7.xi.

1. more about dynamic homology
2. search algorithms
3. traditional search strategies
4. new search strategies
5. tree form and consensus & compromise trees
6. summary

Additional techniques using dynamic homology

AACGGTTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCGGTAGGAT
AACGGTTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATGC

Search-based optimization (SBO)

improved/extended FSO

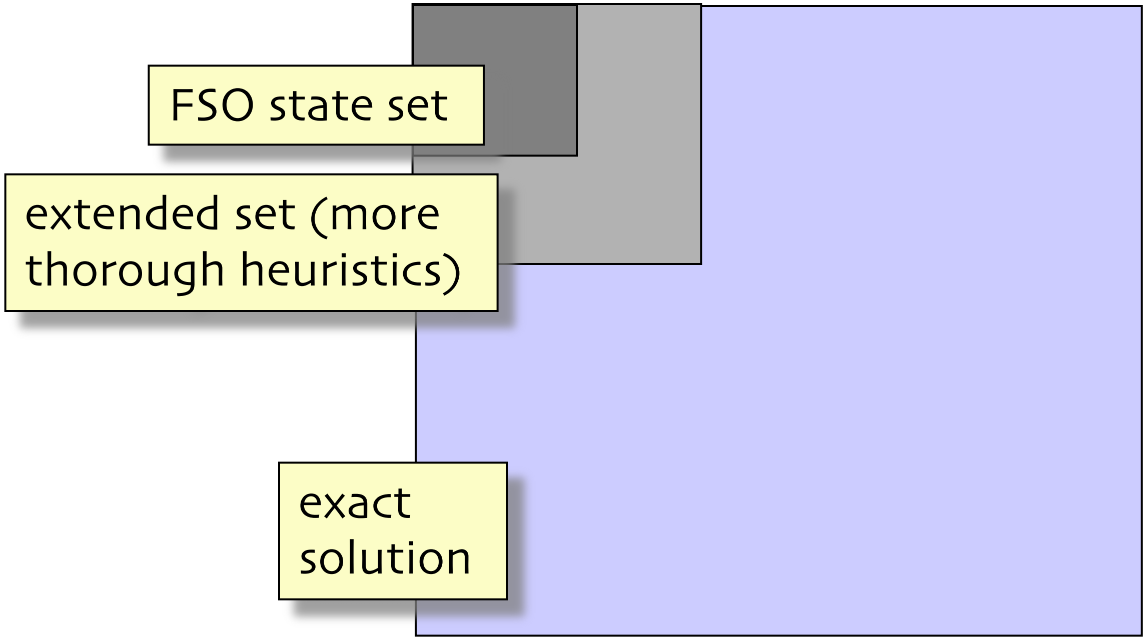
Iterative-pass optimization (IPO)

improved/extended DO

Search-based optimization (SBO)

1) define the set of possible ancestal states

size of this set can vary



FSO state set

extended set (more thorough heuristics)

exact solution

Search-based optimization (SBO)

AACGGTTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCGGTAGGAT
AACGGTTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATGC

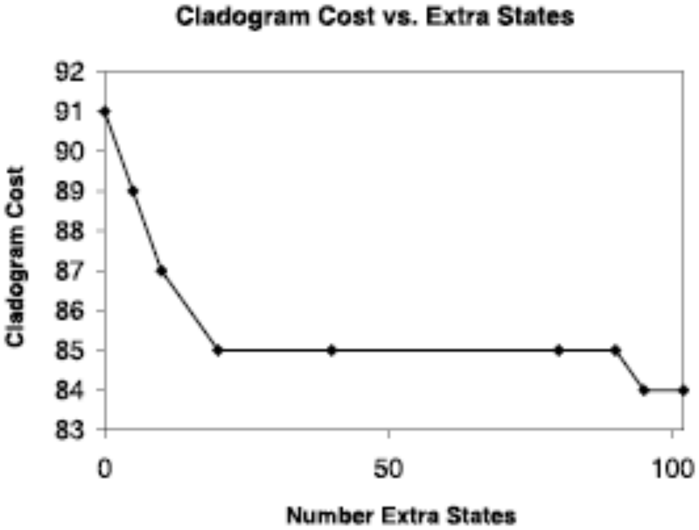
1) define the set of possible ancestal states

size of this set can vary

2) evaluate this set on a given tree

compared to FSO also OTHER sequences than those of the terminals will be used

leads to CONSIDERABLE reduction of tree length



Iterative pass optimization (IPO)

AACGGTTTAAGGTACGGAGAATTAGGCAA GGAT

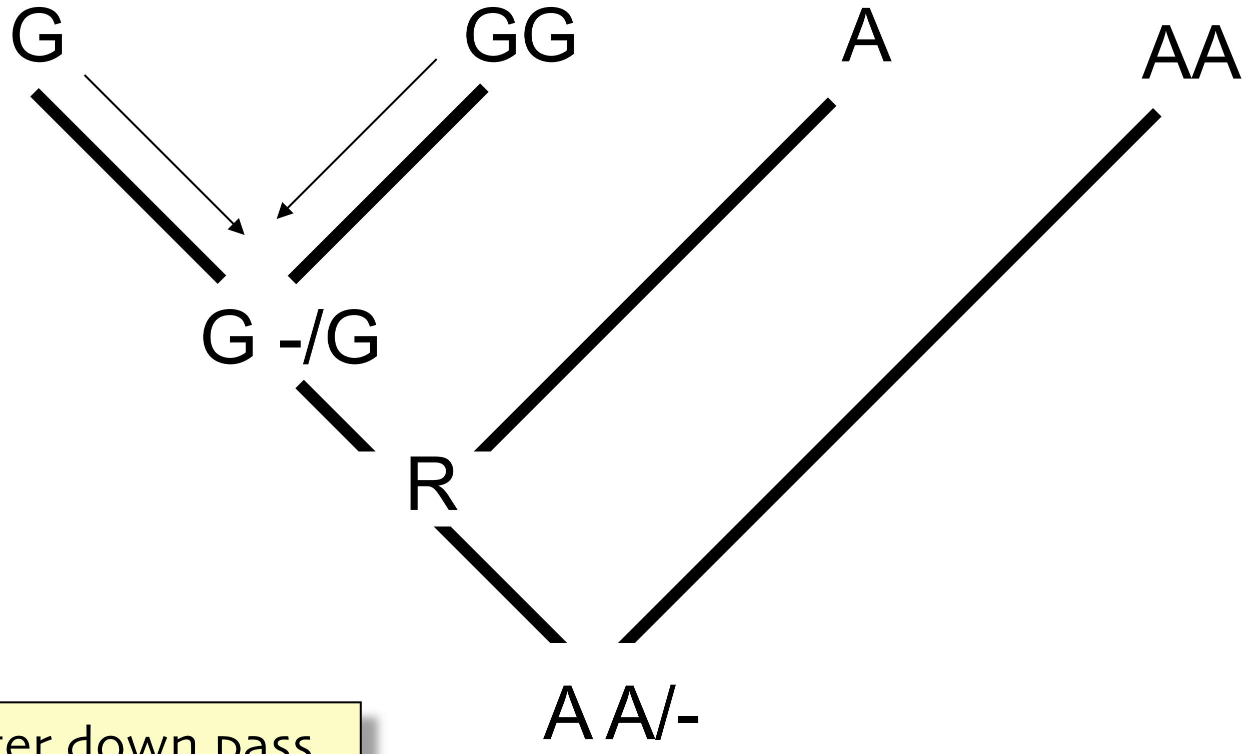
AACGGTTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTA GATGC

Wheeler, W.C. 2003. Cladistics 19: 254-260

WHY ?

Direct optimization (DO)

AACGGTTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCGGTAGGAT
AACGGTTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATGC



obtained after down pass

total cost 2+1+2=5

Iterative pass optimization (IPO)

AACGGTTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCGGTAGGAT
AACGGTTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATGC

1) initiation: sequences for HTU´s obtained e.g with down-pass of DO

2) up-pass is run iteratively, reevaluating each internal node with its 2 descendants & parental node SIMULTANEOUSLY

3 dimensional N-W algorithm

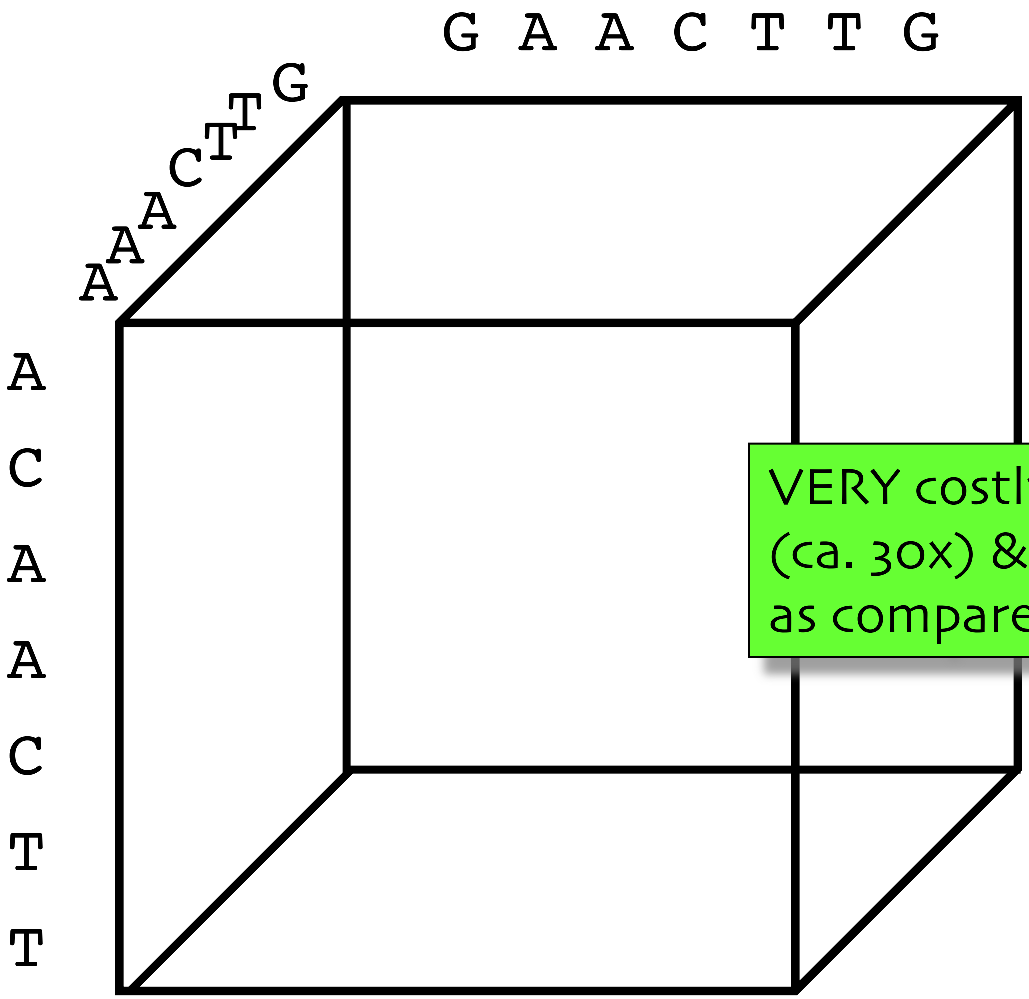
each internal node is connected to 3 other nodes, if 1 is changed also 2 others have to be reevaluated

this is done until solutions stabilize

when initiated with DO normally 3 iterative passes needed

3) cost calculation of the final tree (taking into account int. node sequences inferred using IPO)

compared to DO leads to reduction of tree length



VERY costly in terms of time (ca. 30x) & memory (ca. 9x) as compared to DO

3 sequences with 4 nucleotides ---> 4 x 4 x 4 cells in a matrix = 4^3

Size of the matrix N^M (N sequence length, M number of sequences)

Dynamic homology analyses in practice

AACGGTTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCGGTAGGAT
AACGGTTTAAGGTACGGAGAATTAGGCAACCCTAGGATGCGATGCGCAGAGTTAGGTACGGAGAATTAGGCAACCCTAGGATGC

1) CLUSTAL (or other similarily fast) alignment

2) conserved regions easily found

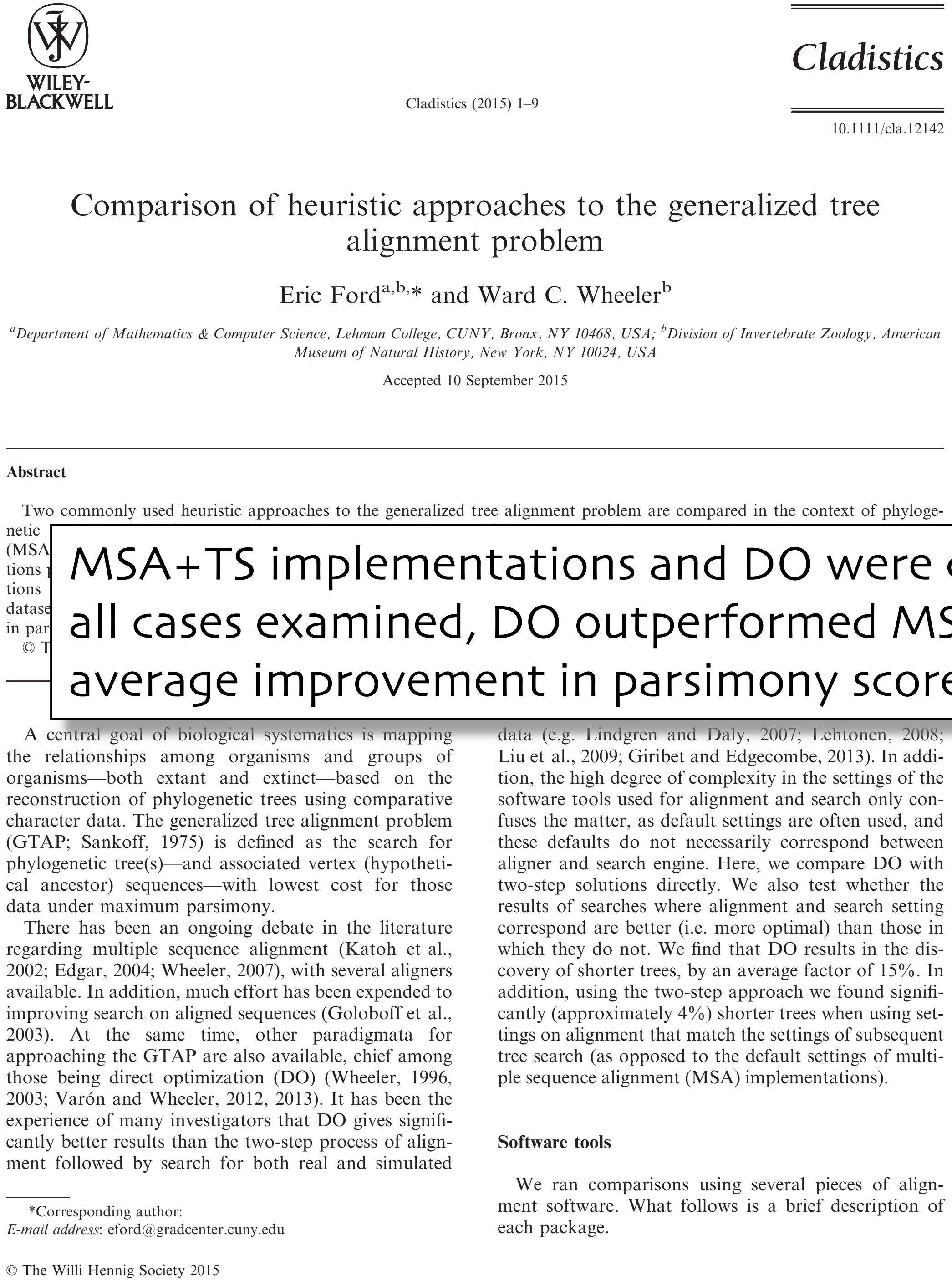
3) sequences cut into fragments using e.g. winclada

should be performed cautiously, only WITHIN conserved regions

can lead to HUGE speed-up of analyses

analyses of only small materials possible using computers with only 1 CPU

parallelization leads to LARGE speed-up of analyses



WILEY-
BLACKWELL

Cladistics (2015) 1–9

Cladistics

10.1111/cla.12142

Comparison of heuristic approaches to the generalized tree alignment problem

Eric Ford[a,b,*] and Ward C. Wheeler[b]

[a]*Department of Mathematics & Computer Science, Lehman College, CUNY, Bronx, NY 10468, USA;* [b]*Division of Invertebrate Zoology, American Museum of Natural History, New York, NY 10024, USA*

Accepted 10 September 2015

Abstract

Two commonly used heuristic approaches to the generalized tree alignment problem are compared in the context of phyloge-
netic
(MSA
tions
tions
datase
in par
© T

MSA+TS implementations and DO were compared… in all cases examined, DO outperformed MSA+TS with average improvement in parsimony score of ca. 15%

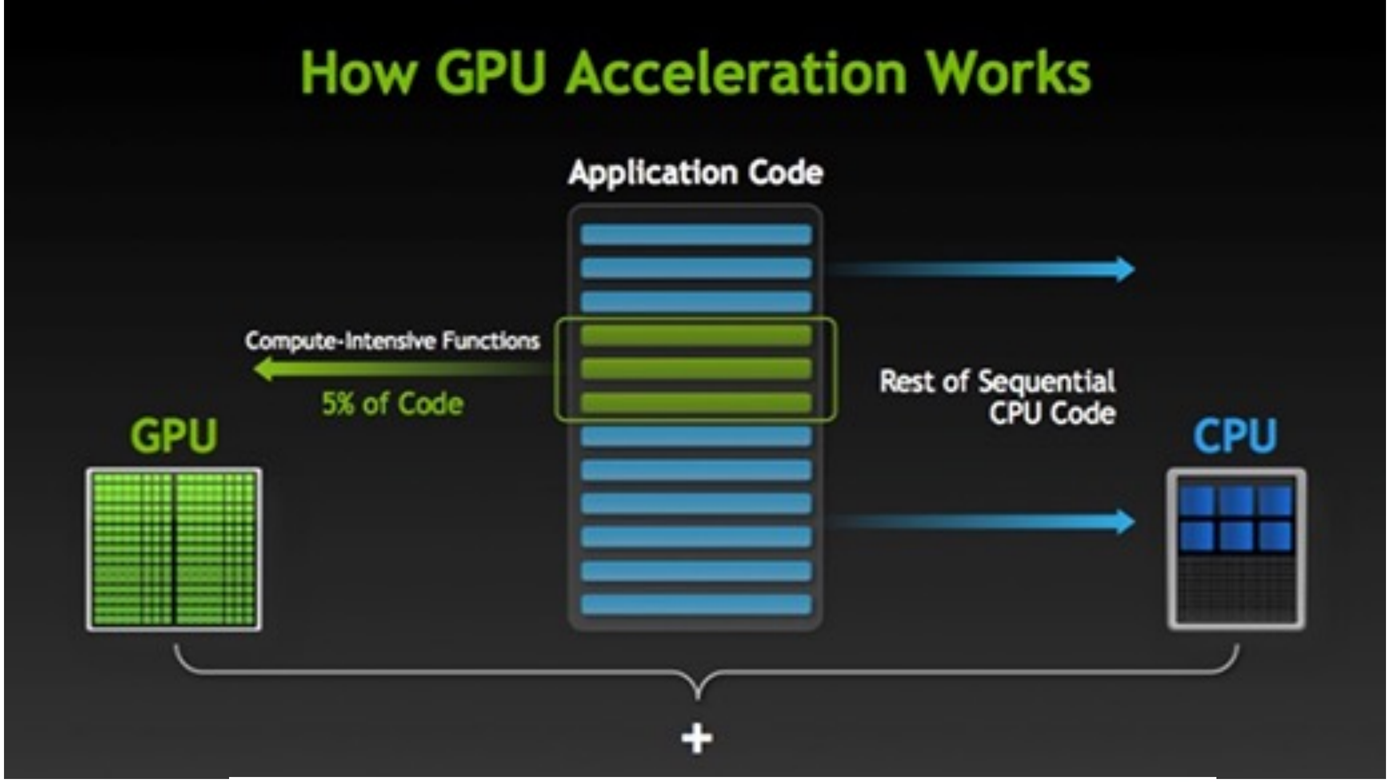
A central goal of biological systematics is mapping the relationships among organisms and groups of organisms—both extant and extinct—based on the reconstruction of phylogenetic trees using comparative character data. The generalized tree alignment problem (GTAP; Sankoff, 1975) is defined as the search for phylogenetic tree(s)—and associated vertex (hypothetical ancestor) sequences—with lowest cost for those data under maximum parsimony.

