Algal Diversity and Phylogeny Part II

What about Algal Phylogeny?

-How algal groups (orders, families or species) have evolved?

-Is it possible to look back into the past?

-Is it possible to discover how the species or families are related to each other?

-Is it possible to uncover when species have appeared and how they become distributed in our planet after millions and millions of years?

Yes, it can be done! just as it is possible to determine if you are related to your family members (paternity tests)

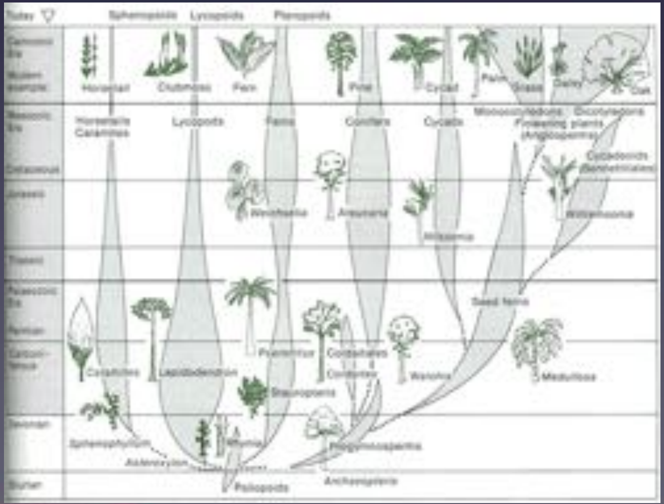
Phylogeny reconstruction

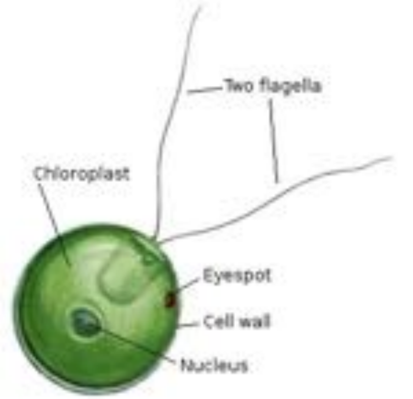
The goal of phylogeny reconstruction is to understand patterns and processes of evolution (to explain the occurrence of particular characteristics in specific groups of organisms)

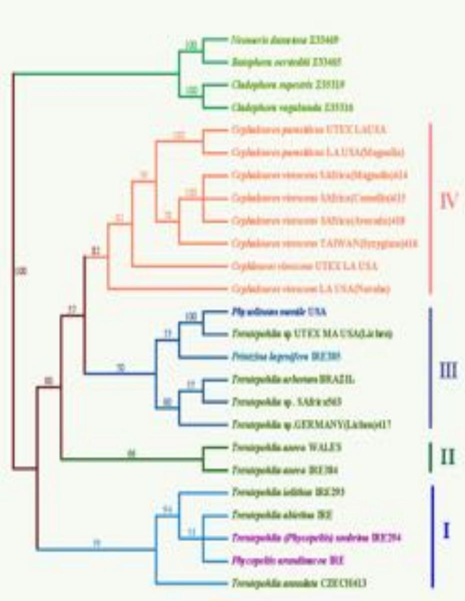
This is particularly important in the absence of fossil record

These characteristics have been typically morphological (for centuries!) but with molecular methods a new field has emerged!

There are a variety of inference methods (phenetics, parsimony, maximum likelihood, and bayesian). These methods result in tree diagrams



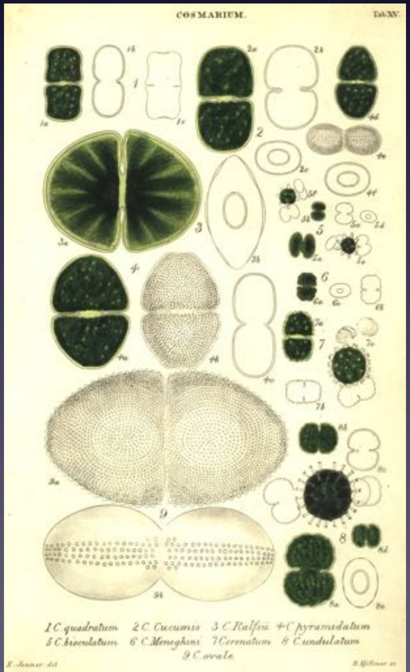




Molecular sequences vs. Structural data

Molecular data are highly numerous (base pairs)

