2nd International Summer School on Networks & Evolution
25/06/2018 - Roscoff

The different types of networks used in evolutionary study: how they are made and what they are useful for

Philippe Gambette

UNIVERSITÉ
PARIS-EST
MARNE-LA-VALLÉE

Gaspard - Monge
Laboratoire d'informatique

Outline

- Phylogenetic networks
- How to build phylogenetic networks?
- Classes of phylogenetic networks
- Unexpected properties of networks
- The research community on phylogenetic networks

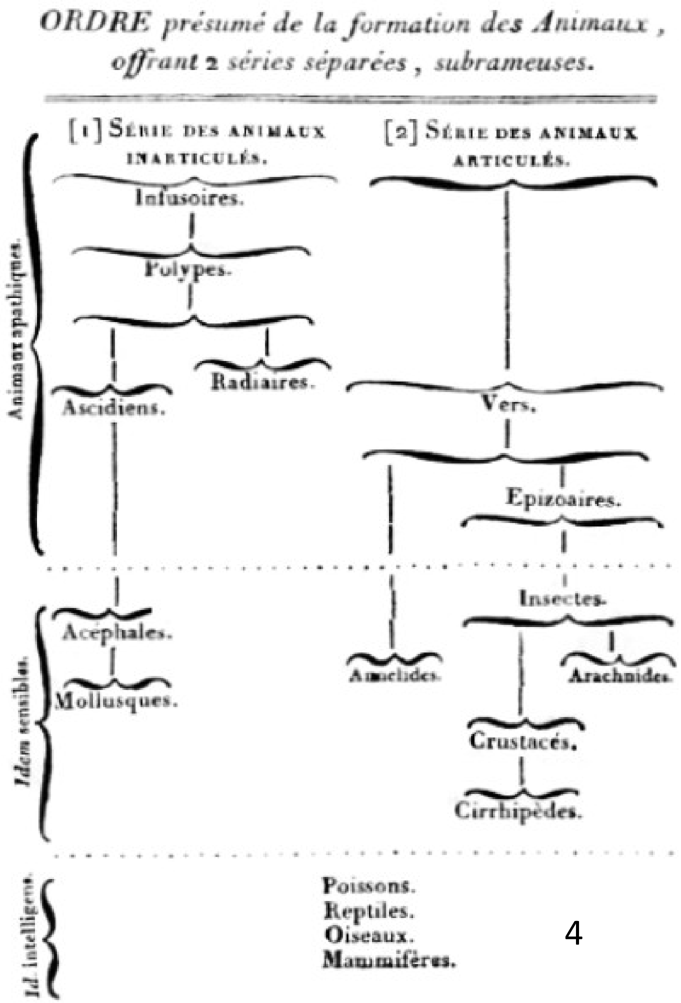
Outline

- Phylogenetic networks
- How to build phylogenetic networks?
- Classes of phylogenetic networks
- Unexpected properties of networks
- The research community on phylogenetic networks

Phylogenetic trees

Phylogenetic tree of a set of species:

- **Classify them depending on common characters** classification
- Describe their evolution



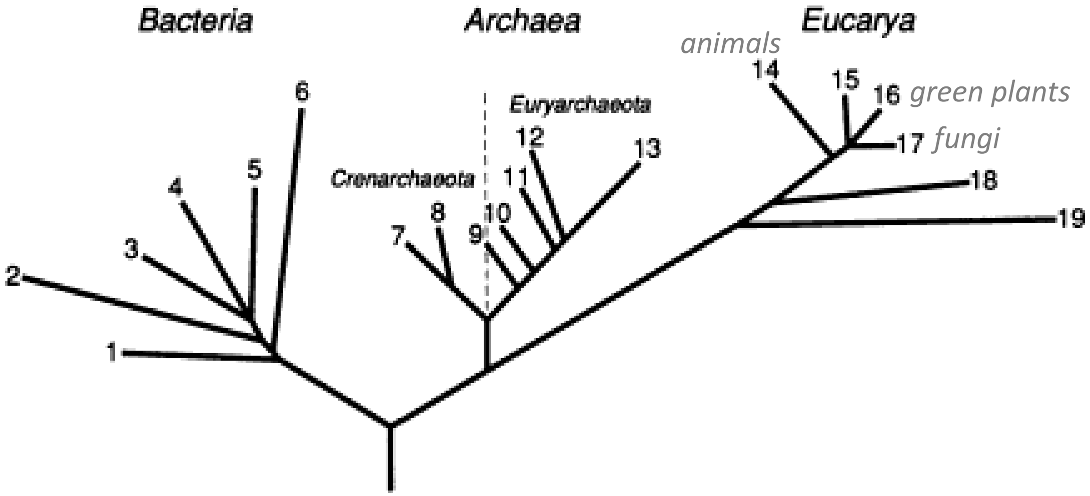
Lamarck : Histoire naturelle des animaux sans vertèbres (1815)

Phylogenetic trees

Phylogenetic tree of a set of species:

- Classify them depending on common characters
- **Describe their evolution**

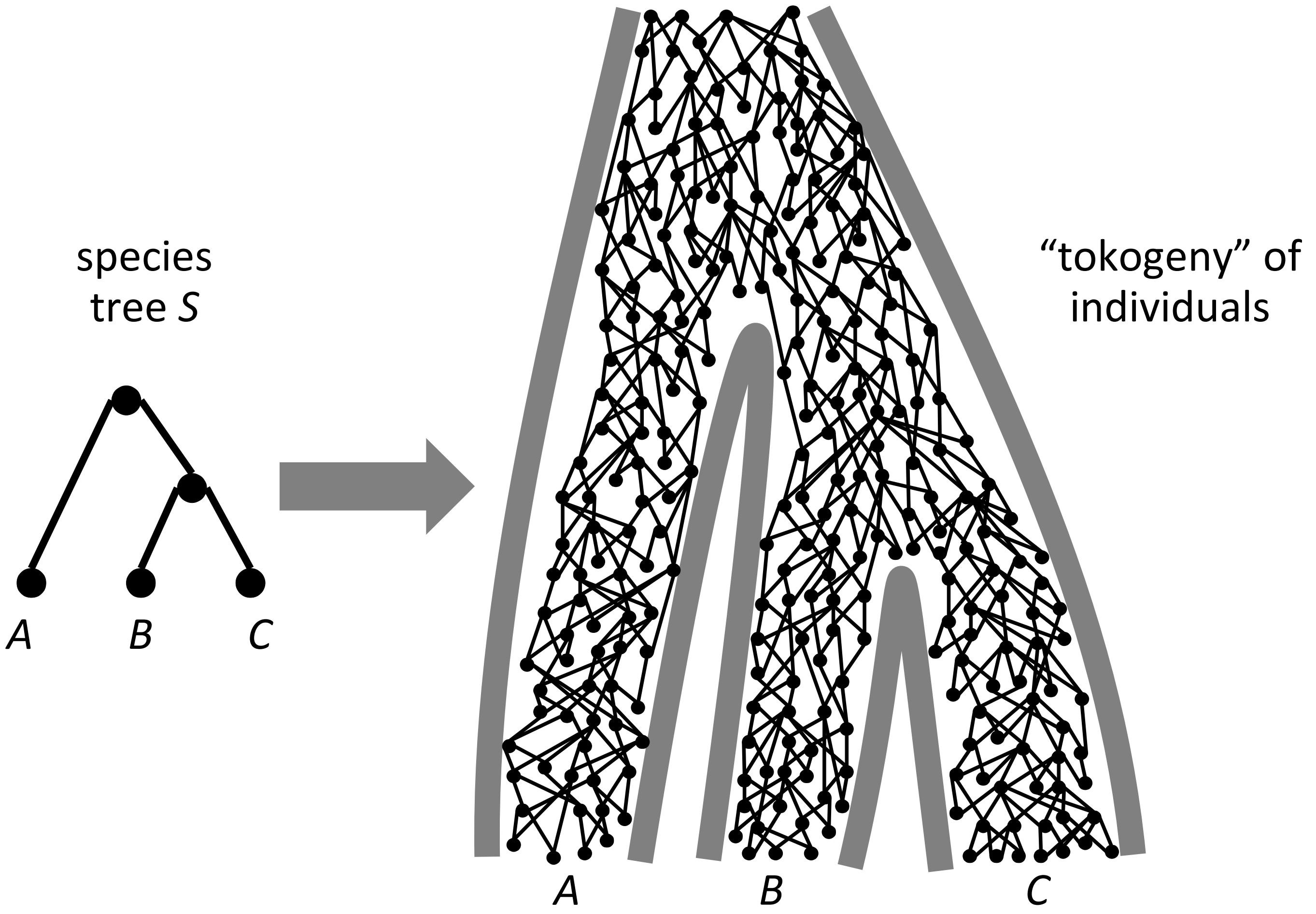
modelization



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)

Phylogenetic trees... and networks

Phylogenetic tree of a set of species



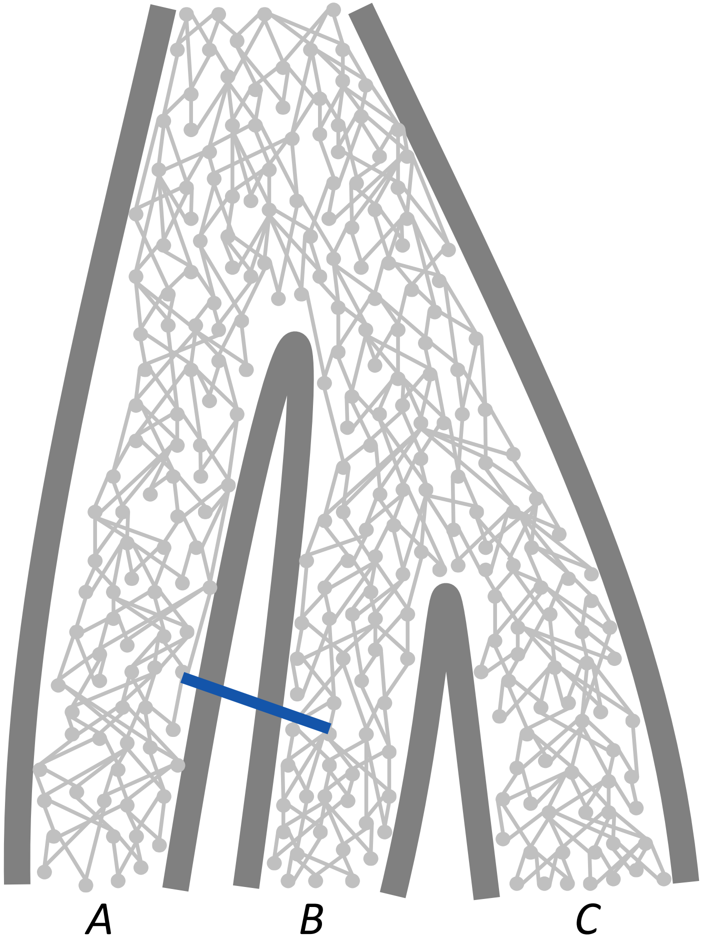
species
tree *S*

“tokogeny” of
individuals

Genetic material transfers

Transfers of genetic material between coexisting species:

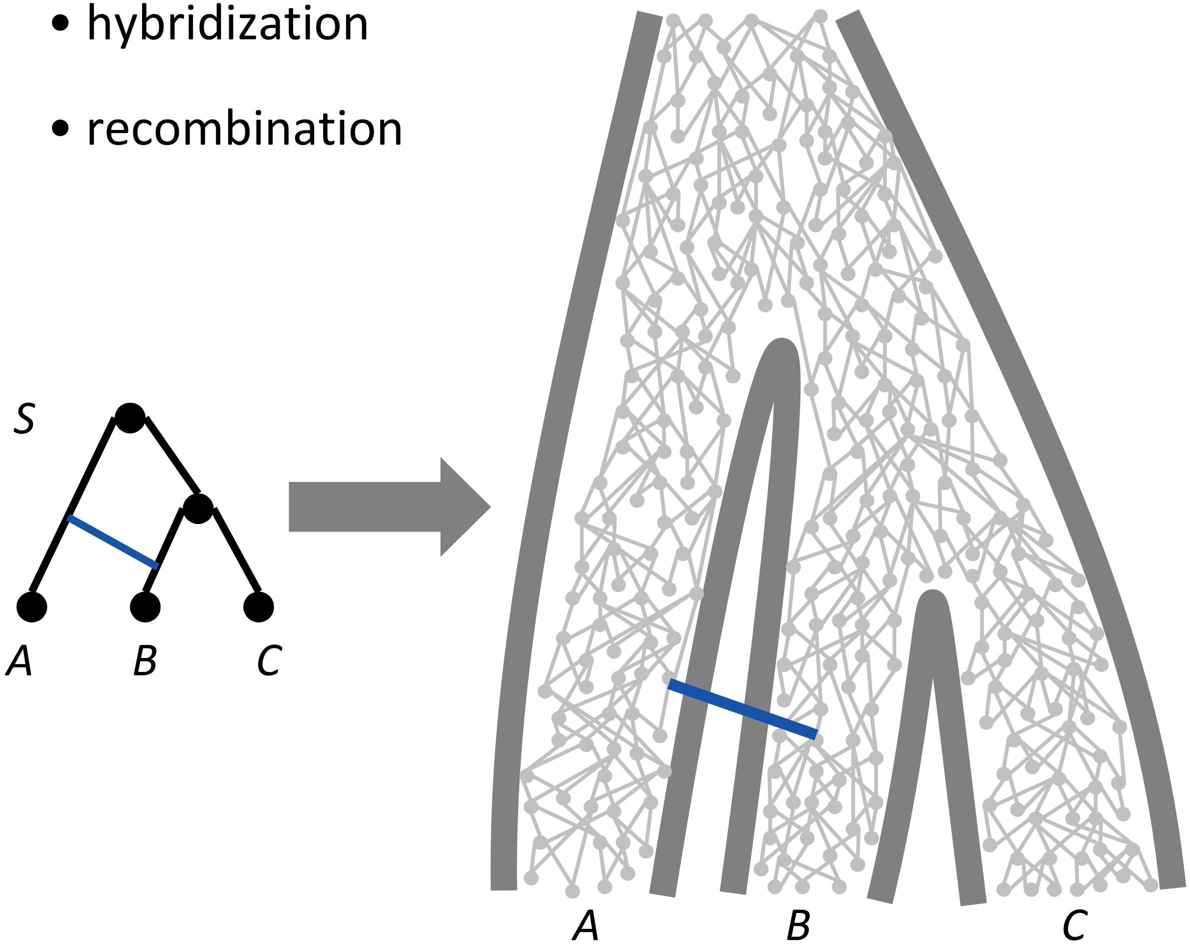
- lateral gene transfer
- hybridization
- recombination



Genetic material transfers

Transfers of genetic material between coexisting species:

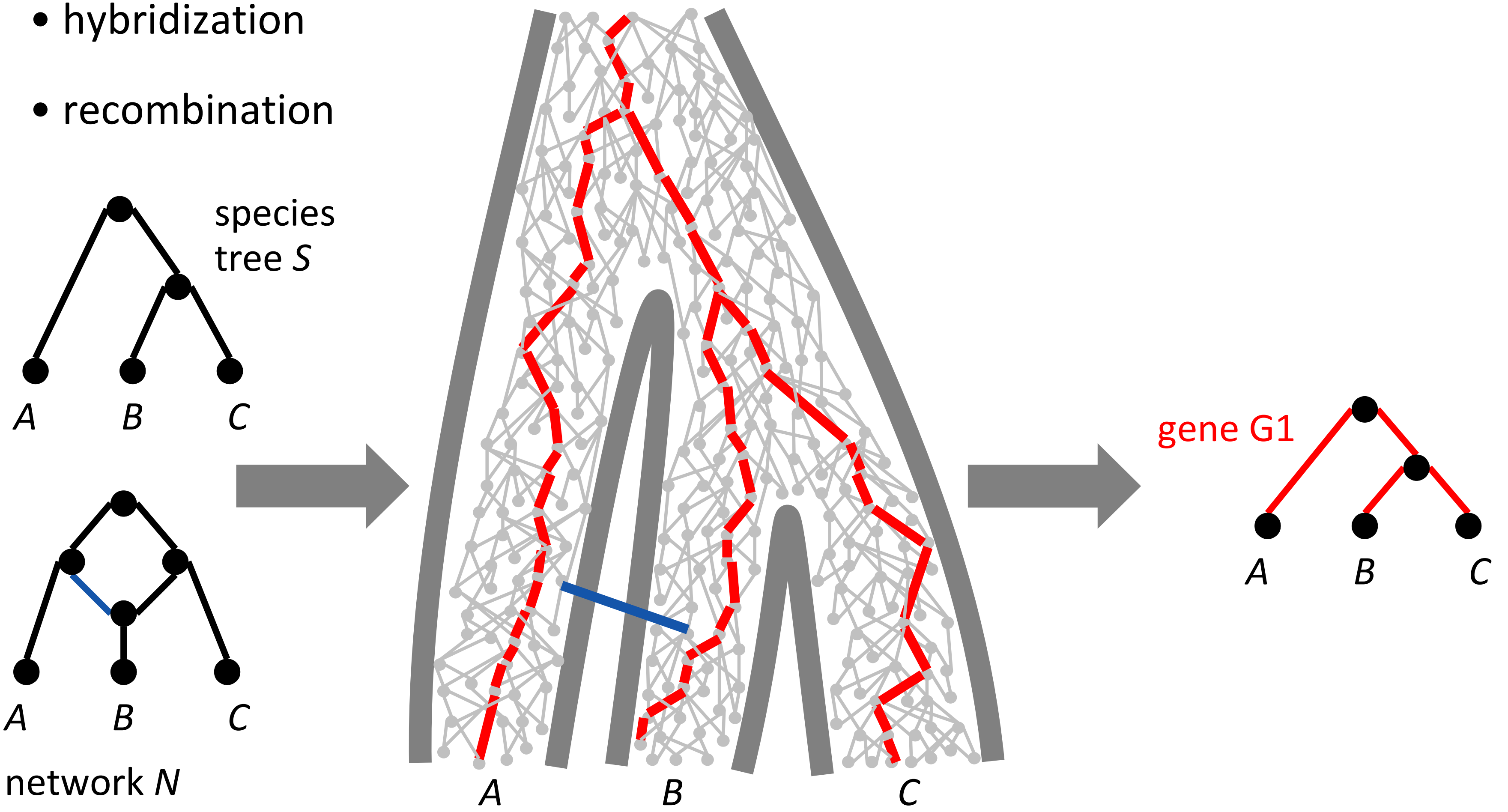
- lateral gene transfer
- hybridization
- recombination



Genetic material transfers

Transfers of genetic material between coexisting species:

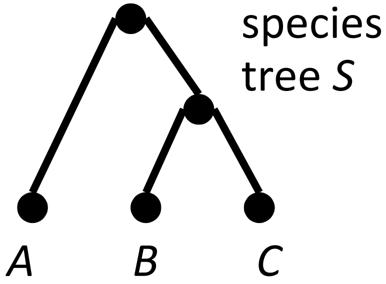
- lateral gene transfer
- hybridization
- recombination



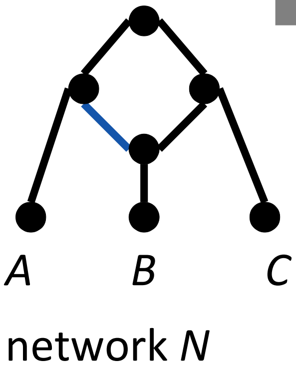
Genetic material transfers

Transfers of genetic material between coexisting species:

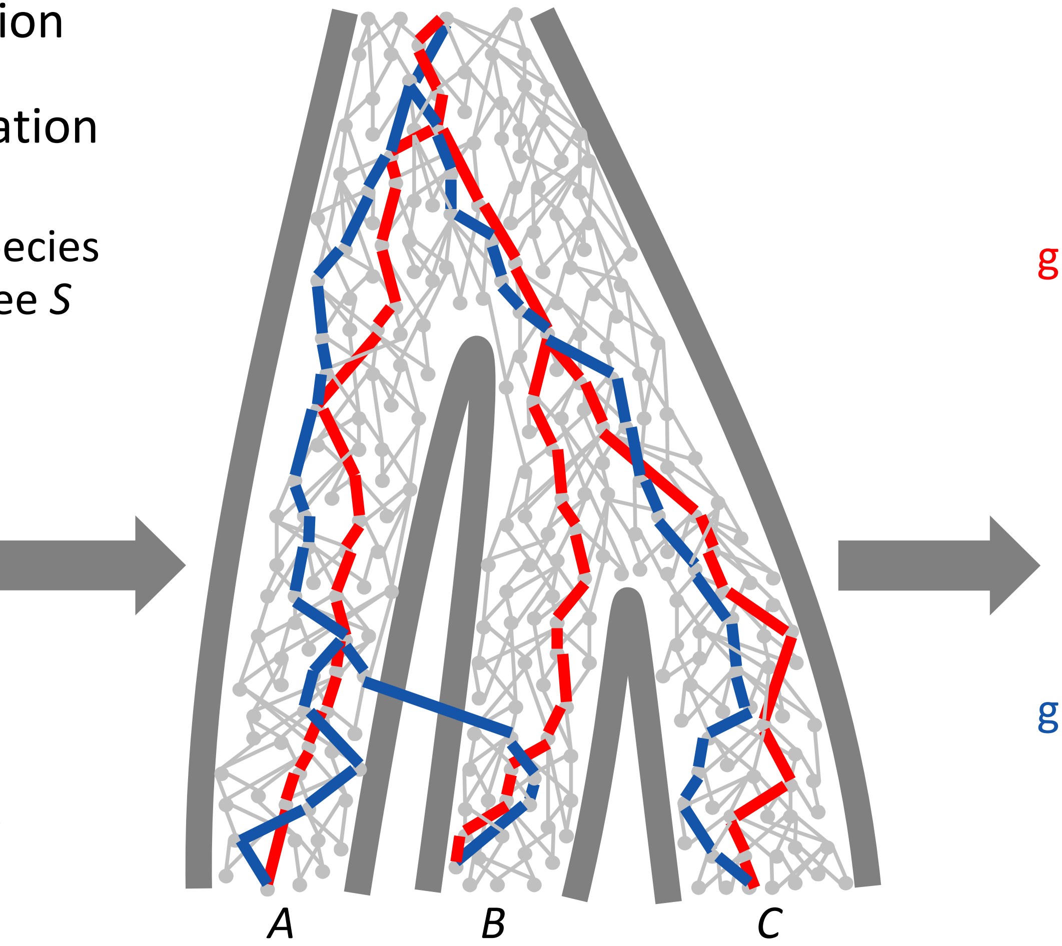
- lateral gene transfer
- hybridization
- recombination

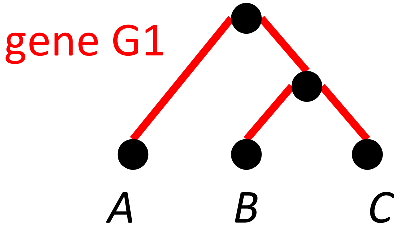


species tree *S*

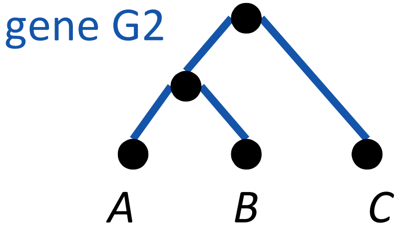


network *N*





incompatible
gene trees



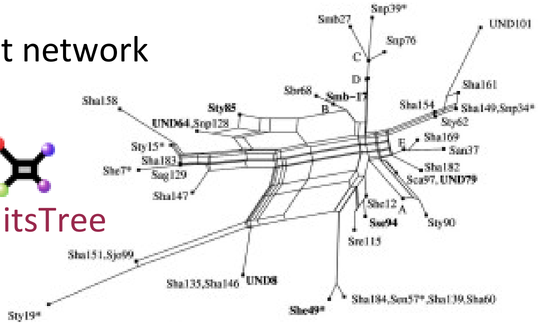
Phylogenetic networks

Phylogenetic network: network representing evolution data

- **abstract / implicit / data-display** phylogenetic networks:
to **classify**, **visualize** **data**

