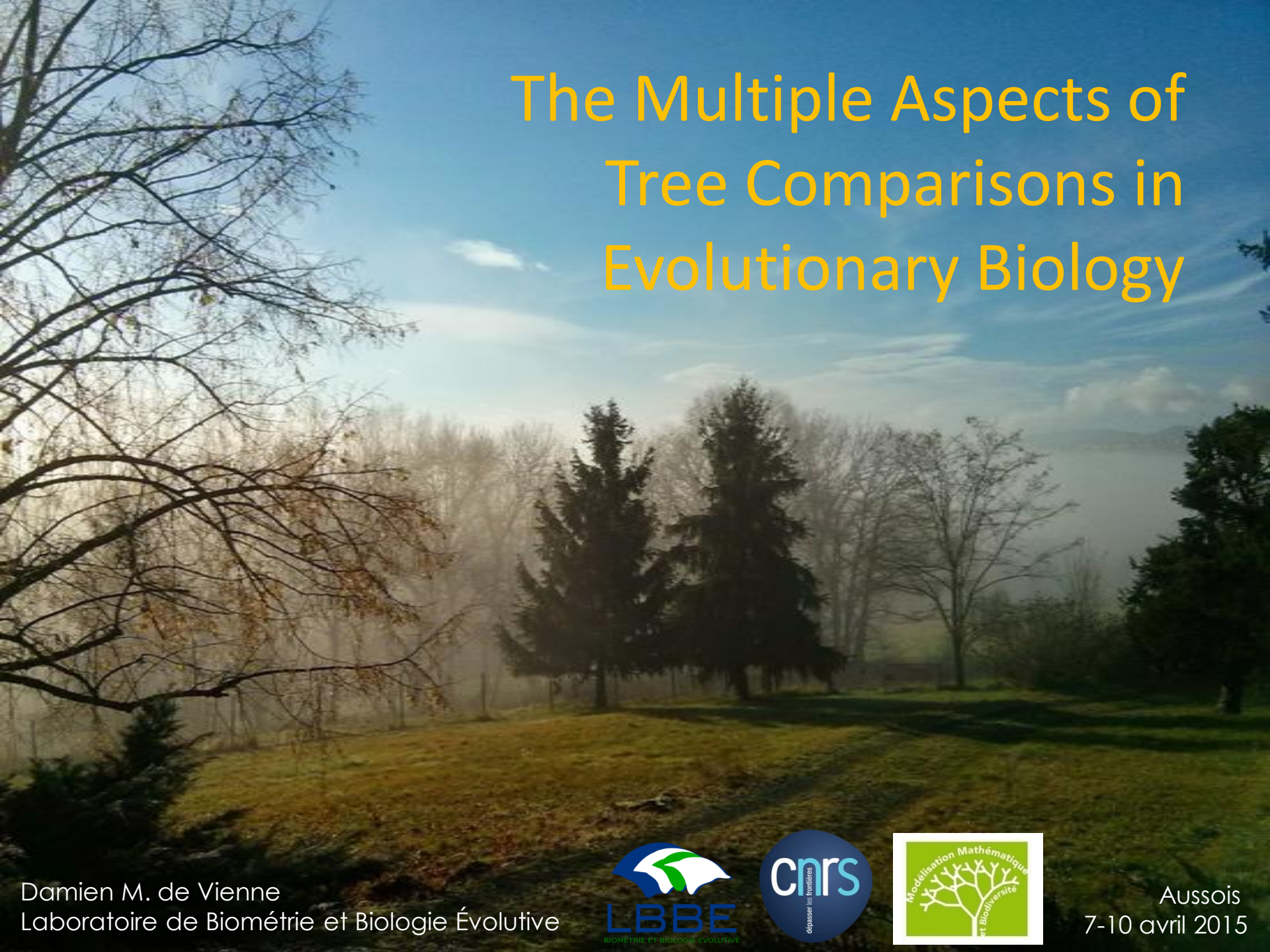


The Multiple Aspects of Tree Comparisons in Evolutionary Biology

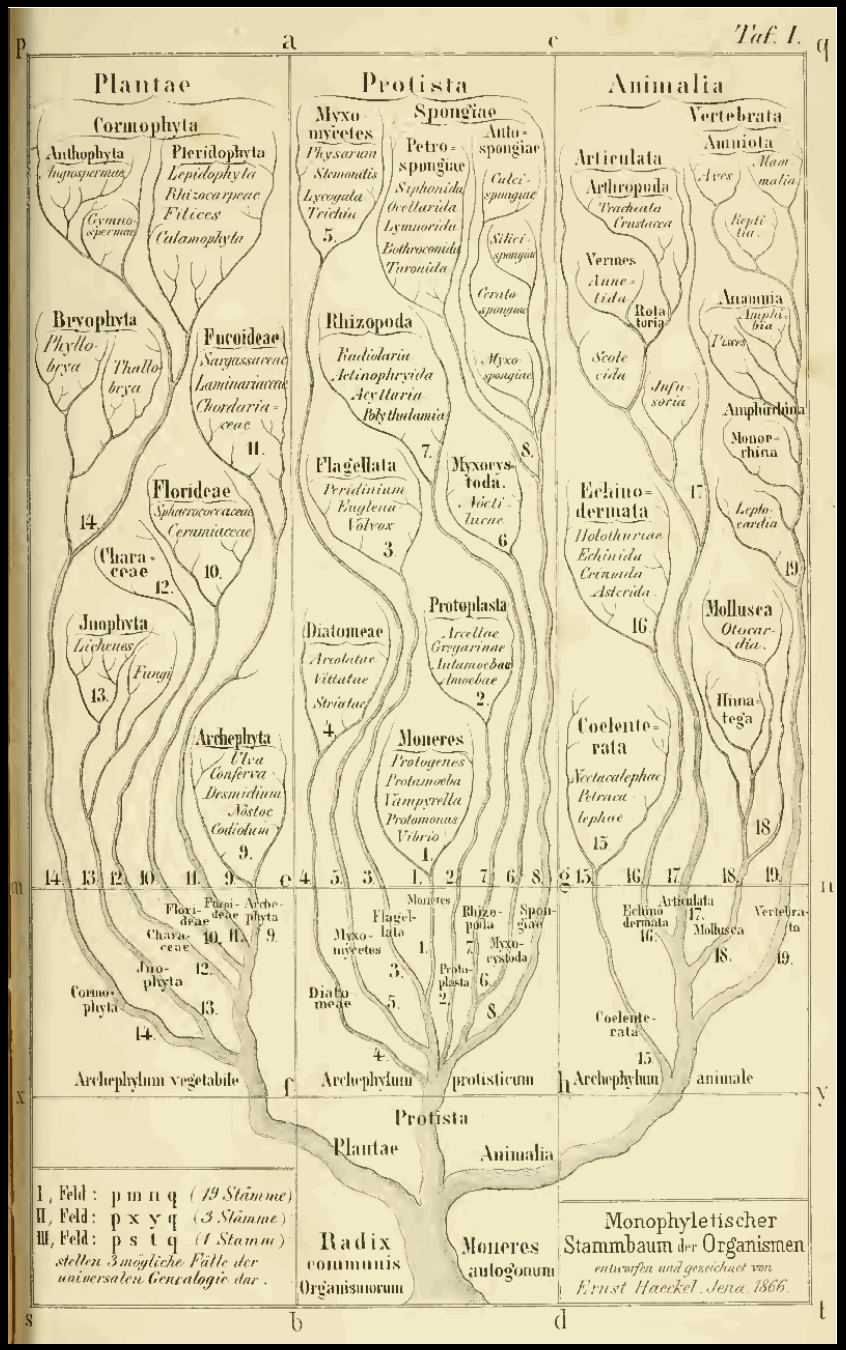
Damien M. de Vienne
Laboratoire de Biométrie et Biologie Évolutive

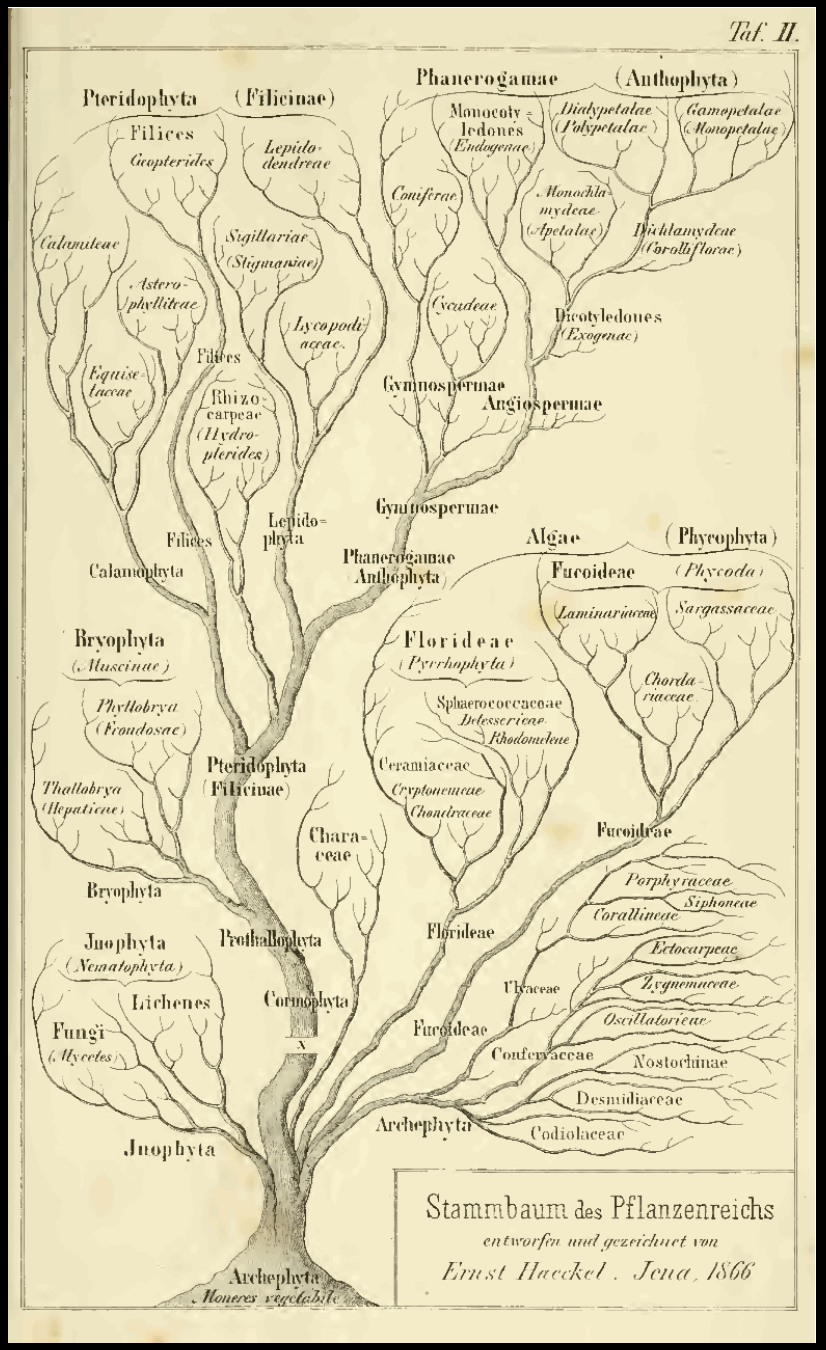
Aussois
7-10 avril 2015

phylogeny

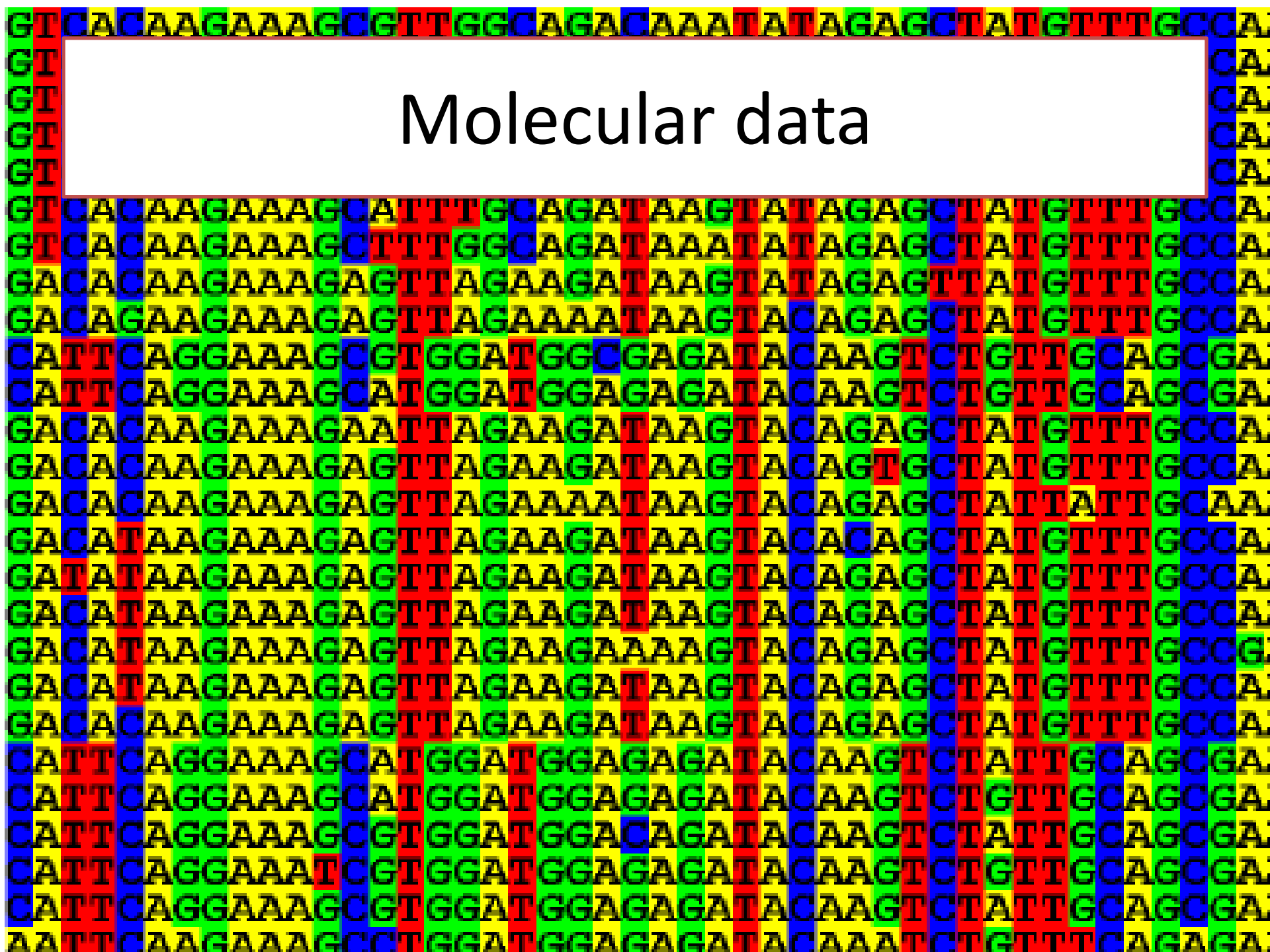
- "genesis and evolution of a phylum," 1869, from German Phylogenie, coined 1866 by German biologist Ernst Heinrich Haeckel (1834-1919) from Greek phylon "race" (see phylo-) + -geneia "origin"

Darwin, notebook B, 1837

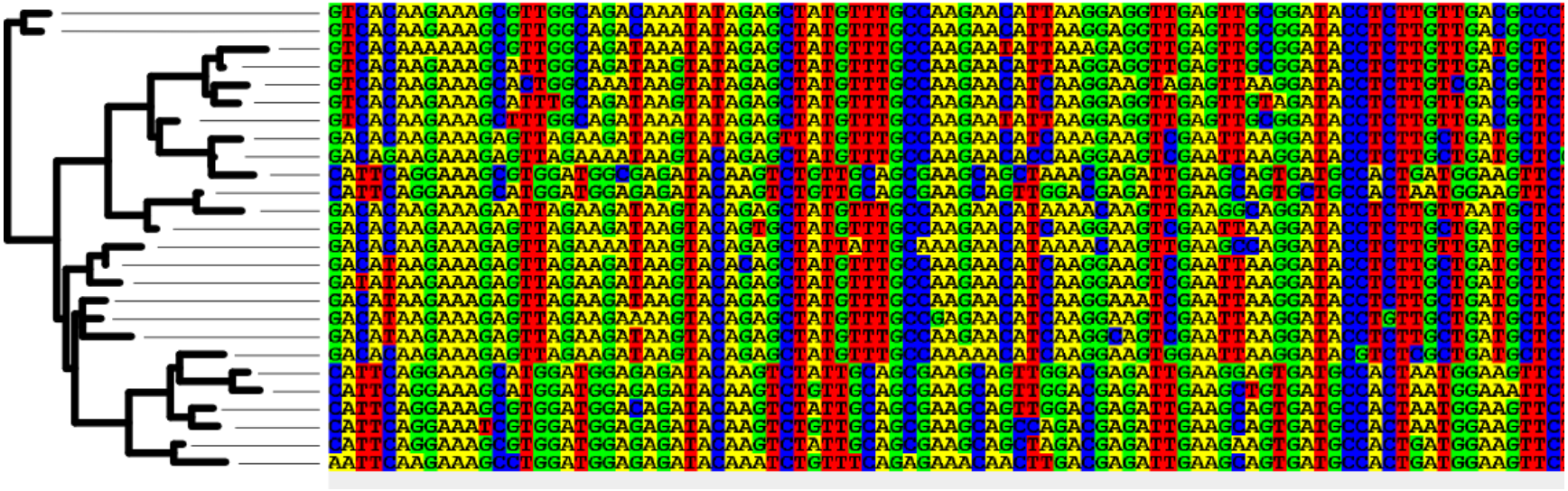

Taf. I.
Plantae
Protista
Animalia
Cormophyta
Anthophyta
Angiospermae
Gymnospermae
Pteridophyta
Lepidophyta
Rhizocarpeae
Filices
Calamophyta
Bryophyta
Phyllobrya
Thallobrya
Fucoideae
Sargassaceae
Laminariaceae
Chordariaceae
Florideae
Sphaerococcaceae
Ceramiaceae
Characeae
Juophyta
Lichenes
Fungi
Archephyta
Ulva
Conferva
Desmidium
Nostoc
Codiolum
Myxomycetes
Physarum
Stemonitis
Lycogala
Trichia
Spongiae
Petrospongiae
Siphonida
Ocellarida
Lymnorida
Bothroconida
Turonida
Autospongiae
Calcispongiae
Silicispongiae
Cornispongiae
Myxospongiae
Rhizopoda
Radiolaria
Actinophryida
Acyttaria
Polythalamia
Flagellata
Peridinium
Euglena
Volvox
Myxomycetoda.
Noctiluca
Diatomeae
Arcolatae
Vittatae
Striatae
Protoplasta
Arcellae
Gregarinae
Antamoebae
Amoebae
Moneres
Protogenes
Protamoeba
Vampyrella
Protomonas
Vibrio
Vertebrata
Anniota
Mammalia
Aves
Reptilia
Anamnia
Amphibia
Pisces
Amphioxus
Monorhina
Leptocardia
Articulata
Arthropoda
Tracheata
Crustacea
Vermes
Annelida
Rotatoria
Scolecida
Infusoria
Echinodermata
Holothuriae
Echinida
Crinoida
Asterida
Coelenterata
Nectacalephae
Petraca
lephae
Mollusca
Otocardia.
Hinnatega
Archephylum vegetabile
Archephylum protisticum
Archephylum animale
Protista
Plantae
Animalia
1. Feld: p m n q (19 Stämme)
II. Feld: p x y q (3 Stämme)
III. Feld: p s t q (1 Stamm)
stellen 3 mögliche Fälle der universalen Genealogie dar.
Radix communis Organismorum
Moneres autogonum
Monophyletischer Stammbaum der Organismen
entworfen und gezeichnet von
Ernst Haeckel. Jena. 1866.


Taf. II.
Pteridophyta (Filicinae)
Filices
Geopterides
Lepidodendreae
Calamiteae
Asterophylliteae
Sigillariae (Stigmariae)
Lycopodiaceae.
Equisetaceae
Rhizocarpeae (Hydropterides)
Lepidophyta
Calamophyta
Phanerogamae (Anthophyta)
Monocotyledones (Endogenae)
Dialypetalae (Polypetalae)
Gamopetalae (Monopetalae)
Coniferae
Monochlamydeae (Apetalae)
Dichlamydeae (Corolliflorae)
Cycadeae
Dicotyledones (Exogenae)
Gymnospermae
Angiospermae
Phanerogamae Anthophyta
Algae (Phycophyta)
Fucoideae (Phycoidae)
Laminariaceae
Sargassaceae
Bryophyta (Muscinae)
Florideae (Pyrrhophyta)
Phyllobrya (Frondosae)
Sphaerococcaceae Delesseriae Rhodomelae
Chordaceae
Ceramiaceae
Cryptonemeae Chondriaceae
Thallobrya (Hepaticae)
Pteridophyta Filicinae
Characeae
Fucoideae
Bryophyta
Porphyraceae
Siphoneae
Corallineae
Florideae
Ectocarpeae
Juophyta (Nematophyta)
Protophyta
Ulvaceae
Zygnemaceae
Lichenes
Cormophyta
Oscillatorieae
Fucoideae
Fungi (Myycetes)
Confervaceae
Nostochinae
Desmidiaceae
Archephyta
Codiolaceae
Juophyta
Archephyta Moneres vegetabile
Stammbaum des Pflanzenreichs
entworfen und gezeichnet von
Ernst Haeckel. Jena. 1866

