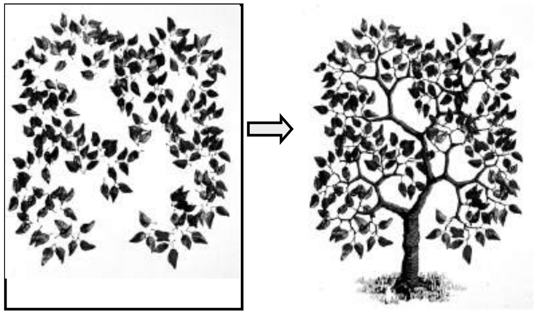
Phylogenetics

Phylogenetics is the estimation of the
“tree” through “time”

Phylogenetics

Phylogenetics is the estimation of the
“tree” through “time” knowing only the “leaves”



Phylogenetics

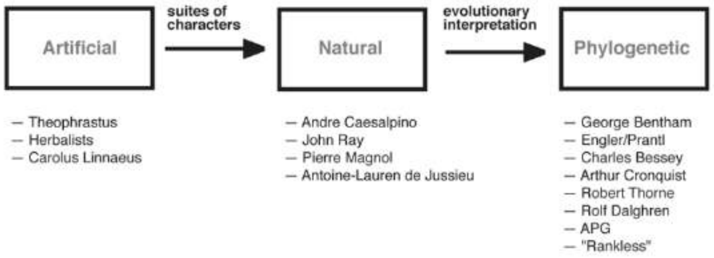
However, the “leaves” are scattered over “space”. Some areas have related “leaves”, others have unrelated “leaves”. Thus, phylogenetics is compounded by issues of both “time” and “space”.

Phylogenetics

Additionally, many related “leaves” diverge in “form”, while other unrelated “leaves” converge in “form”. Thus, phylogenetics is compounded by issues of “time” and “space” and “form”.

Phylogenetics

In natural and phylogenetic systems of classification, characters are selected *a posteriori* for their value in correlating with other characters to form hierarchical structure of groups



Artificial → suites of characters → Natural → evolutionary interpretation → Phylogenetic

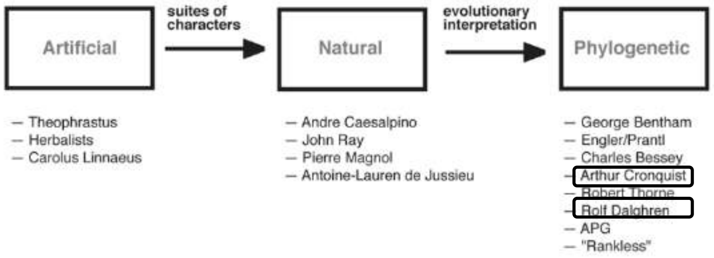
Artificial:
- Theophrastus
- Herbalists
- Carolus Linnaeus

Natural:
- Andre Caesalpino
- John Ray
- Pierre Magnol
- Antoine-Lauren de Jussieu

Phylogenetic:
- George Bentham
- Engler/Prantl
- Charles Bessey
- Arthur Cronquist
- Robert Thorne
- Rolf Dalghren
- APG
- "Rankless"

Phylogenetics

What characters are selected or even considered, has been very subjective. Consider Cronquist and Dalghren with mustard oil families . . .



Artificial → suites of characters → Natural → evolutionary interpretation → Phylogenetic

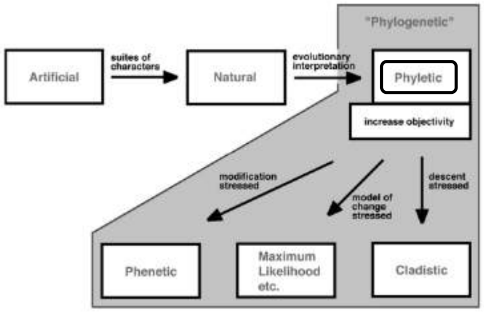
Artificial:
- Theophrastus
- Herbalists
- Carolus Linnaeus

Natural:
- Andre Caesalpino
- John Ray
- Pierre Magnol
- Antoine-Lauren de Jussieu

Phylogenetic:
- George Bentham
- Engler/Prantl
- Charles Bessey
- Arthur Cronquist
- Robert Thorne
- Rolf Dalghren
- APG
- "Rankless"