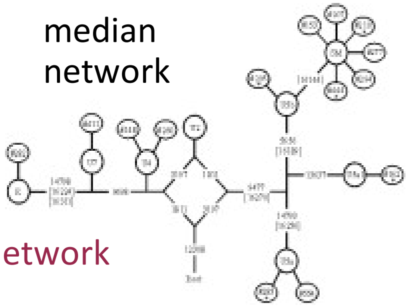
split network

SplitsTree



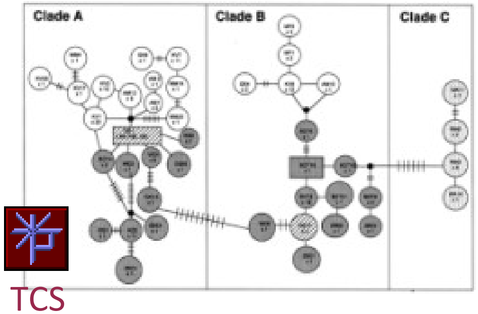
median network

Network

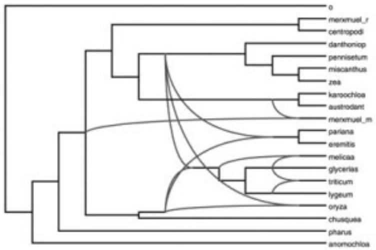


minimum spanning network

TCS



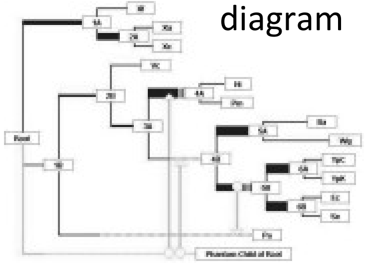
- **explicit** phylogenetic networks:
to **model** **evolution**



galled network

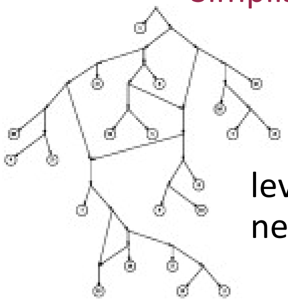
Dendroscope

synthesis diagram



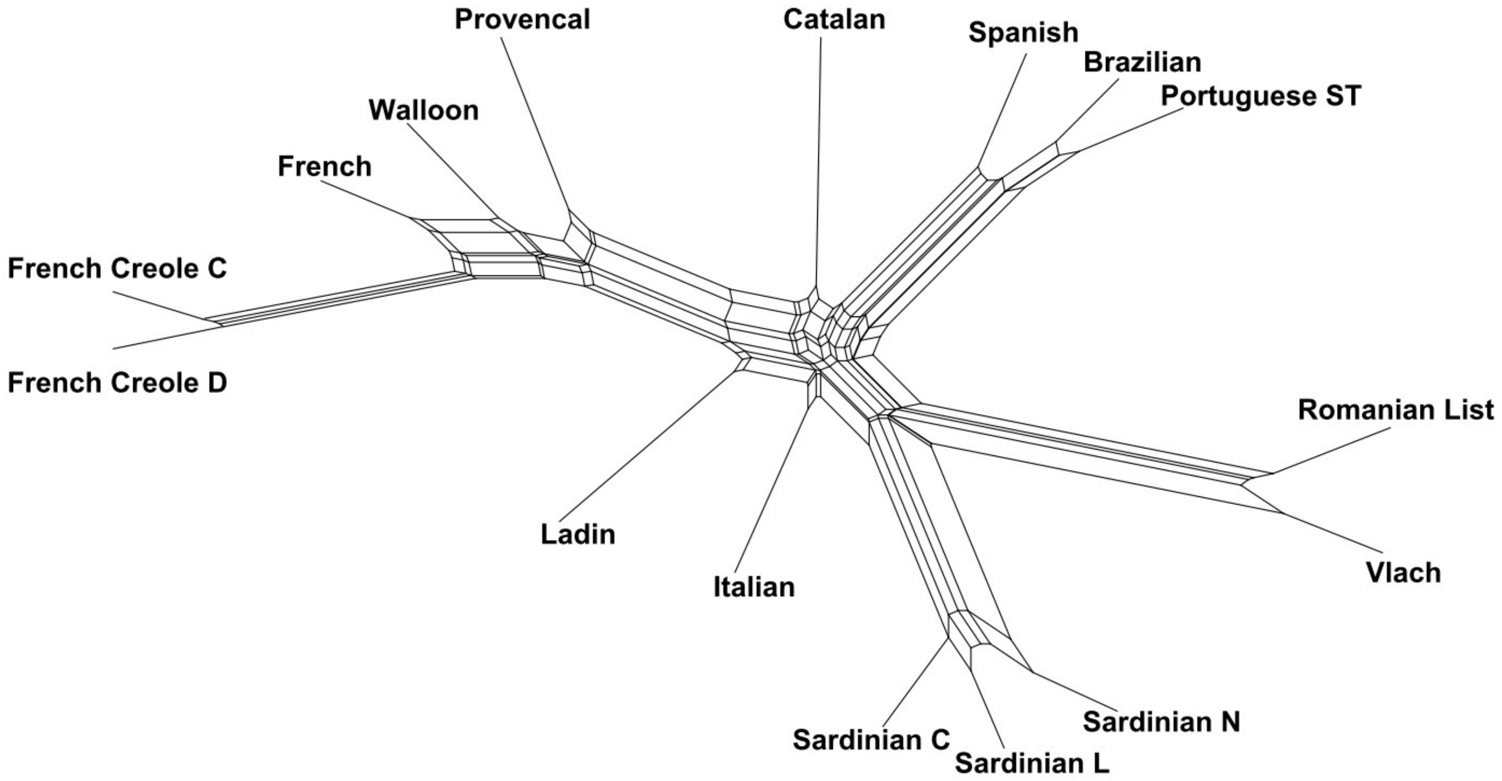
HorizStory

Simplistic



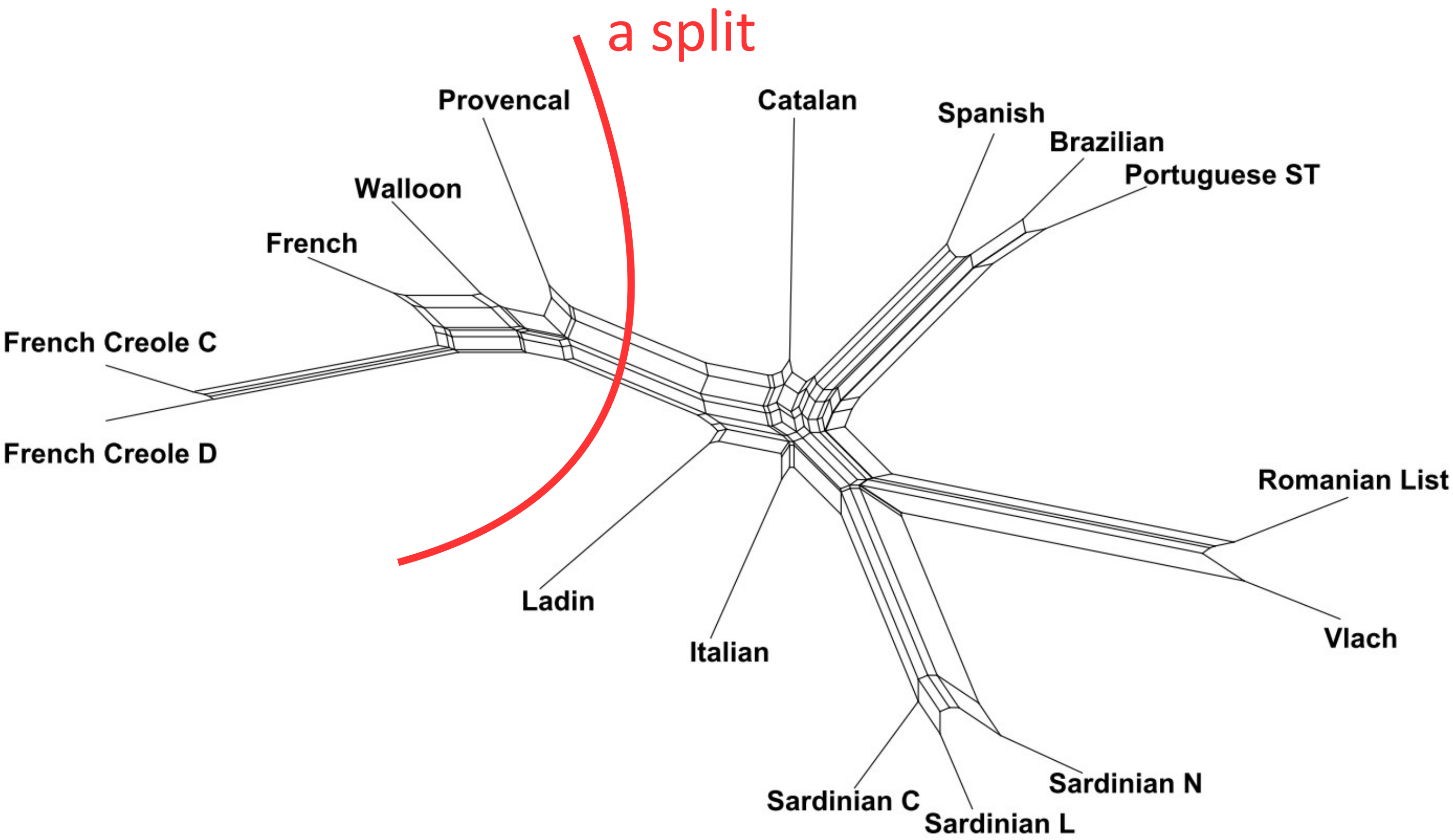
level-2 network

Abstract phylogenetic networks: split networks



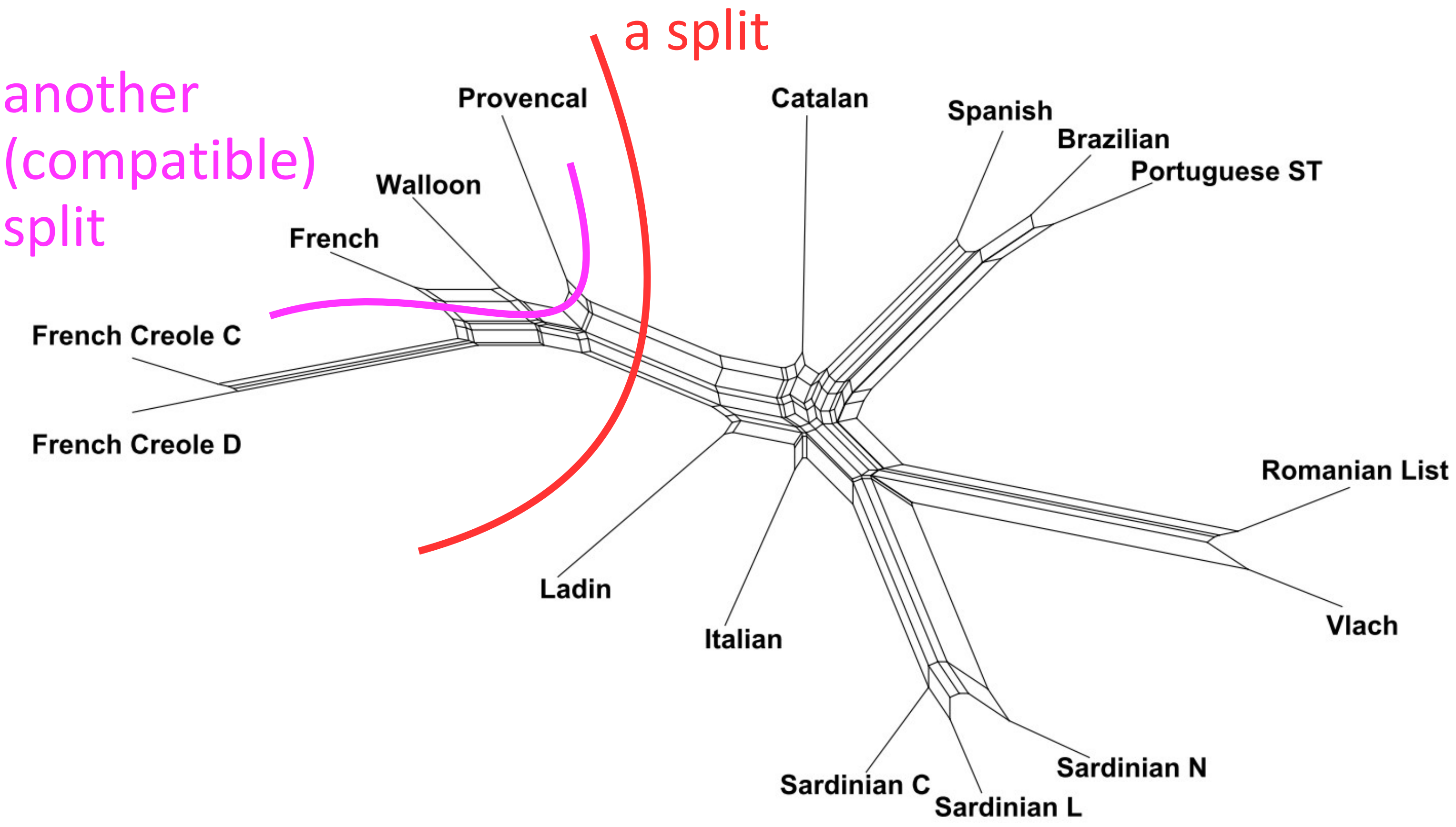
Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M. & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. BMC Evolutionary Biology, 16(1), 180.

Abstract phylogenetic networks: split networks



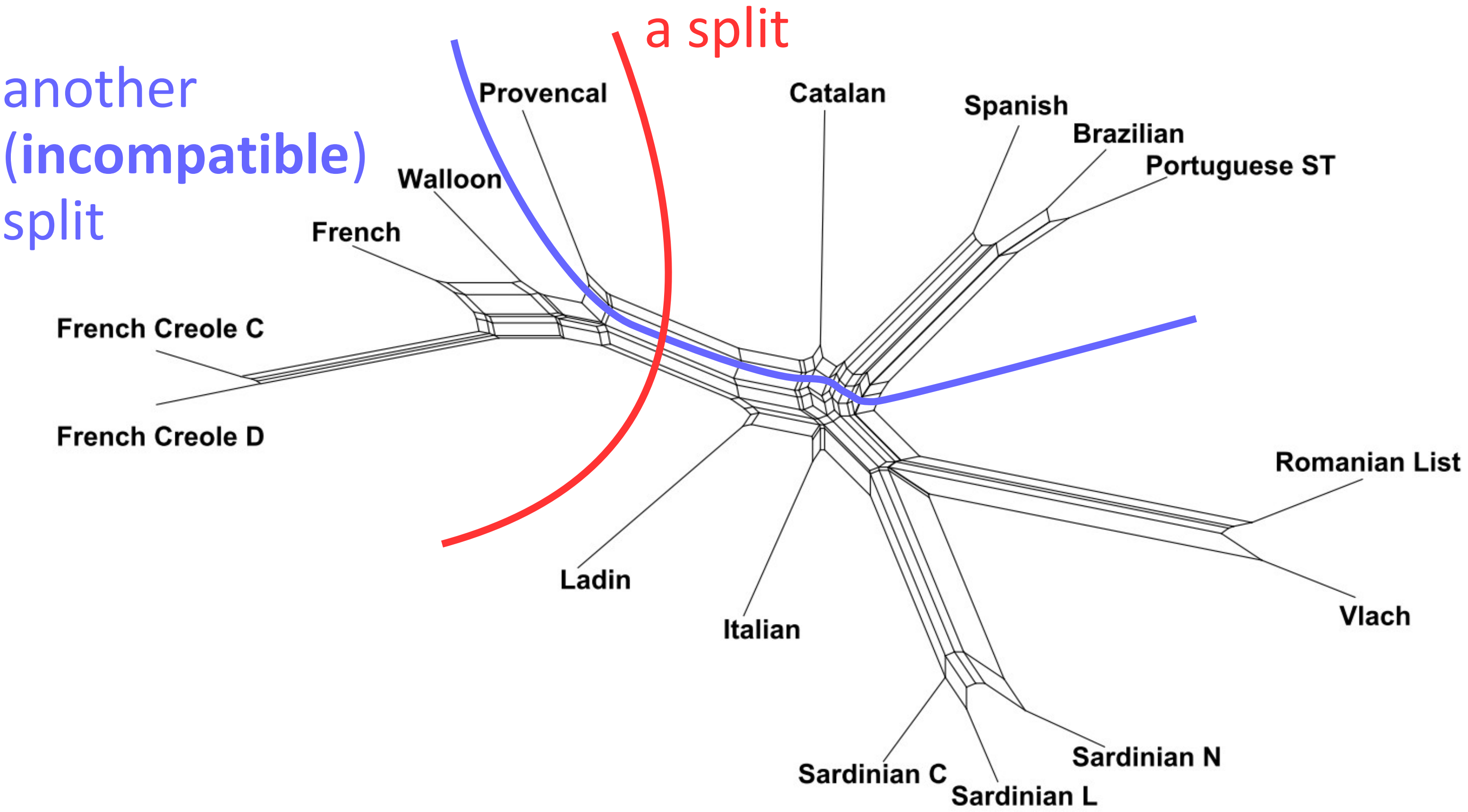
Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M. & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. BMC Evolutionary Biology, 16(1), 180.

Abstract phylogenetic networks: split networks



Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M. & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. BMC Evolutionary Biology, 16(1), 180.

Abstract phylogenetic networks: split networks

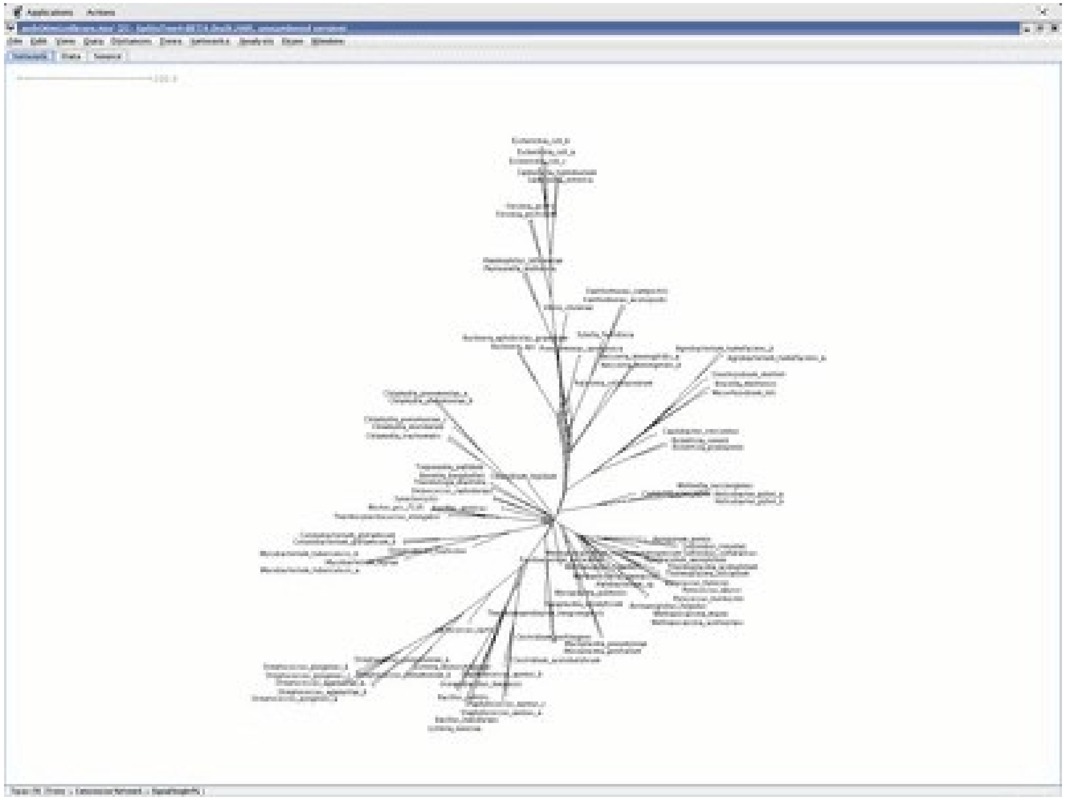


Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Di Sciullo, A. M. & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. BMC Evolutionary Biology, 16(1), 180.

Abstract phylogenetic networks: split networks

The “boxes” in split networks: incompatible splits

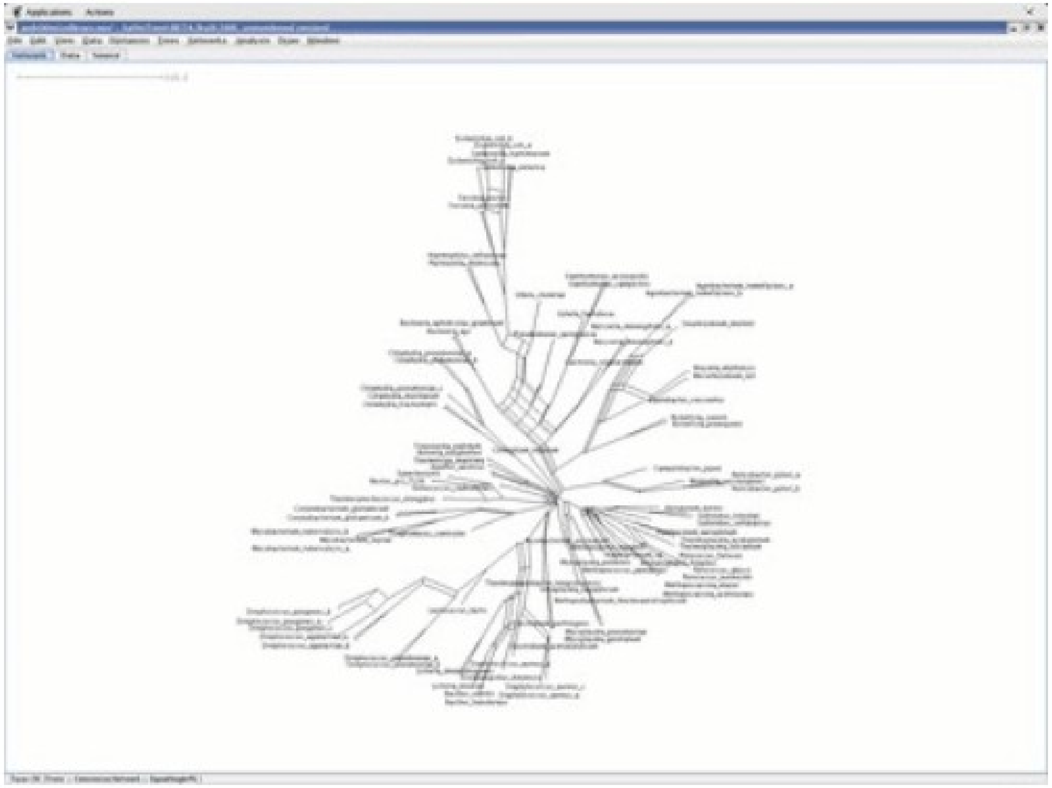
→ open them as much as possible in SplitsTree!



Abstract phylogenetic networks: split networks

The “boxes” in split networks: incompatible splits

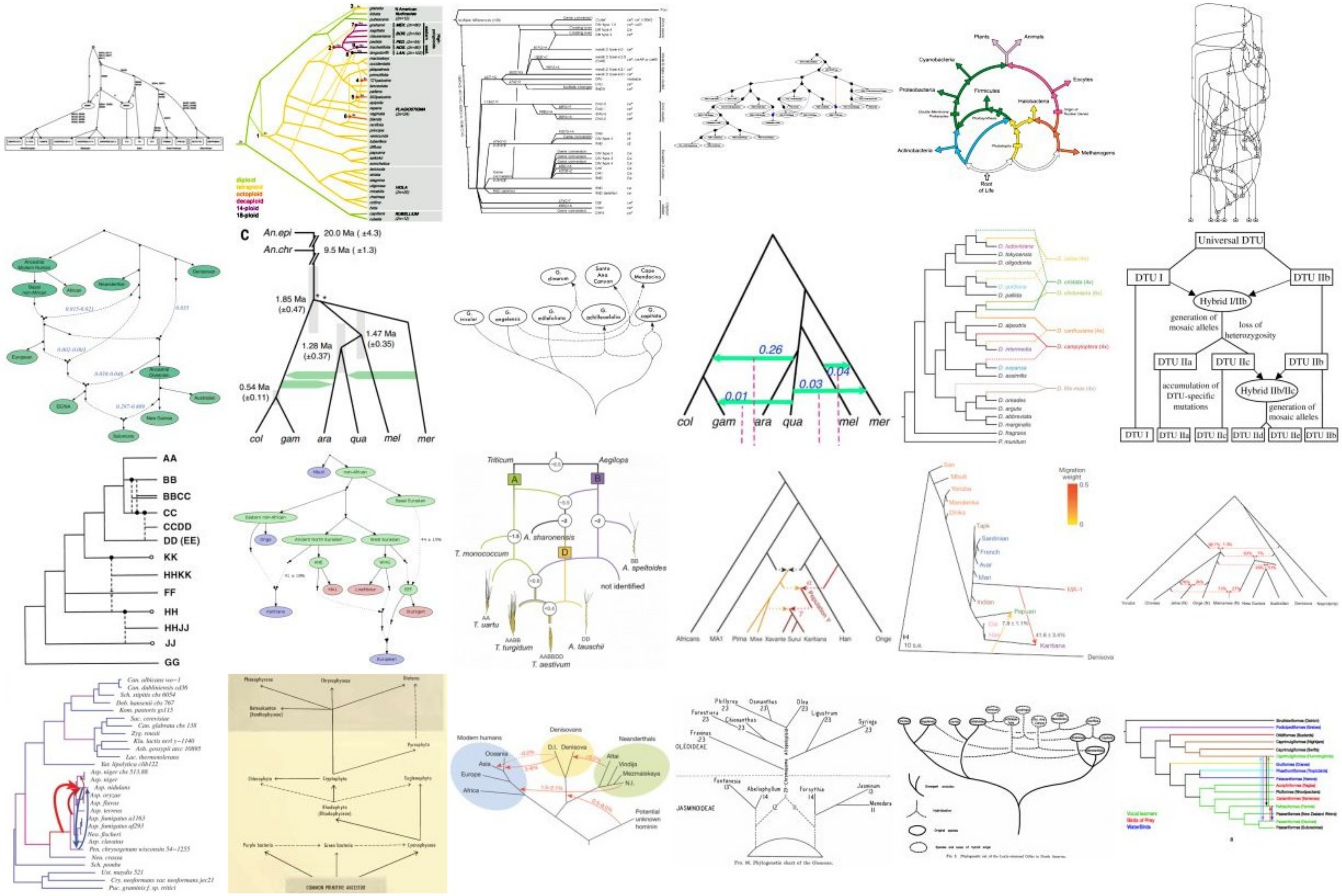
→ open them as much as possible in SplitsTree!



