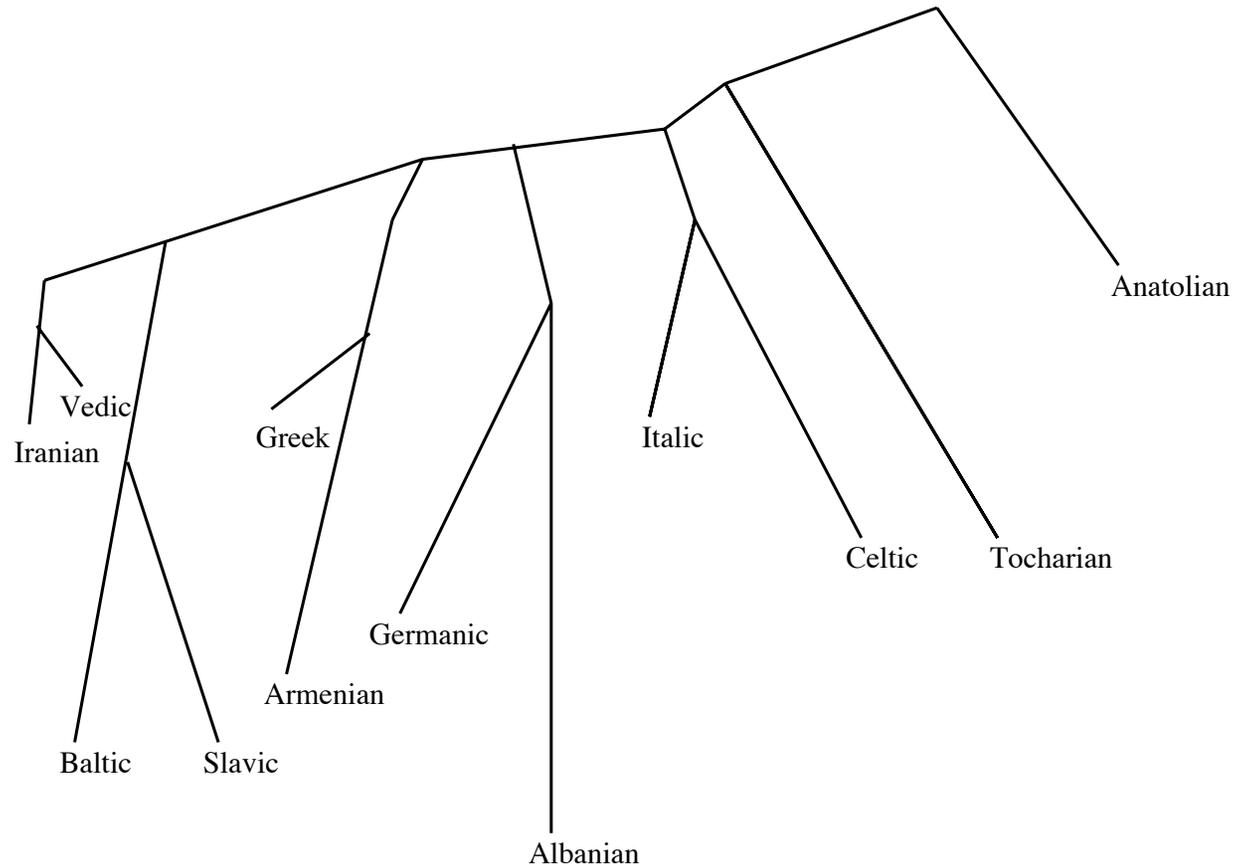


Chordal Graphs and Evolutionary Trees

Tandy Warnow

Possible Indo-European tree (Ringe, Warnow and Taylor 2000)