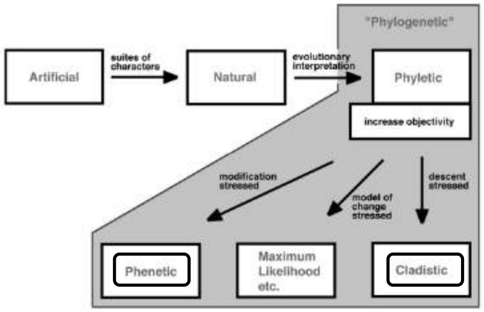
Phylogenetics

These first phylogenetic classifications were “phyletic” - involving a subjective selection of characters for classification



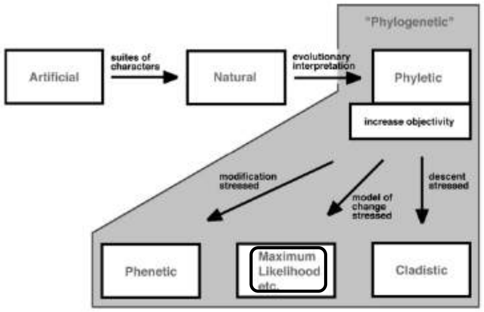
Phylogenetics

In the 1960s, two main groups of systematists became dissatisfied with the phyletic approach and developed more objective methods: phenetic and cladistic



Phylogenetics

With the rise of molecular phylogenetics in the 1980s, additional approaches are now invoked (ML, Bayesian) - a continuum of models are now seen



Phenetics vs. Cladistics

- Phenetics uses "overall similarity" - all characters used ("distance" approaches)
- species similarity (or differences) often scaled from 0 to 1

- Cladistics uses only "phylogenetically informative" characters
- derived state is shared by at least 2 but not all taxa - "shared derived character states"

Phenetics

Data Matrix
taxa
characters
states

	vessels	apocarpy	sympetaly	epipetaly	trees	epigyny	beetle poll.	tepals	bicarpellate	heterostyly
1. ***Magnolia***	+	+	-	-	+	-	+	+	-	-
2. ***Nymphaea***	-	+	-	-	-	-	+	+	-	-
3. ***Rosa***	+	+	-	-	-	-	-	-	-	-
4. ***Primula***	+	-	+	+	-	-	-	-	-	+
5. ***Gentiana***	+	-	+	+	-	-	-	-	+	-
6. ***Aster***	+	-	+	+	-	+	-	-	+	-

Phenetics

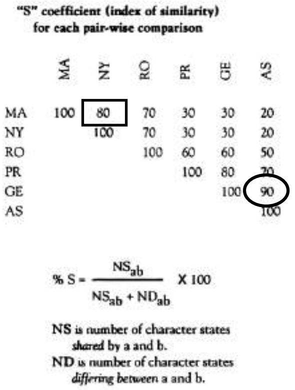
Data Matrix

UPGMA cluster analysis

- convert data matrix into pair-wise matrix based on overall similarity

1. ***Magnolia***	+	+	-	-	+	-	+	+	-	-
2. ***Nymphaea***	-	+	-	-	-	-	+	+	-	-
3. ***Rosa***	+	+	-	-	-	-	-	-	-	-
4. ***Primula***	+	-	+	+	-	-	-	-	-	+
5. ***Gentiana***	+	-	+	+	-	-	-	-	+	-
6. ***Aster***	+	-	+	+	-	+	-	-	+	-

Phenetics



"S" coefficient (index of similarity)
for each pair-wise comparison

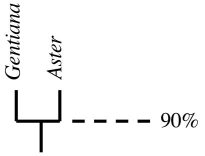
	MA	NY	RO	PR	GE	AS
MA	100	80	70	30	30	20
NY		100	70	30	30	20
RO			100	60	60	50
PR				100	80	70
GE					100	90
AS						100

$$\% S = \frac{NS_{ab}}{NS_{ab} + ND_{ab}} \times 100$$

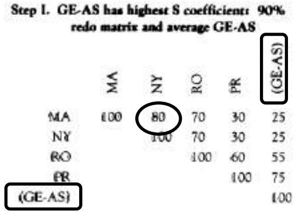
NS is number of character states *shared* by a and b.
ND is number of character states *differing* between a and b.

