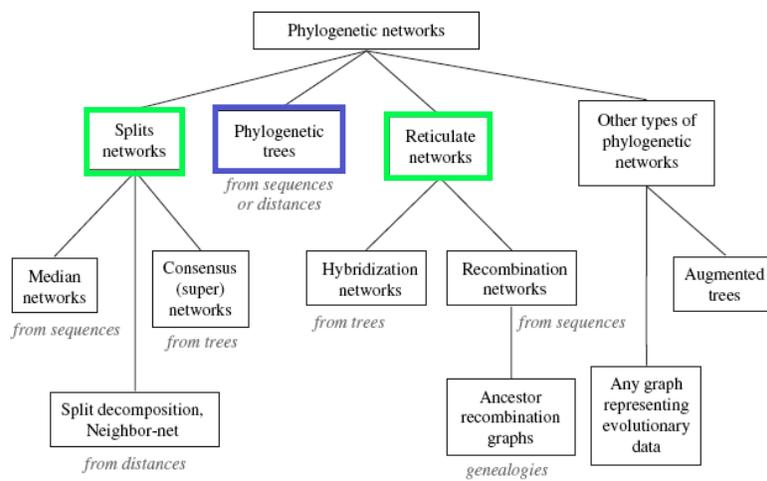# Hybridization and reticulation

Trees, Networks, and Simulations

Phylogenetic networks

Huson and Bryant (2005)

Hybridization

"any crossing of individuals of different genetic composition, *typically belonging to separate species*, resulting in hybrid offspring"

"the crossing between individuals belonging to separate populations which have different adaptive norms"

Reticulation

"…repeated intercrossing between a number of lineages producing a network of relationships"

"…recombination between genes and hybridization between lineages within a population"

Introgression

"invasion of foreign genetic material into a genome. The method of introduction is usually by sexual contact or hybridization"

Outline

Hybridization and reticulation in nature

Tree based methods to identify hybrids

Topological effects of hybridization on UPGMA, NJ, and Parsimony trees

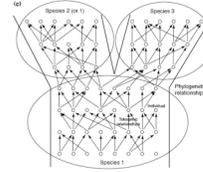Networks

Reticulograms and split networks

Empirical examples

Bees and muskrats

Simulations

Reeves and Richards (2007)

Hybridization and reticulation in nature

- Interspecific vs. intraspecific

- Lateral gene transfer in bacterial evolution
    - -incorporating genetic material from an organism without being offspring of that organism
- Hybridization between "species" of animals and plants
    - More common than we think…
    - (grasshoppers, voles, whitefish, fruit flies, chub fish, crickets, treefrogs, hares, mussels, minnows, house mice, pocket gophers, deer, cichlid fishes, chipmunks, horned lizards, earwigs, avocados, peas, oaks, willows, corn, cabbage, cotton, pines….)
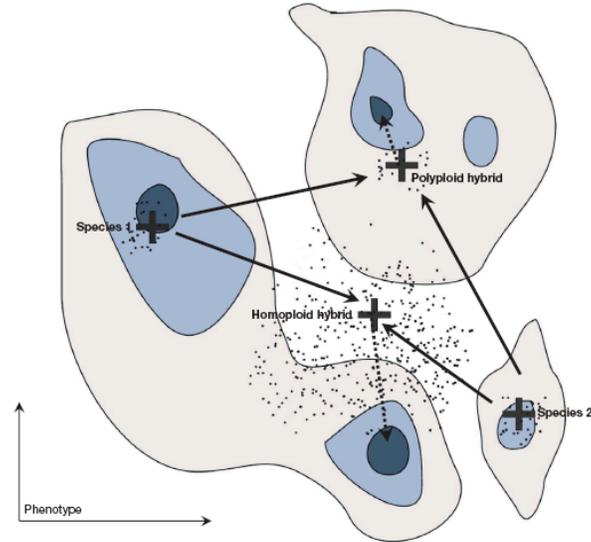