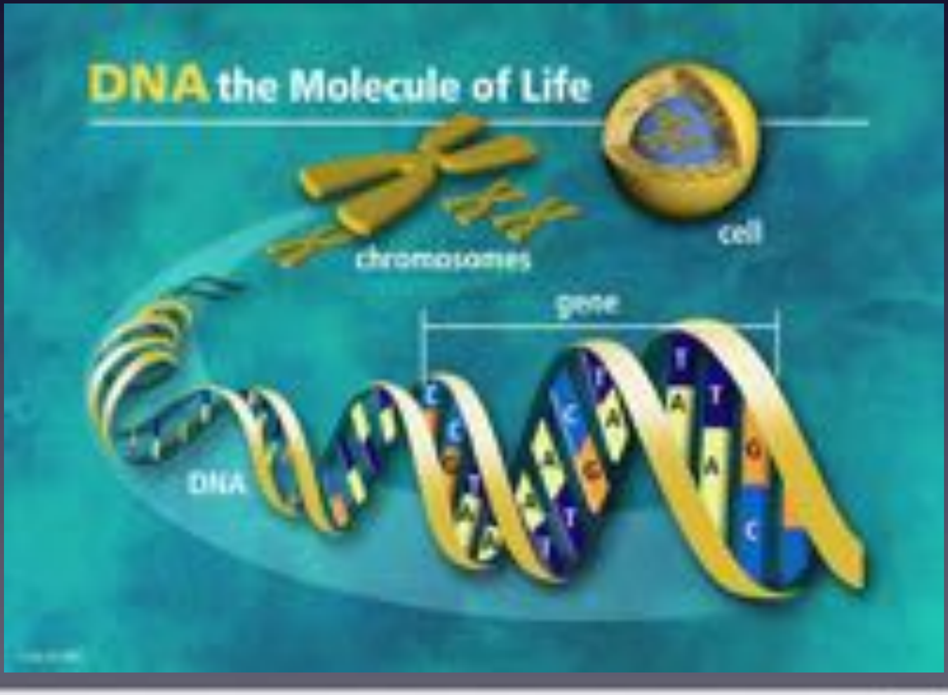
Although molecular data undergo parallel or convergent evolution, just as morphological data, its use often results in greater phylogenetic resolution than does use of morphological data alone


COSMARIUM.
Tab XV.
1 C. quadratum 2 C. Cucumis 3 C. Ralfsii 4 C. pyramidatum
5 C. biocnlatum 6 C. Meneghini 7 C. crenatum 8 C. undulatum
9 C. ovale
E. Jenner del
B. Hillmer sc

Cosmarium
Num. of recognized species: over 500
Few morphological characters

Versus

Potentially over 3 billions bp available for comparison



Choosing molecular phylogenetic approaches

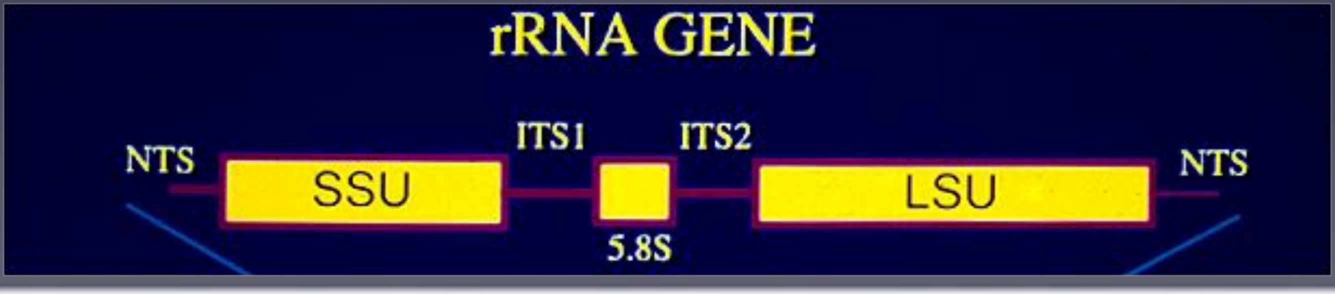
Many different molecules are available for comparison

The choose of the molecule should be adequate for the specific goals

Level of variation is important

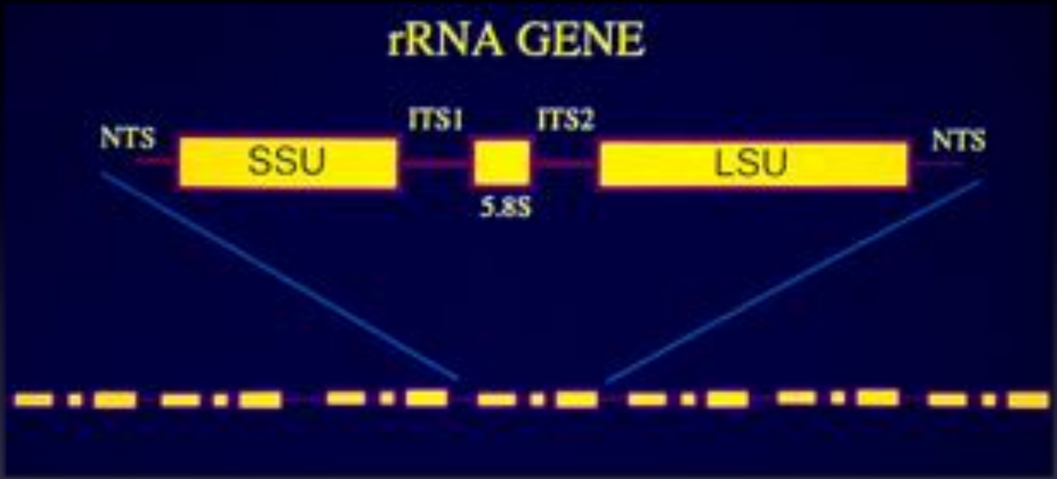
Ribosomal RNA sequences (rDNA) are widely used:

- **Large subunit (LSU or 28S) LESS VARIABLE**
- **Small subunit (SSU or 18S) LESS VARIABLE with regions of high variability**
- **Internal transcribed spacer (ITS) HIGHLY VARIABLE**



Ribosomal RNA gene (rDNA sequences, nuclear-encoded):

Divergences greater than 500 million years ago (major algal divisions) requires sequencing of very slowly evolving genes like the Small and Large (SSU and LSU) subunits ribosomal RNA