UPGMA cluster analysis

- convert data matrix into pair-wise matrix based on overall similarity
- identify most similar pair of taxa and cluster them



Phenetics

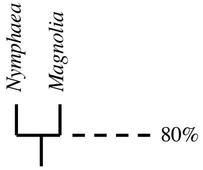


Step I. GE-AS has highest S coefficient: 90%
redo matrix and average GE-AS

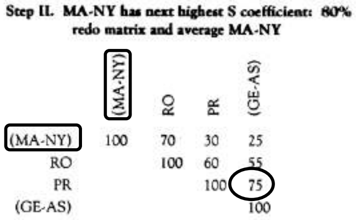
	MA	NY	RO	PR	(GE-AS)
MA	100	80	70	30	25
NY		100	70	30	25
RO			100	60	55
PR				100	75
(GE-AS)					100

UPGMA cluster analysis

- reduce overall similarity matrix by clustering together *Gentiana* and *Aster* and recalculate similarity values
- identify most similar pair of taxa and cluster them



Phenetics

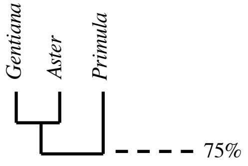


Step II. MA-NY has next highest S coefficient: 80%
redo matrix and average MA-NY

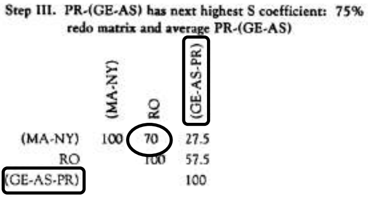
	(MA-NY)	RO	PR	(GE-AS)
(MA-NY)	100	70	30	25
RO		100	60	55
PR			100	75
(GE-AS)				100

UPGMA cluster analysis

- reduce overall similarity matrix by clustering together *Magnolia* and *Nymphaea* and recalculate similarity values
- identify most similar pair of taxa and cluster them



Phenetics

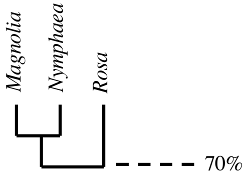


Step III. PR-(GE-AS) has next highest S coefficient: 75%
redo matrix and average PR-(GE-AS)

	(MA-NY)	RO	(GE-AS-PR)
(MA-NY)	100	70	27.5
RO		100	57.5
(GE-AS-PR)			100

UPGMA cluster analysis

- cluster together *Gentiana*, *Aster*, and *Primula* and recalculate values
- identify most similar pair of taxa and cluster them



Phenetics

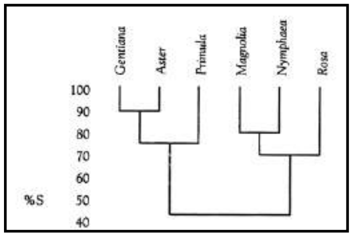
UPGMA cluster analysis

- cluster together *Magnolia*, *Nymphaea*, and *Rosa* and recalculate values

Step IV. RO-(MA-NY) has next highest S coefficient: 70%
redo matrix and average RO-(MA-NY)

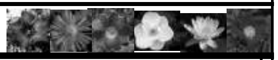
	(MA-NY-RO)	(GE-AS-PR)
(MA-NY-RO)	100	42.5
(GE-AS-PR)		100

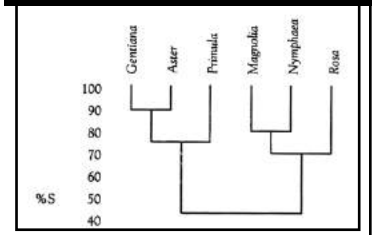
- cluster the two remaining larger groups at 42.5% to make final phenogram



Phenetics

- many different methods based on similarity or differences (including multiple components, ordination, etc.)
- in lab you will be using UPGMA & Neighbor-joining using a computer program PAUP





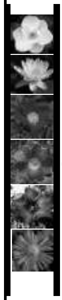
Cladistics

How do you analyze this same data based on cladistics - “shared derived character states”?