Philippe Gambette & Daniel H. Huson, Improved Layout of Phylogenetic Networks, IEEE/ACM TCBB 5(3), p. 472-479

Explicit phylogenetic networks

A gallery of explicit phylogenetic networks :

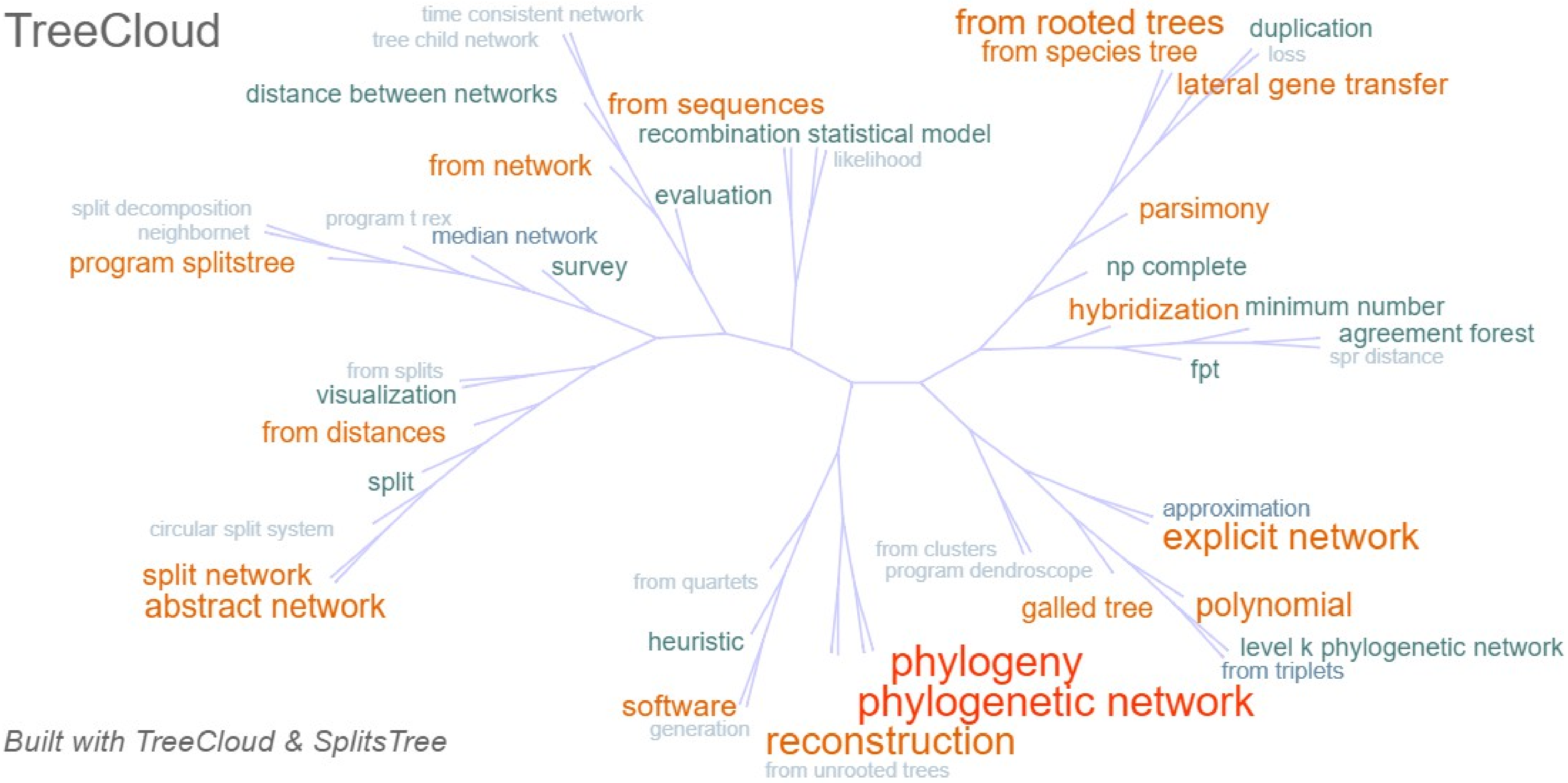


http://phylnet.univ-mlv.fr/recophync/networkDraw.php

Outline

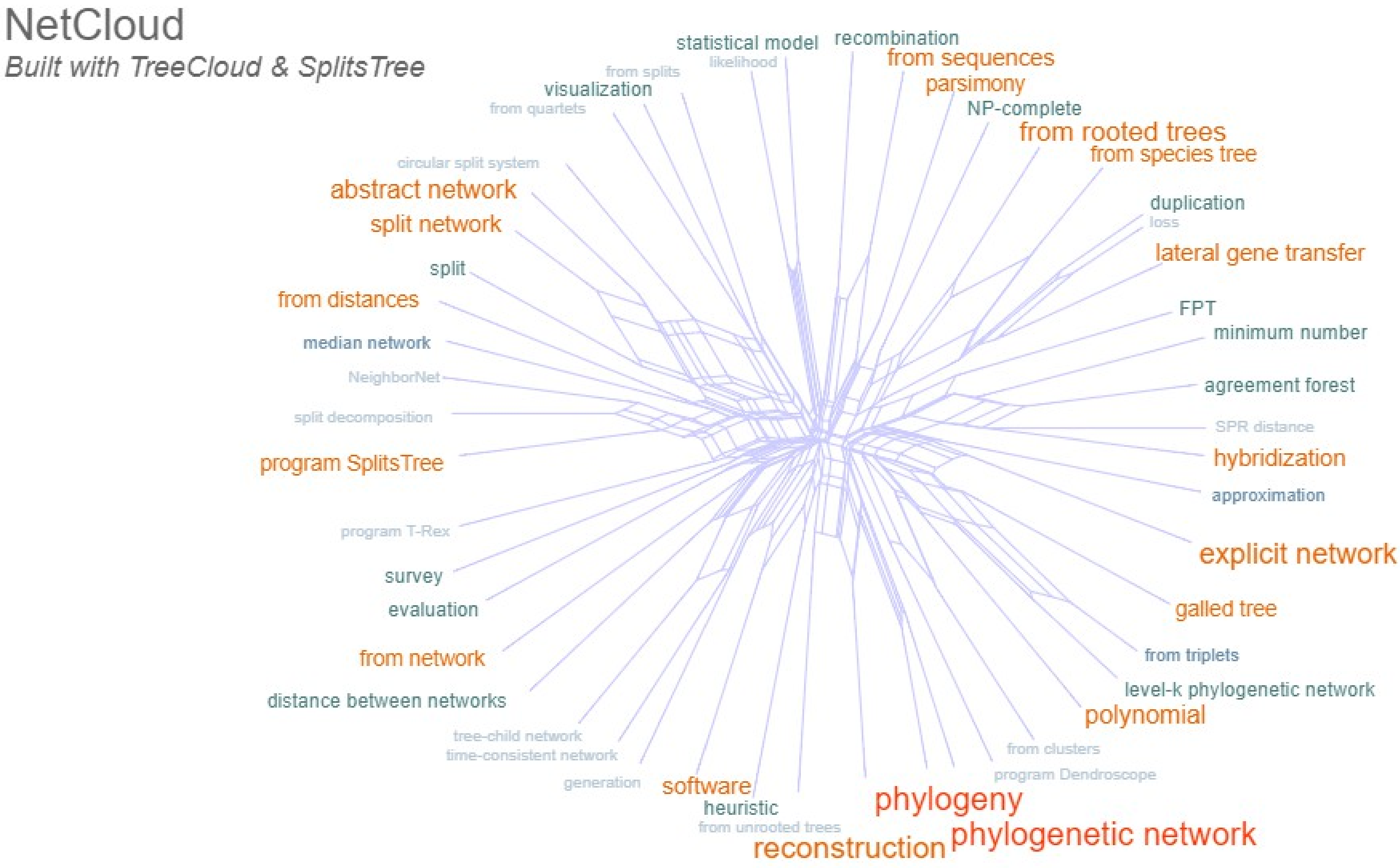
- Phylogenetic networks
- How to build phylogenetic networks?
- Classes of phylogenetic networks
- Unexpected properties of networks
- The research community on phylogenetic networks

An example: keywords about phylogenetic networks



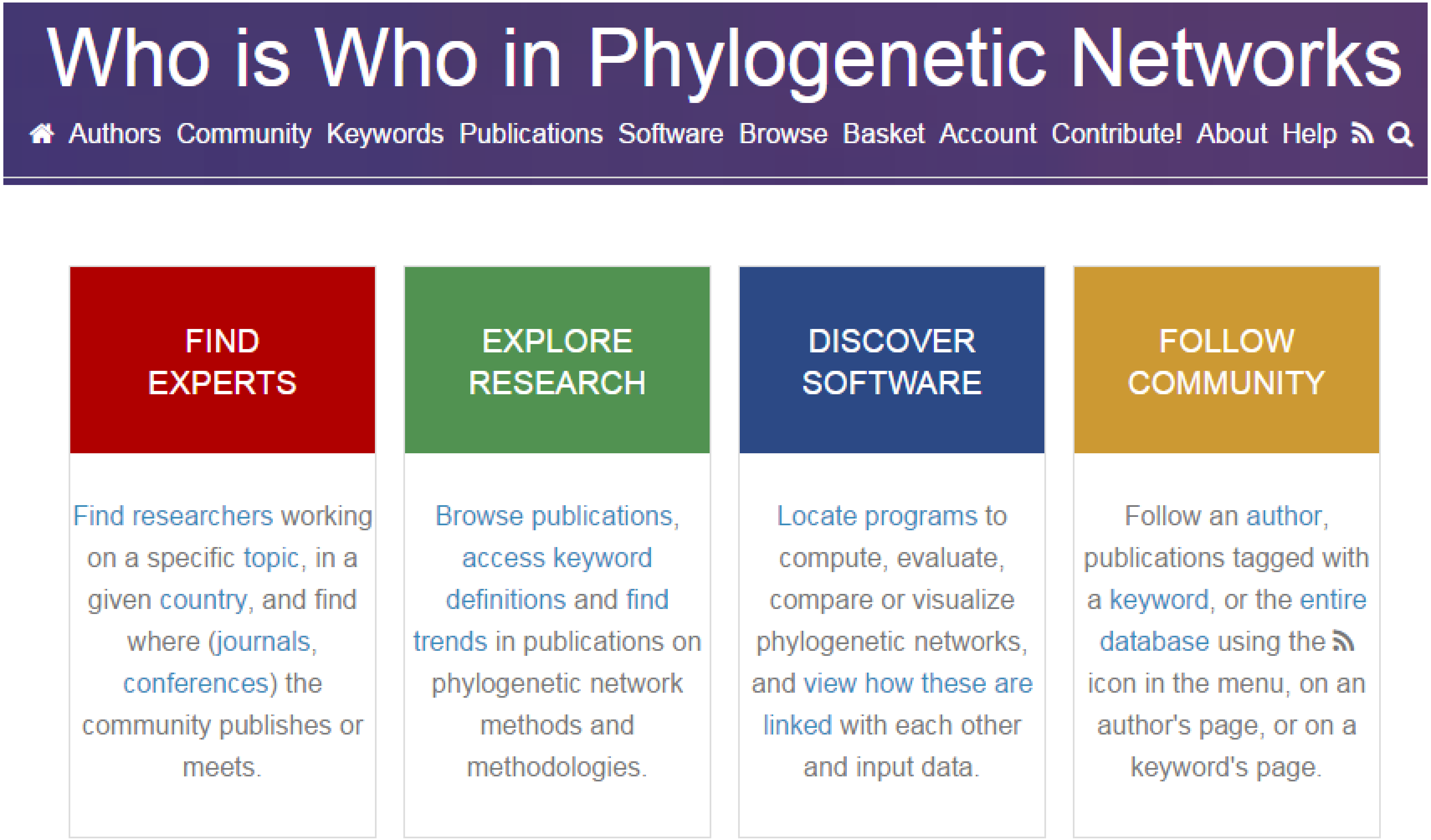
http://phylnet.univ-mlv.fr/keywords.php

An example: keywords about phylogenetic networks



http://phylnet.univ-mlv.fr/keywords.php

Who is who in Phylogenetic Networks?



Who is Who in Phylogenetic Networks

FIND EXPERTS

Find researchers working on a specific topic, in a given country, and find where (journals, conferences) the community publishes or meets.

EXPLORE RESEARCH

Browse publications, access keyword definitions and find trends in publications on phylogenetic network methods and methodologies.

DISCOVER SOFTWARE

Locate programs to compute, evaluate, compare or visualize phylogenetic networks, and view how these are linked with each other and input data.

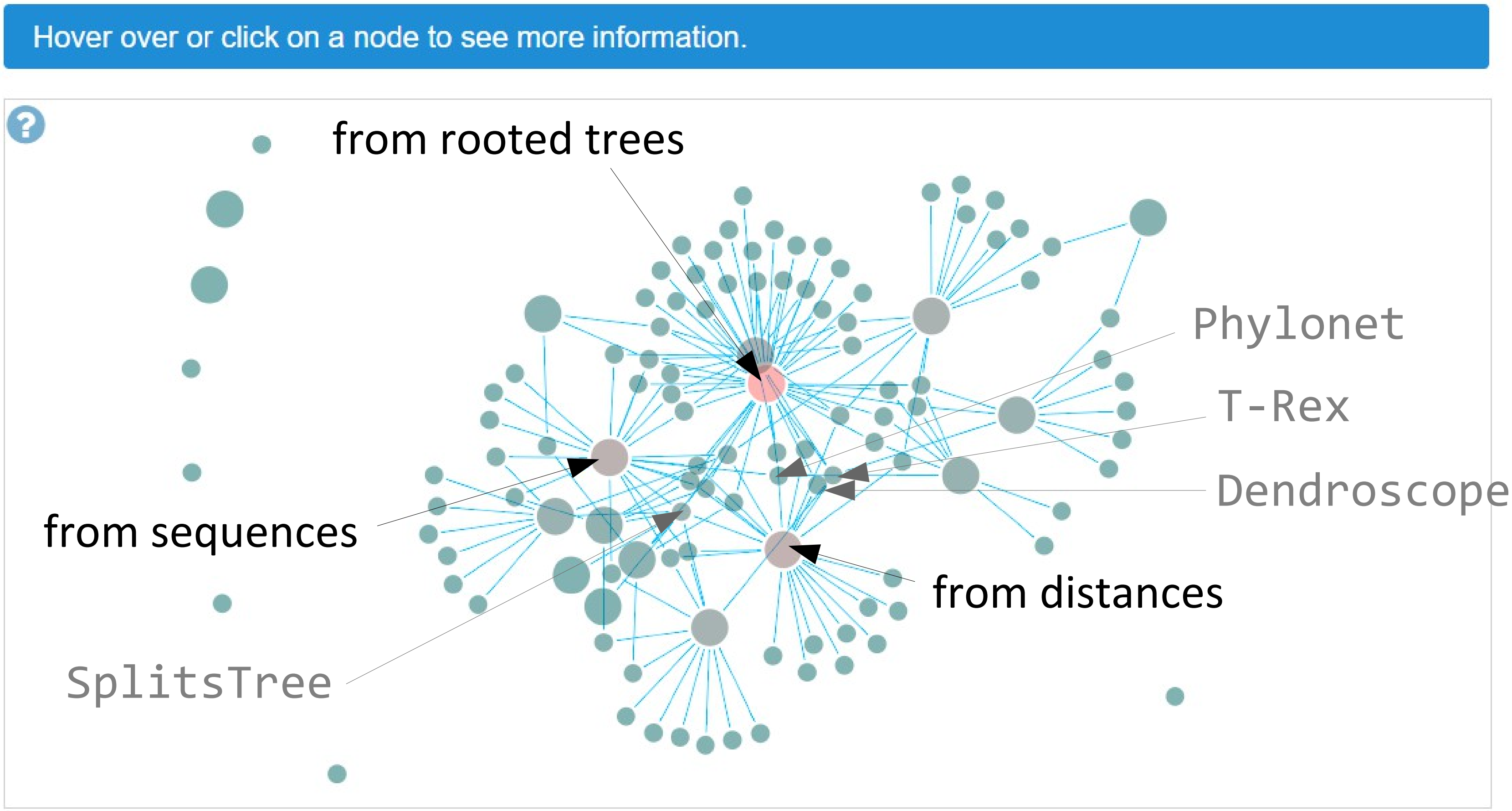
FOLLOW COMMUNITY

Follow an author, publications tagged with a keyword, or the entire database using the ⊃ icon in the menu, on an author's page, or on a keyword's page.

Based on BibAdmin by Sergiu Chelcea
+ tag clouds, date histograms, journal lists, keyword definitions, co-author graphs
(work of Tushar Agarwal)

http://phylnet.univ-mlv.fr

Who is who in Phylogenetic Networks? Software!



Hover over or click on a node to see more information.

from rooted trees

Phylonet

T-Rex

Dendroscope

from sequences

from distances

SplitsTree

http://phylnet.univ-mlv.fr/show.php?keyword=programs

Who is who in Phylogenetic Networks? Keywords!

abstract-network(88) agreement-forest(29) approximation(18) APX-hard(4) ARG(6) bayesian(4) block-realization(1) bootstrap(4) bound(4) branch-and-bound(1) cactus-graph(1) characterization(11) circular-split-system(13) clustering(3) coalescent(10) consensus(8) consistency(2) cophylogeny(1) counting(3) database(1) distance-between-networks(30) diversity(5) duplication(33) dynamic-programming(10) enumeration(4) evaluation(25) explicit-network(191) exponential-algorithm(6) FPT(31) from-clusters(16) from-continuous-characters(1) from-distances(45) from-gene-order(1) from-multilabeled-tree(9) from-network(39) from-NGS-data(1) from-quartets(16) from-rooted-trees(122) from-sequences(51) from-species-tree(46) from-splits(13) from-trees(7) from-trinets(2) from-triplets(24) from-unrooted-trees(16) galled-network(7) galled-tree(38) generation(12) haplotype-network(2) haplotyping(1) heuristic(32) HMM(2) hybridization(52) inapproximability(5) integer-linear-programming(3) isomorphism(3) k-reticulated(1) kernelization(2) labeling(4) lateral-gene-transfer(56) level-k-phylogenetic-network(30) likelihood(16) lineage-sorting(10) loss(16) MASN(4) median-network(16) MedianJoining(2) Minimal-lateral-network(1) minimum-contradiction(2) minimum-number(33) minimum-spanning-network(2) model-selection(2) mu-distance(2) NeighborNet(14) nested-network(2) netting(3) normal-network(9) NP-complete(33) optimal-realization(3) parsimony(39) pedigree(1) perfect(5) phylogenetic-network(402) phylogeny(387) polynomial(70) population-genetics(5) Program-AdmixTools(1) Program-ALE(1) Program-Angst(2) Program-Arlequin(5) Program-Beagle(3) Program-BIMLR(1) Program-Bio-PhyloNetwork(4) Program-Clustistic(2) Program-CMPT(1) Program-CombineTrees(2) Program-ConsensusNetwork(1) Program-constNJ(1) Program-CycleKiller(3) Program-Dendroscope(14) Program-EEEP(3) Program-FastHN(1) Program-FlatNJ(1) Program-Fylogenetica(2) Program-GalledTree(1) Program-GraphDTL(2) Program-HapBound(1) Program-HGT_simul(1) Program-HiDe(1) Program-HorizStory(2) Program-Hybrid-Lambda(1) Program-HybridInterleave(5) Program-HybridNET(2) Program-HybridNumber(3) Program-Hybroscale(3) Program-JML(1) Program-LatTrans(5) Program-LEV1ATHAN(1) Program-Lev1Generator(1) Program-Level2(2) Program-lingpy(1) Program-LNetwork(2) Program-MaafB(2) Program-Marlon(3) Program-MC-Net(1) Program-McKiTscH(1) Program-Mowgli(5) Program-MowgliNNI(2) Program-MPNet(1) Program-MY-CLOSURE(1) Program-Nepal(7) Program-NetGen(3) Program-NetTest(1) Program-NetView(1) Program-Network(5) Program-Notung(1) Program-PADRE(7) Program-Phangorn(2) Program-PhippsNetwork(2) Program-PhyloNet(11) Program-PhyloNet-HMM(1) Program-PIRN(3) Program-Prunier(2) Program-Pyramids(3) Program-QNet(4) Program-Quartet(1) Program-Quartet-Decomposition(1) Program-QuartetMethods(1) Program-QuartetNet(2) Program-QuasiDec(1) Program-QuickCass(1) Program-RANGER-DTL(4) Program-RecMin(1) Program-Recodon(3) Program-RecPars(1) Program-Reticlad(2) Program-SAGE(2) Program-SAQ-Net(1) Program-Serial-NetEvolve(1) Program-SHRUB(3) Program-Simplistic(3) Program-Sliding-MinPD(1) Program-SNSA(2) Program-Spectronet(4) Program-SplitsTree(36) Program-SPNet(5) Program-SPRDist(1) Program-SuperQ(1) Program-T-REX(13) Program-TCS(8) Program-TERA(2) Program-TerminusEst(2) Program-TreeFix-DTL(1) Program-TreeMix(1) Program-Treevolve(2) Program-TripNet(2) Program-ultra-Net(1) Program-Ultranet(1) Program-WeakHierarchies(2) Program-Xscape(1) pyramid(8) quasi-median-network(3) realization(4) recombination(29) recombination-detection(4) reconstruction(255) regular-network(7) reticulogram(10) serial-evolutionary-networks(1) simulated-annealing(4) simulation(5) site-consistency(1) software(66) split(27) split-decomposition(13) split-network(54) SPR-distance(13) spread(2) statistical-model(32) statistical-parsimony(3) supernetwork(4) survey(31) tanglegram(1) time-consistent-network(12) tree-child-network(15) tree-sibling-network(10) tripartition-distance(9) triplet-distance(3) unicyclic-network(3) visualization(30) weak-hierarchy(8) weakly-compatible(3)

http://phylnet.univ-mlv.fr/keywords.php

Who is who in Phylogenetic Networks? Keywords!

abstract-network(88) agreement-forest(29) approximation(18) APX-hard(4) ARG(6) bayesian(4) block-realization(1) bootstrap(4) bound(4) branch-and-bound(1) cactus-graph(1) characterization(11) circular-split-system(13) clustering(3) coalescent(10) consensus(8) consistency(2) cophylogeny(1) counting(3) database(1) distance-between-networks(30) diversity(5) duplication(33) dynamic-programming(10) enumeration(4) evaluation(25) explicit-network(191) exponential-algorithm(6) FPT(31) from-clusters(16) from-continuous-characters(1) from-distances(45) from-gene-order(1) from-multilabeled-tree(9) from-network(39) from-NGS-data(1) from-quartets(16) from-rooted-trees(122) from-sequences(51) from-species-tree(46) from-splits(13) from-trees(7) from-trinets(2) from-triplets(24) from-unrooted-trees(16) galled-network(7) galled-tree(38) generation(12) haplotype-network(2) haplotyping(1) heuristic(32) HMM(2) hybridization(52) inapproximability(5) integer-linear-programming(3) isomorphism(3) k-reticulated(1) kernelization(2) labeling(4) lateral-gene-transfer(56) level-k-phylogenetic-network(30) likelihood(16) lineage-sorting(10) loss(16) MASN(4) median-network(16) MedianJoining(2) Minimal-lateral-network(1) minimum-contradiction(2) minimum-number(33) minimum-spanning-network(2) model-selection(2) mu-distance(2) NeighborNet(14) nested-network(2) netting(3) normal-network(9) NP-complete(33) optimal-realization(3) parsimony(39) pedigree(1) perfect(5) phylogenetic-network(402) phylogeny(387) polynomial(70) population-genetics(5) Program-AdmixTools(1) Program-ALE(1) Program-Angst(2) Program-Arlequin(5) Program-Beagle(3) Program-BIMLR(1) Program-Bio-PhyloNetwork(4) Program-Clustistic(2) Program-CMPT(1) Program-CombineTrees(2) Program-ConsensusNetwork(1) Program-constNJ(1) Program-CycleKiller(3) Program-Dendroscope(14) Program-EEEP(3) Program-FastHN(1) Program-FlatNJ(1) Program-Fylogenetica(2) Program-GalledTree(1) Program-GraphDTL(2) Program-HapBound(1) Program-HGT_simul(1) Program-HiDe(1) Program-HorizStory(2) Program-Hybrid-Lambda(1) Program-HybridInterleave(5) Program-HybridNET(2) Program-HybridNumber(3) Program-Hybroscale(3) Program-JML(1) Program-LatTrans(5) Program-LEV1ATHAN(1) Program-Lev1Generator(1) Program-Level2(2) Program-lingpy(1) Program-LNetwork(2) Program-MaafB(2) Program-Marlon(3) Program-MC-Net(1) Program-McKiTscH(1) Program-Mowgli(5) Program-MowgliNNI(2) Program-MPNet(1) Program-MY-CLOSURE(1) Program-Nepal(7) Program-NetGen(3) Program-NetTest(1) Program-NetView(1) Program-Network(5) Program-Notung(1) Program-PADRE(7) Program-Phangorn(2) Program-PhippsNetwork(2) Program-PhyloNet(11) Program-PhyloNet-HMM(1) Program-PIRN(3) Program-Prunier(2) Program-Pyramids(3) Program-QNet(4) Program-Quartet(1) Program-Quartet-Decomposition(1) Program-QuartetMethods(1) Program-QuartetNet(2) Program-QuasiDec(1) Program-QuickCass(1) Program-RANGER-DTL(4) Program-RecMin(1) Program-Recodon(3) Program-RecPars(1) Program-Reticlad(2) Program-SAGE(2) Program-SAQ-Net(1) Program-Serial-NetEvolve(1) Program-SHRUB(3) Program-Simplistic(3) Program-Sliding-MinPD(1) Program-SNSA(2) Program-Spectronet(4) Program-SplitsTree(36) Program-SPNet(5) Program-SPRDist(1) Program-SuperQ(1) Program-T-REX(13) Program-TCS(8) Program-TERA(2) Program-TerminusEst(2) Program-TreeFix-DTL(1) Program-TreeMix(1) Program-Treevolve(2) Program-TripNet(2) Program-ultra-Net(1) Program-Ultranet(1) Program-WeakHierarchies(2) Program-Xscape(1) pyramid(8) quasi-median-network(3) realization(4) recombination(29) recombination-detection(4) reconstruction(255) regular-network(7) reticulogram(10) serial-evolutionary-networks(1) simulated-annealing(4) simulation(5) site-consistency(1) software(66) split(27) split-decomposition(13) split-network(54) SPR-distance(13) spread(2) statistical-model(32) statistical-parsimony(3) supernetwork(4) survey(31) tanglegram(1) time-consistent-network(12) tree-child-network(15) tree-sibling-network(10) tripartition-distance(9) triplet-distance(3) unicyclic-network(3) visualization(30) weak-hierarchy(8) weakly-compatible(3)

input

software

http://phylnet.univ-mlv.fr/keywords.php

Which methods to build networks?

