Credit

Nel giardino della stemmatologia digitale: alberi, cespugli e ragnatele

Marina Buzzoni
Padova 10-02-2026

ASSOCIAZIONE PER L'INFORMATICA UMANISTICA E LA CULTURA DIGITALE

AIUCD diffonde la riflessione metodologica e teorica, la collaborazione scientifica e lo sviluppo di pratiche, risorse e strumenti condivisi nel campo dell'informatica umanistica e nell'uso delle applicazioni digitali in tutte le aree delle scienze umane. Promuove inoltre la riflessione sui fondamenti umanistici delle metodologie informatiche nel campo delle culture di rete.

AIUCD fosters methodological and theoretical research, scientific collaboration and the development of shared practices, resources and tools in the Digital Humanities, as well as digital applications in all Humanities disciplines. Additionally, it seeks to stimulate reflection on the theoretical foundations of computational methods in Internet culture.

https://www.aiucd.it/

Digitale e *Public Engagement*: pratiche e prospettive nelle *Digital Humanities*

XV convegno annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD)

Il XV convegno **AIUCD** 2026 esplora come le tecnologie digitali aprono nuove forme di partecipazione attiva nella ricerca umanistica. Dalla co-creazione con le comunità ai processi collaborativi di documentazione e valorizzazione del patrimonio culturale, il convegno mette al centro pratiche, strumenti e metodologie che rendono la conoscenza più aperta, condivisa e inclusiva

Update: Nuova data per l'invio delle proposte

Cagliari **3–5 giugno** 2026

https://www.aiucd2026.unica.it/

Stemmatological analysis

Stemmatology is an umbrella term for all scholarly and scientific studies focused on **textual genealogy** […]. It is usually concerned with reconstructing a specific *stemma* starting from the surviving witnesses, though it may also deal with *stemmata* in a more abstract sense.

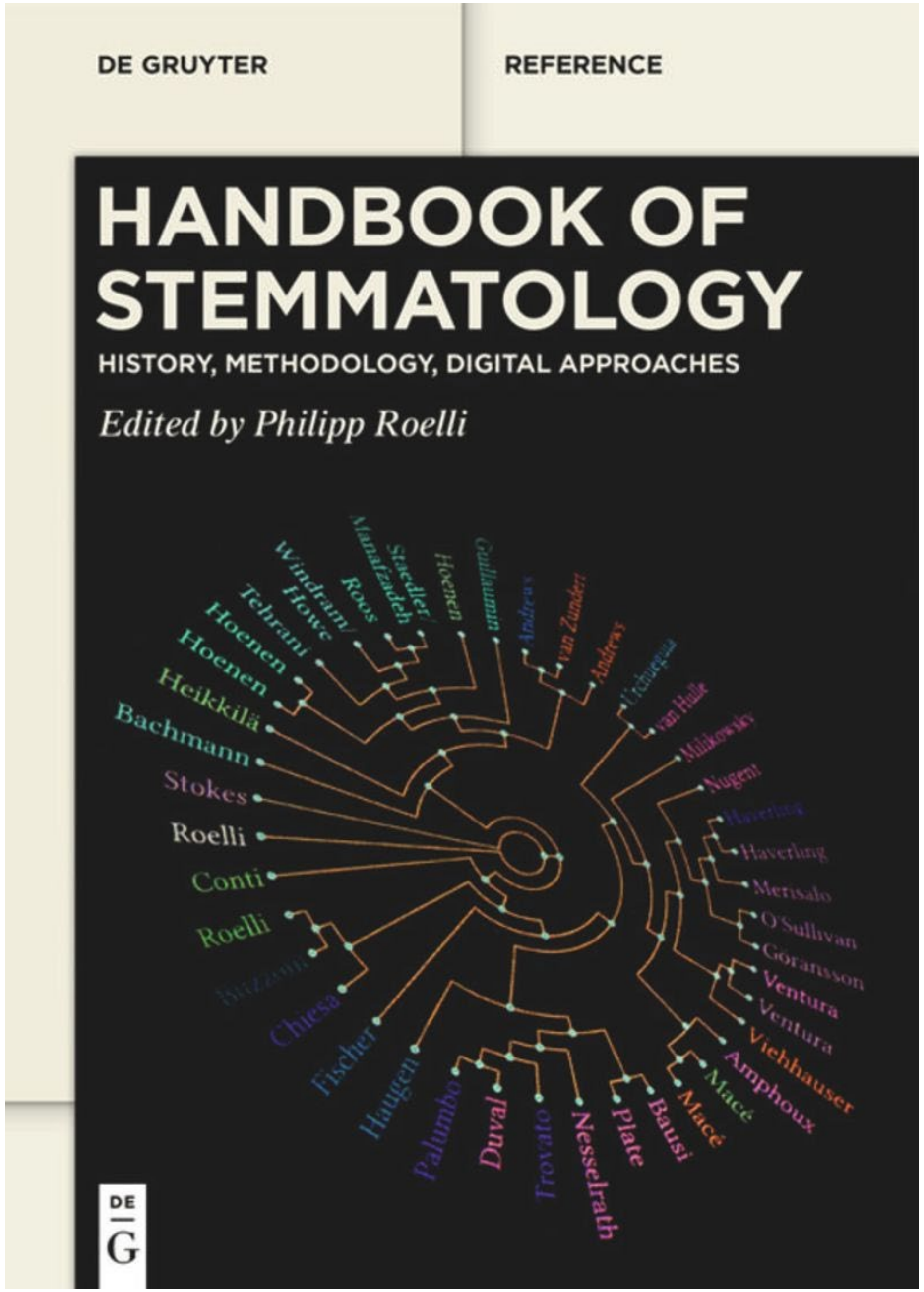
(adapted from *PLS*)