Lateral gene transfer in bacteria
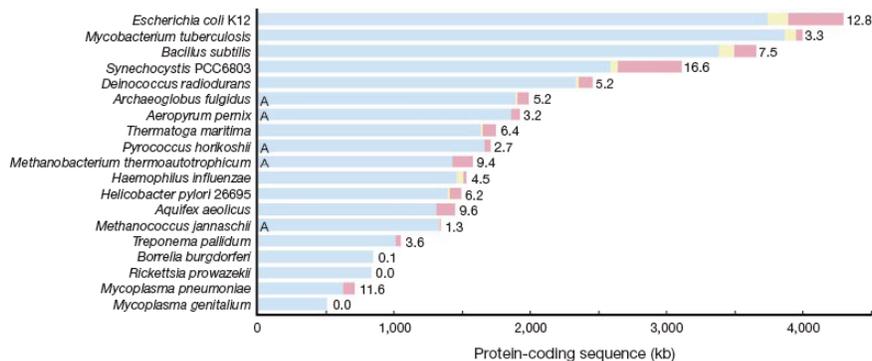
-percentage of foreign DNA in red
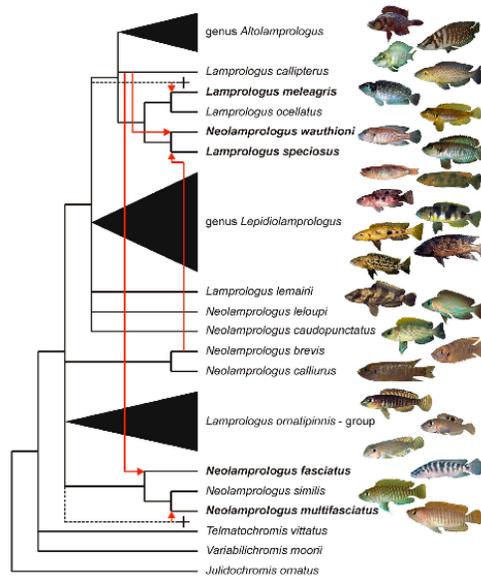
Ochman et al. (2000)

Hybrid speciation

(Mallet 2007)

Reticulate phylogeny of cichlids

-Hybrid species in bold

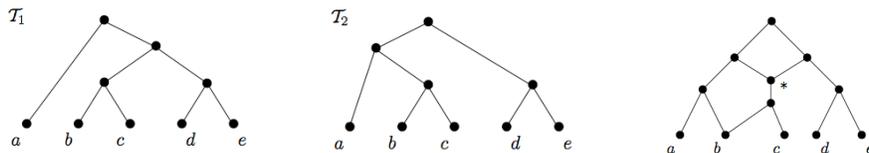-Red lines indicate mitochondrial introgression

Koblmuller et al. (2007)
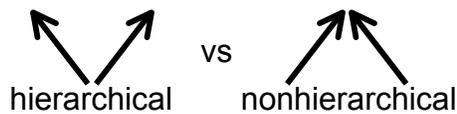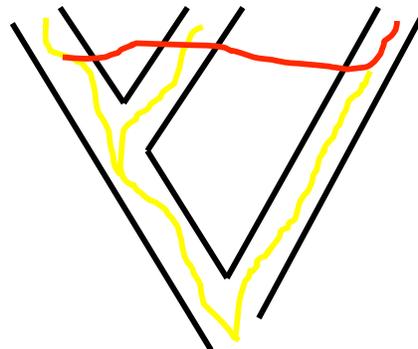
Hybrids and tree based methods

"Even when evolution proceeds in a tree-like manner, analysis of the data may not be best served by forcing the data onto a tree or assuming a tree like model."

Huson and Bryant (2006)

Hybrids and tree based methods

- Assumptions underlying tree based methods that produce hierarchical and bifurcating trees are violated in the presence of hybridization

hierarchical   vs   nonhierarchical

Hybrids and tree based methods

- Parsimony: Hybrids can decrease consistency index (CI) and increase the number of parsimonious trees.