- from distances:

NeighborNet (split networks), T-Rex (reticulograms), etc.

- from sequences:

Parsimony, likelihood.

- from trees (gene trees, species trees):

Consensus networks, supernetworks, reconciliation scenarios with duplication loss and transfer costs.

Building explicit networks from distances

- distance of one of the paths between two leaves

(Chan et al., JBCB, 2006)

- minimum distance

(Bordewich, Huber, Moulton & Semple, JMB, 2018)

- all path distances

(Bordewich & Tokac, DAM, 2016)

- weighted distance

(Willson, BMB, 2013; Francis & Steel, MB, 2015; Willems, Tahiri & Makarenkov, JBCB, 2014)

Building explicit networks from distances

- distance of one of the paths between two leaves

(Chan et al., JBCB, 2006)

- minimum distance

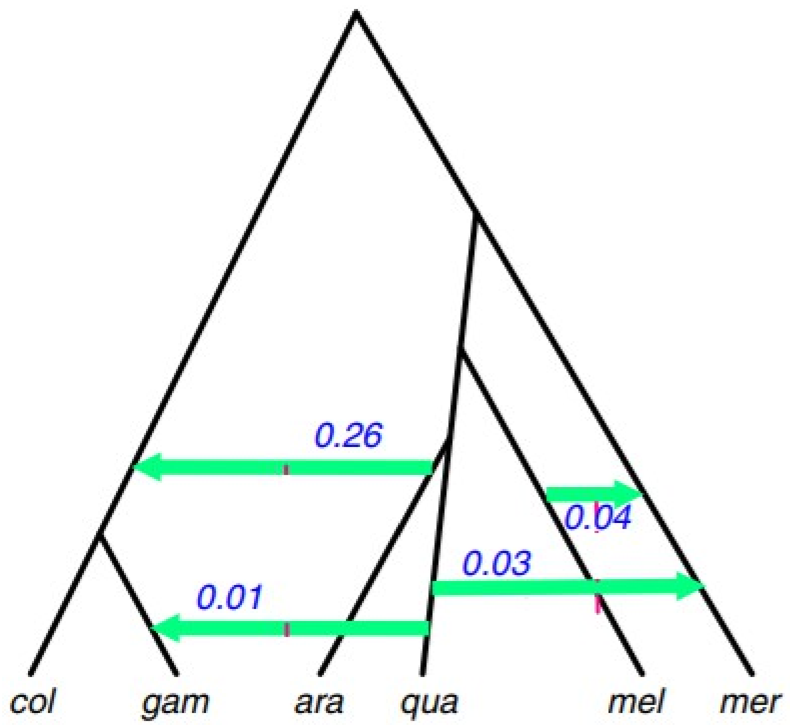
(Bordewich, Huber, Moulton & Semple, JMB, 2018)

- all path distances

(Bordewich & Tokac, DAM, 2016)

- weighted distance

(Willson, BMB, 2013; Francis & Steel, MB, 2015; Willems, Tahiri & Makarenkov, JBCB, 2014)



Adapted from Figure 6 of Wen D, Yu Y, Hahn MW, Nakhleh L (2016). Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Molecular Ecology* 25(11):2361-2372

Building explicit networks from distances

- distance of one of the paths between two leaves

(Chan et al., JBCB, 2006)

- minimum distance

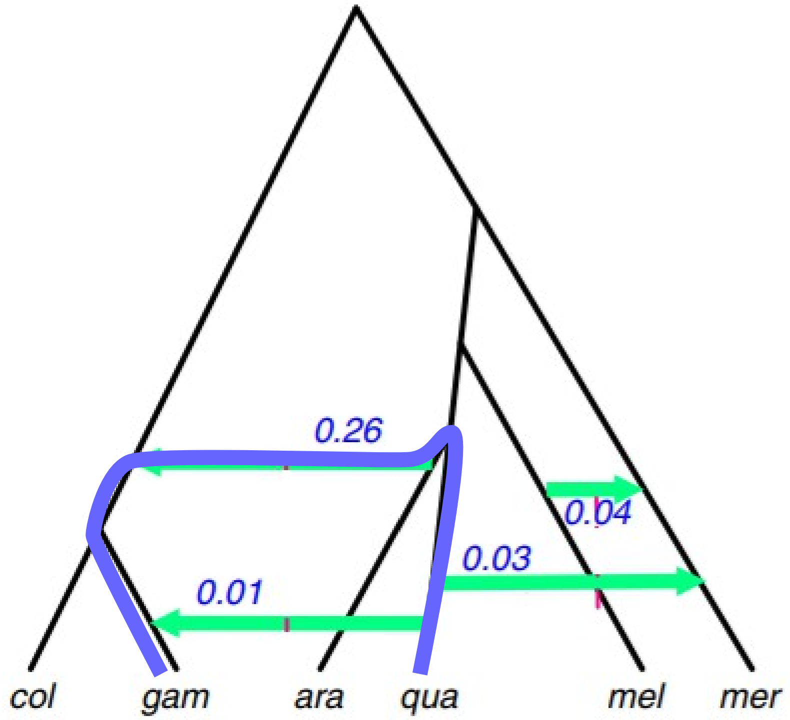
(Bordewich, Huber, Moulton & Semple, JMB, 2018)

- all path distances

(Bordewich & Tokac, DAM, 2016)

- weighted distance

(Willson, BMB, 2013; Francis & Steel, MB, 2015; Willems, Tahiri & Makarenkov, JBCB, 2014)

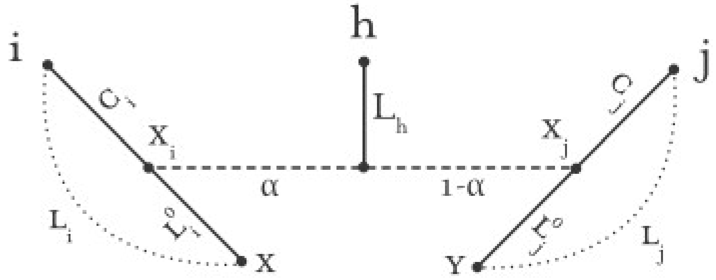


Adapted from Figure 6 of Wen D, Yu Y, Hahn MW, Nakhleh L (2016). Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Molecular Ecology* 25(11):2361-2372

Building explicit networks from distances

Weighted distance:

- proportion α of the genetic material coming from the left
- proportion 1-α coming from the right



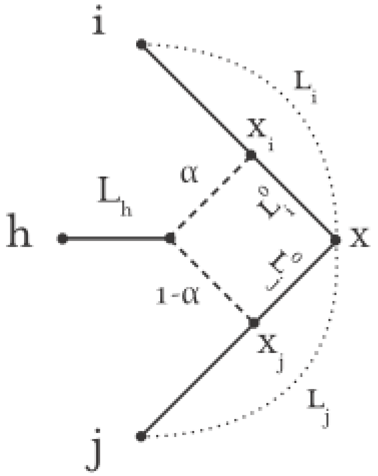
Willems, M., Tahiri, N. & Makarenkov, V. (2014). A new efficient algorithm for inferring explicit hybridization networks following the neighbor-joining principle. Journal of Bioinformatics and Computational Biology, 12(05), 1450024.

Building explicit networks from distances

Weighted distance:

- proportion α of the genetic material coming from the left
- proportion 1-α coming from the right

No unique solution for hybrids between neigbors:



Willems, M., Tahiri, N. & Makarenkov, V. (2014). A new efficient algorithm for inferring explicit hybridization networks following the neighbor-joining principle. Journal of Bioinformatics and Computational Biology, 12(05), 1450024.

Phylogenetic network reconstruction

```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

G1 G2

{gene sequences}

distance methods

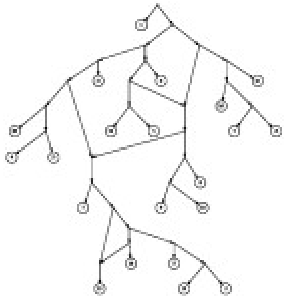
Bandelt & Dress 1992 - Legendre & Makarenkov 2000 - Bryant & Moulton 2002 - Chan, Jansson, Lam & Yiu 2006 - Willems, Tahiri & Makarenkov 2014

parsimony methods

Hein 1990 - Kececioglu & Gusfield 1994 - Jin, Nakhleh, Snir, Tuller 2009 - Park, Jin & Nakhleh 2010 - Kannan & Wheeler, 2012 - Kelk, Pardio, Scornavacca & van Iersel, 2017

likelihood methods

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



network *N*

Phylogenetic network reconstruction

**Problem: methods are usually slow,
especially with rapidly increasing sequence length.**

```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

G1 G2

{gene sequences}

distance methods

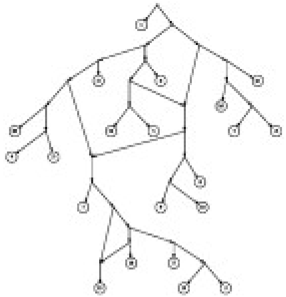
Bandelt & Dress 1992 - Legendre & Makarenkov 2000 - Bryant & Moulton 2002 - Chan, Jansson, Lam & Yiu 2006 - Willems, Tahiri & Makarenkov 2014

parsimony methods

Hein 1990 - Kececioglu & Gusfield 1994 - Jin, Nakhleh, Snir, Tuller 2009 - Park, Jin & Nakhleh 2010 - Kannan & Wheeler, 2012 - Kelk, Pardio, Scornavacca & van Iersel, 2017