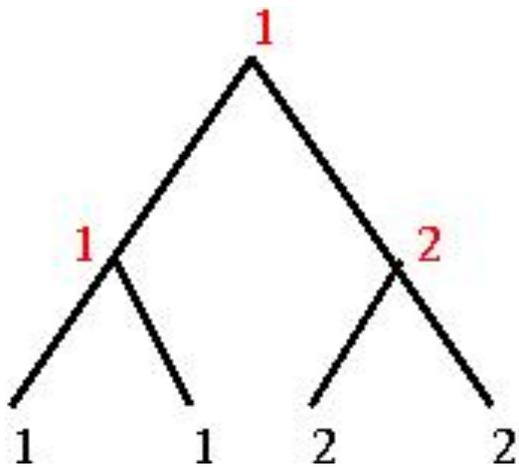


Phylogenies of Languages

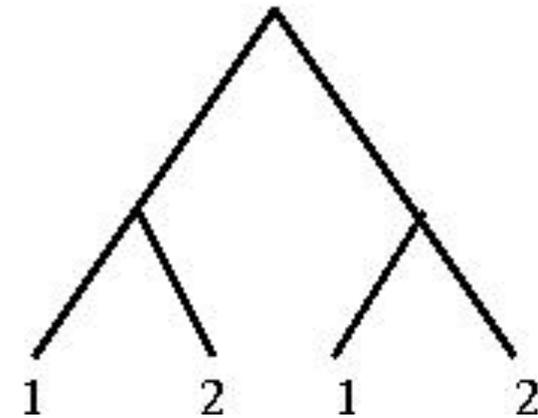
- Languages evolve over time, just as biological species do (geographic and other separations induce changes that over time make different dialects incomprehensible -- and new languages appear)
- The result can be modelled as a rooted tree
- The interesting thing is that many characteristics of languages evolve without back mutation or parallel evolution -- so a “perfect phylogeny” is possible!

“Homoplasy-Free” Evolution (perfect phylogenies)

YES



NO



Historical Linguistic Data

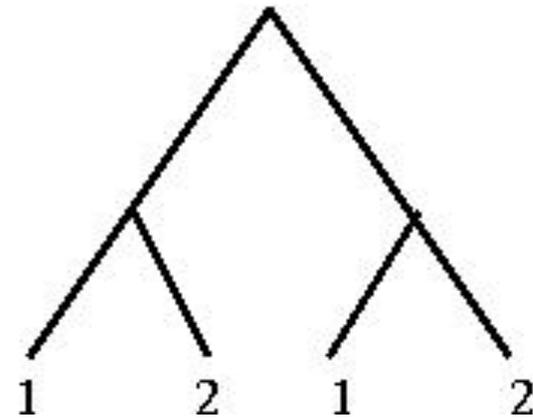
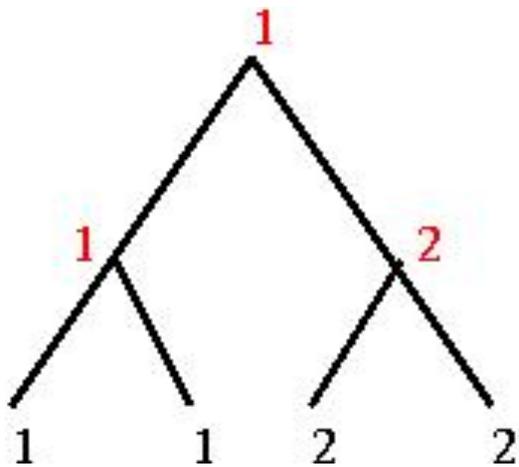
- A character is a function that maps a set of languages, L , to a set of states.
- Three kinds of characters:
 - Phonological (sound changes)
 - Lexical (meanings based on a wordlist)
 - Morphological (grammatical features)

Cognate Classes

- Two words w_1 and w_2 are in the same cognate class, if they evolved from the same word through sound changes.
- French “champ” and Italian “champo” are both descendants of Latin “campus”; thus the two words belong to the same cognate class.
- Spanish “mucho” and English “much” are not in the same cognate class.

The Ringe-Warnow Model of Language Evolution

- The nodes of the tree which contain elements of the same cognate class should form a rooted connected subgraph of the true tree
- The model is known as the **Character Compatibility** or **Perfect Phylogeny**.



Perfect Phylogeny

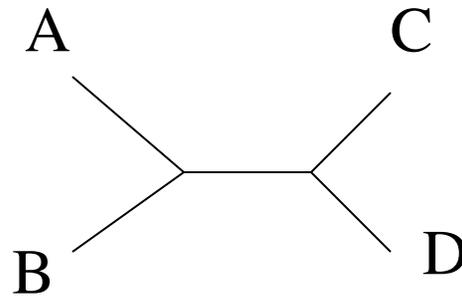
- A phylogeny T for a set S of taxa is a **perfect phylogeny** if each state of each character occupies a subtree (no character has back-mutations or parallel evolution)

Perfect phylogenies, cont.