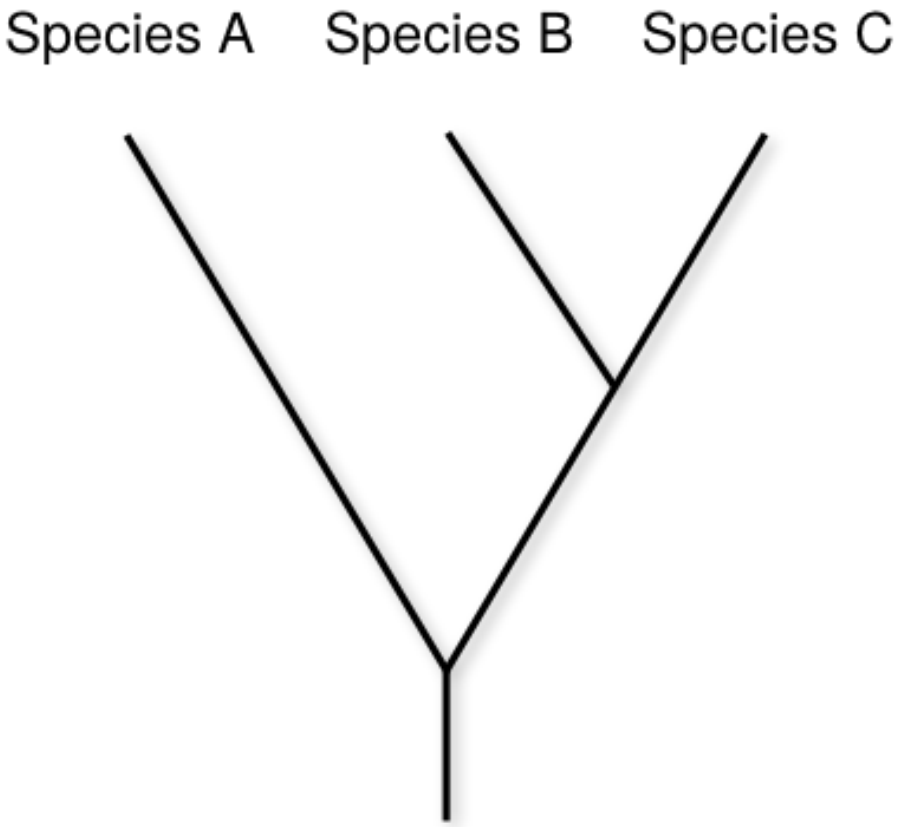
Haekel, 1866


Molecular data

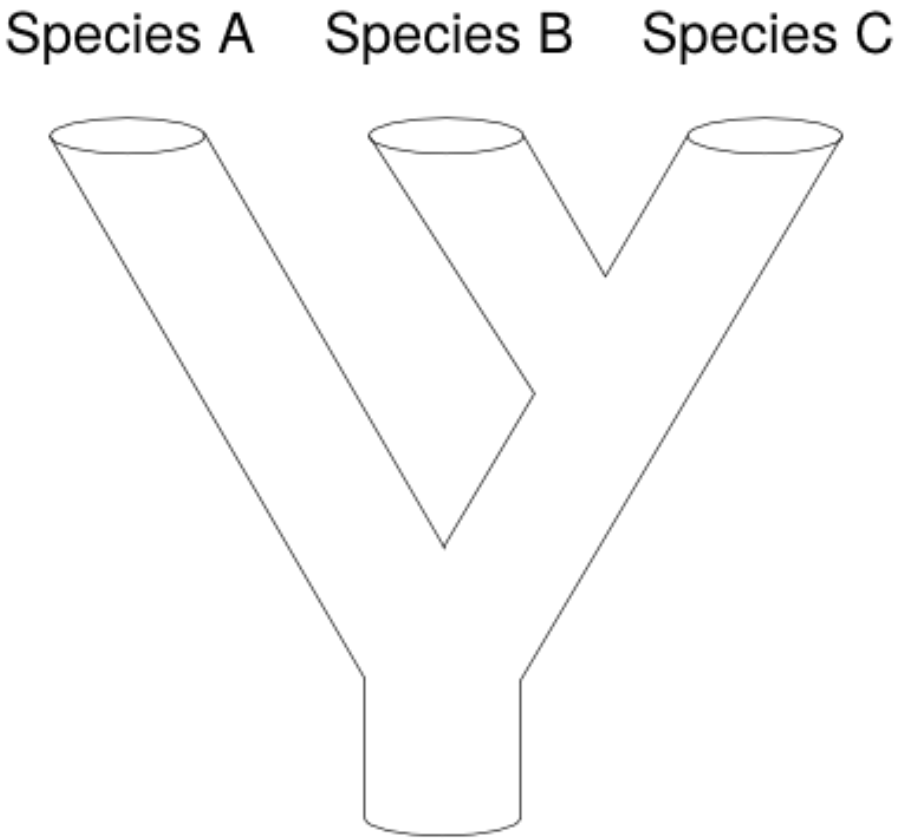
Molecular Phylogenies



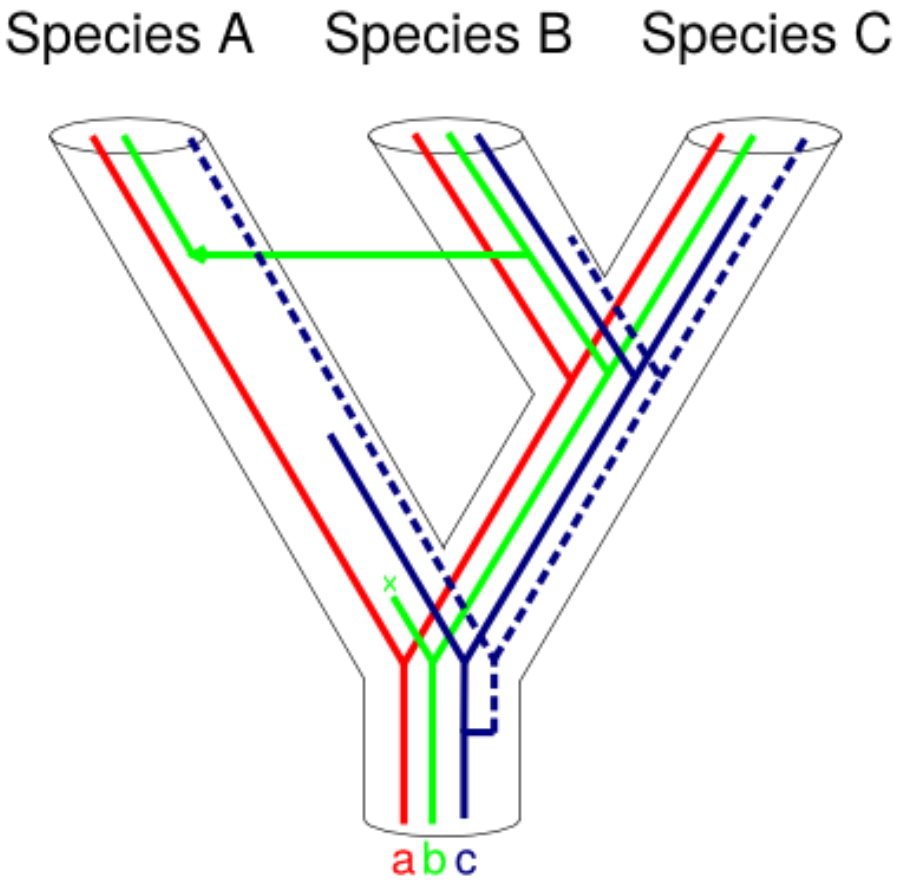
Species trees vs gene trees



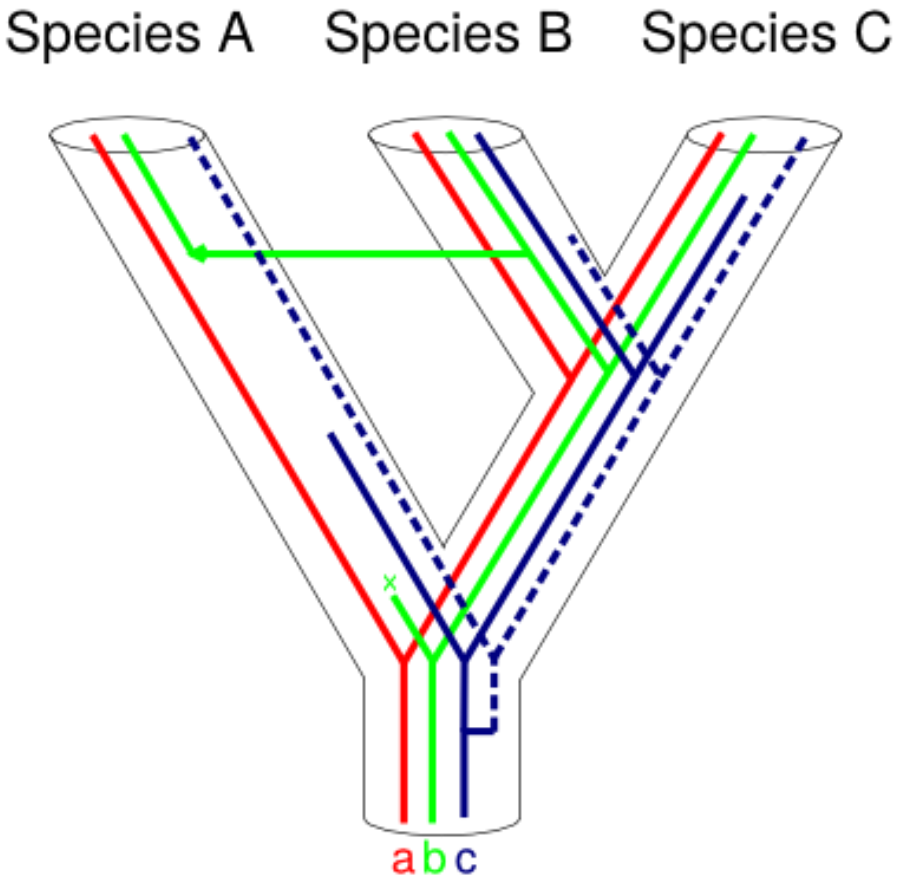
Species trees vs gene trees

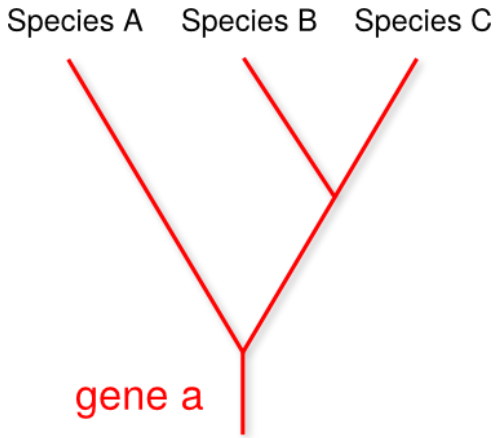


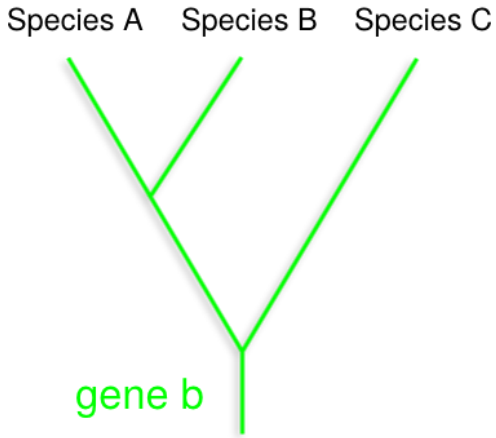
Species trees vs gene trees

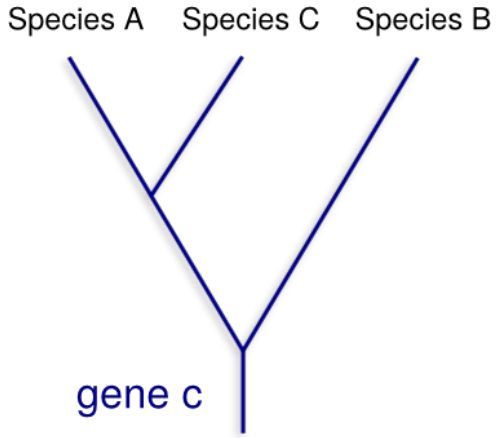


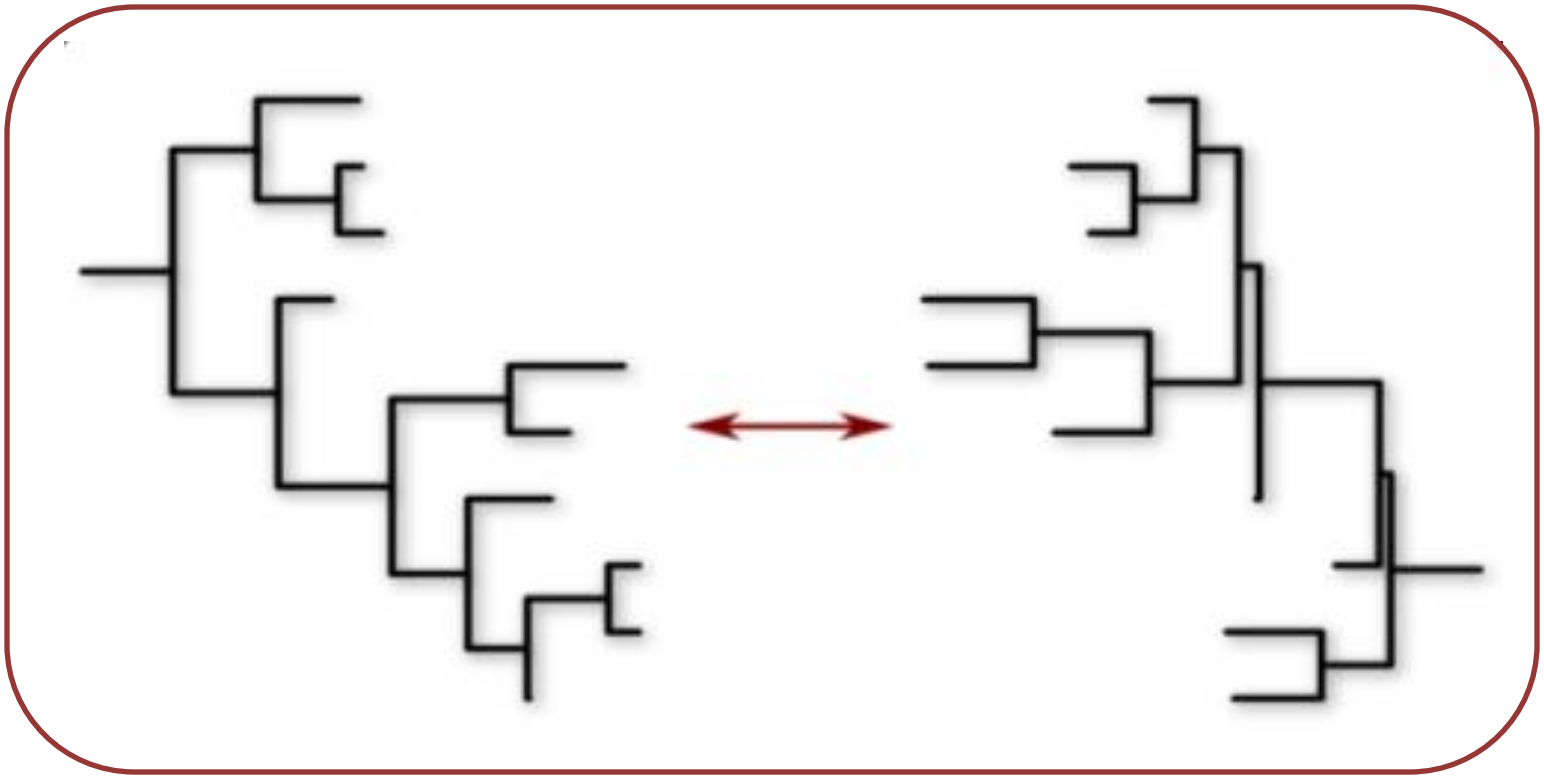
Species trees vs gene trees

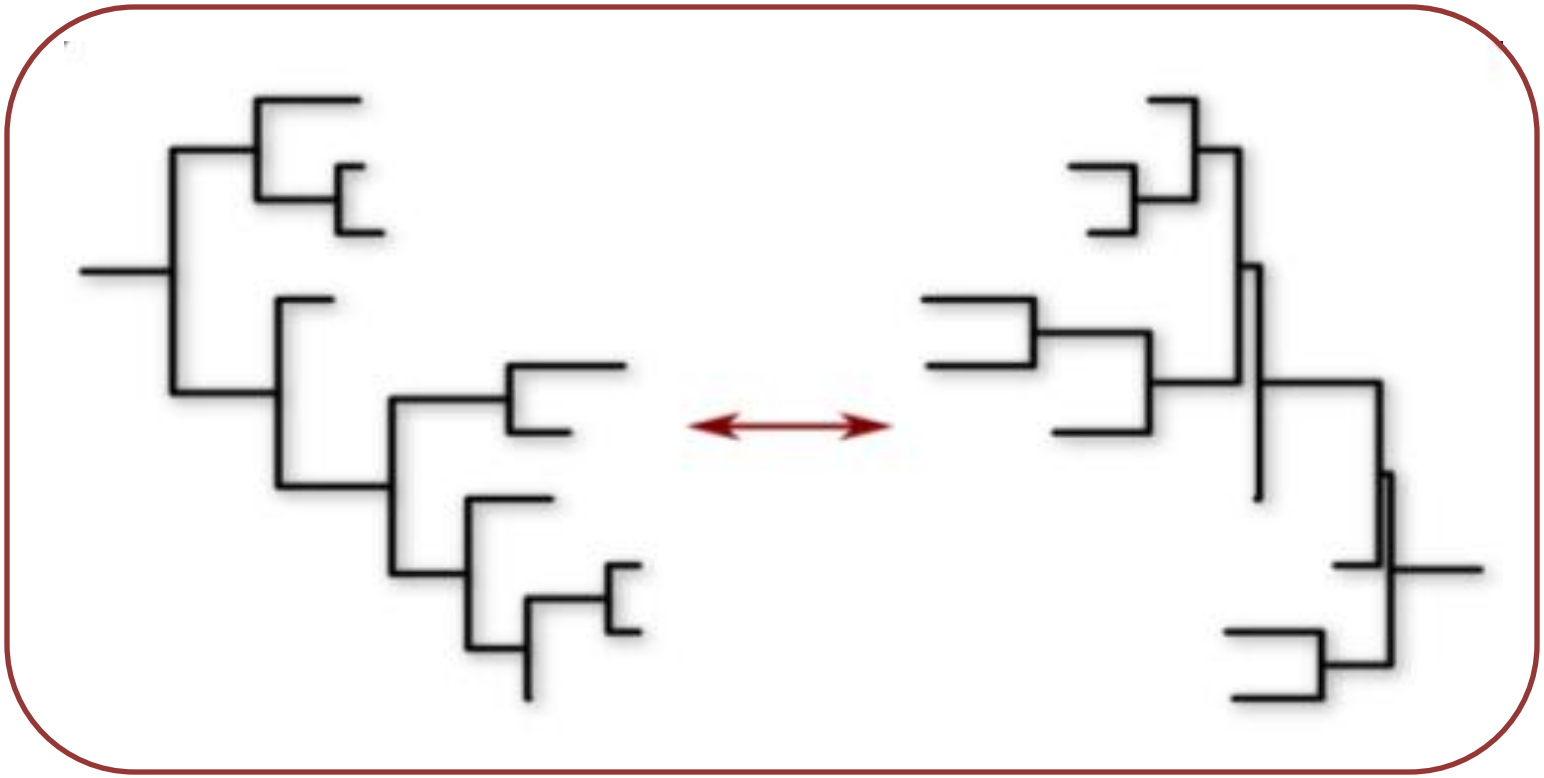




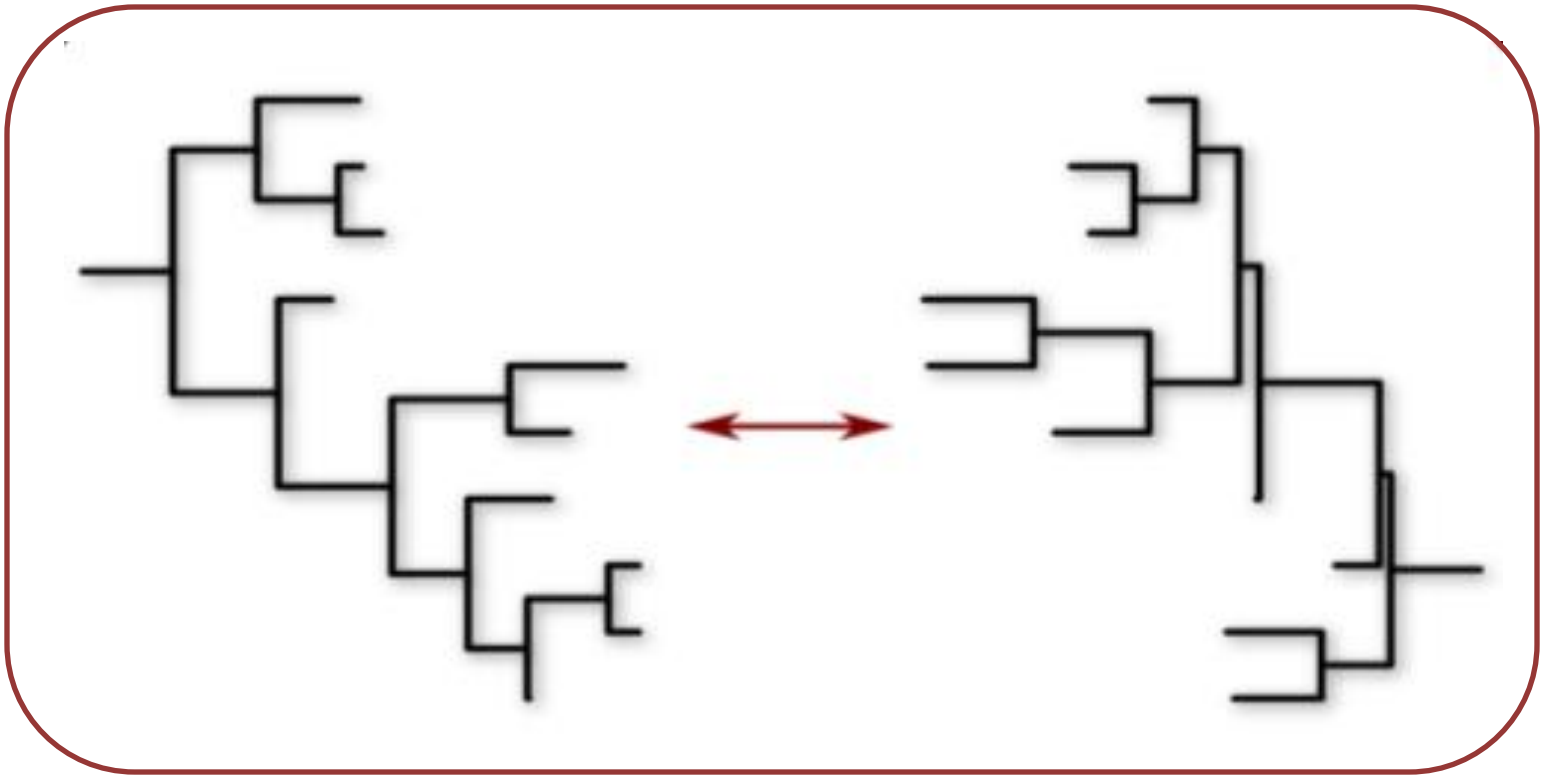








Species tree ↔ Species tree	Host parasite coevolutionary studies
Gene tree ↔ Gene tree	Protein-Protein interaction detection
Species tree ↔ Gene tree	Reconciliation analyses
Multiple gene trees	Phylogenomic studies

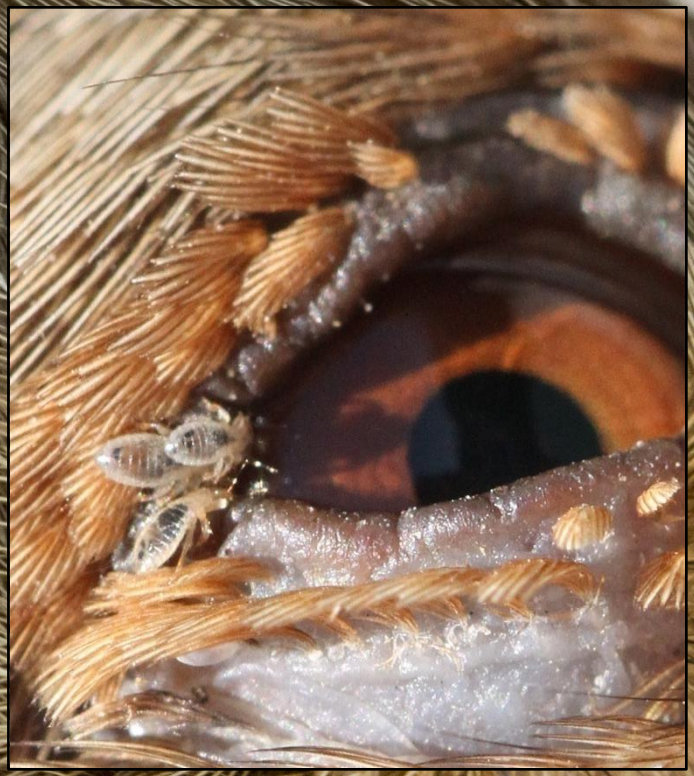


Species tree ↔ Species tree	**Host parasite coevolutionary studies**
Gene tree ↔ Gene tree	**Protein-Protein interaction detection**
Species tree ↔ Gene tree	**Reconciliation analyses**
Multiple gene trees	**Phylogenomic studies**

Fahrenholz, 1913

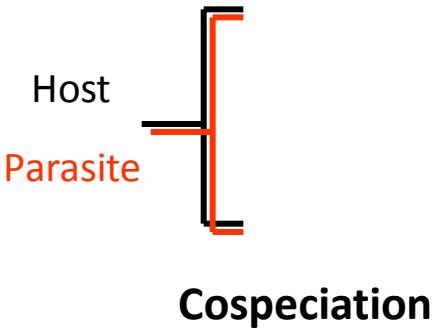
Fahrenholz's rule: "parasite phylogeny mirrors that of its host"

Fahrenholz, 1913



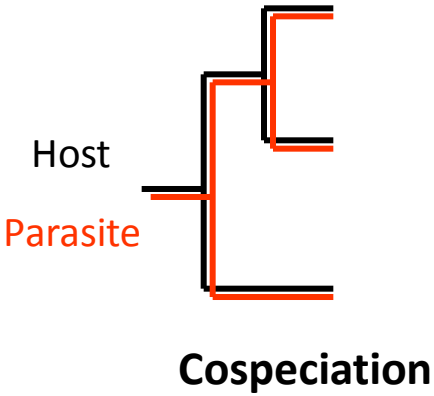
Fahrenholz's rule: "parasite phylogeny mirrors that of its host"

Cospeciation leads to congruent phylogenies



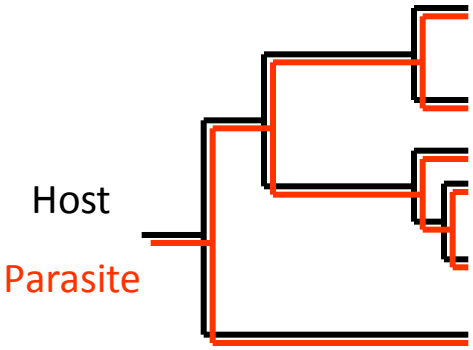
Cospeciation

Cospeciation leads to congruent phylogenies



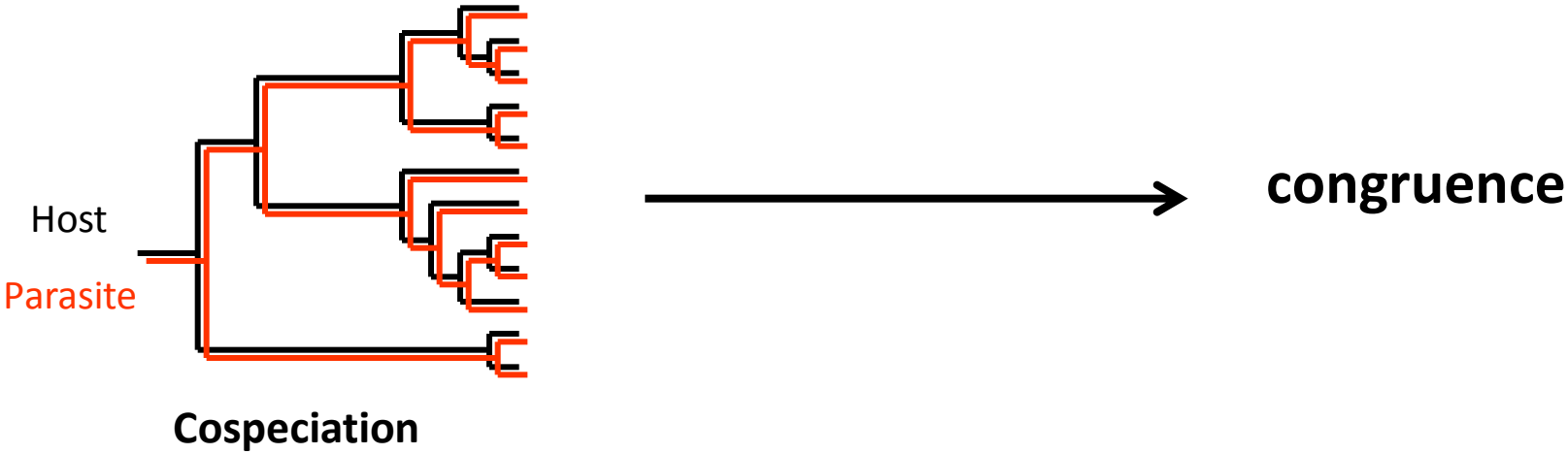
Cospeciation

Cospeciation leads to congruent phylogenies

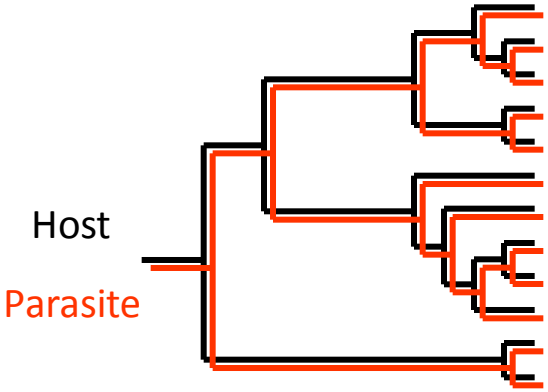


Cospeciation

Cospeciation leads to congruent phylogenies

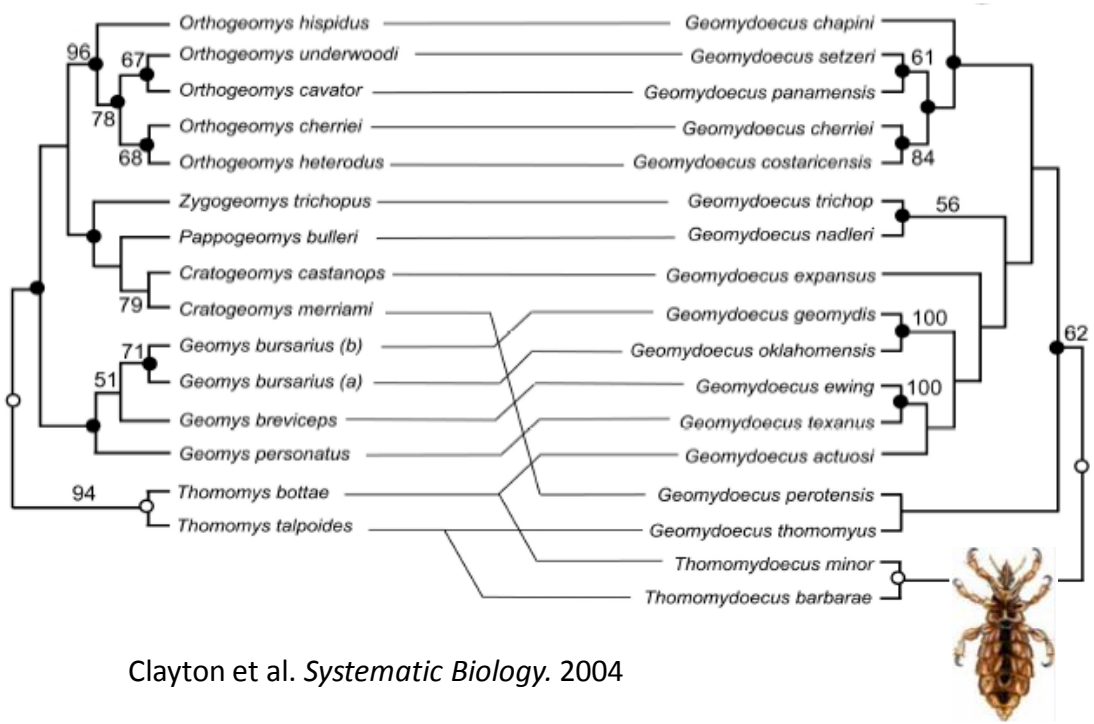


Cospeciation leads to congruent phylogenies



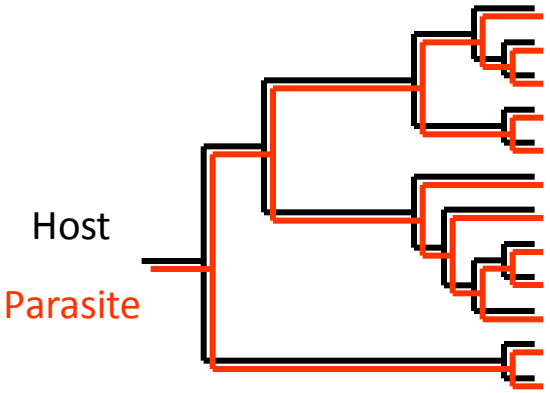
Cospeciation

congruence



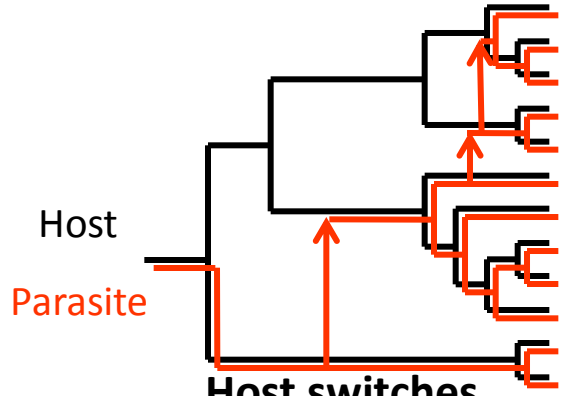
Clayton et al. *Systematic Biology.* 2004

Other events lead to **in**congruent phylogenies



Cospeciation

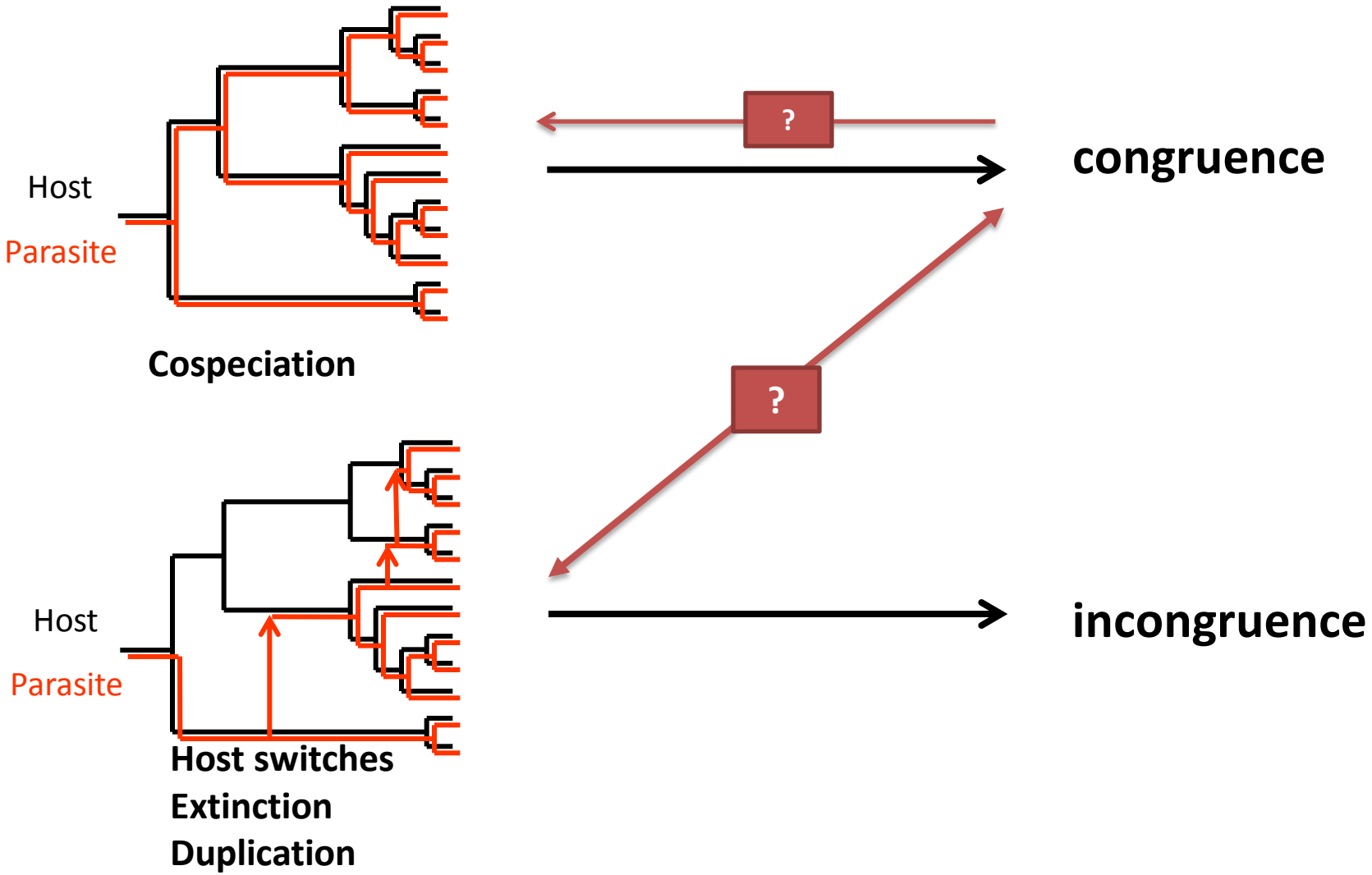
congruence



Host switches
Extinction
Duplication

incongruence

Does congruence imply cospeciations?



Can host-switches produce congruent phylogenies?

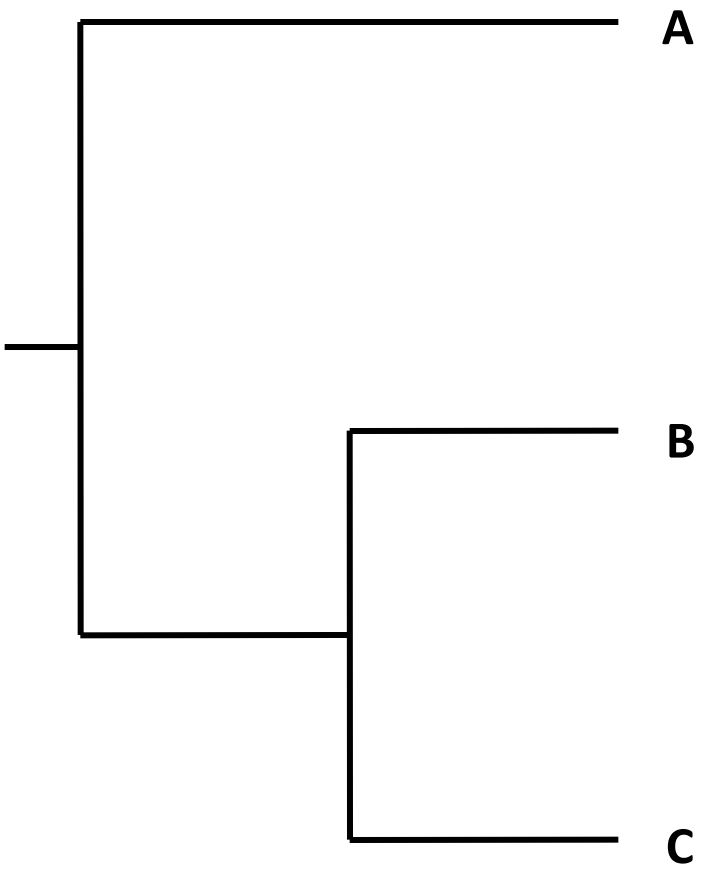
Can congruence be due to host-switches alone?

Adaptive radiation on a group of pre-existing host species

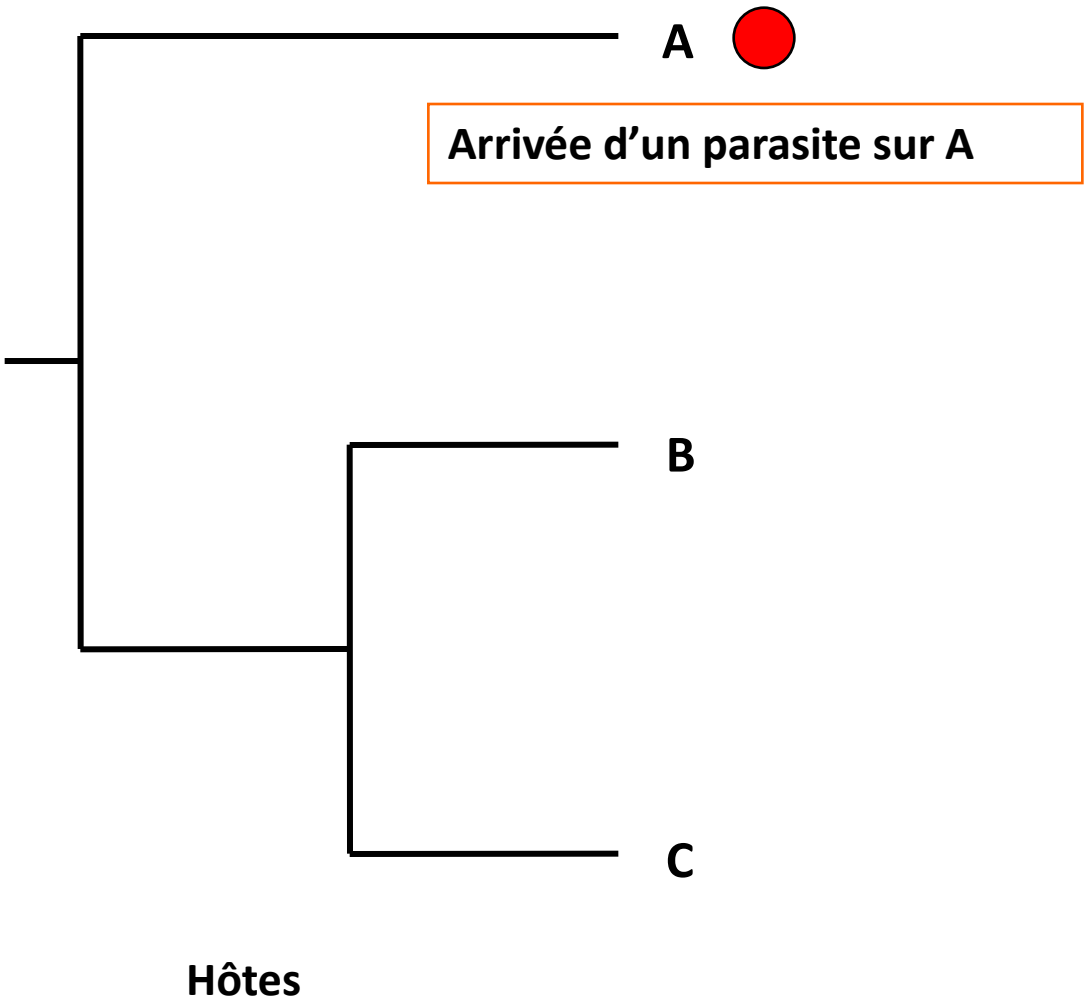
Under what conditions?