likelihood methods

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



network *N*

Phylogenetic network reconstruction from trees

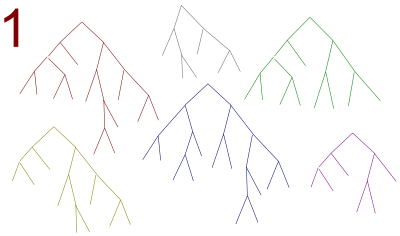
```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

G1 G2

{gene sequences}

Reconstruction of a tree for each gene present in several species

Guindon & Gascuel, SB, 2003



T2

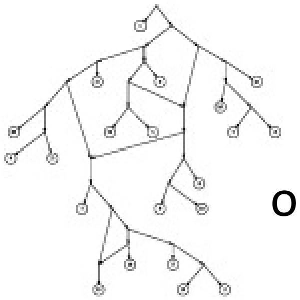
{trees}

HOGENOM Database

Dufayard, Duret, Penel, Gouy, Rechenmann & Perrière, BioInf, 2005

Tree reconciliation or consensus

explicit network



optimal super-network *N*:
- contains the input trees
- has the smallest number of reticulations

http://doua.prabi.fr/databases/hogenom/home.php?contents=hogenom4

Phylogenetic network reconstruction from trees

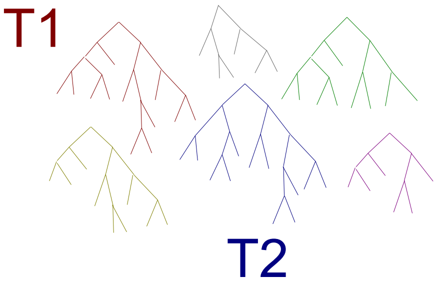
```
espèce 1 : AATTGCAG TAGCCCAAAAT
espèce 2 : ACCTGCAG TAGACCAAT
espèce 3 : GCTTGCCG TAGACAAGAAT
espèce 4 : ATTTGCAG AAGACCAAAT
espèce 5 :          TAGACAAGAAT
espèce 6 : ACTTGCAG TAGCACAAAAT
espèce 7 : ACCTGGTG TAAAAT
```

G1 G2

{gene sequences}

Reconstruction of a tree for each gene present in several species

Guindon & Gascuel, SB, 2003



{trees}

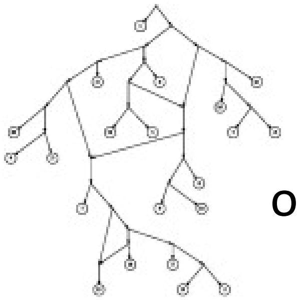
HOGENOM Database

Dufayard, Duret, Penel, Gouy, Rechenmann & Perrière, BioInf, 2005

1470 species, >290 000 trees

Tree reconciliation or consensus

explicit network



optimal super-network *N*:
- contains the input trees
- has the smallest number of reticulations

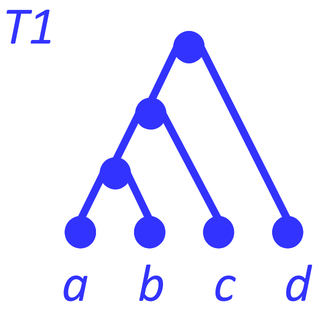
http://doua.prabi.fr/databases/hogenom/home.php?contents=hogenom4

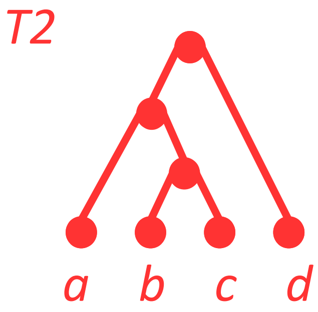
Phylogenetic network reconstruction from trees

The **“hybridization network” problem**:

given 2 trees, find the smallest network containing both of them with the minimum number of hybrid vertices

Easy to find a network containing the two trees!



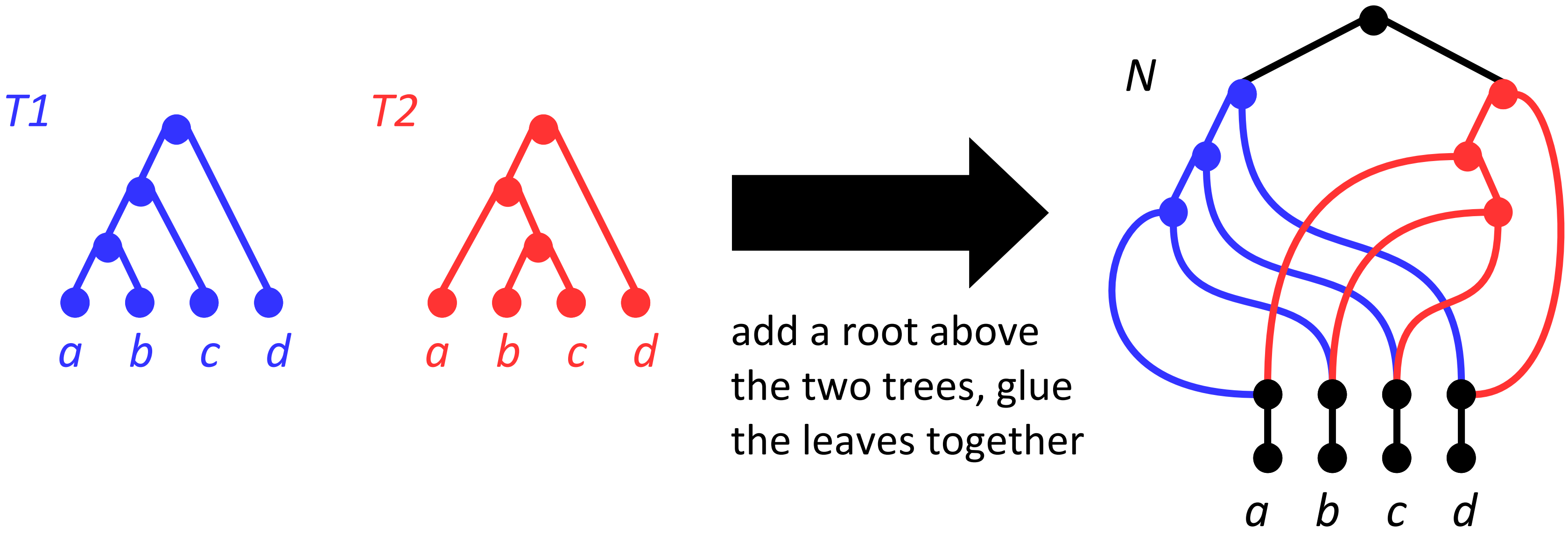


Phylogenetic network reconstruction from trees

The **“hybridization network” problem**:

given 2 trees, find the smallest network containing both of them with the minimum number of hybrid vertices

Easy to find a network containing the two trees!



add a root above the two trees, glue the leaves together

But *n* hybrid vertices for trees with *n* leaves: not optimal!

Phylogenetic network reconstruction from trees

The **“hybridization network” problem**:

given 2 trees, find the smallest network containing both of them with the minimum number of hybrid vertices

NP-hard to minimize the number of hybrid vertices

Still, a few practically efficient algorithms:

- Huson & Linz, Autumn Algorithm, 2018
- Mirzaei & Wu, PIRN, 2016
- Chen & Wang, ultra-Net, 2013

Phylogenetic network reconstruction from trees

The **“hybridization network” problem**:

given 2 trees, find the smallest network containing both of them with the minimum number of hybrid vertices

Even **checking a solution** is **hard**!

The Tree Containment Problem (T.C.P.)

Input: A binary phylogenetic network *N* and a tree *T* over the same set of taxa.

Question: Does *N* display *T*?

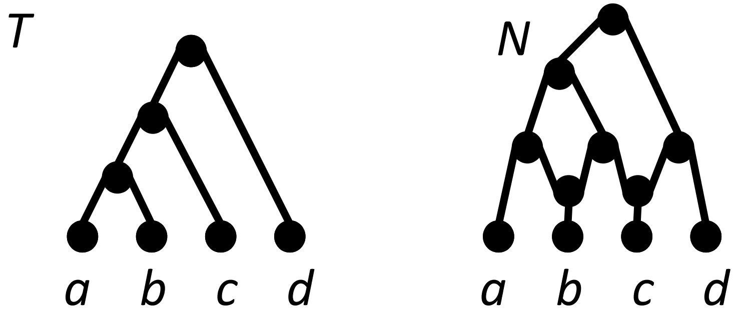
The Tree Containment Problem (T.C.P.)

Input: A binary phylogenetic network *N* and

a tree *T* over the same set of taxa.

Question: Does *N* display *T*?

→ Can we remove one incoming arc, for each vertex with >1 parent in *N*, such that the obtained tree is equivalent to *T*?



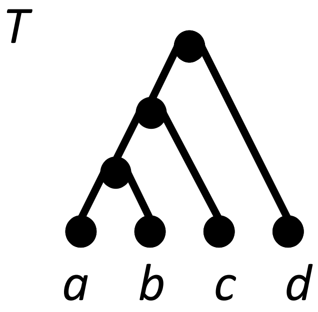
The Tree Containment Problem (T.C.P.)

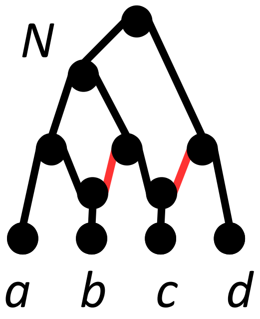
Input: A binary phylogenetic network *N* and

a tree *T* over the same set of taxa.

Question: Does *N* display *T*?

→ Can we **remove one incoming arc**, for each vertex with >1 parent in *N*, such that the obtained tree is equivalent to *T*?





The Tree Containment Problem (T.C.P.)

Input: A binary phylogenetic network *N* and

a tree *T* over the same set of taxa.

Question: Does *N* display *T*?

- **NP-complete** in general (Kanj, Nakhleh, Than & Xia, 2008)

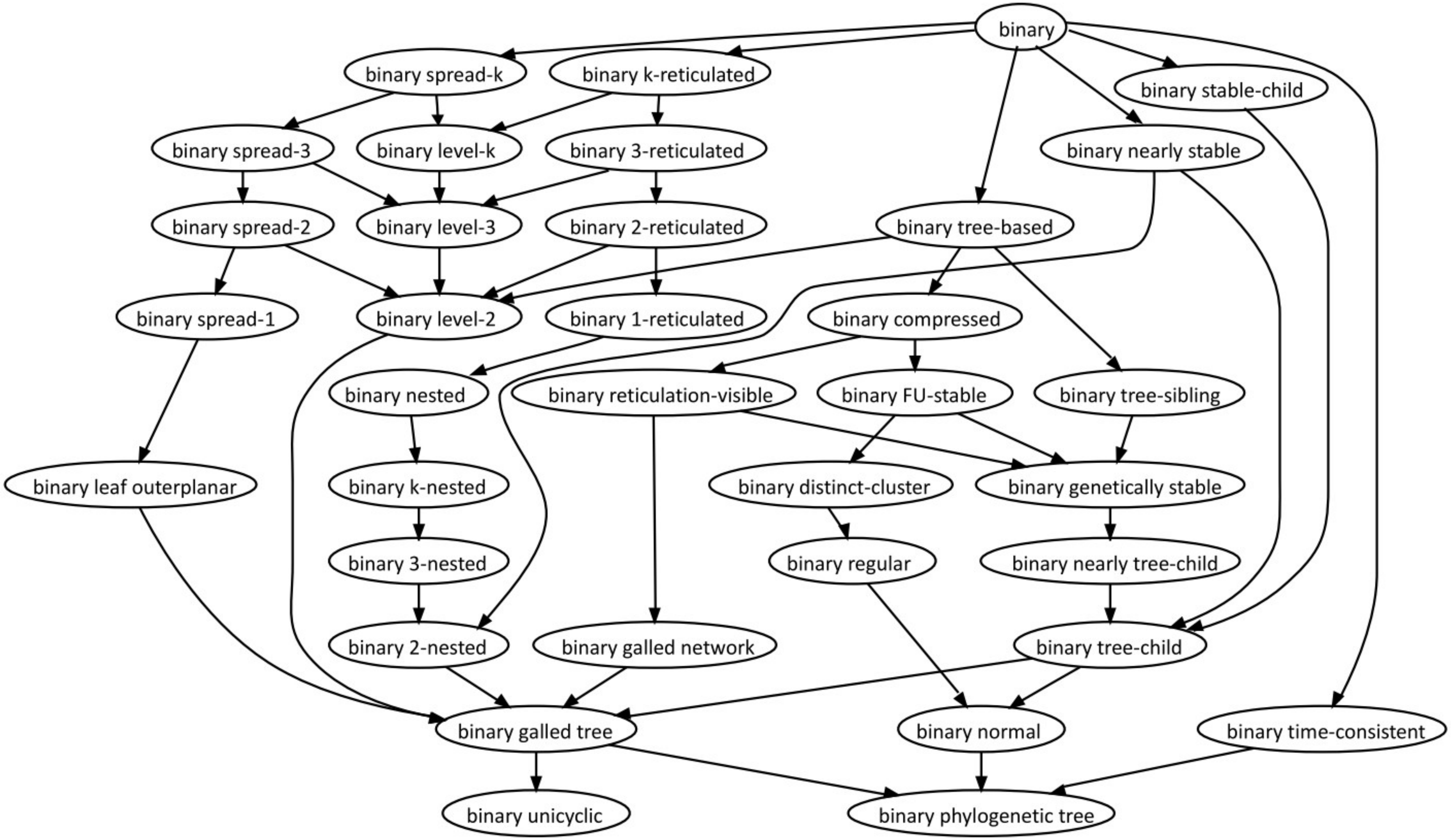
- **NP-complete** for tree-sibling, time-consistent, regular networks (Iersel, Semple & Steel, 2010)

- **Polynomial-time solvable** for normal networks, for binary tree-child networks, and for level-k networks (Iersel, Semple & Steel, 2010)

Outline

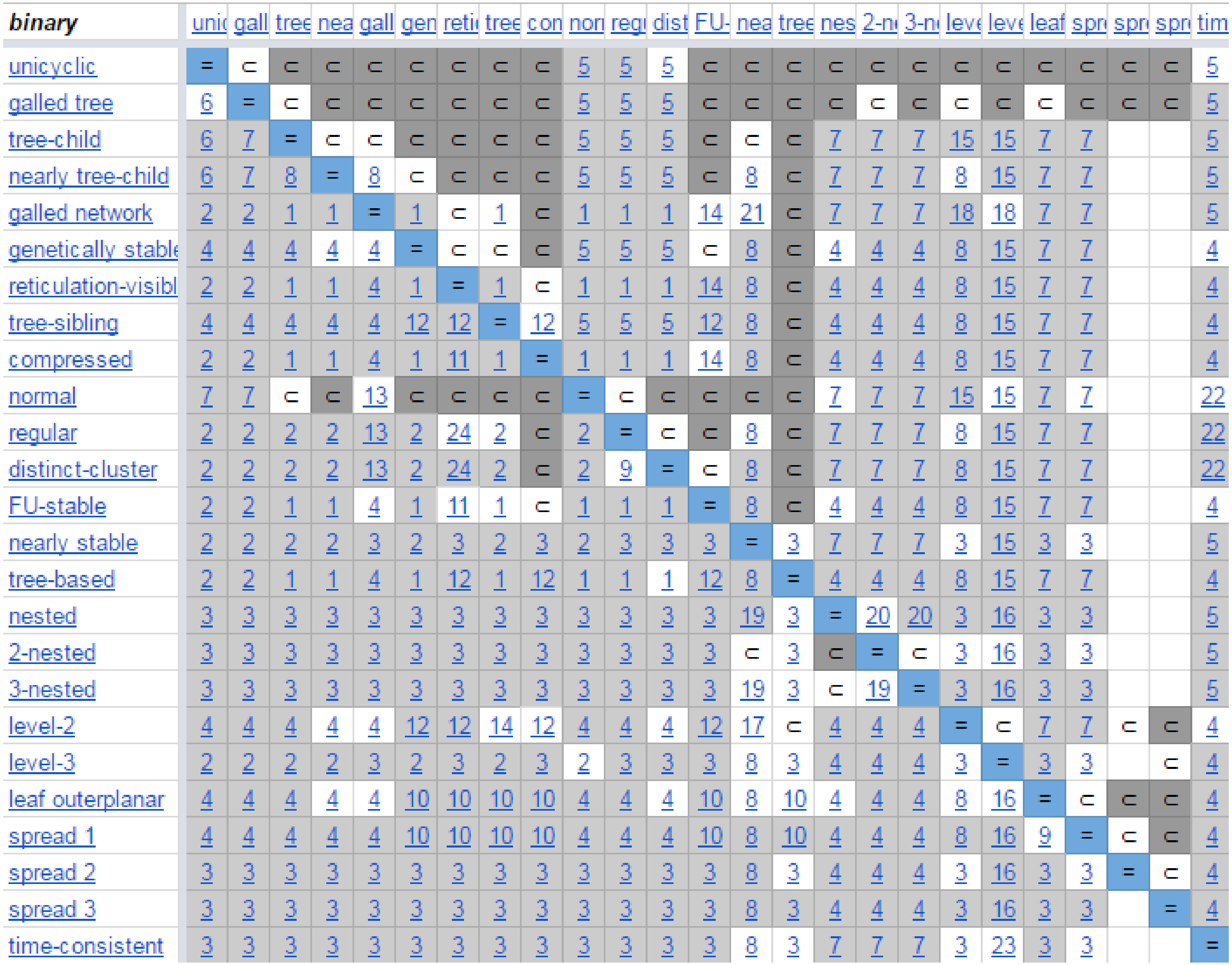
- Phylogenetic networks
- How to build phylogenetic networks?
- Classes of phylogenetic networks
- Unexpected properties of networks
- The research community on phylogenetic networks

Classes of phylogenetic networks



joint work with Maxime Morgado and Narges Tavassoli
http://phylnet.univ-mlv.fr/isiphync/

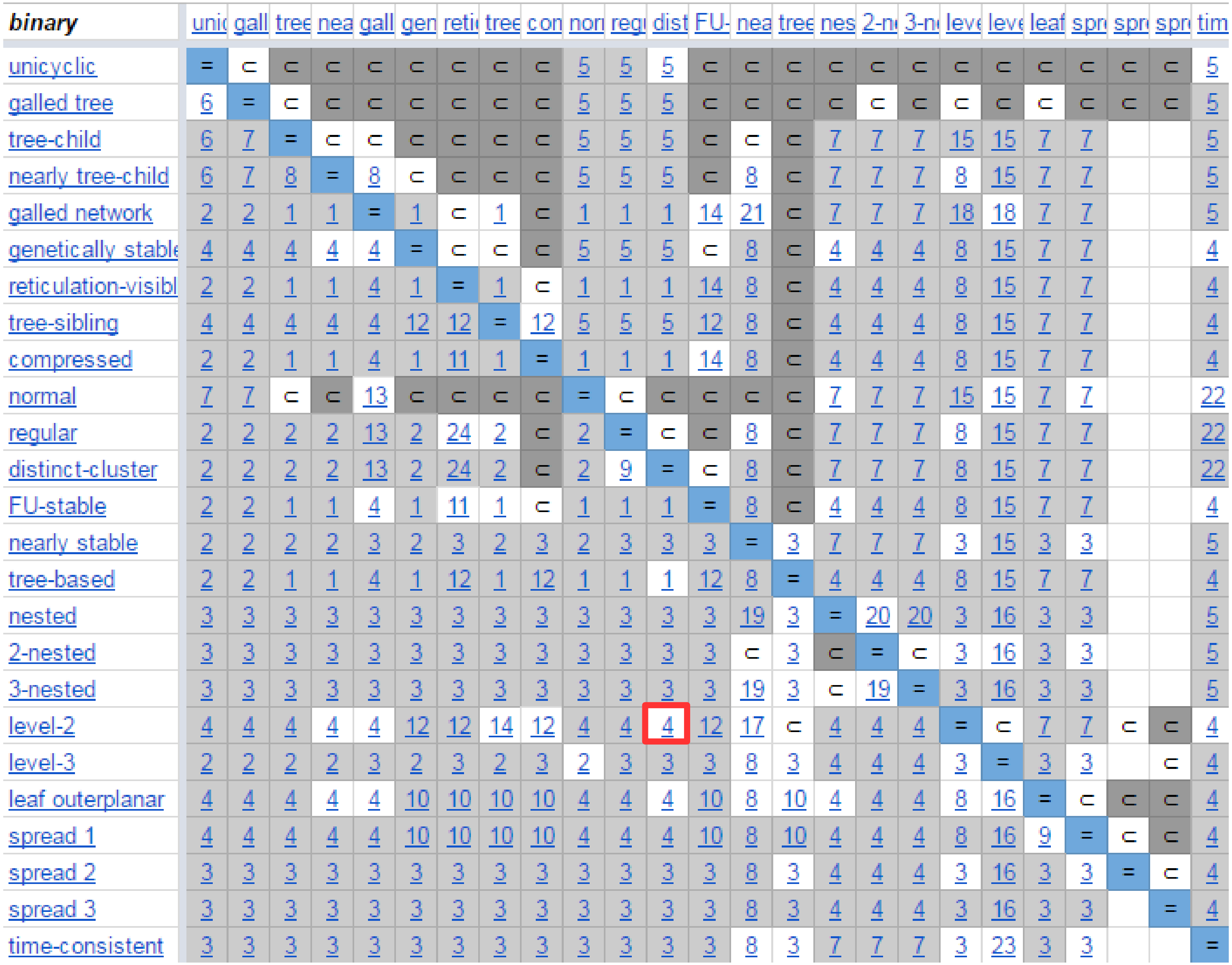
Classes of phylogenetic networks: inclusions



binary	unic	gall	tree	nea	gall	gen	reti	tree	con	nor	regi	dist	FU-	nea	tree	nes	2-n	3-n	leve	leve	leaf	spr	spr	spr	tim
unicyclic	=	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5
galled tree	6	=	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5
tree-child	6	7	=	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	7	7	7	15	15	7	7			5
nearly tree-child	6	7	8	=	8	⊂	⊂	⊂	⊂	5	5	5	⊂	8	⊂	7	7	7	8	15	7	7			5
galled network	2	2	1	1	=	1	⊂	1	⊂	1	1	1	14	21	⊂	7	7	7	18	18	7	7			5
genetically stable	4	4	4	4	4	=	⊂	⊂	⊂	5	5	5	⊂	8	⊂	4	4	4	8	15	7	7			4
reticulation-visibl	2	2	1	1	4	1	=	1	⊂	1	1	1	14	8	⊂	4	4	4	8	15	7	7			4
tree-sibling	4	4	4	4	4	12	12	=	12	5	5	5	12	8	⊂	4	4	4	8	15	7	7			4
compressed	2	2	1	1	4	1	11	1	=	1	1	1	14	8	⊂	4	4	4	8	15	7	7			4
normal	7	7	⊂	⊂	13	⊂	⊂	⊂	⊂	=	⊂	⊂	⊂	⊂	⊂	7	7	7	15	15	7	7			22
regular	2	2	2	2	13	2	24	2	⊂	2	=	⊂	⊂	8	⊂	7	7	7	8	15	7	7			22
distinct-cluster	2	2	2	2	13	2	24	2	⊂	2	9	=	⊂	8	⊂	7	7	7	8	15	7	7			22
FU-stable	2	2	1	1	4	1	11	1	⊂	1	1	1	=	8	⊂	4	4	4	8	15	7	7			4
nearly stable	2	2	2	2	3	2	3	2	3	2	3	3	3	=	3	7	7	7	3	15	3	3			5
tree-based	2	2	1	1	4	1	12	1	12	1	1	1	12	8	=	4	4	4	8	15	7	7			4
nested	3	3	3	3	3	3	3	3	3	3	3	3	3	19	3	=	20	20	3	16	3	3			5
2-nested	3	3	3	3	3	3	3	3	3	3	3	3	3	⊂	3	⊂	=	⊂	3	16	3	3			5
3-nested	3	3	3	3	3	3	3	3	3	3	3	3	3	19	3	⊂	19	=	3	16	3	3			5
level-2	4	4	4	4	4	12	12	14	12	4	4	4	12	17	⊂	4	4	4	=	⊂	7	7	⊂	⊂	4
level-3	2	2	2	2	3	2	3	2	3	2	3	3	3	8	3	4	4	4	3	=	3	3		⊂	4
leaf outerplanar	4	4	4	4	4	10	10	10	10	4	4	4	10	8	10	4	4	4	8	16	=	⊂	⊂	⊂	4
spread 1	4	4	4	4	4	10	10	10	10	4	4	4	10	8	10	4	4	4	8	16	9	=	⊂	⊂	4
spread 2	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	4	4	4	3	16	3	3	=	⊂	4
spread 3	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	4	4	4	3	16	3	3		=	4
time-consistent	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	7	7	7	3	23	3	3			=

joint work with Maxime Morgado and Narges Tavassoli

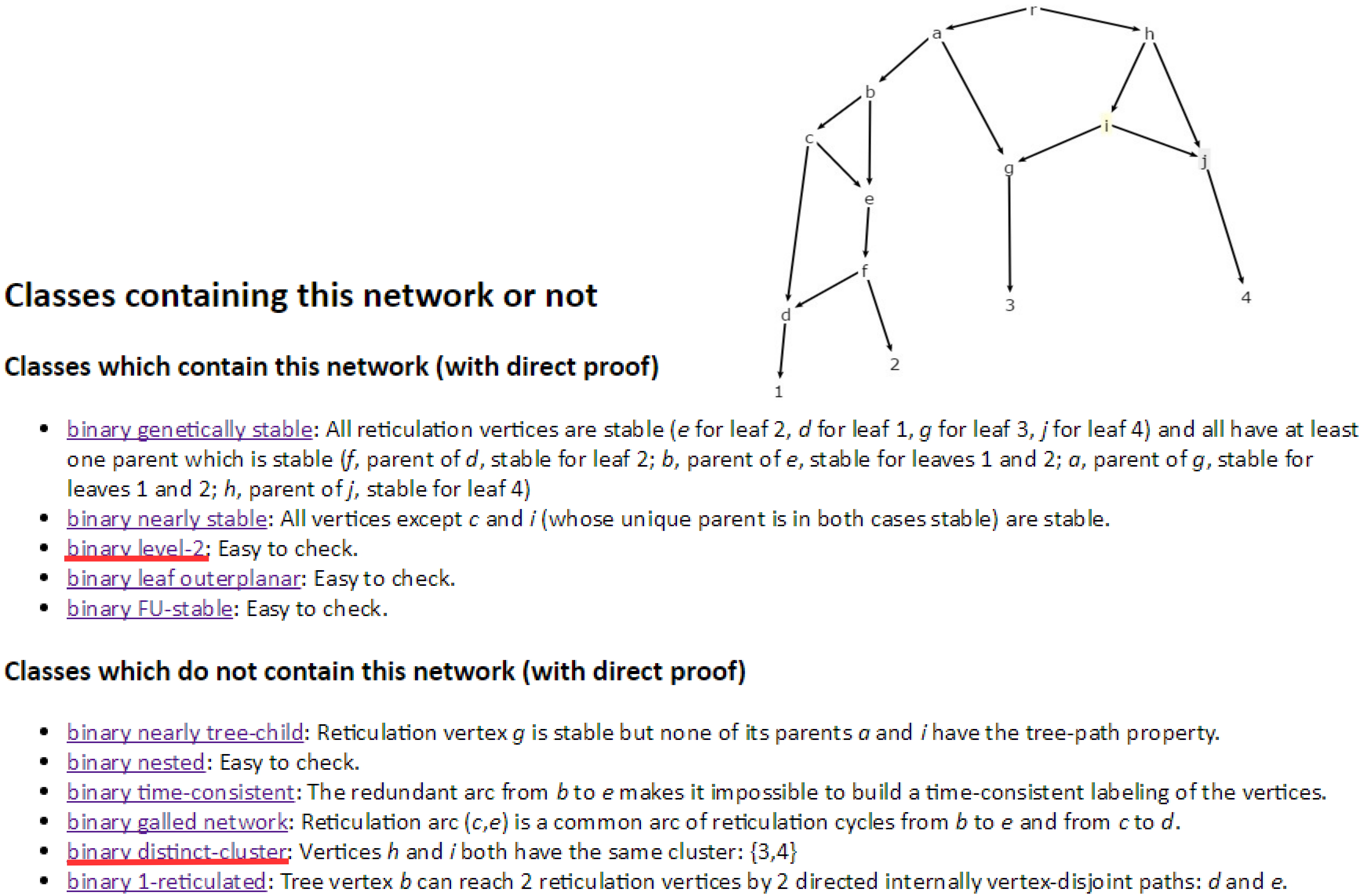
Classes of phylogenetic networks: inclusions



binary	unic	gall	tree	nea	gall	gen	reti	tree	con	non	reg	dist	FU-	nea	tree	nes	2-n	3-n	leve	leve	leaf	spr	spr	spr	tim
unicyclic	=	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5
galled tree	6	=	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	⊂	5
tree-child	6	7	=	⊂	⊂	⊂	⊂	⊂	⊂	5	5	5	⊂	⊂	⊂	7	7	7	15	15	7	7			5
nearly tree-child	6	7	8	=	8	⊂	⊂	⊂	⊂	5	5	5	⊂	8	⊂	7	7	7	8	15	7	7			5
galled network	2	2	1	1	=	1	⊂	1	⊂	1	1	1	14	21	⊂	7	7	7	18	18	7	7			5
genetically stabl	4	4	4	4	4	=	⊂	⊂	⊂	5	5	5	⊂	8	⊂	4	4	4	8	15	7	7			4
reticulation-visibl	2	2	1	1	4	1	=	1	⊂	1	1	1	14	8	⊂	4	4	4	8	15	7	7			4
tree-sibling	4	4	4	4	4	12	12	=	12	5	5	5	12	8	⊂	4	4	4	8	15	7	7			4
compressed	2	2	1	1	4	1	11	1	=	1	1	1	14	8	⊂	4	4	4	8	15	7	7			4
normal	7	7	⊂	⊂	13	⊂	⊂	⊂	⊂	=	⊂	⊂	⊂	⊂	⊂	7	7	7	15	15	7	7			22
regular	2	2	2	2	13	2	24	2	⊂	2	=	⊂	⊂	8	⊂	7	7	7	8	15	7	7			22
distinct-cluster	2	2	2	2	13	2	24	2	⊂	2	9	=	⊂	8	⊂	7	7	7	8	15	7	7			22
FU-stable	2	2	1	1	4	1	11	1	⊂	1	1	1	=	8	⊂	4	4	4	8	15	7	7			4
nearly stable	2	2	2	2	3	2	3	2	3	2	3	3	3	=	3	7	7	7	3	15	3	3			5
tree-based	2	2	1	1	4	1	12	1	12	1	1	1	12	8	=	4	4	4	8	15	7	7			4
nested	3	3	3	3	3	3	3	3	3	3	3	3	3	19	3	=	20	20	3	16	3	3			5
2-nested	3	3	3	3	3	3	3	3	3	3	3	3	3	⊂	3	⊂	=	⊂	3	16	3	3			5
3-nested	3	3	3	3	3	3	3	3	3	3	3	3	3	19	3	⊂	19	=	3	16	3	3			5
level-2	4	4	4	4	4	12	12	14	12	4	4	4	12	17	⊂	4	4	4	=	⊂	7	7	⊂	⊂	4
level-3	2	2	2	2	3	2	3	2	3	2	3	3	3	8	3	4	4	4	3	=	3	3		⊂	4
leaf outerplanar	4	4	4	4	4	10	10	10	10	4	4	4	10	8	10	4	4	4	8	16	=	⊂	⊂	⊂	4
