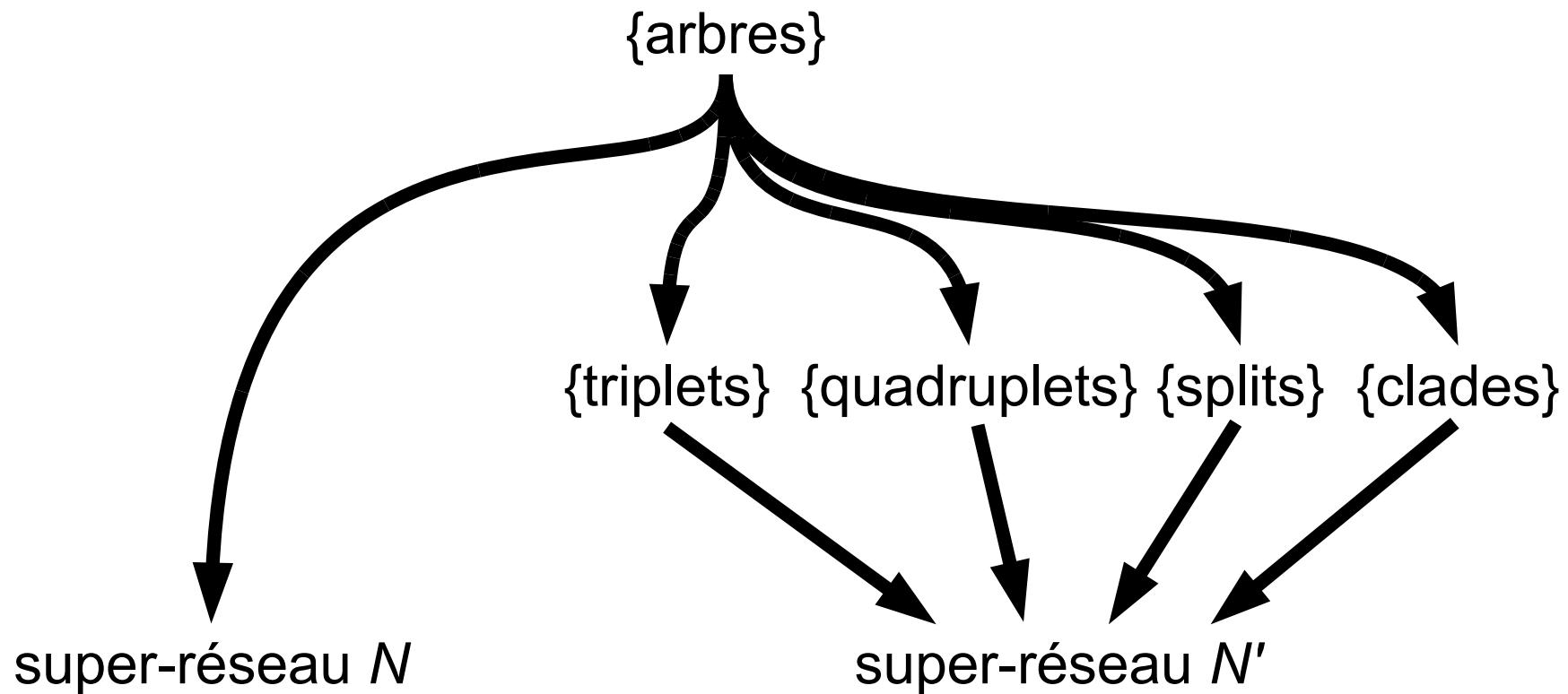
Triplets/quadruplets, splits/clades



$N' = N$?

Pas nécessairement, mais :
 N' complexe \longrightarrow N complexe
 N contient également les triplets, quadruplets...

Triplets/quadruplets, splits/clades



$N' = N ?$

Pas nécessairement, mais :
 N' peut être intéressant en soi...