- $A=(0,0)$, $B=(0,1)$, $C=(1,3)$, $D=(1,2)$ has a perfect phylogeny!
- $A=(0,0)$, $B=(0,1)$, $C=(1,0)$, $D=(1,1)$ does not have a perfect phylogeny!

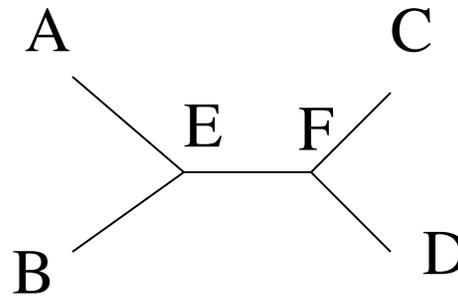
A perfect phylogeny

- $A = 0 \ 0$
- $B = 0 \ 1$
- $C = 1 \ 3$
- $D = 1 \ 2$



A perfect phylogeny

- $A = 0 \ 0$
- $B = 0 \ 1$
- $C = 1 \ 3$
- $D = 1 \ 2$
- $E = 0 \ 3$
- $F = 1 \ 3$

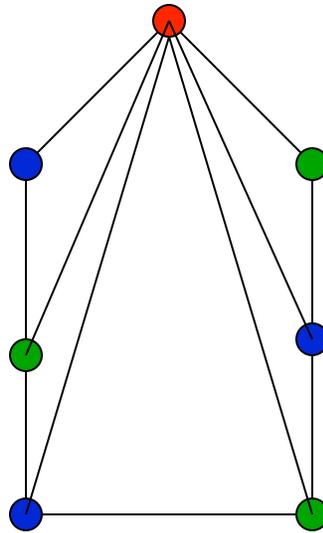


The Perfect Phylogeny Problem

- Given a set S of taxa (species, languages, etc.) determine if a perfect phylogeny T exists for S .
- The problem of determining whether a perfect phylogeny exists is NP-hard (McMorris *et al.* 1994, Steel 1991).

Triangulated Graphs

- A graph is triangulated if it has no simple cycles of size four or more.



Triangulated Graphs

Theorem (Gravil 1974, Buneman 1974): A graph G is triangulated if and only if G is the intersection graph of a set of subtrees of a tree.

Proof: One direction is easy, and the other is not...

Perfect Phylogenies and Triangulated Colored Graphs

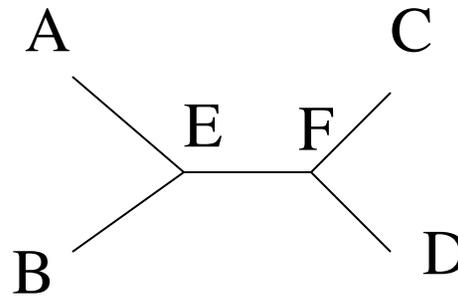
- Suppose M is a character matrix and T is a perfect phylogeny for M .
- Then let M' be the extension of M to include the additional “species” added at the internal nodes.
- Colored graph G based on M' :
 - For each character α and state i in M' , give a vertex $v_{(\alpha,i)}$ and color the vertex with the color for α .
 - Put edges between two vertices if they share any species.
 - G is triangulated and properly colored.

Perfect Phylogenies and Triangulated Colored Graphs

- Suppose M is a character matrix and T is a perfect phylogeny for M .
- Then let M' be the extension of M to include the additional “species” added at the internal nodes.
- The [partition intersection graph](#) G based on M' is triangulated and properly colored. (Why?)
- But if we had based it on M it might not have been triangulated. (Why?)

A perfect phylogeny

- $A = 0 \ 0$
- $B = 0 \ 1$
- $C = 1 \ 3$
- $D = 1 \ 2$
- $E = 0 \ 3$
- $F = 1 \ 3$



Draw the partition intersection graph.

Matrix with a perfect phylogeny

	c1	c2	c3
s1	3	2	1
s2	1	2	2
s3	1	1	3
s4	2	1	1

Matrix with a perfect phylogeny

	c1	c2	c3
s1	3	2	1
s2	1	2	2
s3	1	1	3
s4	2	1	1

Draw the perfect phylogeny and compute the sequences at the internal nodes.