Conservation is derived from their essential role in cell function

rDNA is present in all eukaryotes; with a high number of characters (SSU @ 1800 bp, ITS @600 bp, LSU @ 2800 bp)

It occurs in <u>tandem repeats</u>, so it can be easily amplified

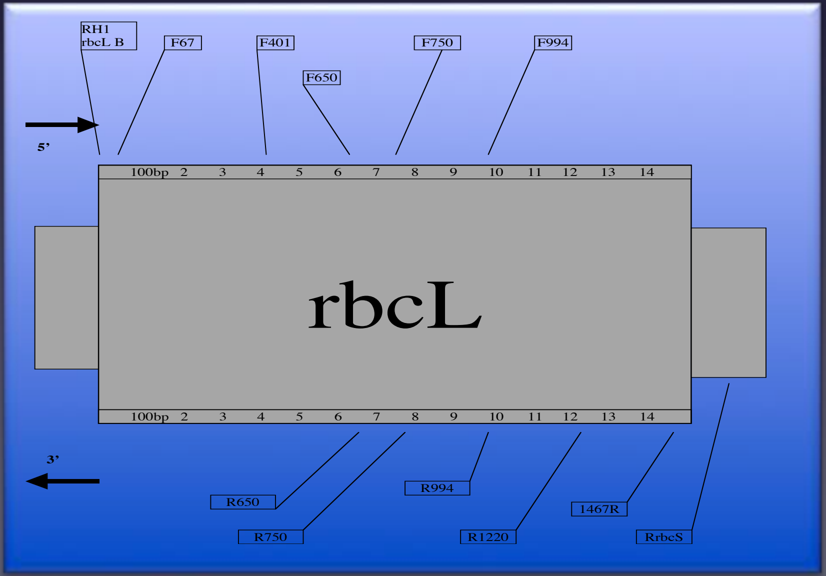
Rubisco gene (chloroplast-encoded):

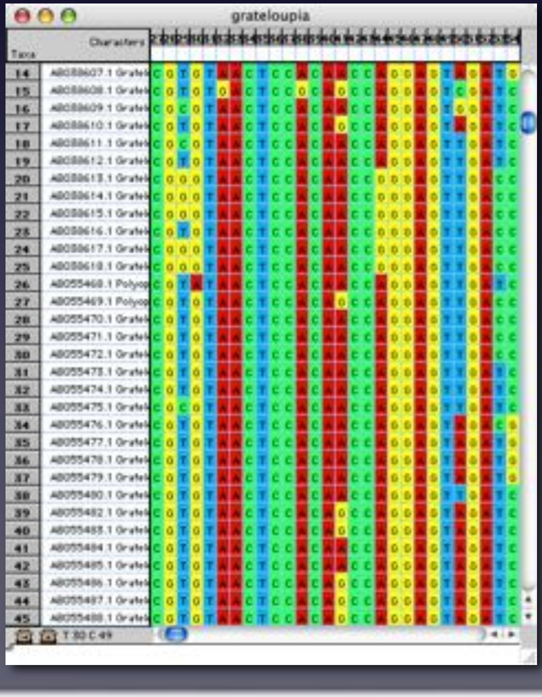
Large subunit rubisco or *rbc*L is commonly used

*rbc*L has a higher base substitution rate and is much more useful for more “recent” divergence events inside of orders, families and particularly genera and species level

No insertions or deletions: 1467 bp

Easy to align


RH1
rbcL B
F67
F401
F650
F750
F994
5’
100bp 2 3 4 5 6 7 8 9 10 11 12 13 14
rbcL
100bp 2 3 4 5 6 7 8 9 10 11 12 13 14
3’
R650
R750
R994
R1220
1467R
RbcS