Plan

- Les réseaux phylogénétiques
- L'arbre en filigrane
- Motivations de l'approche combinatoire
- **Méthodes de reconstruction combinatoire**
- Limites des méthodes combinatoires
- Perspectives

Reconstruction depuis les triplets

{arbres}

Méthodes exactes rapides pour reconstruire un **réseau de niveau 1 et 2** (s'il en existe un) à partir d'un ensemble dense de **triplets**

(Jansson, Nguyen & Sung, SODA'05 : $O(n^3)$ pour niveau 1)
(van Iersel, Kelk & al, RECOMB'08 : $O(n^8)$ pour niveau 2)

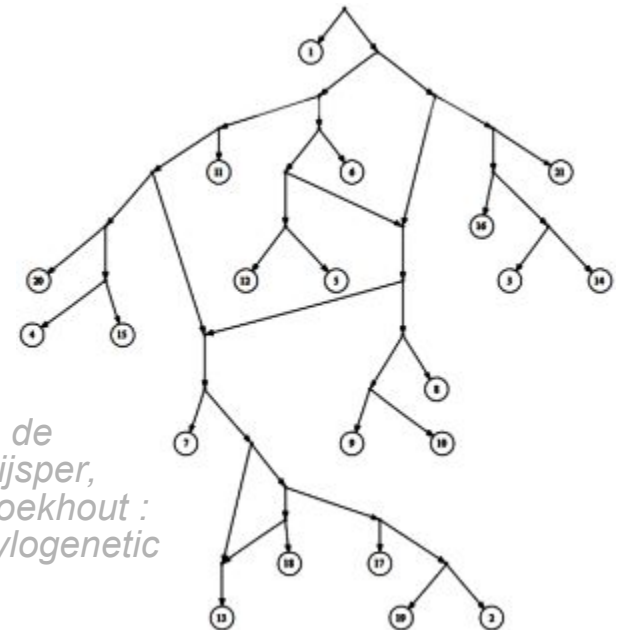
{triplets}

dense =
sur chaque ensemble de 3 feuilles, au moins 1 triplet existe dans T .

Programme Simplistic



N' réseau
de niveau 1
ou 2



Réseau phylogénétique de levures - Van Iersel, Keijsper, Kelk, Stougie, Hagen Boekhout : Constructing level-2 phylogenetic networks from triplets. Recomb 2008

Reconstruction depuis les clades souples

{arbres}



{clades}



N' réseau
“galled
network”

Consensus de clades souples :

Dendroscope 

(Huson, Dezulian, Franz, Rausch, Richter & Rupp, 2007)

Méthode exacte rapide de reconstruction de
réseaux à 1 couche de réticulation à partir
de **clades souples**

(Huson, Rupp, Berry, Gambette & Paul, 2009)

2 étapes :

- choix du plus gros sous-ensemble de taxons où les clades sont compatibles avec un arbre
- ajout du minimum de réticulations pour connecter les autres taxons