There has been an ongoing debate in the literature regarding multiple sequence alignment (Katoh et al., 2002; Edgar, 2004; Wheeler, 2007), with several aligners available. In addition, much effort has been expended to improving search on aligned sequences (Goloboff et al., 2003). At the same time, other paradigmata for approaching the GTAP are also available, chief among those being direct optimization (DO) (Wheeler, 1996, 2003; Varón and Wheeler, 2012, 2013). It has been the experience of many investigators that DO gives significantly better results than the two-step process of alignment followed by search for both real and simulated data (e.g. Lindgren and Daly, 2007; Lehtonen, 2008; Liu et al., 2009; Giribet and Edgecombe, 2013). In addition, the high degree of complexity in the settings of the software tools used for alignment and search only confuses the matter, as default settings are often used, and these defaults do not necessarily correspond between aligner and search engine. Here, we compare DO with two-step solutions directly. We also test whether the results of searches where alignment and search setting correspond are better (i.e. more optimal) than those in which they do not. We find that DO results in the discovery of shorter trees, by an average factor of 15%. In addition, using the two-step approach we found significantly (approximately 4%) shorter trees when using settings on alignment that match the settings of subsequent tree search (as opposed to the default settings of multiple sequence alignment (MSA) implementations).

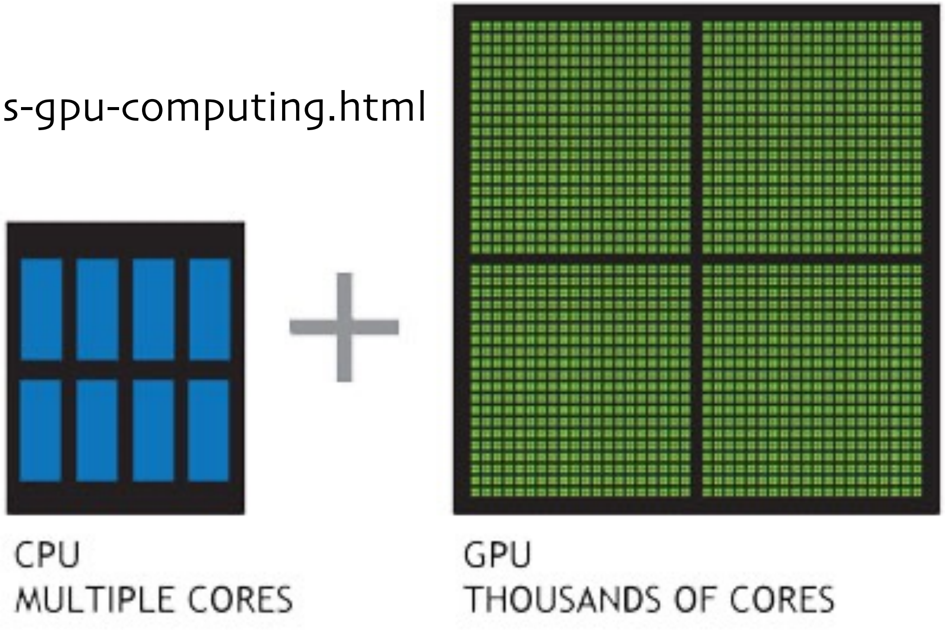
Software tools

We ran comparisons using several pieces of alignment software. What follows is a brief description of each package.

[*]Corresponding author:
E-mail address: eford@gradcenter.cuny.edu

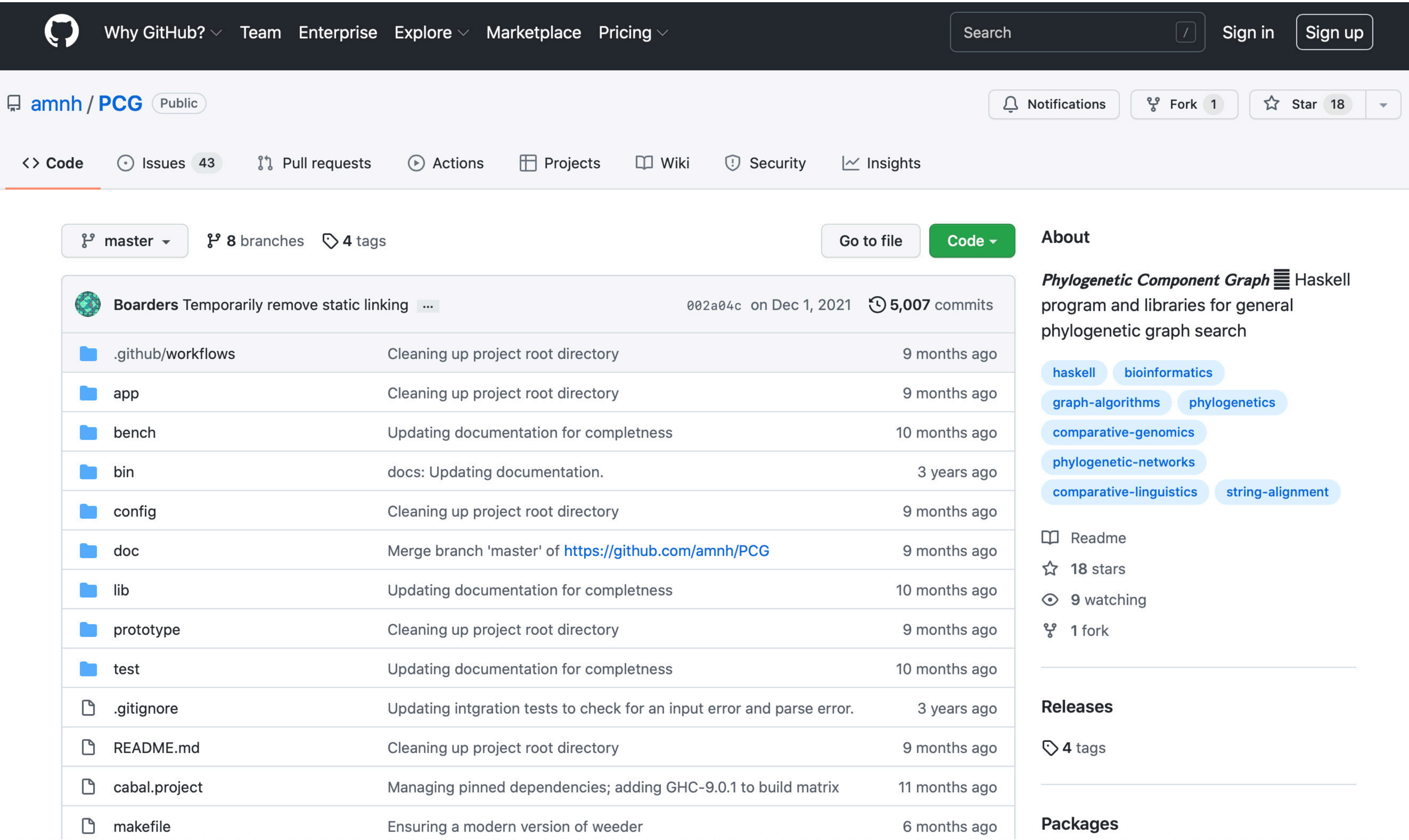

How GPU Acceleration Works
Application Code
Compute-Intensive Functions
5% of Code
GPU
Rest of Sequential CPU Code
CPU
+

nvidia.com/object/what-is-gpu-computing.html



haskell.org

https://github.com/amnh/PCG



Boarders Temporarily remove static linking ... 002a04c on Dec 1, 2021 **5,007** commits

.github/workflows	Cleaning up project root directory	9 months ago
app	Cleaning up project root directory	9 months ago
bench	Updating documentation for completness	10 months ago
bin	docs: Updating documentation.	3 years ago
config	Cleaning up project root directory	9 months ago
doc	Merge branch 'master' of https://github.com/amnh/PCG	9 months ago
lib	Updating documentation for completness	10 months ago
prototype	Cleaning up project root directory	9 months ago
test	Updating documentation for completness	10 months ago
.gitignore	Updating intgration tests to check for an input error and parse error.	3 years ago
README.md	Cleaning up project root directory	9 months ago
cabal.project	Managing pinned dependencies; adding GHC-9.0.1 to build matrix	11 months ago
makefile	Ensuring a modern version of weeder	6 months ago

About

Phylogenetic Component Graph 𝌆 Haskell program and libraries for general phylogenetic graph search

haskell bioinformatics graph-algorithms phylogenetics comparative-genomics phylogenetic-networks comparative-linguistics string-alignment

Readme

18 stars

9 watching

1 fork

Releases

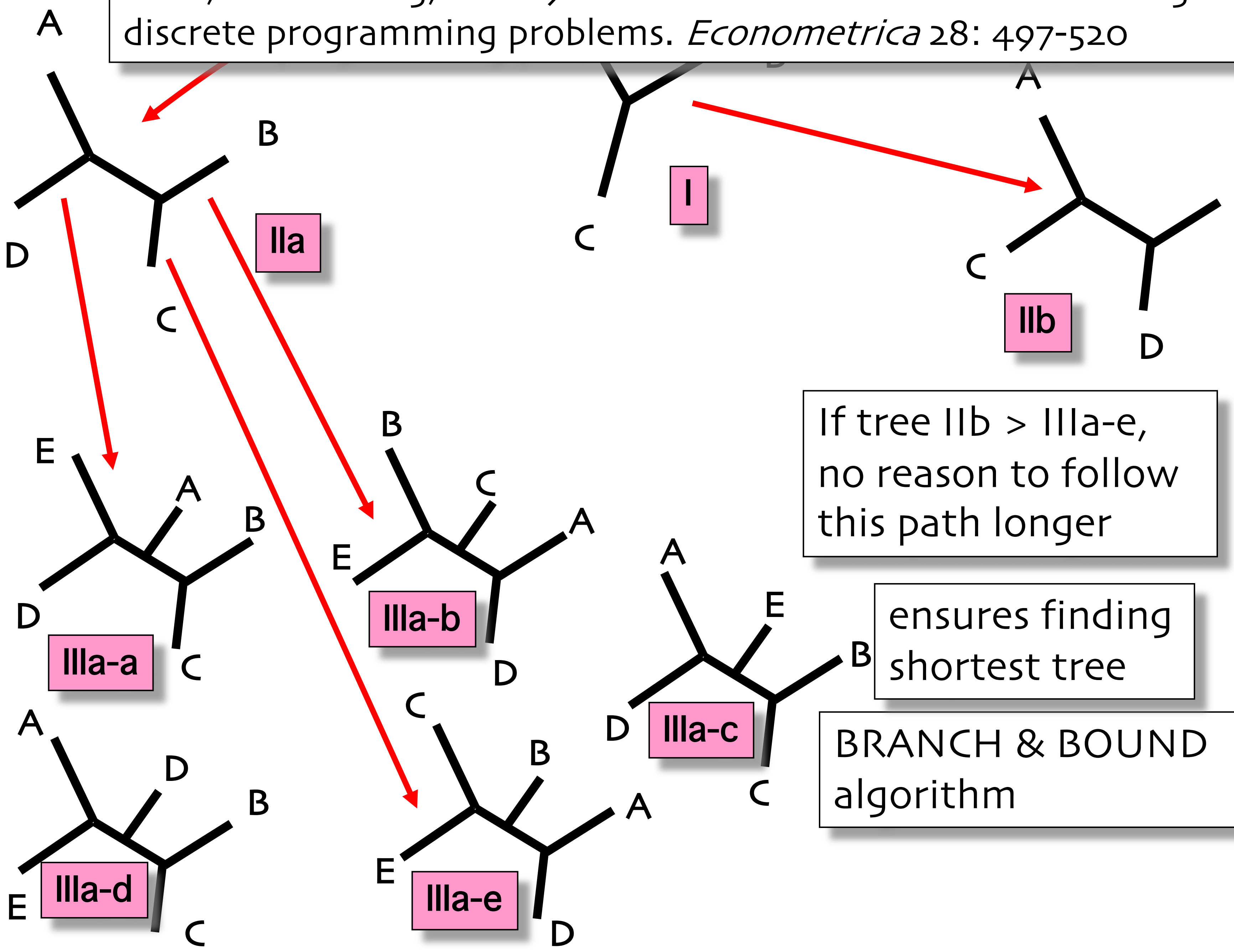
4 tags

Packages

T	B(T)
3	1
4	3
5	15
6	105
7	945
8	10 395
9	135 135
10	2×10^{6}
15	8×10^{12}
20	2×10^{20}
50	3×10^{74}

exhaustive search possible for only VERY limited no. of terminals

Land, A.H. & Doig, A.G. 1960. An automated method of solving discrete programming problems. *Econometrica* 28: 497-520



If tree IIb > IIIa-e, no reason to follow this path longer

ensures finding shortest tree

BRANCH & BOUND algorithm

T	B(T)
3	1
	3
	15
6	105
7	945
8	10 395
9	135 135
10	2×10^{6}
15	8×10^{12}
20	2×10^{20}
50	3×10^{74}

1,7 GHz processor

$1{,}65 \times 10^{6}$ trees/second

BRANCH and BOUND algorithm

can be used for max. ca. 30 terminals

HEURISTIC SEARCH

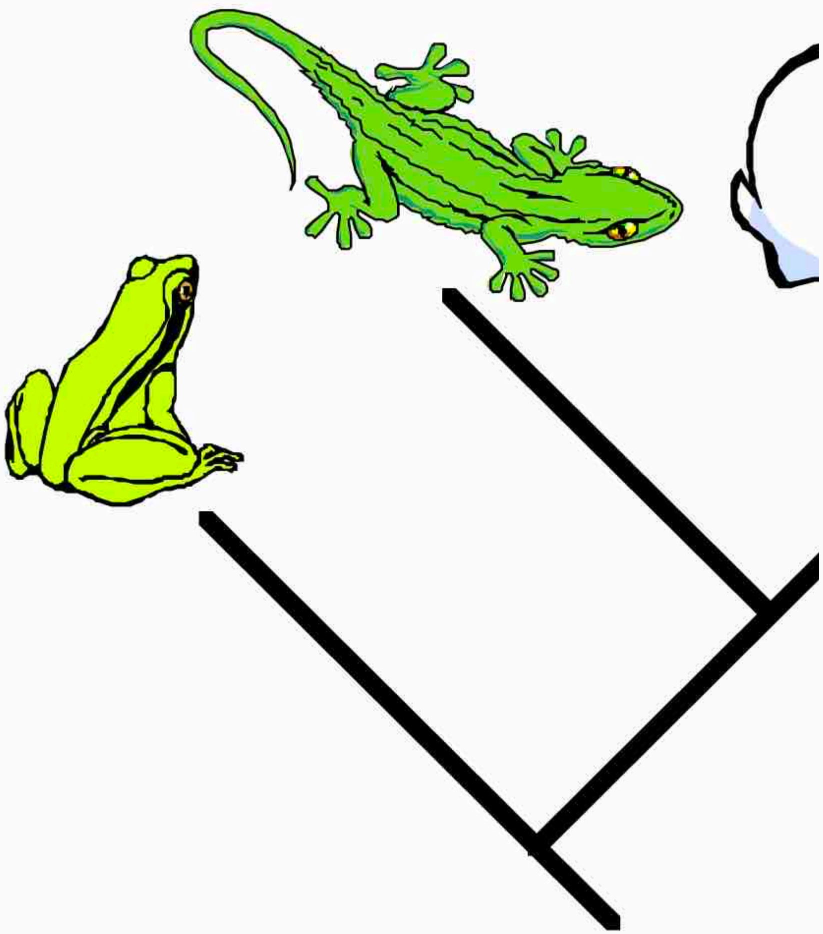
Wagner algorithm

Kluge, A. G. & Farris, J. S. 1969. Quantitative phyletics and the evolution of anurans. *Systematic Zoology* 18:1-32.

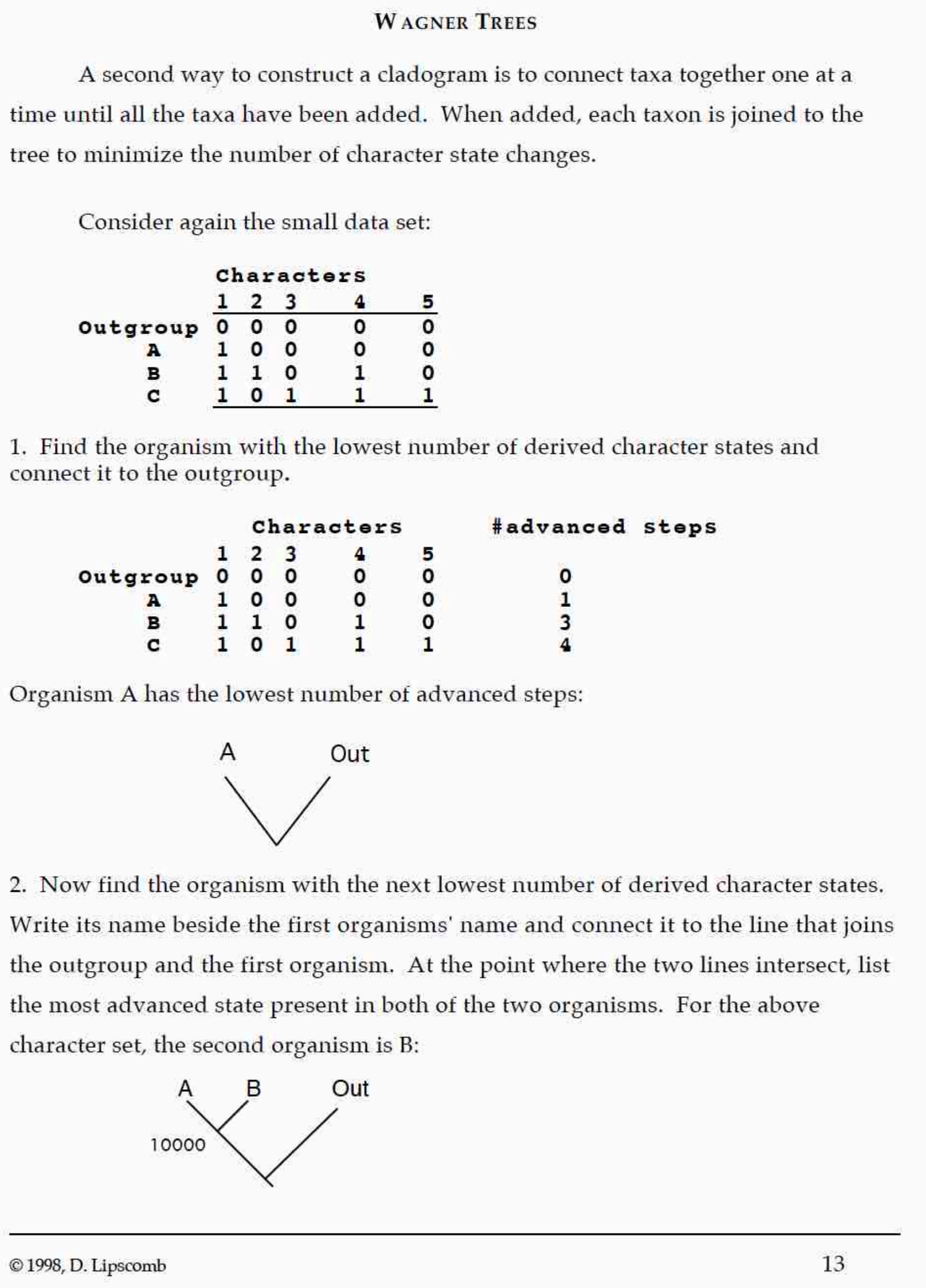
Farris, J. S. 1970. Methods for computing Wagner trees. *Systematic Zoology* 19:83-92.

Basics of Cladistic

pp. 13-18



Diana Lipscomb
George Washington University
Washington D.C.
Copywrite (c) 1998



WAGNER TREES

A second way to construct a cladogram is to connect taxa together one at a time until all the taxa have been added. When added, each taxon is joined to the tree to minimize the number of character state changes.