	vessels	apocarpy	sympetaly	epipetaly	trees	epigyny	beetle poll.	tepals	bicarpellate	heterostyly
1. *Magnolia*	+	+	-	-	+	-	+	+	-	-
2. *Nymphaea*	-	+	-	-	-	-	+	+	-	-
3. *Rosa*	+	+	-	-	-	-	-	-	-	-
4. *Primula*	+	-	+	+	-	-	-	-	-	+
5. *Gentiana*	+	-	+	+	-	-	-	-	+	-
6. *Aster*	+	-	+	+	-	+	-	-	+	-

Cladistics

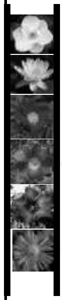
Issue #1- How do you determine what is derived?



	vessels	apocarpy	sympetaly	epipetaly	trees	epigyny	beetle poll.	tepals	bicarpellate	heterostyly
1. *Magnolia*	+	+	-	-	+	-	+	+	-	-
2. *Nymphaea*	-	+	-	-	-	-	+	+	-	-
3. *Rosa*	+	+	-	-	-	-	-	-	-	-
4. *Primula*	+	-	+	+	-	-	-	-	-	+
5. *Gentiana*	+	-	+	+	-	-	-	-	+	-
6. *Aster*	+	-	+	+	-	+	-	-	+	-

Cladistics

Issue #1- ordering or polarizing character states (primitive or derived)

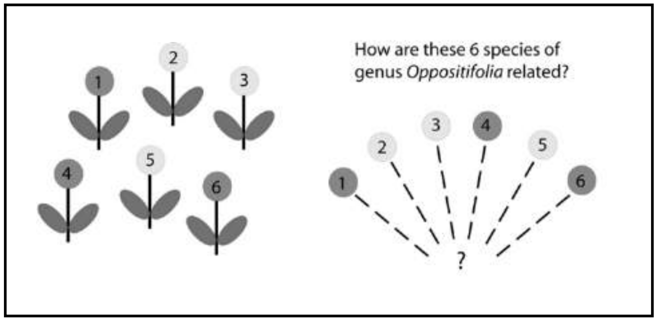


	vessels
1. ***Magnolia***	+
2. ***Nymphaea***	-
3. ***Rosa***	+
4. ***Primula***	+
5. ***Gentiana***	+
6. ***Aster***	+

[redacted]

Cladistics

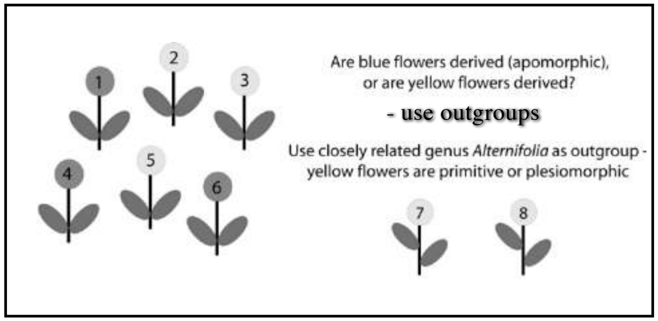
Issue #1- ordering or polarizing character states (primitive or derived)



Cladistics

Issue #1- ordering or polarizing character states (primitive or derived)

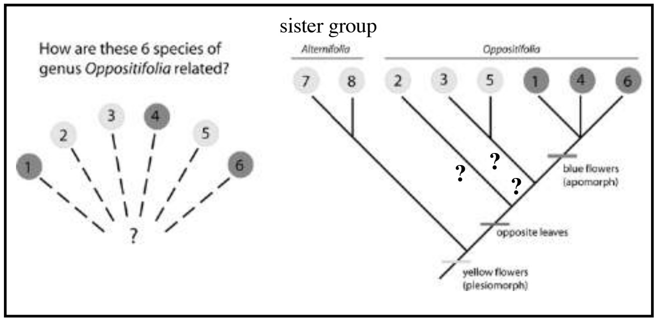
plesiomorph - primitive state apomorph - derived state



Cladistics

Issue #1- ordering or polarizing character states (primitive or derived)