Host tree topology, first host parasitized, host-switch probability, time lag between switch and speciation as a function of the switch distance

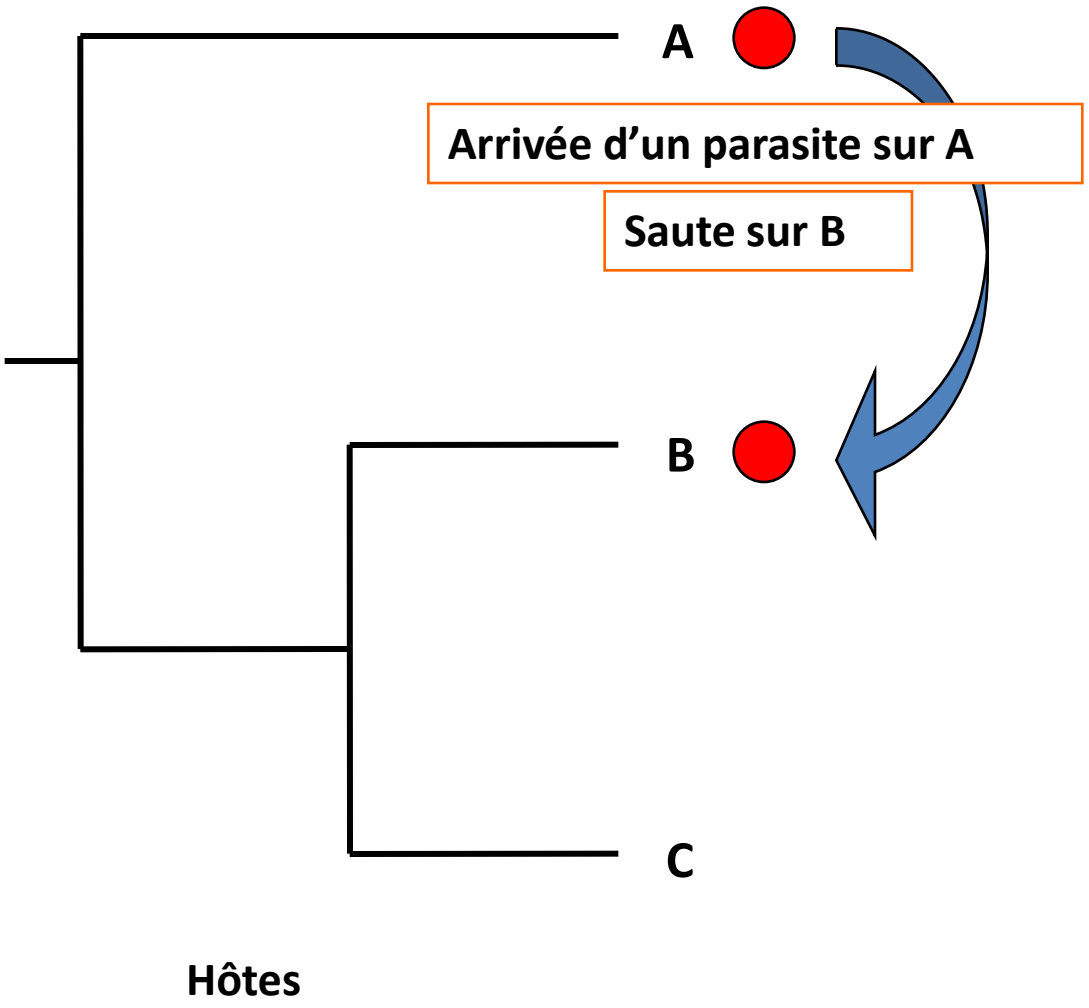
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



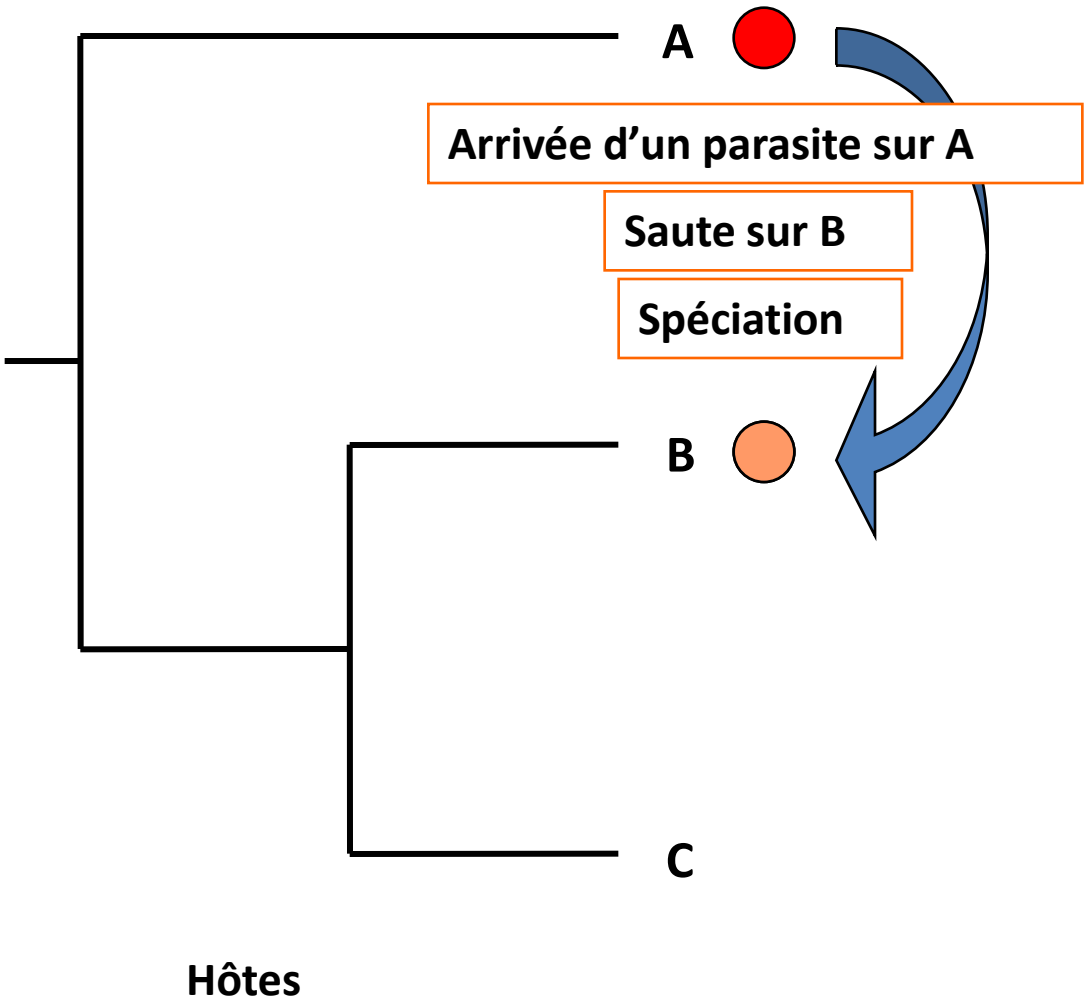
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



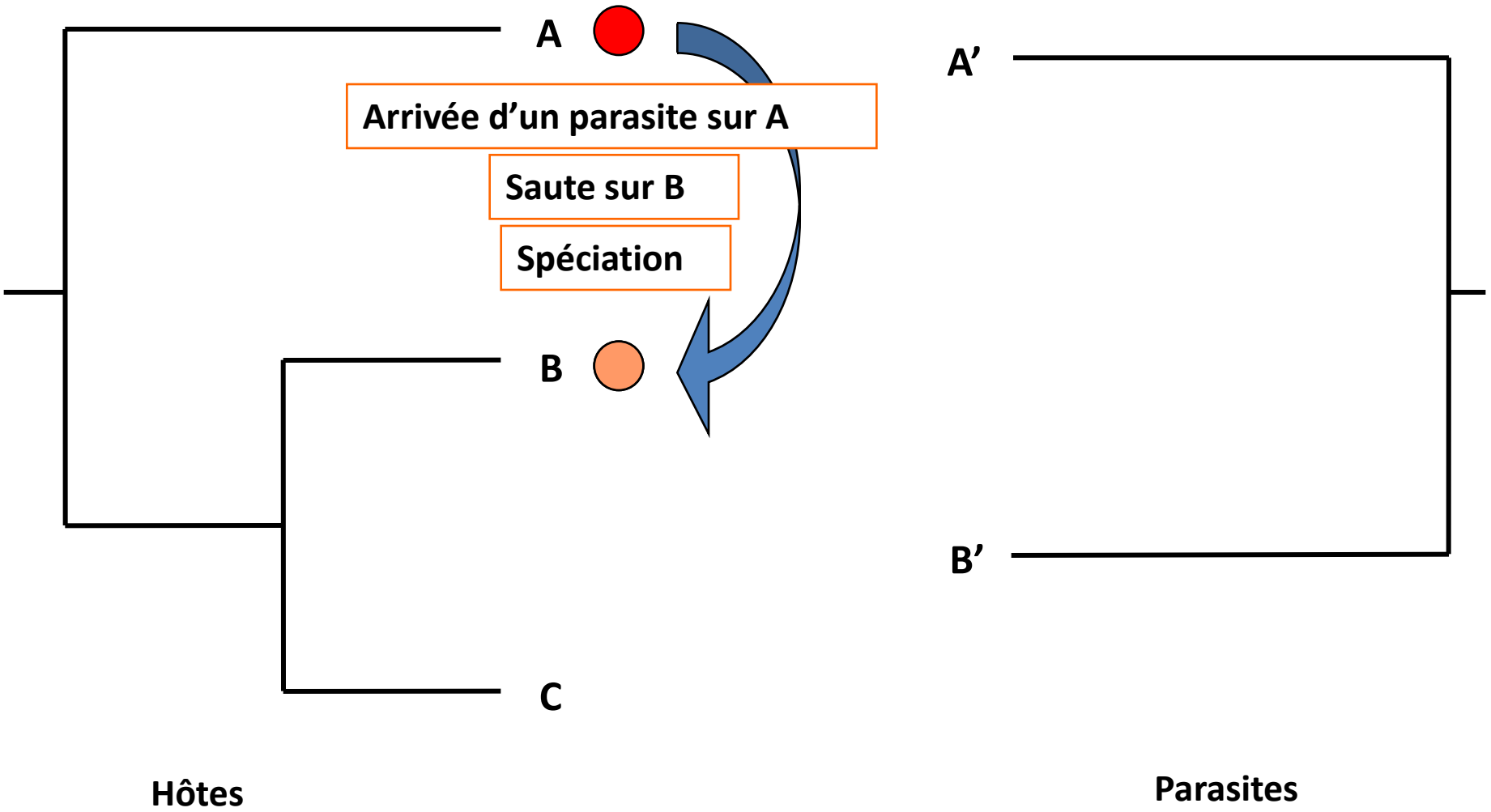
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



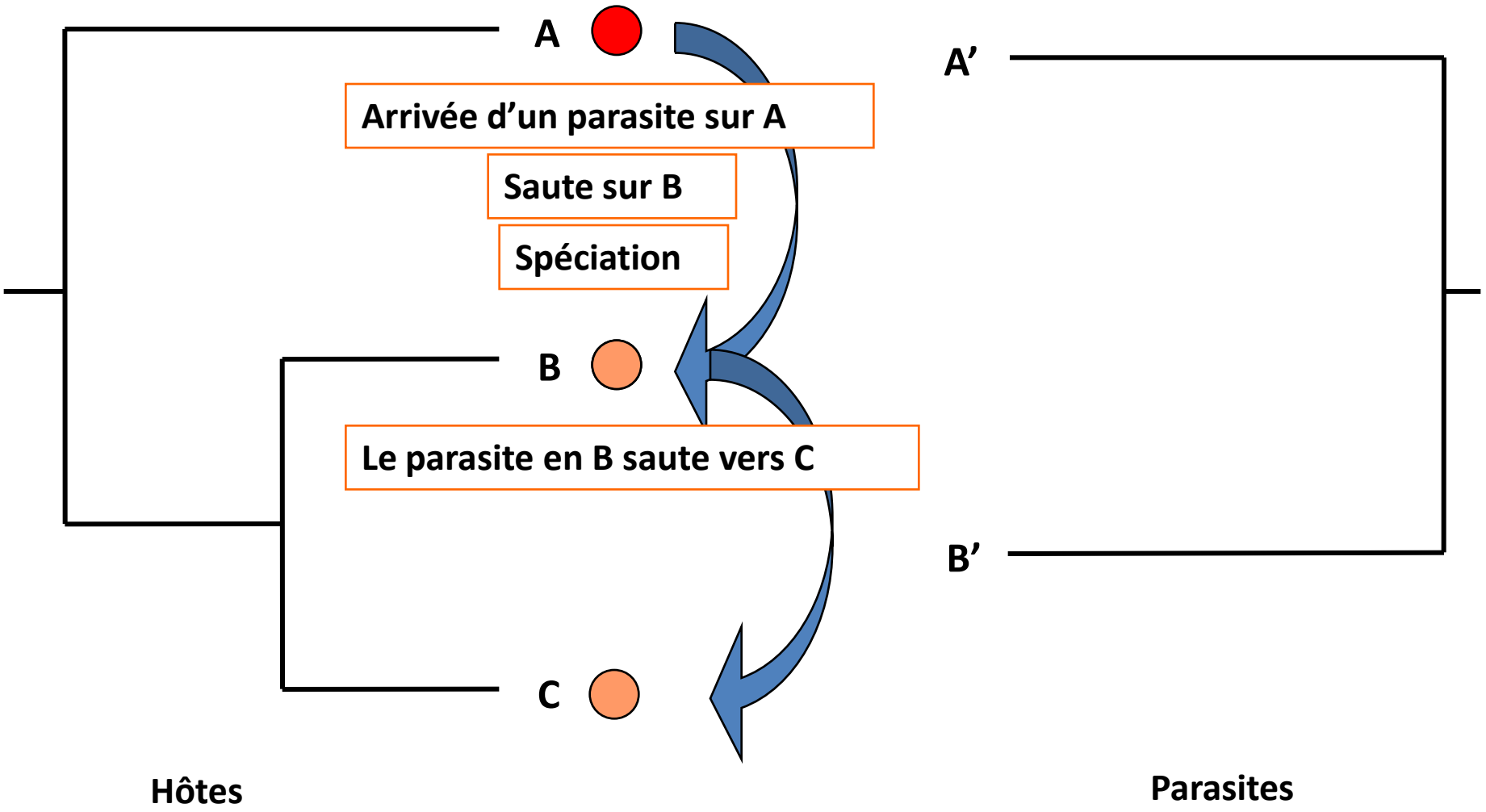
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



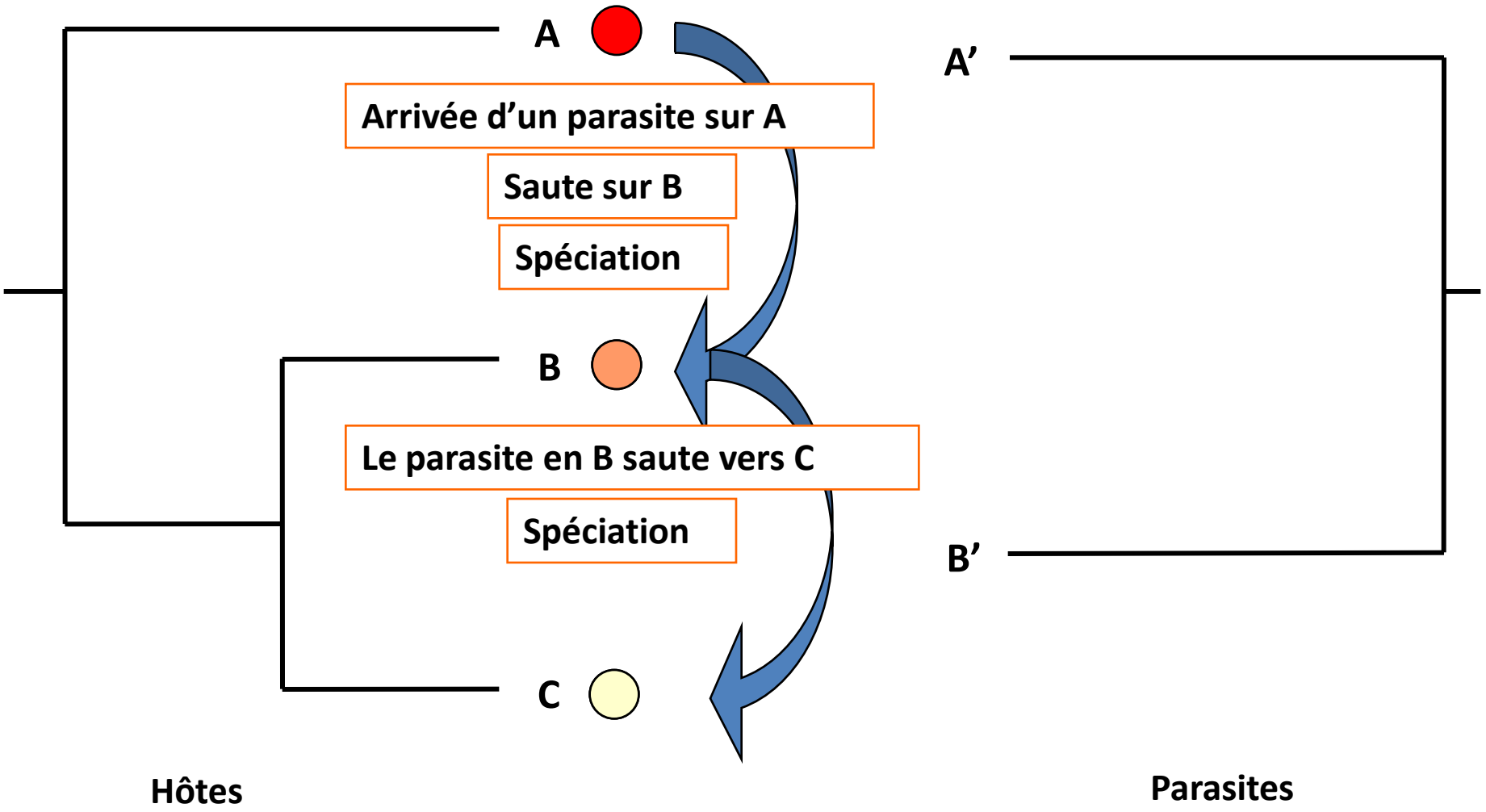
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



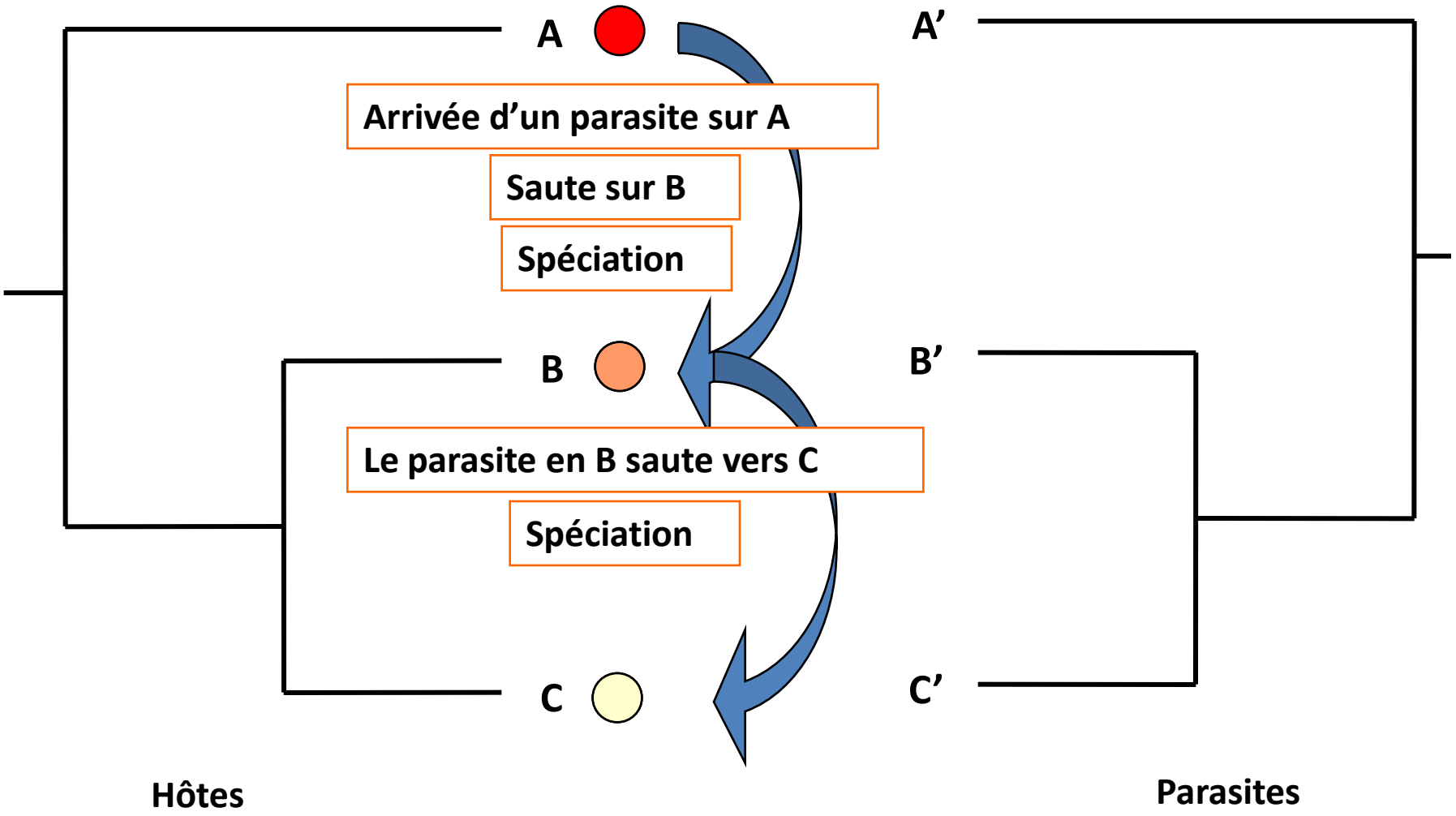
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



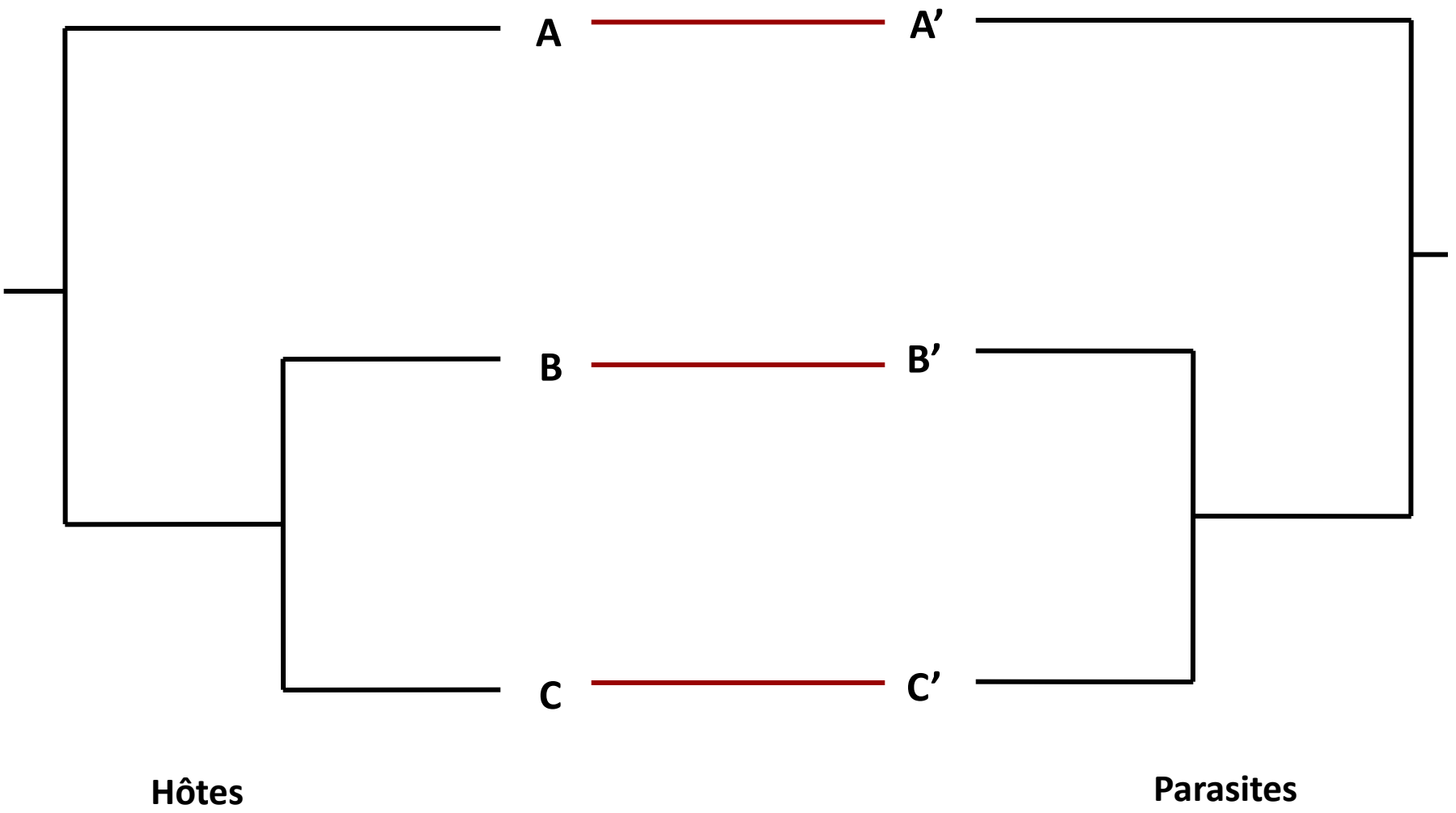
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



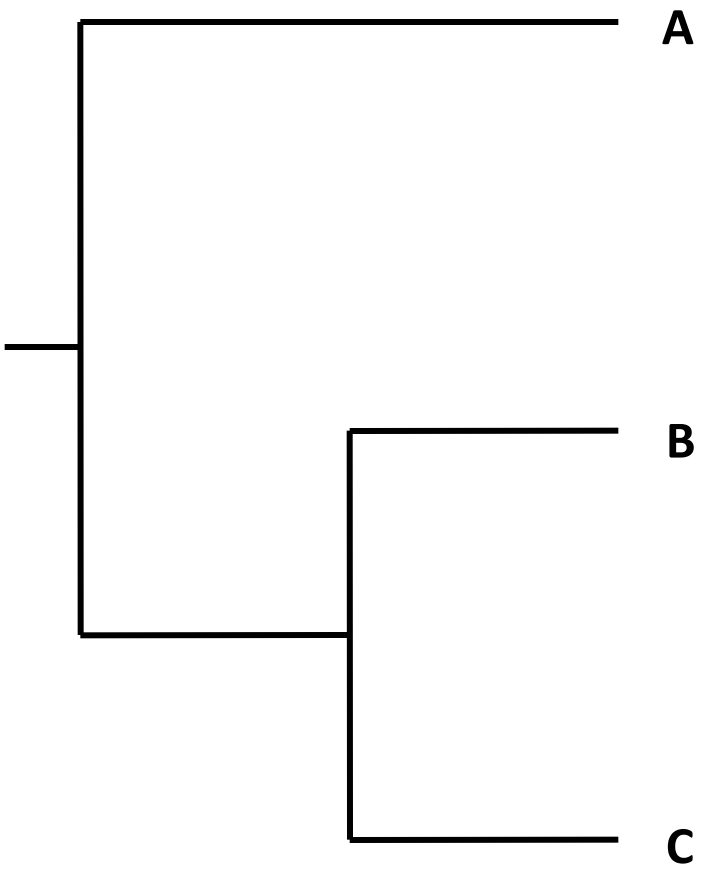
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



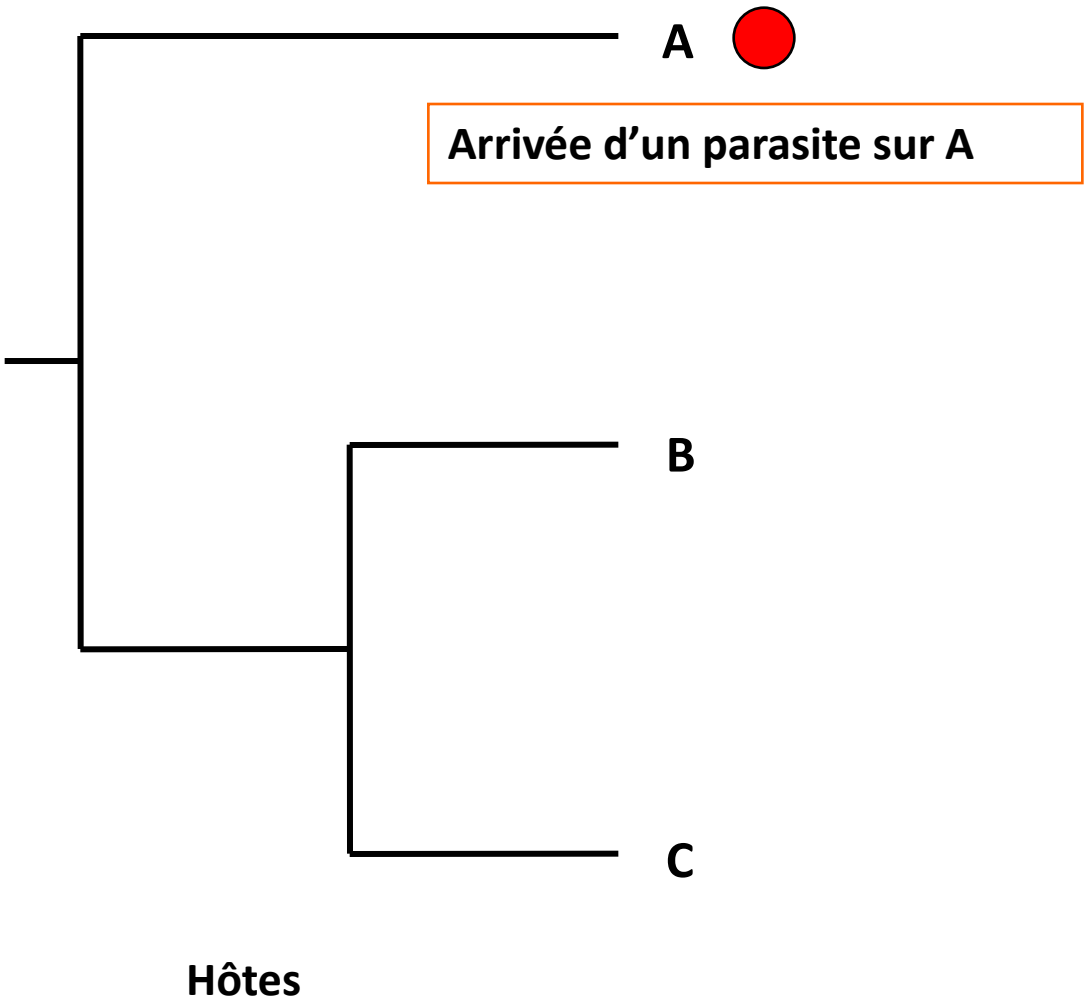
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



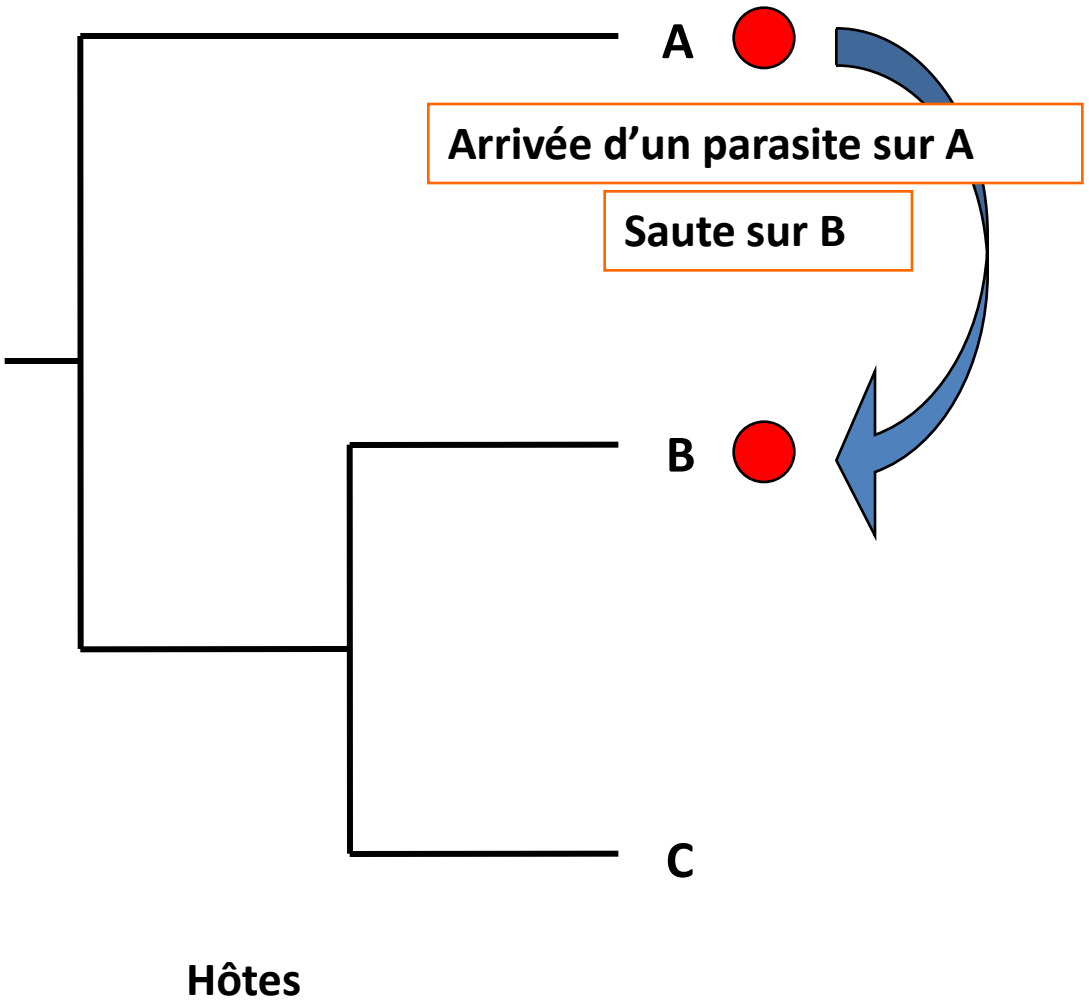
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



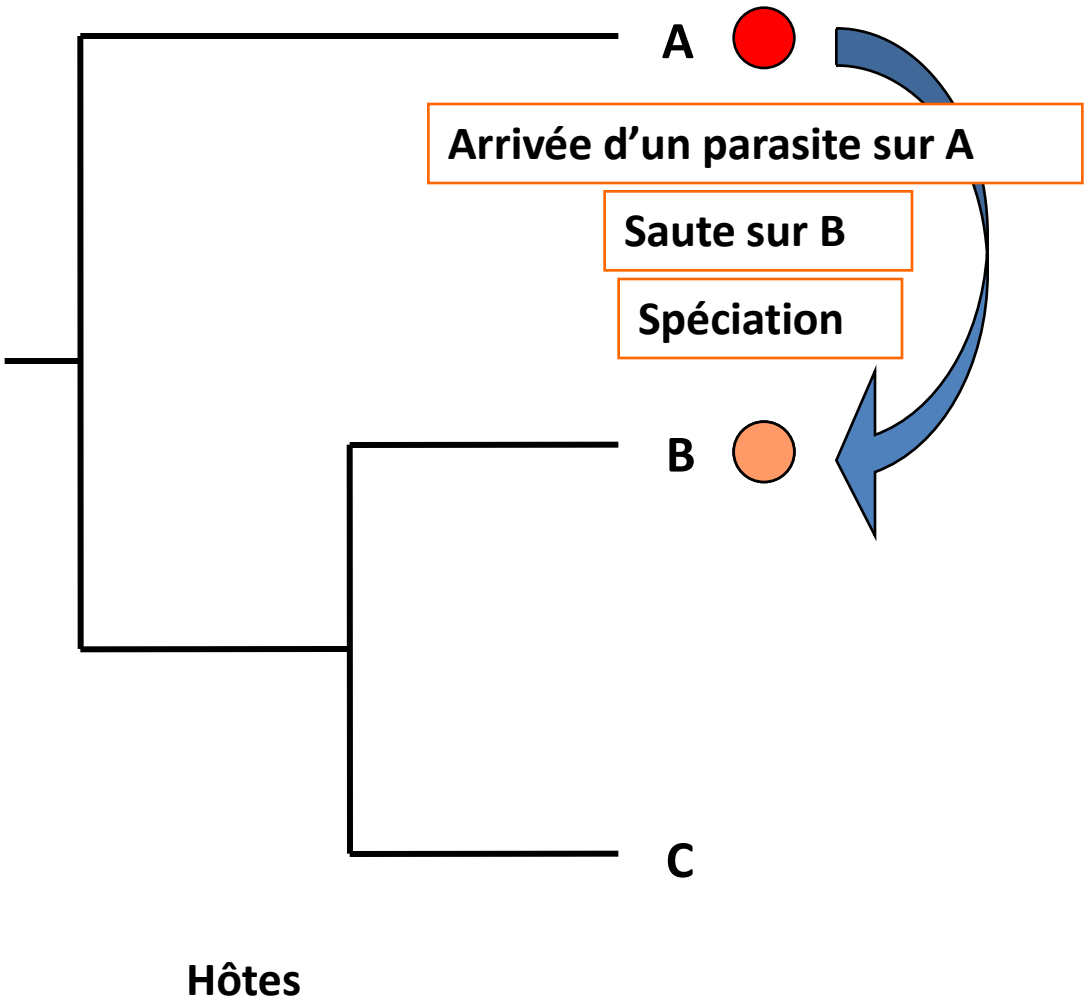
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