https://wiki.helsinki.fi/xwiki/bin/view/stemmatology/Parvum%20lexicon%20stemmatologicum/

On the way to computer-assisted stemmatology

- **Computer-assisted stemmatology**: in the second half of the 20th c. some scholars realized that stemmatology and the study of evolution in e.g. biology have much in common: **the former studies the evolution of texts and the latter that of organisms.**

- These scholars started to think that the use of computerised approaches developed in particular in the field of **phylogenetics** could be potentially very useful for stemmatology as well.



DE GRUYTER

REFERENCE

HANDBOOK OF STEMMATOLOGY

HISTORY, METHODOLOGY, DIGITAL APPROACHES

Edited by Philipp Roelli

https://www.degruyterbrill.com/document/doi/10.1515/9783110684384/html

The underlying idea

«The copying of a manuscript by a scribe with the incorporation of changes that were then propagated when that copy was in turn copied, shows clear parallels to the error-prone replication of DNA.» (Howe & Windram 2011)

How to apply this idea

- The approach for applying phylogenetic methods to texts is simple in principle.
- The texts are **aligned** and then **encoded as a string of characters**, usually with each character corresponding to a word.
- The character strings are then used to **build a file in exactly the same format as used by phylogenetic tree-building programs**, and the file is submitted to the same programs, unaltered.

An example

Howe & Windram 2011: 3

A set of texts is aligned so that that any given column contains the equivalent word from all the texts. An example is given below for a short sequence of words from the poem "His Age"by Robert Herrick (1591-1674) in eleven different texts (denoted by three-letter abbreviations). Examples of three of the texts are given in Figure 1. Aligning the texts is equivalent to aligning biological sequence data so that each column in an alignment contains the homologous amino acid or nucleotide residue, and a method for textual alignment has been developed based explicitly on alignment of biological data [44].

```
BLE thus conspiringe with our mirth   Next to infuse Our better  beere
BLH thus conspiring  with our mirth   next to infuse our better  beere
BoE thus conspiringe wth  our mirth   next to infuse Our better  beare
BoF thus conspiringe with our mirth   Next to infuse Our better  beere
Gre ~    ~           ~    ~   ~       ~    ~  ~      ~   ~       ~
Har thus conspiring  with our mirth   Next to infuse Our better  beere
Hes thus conspiring  with our mirth)  Then to infuse Our browner Ale
Hun this conspiring  with our mirth   next to infuse our better  beare
Osb ~    conspiring  wth  our stealth Next to enfuse Our better  beare
Ros thus conspiring  with our mirth   Next to infuse Our better  beare
SJC ~    conspiringe with our stealth Next to infuse Our better  beere
```

For many texts, spelling variants are not considered to be informative. In which case, the text may be “regularized” by correcting such variants to read the same. Punctuation may similarly be removed.

```
BLE thus conspiring with our mirth   next to  infuse our better  beere
BLH thus conspiring with our mirth   next to  infuse our better  beere
BoE thus conspiring with our mirth   next to  infuse our better  beere
BoF thus conspiring with our mirth   next to  infuse our better  beere
Gre ~    ~          ~    ~   ~       ~    ~   ~      ~   ~       ~
Har thus conspiring with our mirth   next to  infuse our better  beere
Hes thus conspiring with our mirth   Then to  infuse our browner Ale
Hun this conspiring with our mirth   next to  infuse our better  beere
Osb ~    conspiring with our stealth next to  infuse our better  beere
Ros thus conspiring with our mirth   next to  infuse our better  beere
SJC ~    conspiring with our stealth next to  infuse our better  beere
```