Blue flowers = synapomorph - shared derived state



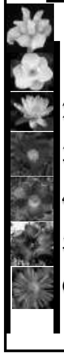
Cladistics

Add in *Amborella* as sister outgroup to rest of angiosperms

	vessels	apocarpy	sympetaly	epipetaly	trees	epigyny	beetle poll.	tepals	bicarpellate	heterostyly
1. ***Magnolia***	+	+	-	-	+	-	+	+	-	-
2. ***Nymphaea***	-	+	-	-	-	-	+	+	-	-
3. ***Rosa***	+	+	-	-	-	-	-	-	-	-
4. ***Primula***	+	-	+	+	-	-	-	-	-	+
5. ***Gentiana***	+	-	+	+	-	-	-	-	+	-
6. ***Aster***	+	-	+	+	-	+	-	-	+	-

Cladistics

Convert data matrix to "0" & "1"

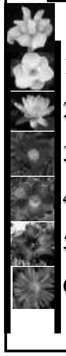


	vessels	apocarpy	sympetaly	epipetaly	trees	epigyny	beetle poll.	tepals	bicarpellate	heterostyly
Amborella	-	+	-	-	+	-	+	+	-	-
1. ***Magnolia***	+	+	-	-	+	-	+	+	-	-
2. ***Nymphaea***	-	+	-	-	-	-	+	+	-	-
3. ***Rosa***	+	+	-	-	-	-	-	-	-	-
4. ***Primula***	+	-	+	+	-	-	-	-	-	+
5. ***Gentiana***	+	-	+	+	-	-	-	-	+	-
6. ***Aster***	+	-	+	+	-	+	-	-	+	-

e.g., "-" to "0" and "+" to "1"

Cladistics

Convert data matrix to "0" & "1"

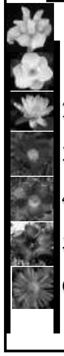


	vessels	apocarpy	sympetaly	epipetaly	trees	epigyny	beetle poll.	tepals	bicarpellate	heterostyly
Amborella	0	1	0	0	1	0	1	1	0	0
1. ***Magnolia***	1	1	0	0	1	0	1	1	0	0
2. ***Nymphaea***	0	1	0	0	0	0	1	1	0	0
3. ***Rosa***	1	1	0	0	0	0	0	0	0	0
4. ***Primula***	1	0	1	1	0	0	0	0	0	1
5. ***Gentiana***	1	0	1	1	0	0	0	0	1	0
6. ***Aster***	1	0	1	1	0	1	0	0	1	0

e.g., "-" to "0" and "+" to "1"

Cladistics

Convert data matrix to "0" & "1"

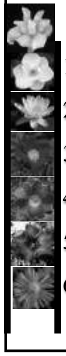


	vessels	apocarpy	sympetaly	epipetaly	trees	epigyny	beetle poll.	tepals	bicarpellate	heterostyly
Amborella	-	+	-	-	+	-	+	+	-	-
1. ***Magnolia***	+	+	-	-	+	-	+	+	-	-
2. ***Nymphaea***	-	+	-	-	-	-	+	+	-	-
3. ***Rosa***	+	+	-	-	-	-	-	-	-	-
4. ***Primula***	+	-	+	+	-	-	-	-	-	+
5. ***Gentiana***	+	-	+	+	-	-	-	-	+	-
6. ***Aster***	+	-	+	+	-	+	-	-	+	-

e.g., *Amborella* state (either "-" or "+") to "0"

Cladistics

Primitive 0
Derived 1

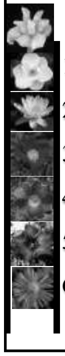


	No vessels - vessels	Apocarpy - syncarpy	Polypetaly - sympetaly	Free stamens - epipetaly	Trees - herbs	Hypogyny - epigyny	beetle poll. - other poll.	Tepals - sepals + peals	Various carpels - bicarpellate	Homostyly - heterostyly
Amborella	0	0	0	0	0	0	0	0	0	0
1. ***Magnolia***	1	0	0	0	0	0	0	0	0	0
2. ***Nymphaea***	0	0	0	0	1	0	0	0	0	0
3. ***Rosa***	1	0	0	0	1	0	1	1	0	0
4. ***Primula***	1	1	1	1	1	0	1	1	0	1
5. ***Gentiana***	1	1	1	1	1	0	1	1	1	0
6. ***Aster***	1	1	1	1	1	1	1	1	1	0

e.g., *Amborella* state (either “-” or “+”) to “0”