  CI = m/s, (s = observed number of changes and m = minimum possible number of changes). Used to measure how well data fits a tree.
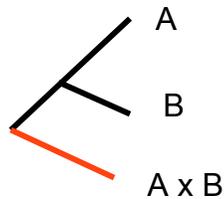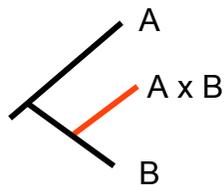
  - No homoplasy and divergently evolving taxa = high CI

McDade (1992)

Hybrids and tree based methods

A

A x B

B
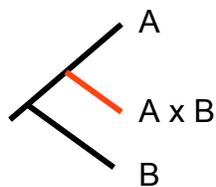
UPGMA: Hybrid between closely related parents jumps between them as sister in cladograms of equal length

A

A x B

B

A

B

A x B

NJ: Hybrid appears basal to parents

McDade (1992)

Hybrids and tree based methods

A

A x B

B
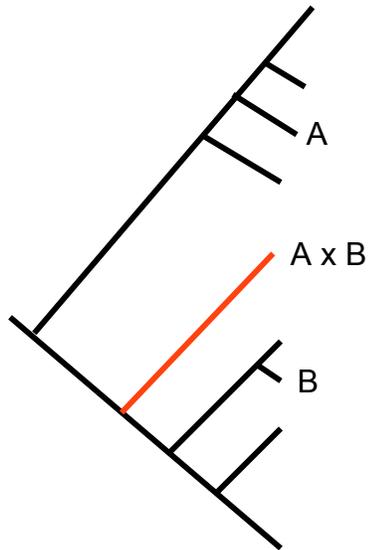
Parsimony: Hybrid between distantly related parents appears basal to the most derived parent

---

Hybrids and trees: what about time?

Topological effect of unidirectional gene flow on parsimony analysis

Reeves and Richards (2007)

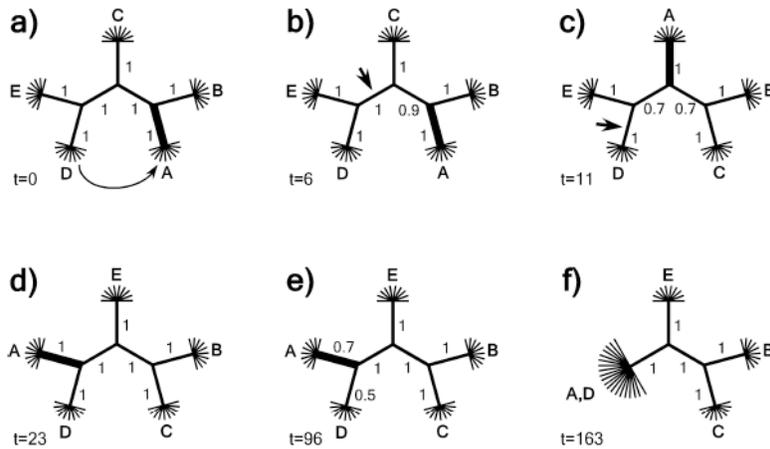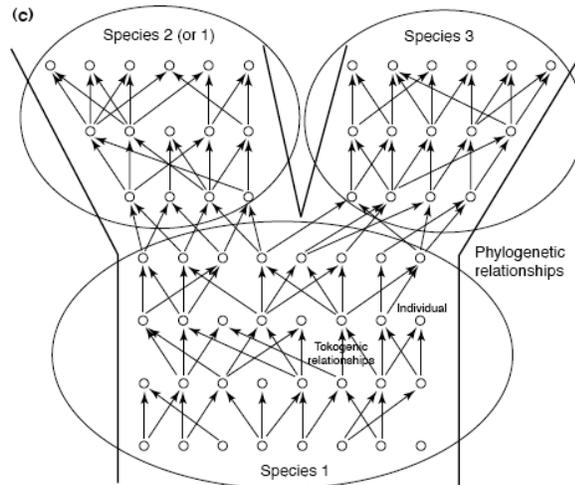# Hybrids and tree based methods

What about intraspecific processes?