spread 1	4	4	4	4	4	10	10	10	10	4	4	4	10	8	10	4	4	4	8	16	9	=	⊂	⊂	4
spread 2	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	4	4	4	3	16	3	3	=	⊂	4
spread 3	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	4	4	4	3	16	3	3		=	4
time-consistent	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	7	7	7	3	23	3	3			=

joint work with Maxime Morgado and Narges Tavassoli

Classes of phylogenetic networks: inclusions



Classes containing this network or not

Classes which contain this network (with direct proof)

- binary genetically stable: All reticulation vertices are stable (*e* for leaf 2, *d* for leaf 1, *g* for leaf 3, *j* for leaf 4) and all have at least one parent which is stable (*f*, parent of *d*, stable for leaf 2; *b*, parent of *e*, stable for leaves 1 and 2; *a*, parent of *g*, stable for leaves 1 and 2; *h*, parent of *j*, stable for leaf 4)
- binary nearly stable: All vertices except *c* and *i* (whose unique parent is in both cases stable) are stable.
- binary level-2: Easy to check.
- binary leaf outerplanar: Easy to check.
- binary FU-stable: Easy to check.

Classes which do not contain this network (with direct proof)

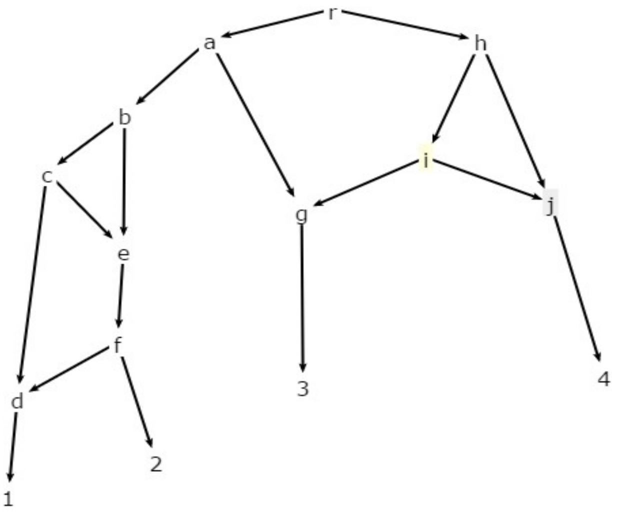
- binary nearly tree-child: Reticulation vertex *g* is stable but none of its parents *a* and *i* have the tree-path property.
- binary nested: Easy to check.
- binary time-consistent: The redundant arc from *b* to *e* makes it impossible to build a time-consistent labeling of the vertices.
- binary galled network: Reticulation arc (*c*,*e*) is a common arc of reticulation cycles from *b* to *e* and from *c* to *d*.
- binary distinct-cluster: Vertices *h* and *i* both have the same cluster: {3,4}
- binary 1-reticulated: Tree vertex *b* can reach 2 reticulation vertices by 2 directed internally vertex-disjoint paths: *d* and *e*.

joint work with Maxime Morgado and Narges Tavassoli
http://phylnet.univ-mlv.fr/isiphync/network.php?id=4

Classes of phylogenetic networks: inclusions

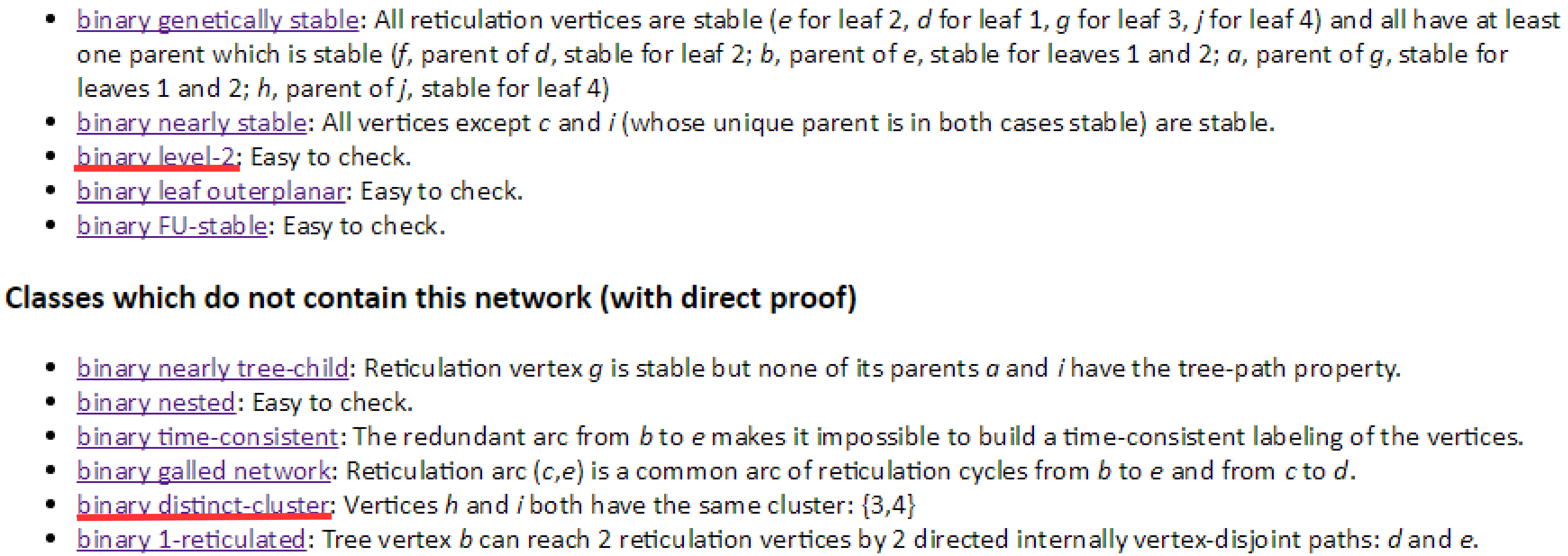
level = maximum number of reticulation vertices among all bridgeless components in the network

cluster = set of leaves below a vertex



Classes containing this network or not

Classes which contain this network (with direct proof)



- binary genetically stable: All reticulation vertices are stable (*e* for leaf 2, *d* for leaf 1, *g* for leaf 3, *j* for leaf 4) and all have at least one parent which is stable (*f*, parent of *d*, stable for leaf 2; *b*, parent of *e*, stable for leaves 1 and 2; *a*, parent of *g*, stable for leaves 1 and 2; *h*, parent of *j*, stable for leaf 4)
- binary nearly stable: All vertices except *c* and *i* (whose unique parent is in both cases stable) are stable.
- binary level-2: Easy to check.
- binary leaf outerplanar: Easy to check.
- binary FU-stable: Easy to check.

Classes which do not contain this network (with direct proof)

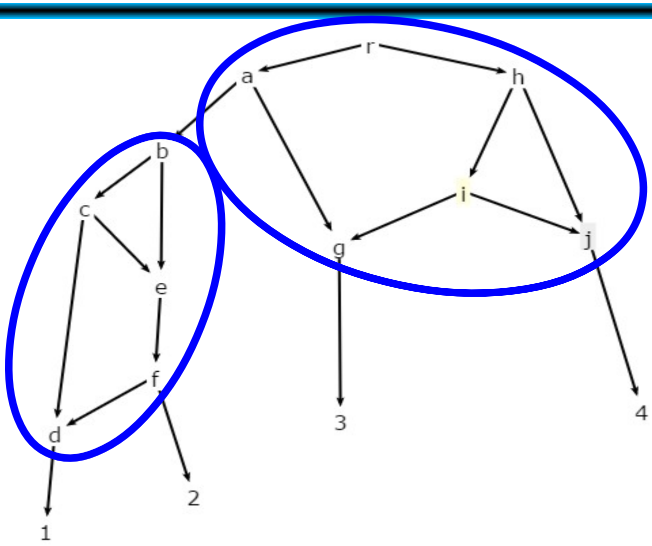
- binary nearly tree-child: Reticulation vertex *g* is stable but none of its parents *a* and *i* have the tree-path property.
- binary nested: Easy to check.
- binary time-consistent: The redundant arc from *b* to *e* makes it impossible to build a time-consistent labeling of the vertices.
- binary galled network: Reticulation arc (*c*,*e*) is a common arc of reticulation cycles from *b* to *e* and from *c* to *d*.
- binary distinct-cluster: Vertices *h* and *i* both have the same cluster: {3,4}
- binary 1-reticulated: Tree vertex *b* can reach 2 reticulation vertices by 2 directed internally vertex-disjoint paths: *d* and *e*.

joint work with Maxime Morgado and Narges Tavassoli
http://phylnet.univ-mlv.fr/isiphync/network.php?id=4

Classes of phylogenetic networks: inclusions

level = maximum number of reticulation vertices among all **bridgeless components** in the network

cluster = set of leaves below a vertex



Classes containing this network or not

Classes which contain this network (with direct proof)

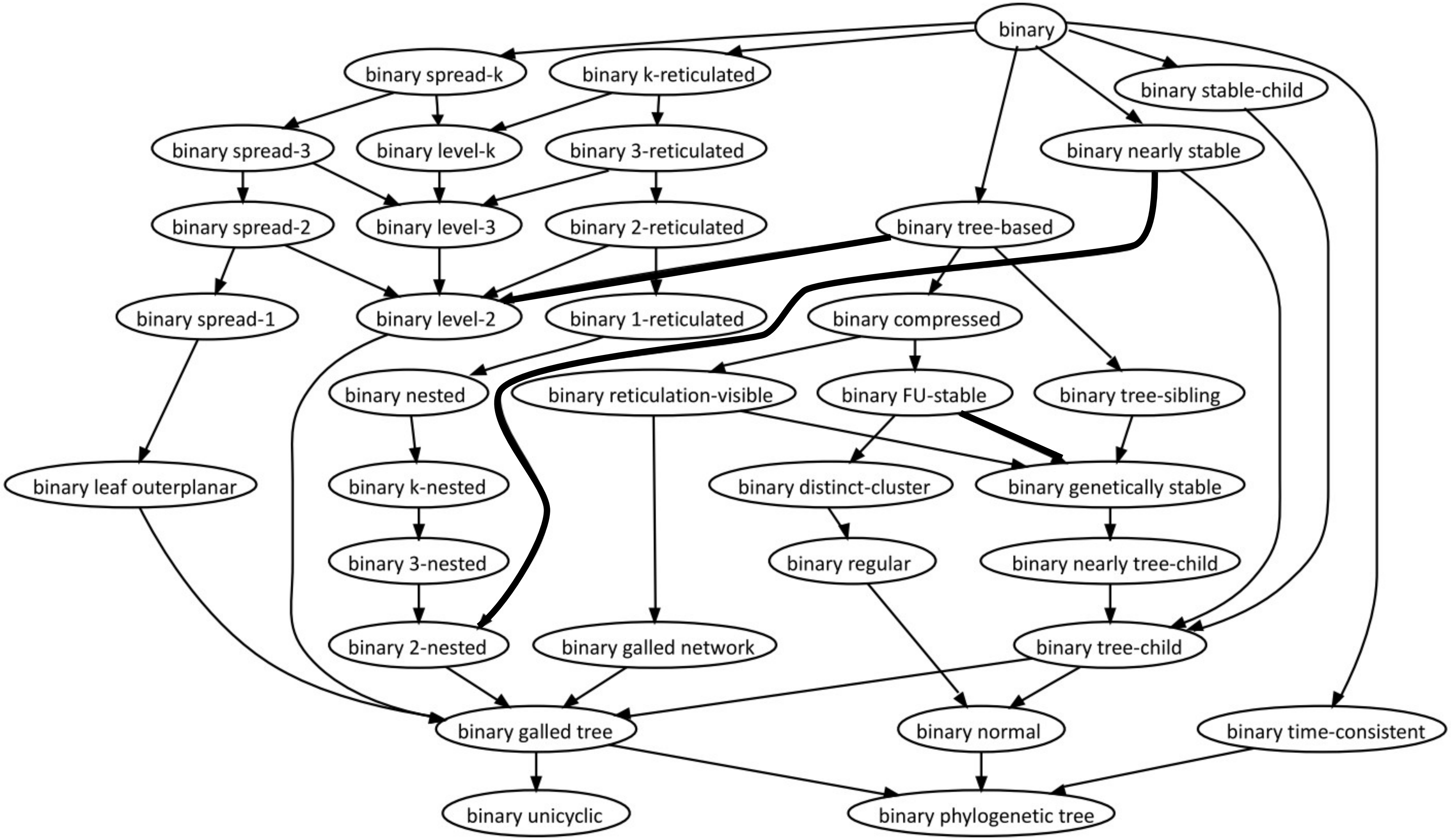
- binary genetically stable: All reticulation vertices are stable (*e* for leaf 2, *d* for leaf 1, *g* for leaf 3, *j* for leaf 4) and all have at least one parent which is stable (*f*, parent of *d*, stable for leaf 2; *b*, parent of *e*, stable for leaves 1 and 2; *a*, parent of *g*, stable for leaves 1 and 2; *h*, parent of *j*, stable for leaf 4)
- binary nearly stable: All vertices except *c* and *i* (whose unique parent is in both cases stable) are stable.
- binary level-2: Easy to check.
- binary leaf outerplanar: Easy to check.
- binary FU-stable: Easy to check.

Classes which do not contain this network (with direct proof)

- binary nearly tree-child: Reticulation vertex *g* is stable but none of its parents *a* and *i* have the tree-path property.
- binary nested: Easy to check.
- binary time-consistent: The redundant arc from *b* to *e* makes it impossible to build a time-consistent labeling of the vertices.
- binary galled network: Reticulation arc (*c*,*e*) is a common arc of reticulation cycles from *b* to *e* and from *c* to *d*.
- binary distinct-cluster: Vertices *h* and *i* both have the same cluster: {3,4}
- binary 1-reticulated: Tree vertex *b* can reach 2 reticulation vertices by 2 directed internally vertex-disjoint paths: *d* and *e*.

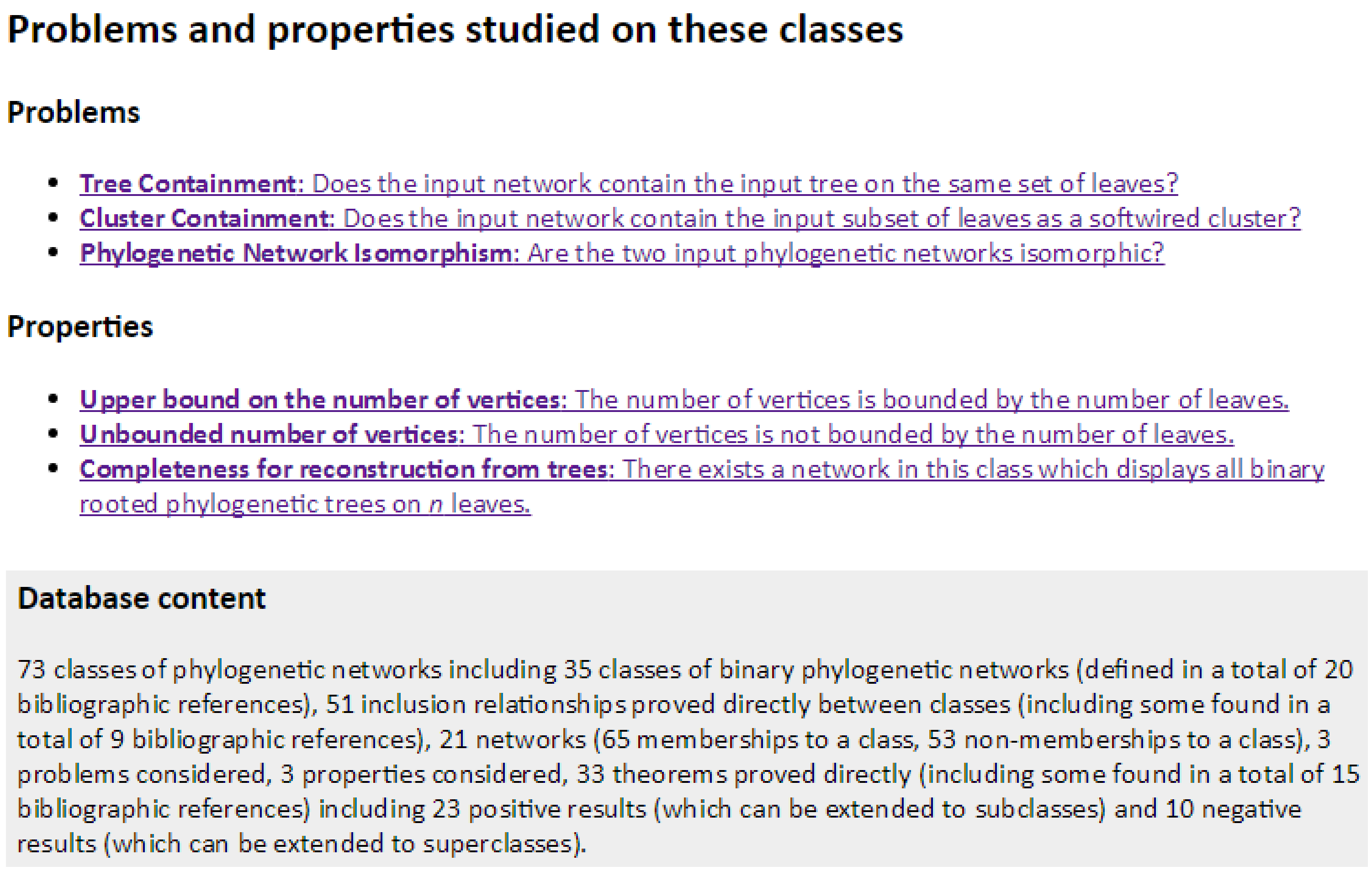
joint work with Maxime Morgado and Narges Tavassoli
http://phylnet.univ-mlv.fr/isiphync/network.php?id=4

Classes of phylogenetic networks: new inclusions



joint work with Maxime Morgado and Narges Tavassoli

Classes of phylogenetic networks



Problems and properties studied on these classes

Problems

- **Tree Containment**: Does the input network contain the input tree on the same set of leaves?
- **Cluster Containment**: Does the input network contain the input subset of leaves as a softwired cluster?
- **Phylogenetic Network Isomorphism**: Are the two input phylogenetic networks isomorphic?

Properties

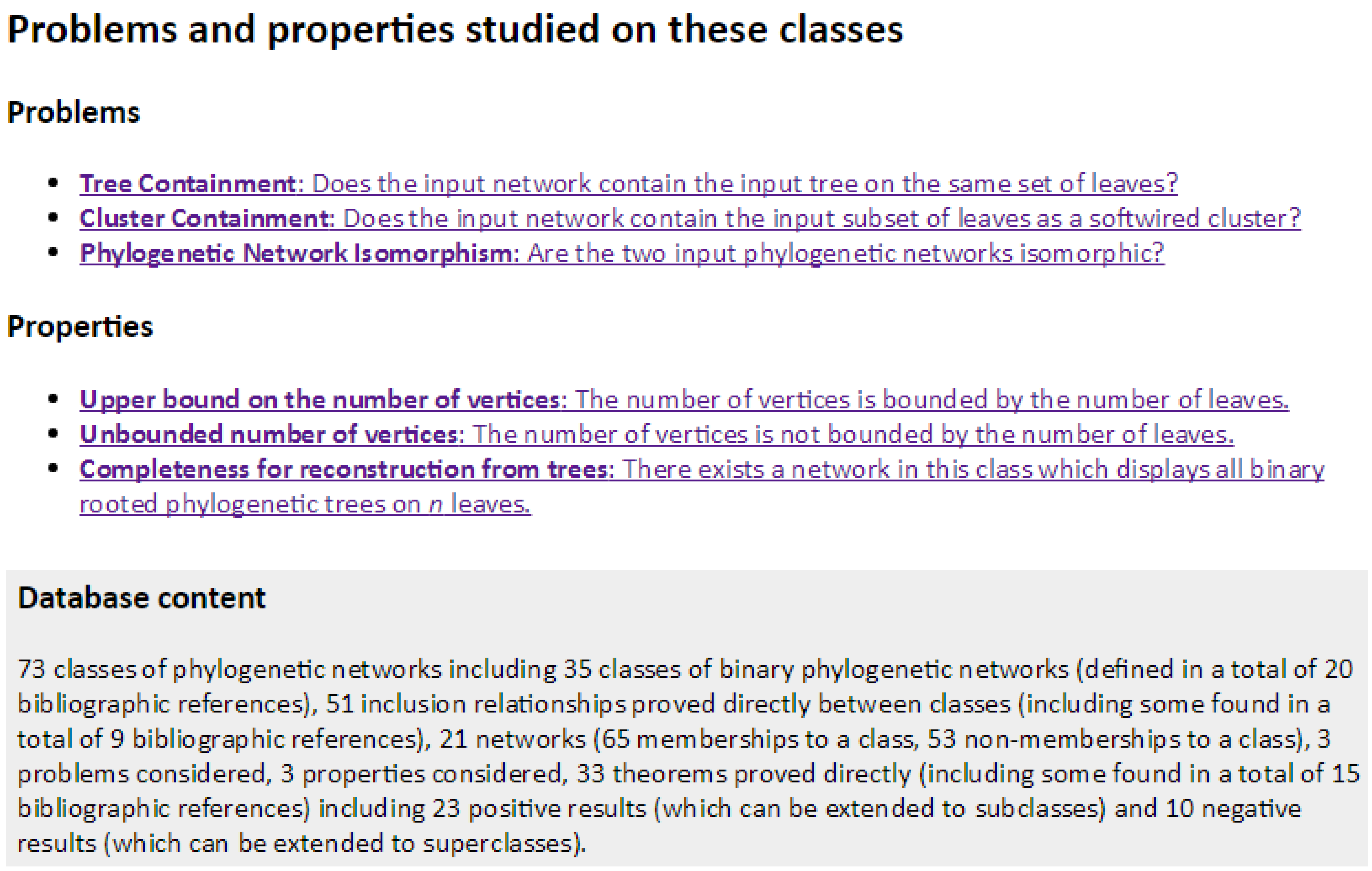
- **Upper bound on the number of vertices**: The number of vertices is bounded by the number of leaves.
- **Unbounded number of vertices**: The number of vertices is not bounded by the number of leaves.
- **Completeness for reconstruction from trees**: There exists a network in this class which displays all binary rooted phylogenetic trees on *n* leaves.

Database content

73 classes of phylogenetic networks including 35 classes of binary phylogenetic networks (defined in a total of 20 bibliographic references), 51 inclusion relationships proved directly between classes (including some found in a total of 9 bibliographic references), 21 networks (65 memberships to a class, 53 non-memberships to a class), 3 problems considered, 3 properties considered, 33 theorems proved directly (including some found in a total of 15 bibliographic references) including 23 positive results (which can be extended to subclasses) and 10 negative results (which can be extended to superclasses).

joint work with Maxime Morgado and Narges Tavassoli

Classes of phylogenetic networks



Problems and properties studied on these classes

Problems

- **Tree Containment:** Does the input network contain the input tree on the same set of leaves?
- **Cluster Containment:** Does the input network contain the input subset of leaves as a softwired cluster?
- **Phylogenetic Network Isomorphism:** Are the two input phylogenetic networks isomorphic?

Properties

- **Upper bound on the number of vertices:** The number of vertices is bounded by the number of leaves.
- **Unbounded number of vertices:** The number of vertices is not bounded by the number of leaves.
- **Completeness for reconstruction from trees:** There exists a network in this class which displays all binary rooted phylogenetic trees on *n* leaves.

Database content

73 classes of phylogenetic networks including 35 classes of binary phylogenetic networks (defined in a total of 20 bibliographic references), 51 inclusion relationships proved directly between classes (including some found in a total of 9 bibliographic references), 21 networks (65 memberships to a class, 53 non-memberships to a class), 3 problems considered, 3 properties considered, 33 theorems proved directly (including some found in a total of 15 bibliographic references) including 23 positive results (which can be extended to subclasses) and 10 negative results (which can be extended to superclasses).