Reconstruction depuis les clades souples

{arbres}



{clades}

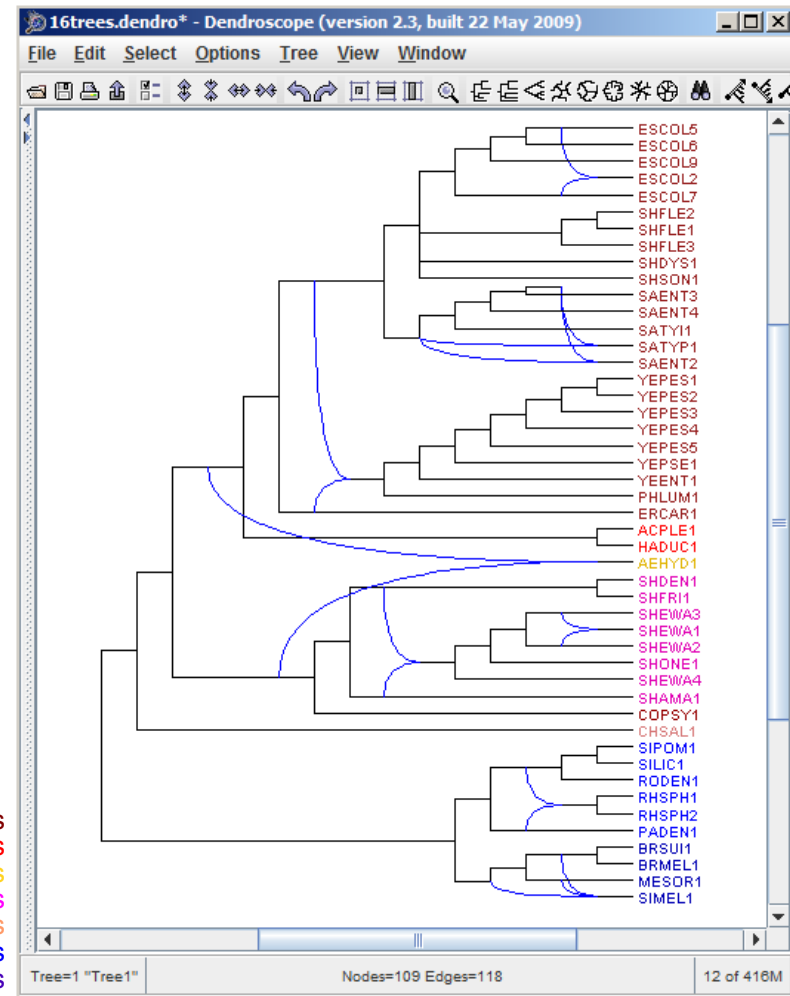


N' réseau
"galled
network"