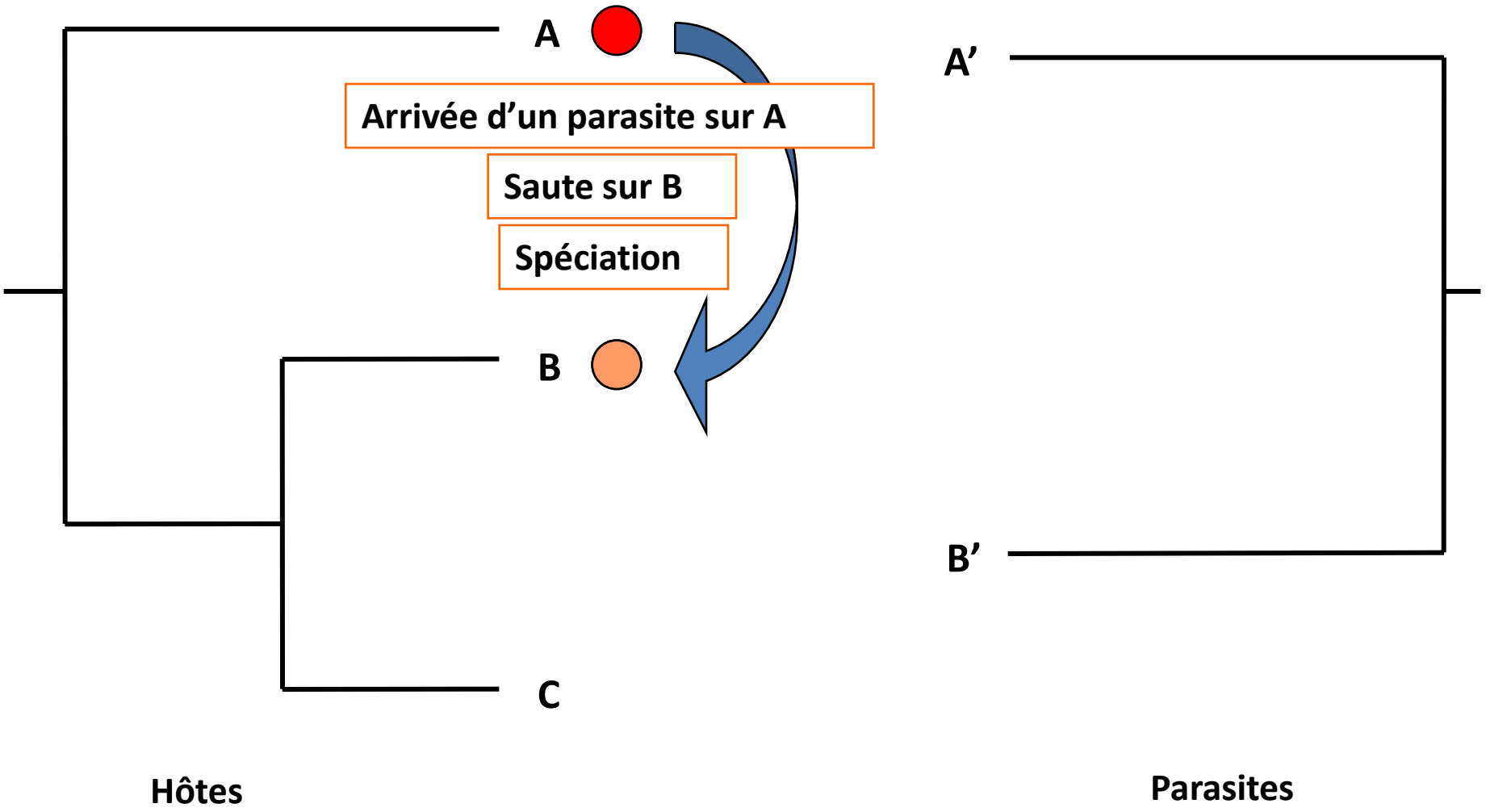
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



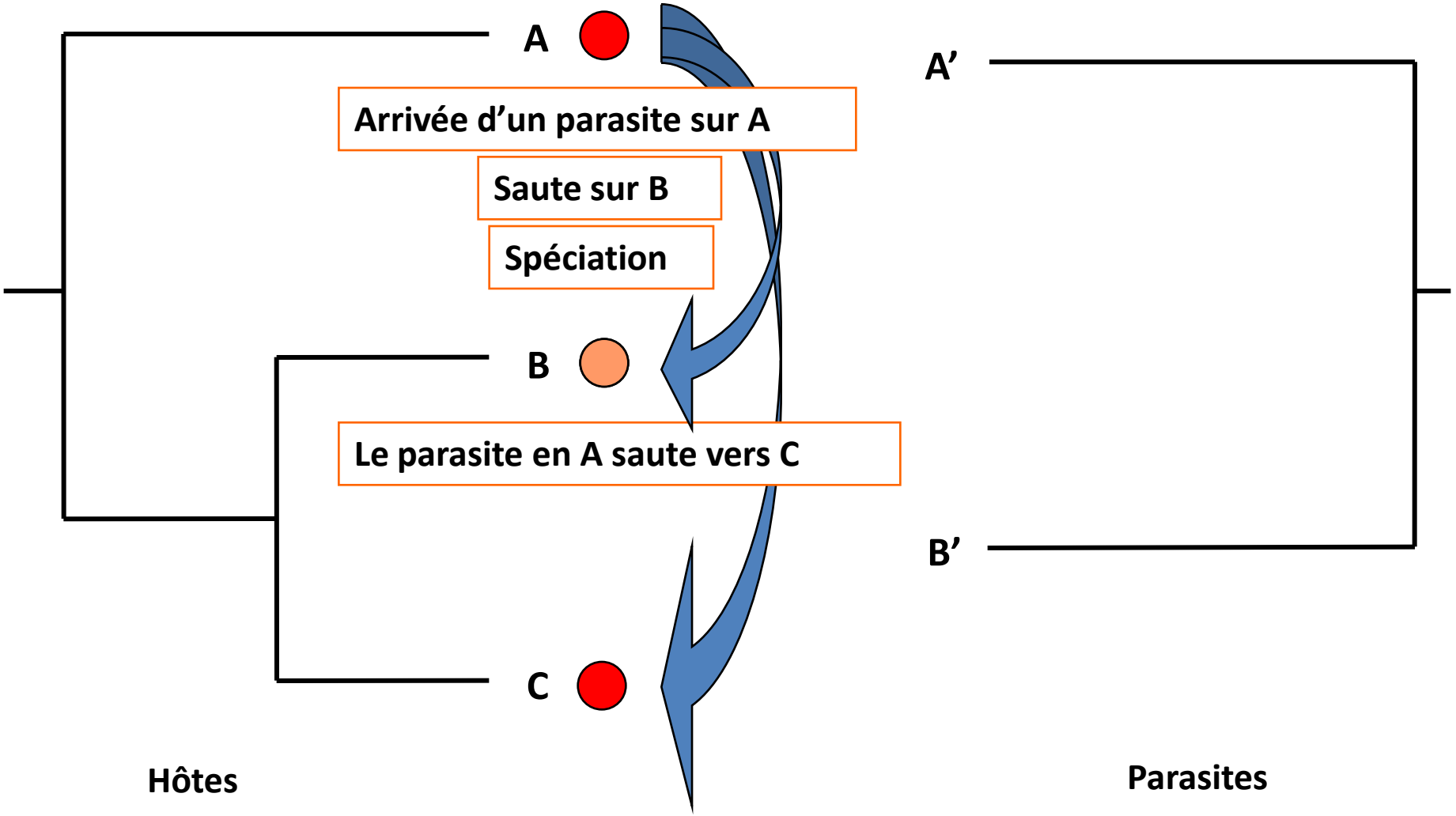
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



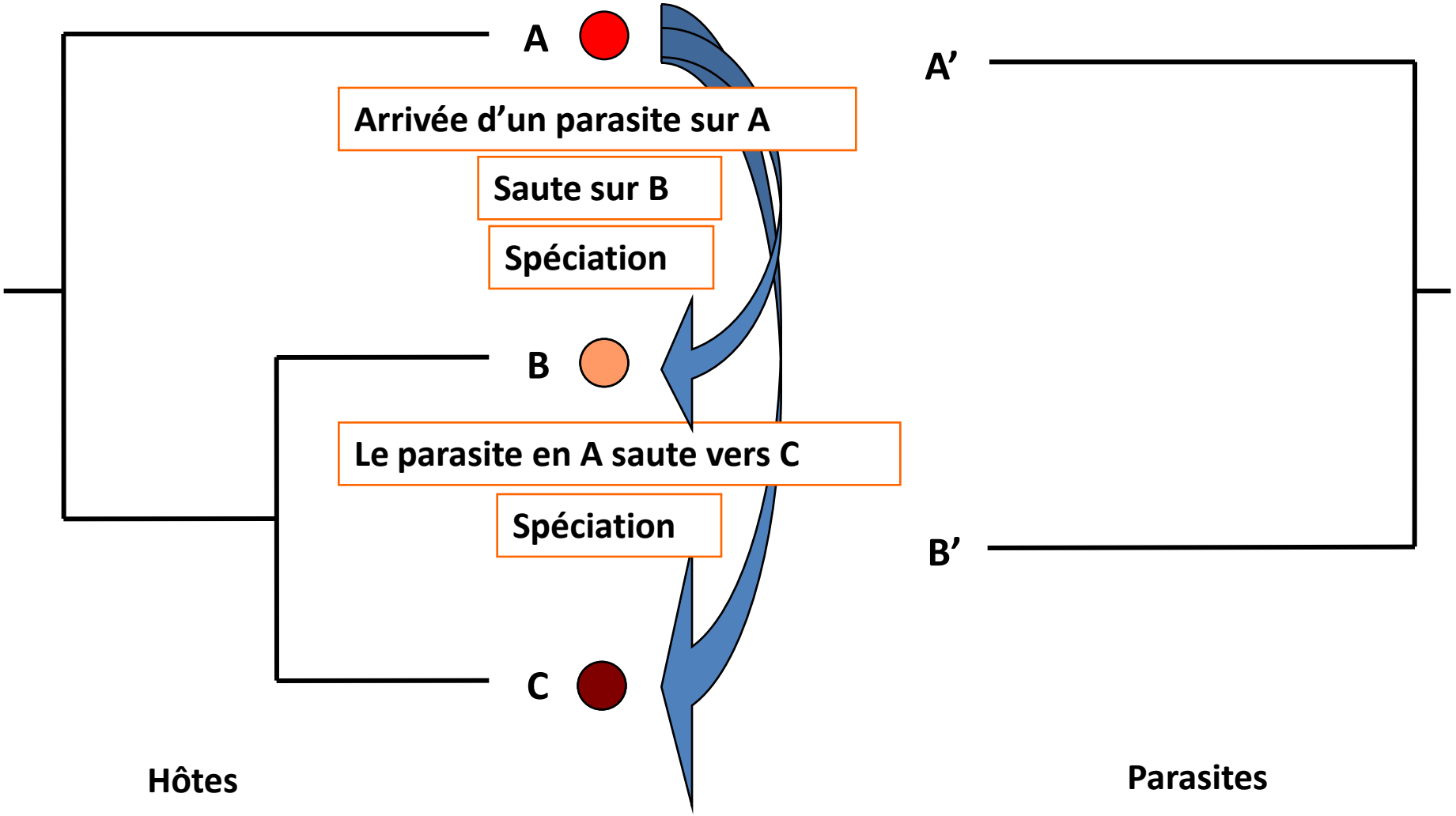
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



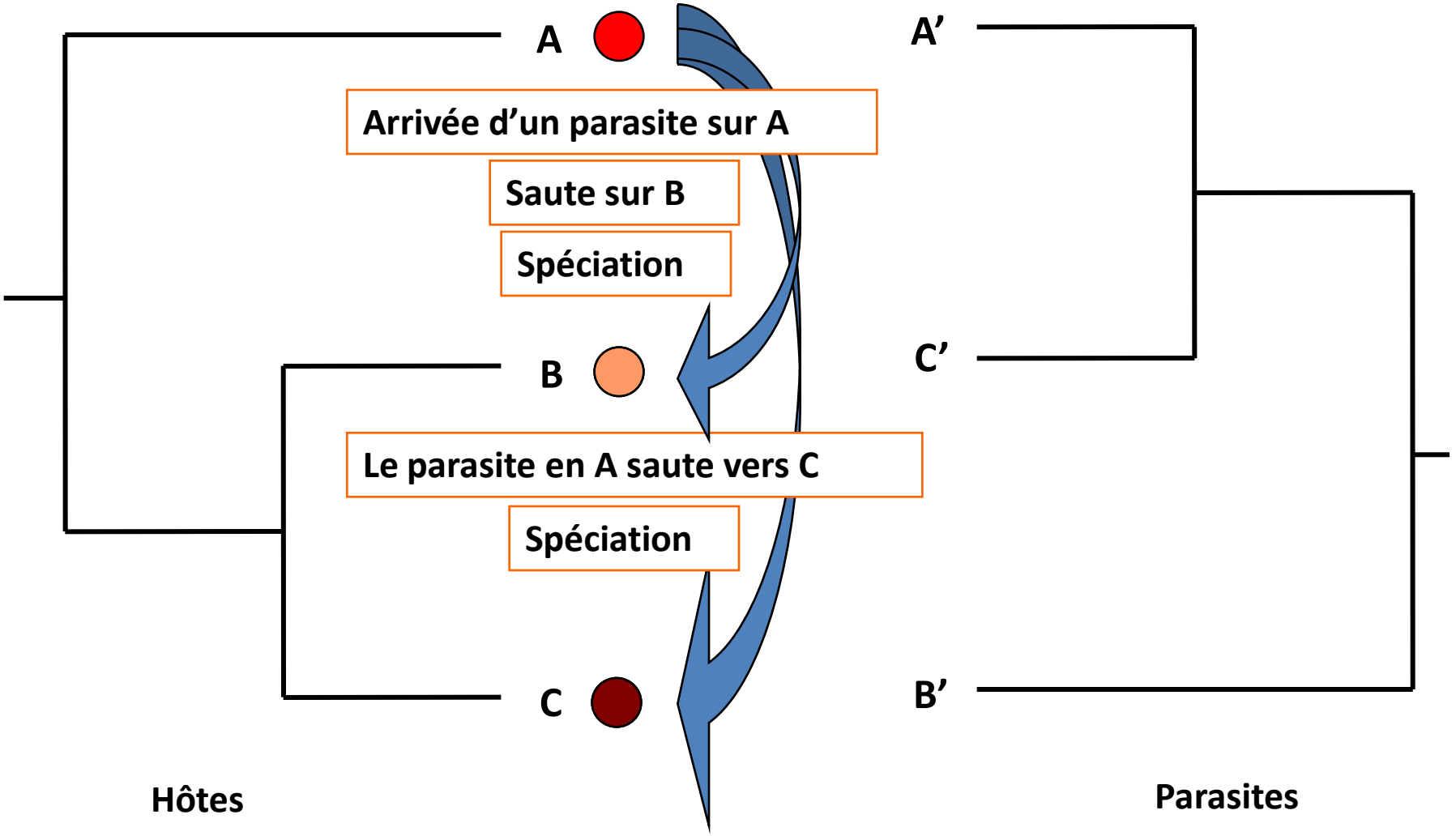
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



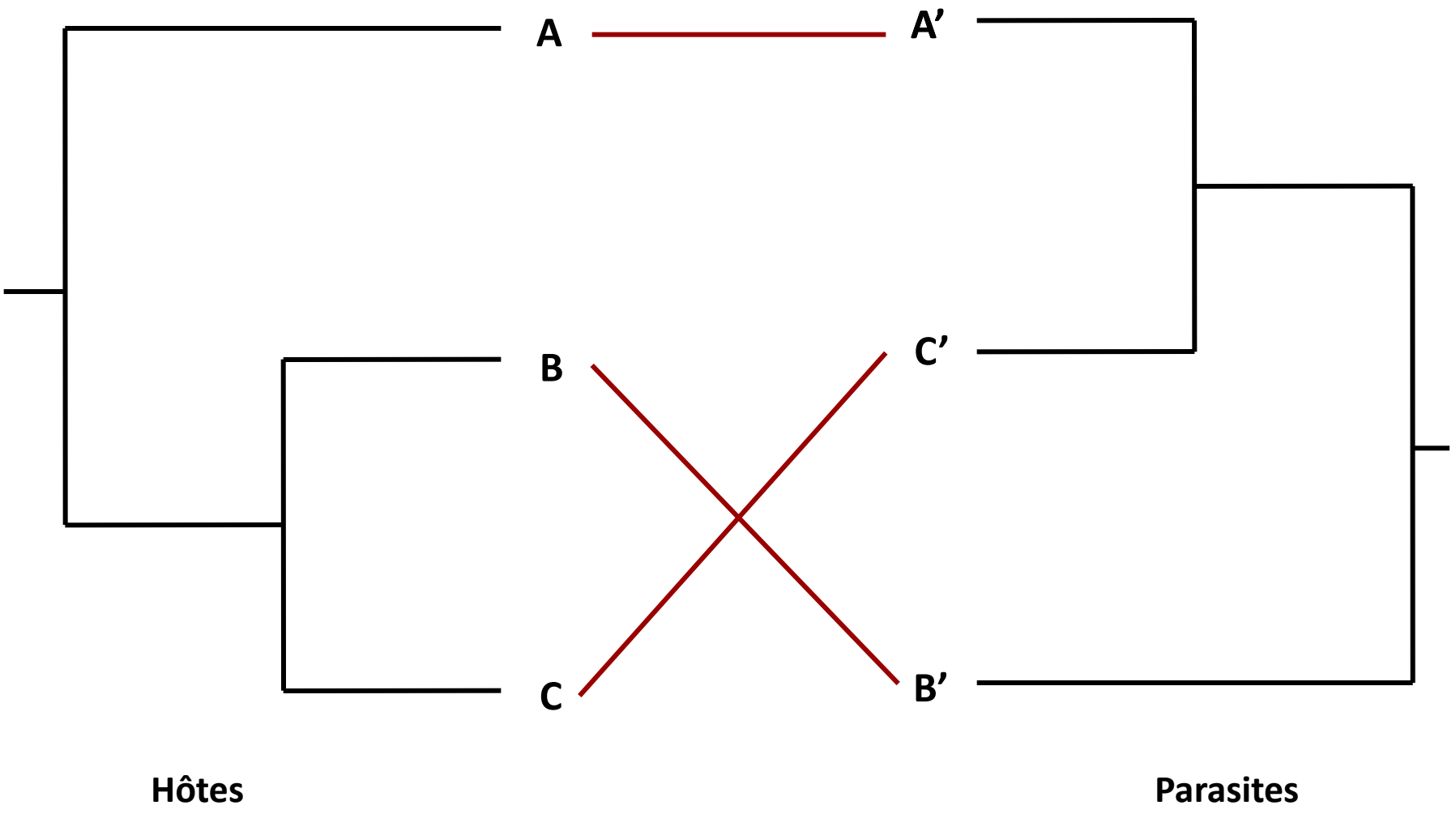
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



Simulate adaptive radiation of a parasite on a group of pre-existing hosts



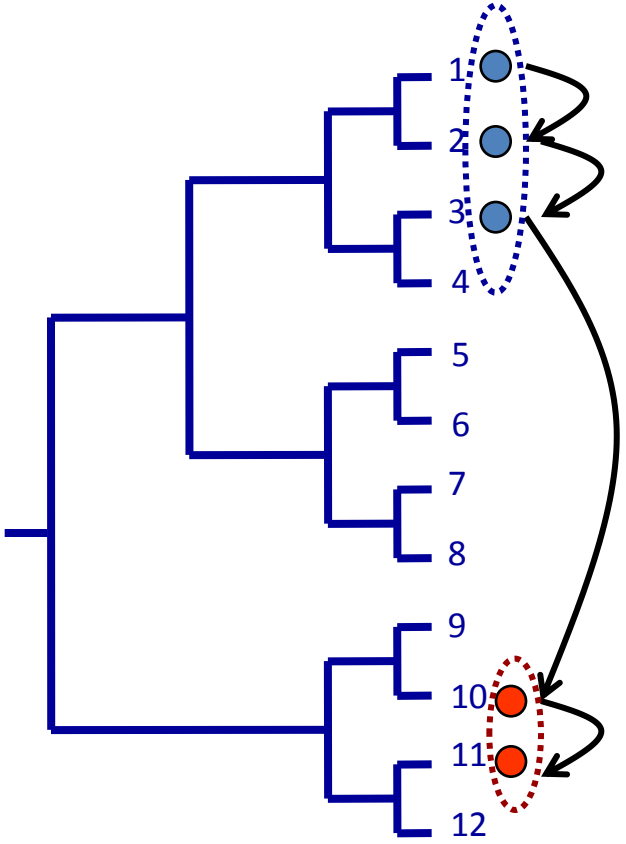
Simulate adaptive radiation of a parasite on a group of pre-existing hosts



High congruence between host and parasite phylogenies can be obtained without cospeciation under plausible conditions

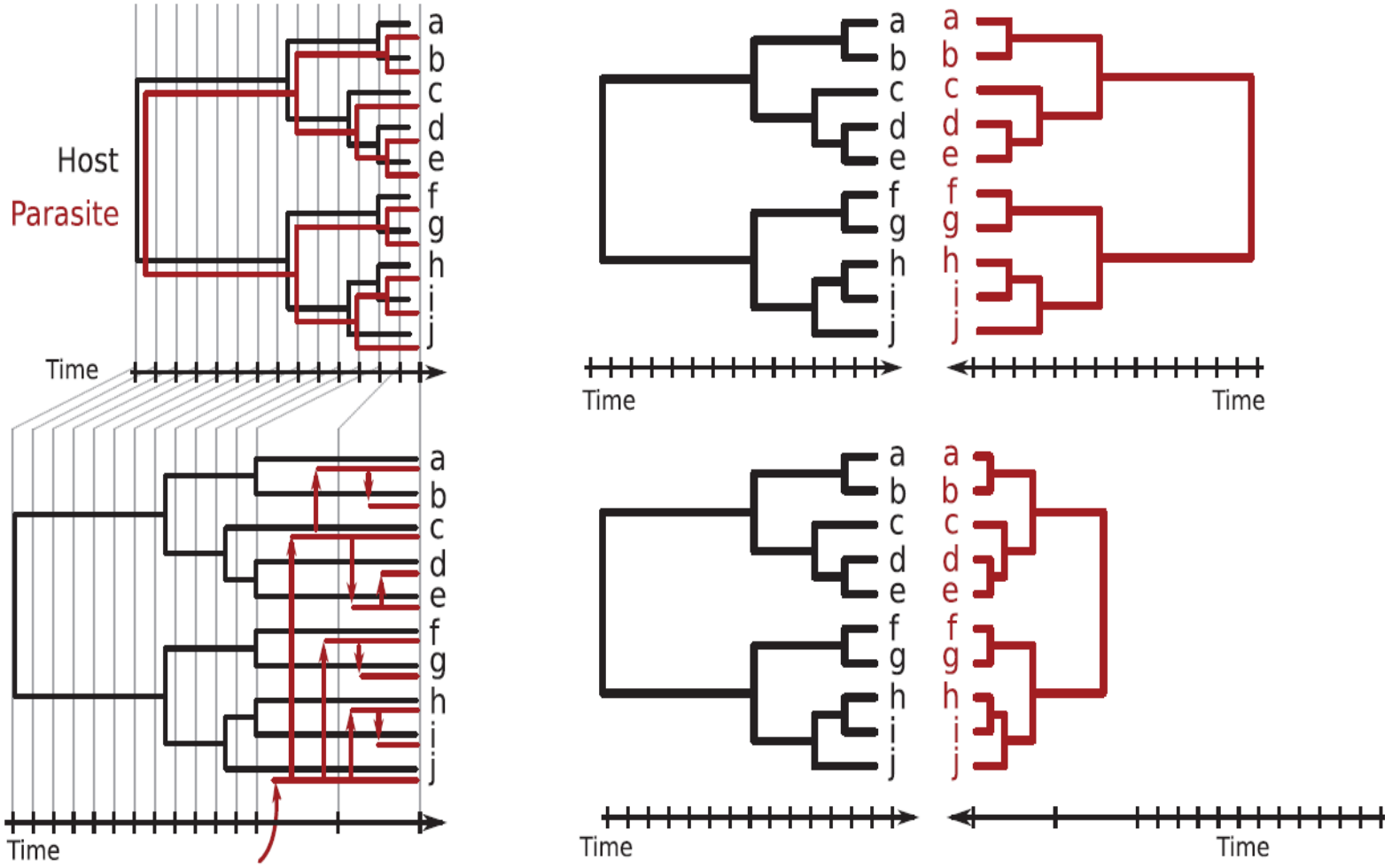
Highest probability for switches to closely related hosts

Faster speciation after distant switches than after close ones



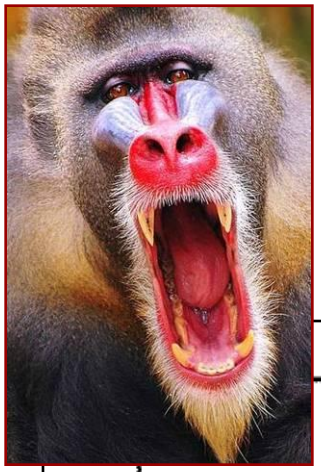
de Vienne et al. JEB. 2007

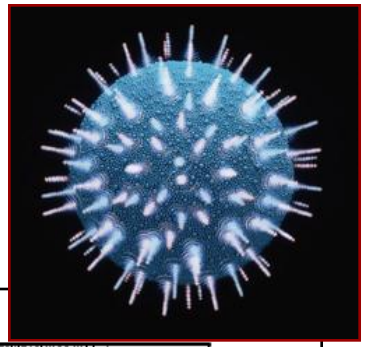
The main difference between cospeciation-based and host-shift based congruence is the age of nodes in the trees

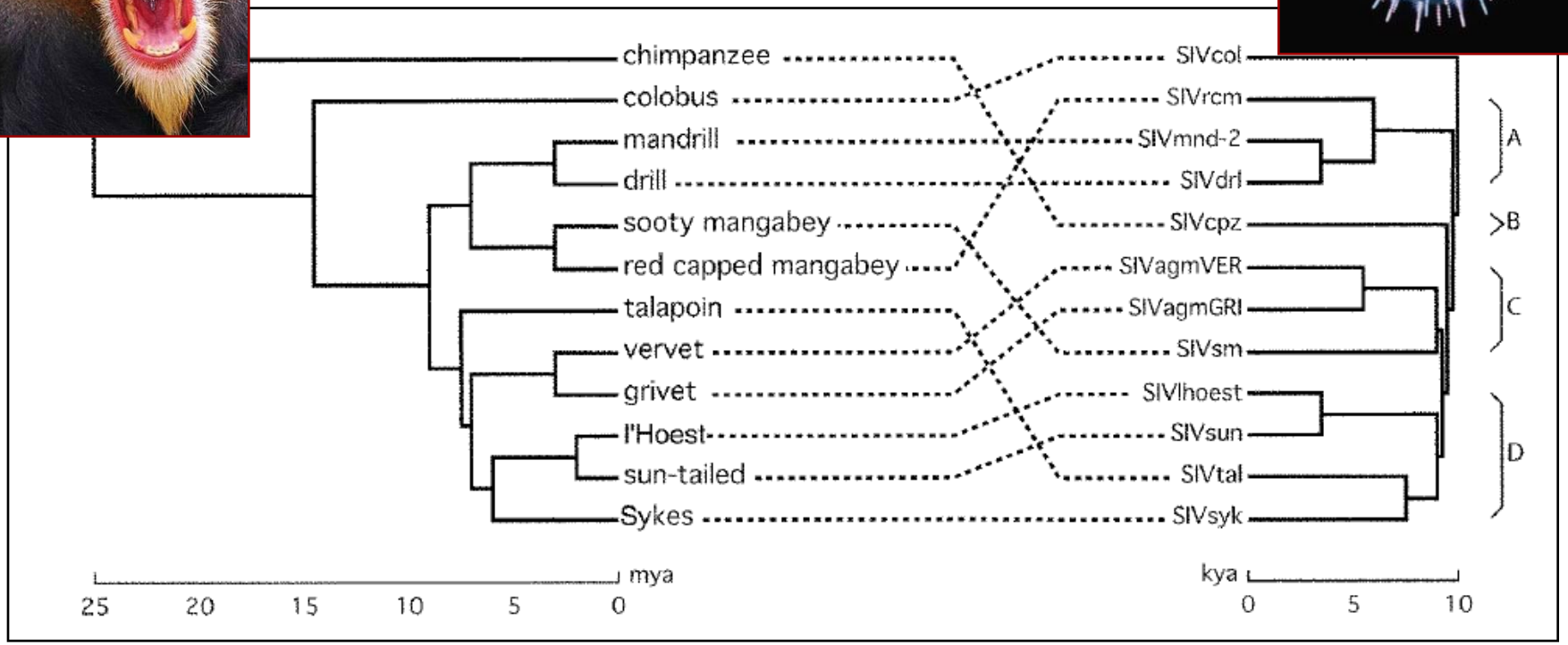


de Vienne et al. New Phytologist. 2013

High congruence between host and parasite phylogenies can be obtained without cospeciation under plausible conditions







Charleston and Robertson. *Systematic Biology.* 2002

Co-phylogenetic methods

- Event- or cost-based
  - Estimate a scenario
  - Costs are associated to each event (cospeciation/duplication/host-shift/extinction)
  - Try to minimize cospeciation
  - Cospeciation is always less costly

- Topology- and distance-based methods
  - Based on comparison of a score with its distribution after permutations
  - No *a priori* on the reasons for the overall congruence (no events)
  - Often, *a posteriori* interpretation that congruence = cospeciation

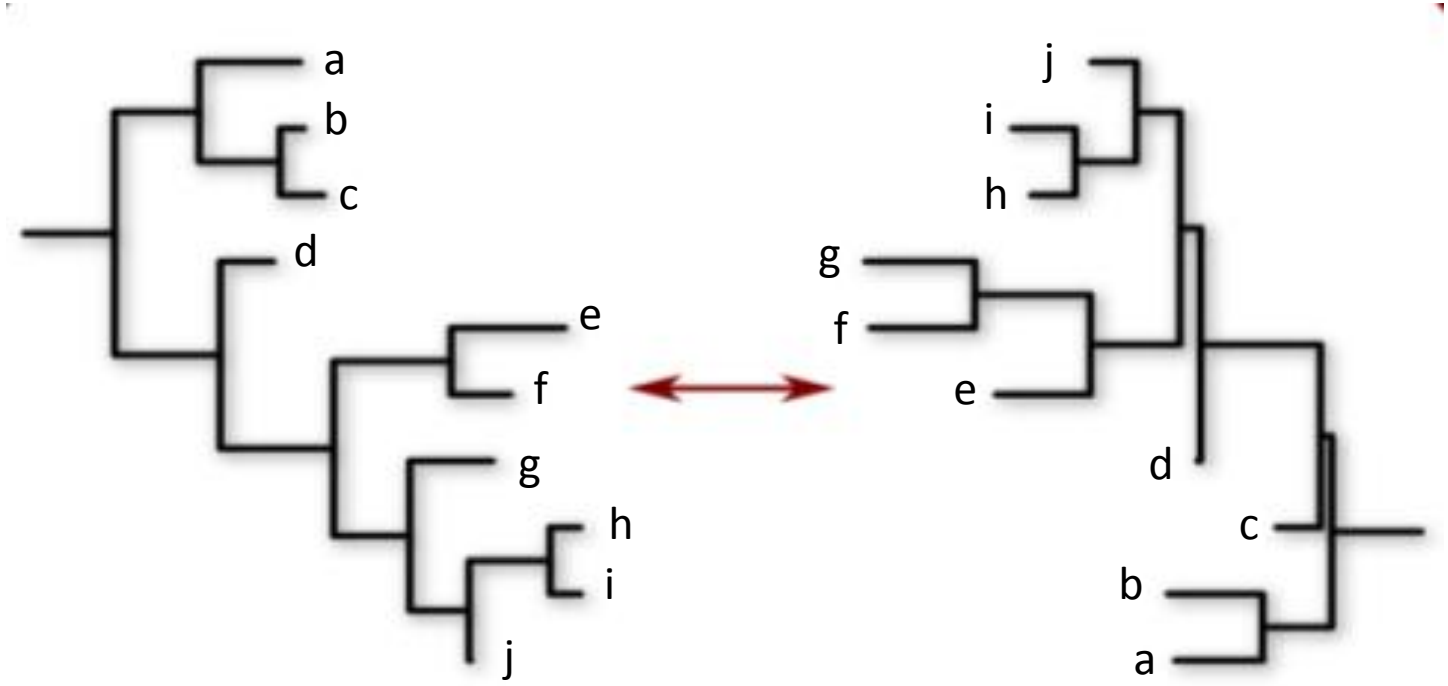
Co-phylogenetic methods

- Event- or cost-based
  - Estimate a scenario
  - Costs are associated to each event (cospeciation/duplication/host-shift/extinction)
  - Try to minimize cospeciation
  - Cospeciation is always less costly