Consider again the small data set:

	Characters 1	2	3	4	5
Outgroup	0	0	0	0	0
A	1	0	0	0	0
B	1	1	0	1	0
C	1	0	1	1	1

1. Find the organism with the lowest number of derived character states and connect it to the outgroup.

	Characters 1	2	3	4	5	#advanced steps
Outgroup	0	0	0	0	0	0
A	1	0	0	0	0	1
B	1	1	0	1	0	3
C	1	0	1	1	1	4

Organism A has the lowest number of advanced steps:

2. Now find the organism with the next lowest number of derived character states. Write its name beside the first organisms' name and connect it to the line that joins the outgroup and the first organism. At the point where the two lines intersect, list the most advanced state present in both of the two organisms. For the above character set, the second organism is B:

© 1998, D. Lipscomb

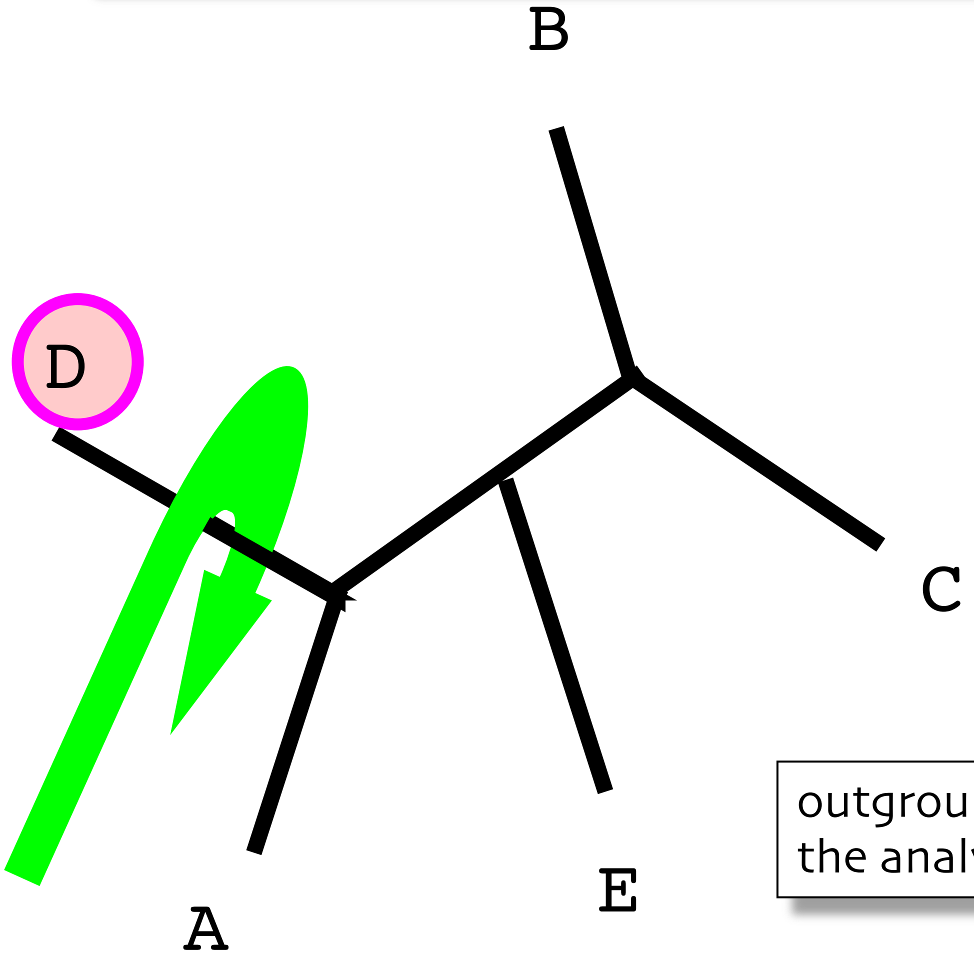
Wagner algorithm

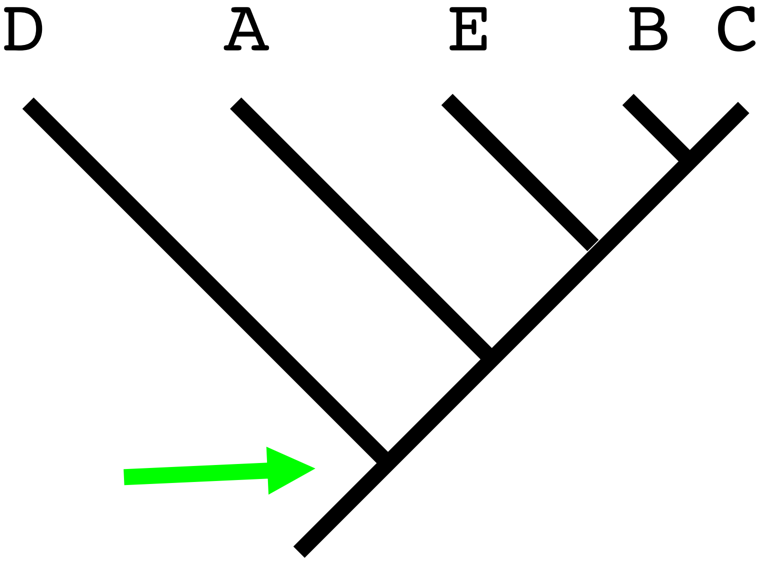
page 13:

1. Find the organism with the lowest number of derived character states and connect it to the outgroup

in practice ANY of the organisms can be used as outgroup

Nixon, K. & Carpenter, J. 1993. On outgroups. Cladistics 9: 413-426.





outgroup terminals should be treated in the analyses AS ANY OTHER TERMINAL

they should be included in the analyses FROM THE VERY START

Wagner algorithm

page 13:

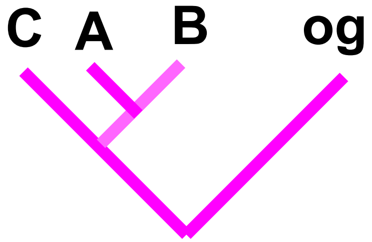
1. Find the organism with the lowest number of ~~derived character states~~ and connect it to the outgroup

in practice ANY of the organisms can be used as outgroup

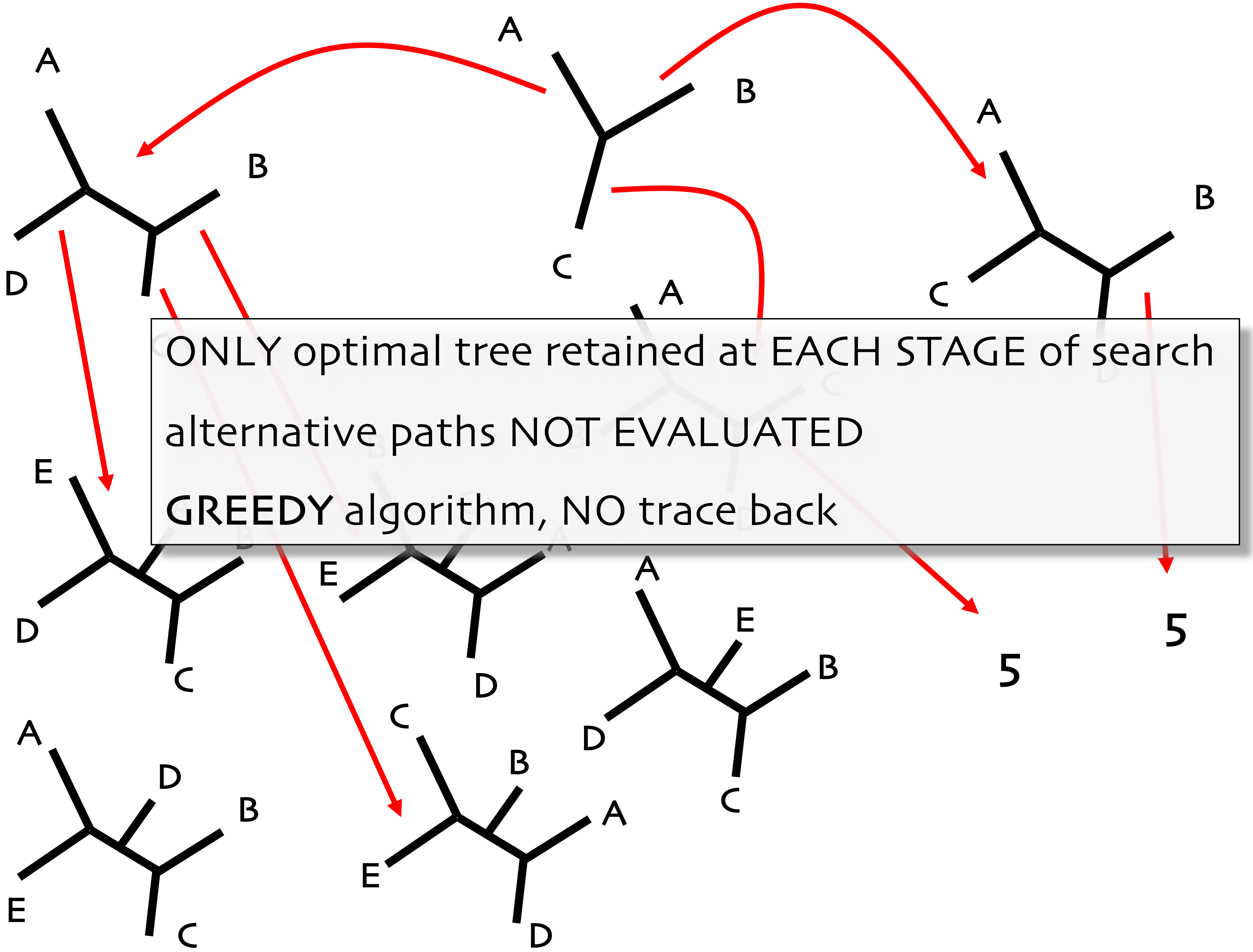
next search for a terminal to pair with it so that the number of observed **DIFFERENCES** between the two are minimized

	1	2	3	4	5	6	7	8	9	10	differences from outgroup
og	0	0	0	0	0	0	0	0	0	0	-
A	1	0	0	0	1	1	0	0	0	1	4
B	1	0	0	1	0	0	0	0	0	1	~~3~~
C	0	0	0	0	0	0	0	0	1	1	~~2~~
D	0	1	1	0	0	0	1	1	0	1	5
E	0	1	1	1	0	0	0	1	0	1	5

the problem of Wagner algorithm is that the order where terminals are added to tree affects the tree finally obtained



continued until all terminals are included



ONLY optimal tree retained at EACH STAGE of search

alternative paths NOT EVALUATED

GREEDY algorithm, NO trace back

HEURISTIC SEARCH

RAS = random addition sequence

1. starting trees built by adding terminals one by one

    matrix/random order

    ✓ Wagner algorithm

    phenetic clustering

all these can be combined to build starting trees

leads RARELY in finding shortest tree

HEURISTIC SEARCH

1. starting trees built by adding terminals one by one

WAGNER algorithm

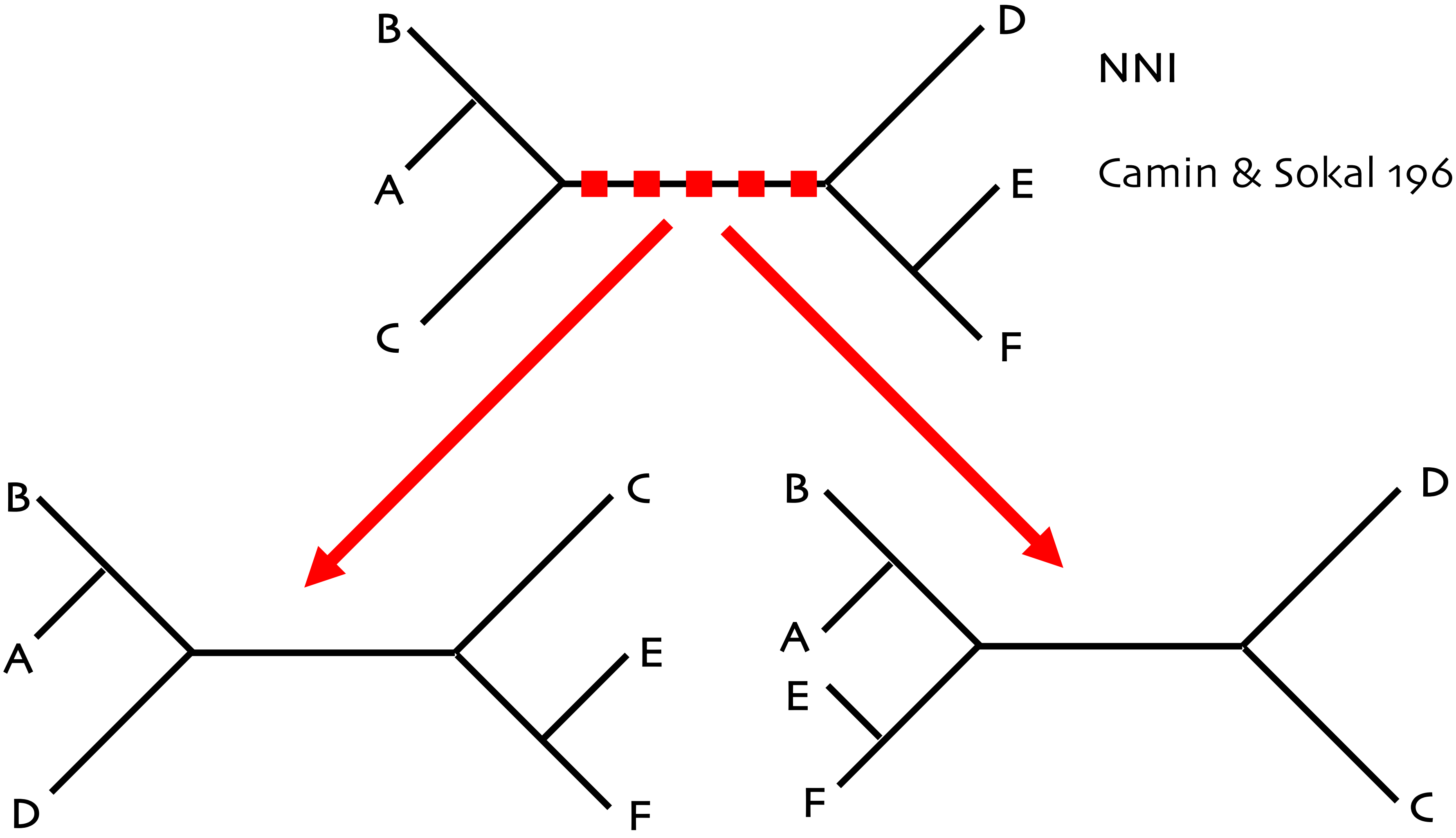
2. after this branches of the tree(s) are moved

branch SWAPPING

Nearest-neighbor interchange

NNI

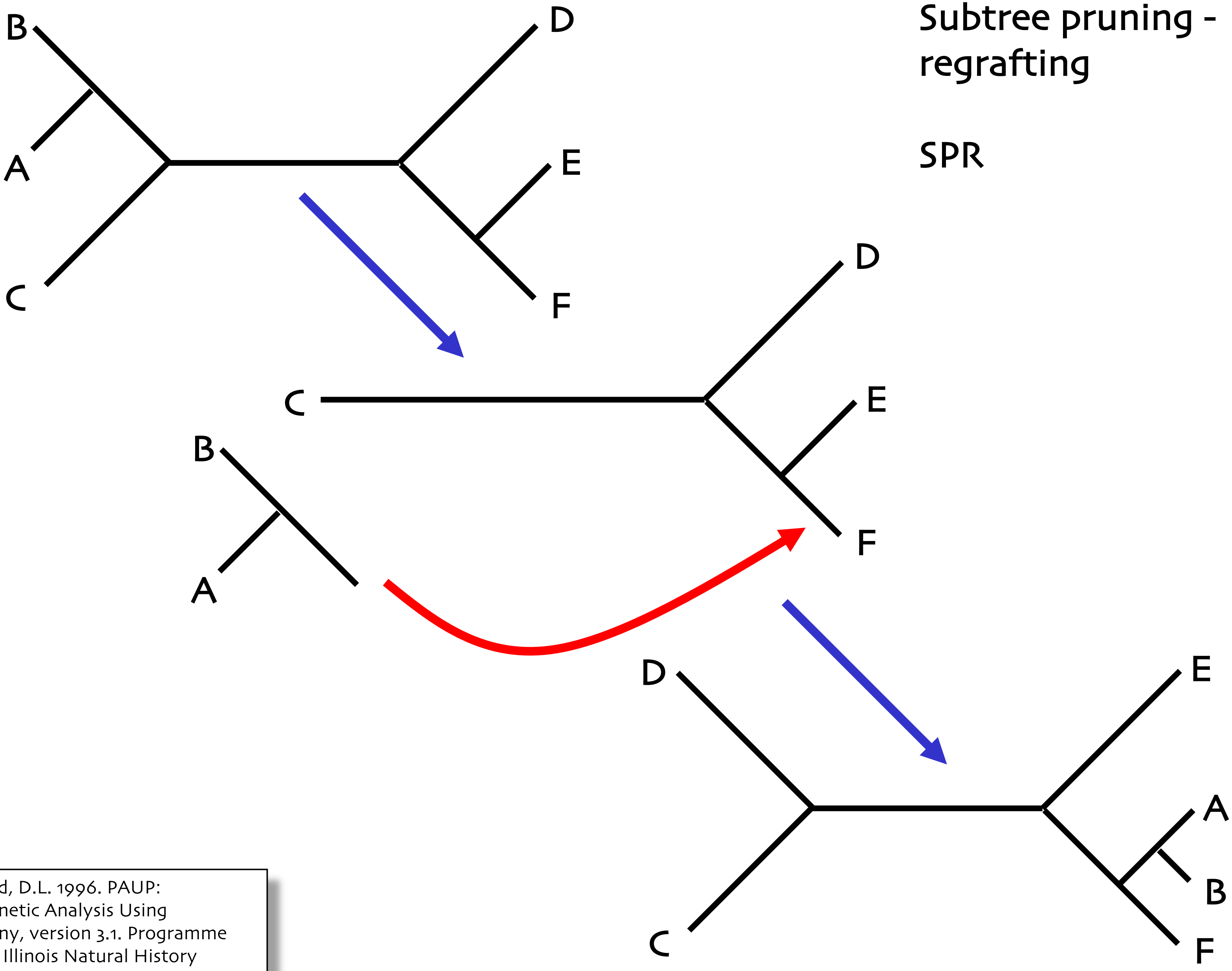
Camin & Sokal 1965



Swofford, D.L. 1996. PAUP: Phylogenetic Analysis Using Parsimony, version 3.1. Programme manual. Illinois Natural History Survey.

Subtree pruning - regrafting

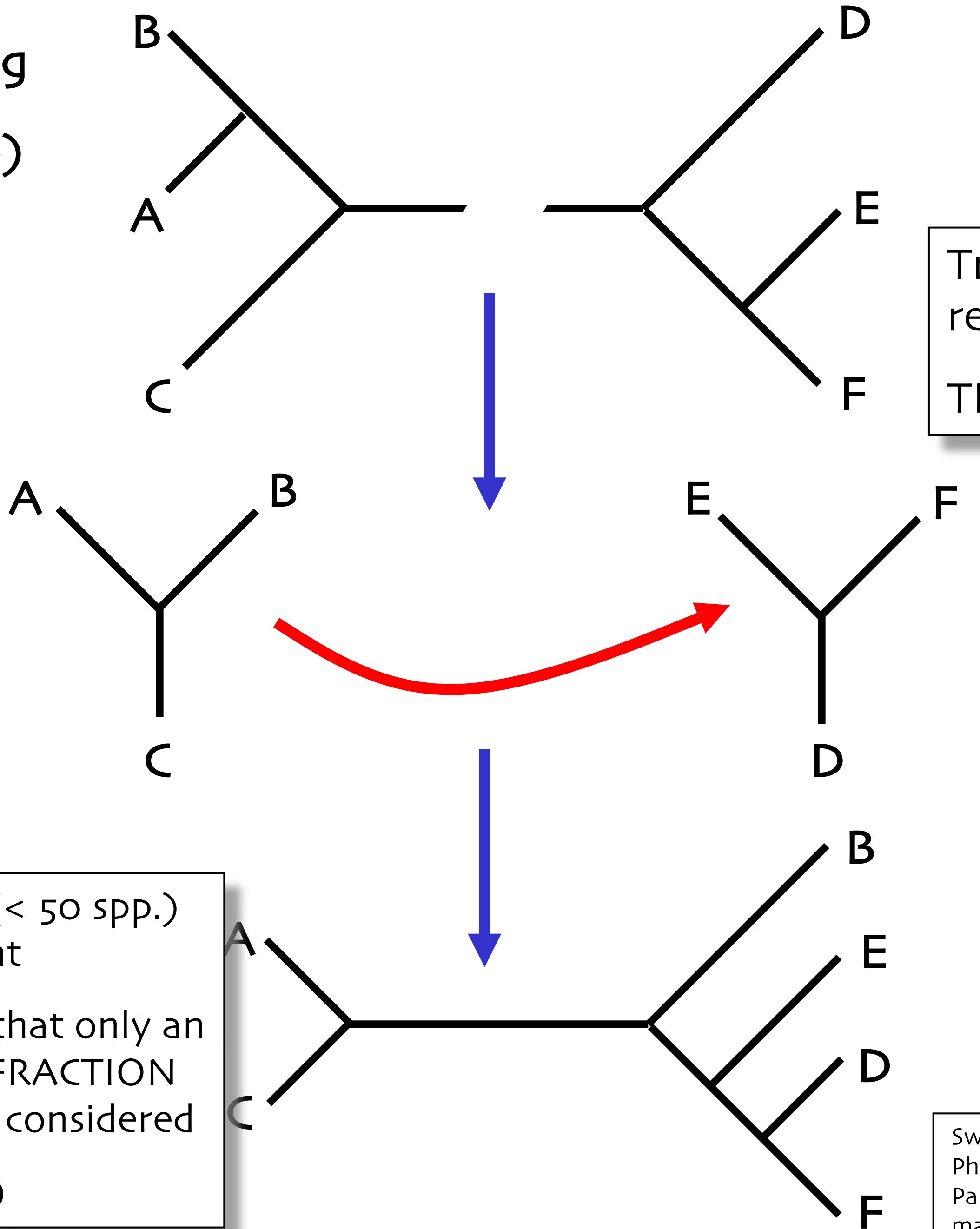
SPR



Swofford, D.L. 1996. PAUP: Phylogenetic Analysis Using Parsimony, version 3.1. Programme manual. Illinois Natural History Survey.