Cladistics

Primitive 0
Derived 1

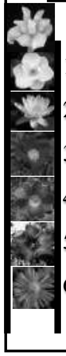


	No vessels - vessels	Apocarpy - syncarpy	Polypetaly - sympetaly	Free stamens - epipetaly	Trees - herbs	Hypogyny - epigyny	beetle poll. - other poll.	Tepals - sepals + peals	Various carpels - bicarpellate	Homostyly - heterostyly
Amborella	0	0	0	0	0	0	0	0	0	0
1. ***Magnolia***	1	0	0	0	0	0	0	0	0	0
2. ***Nymphaea***	0	0	0	0	1	0	0	0	0	0
3. ***Rosa***	1	0	0	0	1	0	1	1	0	0
4. ***Primula***	1	1	1	1	1	0	1	1	0	1
5. ***Gentiana***	1	1	1	1	1	0	1	1	1	0
6. ***Aster***	1	1	1	1	1	1	1	1	1	0

“shared derived” character states

Cladistics

Note: 2
uniformative
characters



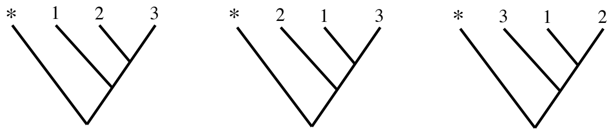
	No vessels - vessels	Apocarpy - syncarpy	Polypetaly - sympetaly	Free stamens - epipetaly	Trees - herbs	Hypogyny - epigyny	beetle poll. - other poll.	Tepals - sepals + peals	Various carpels - bicarpellate	Homostyly - heterostyly
Amborella	0	0	0	0	0	0	0	0	0	0
1. ***Magnolia***	1	0	0	0	0	0	0	0	0	0
2. ***Nymphaea***	0	0	0	0	1	0	0	0	0	0
3. ***Rosa***	1	0	0	0	1	0	1	1	0	0
4. ***Primula***	1	1	1	1	1	0	1	1	0	1
5. ***Gentiana***	1	1	1	1	1	0	1	1	1	0
6. ***Aster***	1	1	1	1	1	1	1	1	1	0

not “shared derived” character states

Cladistics

Issue #2 - how do you select the “best” tree?

- with 3 ingroup species and one outgroup (*), there are 3 trees possible



Cladistics

Issue #2 - how do you select the “best” tree?

- estimation procedure, however, usually involves vast number of possible “trees”
- this study with 7 taxa - there are 10,395 possible tree topologies
- examining all trees is possible here, but with larger numbers of taxa (as the 14 taxa used in lab this week – 7.9 trillion trees!) a heuristic approach is required

---

Cladistics

Issue #2 - how do you select the “best” tree?

- estimation procedure, however, usually involves vast number of possible “trees”
- **for a study with 50 taxa - there are 3 X 10^{74} possible trees or approaching number of atoms in universe (10^{79})!**
- **landmark paper in 1993 for angiosperms had 499 taxa - astronomical number of possible trees! >> 10^{1000}**
- **for a study of the Tree of Life - $10^{70,000,000}$**

Cladistics

Issue #2 - how do you select the “best” tree?

- the “best” tree is dependent on assumption of an optimality criterion: e.g., likelihood, parsimony
- cladistics (morphology) often uses parsimony - based on “Ockham’s Razor”

William of Ockham – *Entia non sunt multiplicanda praeter necessitatem* or “Entities should not be multiplied unnecessarily”

Cladistics

Issue #2 - how do you select the “best” tree?

- in the context of evolution, maximum parsimony = choosing the tree that requires the fewest number of evolutionary changes (apomorphies)
- choose the tree with the least amount of homoplasy - convergences or reversals or character conflict
- choose the shortest, simplest, most efficient tree

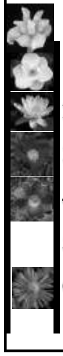
Cladistics

Issue #2 - how do you select the “best” tree?

http://evolution.genetics.washington.edu/phylip/software.html

- **36 of the around 370 phylogenetic software programs available!**
- **many can be used on about 50 free web servers (including supercomputers or tera-grids)**
- **in lab we will use two programs**

Cladistics



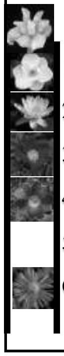
	No vessels - vessels	Apocarpy - syncarpy	Polypetaly - sympetaly	Free stamens - epipetaly	Trees - herbs	Hypogyny - epigyny	beetle poll. - other poll.	Tepals - sepals + peals	Various carpels - bicarpellate	Homostyly - heterostyly
Amborella										
1. *Magnolia*										
2. *Nymphaea*										
3. *Rosa*										
4. *Primula*										
5. *Gentiana*										
6. *Aster*										

- use maximum parsimony to find the “best” of 10,395 possible trees

---

Cladistics

2 characters conflict!