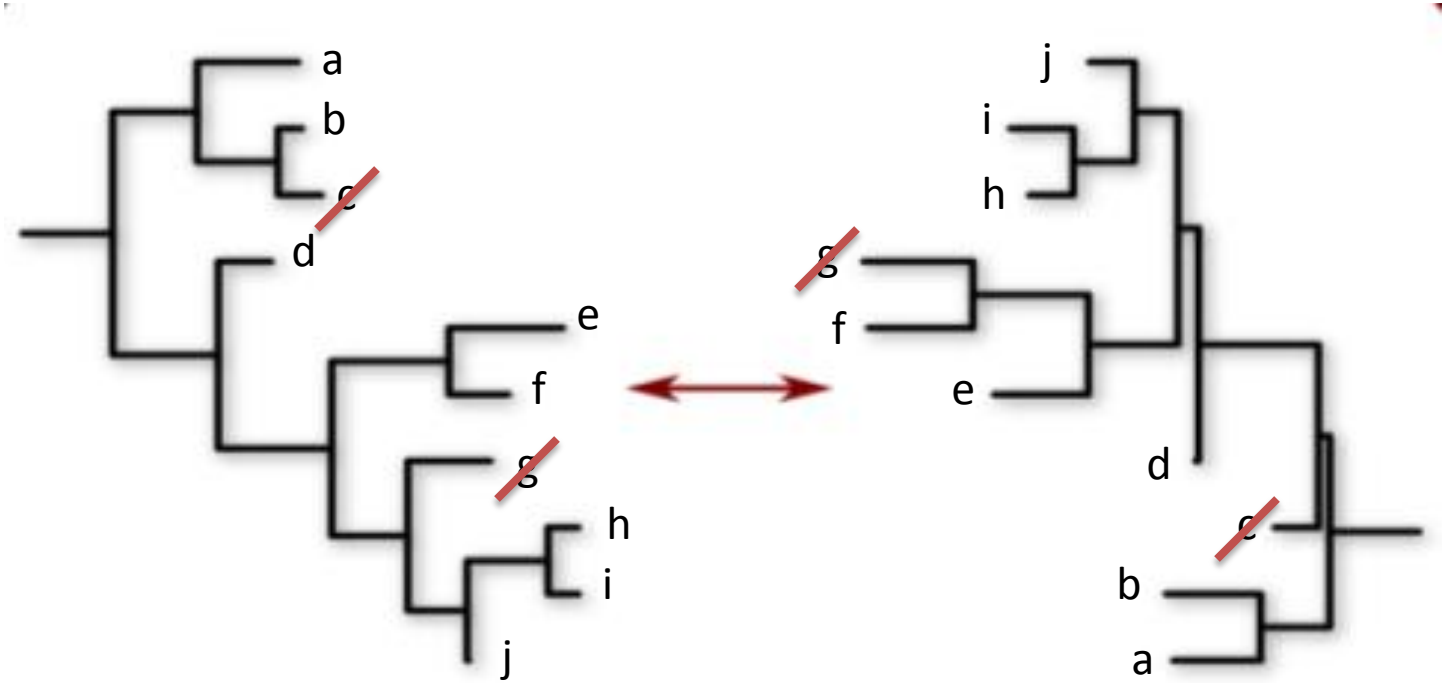
- Topology- and distance-based methods
  - Based on comparison of a score with its distribution after permutations
  - No *a priori* on the reasons for the overall congruence (no events)
  - Often, *a posteriori* interpretation that congruence = cospeciation

I_{cong}: New topological distance based on expected size of the MAST (no permutations)

Maximum Agreement Subtree (MAST)



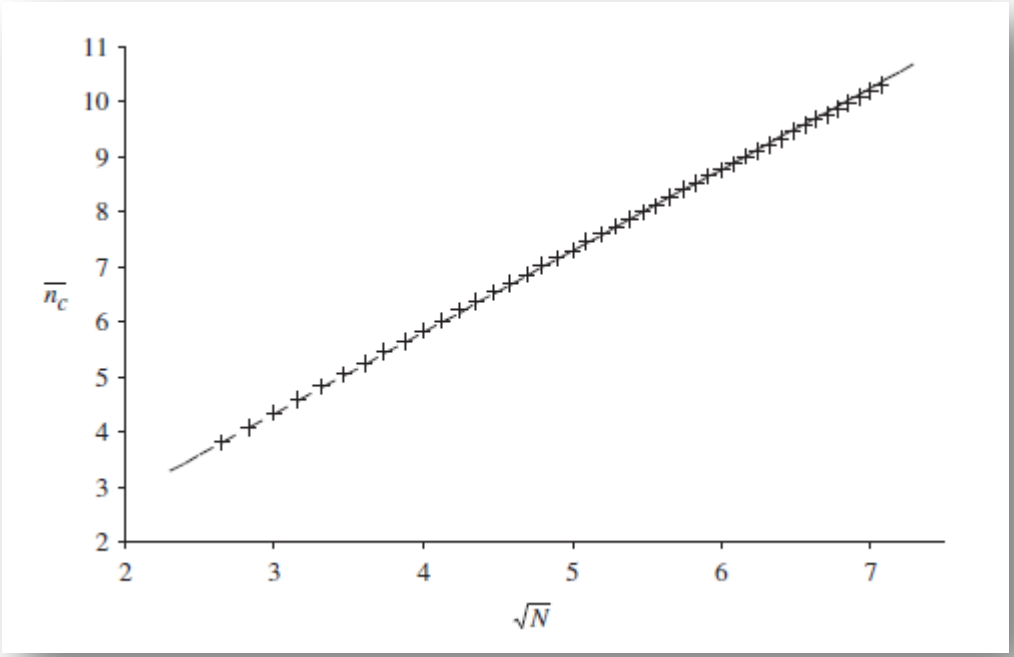
Maximum Agreement Subtree (MAST)



N = 10

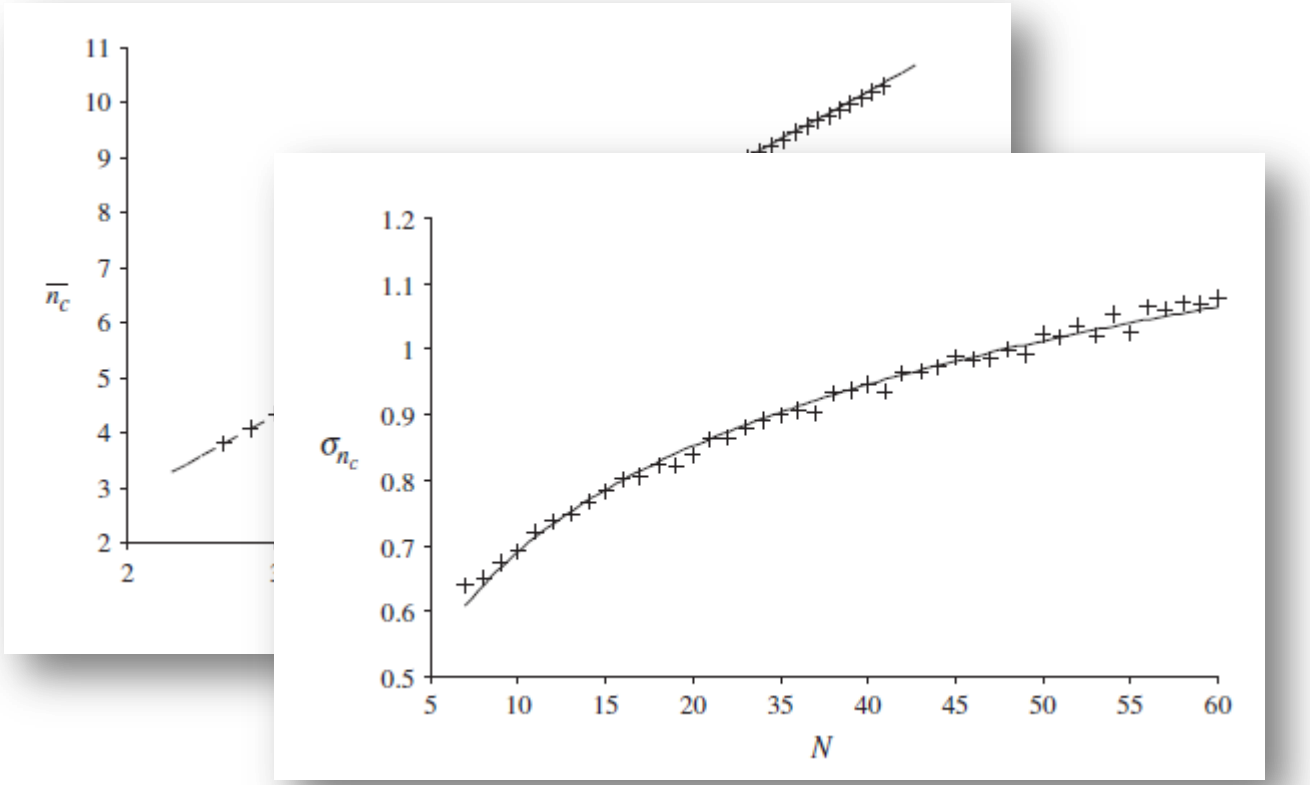
Size of the MAST (n_c) = 8

I_{cong}



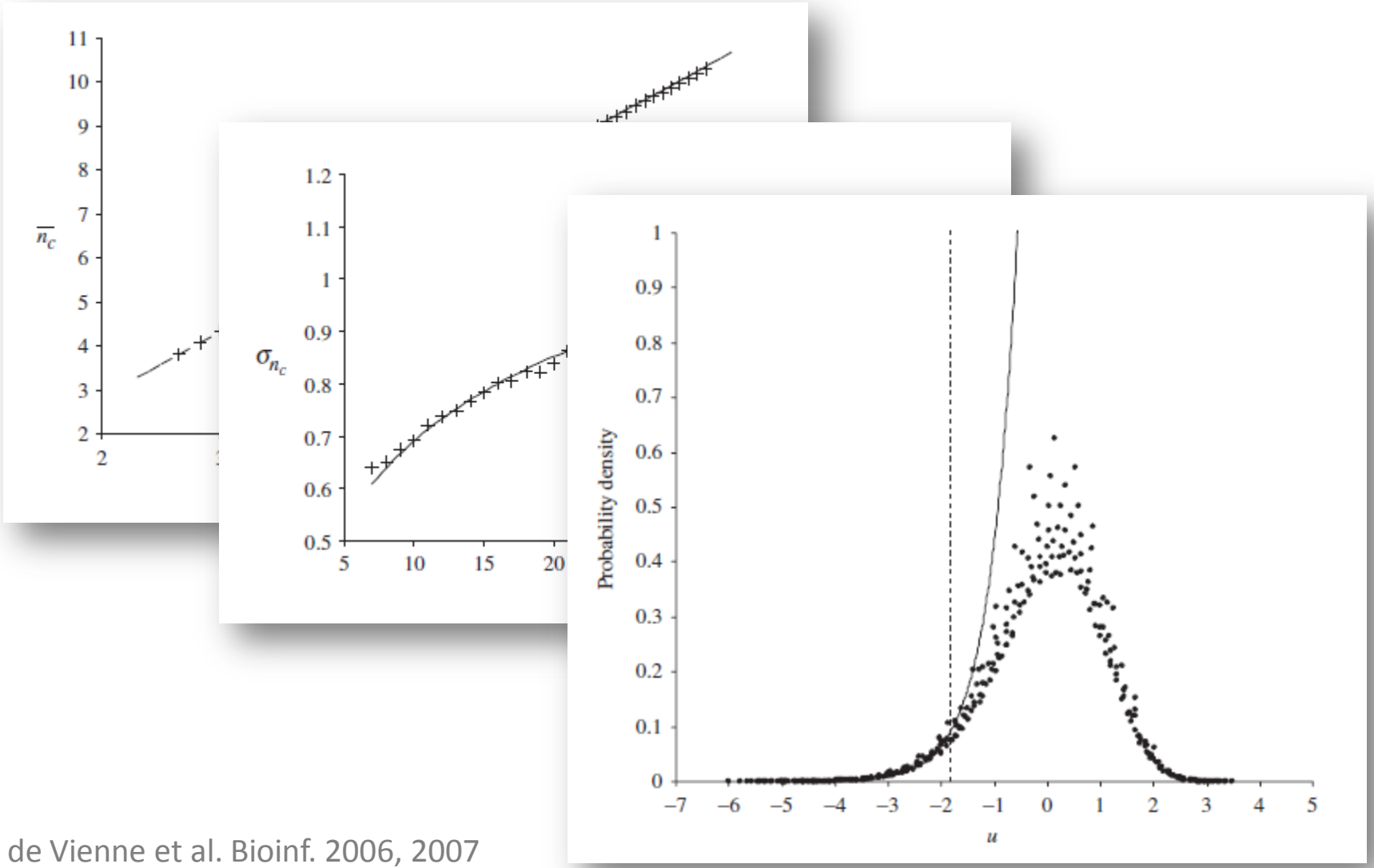
de Vienne et al. Bioinf. 2006, 2007

I_{cong}



de Vienne et al. Bioinf. 2006, 2007

I_{cong}



de Vienne et al. Bioinf. 2006, 2007

Literature survey reveals prevalence of host-shift speciation over cospeciation

New Phytologist

Tansley review Review 379

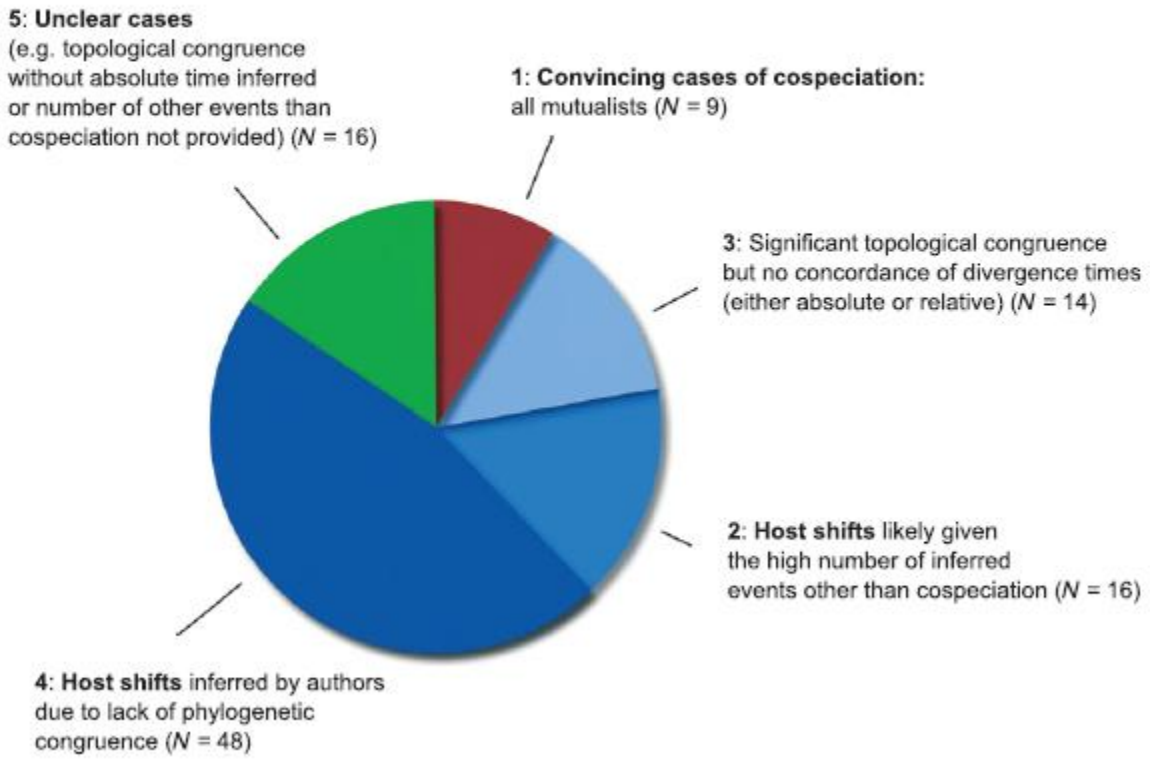


Fig. 5 Illustration of the literature survey in Table 2, with number of cases representing either convincing cases of cospeciation (in red), cases of host shifts inferred from incongruent topologies, discordant times of divergence or likely given the high number of duplication and extinction inferred (in blue), or finally unclear cases (in green).

de Vienne et al. New Phytologist. 2013

Link between [coevolution, specialization, speciation] and [cospeciation or host-shift speciation]

- Coevolution -> specialization -> speciation

(reviewed in Summers et al. 2003)

Speed depends on parasite and host generation time, dispersal rates, effective pop size, etc… (Huyse et al. 2005)

→Specialization of two parasite lineages on sister host species may result in a cospeciation event.

BUT is it what prevails in the long term?

MAYBE NOT / APPARENTLY NOT

Link between [coevolution, specialization, speciation] and [cospeciation or host-shift speciation]

When new species are formed:

loss of associated parasites (Enemy release, Kean & Crowley 2002, Genton et al. 2005)

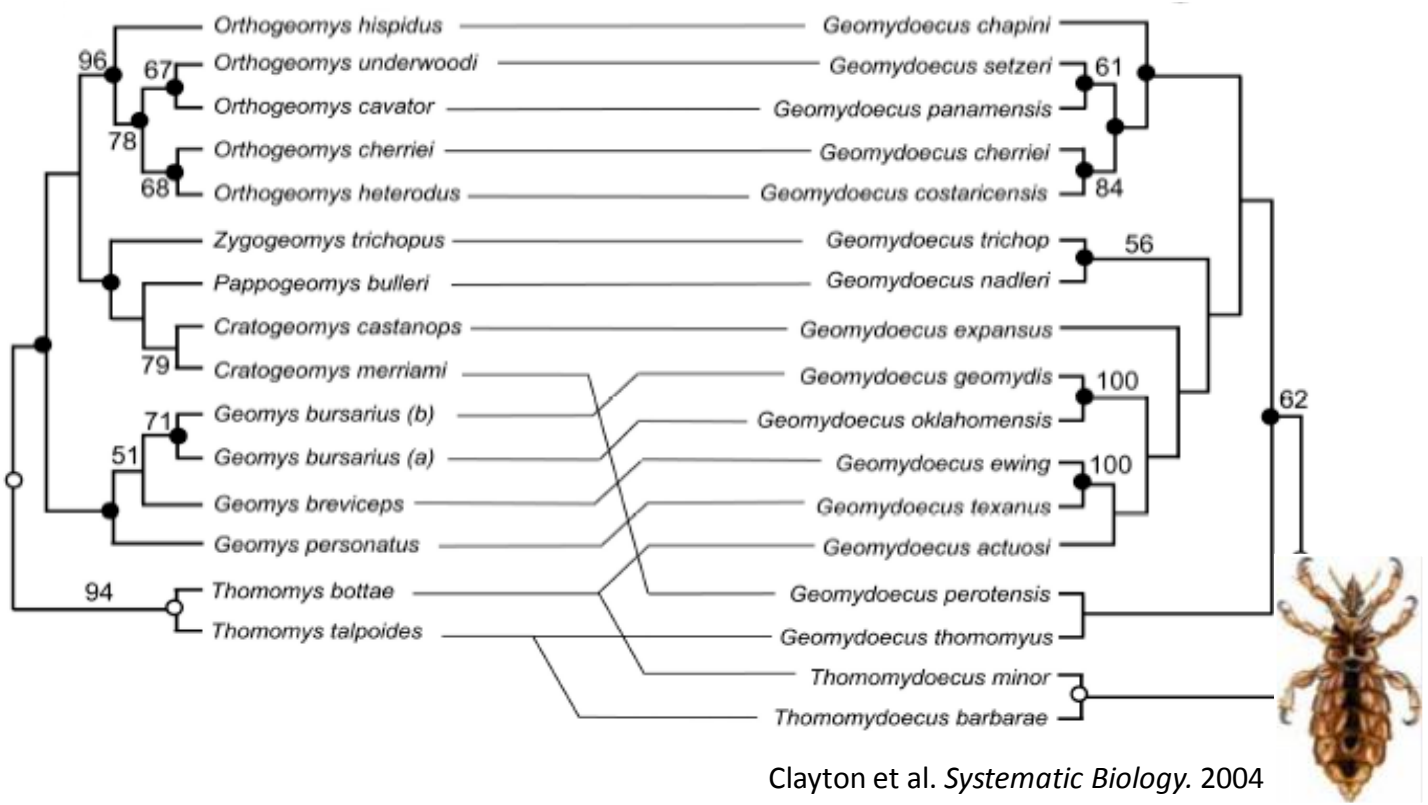
smaller population -> not compatible with specialist parasites (de Castro & Bolker, 2005)

but coevolution hinders the persistence of generalists…

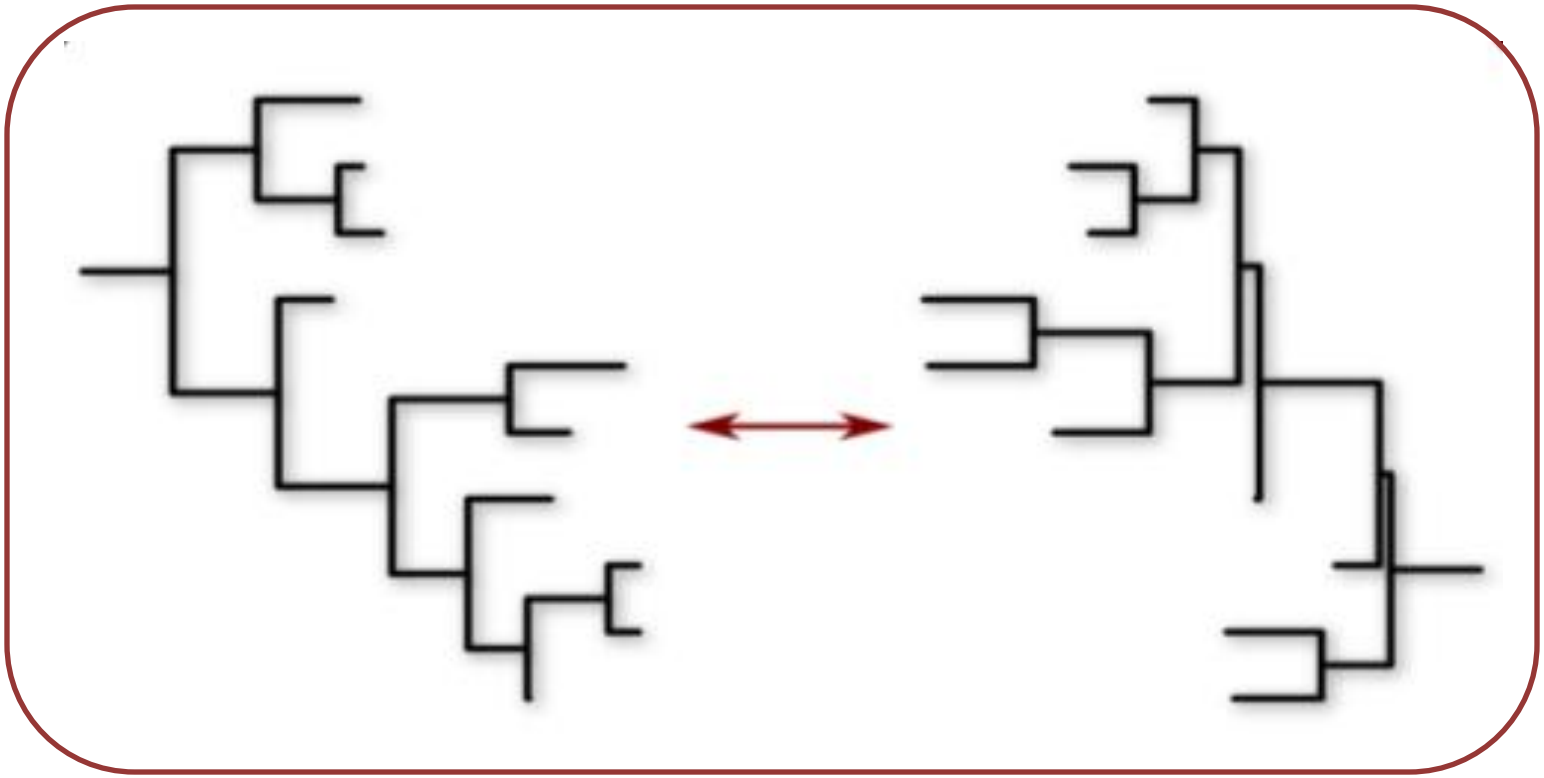
Against the idea coevolution leading to cospeciation.

MODEL? Under what conditions does host-parasite coevolution leads to cospeciation or to speciation by host-shifts?

Link between
[coevolution, specialization, speciation]
and
[cospeciation or host-shift speciation]



Clayton et al. *Systematic Biology.* 2004



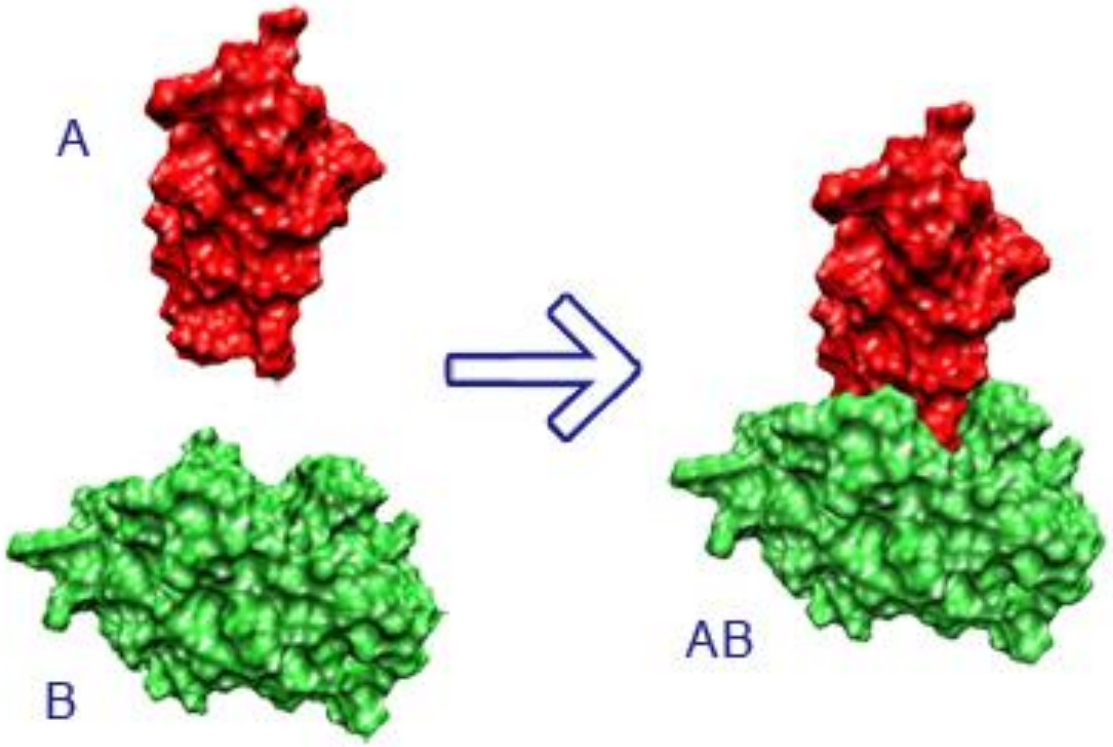
Species tree ↔ Species tree — Host parasite coevolutionary studies

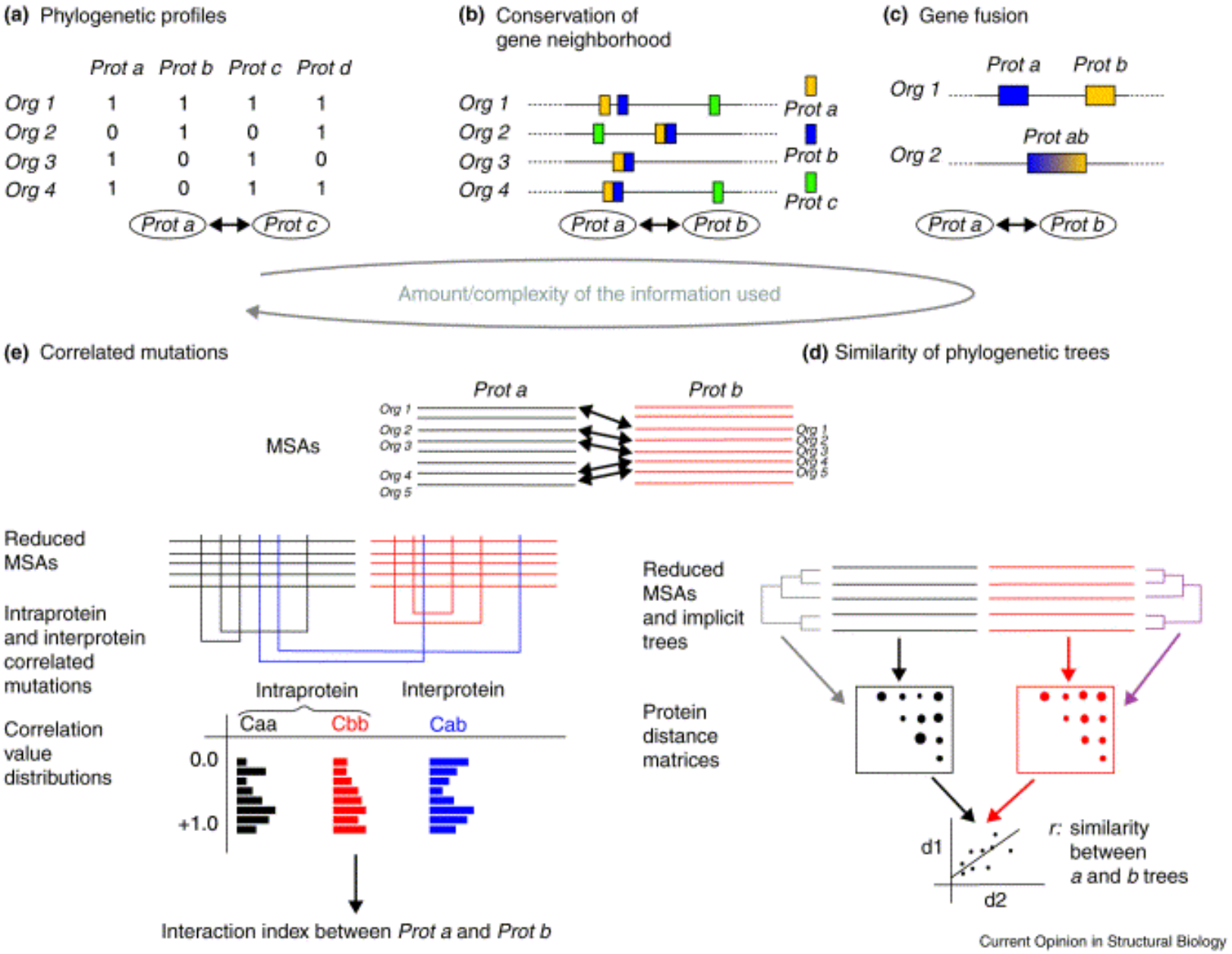
Gene tree ↔ Gene tree — Protein-Protein interaction detection

Species tree ↔ Gene tree — Reconciliation analyses

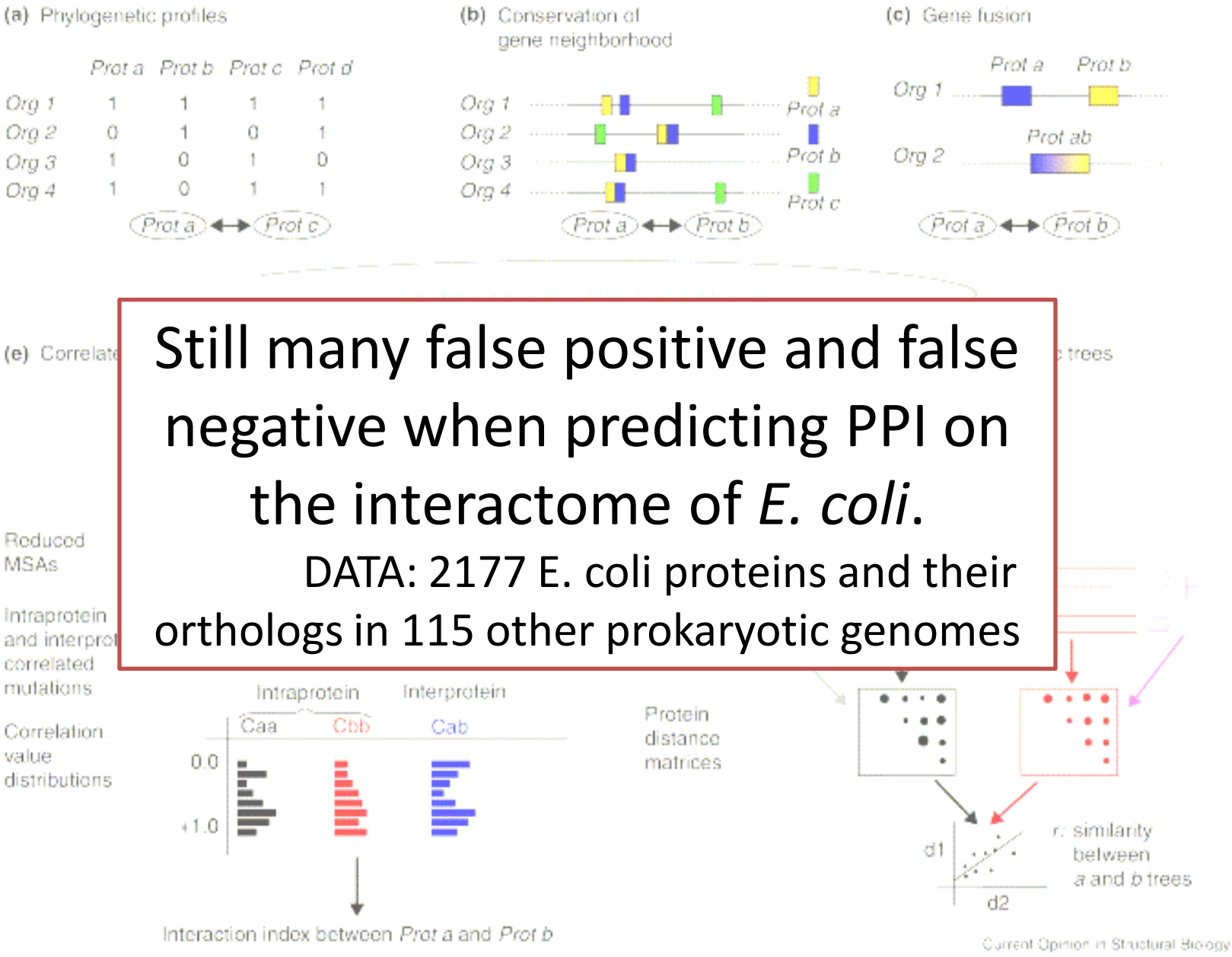
Multiple gene trees — Phylogenomic studies