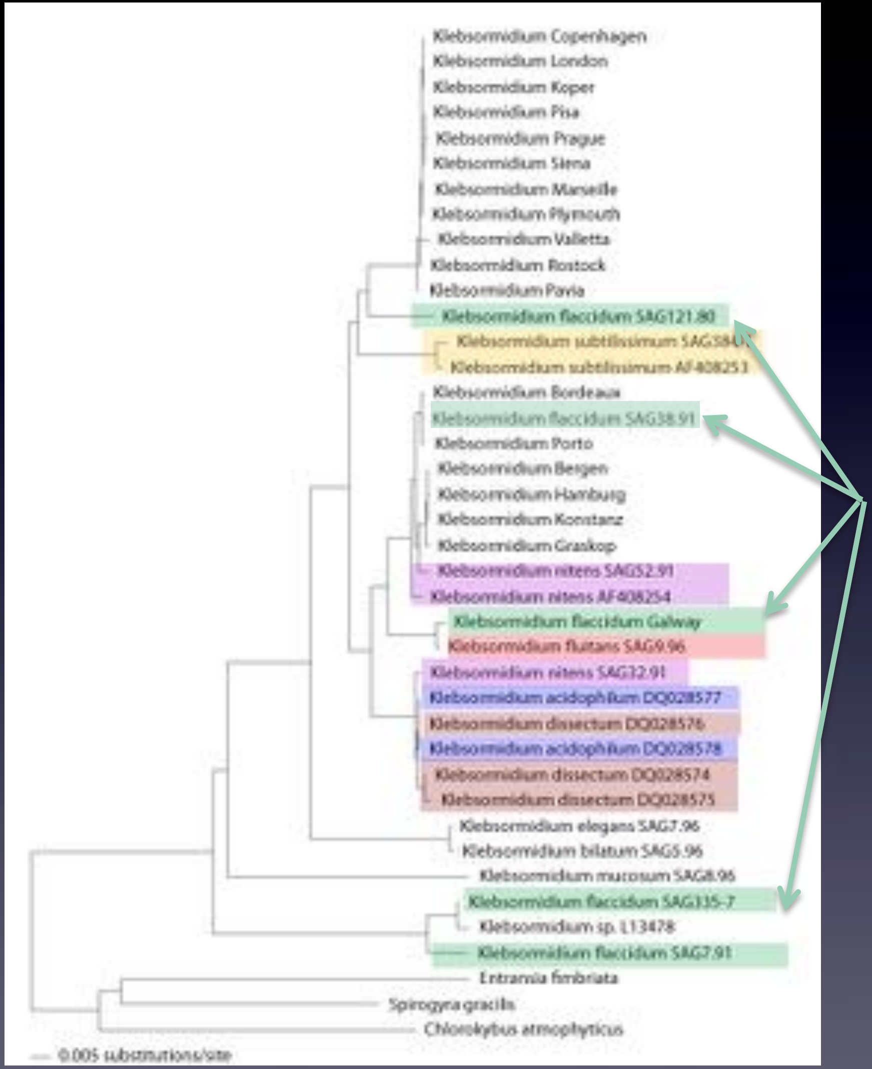


Generating, identifying, and evaluating optimal phylogenetic trees

Trees can be generated by

a) Distance methods

b) Maximum parsimony (cladistic methods)


Klebsormidium Copenhagen
Klebsormidium London
Klebsormidium Koper
Klebsormidium Pisa
Klebsormidium Prague
Klebsormidium Siena
Klebsormidium Marseille
Klebsormidium Plymouth
Klebsormidium Valletta
Klebsormidium Rostock
Klebsormidium Pavia
Klebsormidium flaccidum SAG121.80
Klebsormidium subtilissimum SAG384
Klebsormidium subtilissimum AF408253
Klebsormidium Bordeaux
Klebsormidium flaccidum SAG38.91
Klebsormidium Porto
Klebsormidium Bergen
Klebsormidium Hamburg
Klebsormidium Konstanz
Klebsormidium Graskop
Klebsormidium nitens SAG52.91
Klebsormidium nitens AF408254
Klebsormidium flaccidum Galway
Klebsormidium fluitans SAG9.96
Klebsormidium nitens SAG32.91
Klebsormidium acidophilum DQ028577
Klebsormidium dissectum DQ028576
Klebsormidium acidophilum DQ028578
Klebsormidium dissectum DQ028574
Klebsormidium dissectum DQ028575
Klebsormidium elegans SAG7.96
Klebsormidium bilatum SAG5.96
Klebsormidium mucosum SAG8.96
Klebsormidium flaccidum SAG335-7
Klebsormidium sp. L13478
Klebsormidium flaccidum SAG7.91
Entransia fimbriata
Spirogyra gracilis
Chlorokybus atmophyticus
0.005 substitutions/site

Samples originally corresponding to *Klebsormidium flaccidum* are spread along the *rbc*L tree

In Distance methods the number of differences between all pairs of taxa are determined

Then these numbers are are used to group taxa to form a Dendrogram (tree diagram) which attempts to accommodate all of the pairwise distances

Species	1	2	3	4	5	6	7	8	9	10
I	A	T	A	T	-	C	G	T	A	T
II	G	T	A	T	A	C	G	T	A	T
III	A	T	G	G	A	C	G	T	G	C
Outgroup	G	C	G	T	A	T	G	C	A	C

	I	II	III	Outgroup
I	-	2/10 = 0.20	5/10 = 0.50	0.70
II		-	0.50	0.50
III			-	0.60
Outgroup				-

P= # differences/10 sites

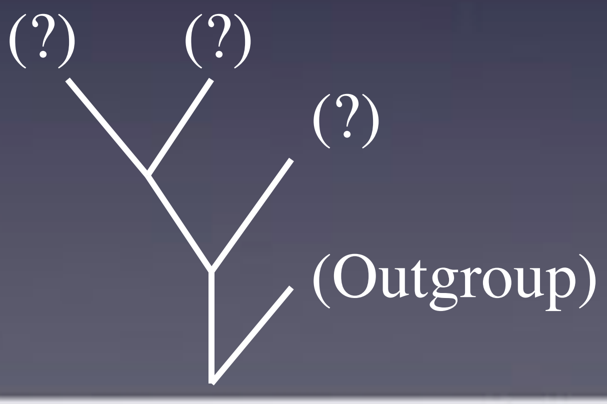
The table above summarizes the calculation of pairwise distances between the gene sequences for four hypothetical species

The coefficients provide a simple summary of how similar (or different) each sequence is from the other
Sequence I and II are more alike to each other than either is to III

Parsimony methods operate under the assumption (which is not always correct) that evolution operates in the most efficient (parsimonious) manner, i.e., the most accurate phylogeny is the one requiring the fewest number of changes