	No vessels - vessels	Apocarpy - syncarpy	Polypetaly - sympetaly	Free stamens - epipetaly	Trees - herbs	Hypogyny - epigyny	beetle poll. - other poll.	Tepals - sepals + peals	Various carpels - bicarpellate	Homostyly - heterostyly
Amborella	0	0	0	0	0	0	0	0	0	0
1. *Magnolia*	1	0	0	0	0	0	0	0	0	0
2. *Nymphaea*	0	0	0	0	1	0	0	0	0	0
3. *Rosa*	1	0	0	0	1	0	1	1	0	0
4. *Primula*	1	1	1	1	1	0	1	1	0	1
5. *Gentiana*	1	1	1	1	1	0	1	1	1	0
6. *Aster*	1	1	1	1	1	1	1	1	1	0

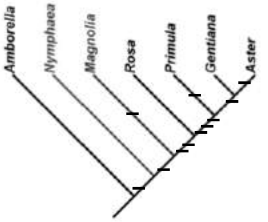
Cladistics

- **two trees are equally parsimonious**
- with character conflict, each is 11 steps long and not the expected 10

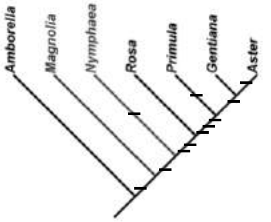
Consistency Index = 0.91

$$\frac{\text{\# changes minimally expected}}{\text{\# changes occurred on tree}}$$

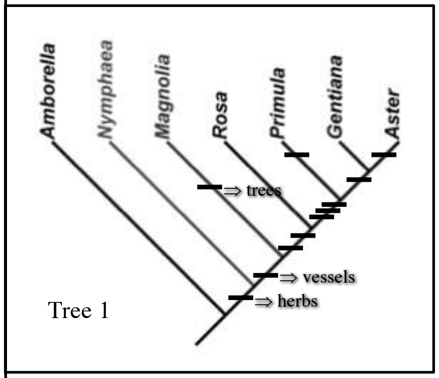
Most parsimonious tree 1



Most parsimonious tree 2

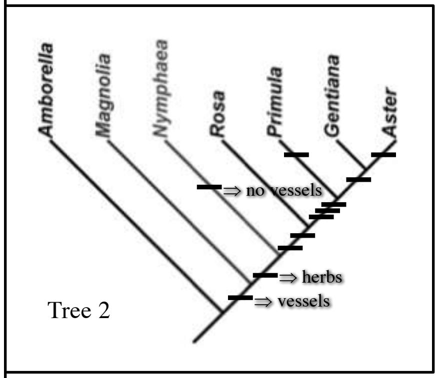


Cladistics



- **tree 1 has vessels as synapomorphy for all taxa except outgroup + *Nymphaea***
- habit shows homoplasy ("messy") with an origin to herb and then reversal back to tree

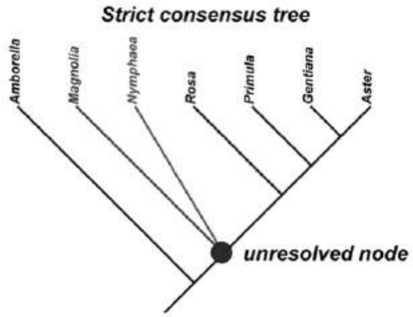
Cladistics



- **tree 2 has herbs as synapomorphy for all taxa except outgroup + *Magnolia***
- vessels shows homoplasy ("messy") with an origin to vessels and then reversal back to vessel-less

Cladistics

- a consensus tree depicts the maximum information possible from all most parsimonious trees (note: not equal to phenogram)