Two proteins involved in the same complex/pathway are expected to be coevolving





Valencia A, & Pazos F. Curr. Opin. Struct. Biol. 2002

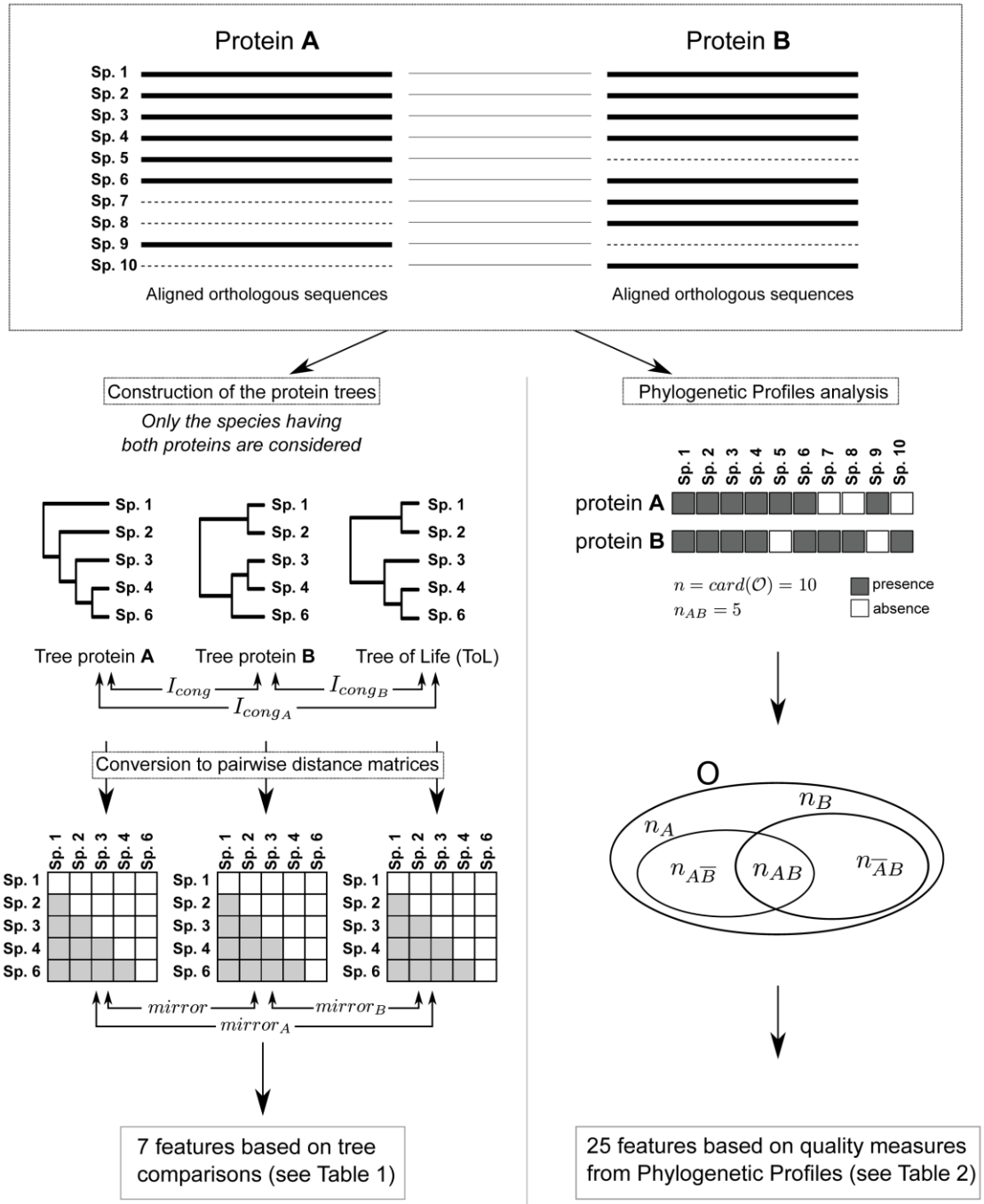


Valencia A, & Pazos F. Curr. Opin. Struct. Biol. 2002

- Development of new descriptors of coevolution

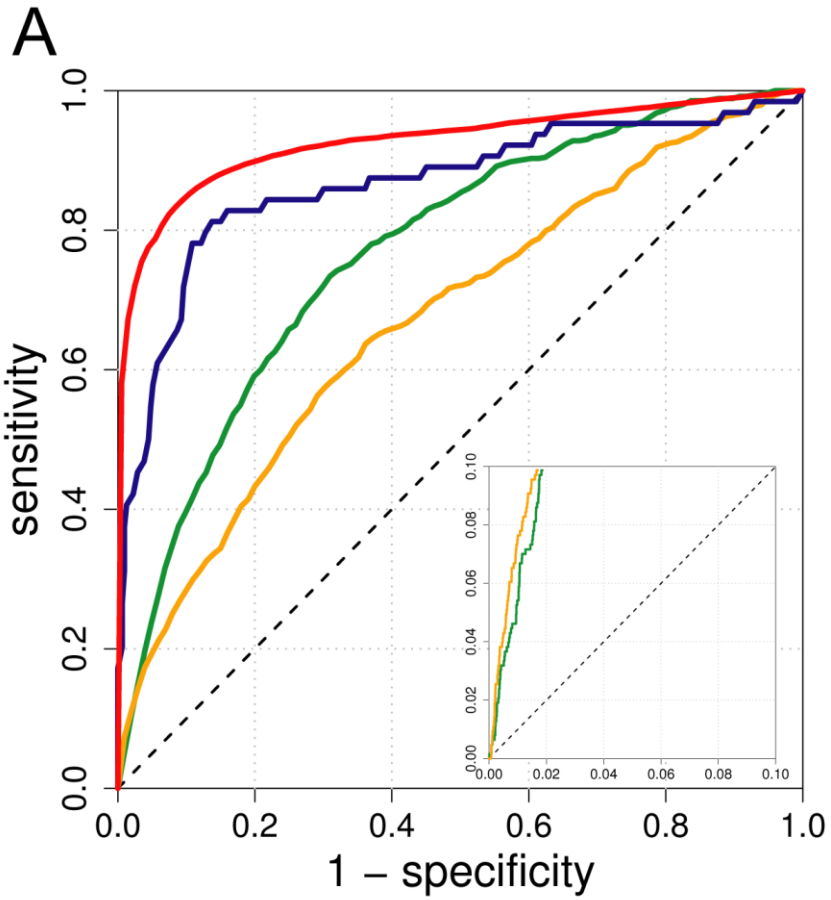
-Combination of all these features in a Machine Learning framework

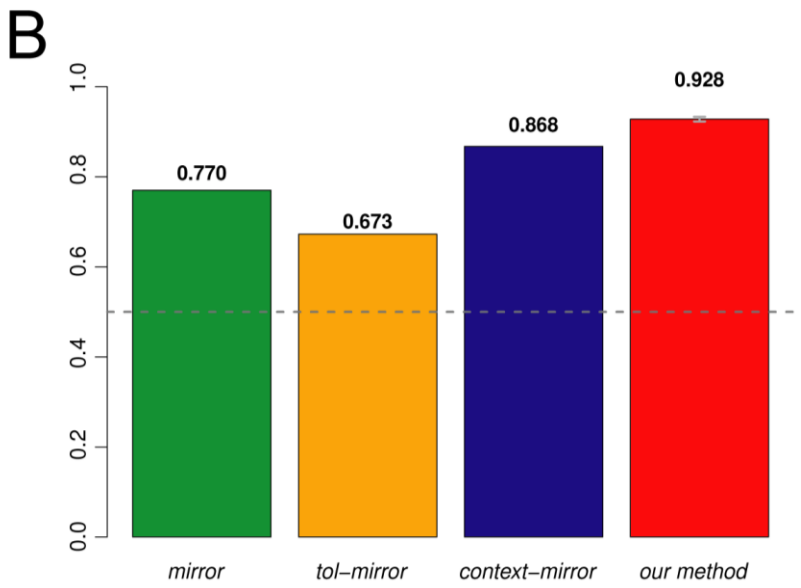
-Look at the capacity to correctly sort positive and negative pairs



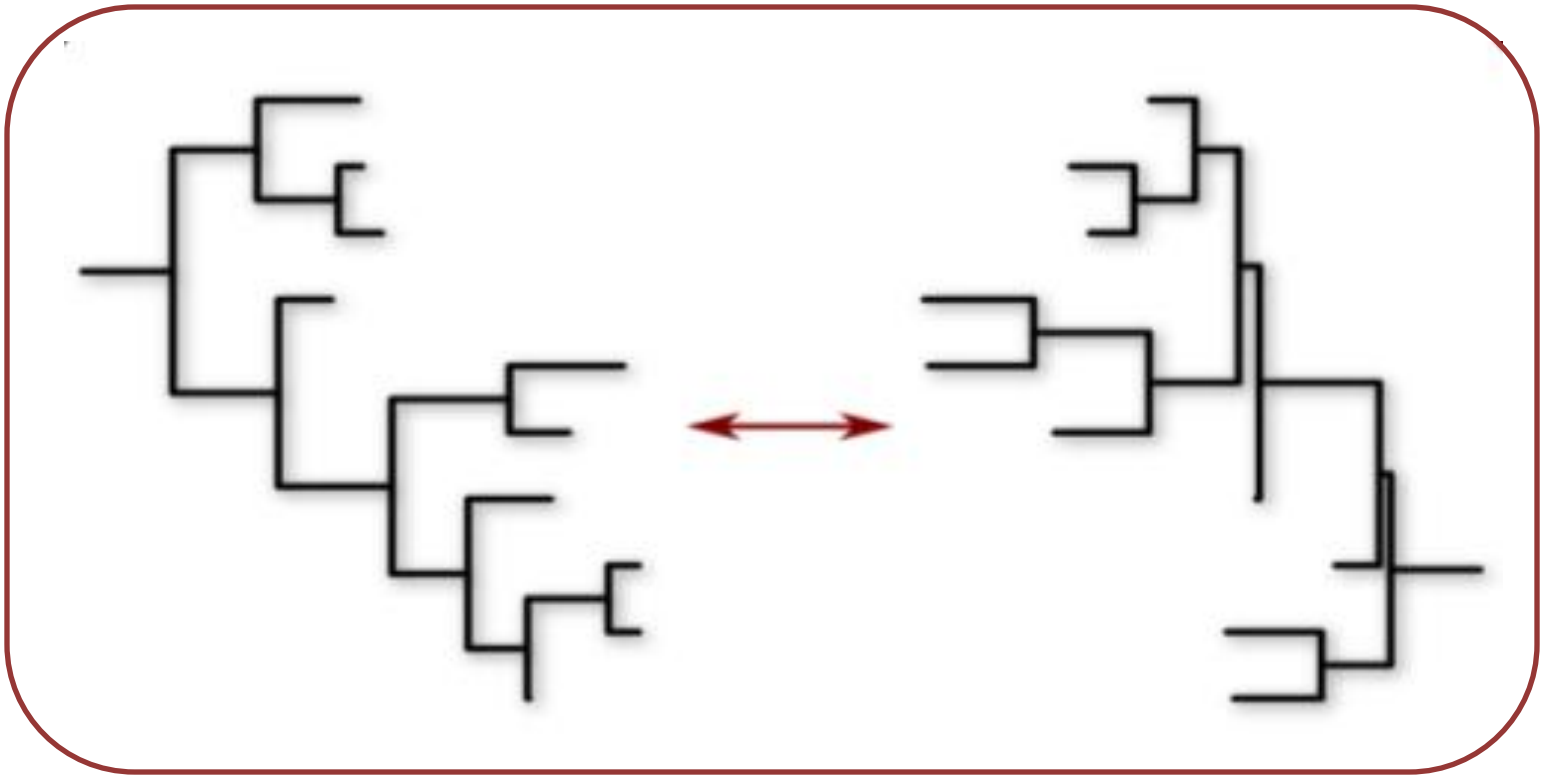
de Vienne and Azé. PLoS ONE. 2012

Combining multiple classifiers allows improving PPI prediction

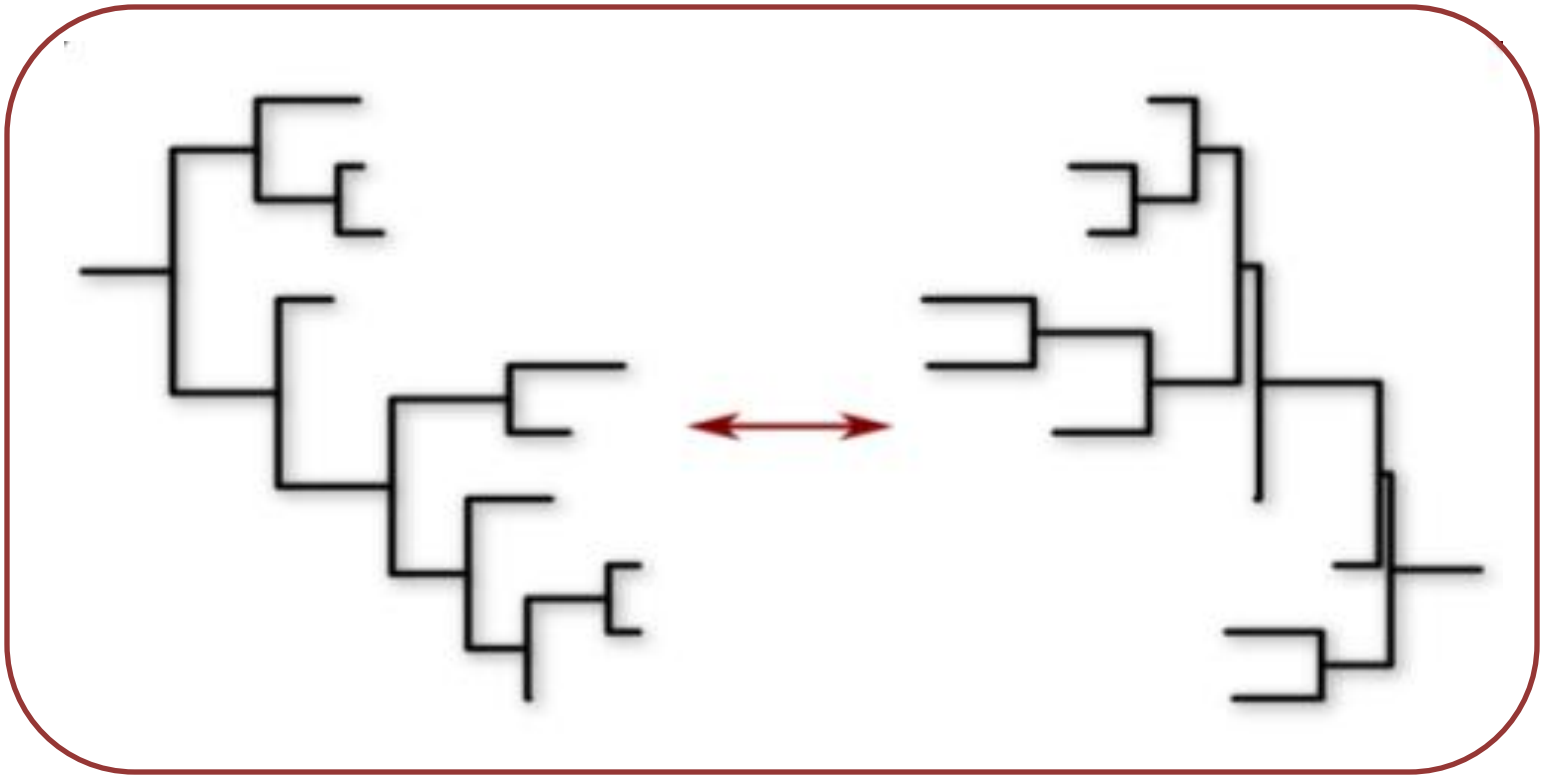




de Vienne and Azé. PLoS ONE. 2012



Species tree ↔ Species tree	**Host parasite coevolutionary studies**
Gene tree ↔ Gene tree	**Protein-Protein interaction detection**
Species tree ↔ Gene tree	**Reconciliation analyses**
Multiple gene trees	**Phylogenomic studies**

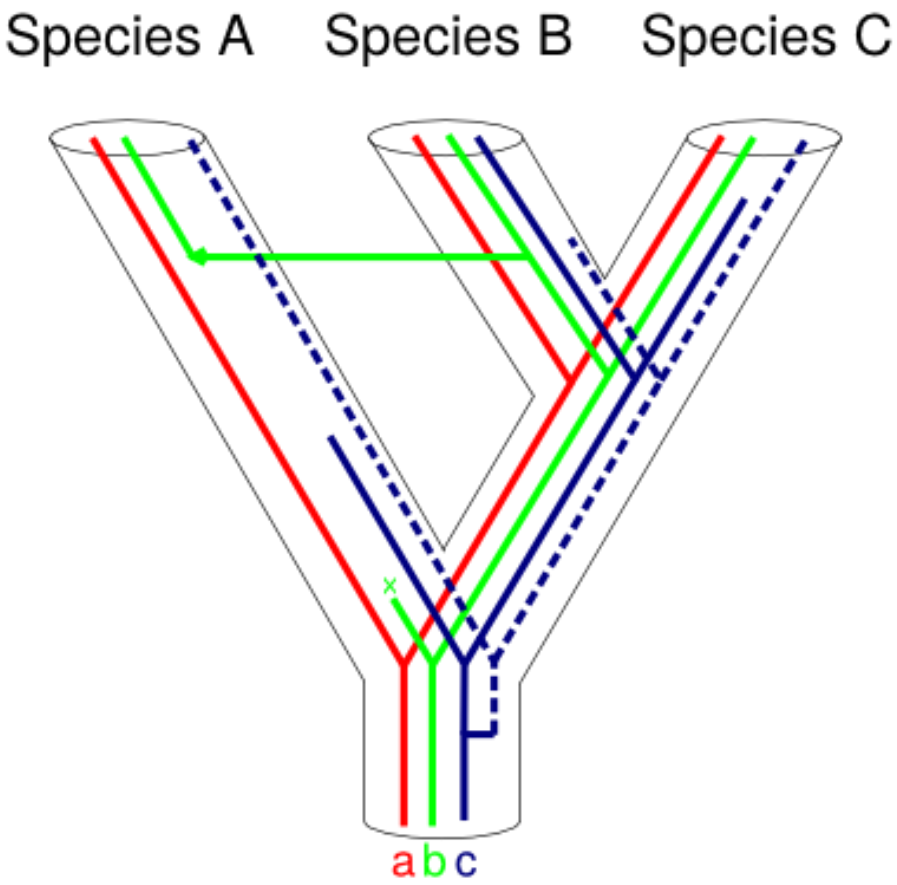


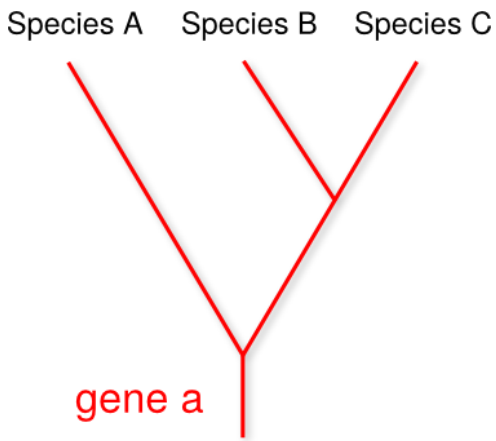
Species tree ↔ Species tree — Host parasite coevolutionary studies

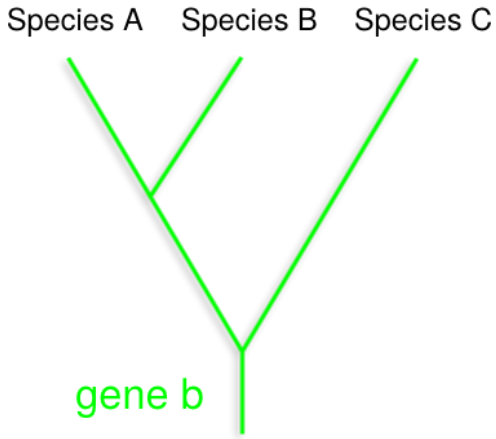
Gene tree ↔ Gene tree — Protein-Protein interaction detection

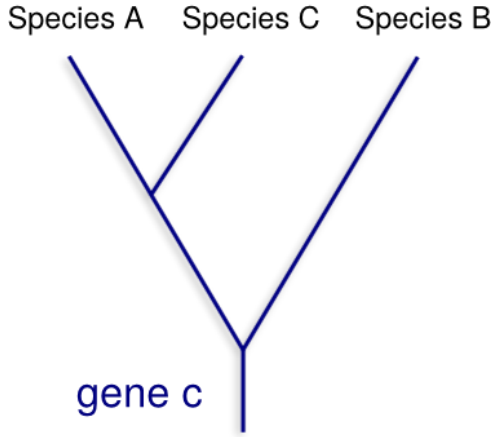
Species tree ↔ Gene tree — Reconciliation analyses

Multiple gene trees — Phylogenomic studies

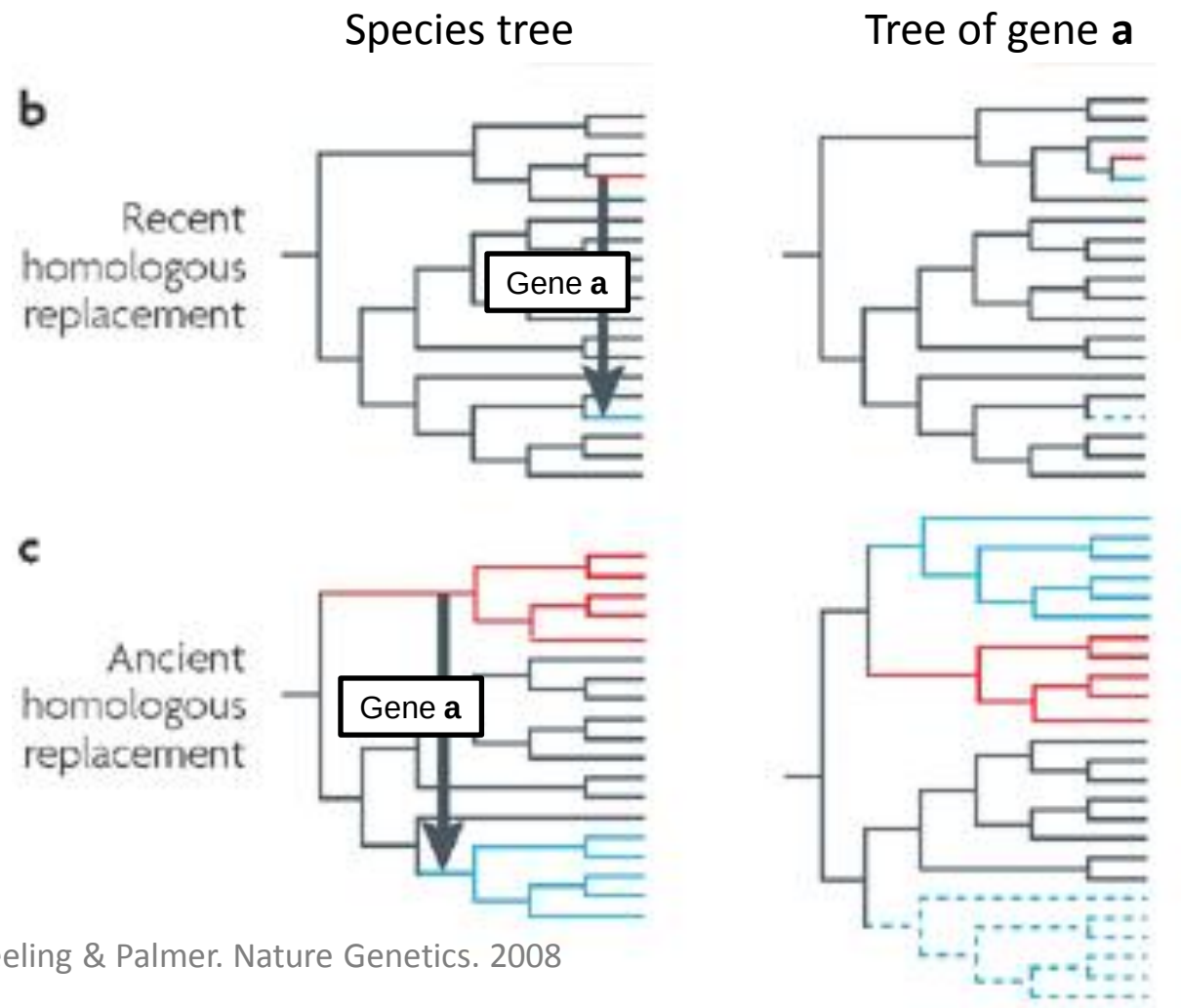

Species A
Species B
Species C
x
a b c


Species A
Species B
Species C
gene a


Species A
Species B
Species C
gene b

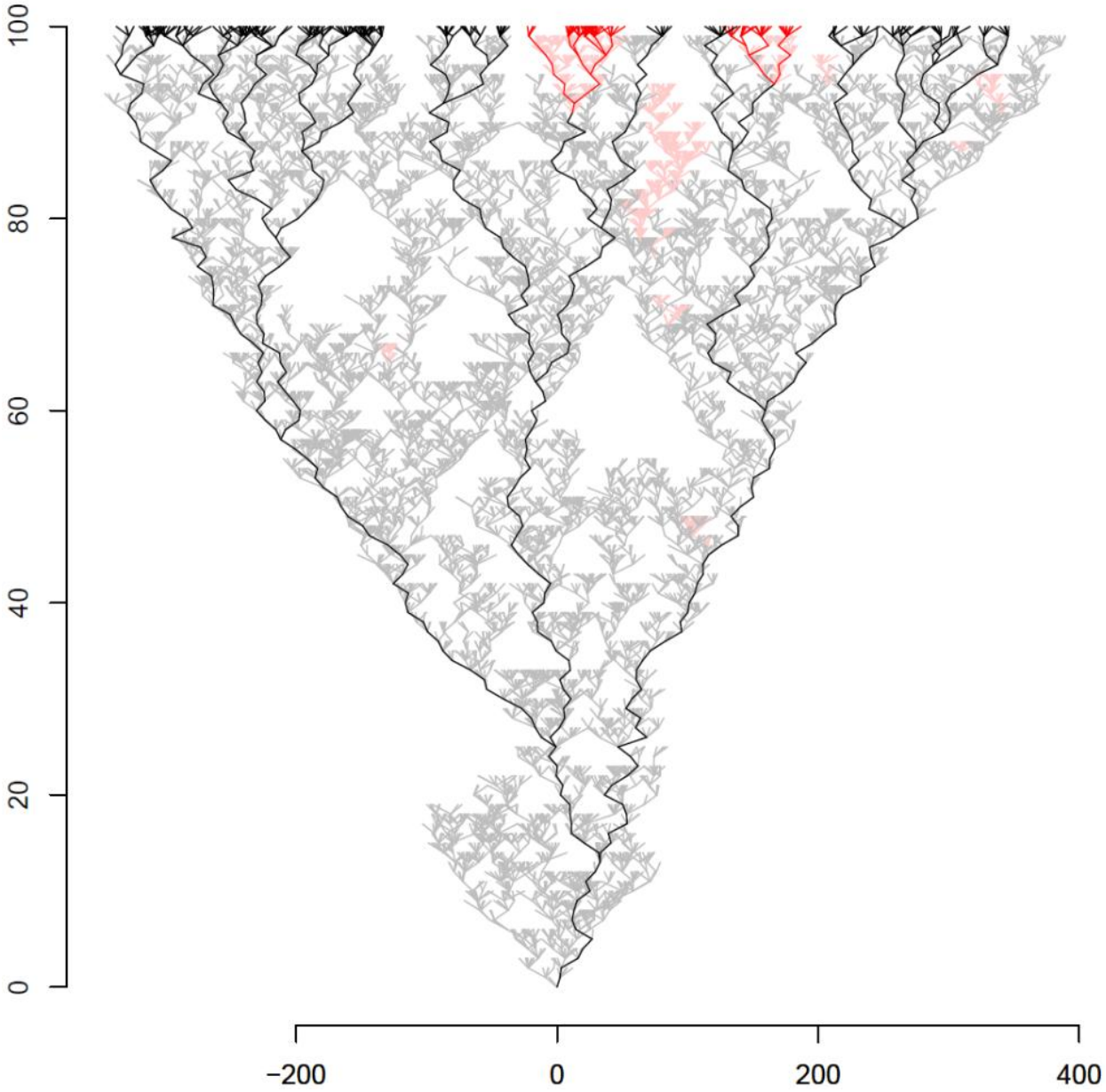

Species A
Species C
Species B
gene c

Consequence of horizontal gene transfer on gene tree topology



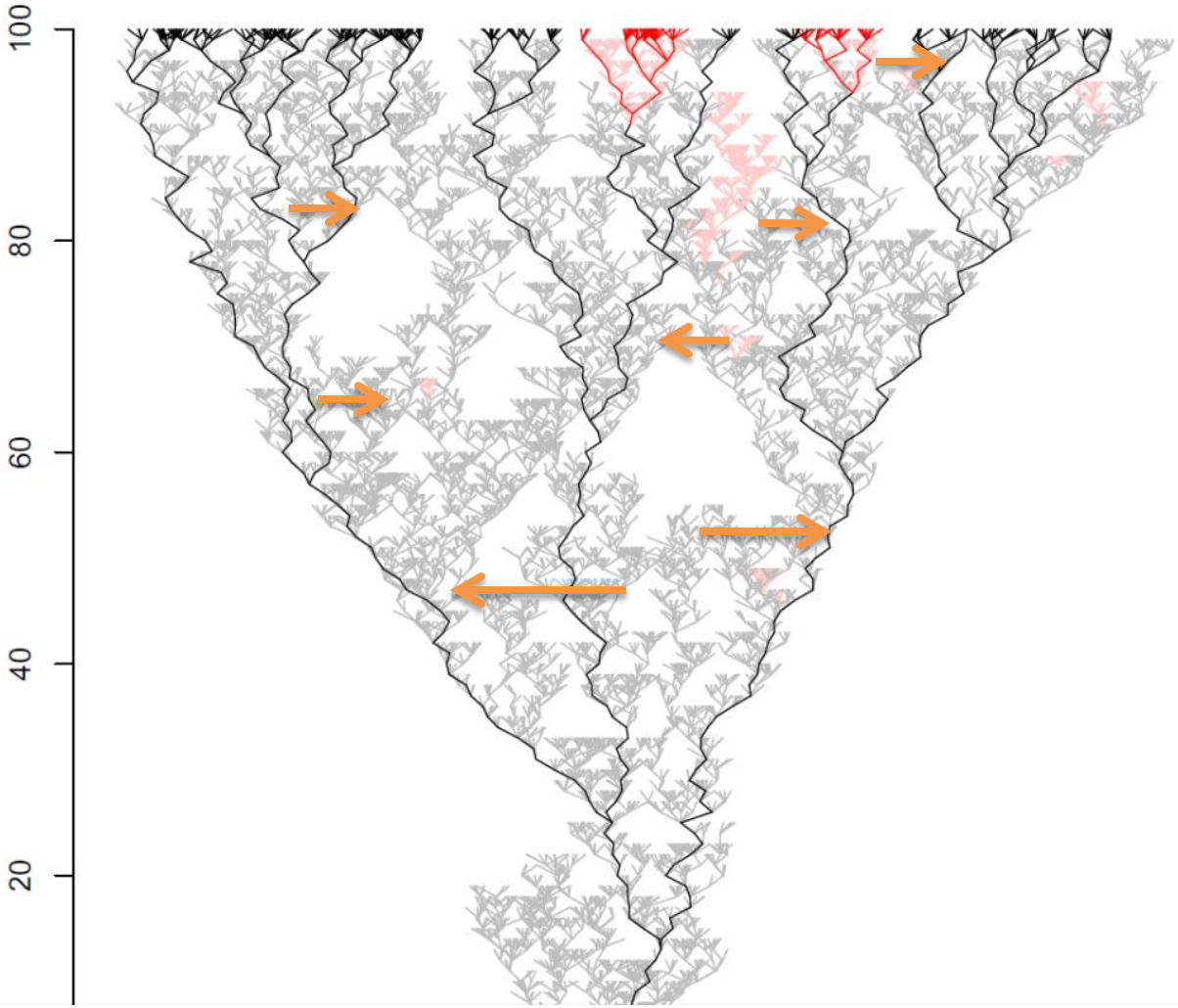
Modified from Keeling & Palmer. Nature Genetics. 2008

“To a first approximation, all species are extinct”

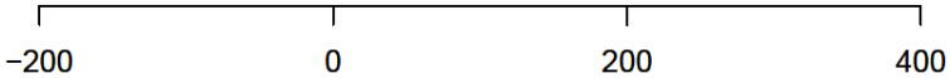


de Vienne, Giraud, Gouyon. PLoS ONE. 2013

“To a first approximation, all species are extinct”

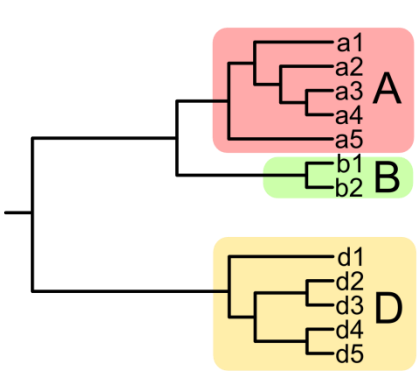


MANY genes were transferred from now extinct of unknown species

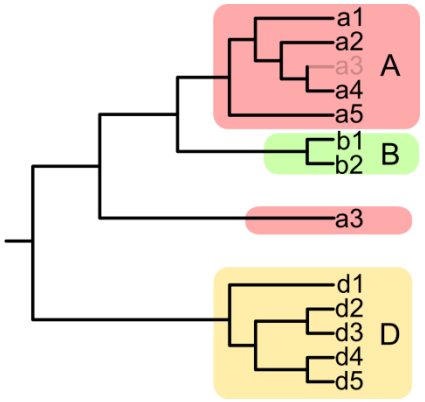


de Vienne, Giraud, Gouyon. PLoS ONE. 2013

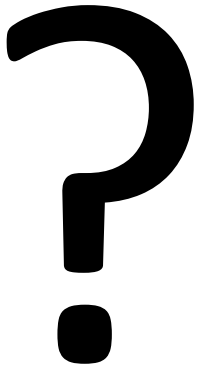
CAN we use HGT detection to explore extinct or unknown diversity?