Matrix with a perfect phylogeny

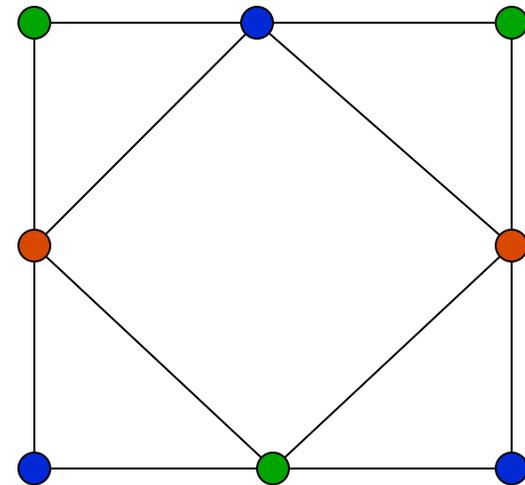
	c1	c2	c3
s1	3	2	1
s2	1	2	2
s3	1	1	3
s4	2	1	1

Draw the partition intersection graph for the extended matrix (including the sequences at the internal nodes).

The partition intersection graph

“Yes” Instance of PP:

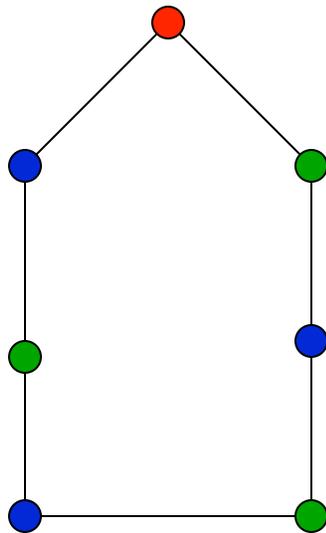
	c1	c2	c3
s1	3	2	1
s2	1	2	2
s3	1	1	3
s4	2	1	1



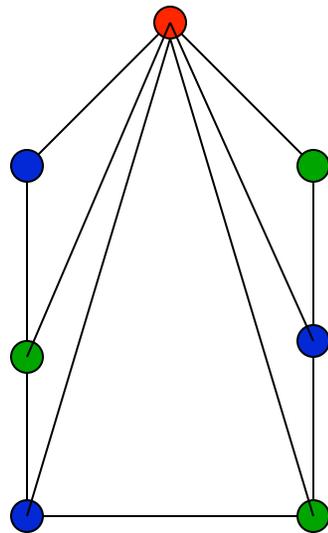
Triangulating colored graphs

- Let $G=(V,E)$ be a graph and c be a vertex coloring of G . Then G can be c -triangulated if a supergraph $G'=(V,E')$ exists that is triangulated and where the coloring c is proper.
- In other words, G can be c -triangulated if and only if we can add edges to G to make it triangulated without adding edges between vertices of the same color.

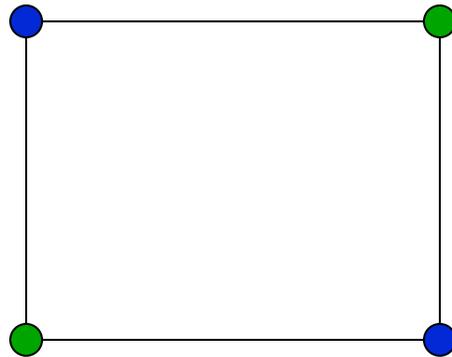
A graph that can be c-triangulated



A graph that can be c-triangulated



A graph that cannot be c-triangulated



Triangulating Colored Graphs (TCG)

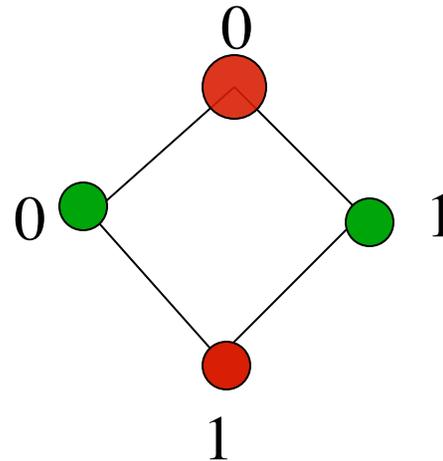
Triangulating Colored Graphs: given a vertex-colored graph G , determine if G can be c -triangulated.