Finally, each aligned and regularized text is converted to a string of characters, with each text represented as a row of characters, with each character usually corresponding to a word in the text. Sometimes a character corresponds to a more complex feature, such as a rearrangement of word order in a sentence. Strings of characters are then converted to a NEXUS file, the format widely used by phylogenetic programs. In this file ? represents missing data, - is a deletion and 0 or 1 are different character states. This file can then be used directly for phylogenetic analysis.

```
BLE  00000000000
BLH  00000000000
BoE  00000000000
BoF  00000000000
Gre  ???????????
Har  00000000000
Hes  00000100011
Hun  10000000000
Osb  -0001000000
Ros  00000000000
SJC  -0001000000
```

Estimated advantages

Possibly the greatest advantages of computer-assisted methods over traditional ones are their speed (in the final part of the process), calculation power, the possibility of using different methods on the same data, of assessing the results and of easily redoing the calculation whenever necessary (after having added a new witness, for example).

Potential difficulties

- The quality of the input is of primary importance for the reliability of the results → The preparatory phase can be very long and time-consuming;

- The normalisation step can be highly problematic;

- The algorithms underlying the methods used are often black boxes (for humanities scholars).

Most common methods

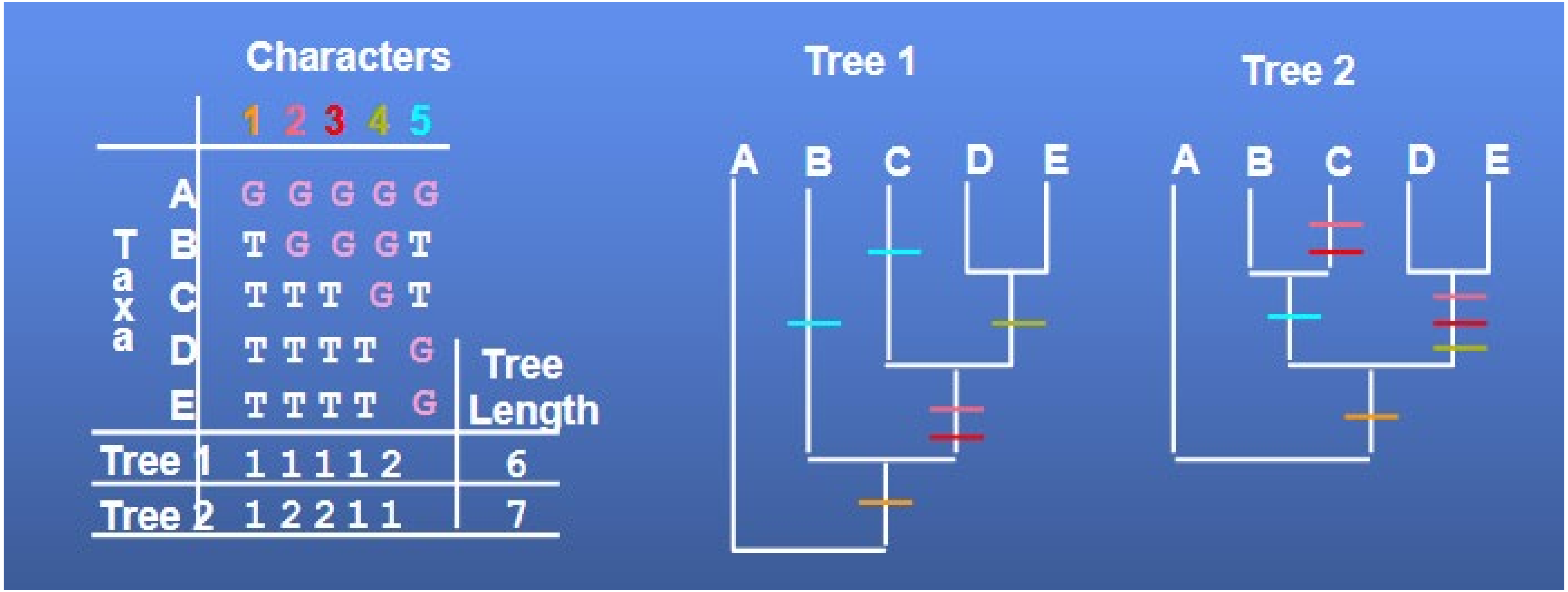
The most common computer-assisted methods (though not the only ones!) are:

- **Maximum Parsimony Method (PAUP)**;
- NeighborJoining;
- **NeighborNet**;
- RHM;
- Semstem.