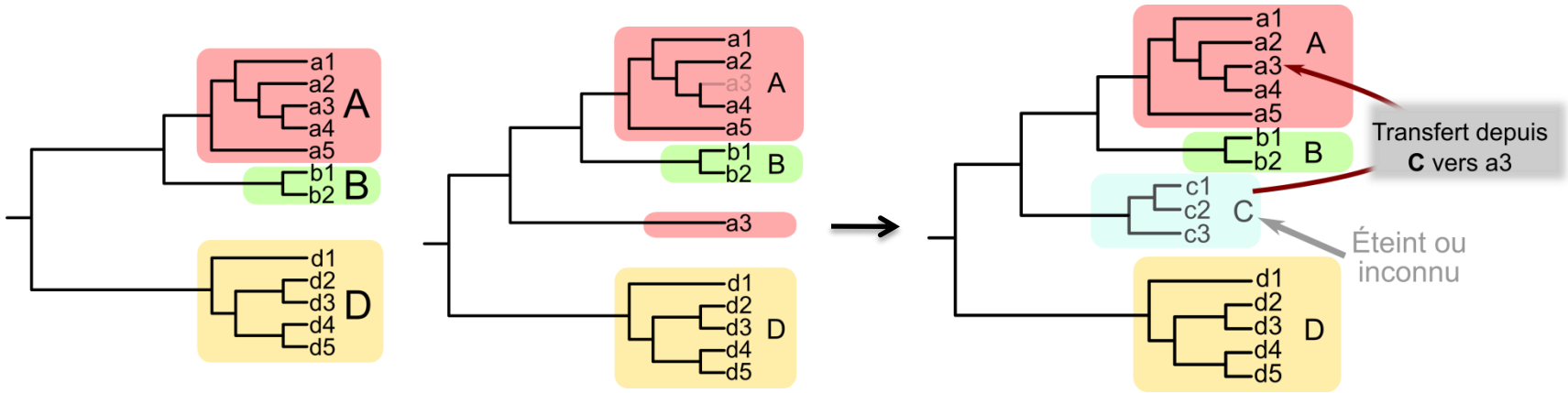
Species tree



Gene tree

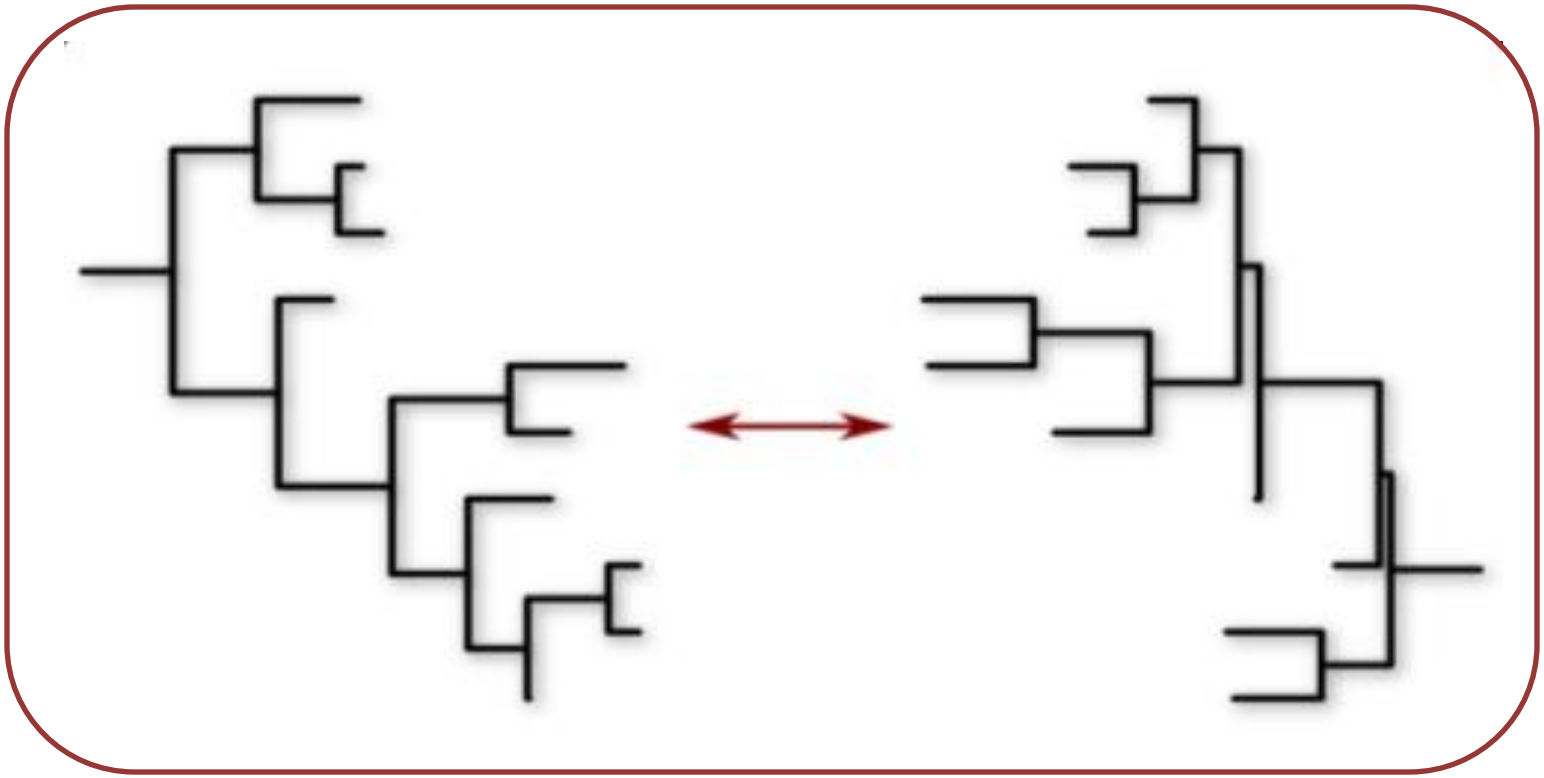


CAN we use HGT detection to explore extinct or unknown diversity?



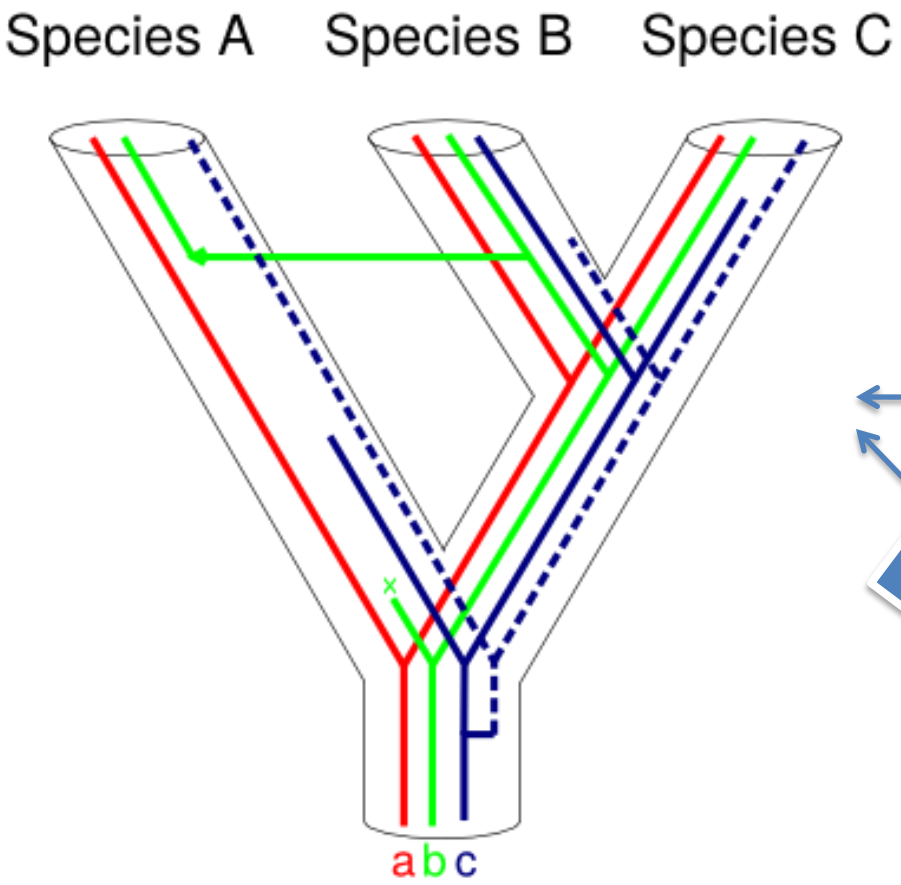
Species tree | Gene tree | Most parsimonious explanation

- Simulate species evolution with speciation, extinction, duplication
- Estimate sets of parameters allowing detection of extinct/unknown diversity
- Test on real bacterial dataset

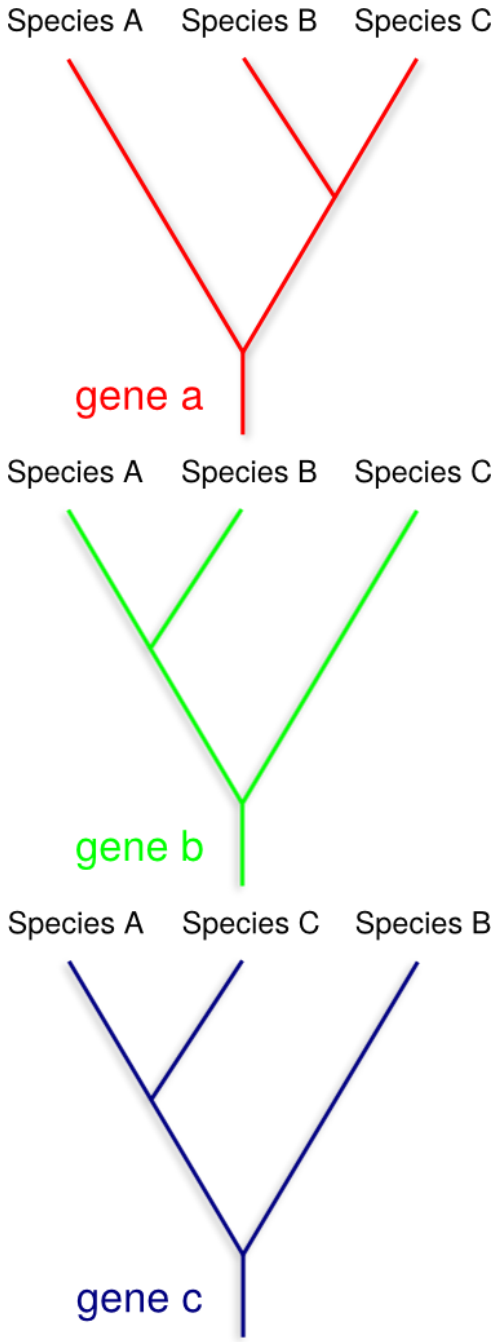


Species tree ↔ Species tree	**Host parasite coevolutionary studies**
Gene tree ↔ Gene tree	**Protein-Protein interaction detection**
Species tree ↔ Gene tree	**Reconciliation analyses**
Multiple gene trees	**Phylogenomic studies**

Incongruences between gene trees are common in phylogenomics



Incongruences



Existing methods for dealing with this variability

- Supermatrices
- Supertree
- Bayesian sampling
- Agreement subtrees
- Networks

Existing methods for dealing with this variability

- Supermatrices
- Supertree
- Bayesian sampling
- Agreement subtrees
- Networks

However, by concatenating the multilocus data, or by summarizing or obtaining a consensus of their individual gene trees, the latter methods lose a wealth of potentially interesting information, especially by removing outlier data or trees.

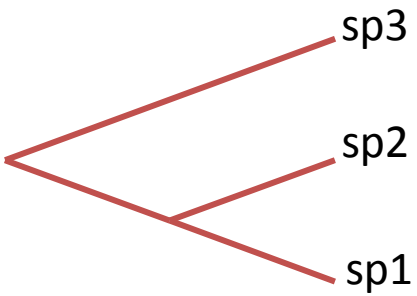
Phylo-MCOA

The goal

- Quickly compare a large number of gene trees
- Visualize the overall evolutionary history of the group analyzed.
- Find gene trees that tell the same story
- Find gene trees that tell different stories (produce discordant topologies)
  - Identify the species responsible for the discordance
  - Identify candidates for interesting biological processes

Principle

Gene tree G1



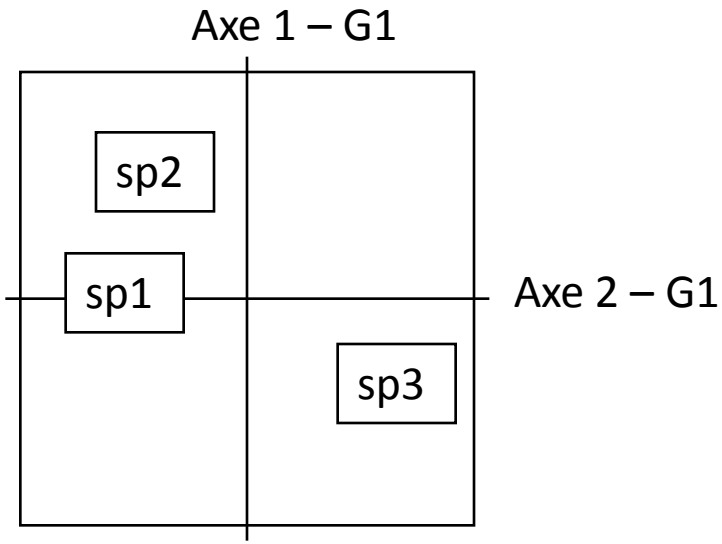
Alignment for gene G1

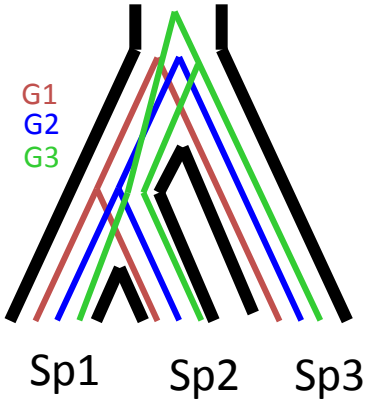
Sp1	AATGTGCATAC
Sp2	AATCTGCATAC
Sp3	AATGTCATTAC

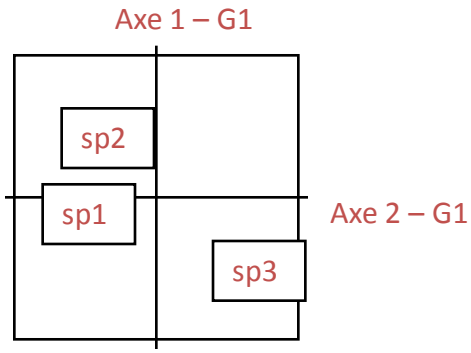
	Sp1	Sp2	Sp3
Sp1	0	x	y
Sp2	x	0	z
Sp3	y	z	0

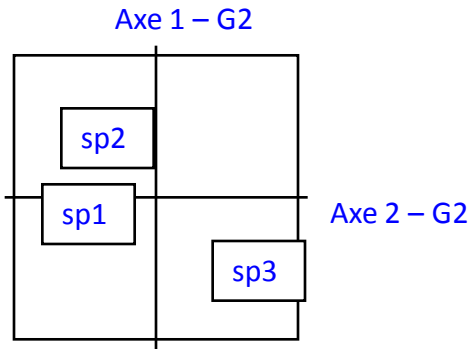
Pairwise distance matrix

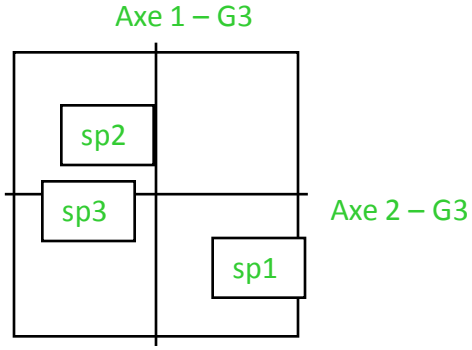
PCO



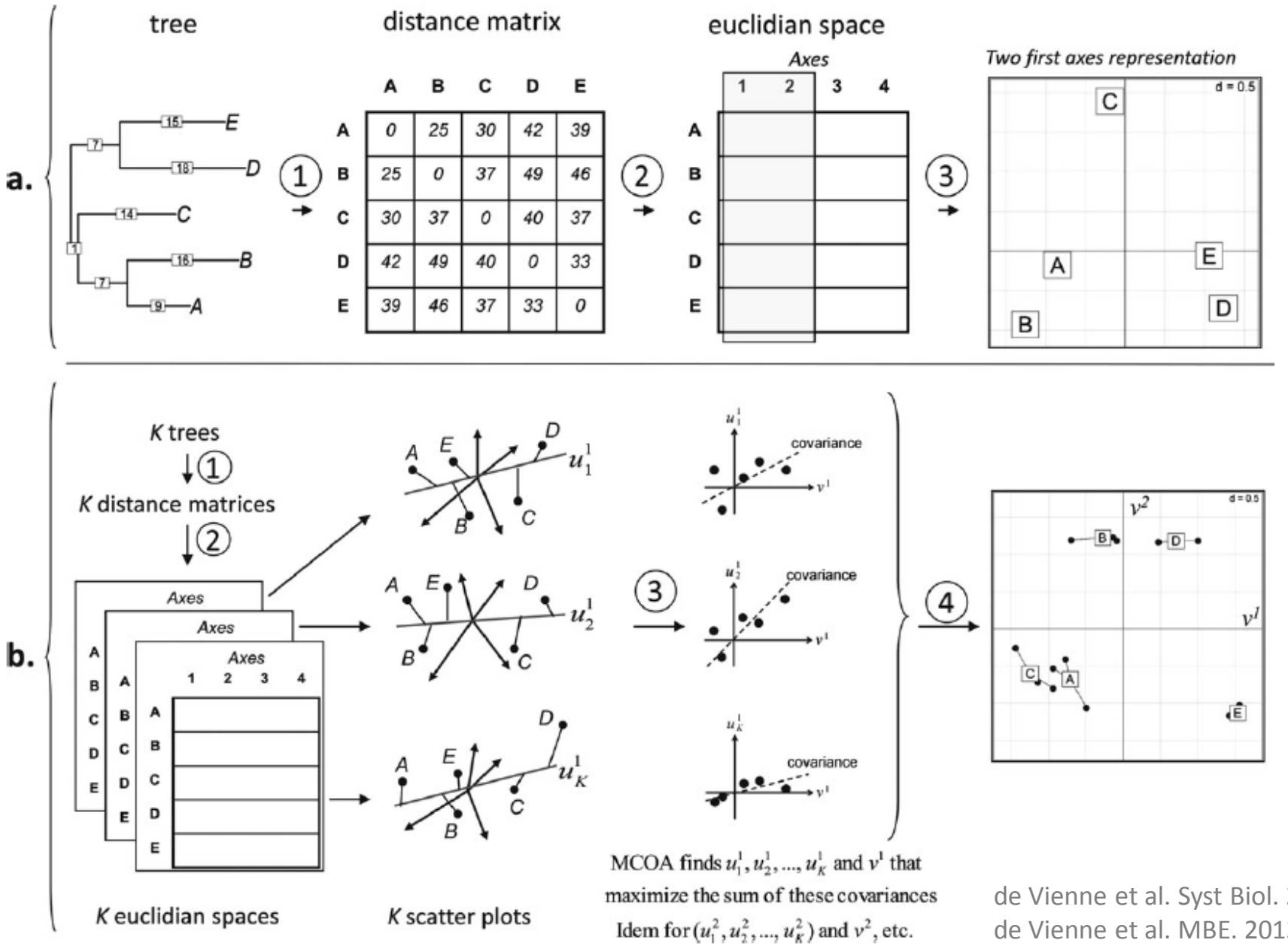




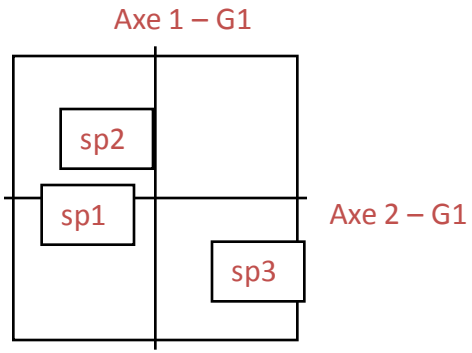


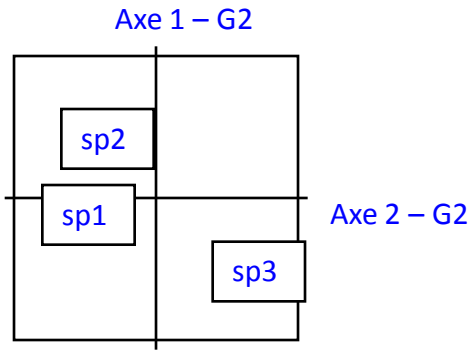


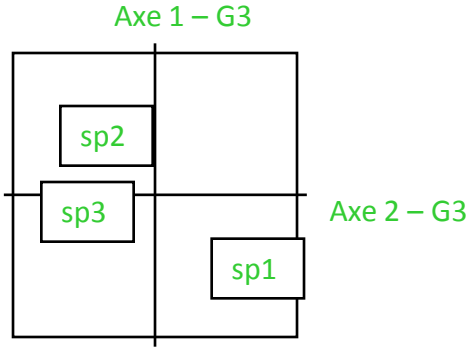
We want to visualize if the three genes tell the same story.
So we want to compare the individual PCOs.
So we need common axes. That's what MCOA does.



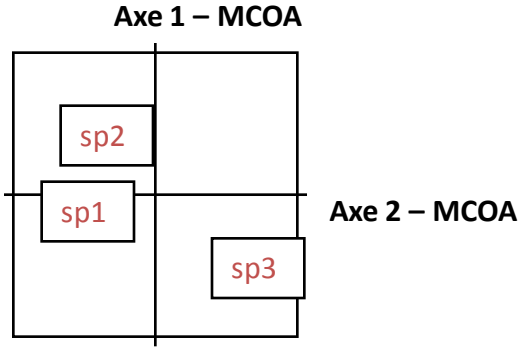
de Vienne et al. Syst Biol. 2012
de Vienne et al. MBE. 2013

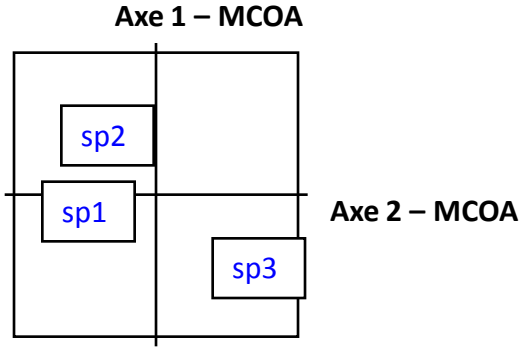


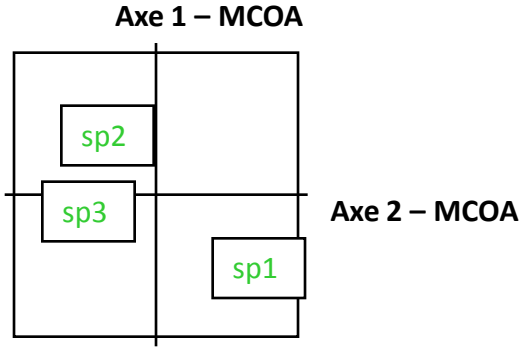




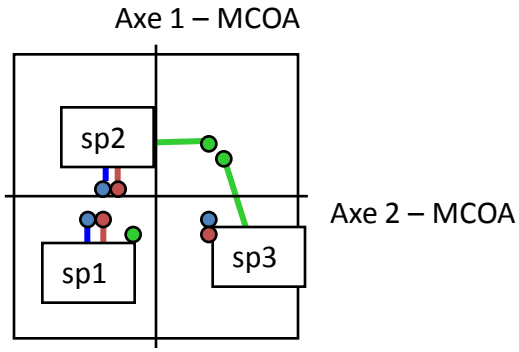
MCOA coordinates the individual axes





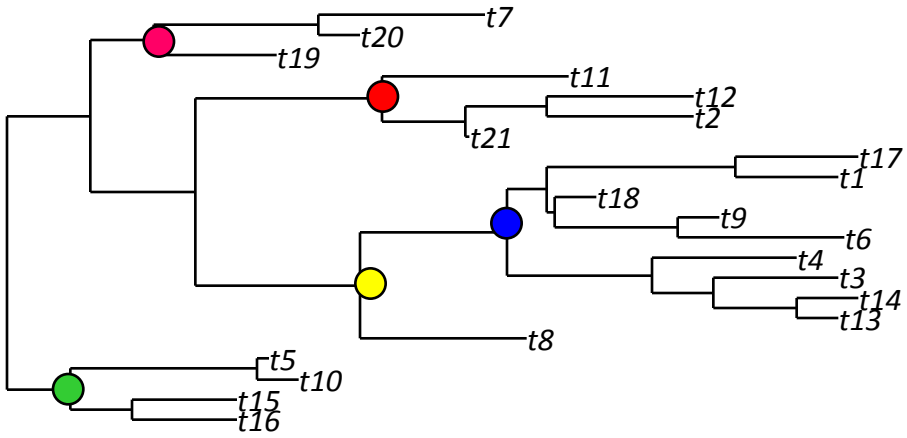


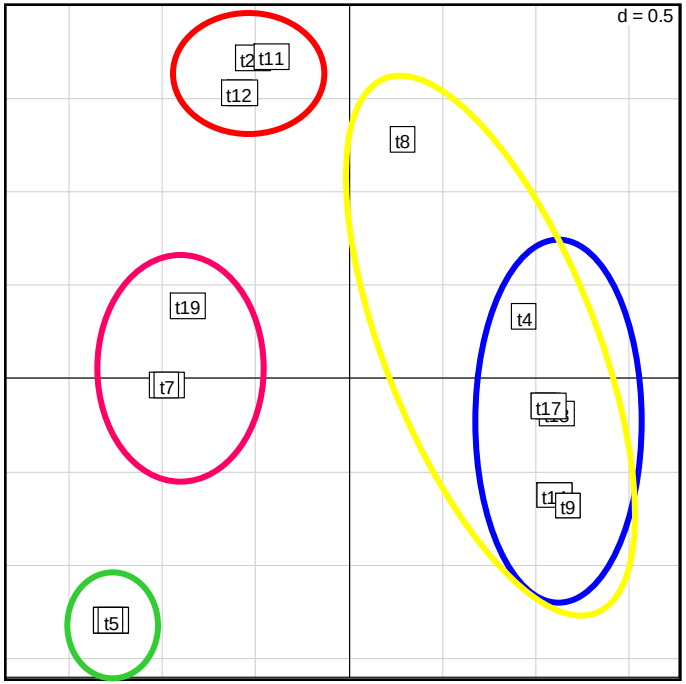
We can produce a single plot to visualize the concordance



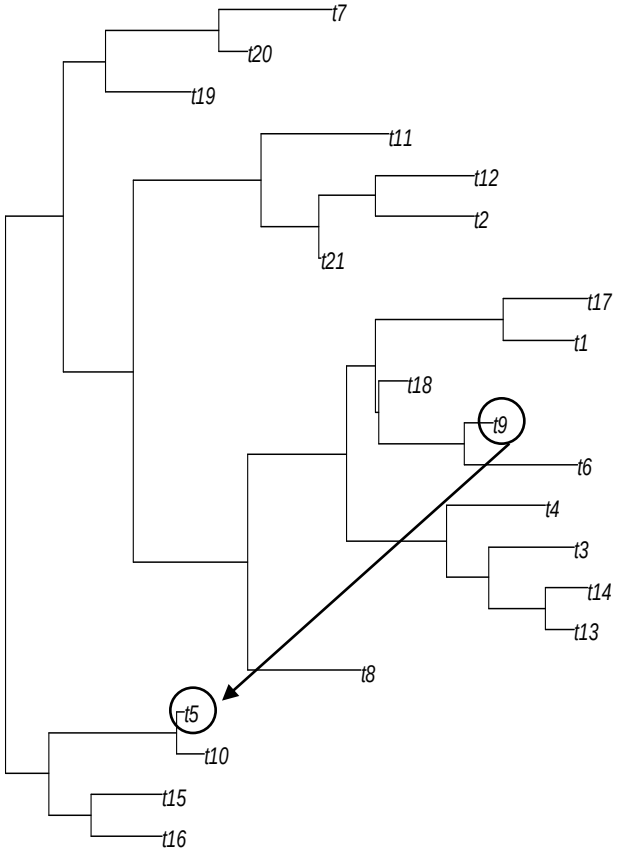
Cohesion plot

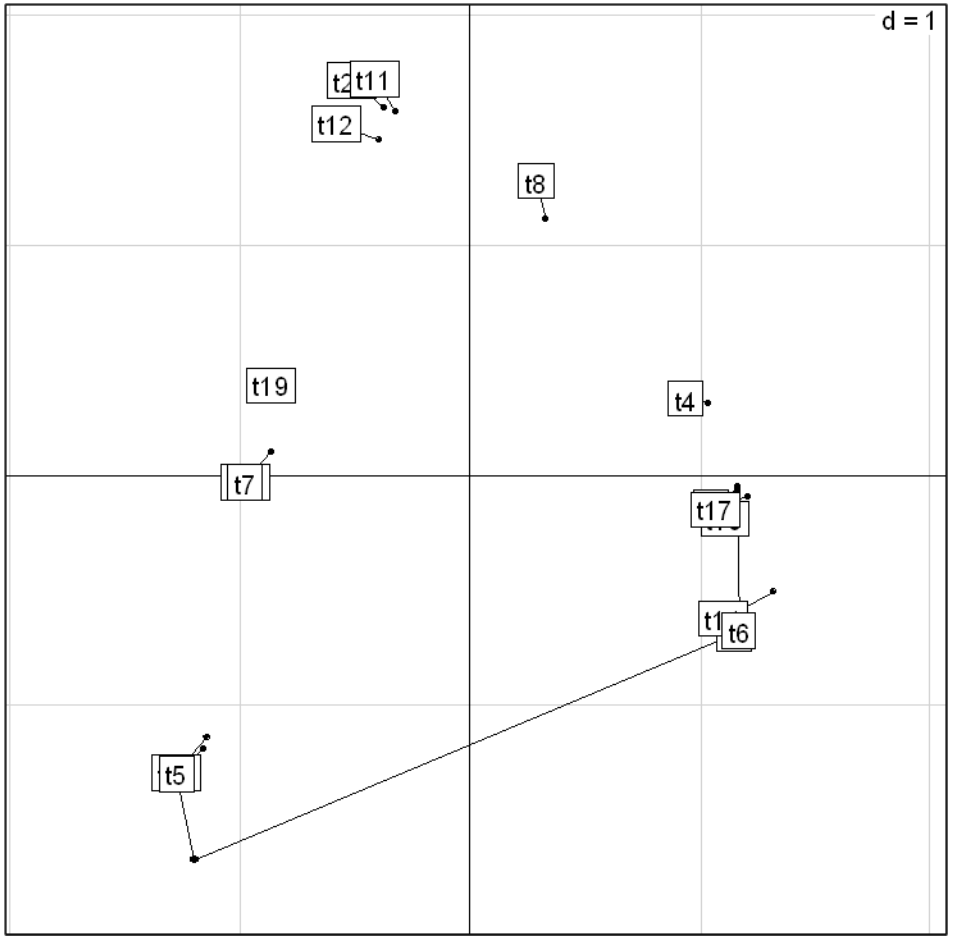
All the genes tell the same story



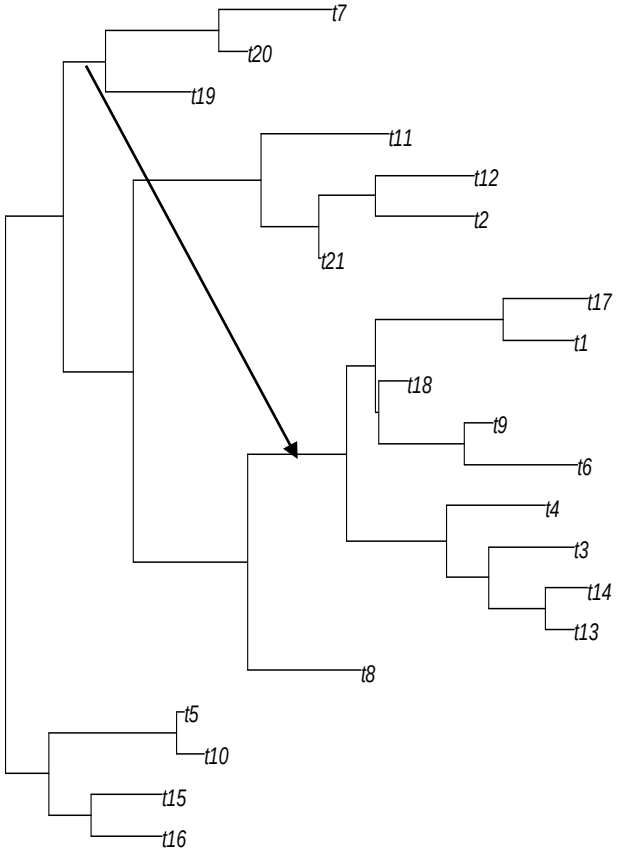


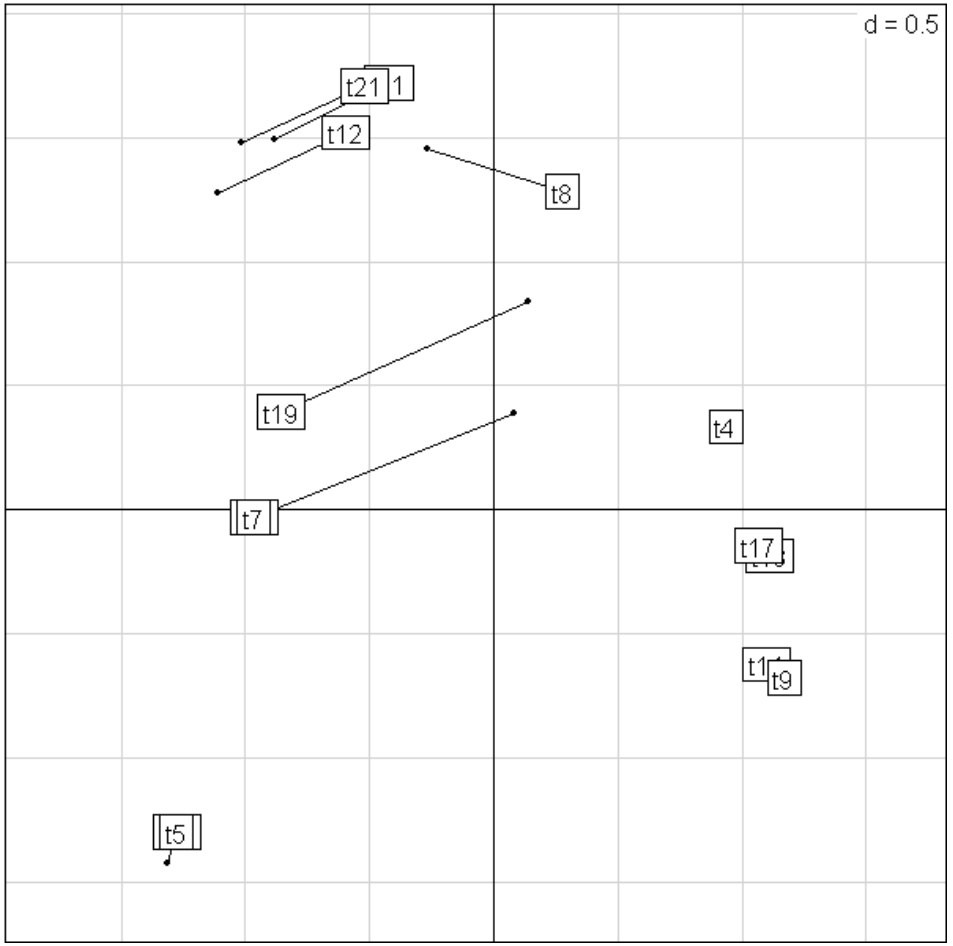
All the genes tell the same story + 1 recent HGT