Branch breaking

BB (Farris 1970)



Tree bisection-reconnection

TBR (Swofford 1990)

for small matrices (< 50 spp.) surprisingly efficient

despite of the fact that only an extremely SMALL FRACTION of all possible trees considered

(50 spp. 1.0 x 10^{-69})

Swofford, D.L. 1996. PAUP: Phylogenetic Analysis Using Parsimony, version 3.1. Programme manual. Illinois Natural History Survey.

River
West New York
Broadway
East River Esplanade and Bikeway
Central Park
Astoria Park
Manhattan Island
Union City
Steinway
1st Ave
12th Ave
East River
Queens Blvd
The High Line
oboken
Greenwich Village
1st Ave
Newtown Creek
Haberman
Greenpoint
East River Park
English Kills
Bushwick Junction
New York

St. John, K. 2017. Review paper: the shape of phylogenetic treespace. *Syst. Biol*. 66: e83-e94

HEURISTIC SEARCH algorithms

NNI	$O\ (n)$
SPR	$O\ (n^2)$
TBR	$O\ (n^3)$

n = no. of lvs/terminals

....and in practice if we use RAS+TBR $O\ (n^4)$

Whidden, C. & Matsen IV, F.A. 2018. Efficiently inferring pairwise subtree prune-and-regraft adjacencies between phylogenetic trees. Proceedings of the meeting on analytical algorithmics and combinatorics: 77-91.

Wheeler 2017 (lecture notes CP217)

HEURISTIC SEARCH

1. starting trees build by adding terminals one by one

Wagner algorithm

2. after this branches of the tree(s) are moved

branch SWAPPING with NNI, SPR, TBR

3. current programs include algorithms that allow evaluation of tree length WITHOUT visiting ALL nodes of trees

save processing time

Goloboff, P.A. 1993. Character optimization and calculation of tree lengths. *Cladistics* 9: 433-436.

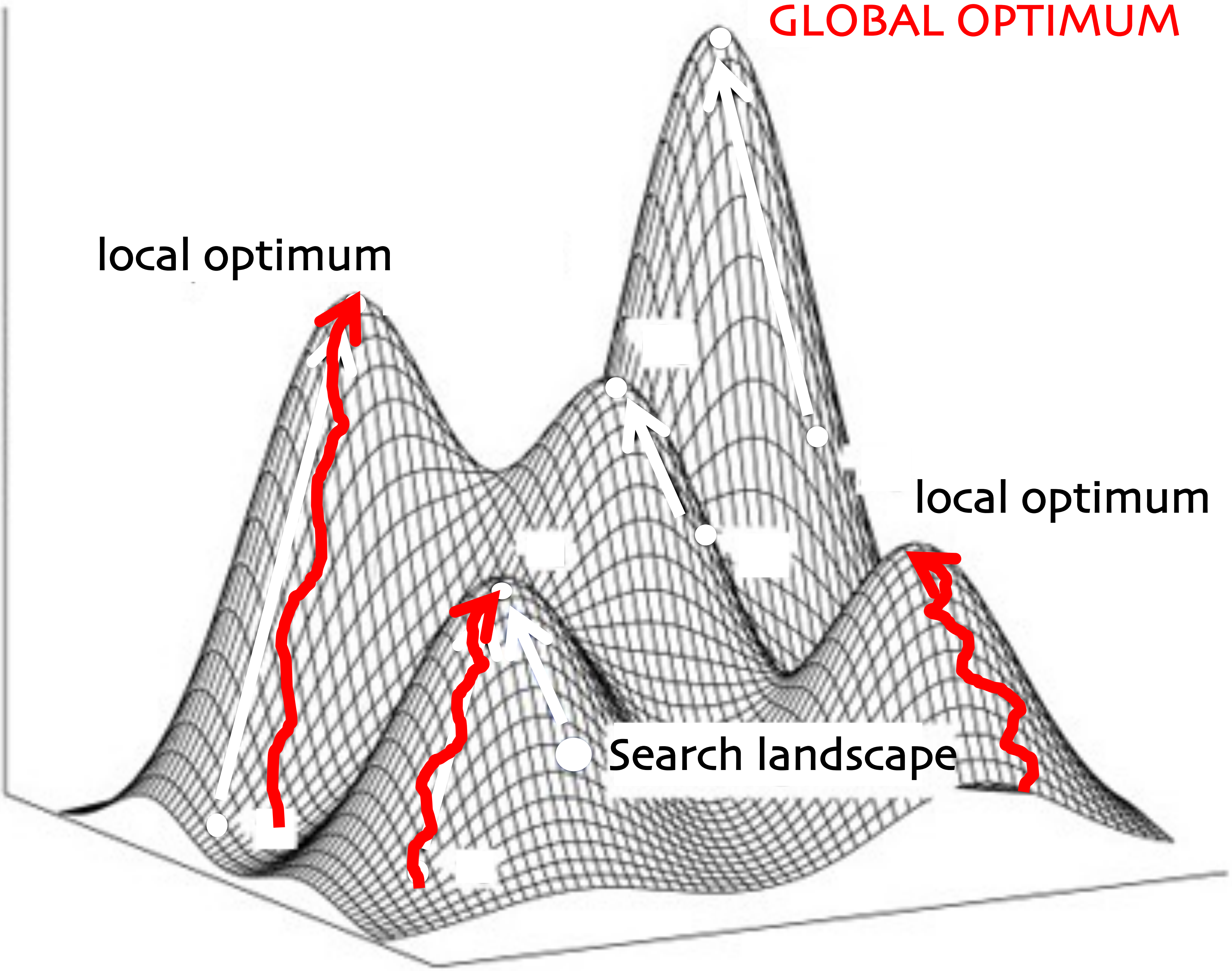
Search strategies

Maddison, D.R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. *Syst. Biol.* 40: 315-328.

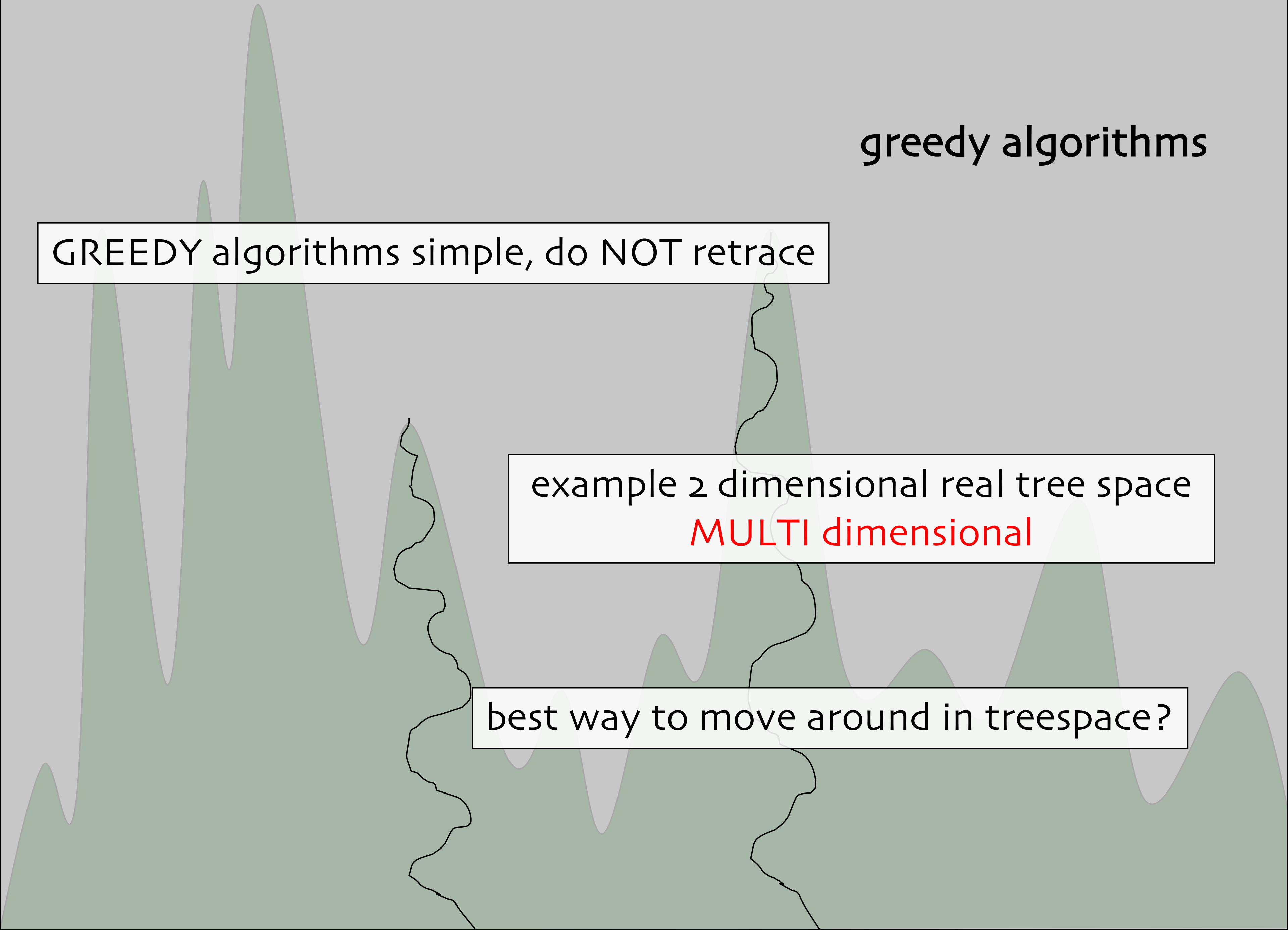
islands of trees in tree “space” (landscape)

problem especially in analyses of larger matrices (> 100-200 terminals)

http://genealogyreligion.net/wp-content/uploads/2010/04/intelligencelandscape.jpg



greedy algorithms



GREEDY algorithms simple, do NOT retrace

example 2 dimensional real tree space
MULTI dimensional

best way to move around in treespace?

TRADITIONAL SEARCH STRATEGIES

search repeated SEVERAL TIMES (10 - 10 000 x)

how many times is sufficient?

has optimal tree be found?

if first 100x give same result

STOP

if only 1/100 give optimal result

CONTINUE

moving branches takes time, thus the number of trees saved at intermediate stages of search kept SMALL, e.g. ≤ 2-10

WHY?

when starting search potentially VERY DIFFERENT trees, after moving branches this is NOT anymore true, no reason to keep MANY SIMILAR trees

most likely will lead finding same optimal tree

TRADITIONAL SEARCH STRATEGIES

if > 10% of random searches give same result ---> search should be extended by expanding the number of trees saved in cache memory (50-100)

if < 10% random searches give same results ---> continue with NUMEROUS (500-1 000) additional random search repetitions with saving few trees in cache

finally 5-10% of trees saved in cache (50-200) analyzed in more detail

Davis, J. & al. 2005. The limits of conventional cladistic analysis. In: Albert, V. (ed.) *Parsimony, phylogeny and genomics*. 229 s. Oxford University Press

TRADITIONAL SEARCH STRATEGIES

NO NEED to find ALL optimal trees (Farris & al. 1996)

for large matrices > 10^6 equally parsimonious trees might exist

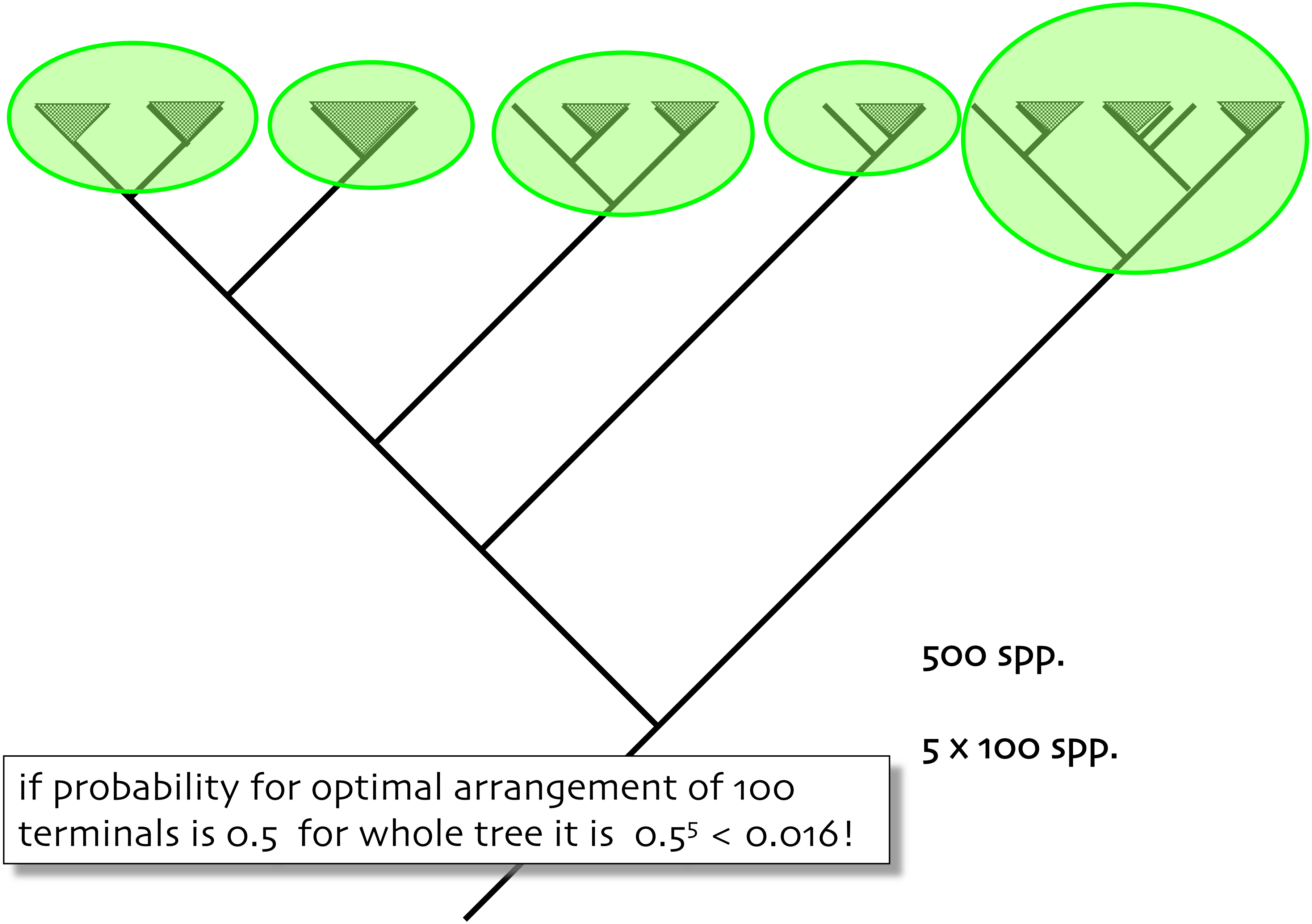
consensus tree based only on small sample of trees might be identical with the one based on ALL trees if those sampled originate from SEPARATE tree groups (islands)

can also be used to "guide" our search, i.e. if consensus does not change, no need to continue

NEW SEARCH STRATEGIES

required for analyses of LARGE (> 500-700 spp.) matrices

in large trees ALL parts should be in optimal arrangement
in order to be optimal as a WHOLE



if probability for optimal arrangement of 100 terminals is 0.5 for whole tree it is $0.5^5 < 0.016$!

GOLOBOFF 2000: dice comparison

www.weld-re1.k12.co.us/webclass/students/reamj/web/dice.gif

NEW SEARCH STRATEGIES

Moilanen, A. 1999. Searching for most parsimonious trees with simulated evolutionary optimization. *Cladistics* 15: 39-50.

Nixon, K.C. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* 15: 407-414.

Goloboff. P.A. 1999. Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics* 15: 415-428.

Nixon, K. Parsimony ratchet

diversityoflife.org/winclada/

1a. starting tree (e.g. Wagner algorithm)
1b. continued using fast & simple branch swapping

weight = cost of ch. state change

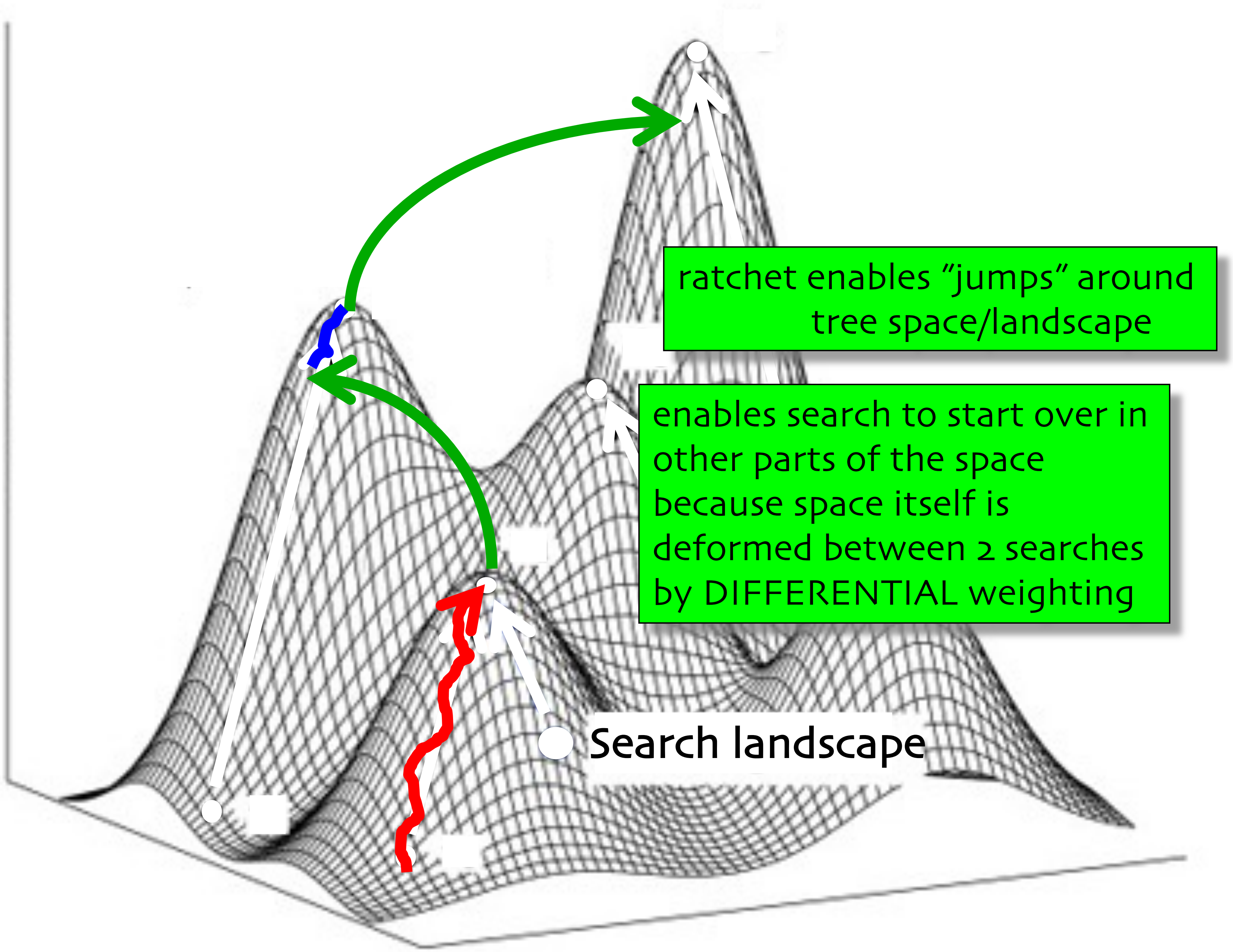
2. weight randomly e.g. 10-30% of characters
3. try to find shortest tree of this NEW matrix (as in 1b)