Tokogenetic relationships: nonhierarchical genetic relationships among individuals. Arising by sexual reproduction.

Networks

- Useful to account for:
  - Lateral gene transfer in bacteria
  - Hybridization between species
  - Microevolution of local populations within a species
  - Homoplasy due to convergence
  - Host parasite relationships
  - Viral recombination
  - Vicariance and dispersal biogeography

Types of Networks

- Pyramids (c)
- Statistical geometry (d)
- Split decomposition (e)
- Minimum spanning (f)
- Median-joining (g)
- Statistical parsimony (h)
- Molecular variance parsimony
- Netting
- Likelihood network
- Reticulogram (i)
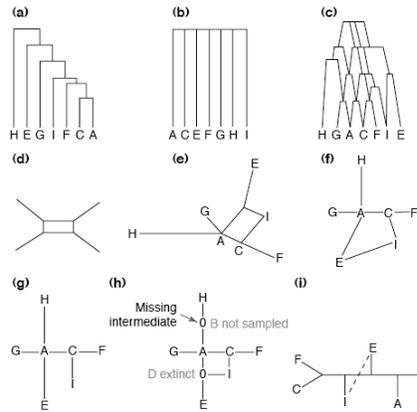
Reticulate Networks
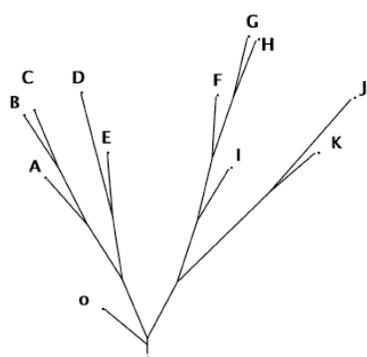
Phylogenetic tree   Reticulate network

Huson and Bryant (2005)

Reticulograms

Reticulogram: reticulated cladogram

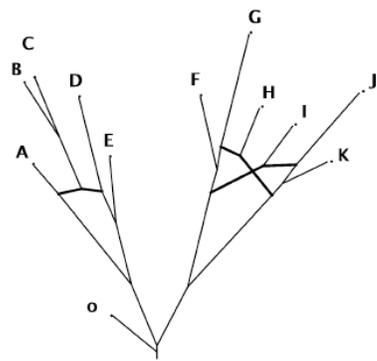-"Distance between a pair of nodes in a reticulogram is the minimum path-length distance **over the set of all paths linking them**."

Legendre and Makarenkov (2002)
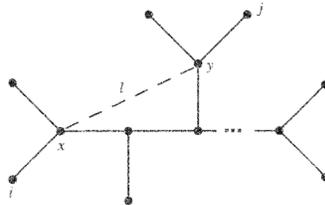
Reticulograms

•Reticulograms…Legendre and Makarenkov (2002)
  –"Reticulate evolution refers to evolutionary processes that cannot be fully represented by the tree model."

Phylogenetic tree   Reticulogram

How to make a reticulogram

- 1) phylogenetic tree inferred from distance matrix

- 2) reticulation branches are added to the tree (with least squares, parsimony, or maximum likelihood criteria to optimize)

- 3) addition of reticulation branches stops when minimum of a goodness of fit function is reached

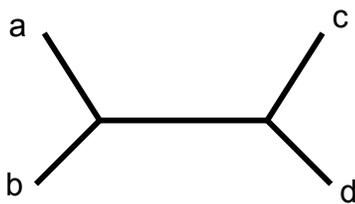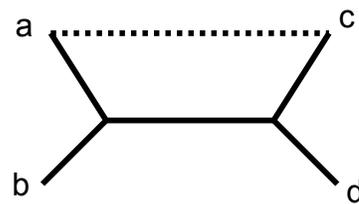# An example from bees (*Apis*)