Phylogenetic Analysis of Asterids

1. Data set for 13 asterids and one rosid outgroup

   report should include data set (characters/states)

Phylogenetic Analysis of Asterids

1. Data set for 13 asterids and one rosid outgroup

   report should include data set (characters/states)

2. Distance (phenetic) approach in PAUP – two different ones

   report should include UPGMA and NJ trees

3. Parsimony (cladistic) approach in PAUP

   report should include strict consensus tree (# trees)

   report optionally include strict consensus tree after weighting characters

---

Phylogenetic Analysis of Asterids

4. Mapping of characters onto DNA tree in MacClade

   report should include discussion of good vs. bad characters (homoplasy)

5. Phylogenetics and classification

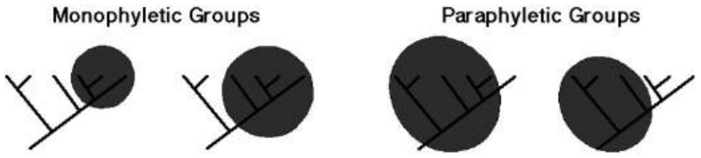
   report should include discussion of how asterids are or should be classified based on YOUR data

---

Cladistics

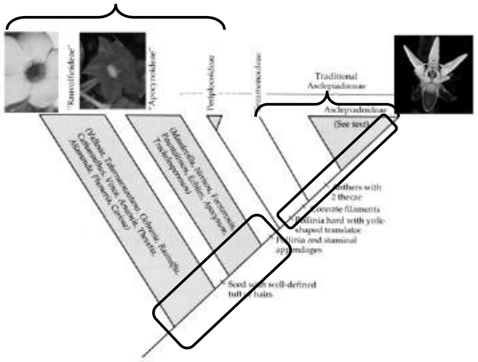
Issue #3 - how do you construct a classification?

- most cladists advocate monophyletic groupings only - stressing primacy of descent
- a different group, evolutionary taxonomists, allow for paraphyletic grouping - stressing both descent and modification



Cladistics

Issue #3 - **how do you construct a classification?**



Milkweeds are a highly derived lineage from within the dogbane family - Apocynaceae

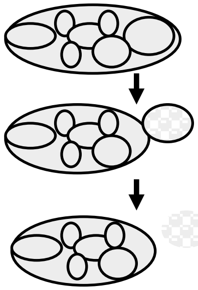
- recognizing Asclepiadaceae makes the Apocynaceae paraphyletic

- some agree since Asclepiadaceae are so divergent

Cladistics

Issue #3 - how do you construct a classification?

Lisianthius in central Panamanian cloud forests



peripheral isolates: new species forms at edge of retained ancestral species

Recognize paraphyletic species?

- island or peripheral geographic speciation is a common model in plants

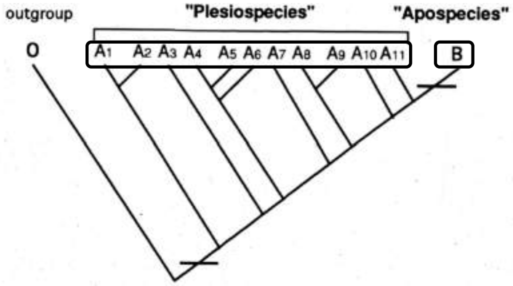
- ancestral species becomes paraphyletic, new species monophyletic

---

Cladistics

Paraphyletic species - 3 options:

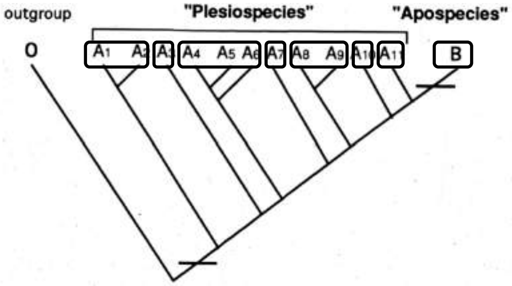
1. Recognize both the derived (apo) species and the paraphyletic ancestral (plesio) species - 2 species



Cladistics

Paraphyletic species - 3 options:

2. Recognize the derived (apo) species and monophyletic units from the ancestral (plesio) species



Cladistics

Paraphyletic species - 3 options:

3. Recognize only one monophyletic species