4. return ORIGINAL weights

5. search for the shortest tree (as in 1b)
6. return to 2. and repeat

> 30x faster than traditional search strategies

http://genealogyreligion.net/wp-content/uploads/2010/04/intelligencelandscape.jpg



ratchet enables “jumps” around tree space/landscape

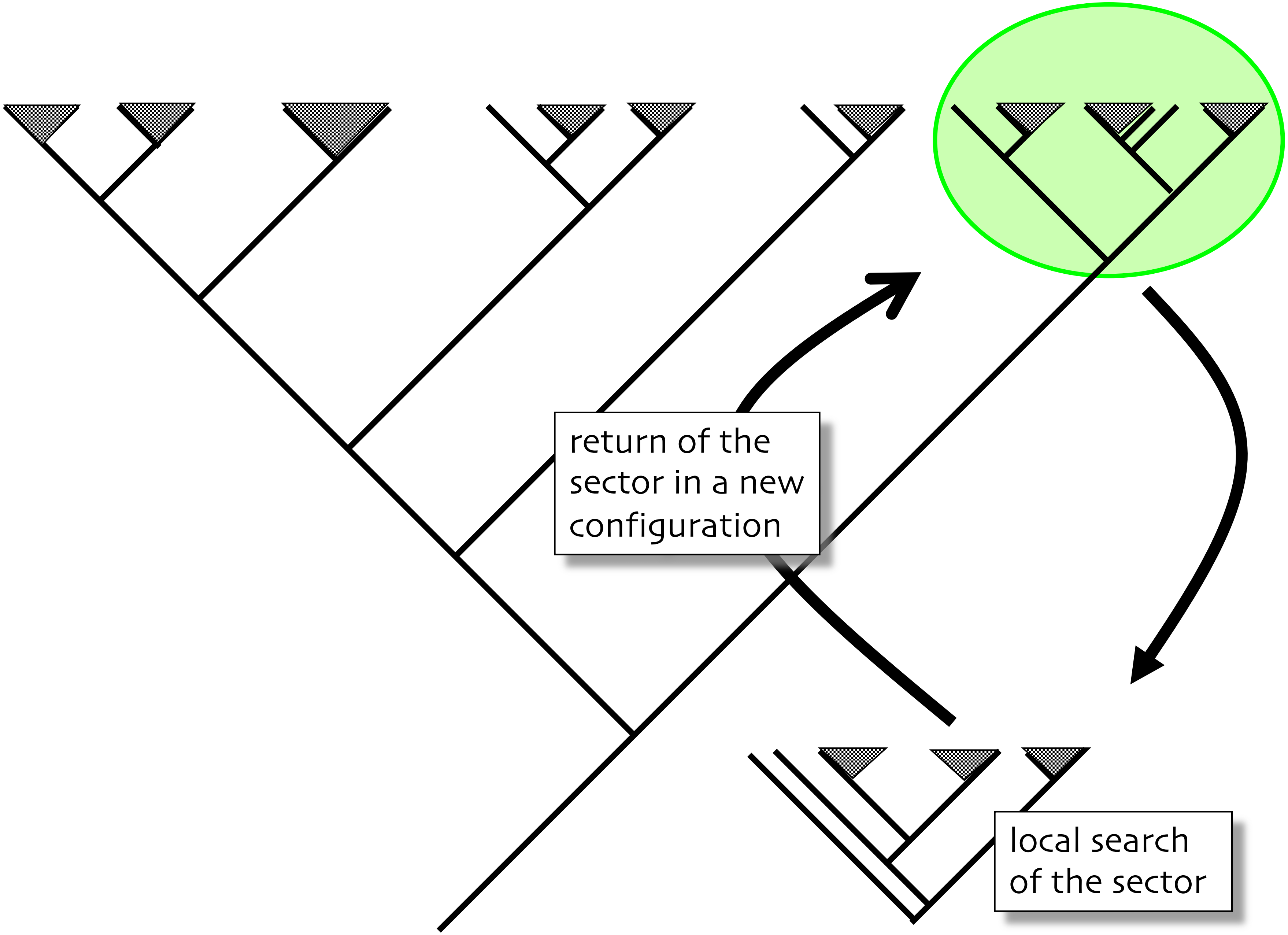
enables search to start over in other parts of the space because space itself is deformed between 2 searches by DIFFERENTIAL weighting

Search landscape

Goloboff, P. Sectorial search

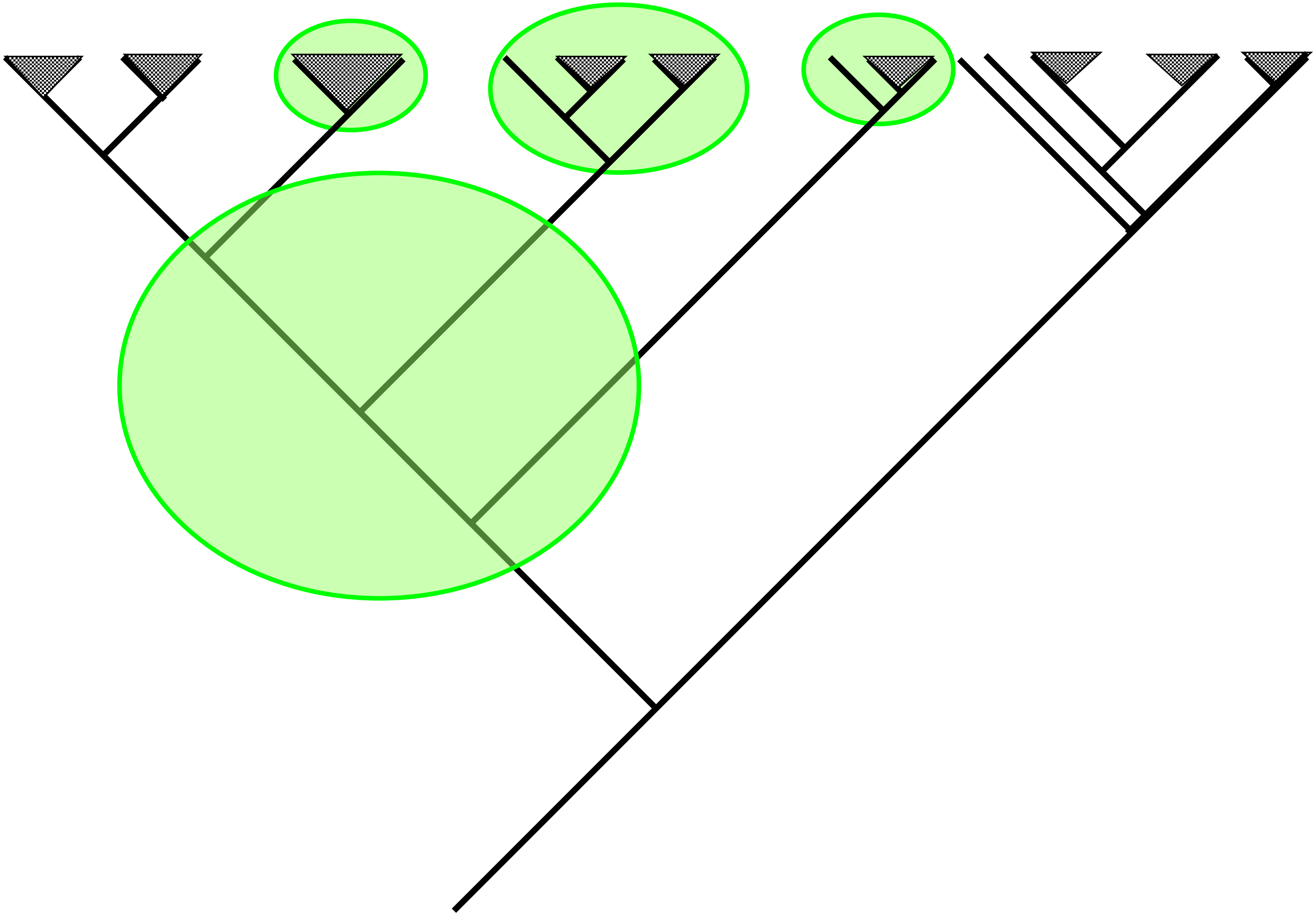
1. part (a sector) of the tree found by traditional search chosen

2. a LOCAL search performed in this part & returned to original tree



return of the sector in a new configuration

local search of the sector



Goloboff, P. Sectorial search

1. part (a sector) of the tree found by traditional search chosen

2. a LOCAL search performed in this part & returned to original tree

3. search performed numerous times with different parts

4. leads only rarely in finding optimal solution but MUCH faster than TBR

alternatives: RSS (random sectorial searches)

CSS (consensus-based sectorial s.)

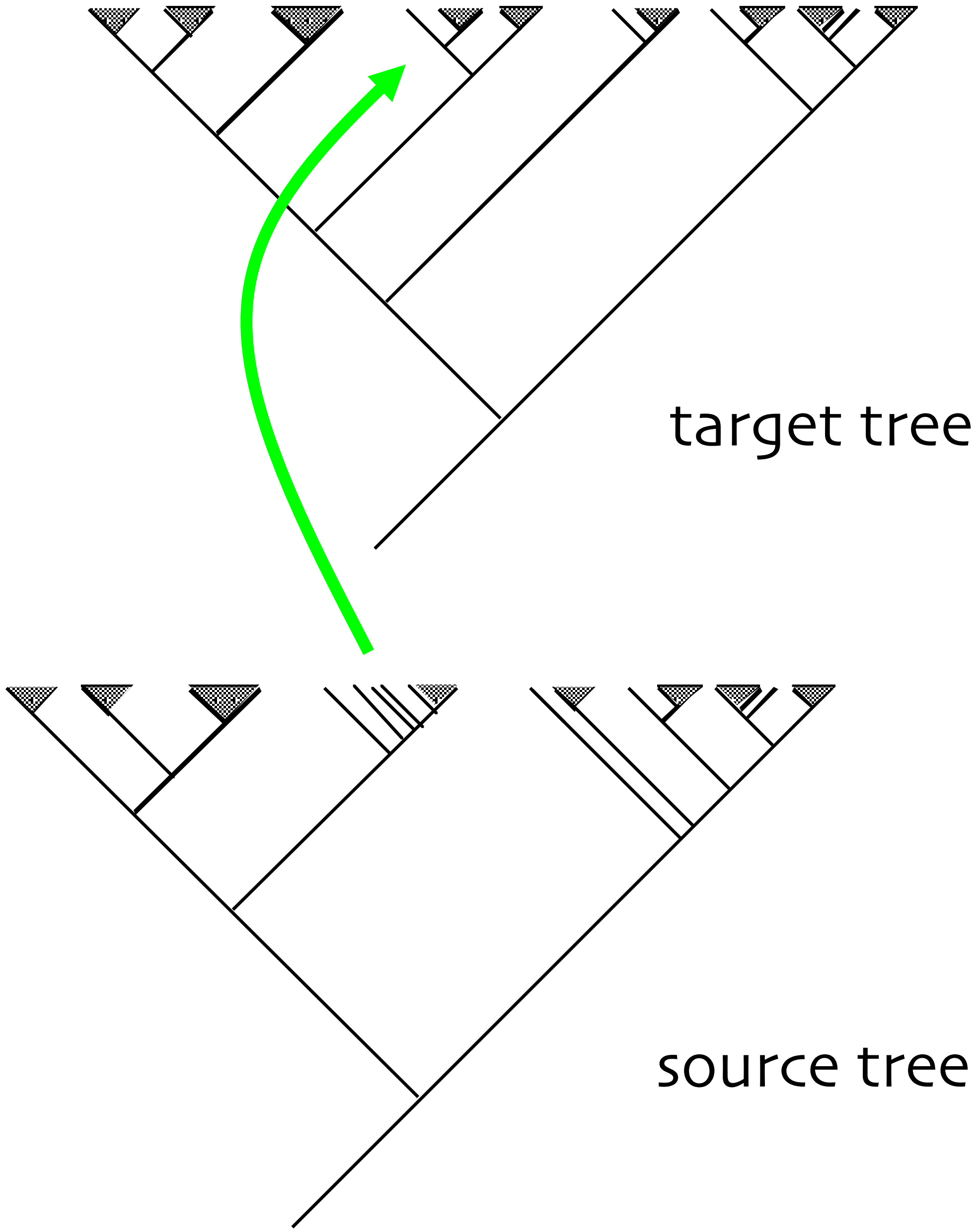
MSS (mixed sectorial s.)

CPU time increases LINEARLY in relation to number of sectors

EXPONENTIALLY in relation to number of terminals in TBR

Goloboff, P. & Moilanen, A. Tree fusing

1. 2 starting trees chosen
2. trees compared sector by sector
3. all sectors of source tree that make target tree shorter are transferred to this latter tree


target tree
source tree

Goloboff, P. (& Moilanen, A.) Tree fusing

1. 2 starting trees chosen
2. trees compared one sector at a time
3. all sectors that reduce tree length transferred from source to target tree
4. a new source tree chosen

initially trees resulting from numerous searches needed

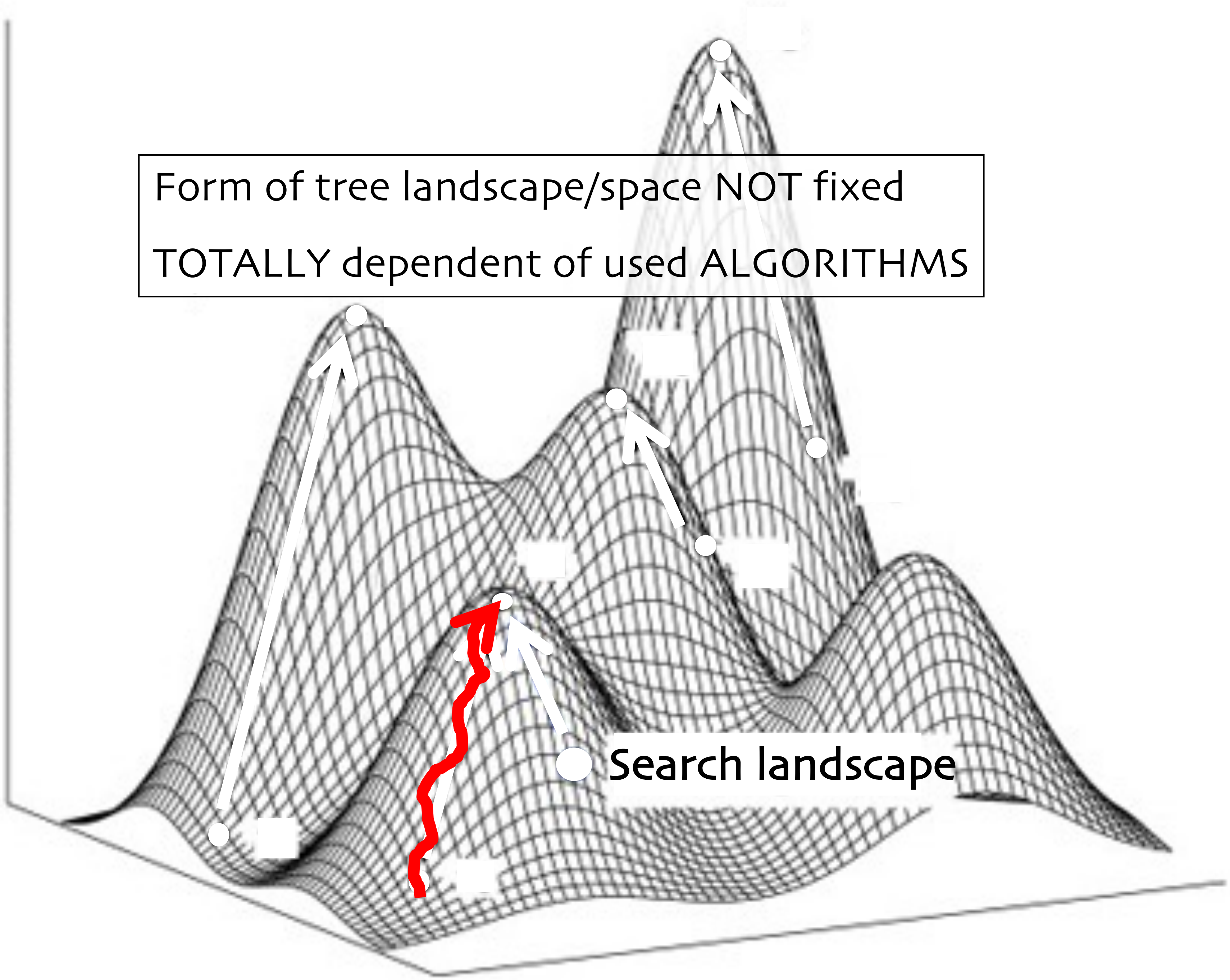
efficiency of the method based on the fact that at least one part of the tree is in OPTIMAL configuration

Goloboff, P. Tree drifting

longer than optimal trees accepted with predefined probability

widely known as “simulated annealing” used for analyses of difficult optimization problems

http://genealogyreligion.net/wp-content/uploads/2010/04/intelligencelandscape.jpg



Form of tree landscape/space NOT fixed

TOTALLY dependent of used ALGORITHMS

Search landscape

WILEY-
BLACKWELL

Cladistics 25 (2009) 1–20

Cladistics

10.1111/j.1096-0031.2009.00255.x

Phylogenetic analysis of 73 060 taxa corroborates major eukaryotic groups

Pablo A. Goloboff[a,*], Santiago A. Catalano[b], J. Marcos Mirande[b], Claudia A. Szumik[a], J. Salvador Arias[a], Mari Källersjö[c] and James S. Farris[d]

[a]*INSUE (Instituto Superior de Entomología), CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Instituto Miguel Lillo, Miguel Lillo 205, 4000 S.M.Tucumán, Argentina;* [b]*CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Fundación Miguel Lillo, Miguel Lillo 251, 4000 S.M.Tucumán, Argentina;* [c]*Göteborgs Botaniska Trädgård (Gothenburgh Botanical Garden), Carl Skottbergs Gata 22A, SE-413 19 Göteborg, Sweden;* [d]*Molekylärsystematiska laboratoriet, Naturhistoriska riksmuseet, Box 50007, 104-05 Stockholm, Sweden*

Accepted 21 February 2009

Abstract

Obtaining a well supported schema of phylogenetic relationships among the major groups of living organisms requires considering as much taxonomic diversity as possible, but the computational cost of calculating large phylogenies has so far been a major obstacle. We show here that the parsimony algorithms implemented in TNT can successfully process the largest phylogenetic data set ever analysed, consisting of molecular sequences and morphology for 73 060 eukaryotic taxa. The trees resulting from molecules alone display a high degree of congruence with the major taxonomic groups, with a small proportion of misplaced species; the combined data set retrieves these groups with even higher congruence. This shows that tree-calculation algorithms effectively

Concluding remarks

solutions to find optimal trees varies according to the data analyzed

e.g. extensive homoplasy vs. **randomly sparse** matrices

increasingly common with large genetic materials

Wagner algorithm used initially to find starting trees modified

selected/informative addition

Goloboff, P.A. 2014. Hide and vanish: data sets where the most parsimonious tree is known but hard to find, and their implications for tree search methods. *Mol. Phyl.& Evol.* 79: 118-131.

7.xi.