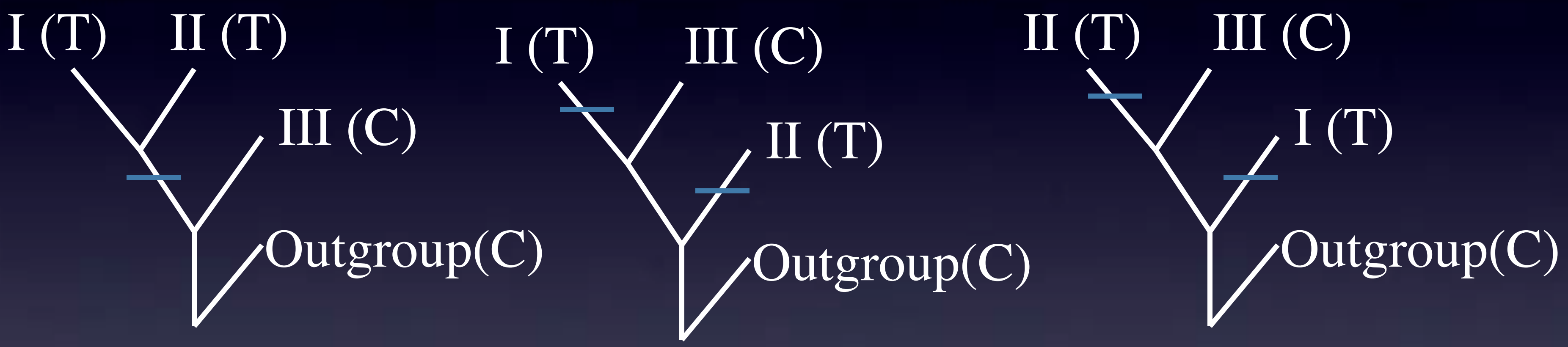
Species	1	2	3	4	5	6	7	8	9	10
I	A	T	A	T	-	C	G	T	A	T
II	G	T	A	T	A	C	G	T	A	T
III	A	T	G	G	A	C	G	T	G	C
Outgroup	G	C	G	T	A	T	G	C	A	C

Given 4 taxa (I, II, III, and outgroup) how many different ways these branches can be arranged in a tree?



There are only 3 possible branch arrangments/combinations

When we superimpose in these three possible trees the changes (-) necessary to explain each tree from the ancestral state (Outgroup) then we get different values (changes) for each tree:



1 change

2 changes

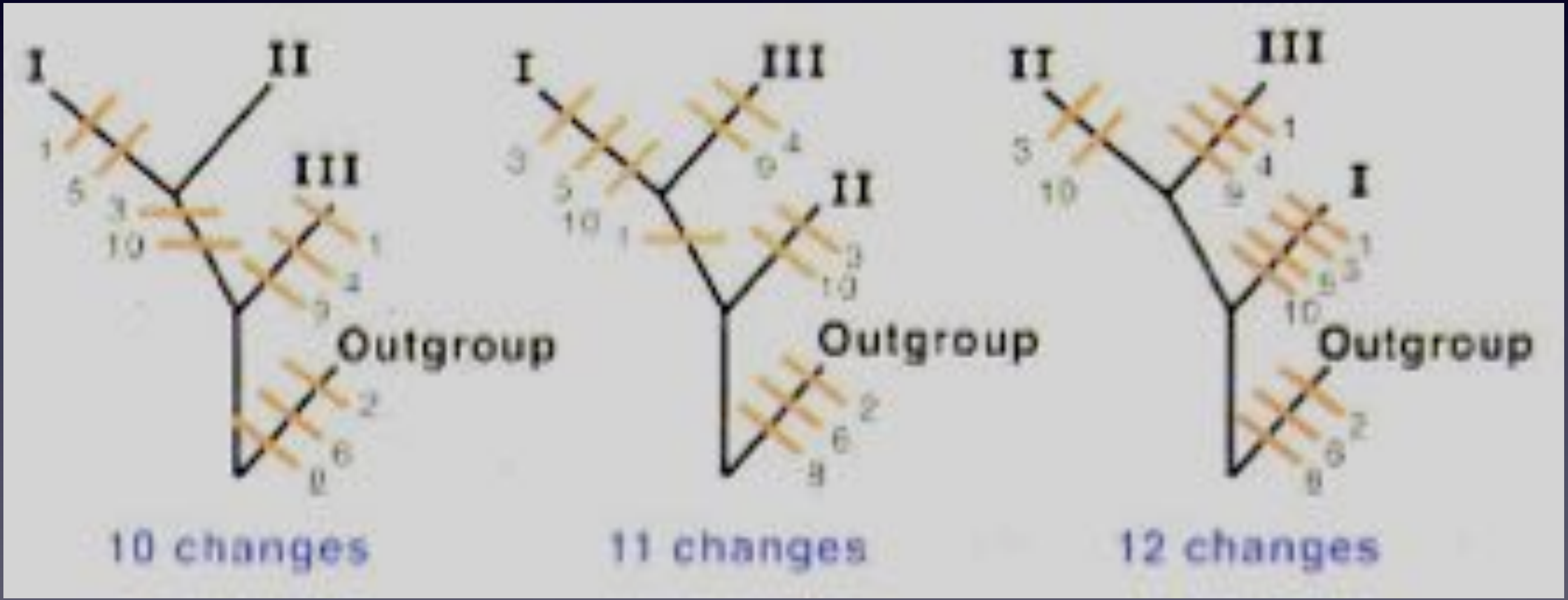
2 changes

— = C to T change

The left tree is the most parsimonious because it requires only 1 change!