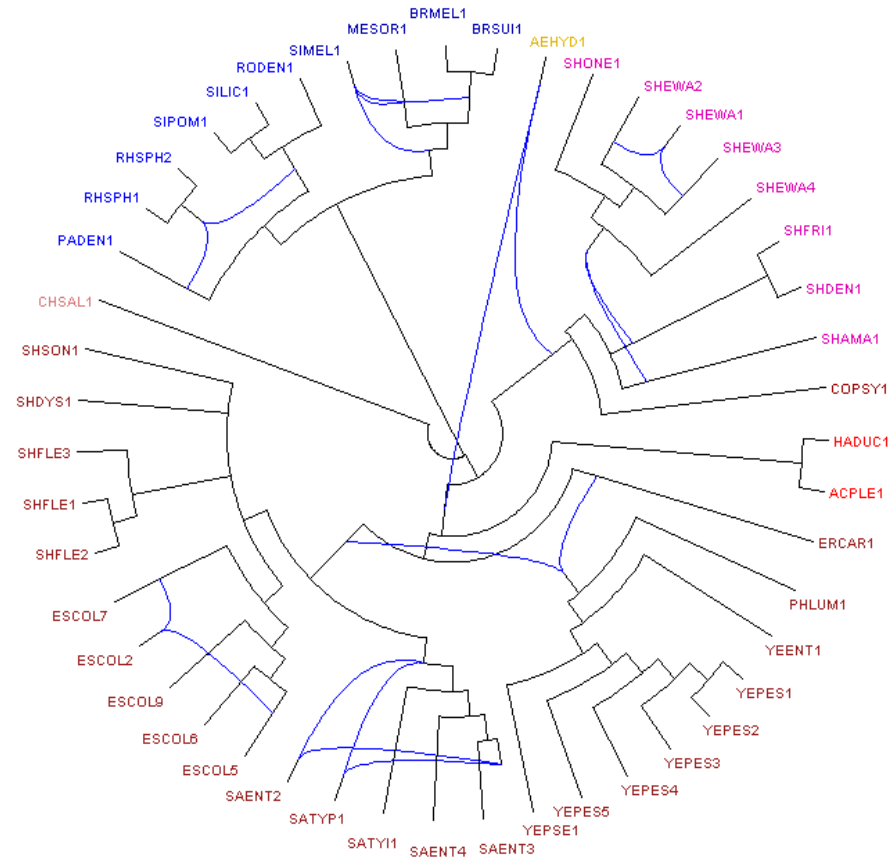
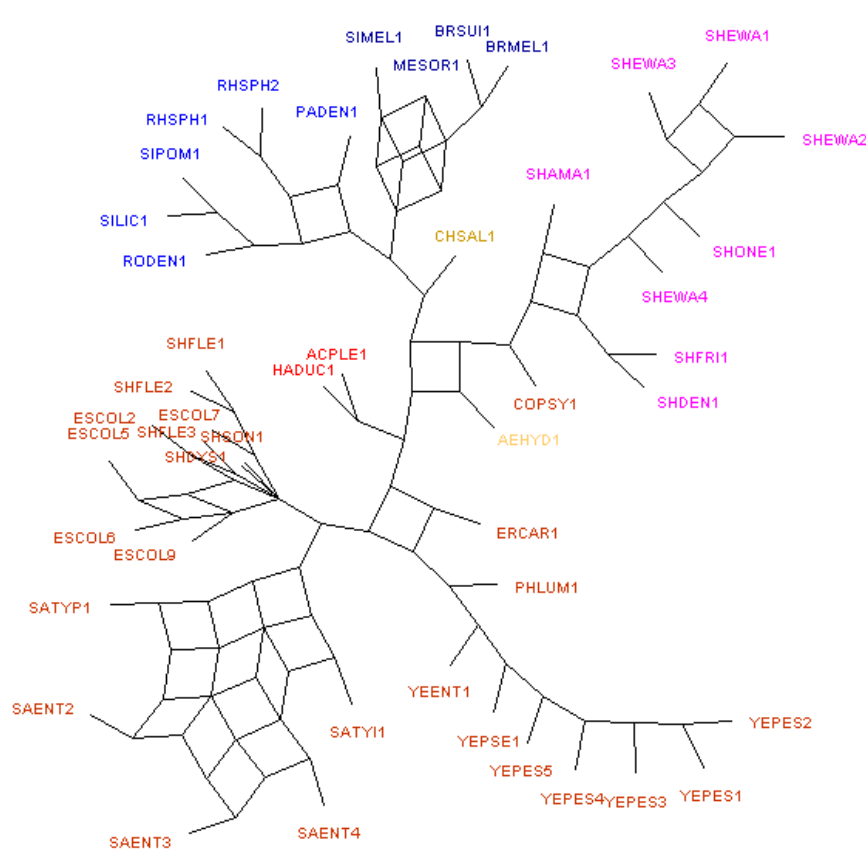
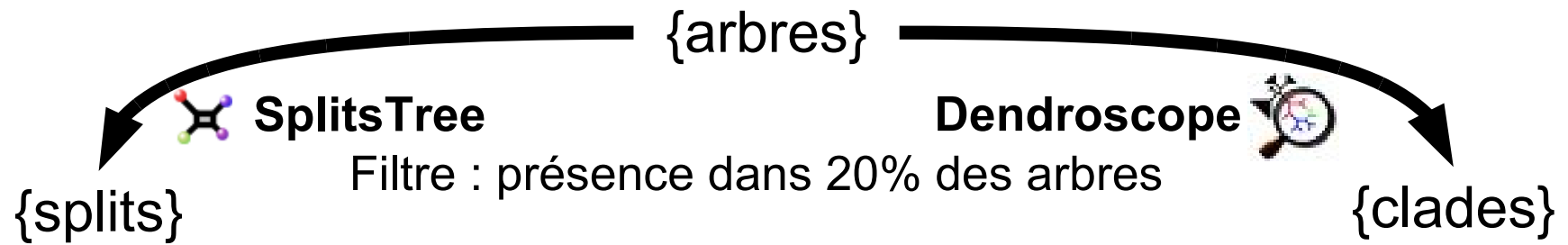
16 arbres de gènes de 46 espèces de bactéries
Réseau "galled network" des clades apparaissant
dans 20% des arbres :

Dendroscope 

Enterobacteriales
Pasteurellales
Aeromonadales
Alteromonadales
Oceanospirillales
Rhodobacterales
Rhizobiales