Trees and bushes

PS = Maximum Parsimony (with PAUP)

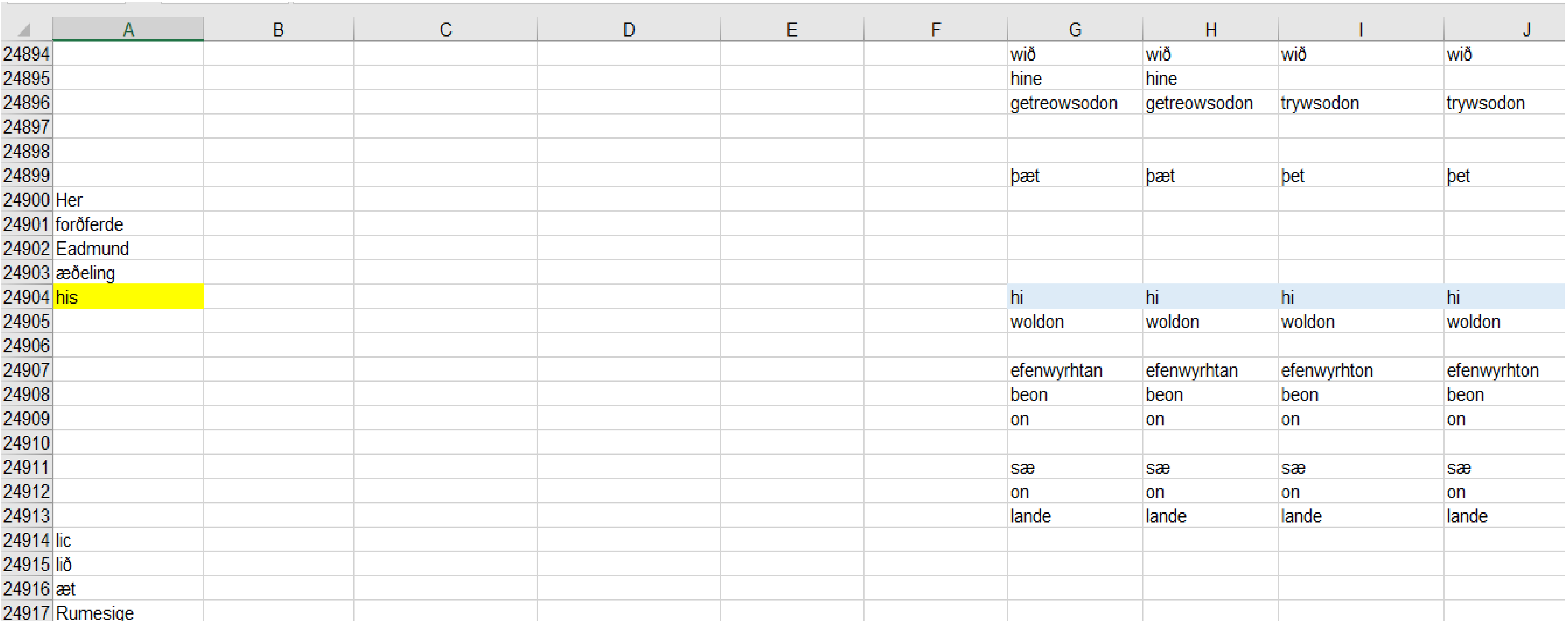
- The principle of parsimony is very important to biological cladistic classification.
- The basic idea of this principle is **that the option with fewest changes of state is always the most likely one** when creating a phylogenetic tree.
- PAUP (acronym of Phylogenetic Analysis Using Parsimony) is a tool based on this method, and created by Dr. David L. Swofford. From version 4.0 (when the program became known as PAUP*) it also supports other methods (e.g. distance matrix and likelihood methods).



Taxa	1	2	3	4	5	Tree Length
A	G	G	G	G	G	
B	T	G	G	G	T	
C	T	T	T	G	T	
D	T	T	T	T	G	
E	T	T	T	T	G	
Tree 1	1	1	1	1	2	6
Tree 2	1	2	2	1	1	7