The entire sequence must be considered in the analysis; in this case, that means all 10 positions, not just the 10th

Once all 10 positions are considered, the tree, called Cladogram, on the left is still considered the most parsimonious, as it takes only 10 changes (rather than 11 or 12)



Cladograms

The number of possible phylogenies (trees) increases with the number of taxa examined. There are 2 million possible trees for 10 taxa; for 50 taxa the number is $3x10^{74}$

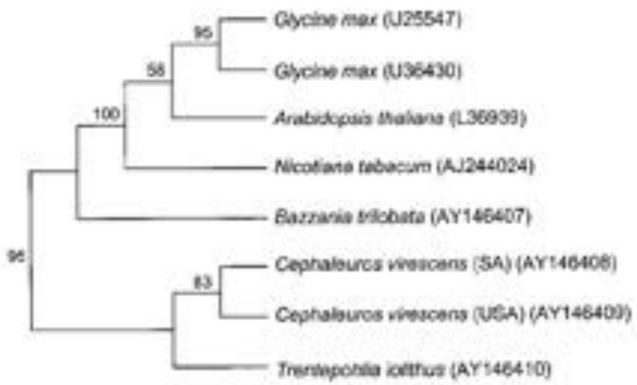
Therefore an Exhaustive search (when every possible tree is analyzed) is performed when the number of taxa is less than 12.

For studies with 12 or more taxa, Heuristic searches (trial and error approaches) are performed.

How reliable a tree can be?

Each individual node in the tree is evaluated by calculation of a Bootstrap value: it simulates the collection of replicate data sets through repeated re-sampling of the data, including some portions and leaving out others each time. The number of times a particular branch is recovered is determined as a percent value. Branches with 50% or less are considered poorly supported

Example of a cladogram, note the similarity among branch lengths



95
58
100
96
83
Glycine max (U25547)
Glycine max (U36430)
Arabidopsis thaliana (L36939)
Nicotiana tabacum (AJ244024)
Bazzania trilobata (AY146407)
Cephaleuros virescens (SA) (AY146408)
Cephaleuros virescens (USA) (AY146409)
Trentepohlia iolithus (AY146410)

Fig. 1. Maximum-parsimony tree based on partial phragmoplastin sequences, estimated from exhaustive search. A total of 584 positions are represented in the phylogram. GenBank accession numbers are given in parentheses. Tree length = 1090; CI = 0·868; RI = 0·636. Bootstrap values represent percentages derived from 10 000 replicates.