Reconstruction depuis les clades souples



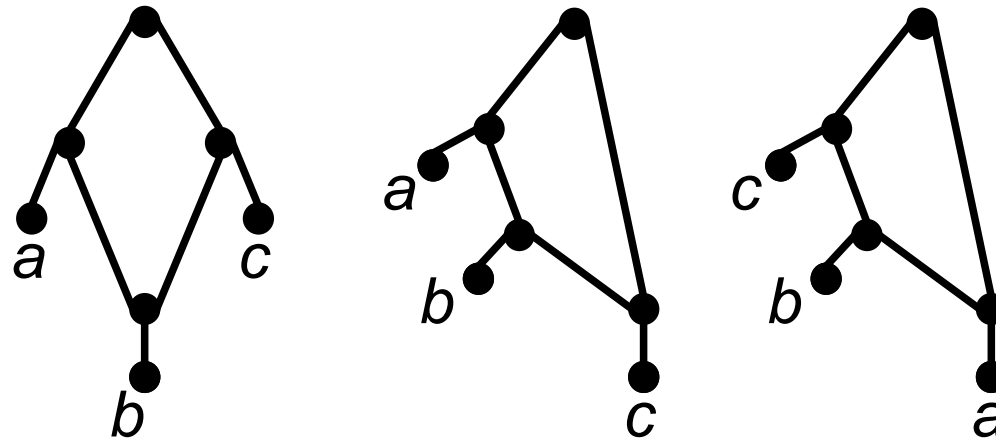
Plan

- Les réseaux phylogénétiques
- L'arbre en filigrane
- Motivations de l'approche combinatoire
- Méthodes de reconstruction combinatoire
- **Limites des méthodes combinatoires**
- Perspectives

Limites des approches combinatoires

Plusieurs réseaux minimaux ont exactement le même ensemble d'arbres, de triplets, de clades.

Gambette & Huber, 2009

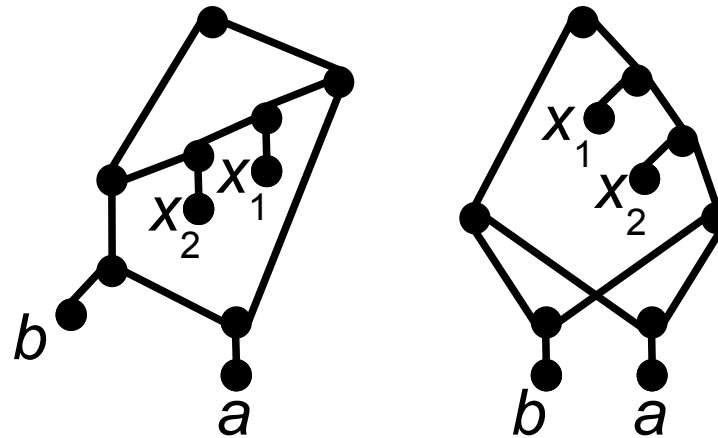


Caractérisation des réseaux de niveau 1 ayant exactement le même ensemble d'arbres, de triplets, de clades.

Limites des approches combinatoires

Plusieurs réseaux minimaux ont exactement le même ensemble d'arbres, de triplets, de clades.

Gambette & Huber, 2009

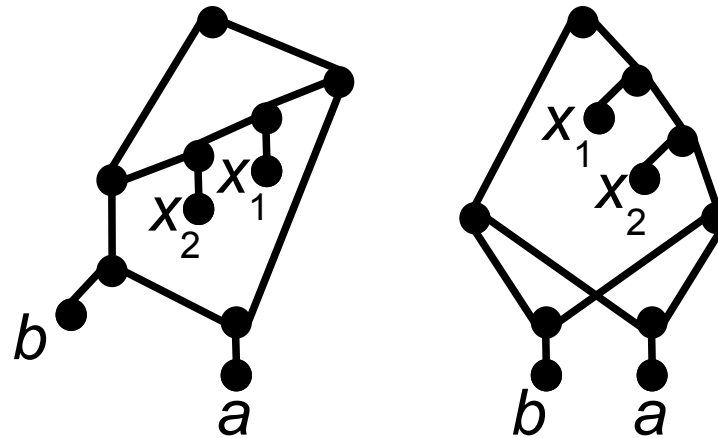


2 réseaux de niveau 2 avec le même ensemble de triplets

Limites des approches combinatoires

Plusieurs réseaux minimaux ont exactement le même ensemble d'arbres, de triplets, de clades.