1. search algorithms

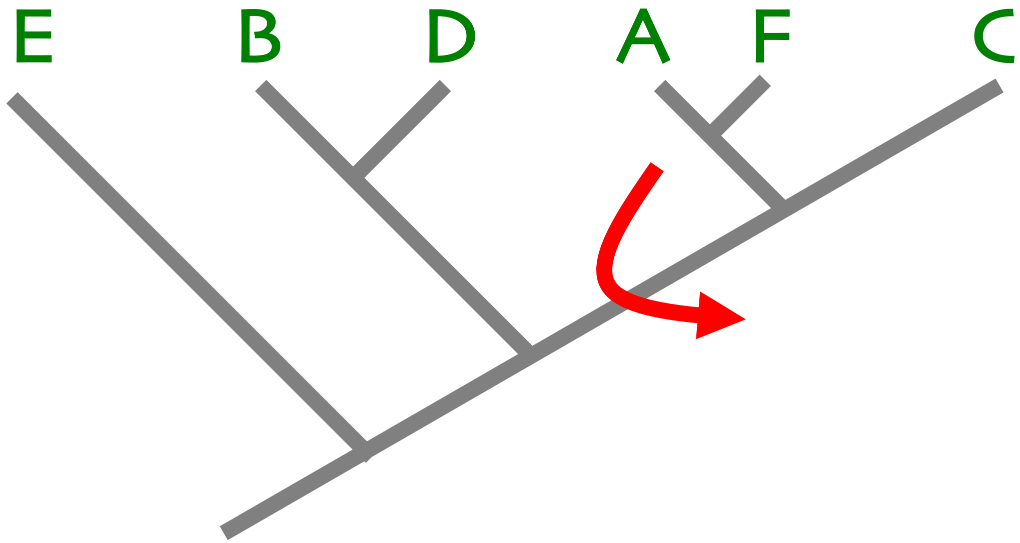
2. traditional search strategies

3. new search strategies

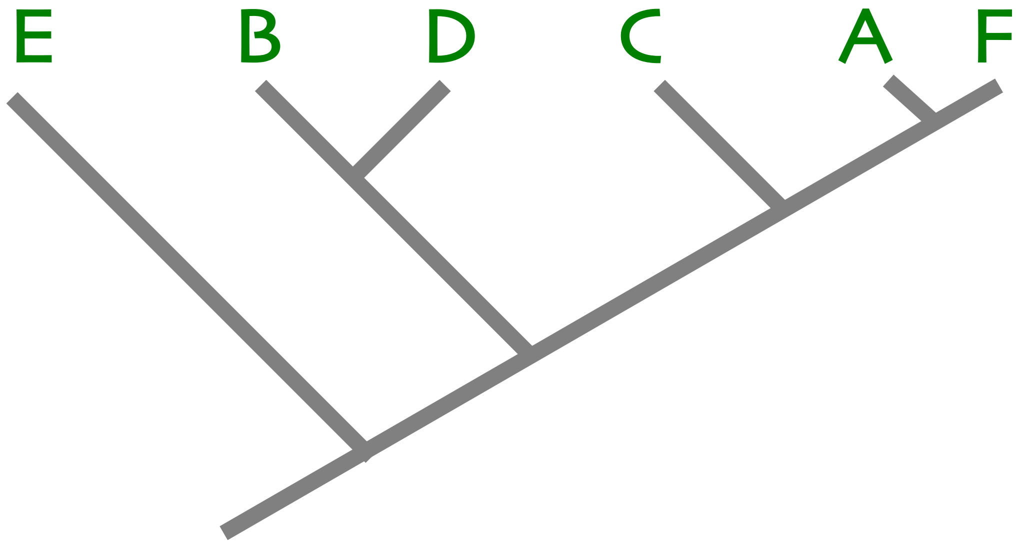
4. **tree form and consensus & compromise trees**

5. summary

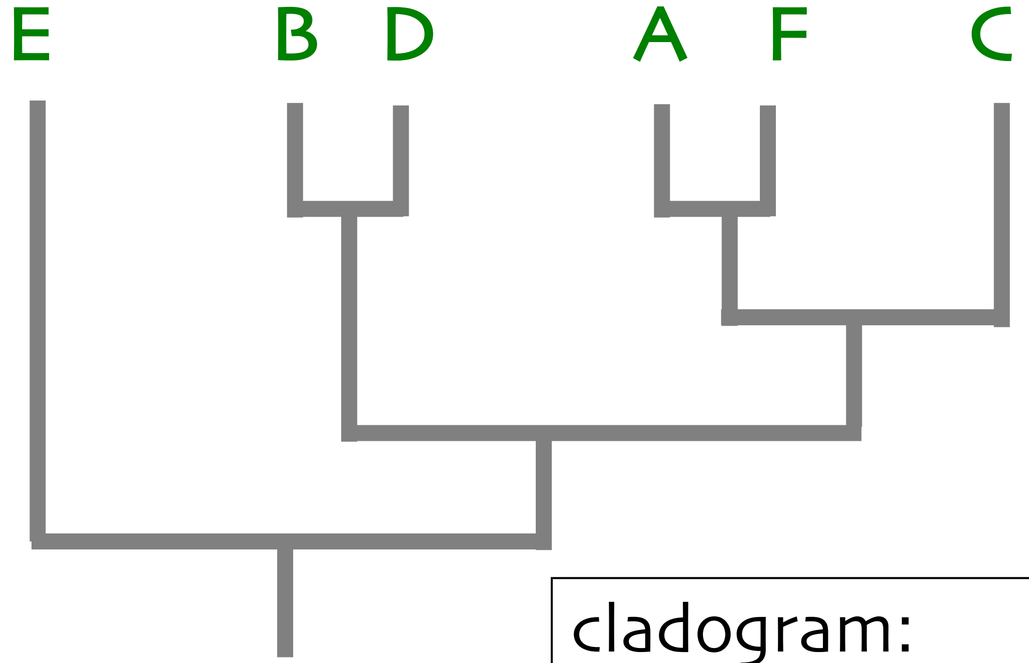
TREES & their form



TREES & their form



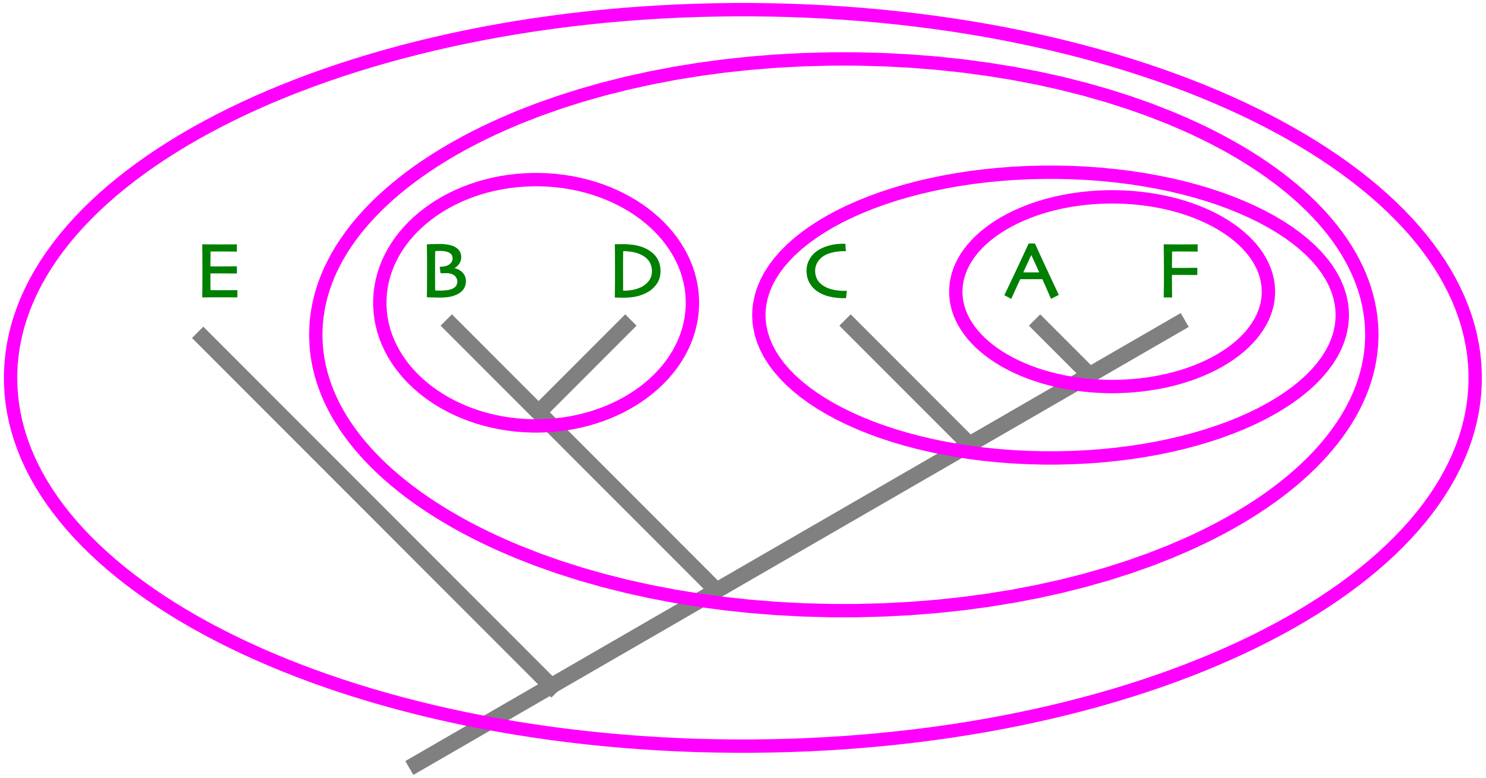
TREES & their form



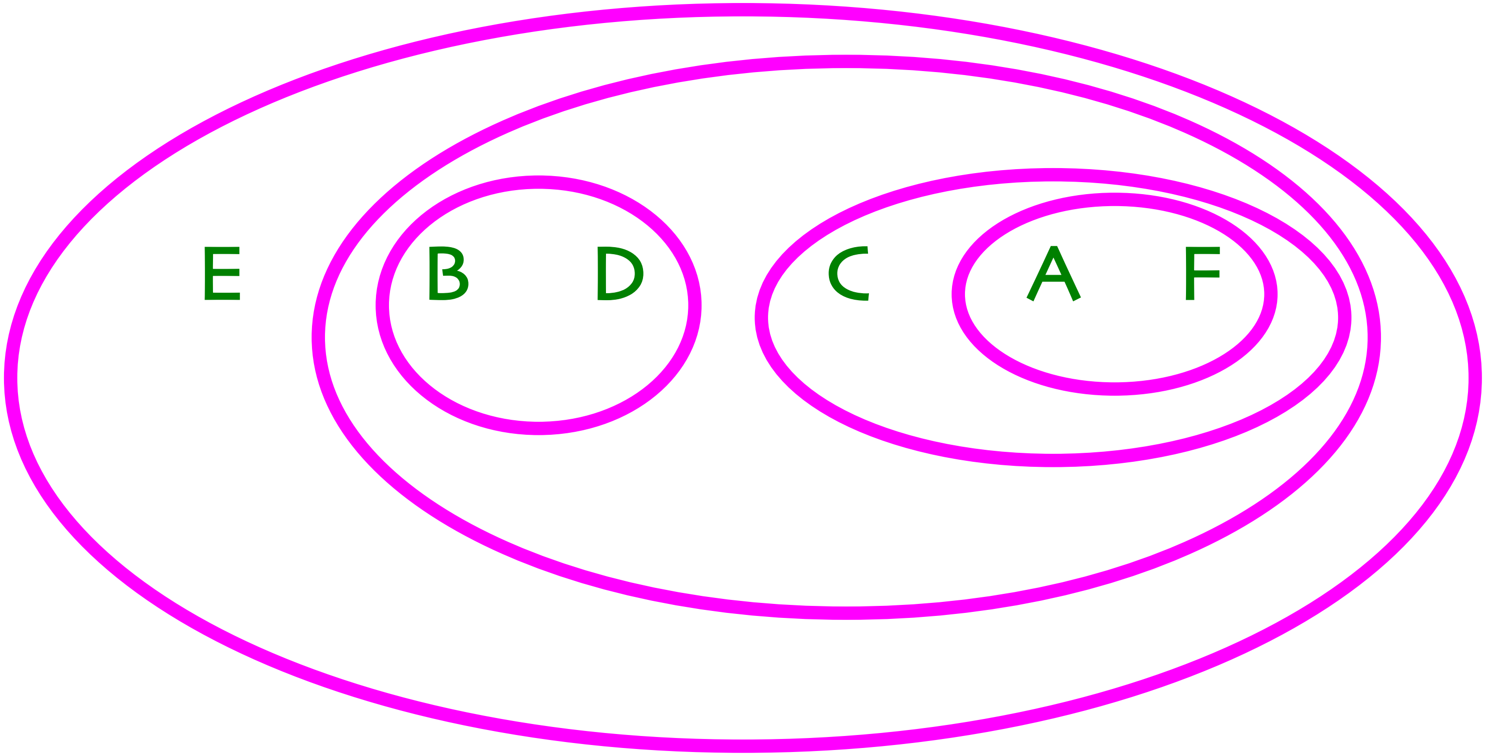
cladogram:

only branching ORDER matter

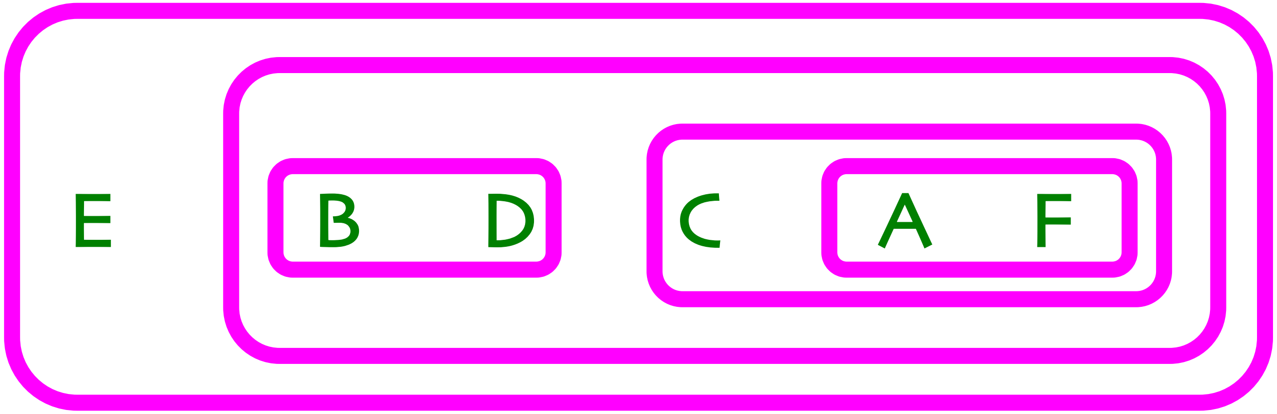
CLADOGRAMS & their form



CLADOGRAMS & their form

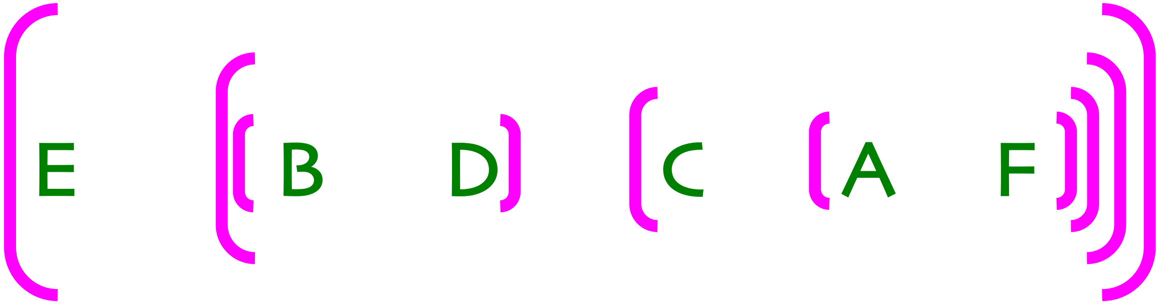


CLADOGRAMS & their form



Venn diagram

CLADOGRAMS & their form



From Venn diagram ---->

CLADOGRAMS & their form

(E ((B D) (C (A F))))

parenthetical notation

CLADOGRAMS & their form

Enables presentation of trees as part of normal text

(E ((B D) (C (A F))))

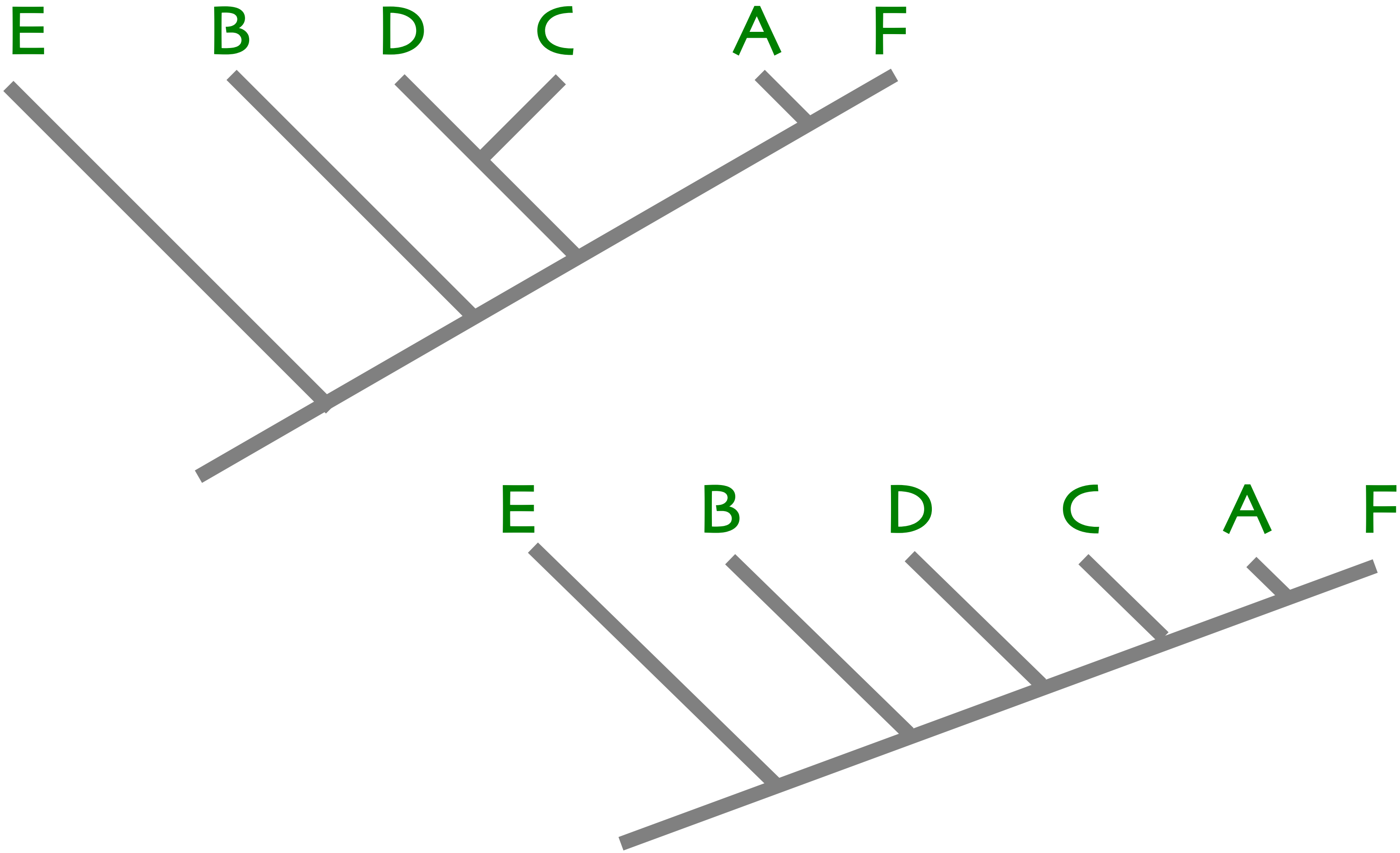
naturally used also in programming

Consensus & compromise trees

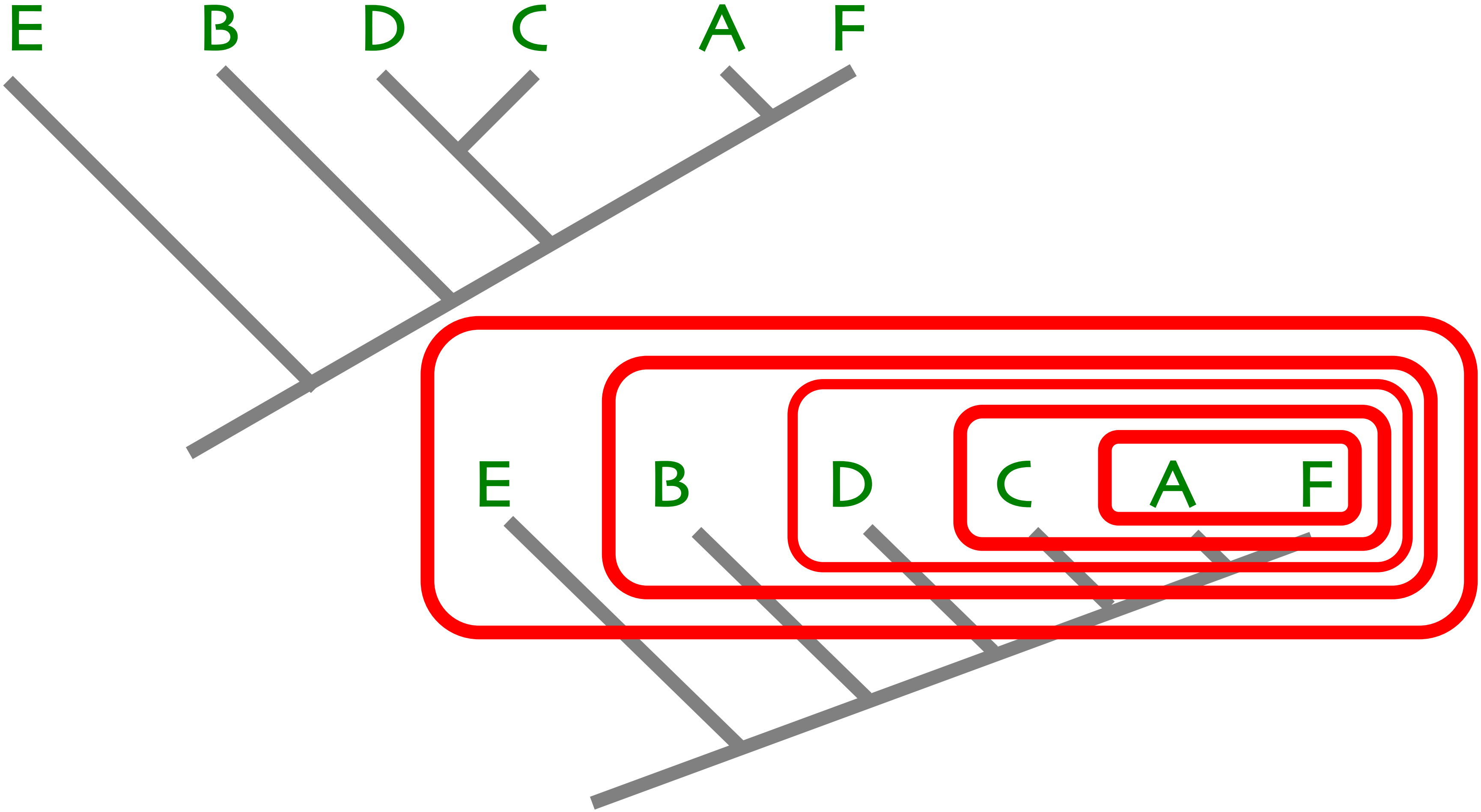
ONLY trees treated & compared, **NOT CHARACTERS**