The PP and TCG Problems

- **Buneman's Theorem:**
A perfect phylogeny exists for a set S *if and only if* the associated character state intersection graph can be c -triangulated.
- The PP and TCG problems are polynomially equivalent and NP-hard.

A no-instance of Perfect Phylogeny

- A = 0 0
- B = 0 1
- C = 1 0
- D = 1 1

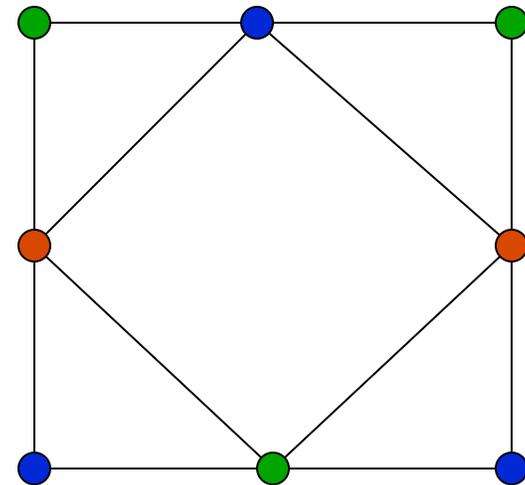


An input to perfect phylogeny (left) of four sequences described by two characters, and its partition intersection graph. Note that the partition intersection graph is 2-colored.

Solving the PP Problem Using Buneman's Theorem

“Yes” Instance of PP:

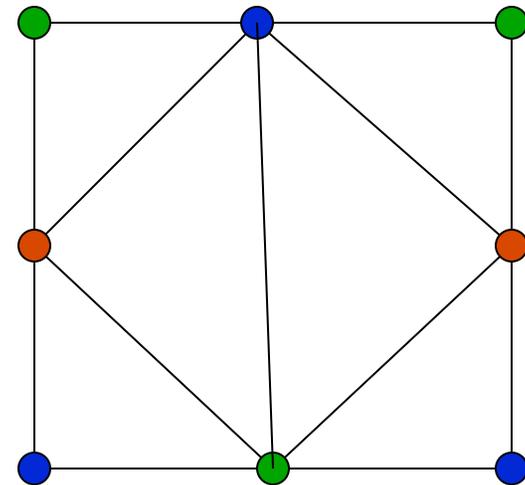
	c1	c2	c3
s1	3	2	1
s2	1	2	2
s3	1	1	3
s4	2	1	1



Solving the PP Problem Using Buneman's Theorem

“Yes” Instance of PP:

	c1	c2	c3
s1	3	2	1
s2	1	2	2
s3	1	1	3
s4	2	1	1



Some special cases are easy

- Binary character perfect phylogeny solvable in linear time
- r -state characters solvable in polynomial time for each r (combinatorial algorithm)
- Two character perfect phylogeny solvable in polynomial time (produces 2-colored graph)
- k -character perfect phylogeny solvable in polynomial time for each k (produces k -colored graphs -- connections to Robertson-Seymour graph minor theory)

Early History

- LeQuesne (1969,1972,1974,1977): initial formulation of perfect phylogenies
- Estabrook (1972), Estabrook et al. (1975): mathematical foundations of perfect phylogenies
- McMorris (1997): binary character compatibility
- Felsenstein (1984): review paper
- Estabrook & Landrum, Fitch 1975: compatibility of two characters
- Steel (1992) and Bodlaender et al. (1992): NP-hardness
- Buneman (1974): reduced perfect phylogeny to triangulating colored graphs
- Kannan and Warnow (1992): established equivalence of TCG and PP

See T. Warnow, 1993. Constructing phylogenetic trees efficiently using compatibility criteria. *New Zealand Journal of Botany*, 31:3, pp. 239-248 (linked off my homepage) for a survey of the early literature and the full citations.