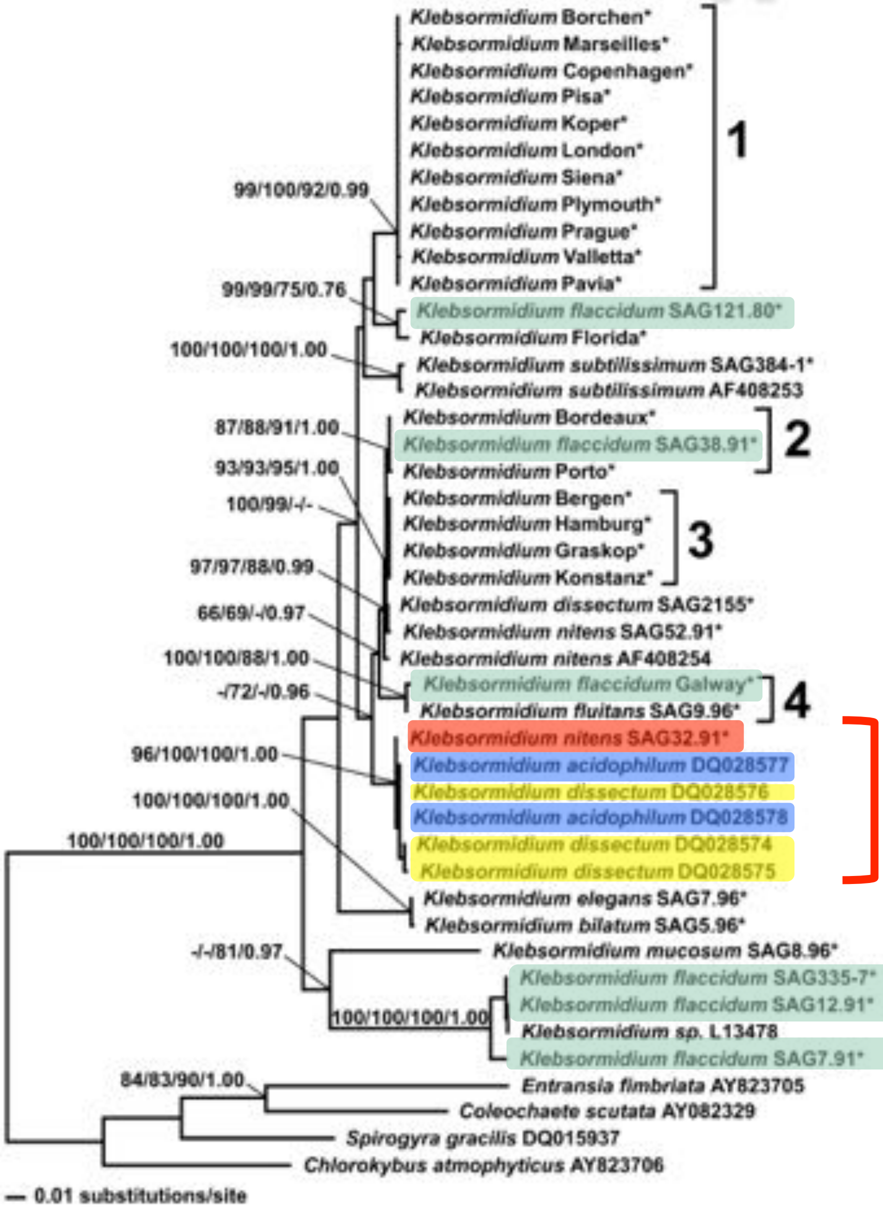

Klebsormidium Borchen*
Klebsormidium Marseilles*
Klebsormidium Copenhagen*
Klebsormidium Pisa*
Klebsormidium Koper*
Klebsormidium London*
Klebsormidium Siena*
Klebsormidium Plymouth*
Klebsormidium Prague*
Klebsormidium Valletta*
Klebsormidium Pavia*
1
99/100/92/0.99
99/99/75/0.76
Klebsormidium flaccidum SAG121.80*
Klebsormidium Florida*
100/100/100/1.00
Klebsormidium subtilissimum SAG384-1*
Klebsormidium subtilissimum AF408253
87/88/91/1.00
Klebsormidium Bordeaux*
Klebsormidium flaccidum SAG38.91*
Klebsormidium Porto*
2
93/93/95/1.00
100/99/-/-
Klebsormidium Bergen*
Klebsormidium Hamburg*
Klebsormidium Graskop*
Klebsormidium Konstanz*
3
97/97/88/0.99
Klebsormidium dissectum SAG2155*
66/69/-/0.97
Klebsormidium nitens SAG52.91*
100/100/88/1.00
Klebsormidium nitens AF408254
-/72/-/0.96
Klebsormidium flaccidum Galway*
Klebsormidium fluitans SAG9.96*
4
Klebsormidium nitens SAG32.91*
96/100/100/1.00
Klebsormidium acidophilum DQ028577
Klebsormidium dissectum DQ028576
100/100/100/1.00
Klebsormidium acidophilum DQ028578
100/100/100/1.00
Klebsormidium dissectum DQ028574
Klebsormidium dissectum DQ028575
Klebsormidium elegans SAG7.96*
Klebsormidium bilatum SAG5.96*
-/-/81/0.97
Klebsormidium mucosum SAG8.96*
Klebsormidium flaccidum SAG335-7*
Klebsormidium flaccidum SAG12.91*
100/100/100/1.00
Klebsormidium sp. L13478
Klebsormidium flaccidum SAG7.91*
84/83/90/1.00
Entransia fimbriata AY823705
Coleochaete scutata AY082329
Spirogyra gracilis DQ015937
Chlorokybus atmophyticus AY823706
0.01 substitutions/site

FIG. 2. Phylogram inferred from maximum-likelihood analysis of the *rbc*L gene in *Klebsormidium* and outgroup taxa, with bootstrap support (BP) and Bayesian posterior probabilities (PP) indicated at the nodes. From left to right, support values at nodes correspond to neighbor-joining BP, maximum-parsimony BP, maximum-likelihood BP, and Bayesian PP. New sequences produced in this study are marked with an asterisk. The four clades containing strains from urban environments are numbered as reported in the text.

Samples originally corresponding to *Klebsormidium flaccidum* are spread along the *rbc*L tree

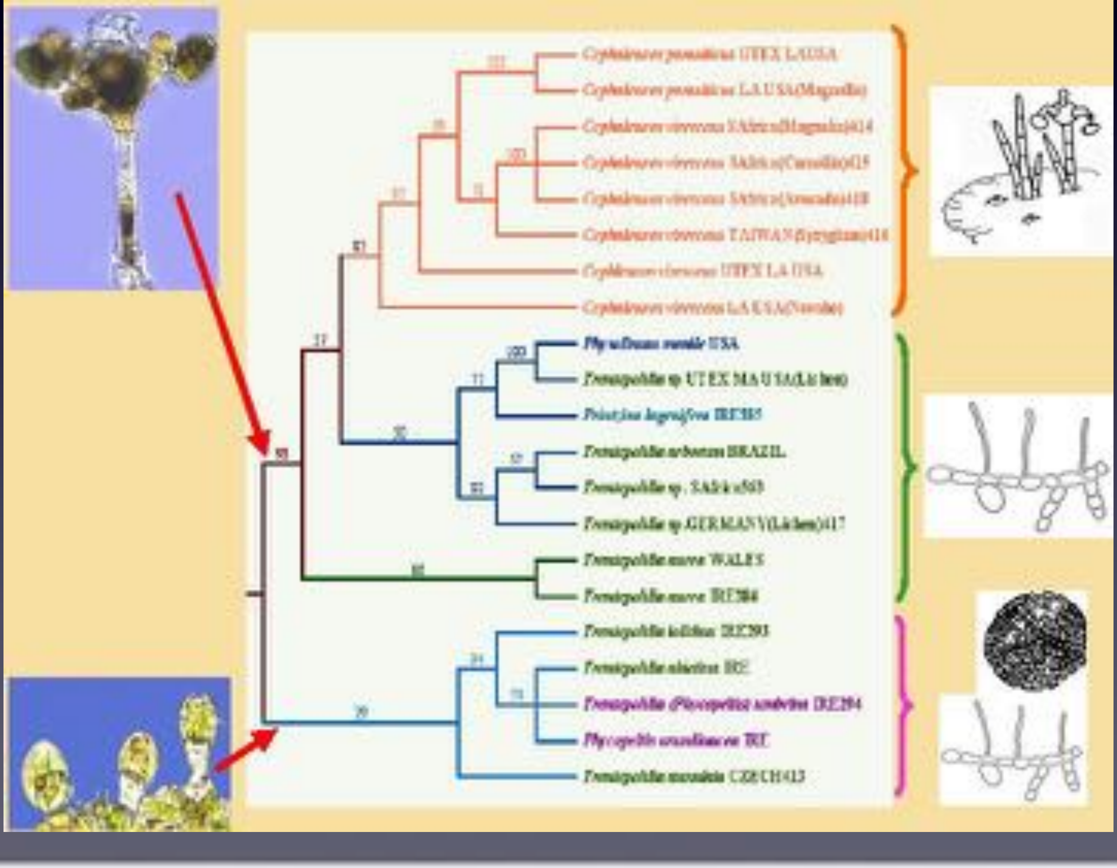
On the other hand, samples corresponding to different morphological species are grouped together