Consensus trees

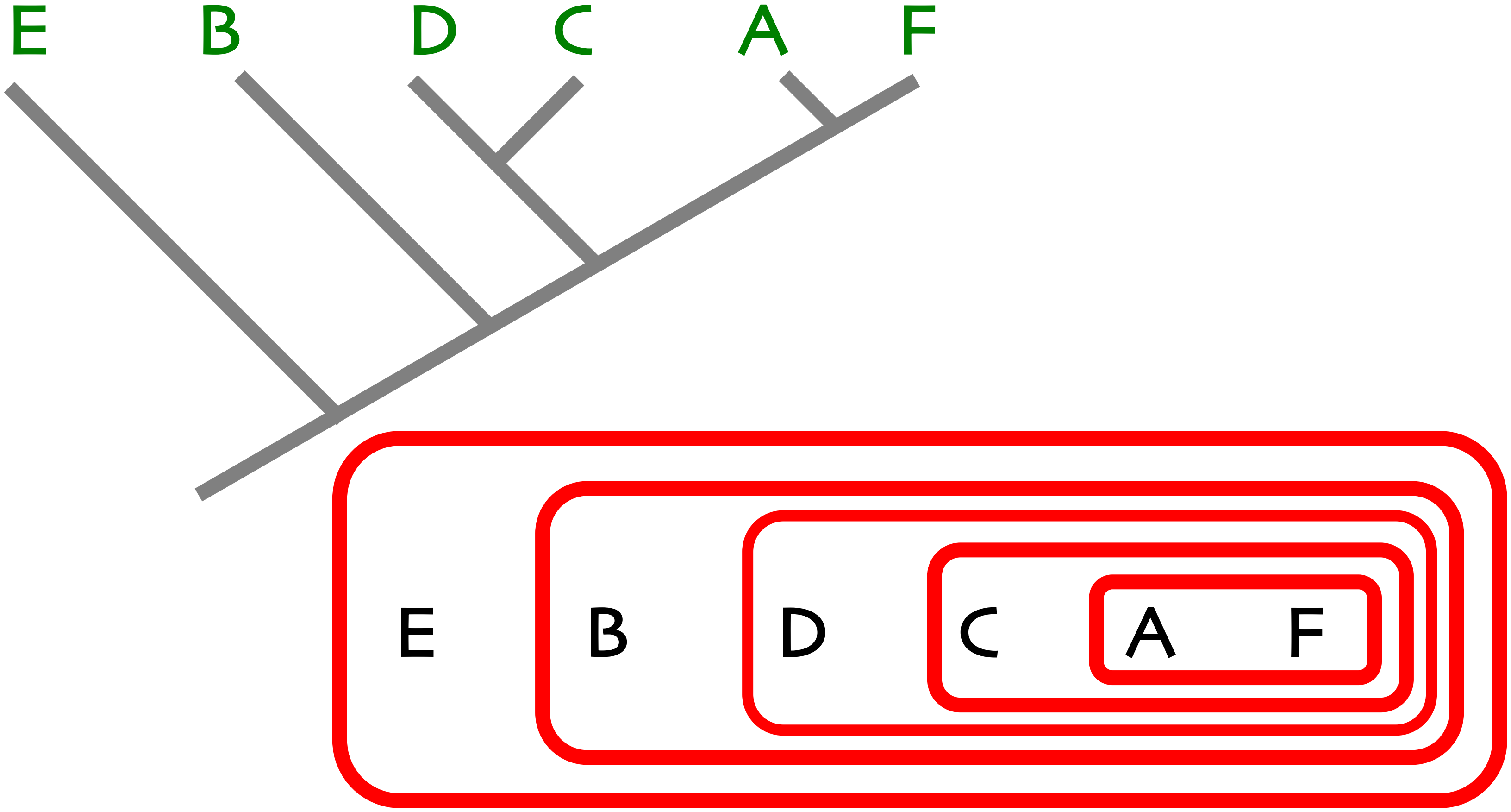
Sokal, R. R. & Rohlf, F. J. 1981. Taxonomic congruence in the Leptopodomorpha re-examined. *Systematic Zoology* 30: 309-325.



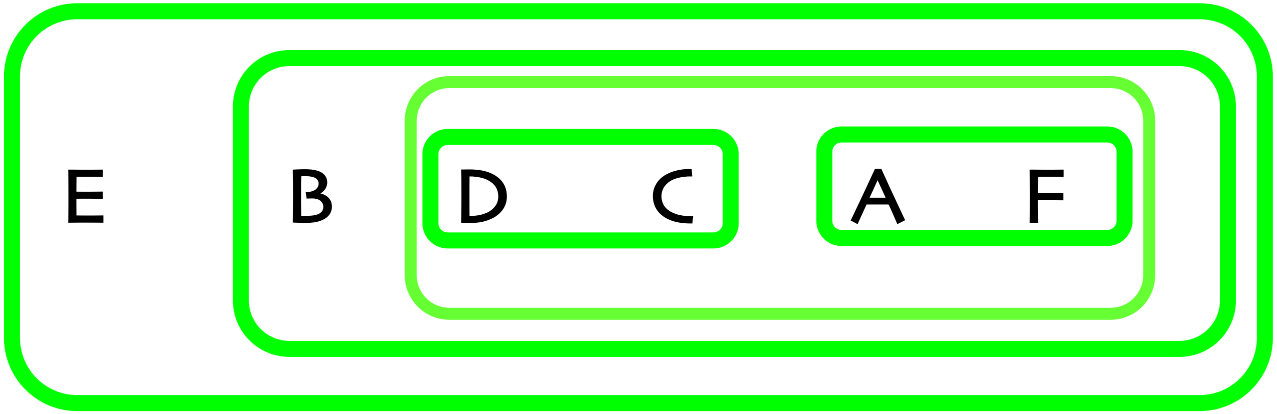
Consensus trees

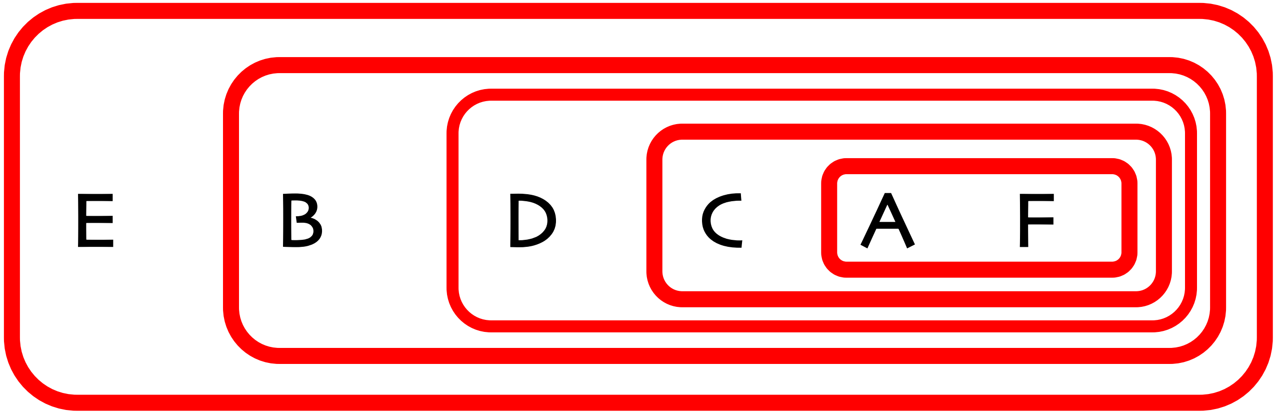


Consensus trees

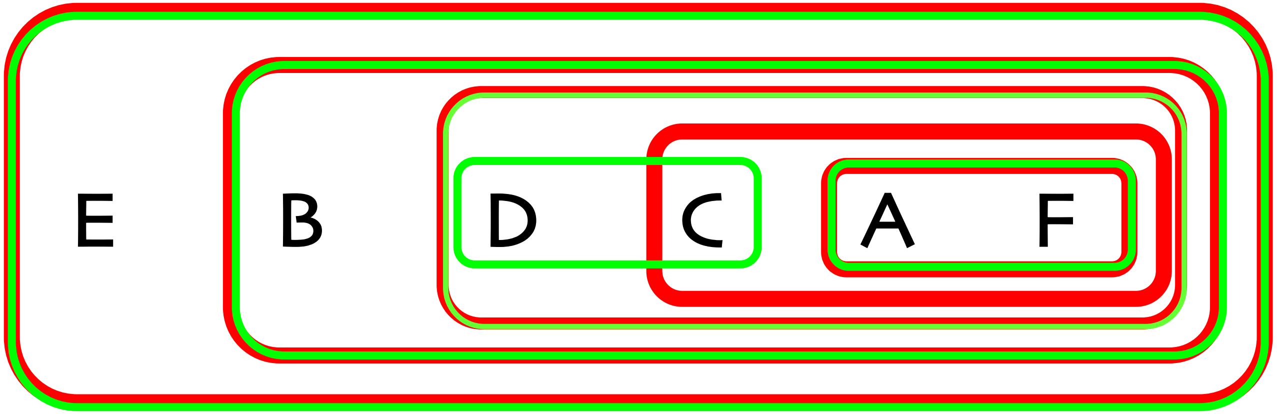


Consensus trees

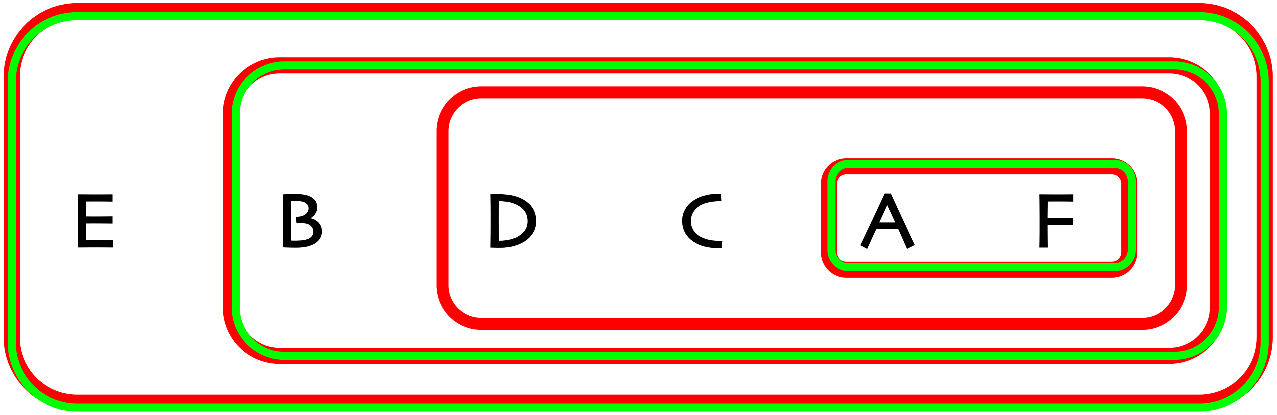




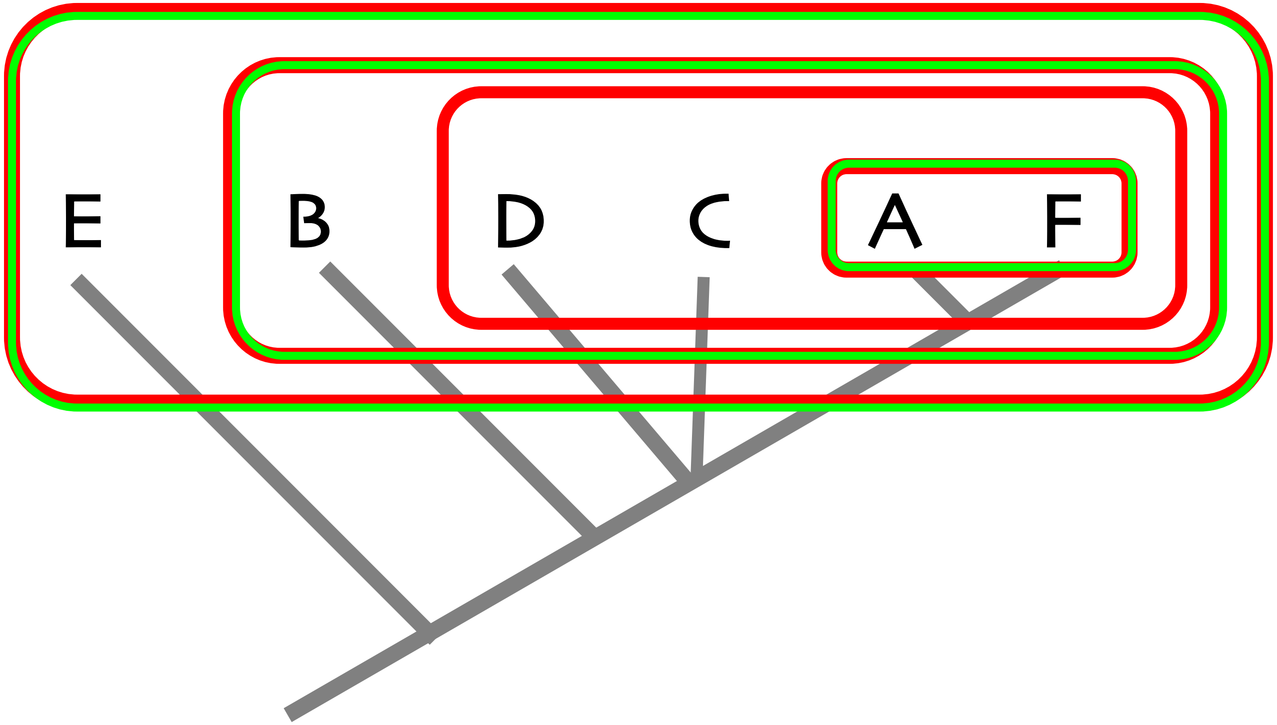
Consensus trees



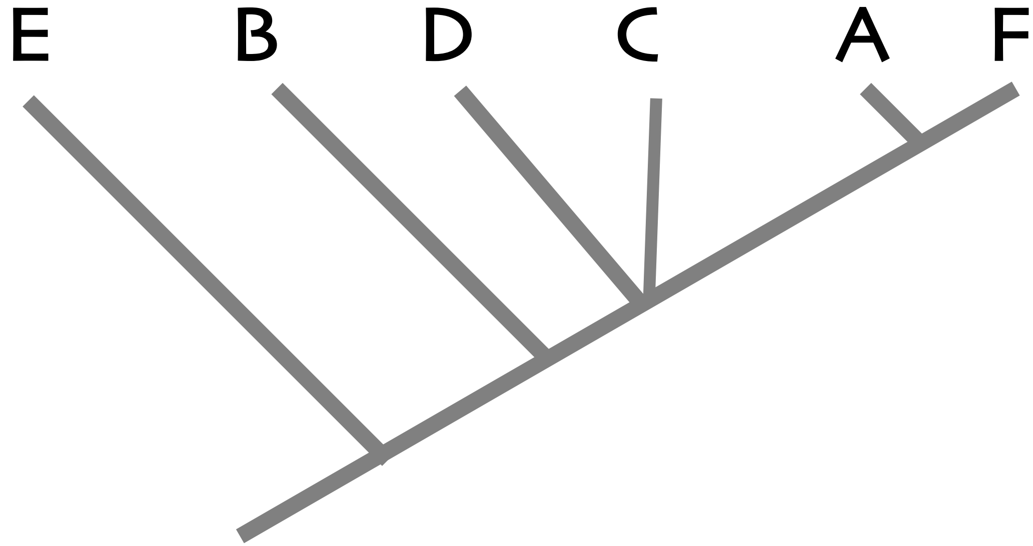
Consensus trees



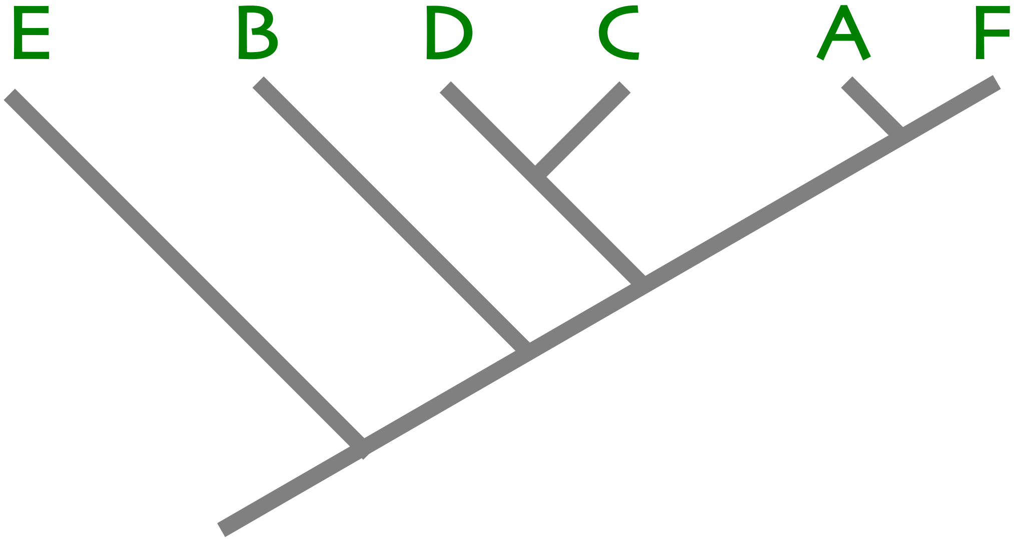
Consensus trees



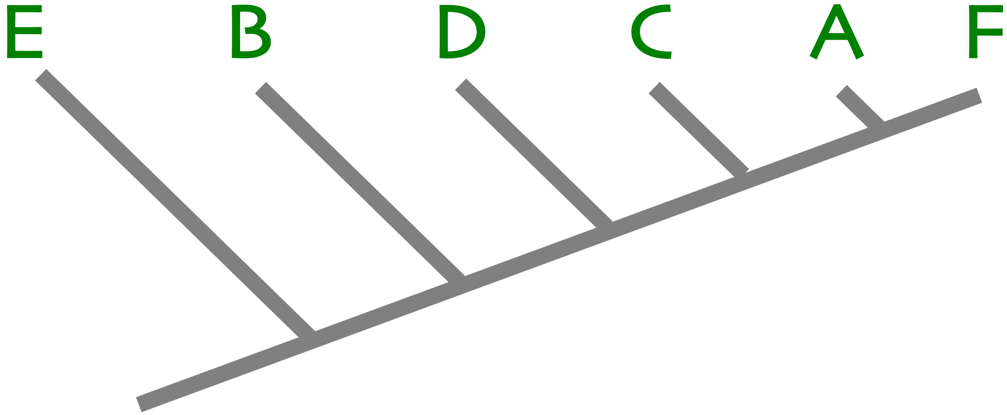
Consensus trees



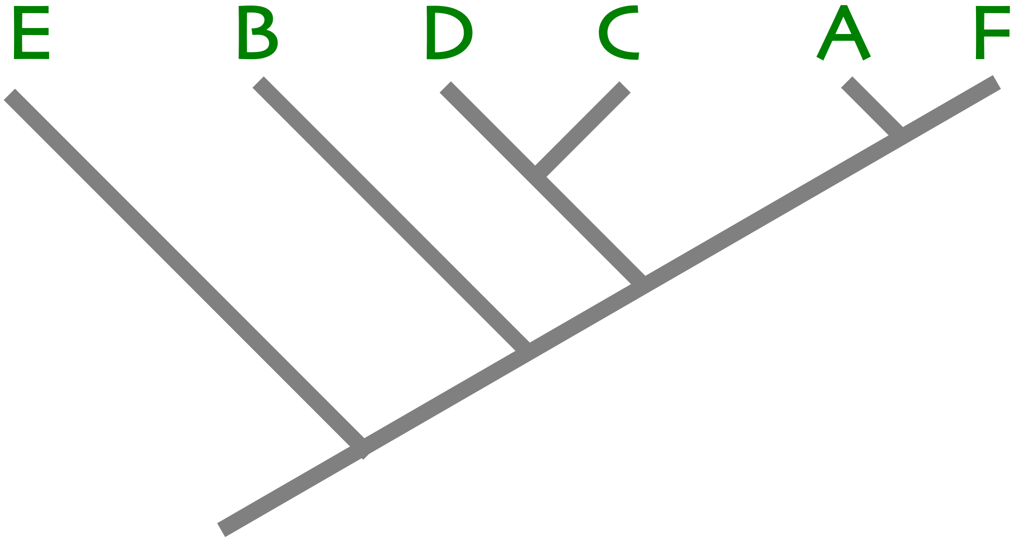
Consensus trees



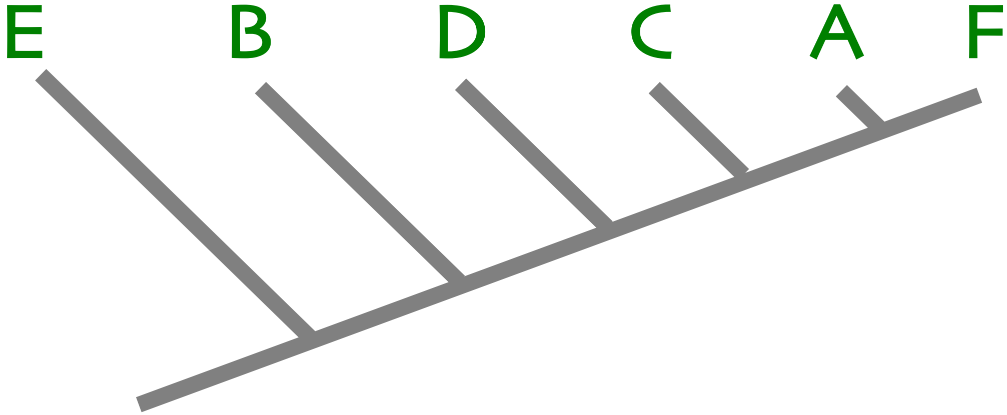
(E (B (D C) (A F))))