Gambette & Huber, 2009



2 réseaux de niveau 2 avec le même ensemble de triplets

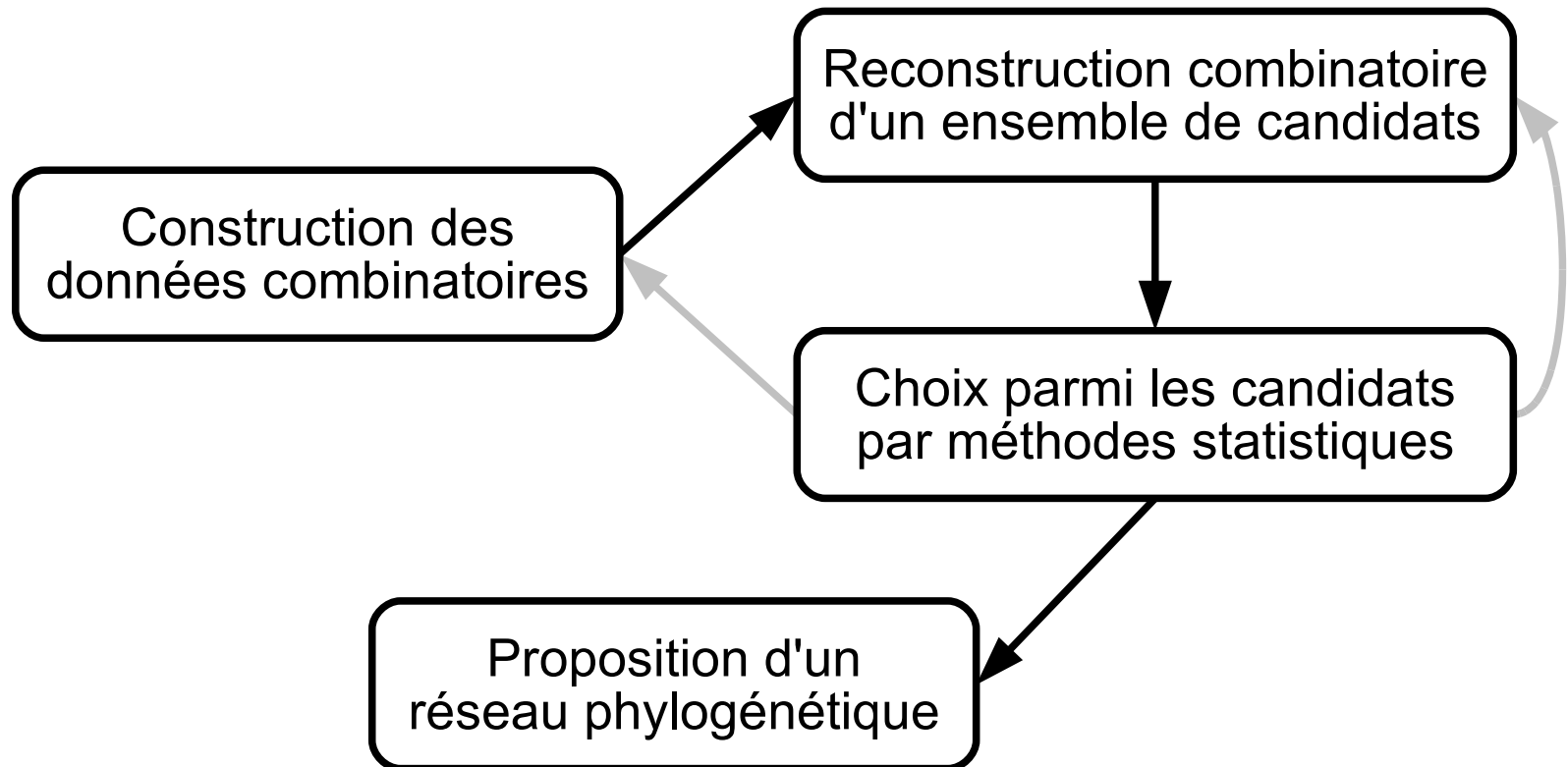
Même avec des données de départ **complètes** et **correctes**,
il est impossible de choisir entre les formes ambiguës !