A. andreniformis	0				
A. mellifera	0.090	0			
A. dorsata	0.103	0.093	0		
A. cerana	0.096	0.090	0.117	0	
A. florea	0.004	0.093	0.106	0.099	0
A. koschevnikovi	0.075	0.100	0.103	0.099	0.078	0

NJ   *mellifera*   *cerana*

        dorsata

                                florea

        koschevnikovi   *andreniformis*

Makarenkov et al. 2004

Reticulograms

mellifera   *cerana*

57%

dorsata

88%

florea

koschevnikovi   *andreniformis*

mellifera   *cerana*

dorsata

florea

koschevnikovi   *andreniformis*

Reticulograms

Reticulation branch lengths indicate homoplasy (convergent evolution) and reveal genetic relationships absent in the phylogenetic tree

A. mellifera   *A. cerana*

Ancestor   Ancestor

mellifera   *cerana*

dorsata

florea

koschevnikovi   *andreniformis*

Makarenkov et al. (2004)

An example from muskrats (*Ondatra*)

-Morphology of muskrats in Belgium

-M and Z separated by short swampy areas

-Small creeks from C/E to M

-Headwaters at N and O but separated by forest
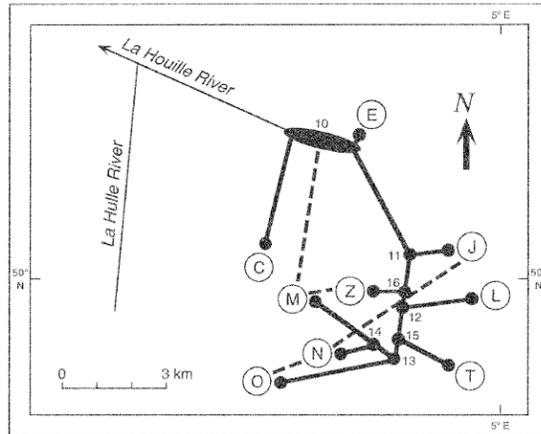
Reticulogram

Le Boulenge et al. (1996); Legendre and Makarenkov (2002)
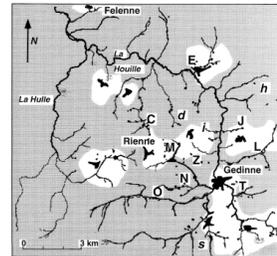
Reticulograms

Useful for:

    -uncovering patterns due to reticulate evolutionary histories

    -representing relationships among organisms that may have more than one path

Split networks

Phylogenetic tree   Split network
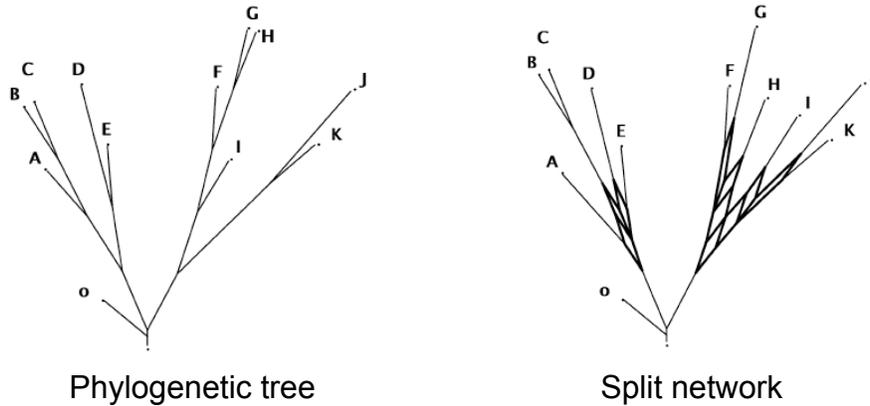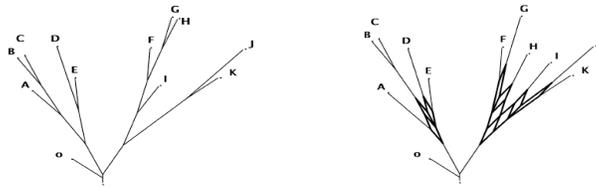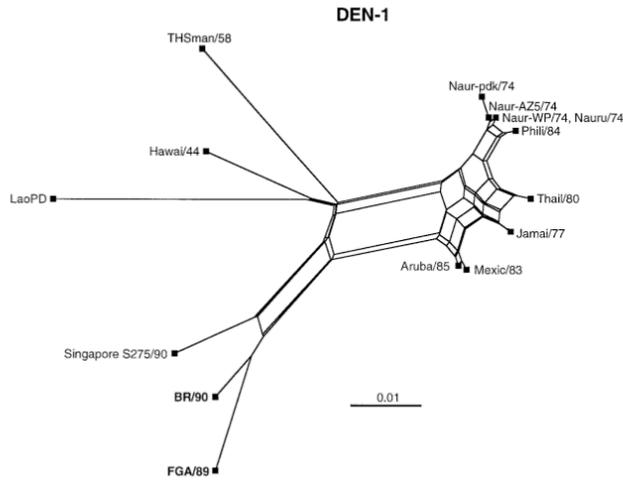
Huson and Bryant (2005)