Problem **easy to solve** on class A → **easy to solve** on **subclass** B
Hard to solve on class B → **hard to solve** on **superclass** A
(similar to ISGCI)

joint work with Maxime Morgado and Narges Tavassoli

Who is who in Phylogenetic Networks?

abstract-network(88) agreement-forest(29) approximation(18) APX-hard(4) ARG(6) bayesian(4) block-realization(1) bootstrap(4) bound(4) branch-and-bound(1) cactus-graph(1) characterization(11) circular-split-system(13) clustering(3) coalescent(10) consensus(8) consistency(2) cophylogeny(1) counting(3) database(1) distance-between-networks(3 diversity(5) duplication(33) dynamic-programming(10) enumeration(4) evaluation(25) explicit-network(191) exponential-algorithm(6) FPT(31) from-clusters(16) from-continuous-characters(1) from-distances(45) from-gene-order(1) from-multilabeled-tree(9) from-network(39) from-NGS-data(1) from-quartets(16) from-rooted-trees(122) from-sequences(51) from-species-tree(46) from-splits(13) from-trees(7) from-trinets(2) from-triplets(24) from-unrooted-trees(16) galled-network(7) galled-tree(38) generation(12) haplotype-network(2) haplotyping(1) heuristic(32) HMM(2) hybridization(52) inapproximability(5) integer-linear-programming(3) isomorphism(3) k-reticulated(1) kernelization(2) labeling(4) lateral-gene-transfer(56) level-k-phylogenetic-network(30) likelihood(16) lineage-sorting(10) loss(16) MASN(4) median-network(16) MedianJoining(2) Minimal-lateral-network(1) minimum-contradiction(2) minimum-number(33) minimum-spanning-network(2) model-selection(2) mu-distance(2) NeighborNet(14) nested-network(2) netting(3) normal-network(9) NP-complete(33) optimal-realization(3) parsimony(39) pedigree(1) perfect(5) phylogenetic-network(402) phylogeny(387) polynomial(70) population-genetics(5) Program-AdmixTools(1) Program-ALE(1) Program-Angst(2) Program-Arlequin(5) Program-Beagle(3) Program-BIMLR(1) Program-Bio-PhyloNetwork(4) Program-Clustistic(2) Program-CMPT(1) Program-CombineTrees(2) Program-ConsensusNetwork(1) Program-constNJ(1) Program-CycleKiller(3) Program-Dendroscope(14) Program-EEEP(3) Program-FastHN(1) Program-FlatNJ(1) Program-Fylogenetica(2) Program-GalledTree(1) Program-GraphDTL(2) Program-HapBound(1) Program-HGT_simul(1) Program-HiDe(1) Program-HorizStory(2) Program-Hybrid-Lambda(1) Program-HybridInterleave(5) Program-HybridNET(2) Program-HybridNumber(3) Program-Hybroscale(3) Program-JML(1) Program-LatTrans(5) Program-LEV1ATHAN(1) Program-Lev1Generator(1) Program-Level2(2) Program-lingpy(1) Program-LNetwork(2) Program-MaafB(2) Program-Marlon(3) Program-MC-Net(1) Program-McKiTscH(1) Program-Mowgli(5) Program-MowgliNNI(2) Program-MPNet(1) Program-MY-CLOSURE(1) Program-Nepal(7) Program-NetGen(3) Program-NetTest(1) Program-NetView(1) Program-Network(5) Program-Notung(1) Program-PADRE(7) Program-Phangorn(2) Program-PhippsNetwork(2) Program-PhyloNet(11) Program-PhyloNet-HMM(1) Program-PIRN(3) Program-Prunier(2) Program-Pyramids(3) Program-QNet(4) Program-Quartet(1) Program-Quartet-Decomposition(1) Program-QuartetMethods(1) Program-QuartetNet(2) Program-QuasiDec(1) Program-QuickCass(1) Program-RANGER-DTL(4) Program-RecMin(1) Program-Recodon(3) Program-RecPars(1) Program-Reticlad(2) Program-SAGE(2) Program-SAQ-Net(1) Program-Serial-NetEvolve(1) Program-SHRUB(3) Program-Simplistic(3) Program-Sliding-MinPD(1) Program-SNSA(2) Program-Spectronet(4) Program-SplitsTree(36) Program-SPNet(5) Program-SPRDist(1) Program-SuperQ(1) Program-T-REX(13) Program-TCS(8) Program-TERA(2) Program-TerminusEst(2) Program-TreeFix-DTL(1) Program-TreeMix(1) Program-Treevolve(2) Program-TripNet(2) Program-ultra-Net(1) Program-Ultranet(1) Program-WeakHierarchies(2) Program-Xscape(1) pyramid(8) quasi-median-network(3) realization(4) recombination(29) recombination-detection(4) reconstruction(255) regular-network(7) reticulogram(10) serial-evolutionary-networks(1) simulated-annealing(4) simulation(5) site-consistency(1) software(66) split(27) split-decomposition(13) split-network(54) SPR-distance(13) spread(2) statistical-model(32) statistical-parsimony(3) supernetwork(4) survey(31) tanglegram(1) time-consistent-network(12 tree-child-network(15) tree-sibling-network(10) tripartition-distance(9) triplet-distance(3) unicyclic-network(3) visualization(30) weak-hierarchy(8) weakly-compatible(3

input

software

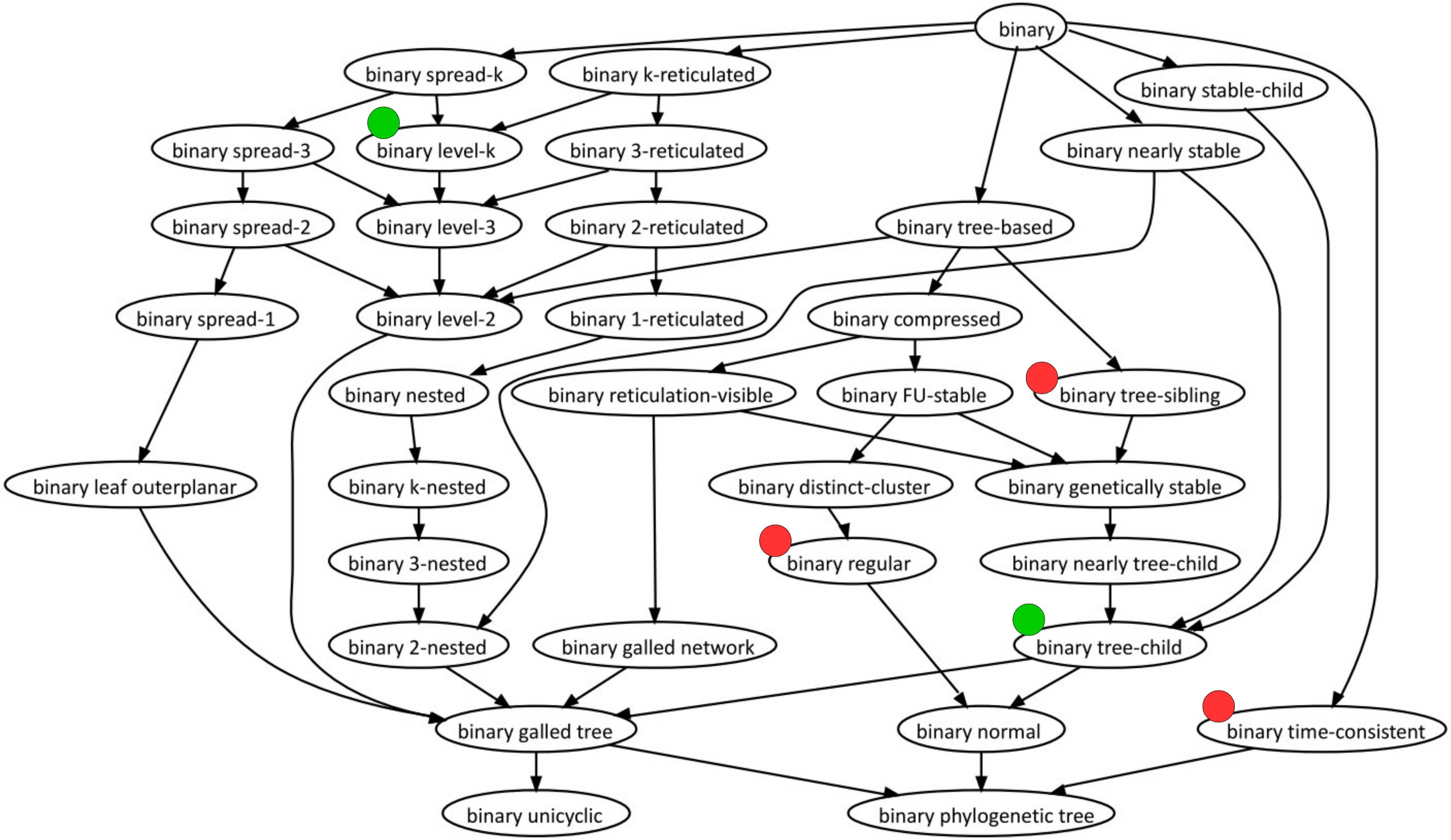
classes

problems

algorithmic properties

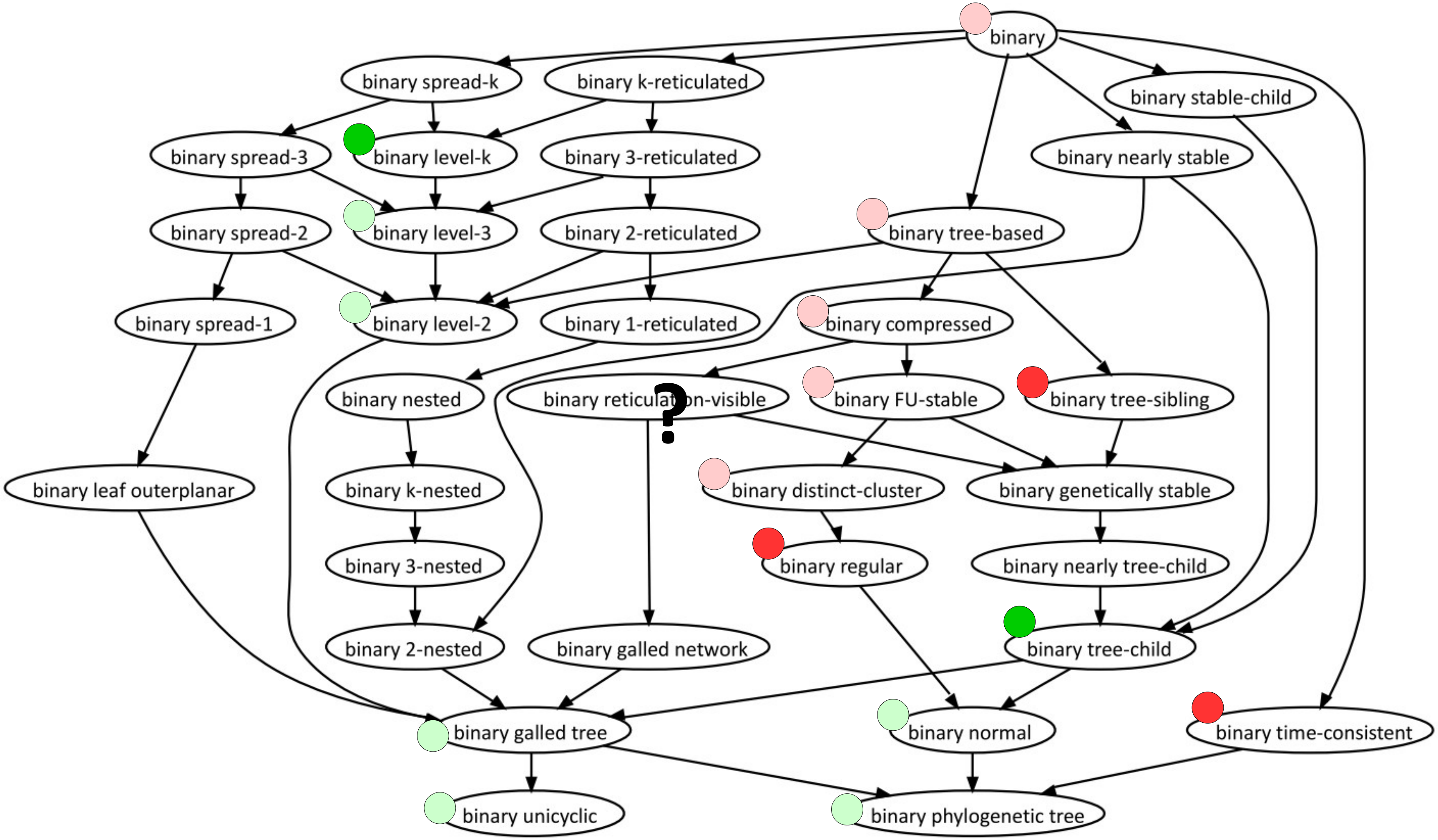
Back to the Tree Containment Problem

State of the art in 2011:

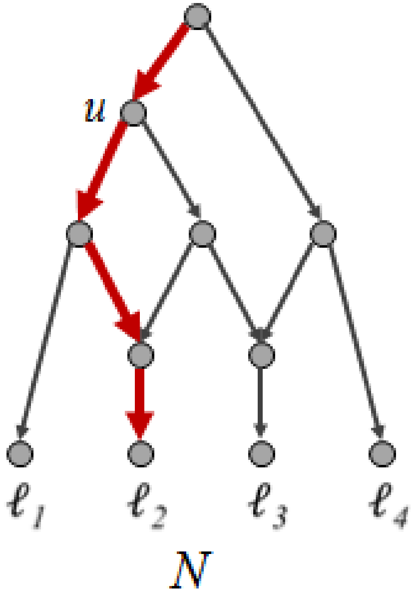


Classes of phylogenetic networks and the T.C.P.

State of the art in 2011:

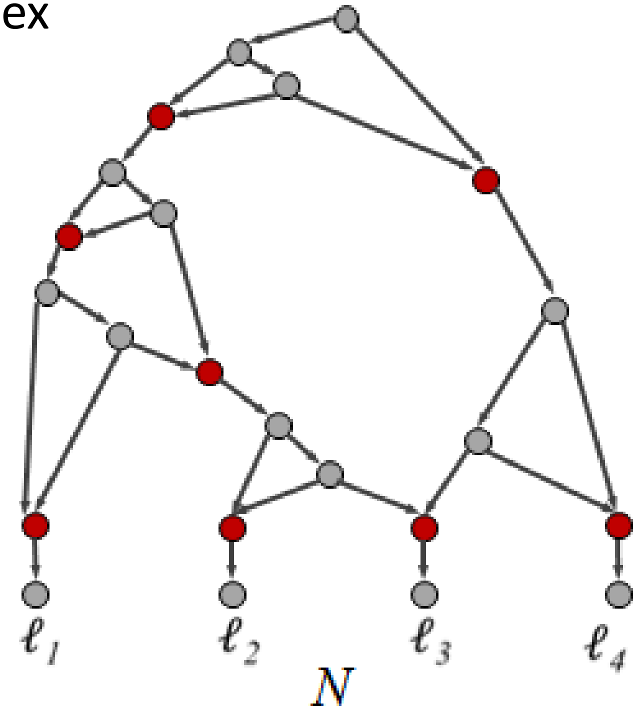


Reticulation-visible and nearly-stable networks



A vertex *u* is **stable** if there exists a leaf *l* such that all paths from the root to *l* go through *u*.

A phylogenetic network is **reticulation-visible** if every reticulation vertex is stable.

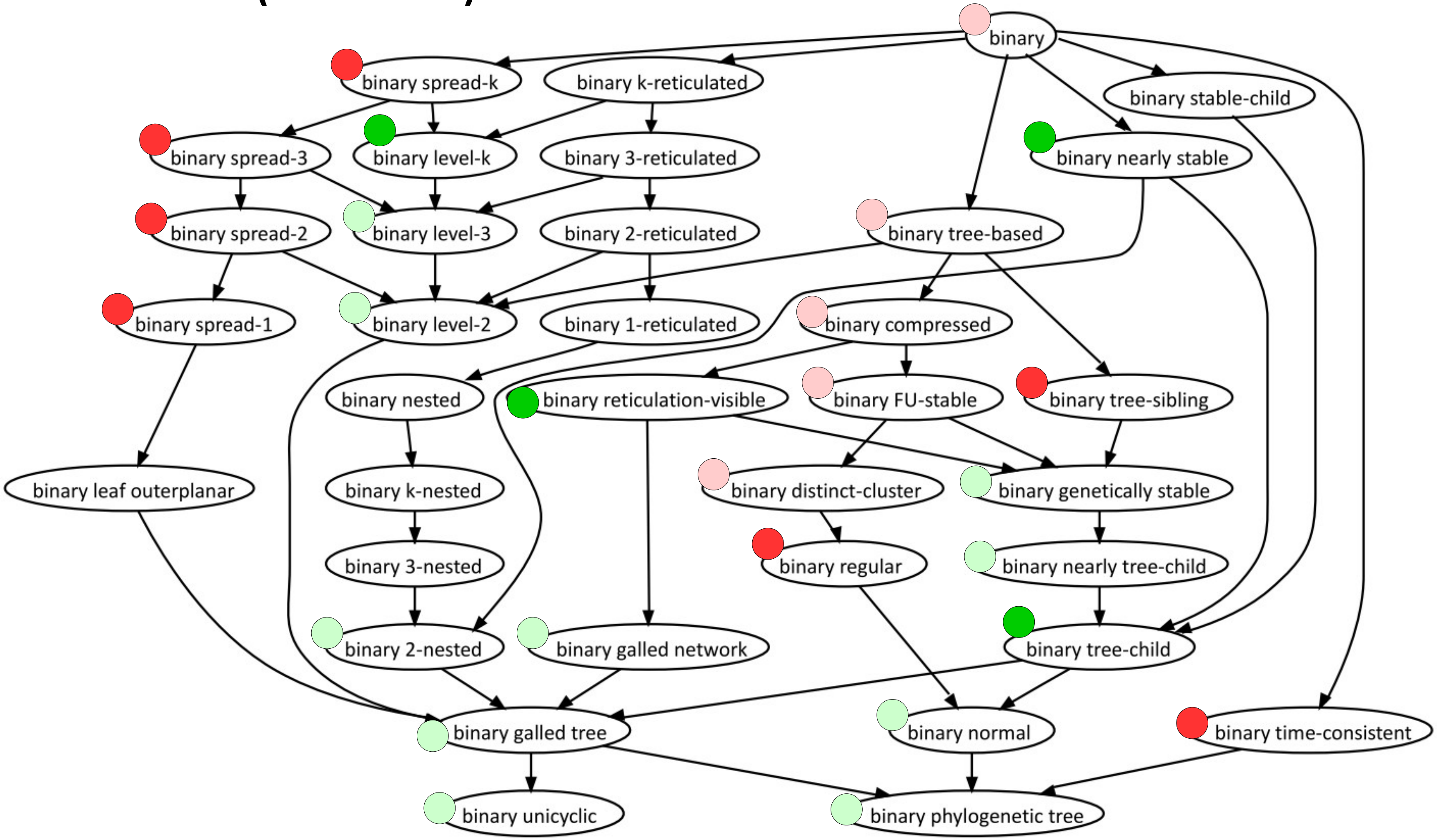


A phylogenetic network is **nearly-stable** if for each vertex, either it is stable or its parents are.

Gambette, Gunawan, Labarre, Vialette & Zhang, RECOMB 2015

Classes of phylogenetic networks and the T.C.P.

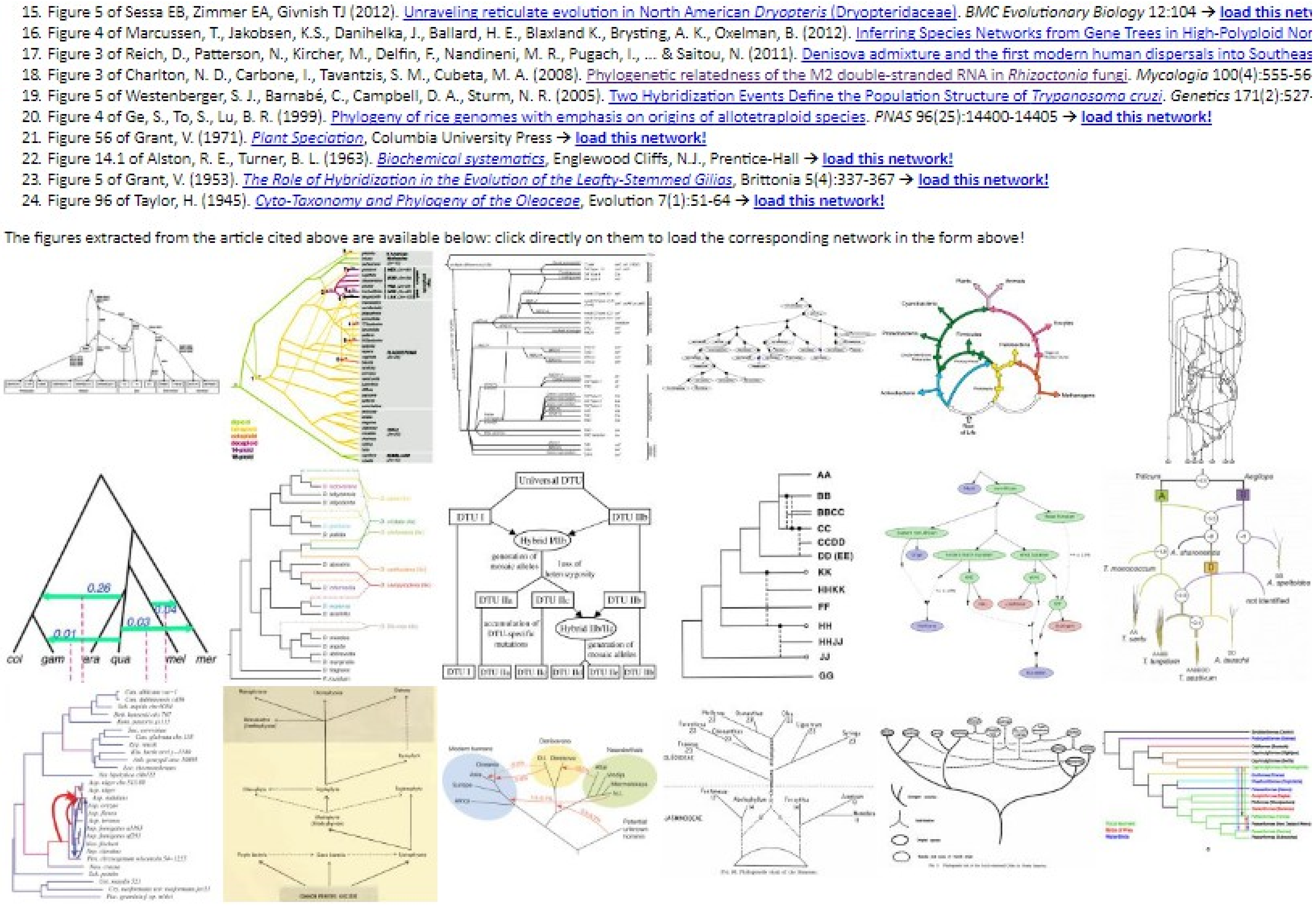
Recent results (since 2015):


binary
binary spread-k
binary k-reticulated
binary stable-child
binary spread-3
binary level-k
binary 3-reticulated
binary nearly stable
binary spread-2
binary level-3
binary 2-reticulated
binary tree-based
binary spread-1
binary level-2
binary 1-reticulated
binary compressed
binary nested
binary reticulation-visible
binary FU-stable
binary tree-sibling
binary leaf outerplanar
binary k-nested
binary distinct-cluster
binary genetically stable
binary 3-nested
binary regular
binary nearly tree-child
binary 2-nested
binary galled network
binary tree-child
binary galled tree
binary normal
binary time-consistent
binary unicyclic
binary phylogenetic tree

Fête de la Science at UPEM

Photos : Campus numérique de l'UPEM

Science académie at UPEM



15. Figure 5 of Sessa EB, Zimmer EA, Givnish TJ (2012). Unraveling reticulate evolution in North American *Dryopteris* (Dryopteridaceae). *BMC Evolutionary Biology* 12:104 → **load this net**
16. Figure 4 of Marcussen, T., Jakobsen, K.S., Danihelka, J., Ballard, H. E., Blaxland K., Brysting, A. K., Oxelman, B. (2012). Inferring Species Networks from Gene Trees in High-Polyploid Nor
17. Figure 3 of Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., ... & Saitou, N. (2011). Denisova admixture and the first modern human dispersals into Southeas
18. Figure 3 of Charlton, N. D., Carbone, I., Tavantzis, S. M., Cubeta, M. A. (2008). Phylogenetic relatedness of the M2 double-stranded RNA in *Rhizoctonia fungi*. *Mycologia* 100(4):555-56
19. Figure 5 of Westenberger, S. J., Barnabé, C., Campbell, D. A., Sturm, N. R. (2005). Two Hybridization Events Define the Population Structure of *Trypanosoma cruzi*. *Genetics* 171(2):527-
20. Figure 4 of Ge, S., To, S., Lu, B. R. (1999). Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *PNAS* 96(25):14400-14405 → **load this network!**
21. Figure 56 of Grant, V. (1971). *Plant Speciation*, Columbia University Press → **load this network!**
22. Figure 14.1 of Alston, R. E., Turner, B. L. (1963). *Biochemical systematics*, Englewood Cliffs, N.J., Prentice-Hall → **load this network!**
23. Figure 5 of Grant, V. (1953). *The Role of Hybridization in the Evolution of the Leafty-Stemmed Gilias*, Brittonia 5(4):337-367 → **load this network!**
24. Figure 96 of Taylor, H. (1945). *Cyto-Taxonomy and Phylogeny of the Oleaceae*, Evolution 7(1):51-64 → **load this network!**

The figures extracted from the article cited above are available below: click directly on them to load the corresponding network in the form above!

Stage Science Académie (association Paris Montagne) de S. Nguyen and A. Sirunyan