Plan

- Les réseaux phylogénétiques
- L'arbre en filigrane
- Motivations de l'approche combinatoire
- Méthodes de reconstruction combinatoire
- Limites des méthodes combinatoires
- **Perspectives**

Perspectives

Reconstruction du réseau par un dialogue entre méthodes combinatoires et méthodes statistiques

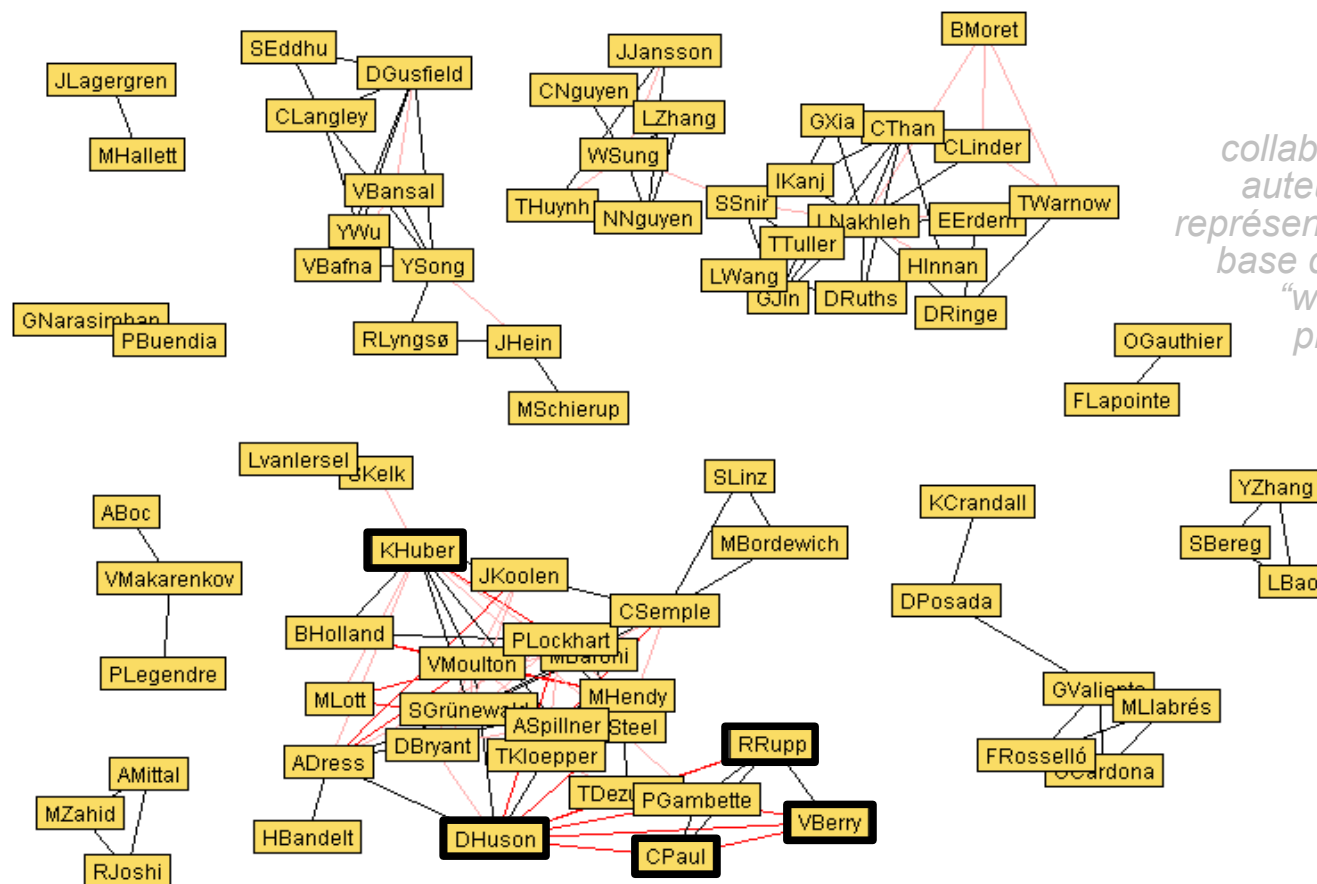


Des questions ?

Merci pour votre attention !

Co-auteurs des résultats présentés :

- Vincent Berry, Christophe Paul (LIRMM, Université Montpellier 2)
- Katharina Huber (Université East Anglia, UK)
- Daniel Huson, Regula Rupp (Université de Tübingen, Allemagne)



Grphe de collaboration des auteurs les plus représentés dans la base de données "who's who in phylogenetic networks".