Split networks

-Used to represent incompatible and ambiguous signals in a data set. Parallel edges, rather than branches, are used to represent splits computed from the data.

-Nodes in a split network do not represent ancestral species. They are an *implicit* rather than *explicit* representation of evolution.
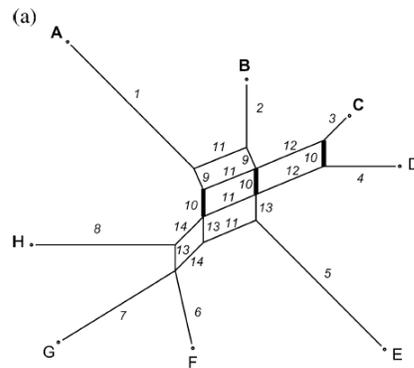
Split networks

DEN-1

THSman/58

Naur-pdk/74
Naur-AZ5/74
Naur-WP/74, Nauru/74
Phili/84

Hawai/44

LaoPD

Thail/80

Jamai/77

Aruba/85 Mexic/83

Singapore S275/90

0.01

BR/90

FGA/89

Viral Studies

-Dengue virus sequences

-Multiple pathways provide evidence of recombination

Holmes et al. (1999)

Split networks

	A	B	C	D	E	F	G	H	weights
1	●	○	○	○	○	○	○	○	7.92
2	○	●	○	○	○	○	○	○	3.31
3	○	○	●	○	○	○	○	○	1.74
4	○	○	○	●	○	○	○	○	3.72
5	○	○	○	○	●	○	○	○	8.94
6	○	○	○	○	○	●	○	○	3.88
7	○	○	○	○	○	○	●	○	5.63
8	○	○	○	○	○	○	○	●	6.21
9	●	●	○	○	○	○	○	○	1.12
10	●	●	●	○	○	○	○	○	1.28
11	○	●	●	●	●	○	○	○	2.83
12	○		●	●	○	○	○	○	3.63
13	○	○	○	○	●	●	●	○	1.28
14	○	○	○	○	○	●	●	●	1.95

(a)

A
B
C
D
E
F
G
H

1 2 3 4 5 6 7 8
9 10 11 12 13 14

- HIV strains from a single patient (data from Shankarappa et al. 1999)

Split networks

Useful for:
- representing incompatible and ambiguous signals in a data set that *potentially* arise from reticulation

Networks

Advantages
- interspecific hybridization and reticulation
- account for intraspecific processes
- population genetics theory