Literature sample

- R. Agarwala and D. Fernandez-Baca, 1994. A polynomial-time algorithm for the perfect phylogeny problem when the number of character states is fixed. *SIAM Journal on Computing*, 23 , 1216–1224.
- R. Agarwala and D. Fernandez-Baca, 1996. Simple algorithms for perfect phylogeny and triangulating colored graphs. *International Journal of Foundations of Computer Science*, 7 , 11–21.
- H.L. Bodlaender, M.R. Fellows, Michael T. Hallett, H. Todd Wareham, and T. Warnow. 2000. The hardness of perfect phylogeny, feasible register assignment and other problems on thin colored graphs, *Theoretical Computer Science* 244 (2000) 167-188
- H.L. Bodlaender and T. Kloks, 1993: A simple linear time algorithm for triangulating three-colored graphs. *Journal of algorithms* J 5: 160-172.
- D. Fernandez-Baca, 2000. The perfect phylogeny problem. Pages 203–234 of: Du, D.-Z., and Cheng, X. (eds), *Steiner Trees in Industries* . Kluwer Academic Publishers.
- D. Gusfield, 1991. Efficient algorithms for inferring evolutionary trees. *Networks* 21, 19–28.
- R. Idury and A. Schaffer. 1993: Triangulating three-colored graphs in linear time and linear space. *SIAM journal on discrete mathematics* 6: 289-294.
- S. Kannan and T. Warnow, 1992. Triangulating 3-colored graphs. *SIAM J. on Discrete Mathematics*, Vol. 5 No. 2, pp. 249-258 (also SODA 1991)
- S. Kannan and T. Warnow, 1997. A fast algorithm for the computation and enumeration of perfect phylogenies when the number of character states is fixed. *SIAM J. Computing*, Vol. 26, No. 6, pp. 1749-1763 (also SODA 1995)
- F.R. McMorris, T. Warnow, and T. Wimer, 1994. Triangulating Vertex Colored Graphs. *SIAM J. on Discrete Mathematics*, Vol. 7, No. 2, pp. 296-306 (also SODA 1993).

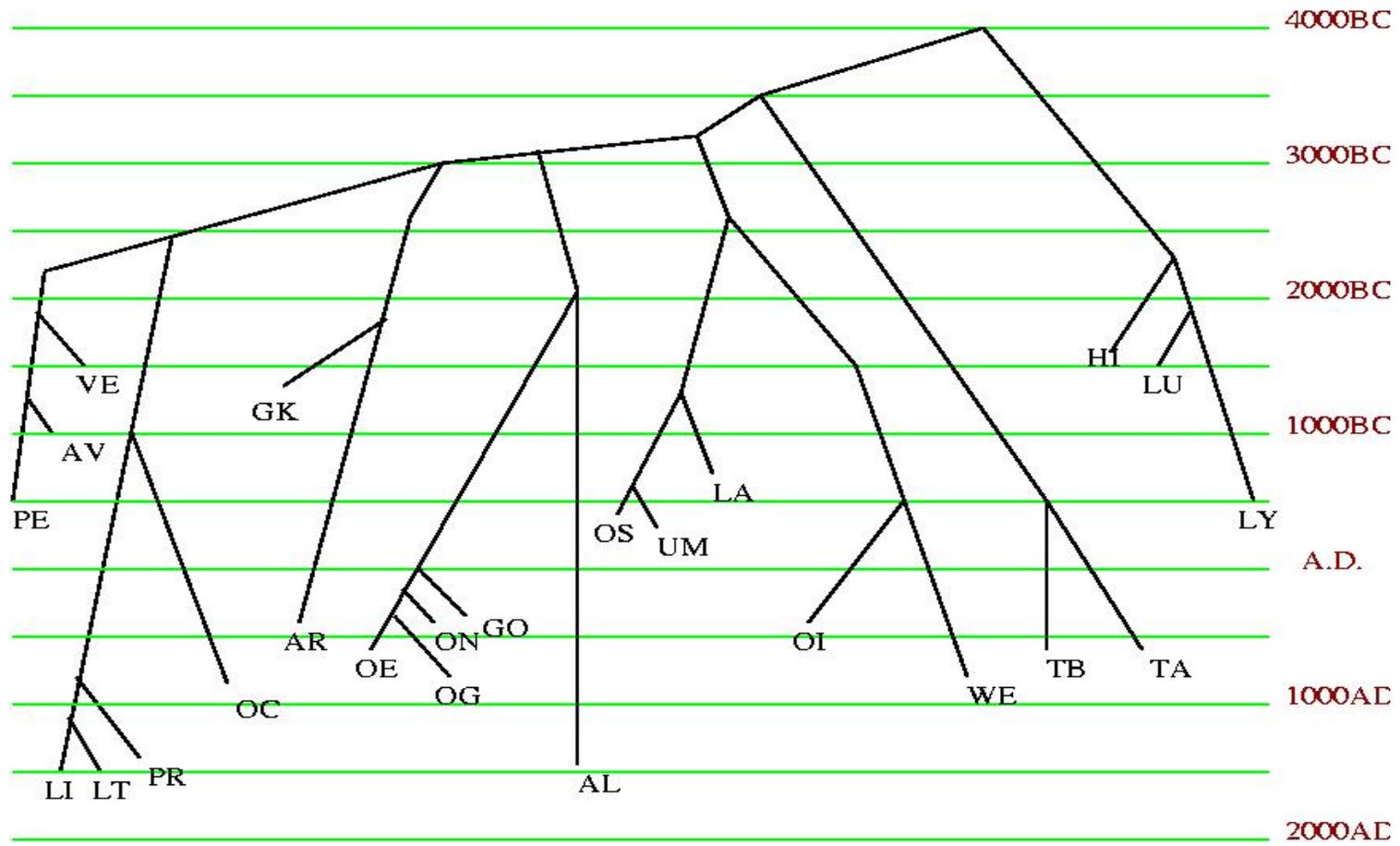
Constructing trees in historical linguistics