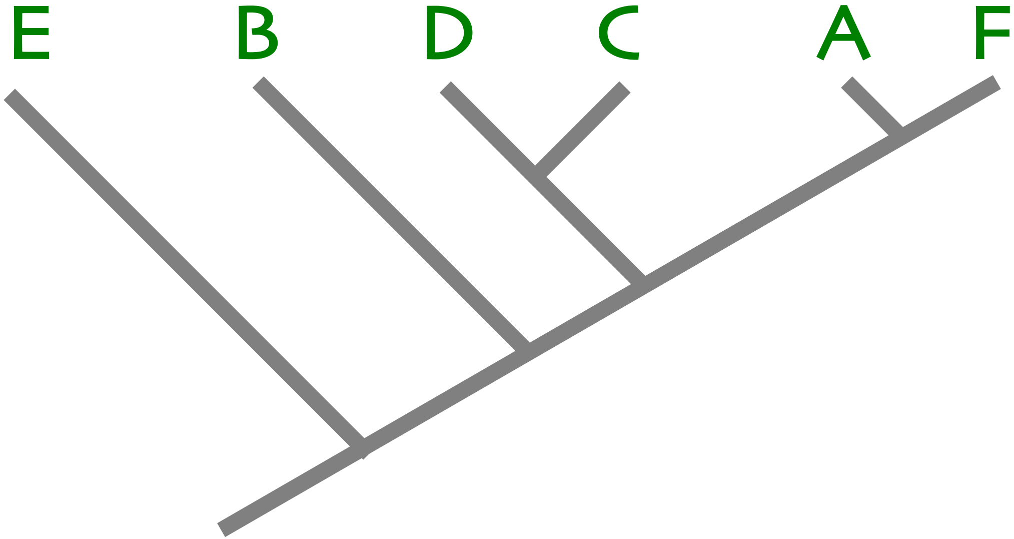
Consensus trees



(E B D C A F)
(B D C A F)
(D C A F)
(D C)
(A F)

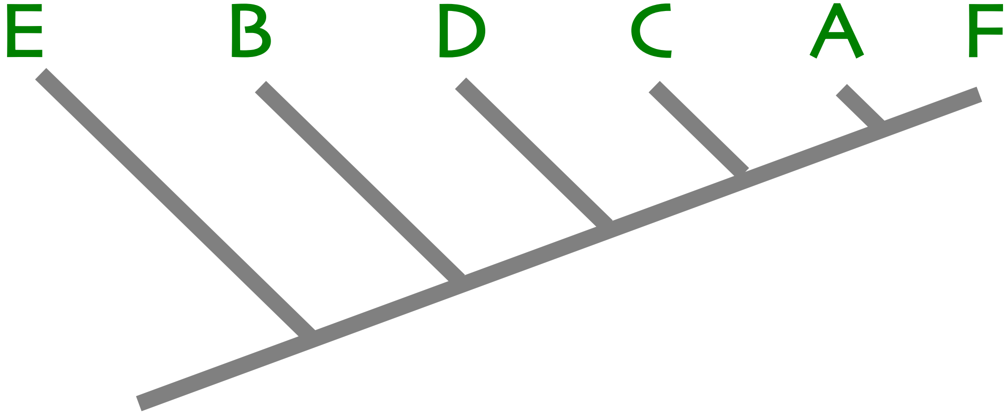


Consensus trees

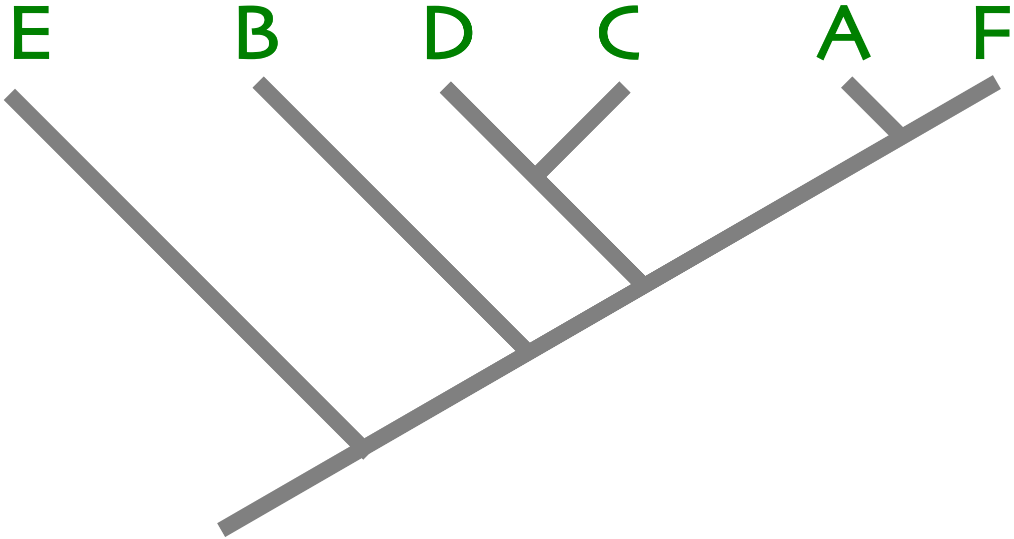


(E B D C A F)
(B D C A F)
(D C A F)
(D C)
(A F)

(E B D C A F)
(B D C A F)
(D C A F)
(C A F)
(A F)

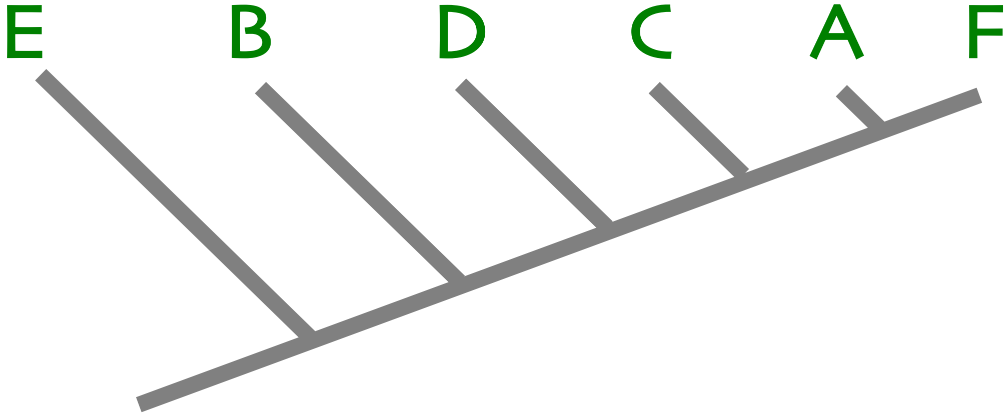


Consensus trees



(E B D C A F)
(B D C A F)
(D C A F)
~~(D C)~~
(A F)

(E B D C A F)
(B D C A F)
(D C A F)
~~(C A F)~~
(A F)



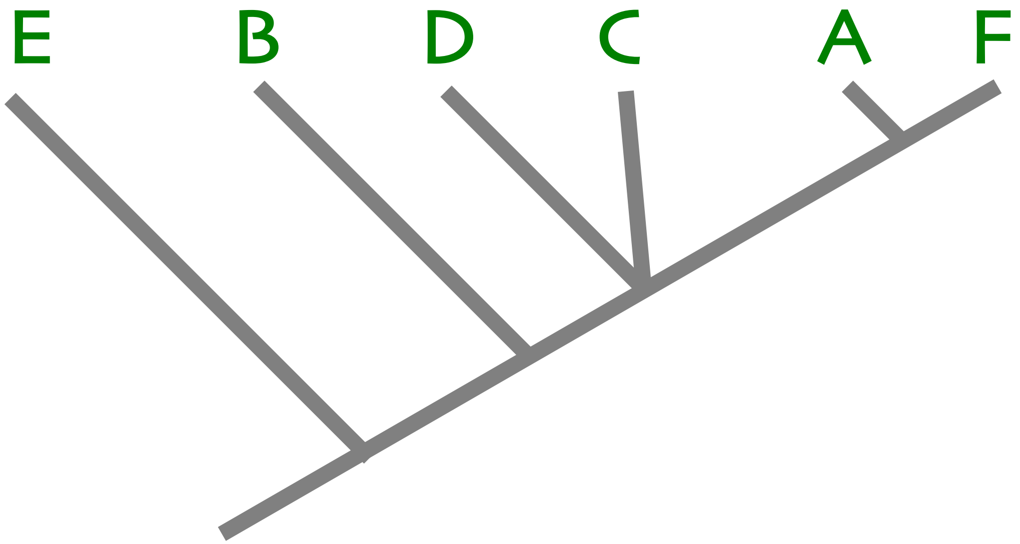
Consensus trees

(E (B (D C (A F))))

(E B D C A F)
(B D C A F)
(D C A F)
(A F)

Consensus trees

(E (B (D C (A F))))

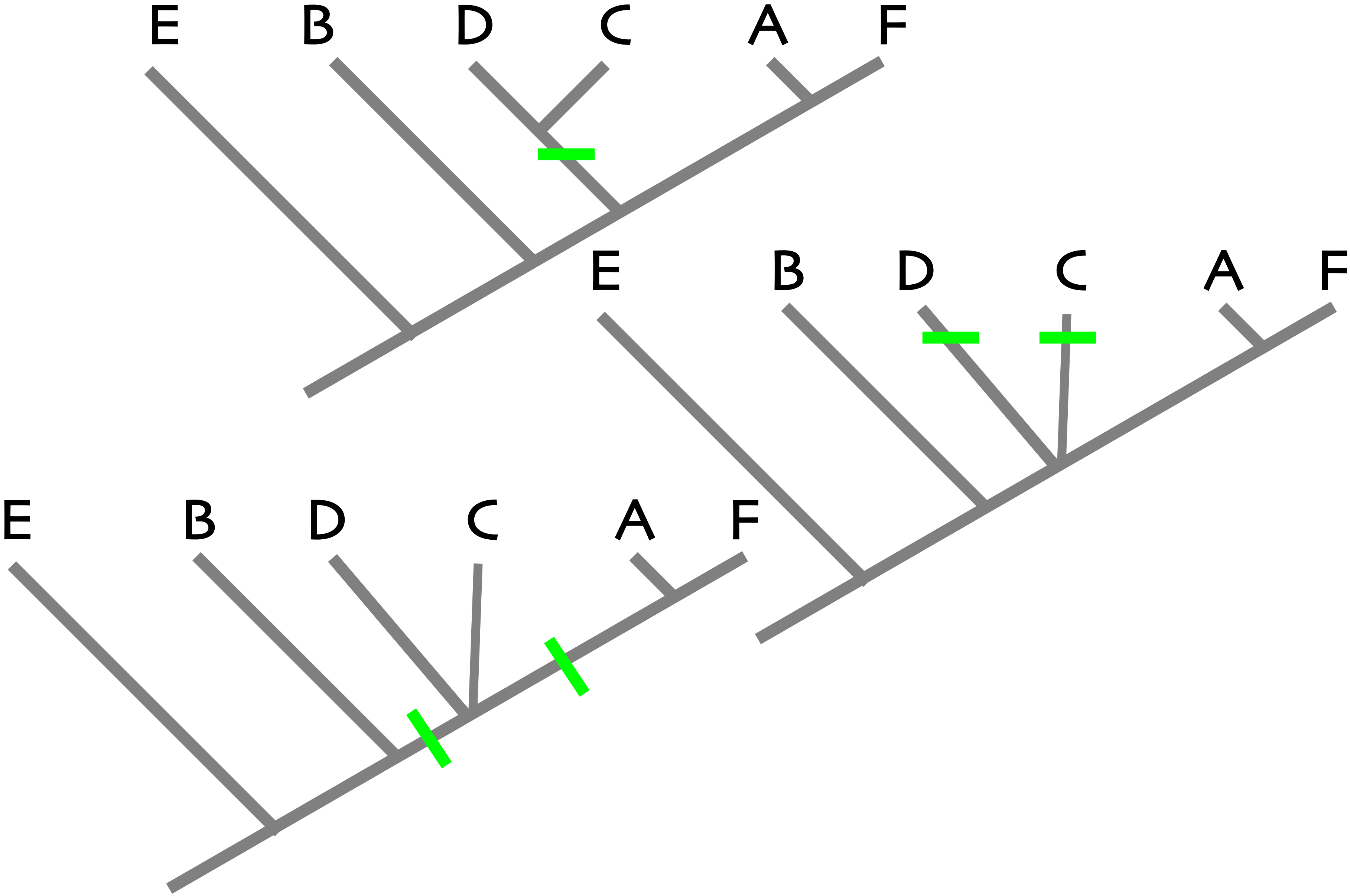


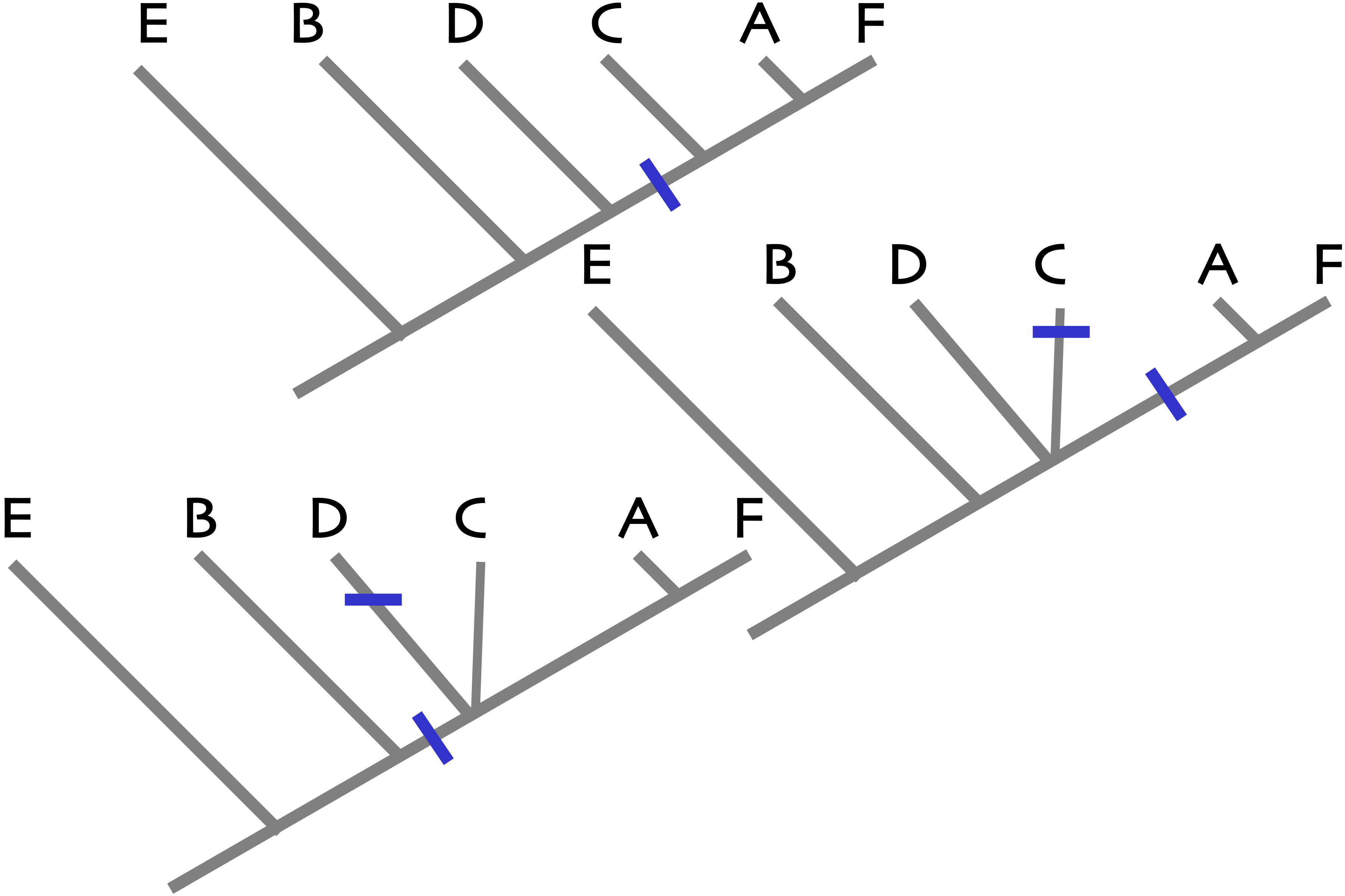
(E B D C A F)
(B D C A F)
(D C A F)
(A F)

Consensus trees

consensus tree is ALWAYS ONLY SUMMARY

it is ALWAYS more complicated than any of the original trees


E
B
D
C
A
F
E
B
D
C
A
F
E
B
D
C
A
F


E
B
D
C
A
F
E
B
D
C
A
F
E
B
D
C
A
F

Consensus trees

consensus tree is ALWAYS ONLY SUMMARY

it is ALWAYS more complicated than any of the original trees

groups shared by ALL trees are presented on 1 tree

COMPROMISE TREES

often referred to as consensus trees

Majority rule compromise

Adams

Combinable component (semistrict)

Nixon, K. C. & Carpenter, J. 1996. On consensus, collabsibility, and clade concordance. *Cladistics* 12: 305-321.

COMPROMISE TREES

Majority rule compromise

Margush, T. & McMorris, F. R. 1981. Consensus n-trees. *Bull. Math. Biol*. 43: 239-244.

Adams

Combinable component (semistrict)

COMPROMISE TREES

commonly used for presentation of support values

mostly those groups present on $\geq$ 50% of original trees included

percentage describing the presence of groups marked on compromise tree (50-100)

COMPROMISE TREES

majority rule compromise

when used as summary of optimal trees it should be noticed that part of the original trees are in CONFLICT with this summary!!

this kind of usage IS NOT RECOMMENDED, simply MISLEADING

SUMMARY

exhaustive & branch and bound searches can be used only for analyses of SMALL matrices

only these two ENSURE finding the optimal tree

heuristic search is based on rearrangement of branches of tree(s)

branch SWAPPING

searches should be planned carefully BEFORE starting them in order to avoid unnecessary analyses of large number of similar trees & use of CPU time

use of best programs & efficient algorithms necessary for analyses of LARGE matrices

PARALLELIZATION have enabled analyses of larger and larger materials

trees can be presented as PARENTHETICAL NOTATIONS

consensus trees are useful **SUMMARIES** of many trees