This is a <u>phylogram</u>, an evolutionary tree where branch lengths are proportional to the amount of character change

The application of Phylogeny

Can be used to define monophyletic groups of algal species with some confidence

When other data (biochemical, ultrastructural, etc) are mapped onto phylogenies, clade-specific attributes are often revealed!


Cephaleuros parasiticus UTEX LA USA
Cephaleuros parasiticus LA USA(Magnolia)
Cephaleuros virescens SAfrica(Magnolia)414
Cephaleuros virescens SAfrica(Camellia)415
Cephaleuros virescens SAfrica(Avocado)418
Cephaleuros virescens TAIWAN(Syzygium)416
Cephaleuros virescens UTEX LA USA
Cephaleuros virescens LA USA(Vernida)
Physolinum monile USA
Trentepohlia sp UTEX MAUI SA(Lichen)
Printzina lagenifera IRE385
Trentepohlia arborum BRAZIL
Trentepohlia sp. SAfrica503
Trentepohlia sp GERMANY(Lichen)417
Trentepohlia aurea WALES
Trentepohlia aurea IRE384
Trentepohlia iolithus IRE393
Trentepohlia abietina IRE
Trentepohlia (Phycopeltis) umbrina IRE394
Phycopeltis arundinacea IRE
Trentepohlia annulata CZECH413

Phylogenetic information can form a robust basis for predicting the extent to which physiologies or ecological behavior can be extrapolated from taxa that have been studied to related form that have not

Evolutionary utility

-Biodiversity

-Evolution of life history and sexual reproduction

-Ecological utility

-Detection of microorganisms in natural environments

-Evaluation of gene expression of primary producers in horizontal, vertical or geographic gradients

Reading: Assembling the Tree of Life

The biological revolution caused by electron microscopy in the 1970's is repeated today by the Gene analysis!

→

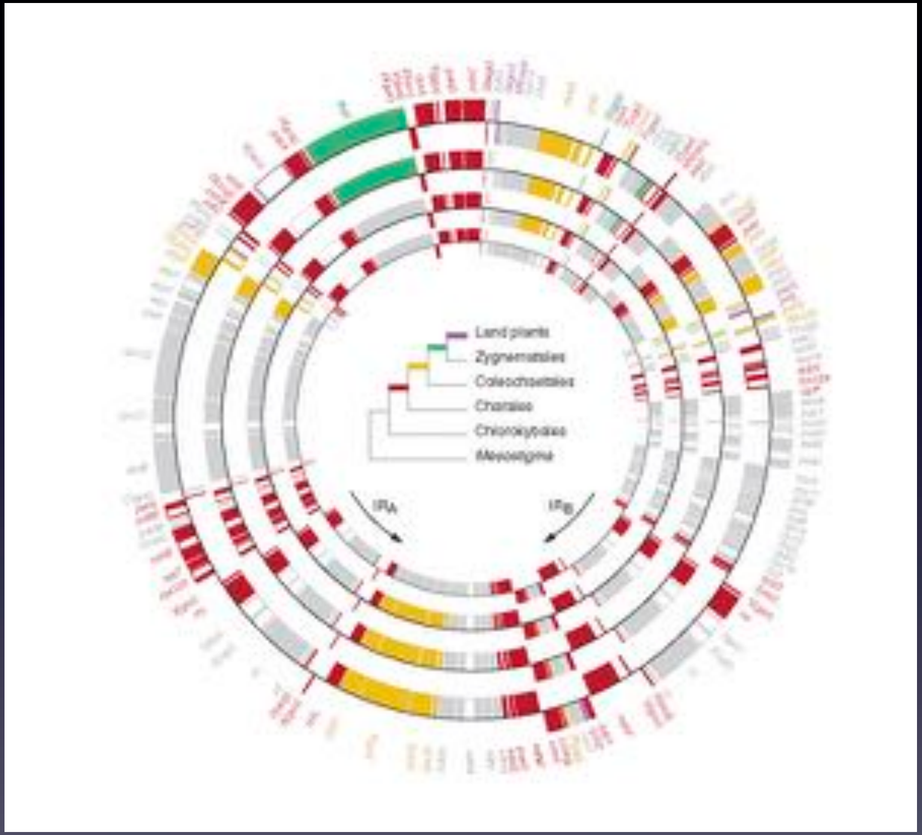
→ ?

What is next?

Maybe Phylogenomics?

Complete genome sequences is allowing to compare organisms in a new way to understand the origin and evolution of the genome

By combining the use of two disciplines, Molecular Phylogenetics (*the study of the evolutionary relationships among organisms and genes*) and Genomics (*the study of the organization and evolution of genes and genomes*)



The green algal ancestry of land plants as revealed by the chloroplast genome, by Turmel *et al*., *in press*