- Maximum Compatibility: given the input matrix for the set S of languages described by the set C of characters, find a tree T leaf-labelled by S on which a maximum number of the characters in C are compatible (i.e., evolve without homoplasy).
- NP-hard.

The Indo-European (IE) Dataset

- 24 languages
- 22 phonological characters, 15 morphological characters, and 333 lexical characters.
- Total number of working characters is 390 (multiple character coding, and parallel development)
- A phylogenetic tree T on the IE dataset (Ringe, Taylor and Warnow)
- T is compatible with all but 16 characters
- *Resolves most of the significant controversies in Indo-European evolution; shows however that Germanic is a problem (not treelike)*

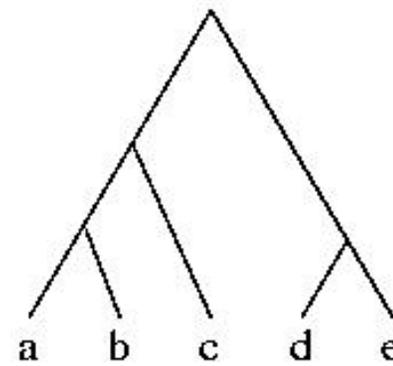
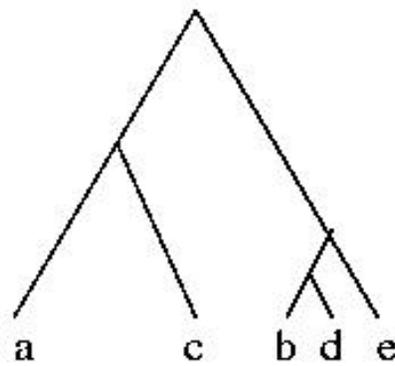
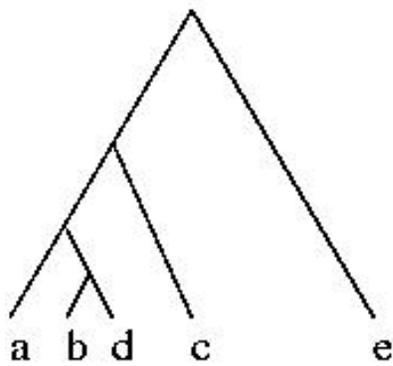
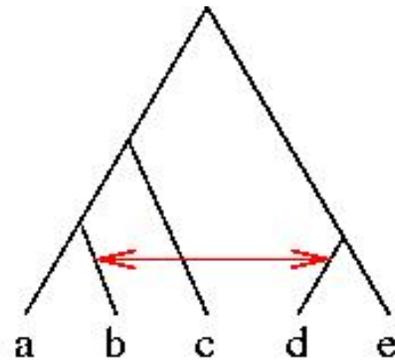
Phylogenetic Tree of the IE Dataset (Ringe, Warnow, and Taylor)



Explaining remaining incompatibilities

- We modelled the remaining incompatibilities as undetected borrowing between languages.
- This leads to the mathematical model of “perfect phylogenetic networks”

Modelling borrowing: Networks and Trees within Networks



“Perfect Phylogenetic Network” for IE

Nakhleh et al., Language 2005