Disadvantages
- rooting network is difficult
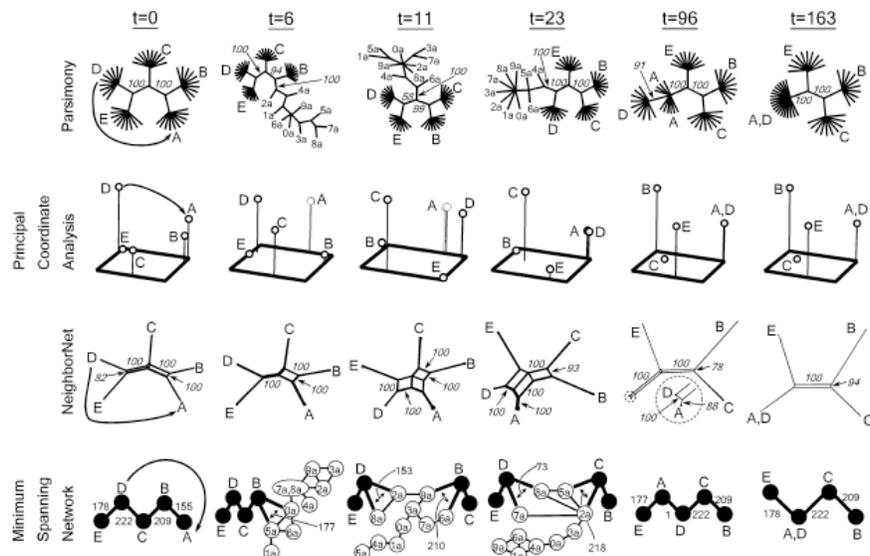- subject to Type I and Type II errors….
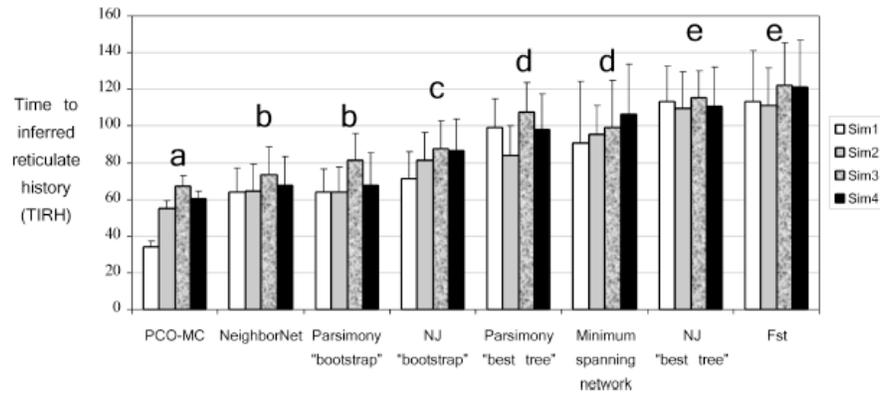
Simulations

Reeves and Richards (2007)

-Simulated gene flow between taxa (200 hybridization events per data set)

-Analyzed data sets using parsimony, neighbor-joining, minimum spanning networks, split networks, principal coordinate analysis
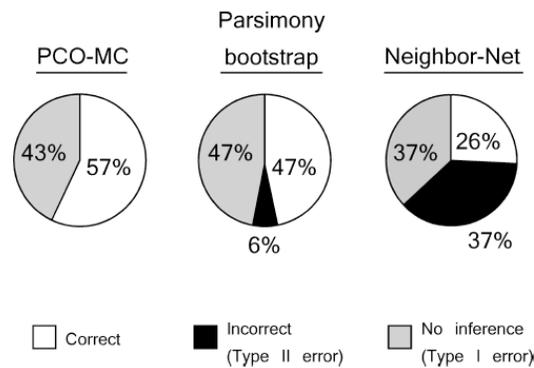
Simulations

Simulations

Performance of analytical procedures to infer reticulate history

Simulations: Type I and II errors

Type I error: finding a taxon to be monophyletic even when it has substantial reticulate history

Type II error: finding that clearly monophyletic groups are a product of reticulation

Conclusions

- Bifurcating tree based methods more appropriate for studying higher level relationships than for lower level studies (hybridization and subsequent reticulation histories)

- Network methods may identify hybridization and reticulation however may be subject to Type I and II errors