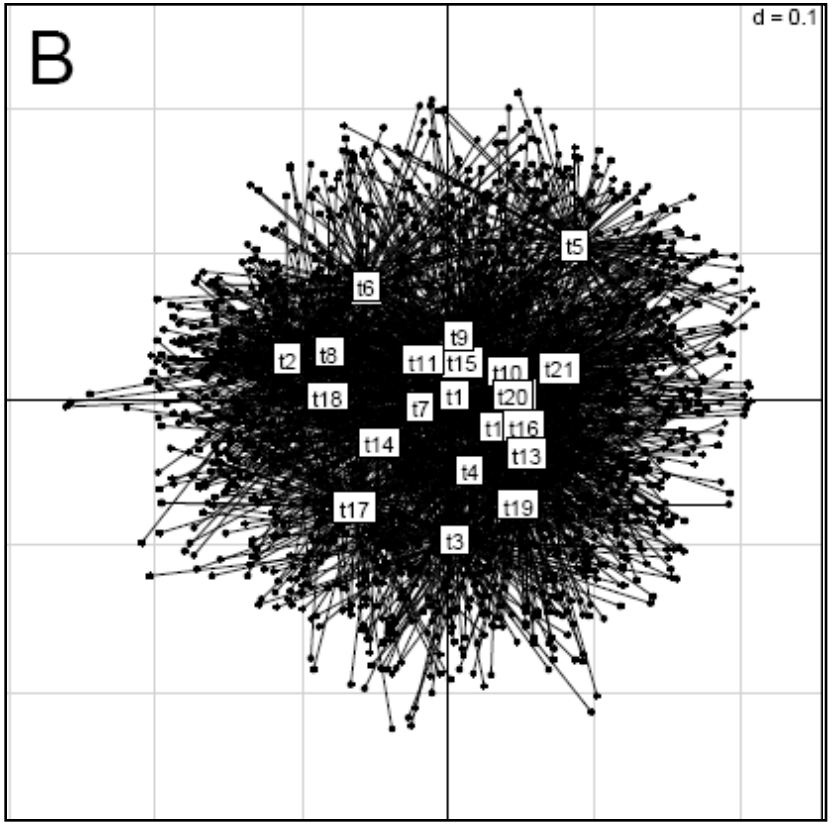


All the genes tell the same story + 1 ancient HGT

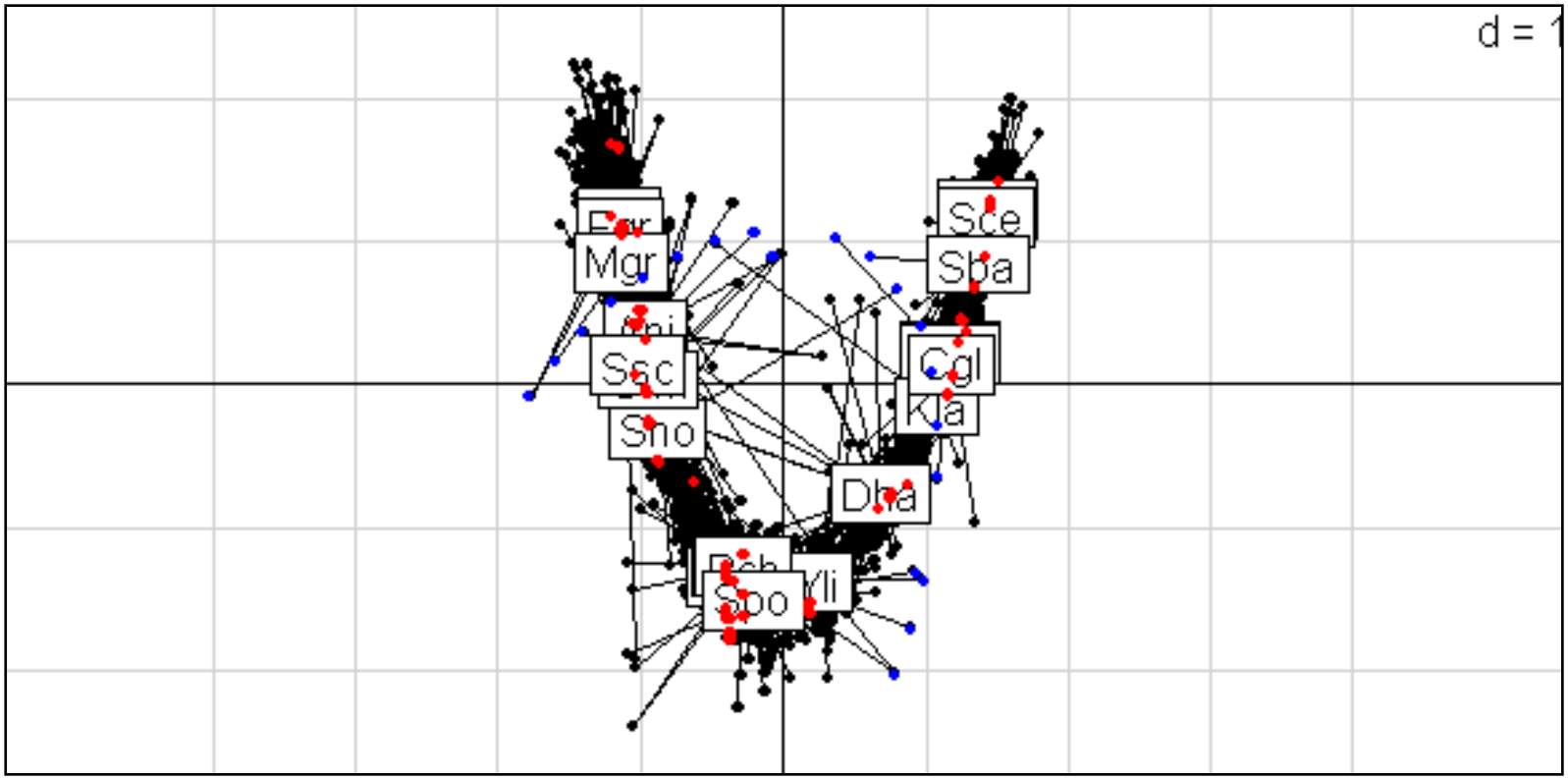




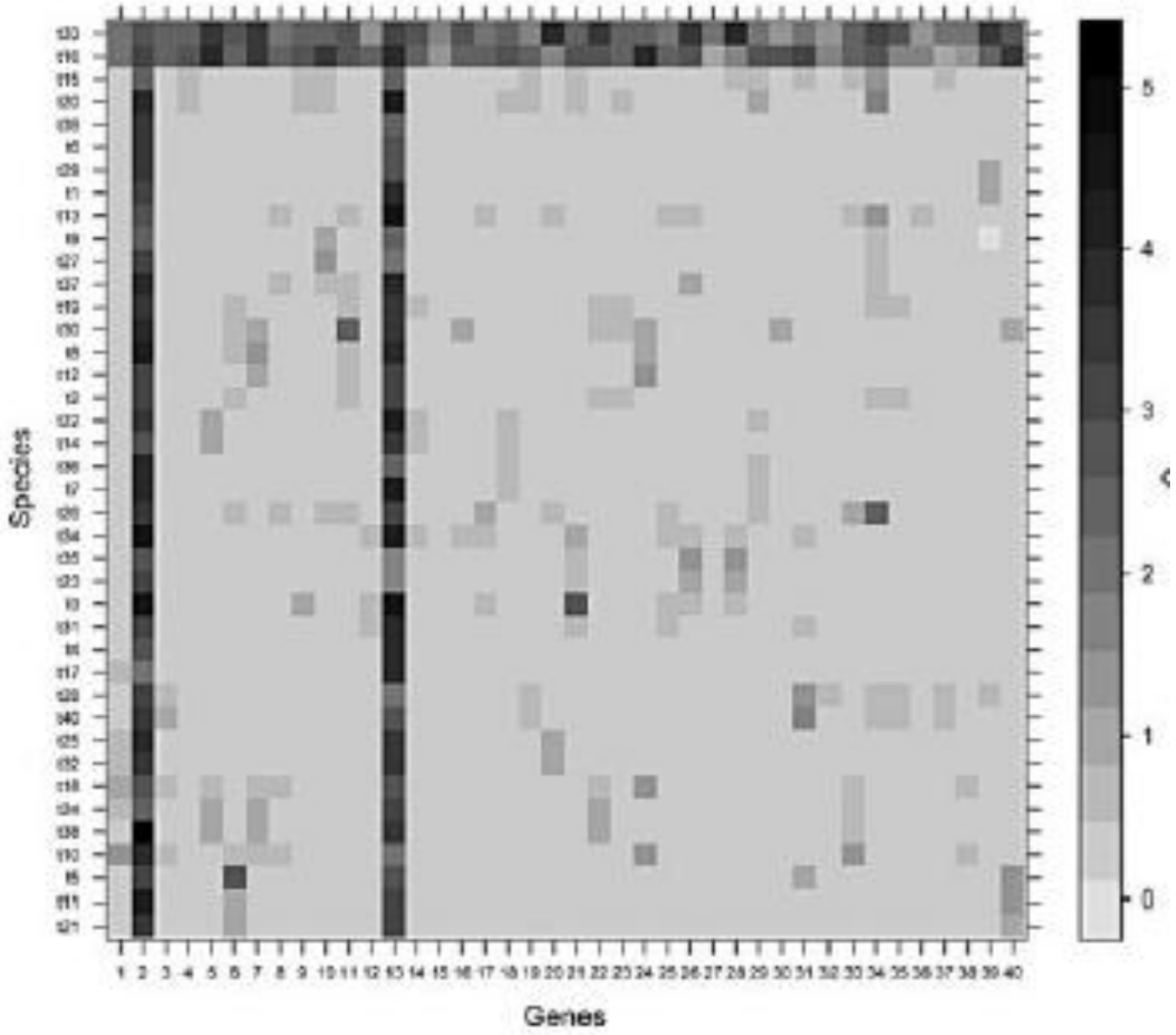
All the genes tell different stories



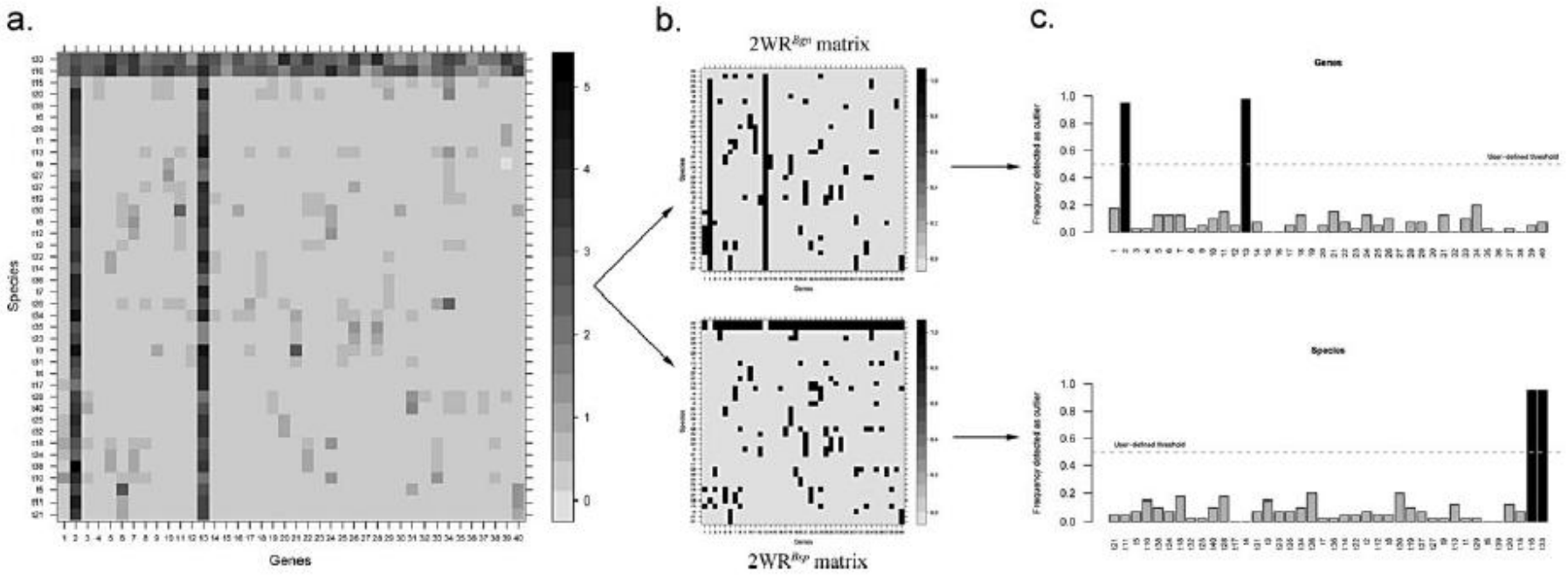
Real example (21 fungal species)



a.

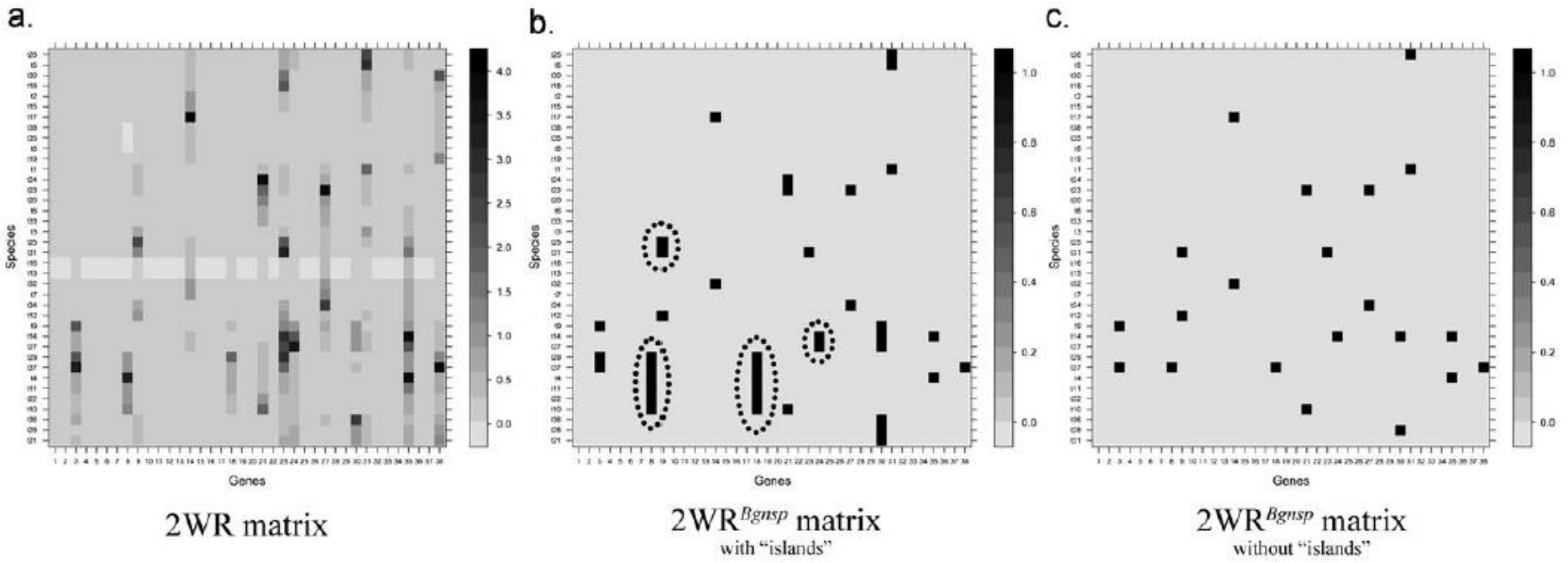

Species
Genes
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
5
4
3
2
1
0

From cohesion plot to 2-way reference matrix



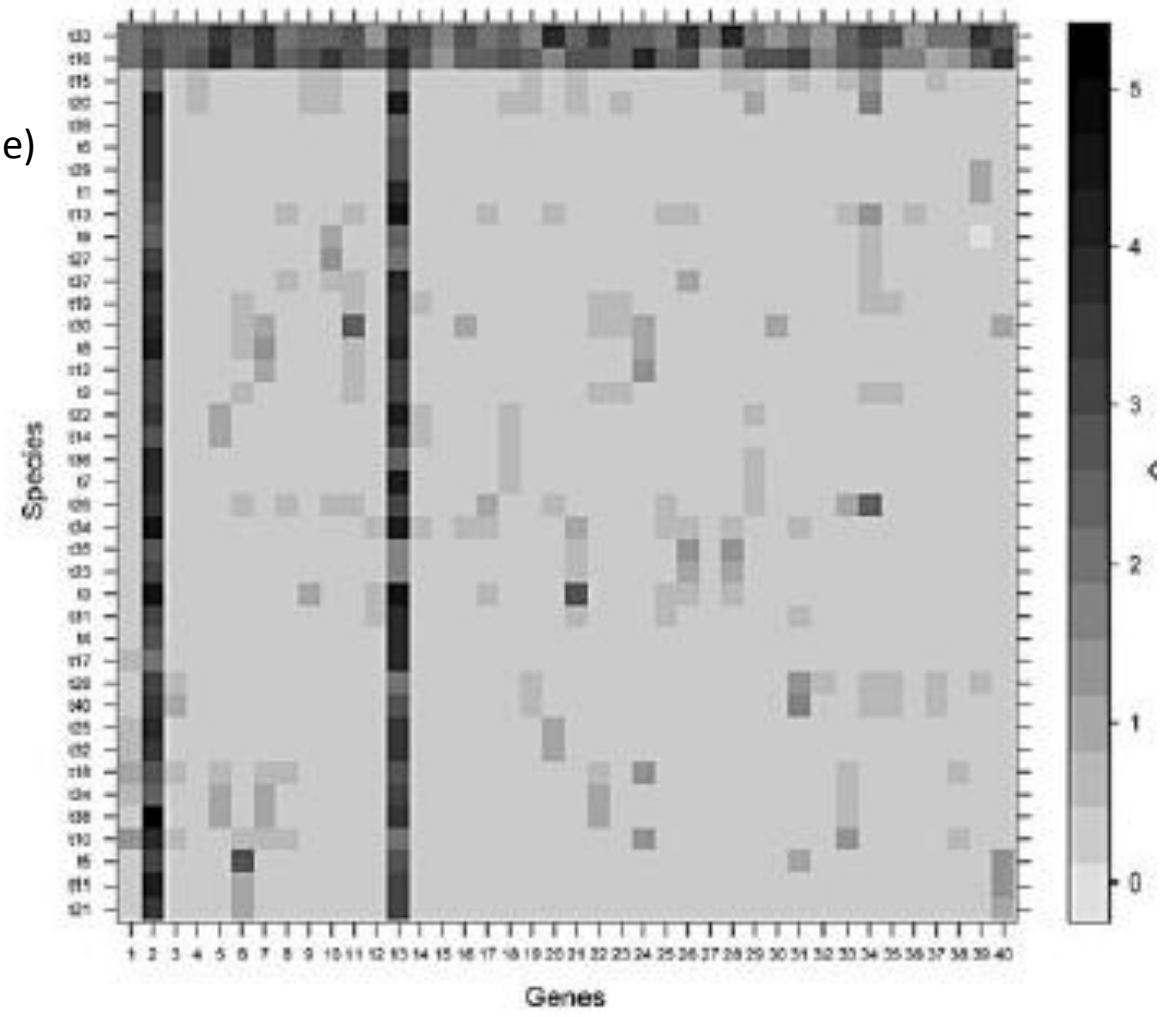
→ Identify outliers

de Vienne et al. MBE. 2013



→ Identify outliers

- Any problem (species identification…)
- “Volatile” species
(acceptors of genes by HGT, for example)



What it means

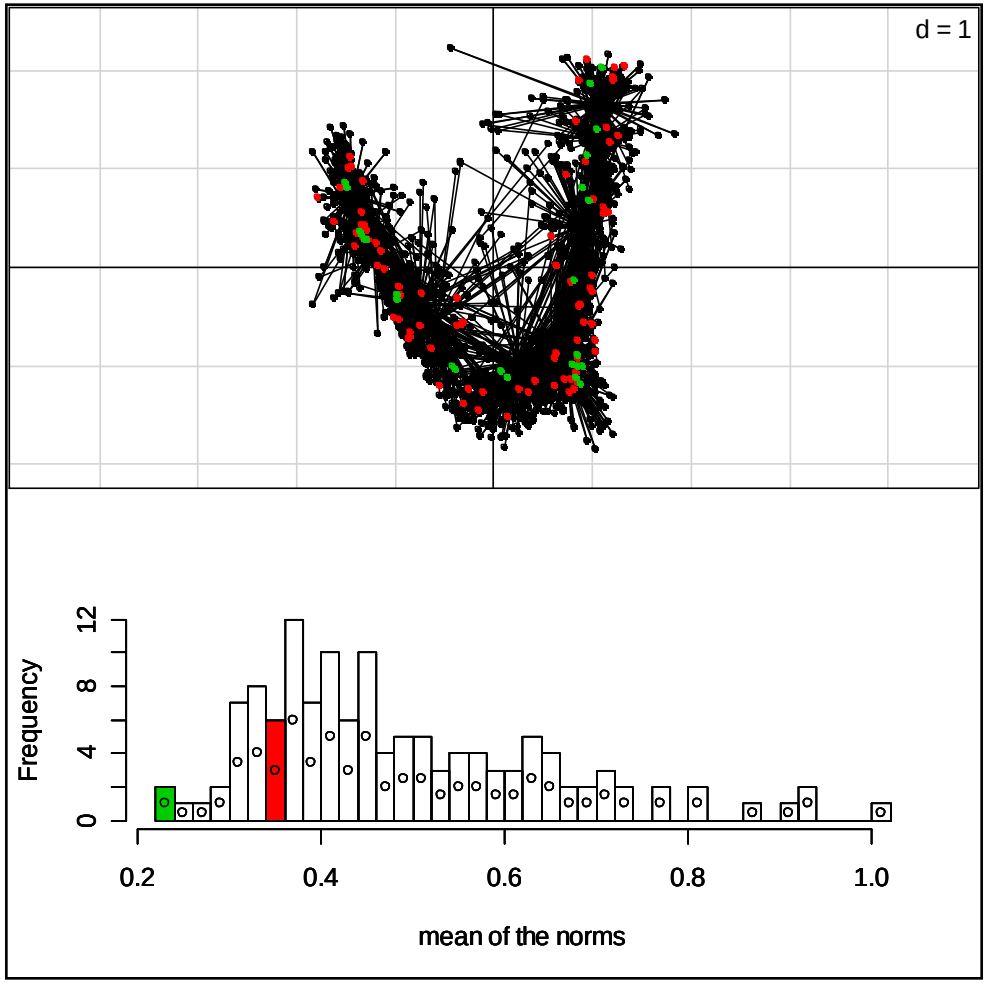
- Any problem (sequencing, paralogy, tree reconstruction…)
- “highliy transferable genes /shared genes

Thanks

- HP interaction
  - Tatiana Giraud
  - Guislaine Refrégier
  - Michael Hood
  - …
- Icong
  - Olivier Martin
- PPI
  - Jérôme Azé
- Phylo-MCOA
  - Gabriela Aguileta
  - Sebastien Ollier

Our method, Phylo-MCOA

- Is fast
- Is user friendly
- Is concordant with other methods
- Is colorful
- Is free

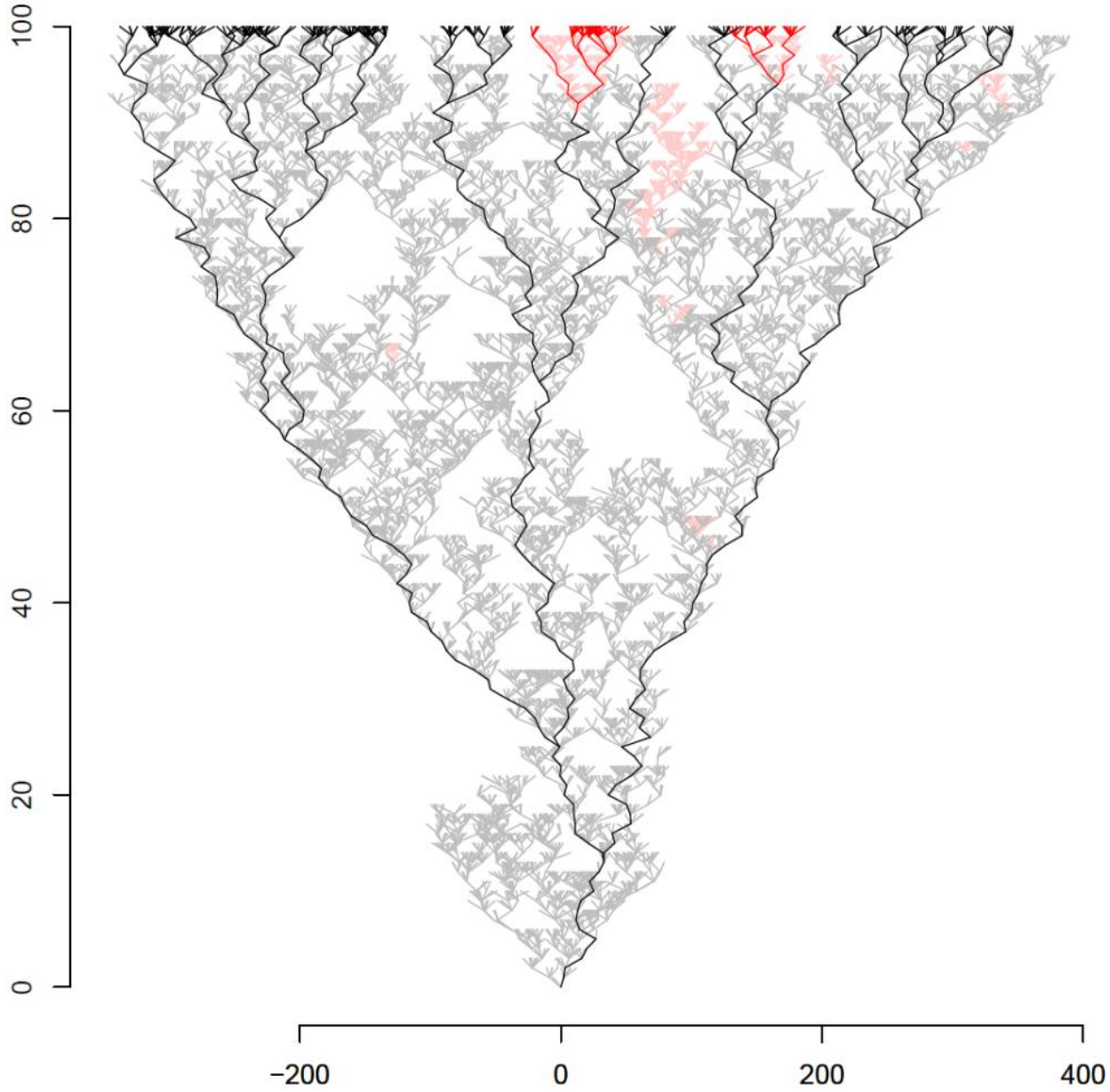


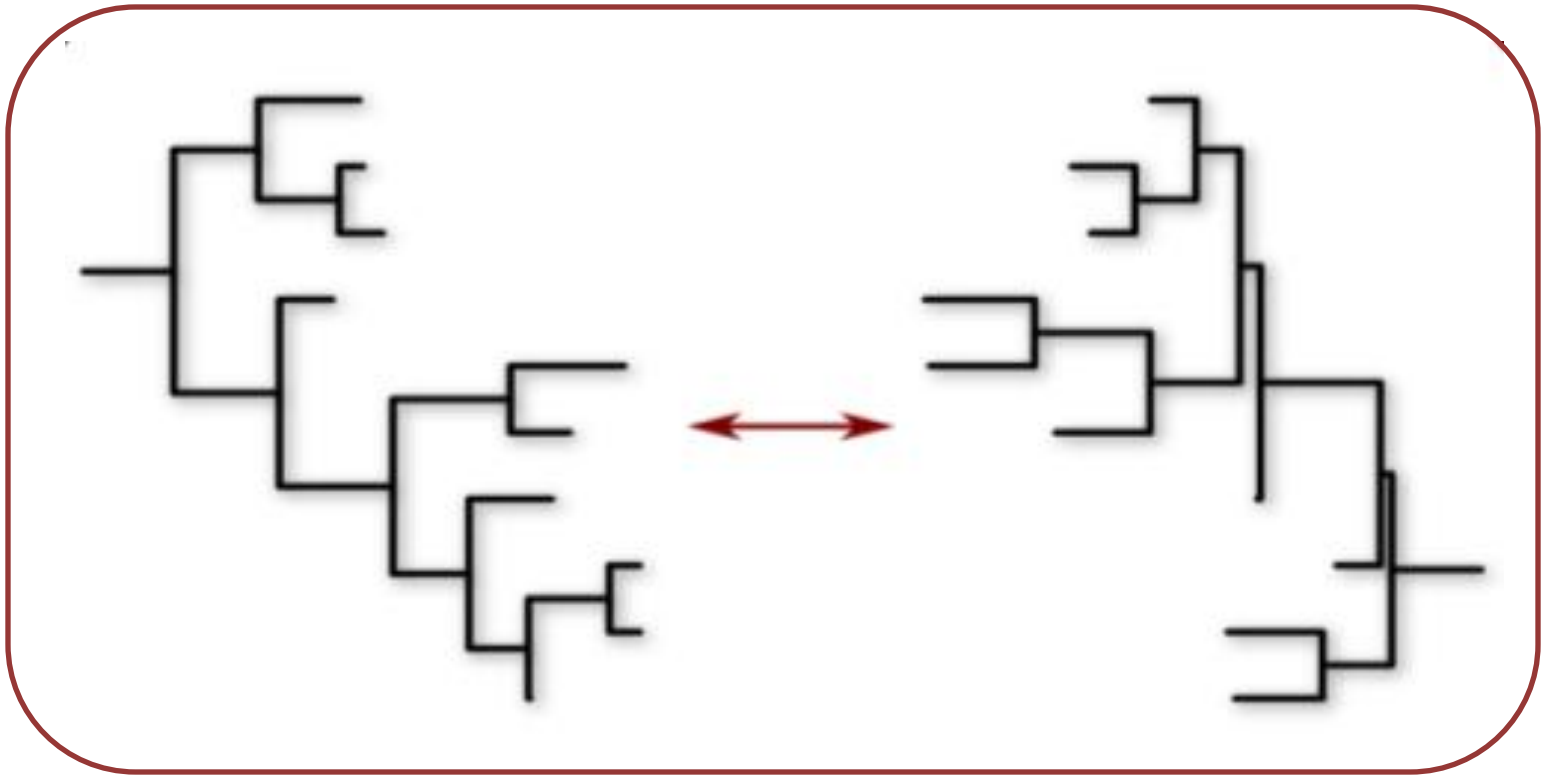
Examples with simulated data

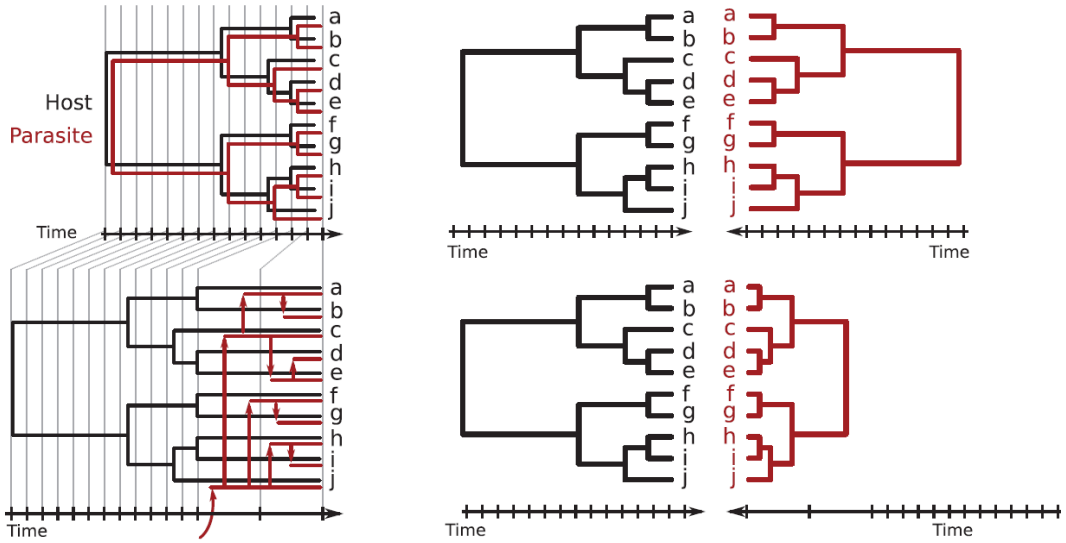
- All genes tell different evolutionary histories
- All genes tell the same evolutionary history
- Recent HGT
- Ancient HGT

- Tree comparisons and coevolution
  - Between host and parasites (JEB)
  - Between genes/proteins (PloS ONE)

- Tree comparisons and species tree/gene tree "reconciliation"
  - Multiple gene trees in comparative genomics


100
80
60
40
20
0
−200
0
200
400




Host
Parasite
Time
a
b
c
d
e
f
g
h
i
j
Time
Time
Time
Time

I_{cong}

ongruence Index - Google Chrome

max2.ese.u-psud.fr/ × | Icong Congruence Ir ×

max2.ese.u-psud.fr/icong/index.htm

Applications | crg | Open Source Wel | GeoServer | Plateforme Alter | Visualiseur des ré | Blender 3D Mode | Sweet Home 3D

ON-LINE CALCULATION OF THE CONGRUENCY INDEX I_{cong}

TO ASSESS THE EXISTENCE OF TOPOLOGICAL CONGRUENCE
BETWEEN TWO TREES

NEW version, goes faster !!

If you need help for the formatting of your trees or if you want to see an example, click here.

If you use this index for a publication, please cite
de Vienne D.M., Giraud, T. and Martin, O.C. 2007. A Congruence Index for Testing Topological Similarity between Trees. *Bioinformatics* **23** (23): 3119-3124.

Type or paste the first tree here:

Type or paste the second tree here:

Calculate Icong and its P-value | Reset

Last updated 3rd of March 2011 - © 2011 Damien de Vienne