Outline

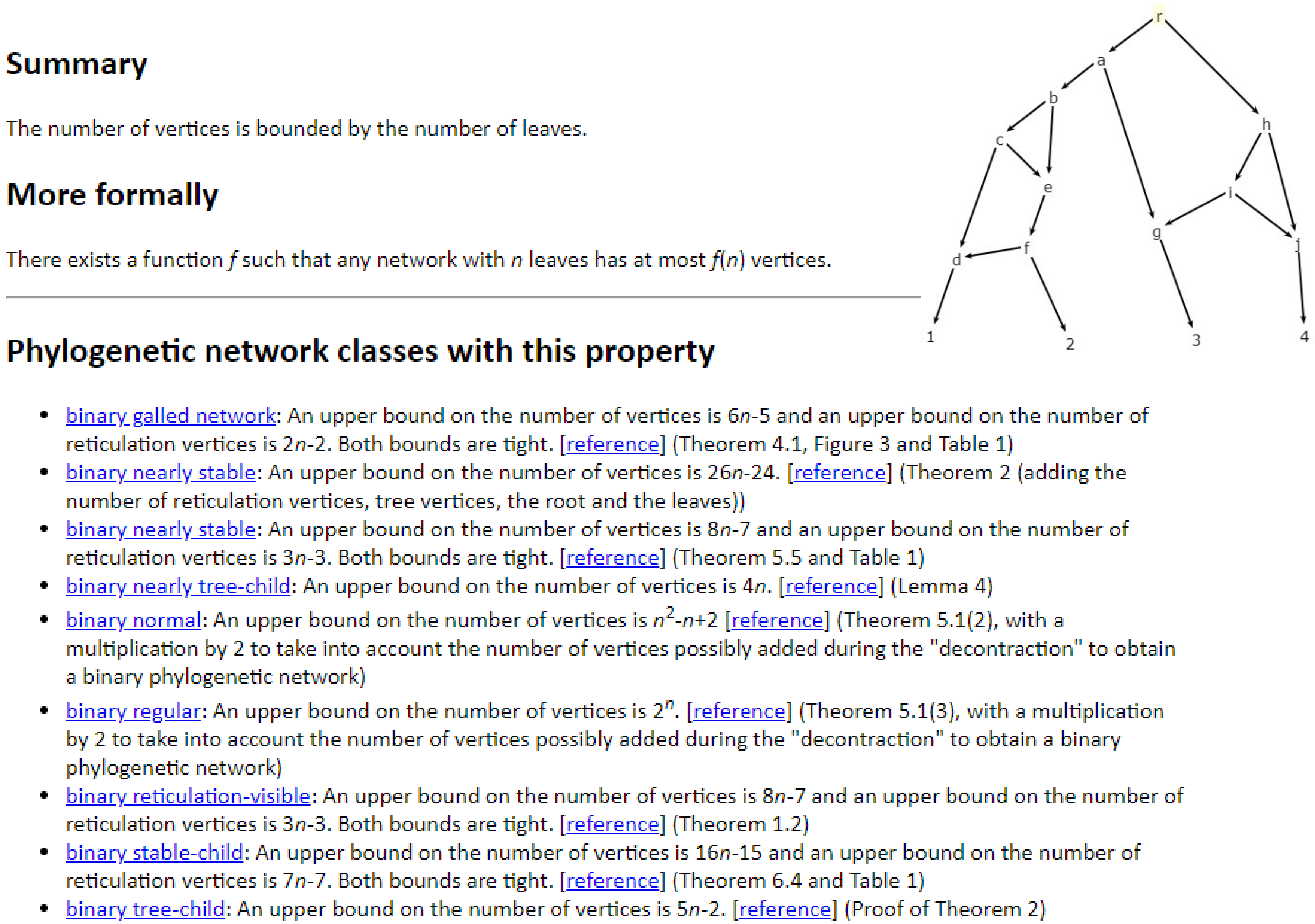
- Phylogenetic networks
- How to build phylogenetic networks?
- Classes of phylogenetic networks
- Unexpected properties of networks
- The research community on phylogenetic networks

Unexpected properties of phylogenetic networks

Properties which differ on phylogenetic trees and networks:

- number of vertices?
- lowest common ancestor?
- decide if two networks are identical?
- distance between phylogenetic networks?
- distinguish two networks from their triplet sets?

Number of vertices



Summary

The number of vertices is bounded by the number of leaves.

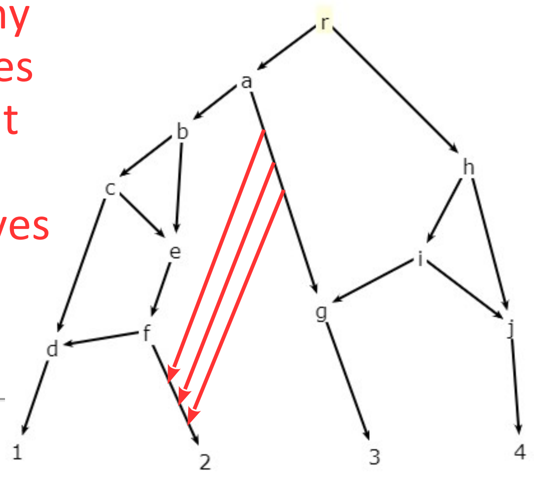
More formally

There exists a function *f* such that any network with *n* leaves has at most *f*(*n*) vertices.

Phylogenetic network classes with this property

- binary galled network: An upper bound on the number of vertices is 6n-5 and an upper bound on the number of reticulation vertices is 2n-2. Both bounds are tight. [reference] (Theorem 4.1, Figure 3 and Table 1)
- binary nearly stable: An upper bound on the number of vertices is 26n-24. [reference] (Theorem 2 (adding the number of reticulation vertices, tree vertices, the root and the leaves))
- binary nearly stable: An upper bound on the number of vertices is 8n-7 and an upper bound on the number of reticulation vertices is 3n-3. Both bounds are tight. [reference] (Theorem 5.5 and Table 1)
- binary nearly tree-child: An upper bound on the number of vertices is 4n. [reference] (Lemma 4)
- binary normal: An upper bound on the number of vertices is n^2-n+2 [reference] (Theorem 5.1(2), with a multiplication by 2 to take into account the number of vertices possibly added during the "decontraction" to obtain a binary phylogenetic network)
- binary regular: An upper bound on the number of vertices is 2^n. [reference] (Theorem 5.1(3), with a multiplication by 2 to take into account the number of vertices possibly added during the "decontraction" to obtain a binary phylogenetic network)
- binary reticulation-visible: An upper bound on the number of vertices is 8n-7 and an upper bound on the number of reticulation vertices is 3n-3. Both bounds are tight. [reference] (Theorem 1.2)
- binary stable-child: An upper bound on the number of vertices is 16n-15 and an upper bound on the number of reticulation vertices is 7n-7. Both bounds are tight. [reference] (Theorem 6.4 and Table 1)
- binary tree-child: An upper bound on the number of vertices is 5n-2. [reference] (Proof of Theorem 2)

Number of vertices



add as many arcs/vertices as you want without adding leaves

Summary

The number of vertices is bounded by the number of leaves.

More formally

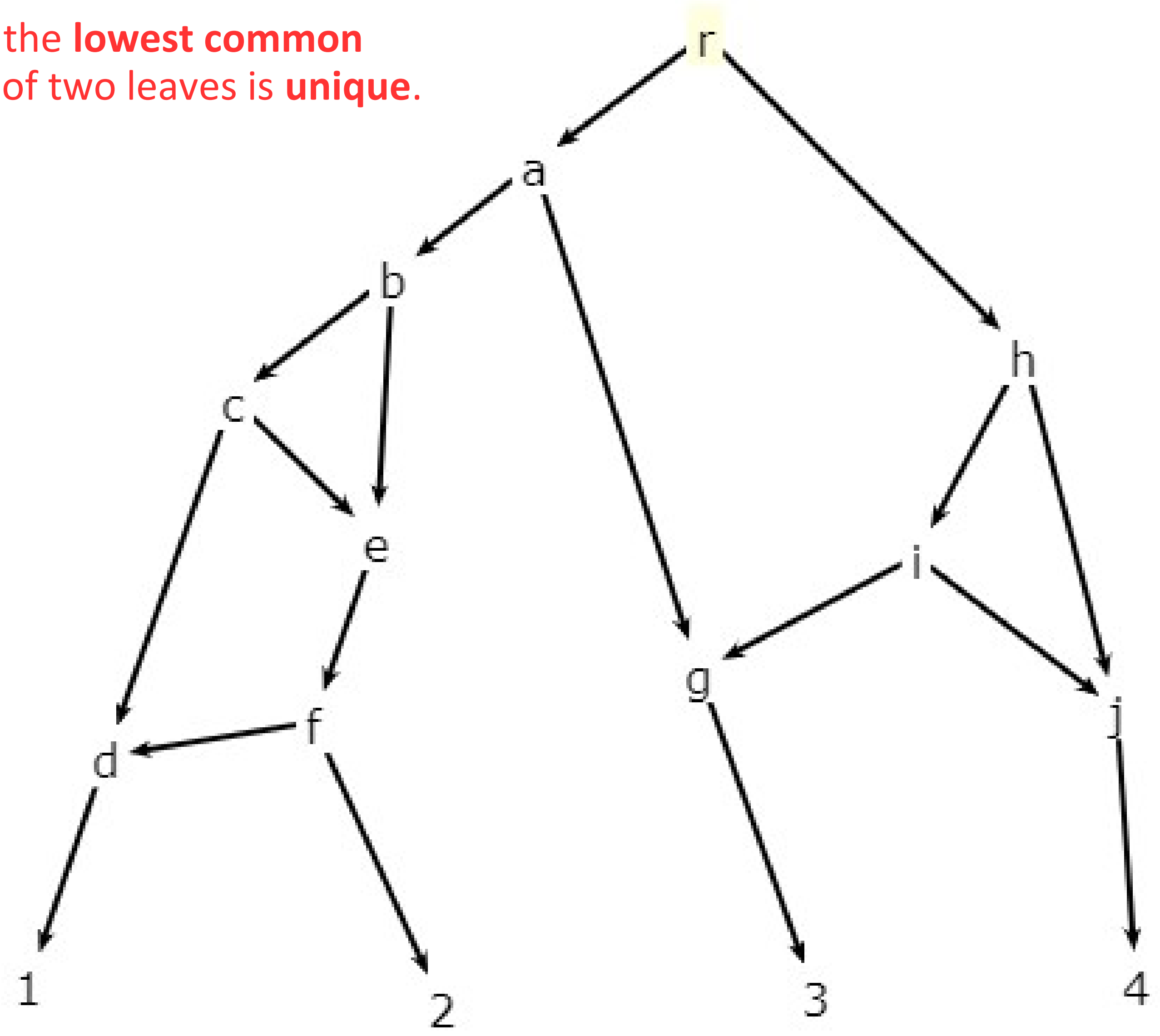
There exists a function *f* such that any network with *n* leaves has at most *f*(*n*) vertices.

Phylogenetic network classes with this property

- binary galled network: An upper bound on the number of vertices is 6n-5 and an upper bound on the number of reticulation vertices is 2n-2. Both bounds are tight. [reference] (Theorem 4.1, Figure 3 and Table 1)
- binary nearly stable: An upper bound on the number of vertices is 26n-24. [reference] (Theorem 2 (adding the number of reticulation vertices, tree vertices, the root and the leaves))
- binary nearly stable: An upper bound on the number of vertices is 8n-7 and an upper bound on the number of reticulation vertices is 3n-3. Both bounds are tight. [reference] (Theorem 5.5 and Table 1)
- binary nearly tree-child: An upper bound on the number of vertices is 4n. [reference] (Lemma 4)
- binary normal: An upper bound on the number of vertices is n^2-n+2 [reference] (Theorem 5.1(2), with a multiplication by 2 to take into account the number of vertices possibly added during the "decontraction" to obtain a binary phylogenetic network)
- binary regular: An upper bound on the number of vertices is 2^n. [reference] (Theorem 5.1(3), with a multiplication by 2 to take into account the number of vertices possibly added during the "decontraction" to obtain a binary phylogenetic network)
- binary reticulation-visible: An upper bound on the number of vertices is 8n-7 and an upper bound on the number of reticulation vertices is 3n-3. Both bounds are tight. [reference] (Theorem 1.2)
- binary stable-child: An upper bound on the number of vertices is 16n-15 and an upper bound on the number of reticulation vertices is 7n-7. Both bounds are tight. [reference] (Theorem 6.4 and Table 1)
- binary tree-child: An upper bound on the number of vertices is 5n-2. [reference] (Proof of Theorem 2)

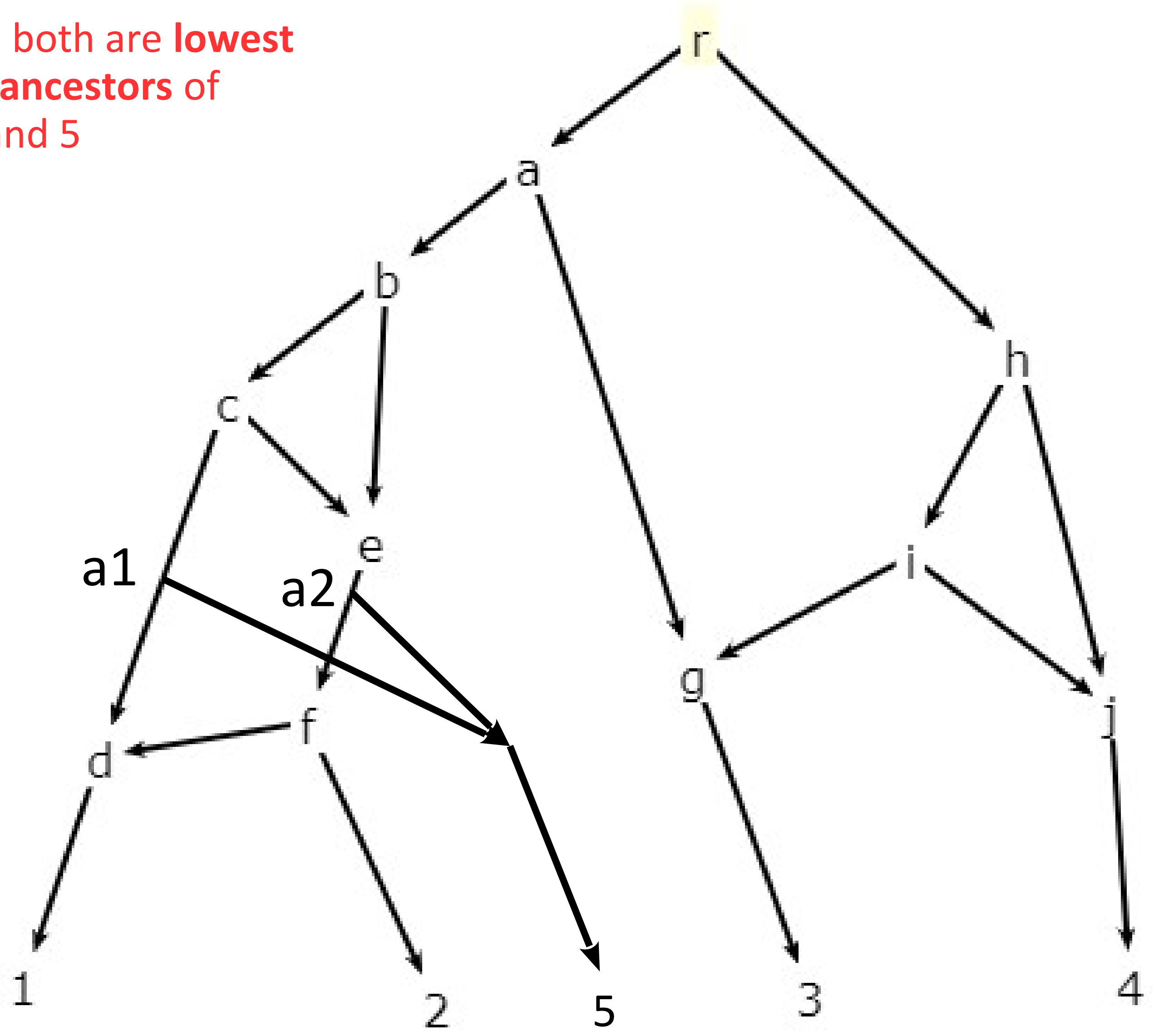
Lowest common ancestor(s)

In a tree, the **lowest common ancestor** of two leaves is **unique**.



Lowest common ancestor(s)

a1 and a2 both are **lowest common ancestors** of leaves 1 and 5



Decide if two networks are identical?

If the networks have bounded maximum degree (maximum number of children and parents for each vertex), it can be checked in polynomial time, but :

- Efficient algorithm difficult to implement

Adrià Alcalà Mena and Francesc Rosselló (2012). Ternary graph isomorphism in polynomial time, after Luks. https://arxiv.org/abs/1209.0871

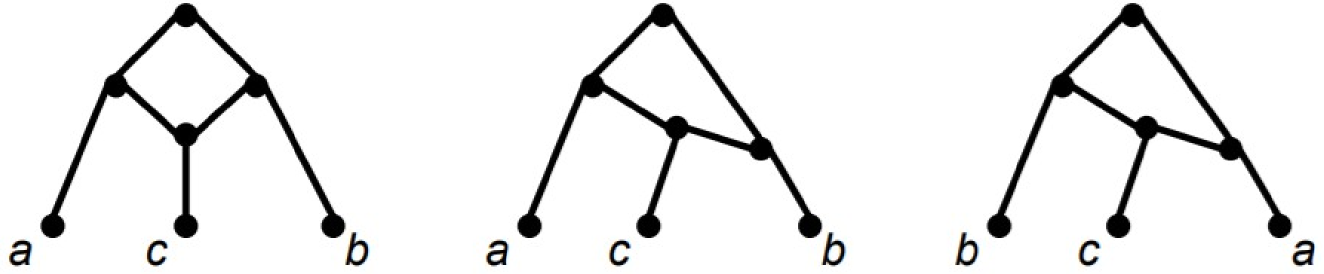
- The problem is Graph-Isomorphism complete (as hard as the graph isomorphism problem) in general

Gabriel Cardona, Mercè Llabrés, Francesc Rosselló and Gabriel Valiente (2014). The comparison of tree-sibling time consistent phylogenetic networks is graph-isomorphism complete. *The Scientific World Journal* 2014(254279):1-6

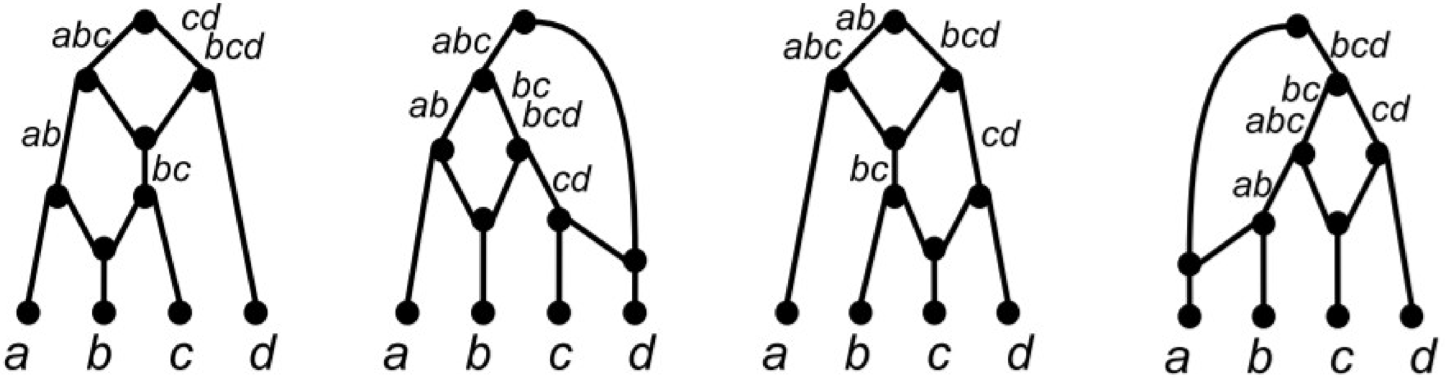
Distinguishing networks from their contained triplets

Impossible in some cases!

Same triplet sets for these 3 networks:



Same triplet sets for these 4 networks:



Philippe Gambette & Katharina Huber (2012)
On Encodings of Phylogenetic Networks of Bounded Level
Journal of Mathematical Biology 61(1):157-180

Distinguishing networks from their contained networks

Impossible in some cases!

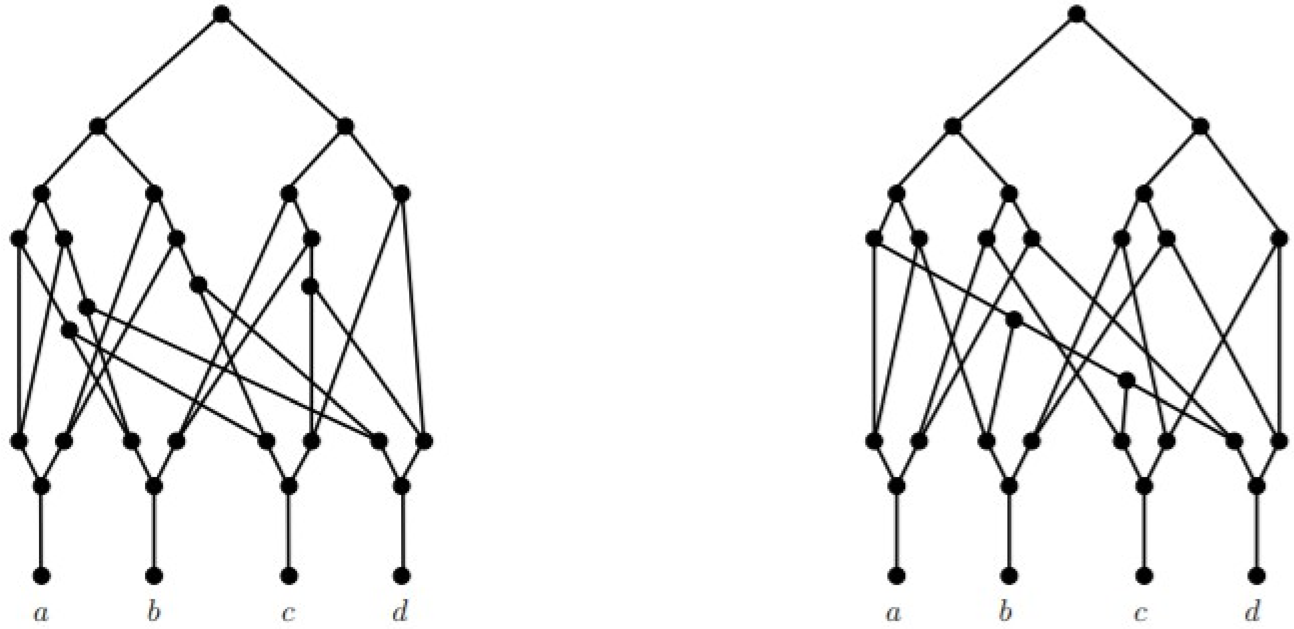


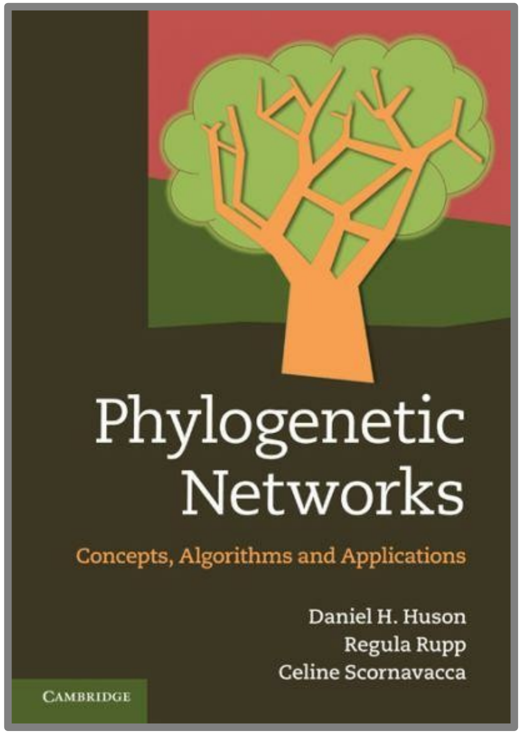
Figure: Two different networks N_1 and N_2 that display the same set of subnetworks.

Katharina Huber, Leo van Iersel, Vincent Moulton and Taoyang Wu (2015)
How much information is needed to infer reticulate evolutionary histories?
Systematic Biology, Vol. 64(1):102-111

Outline

- Phylogenetic networks
- How to build phylogenetic networks?
- Classes of phylogenetic networks
- Unexpected properties of networks
- The research community on phylogenetic networks

Books about phylogenetic networks


Phylogenetic Networks

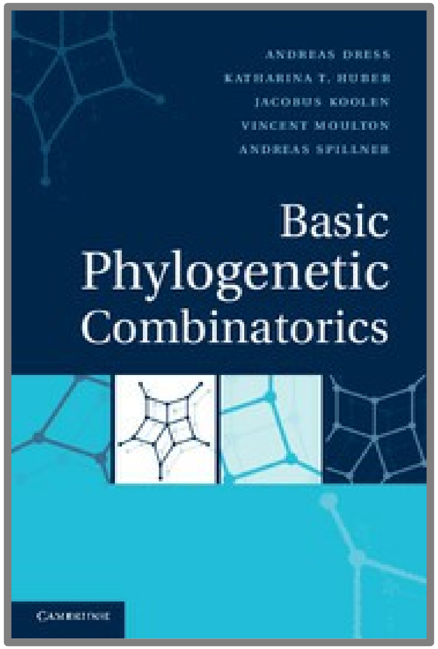
Concepts, Algorithms and Applications

Daniel H. Huson
Regula Rupp
Celine Scornavacca

CAMBRIDGE

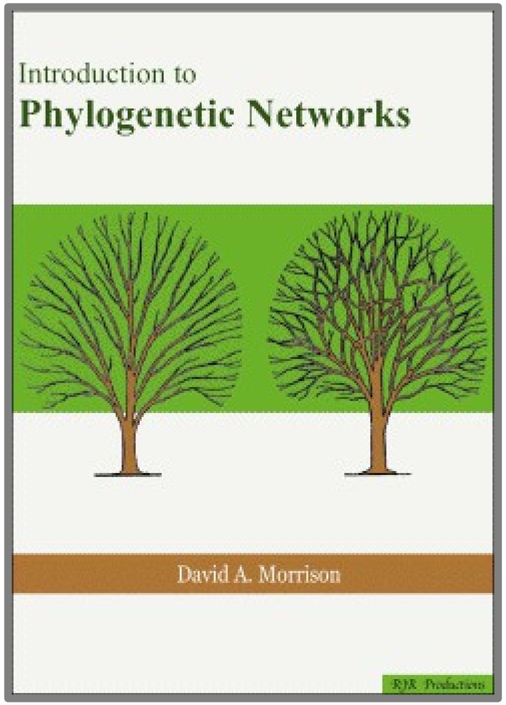
Huson, Rupp,
Scornavacca, 2011

Dress, Huber,
Koolen, Moulton,
Spillner, 2012


ANDREAS DRESS
KATHARINA T. HUBER
JACOBUS KOOLEN
VINCENT MOULTON
ANDREAS SPILLNER

Basic Phylogenetic Combinatorics

CAMBRIDGE

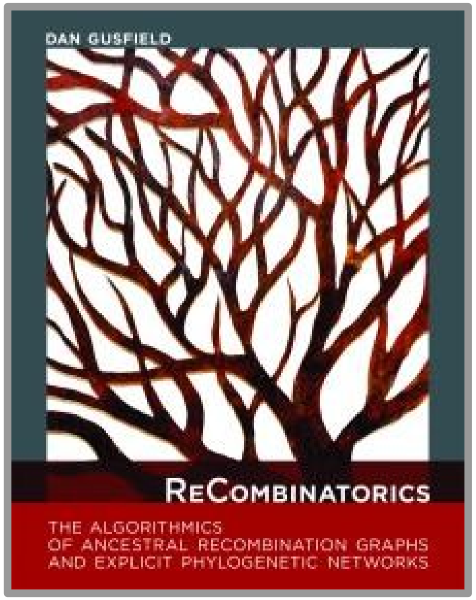

Introduction to
Phylogenetic Networks

David A. Morrison

RJR Productions

Morrison, 2011

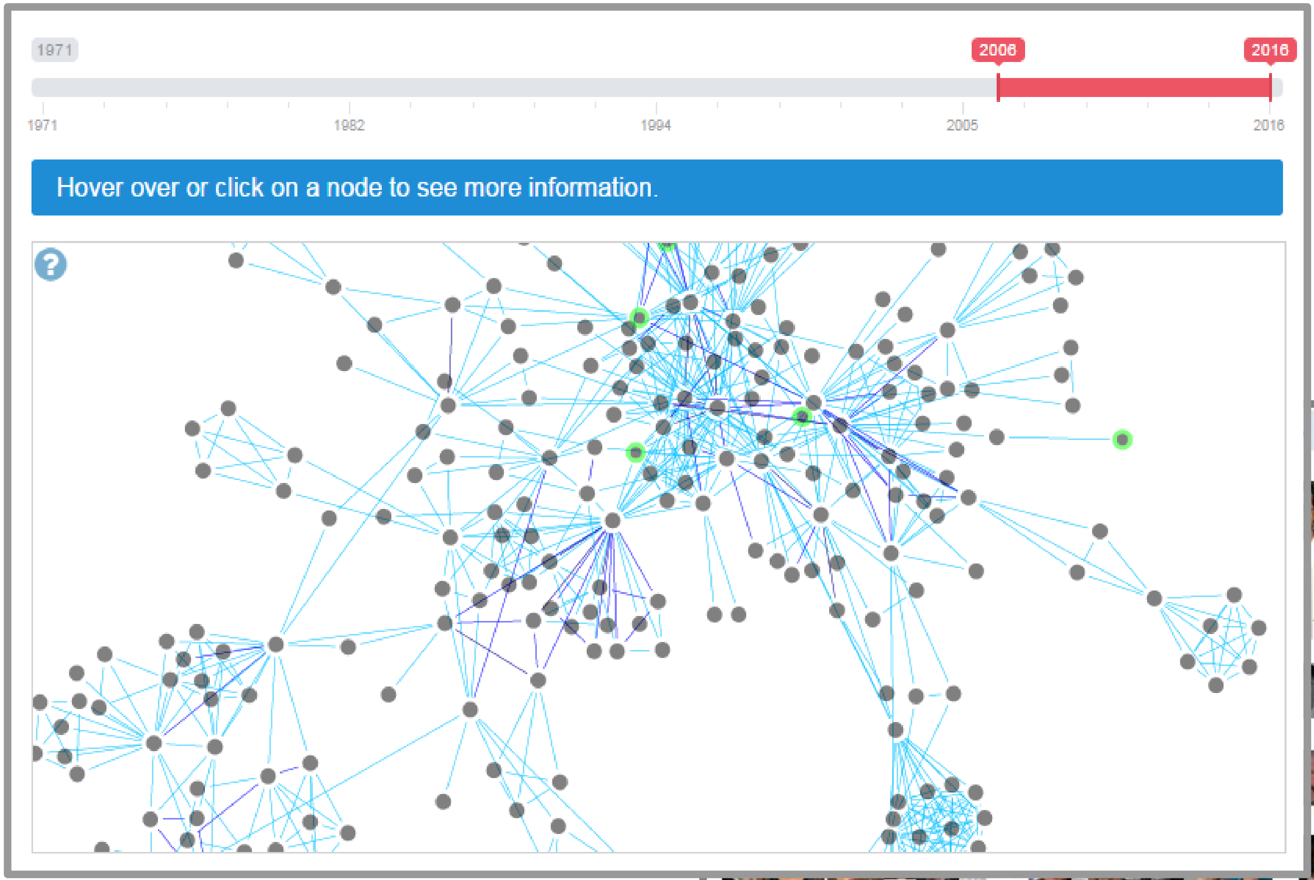
Gusfield, 2014


DAN GUSFIELD

ReCombinatorics

THE ALGORITHMICS
OF ANCESTRAL RECOMBINATION GRAPHS
AND EXPLICIT PHYLOGENETIC NETWORKS

Who is who in Phylogenetic Networks?



Hover over or click on a node to see more information.

Who is Who in Phylogenetic Networks, Articles, **Authors** & Programs

Analysis of the co-author and keyword graphs: internship of **Tushar Agarwal**

Agarwal, Gambette & Morrison, arXiv, 2017

http://phylnet.univ-mlv.fr

Workshops about phylogenetic networks

The Future of Phylogenetic Networks, 15-19 October **2012**,
Lorentz Center
(Leiden, The Netherlands)

Utilizing Genealogical Phylogenetic Networks in Evolutionary Biology: Touching the Data,
7-11 July **2014**, Lorentz Center
(Leiden, The Netherlands)

The Phylogenetic Network Workshop, 27-31 July **2015**,
Institute for Mathematical Science
(National University of Singapore)

Distinguishability in Genealogical Phylogenetic Networks,
13-17 August **2018**
(Leiden, The Netherlands)