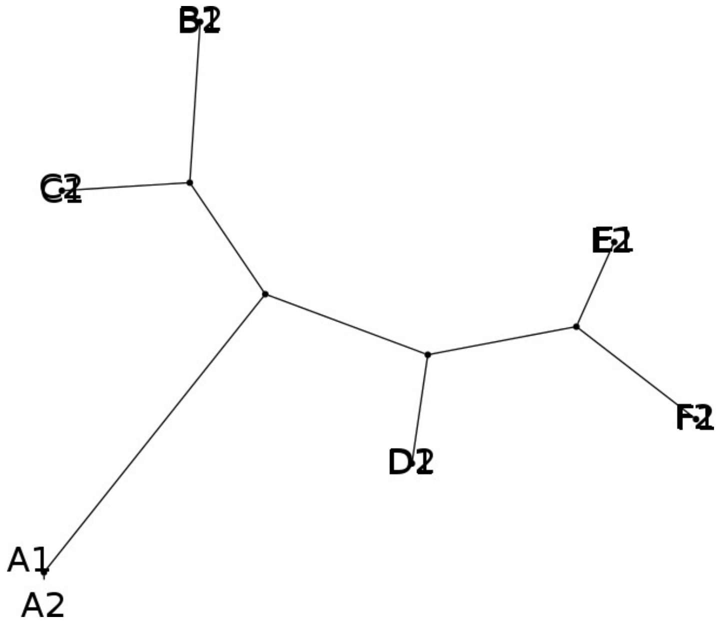
The *A-S Chronicle* tradition in a nutshell

- Seven witnesses: A to G.
- G is clearly a *codex descriptus* derived from A; B is very close to C, both having been copied from the same ancestor; F is a bilingual epitome of E.
- The complex relationships between the extant witnesses, also including some Latin texts that have a version of the Chronicle among their sources – in particular Asser's *Life of Alfred*, Æthelweard's *Chronicon*, and the *Annals of St Neots* – have been variously interpreted and represented by scholars (Whitelock 1979, 118; Cubbin 1996, liii; Buzzoni 2001, 42 and Buzzoni et al. 2016).

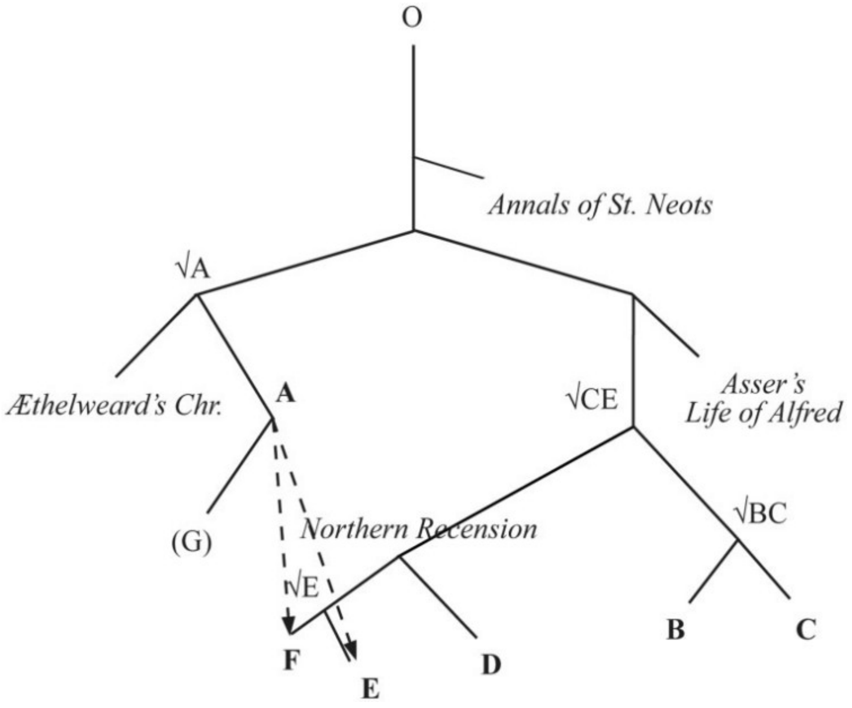


	A	B	C	D	E	F	G	H	I	J
24894							wið	wið	wið	wið
24895							hine	hine		
24896							getreowsodon	getreowsodon	trywsodon	trywsodon
24897										
24898										
24899							þæt	þæt	þet	þet
24900	Her									
24901	forðferde									
24902	Eadmund									
24903	æðeling									
24904	his						hi	hi	hi	hi
24905							woldon	woldon	woldon	woldon
24906										
24907							efenwyrhtan	efenwyrhtan	efenwyrhton	efenwyrhton
24908							beon	beon	beon	beon
24909							on	on	on	on
24910										
24911							sæ	sæ	sæ	sæ
24912							on	on	on	on
24913							lande	lande	lande	lande
24914	lic									
24915	lið									
24916	æt									
24917	Rumesige									

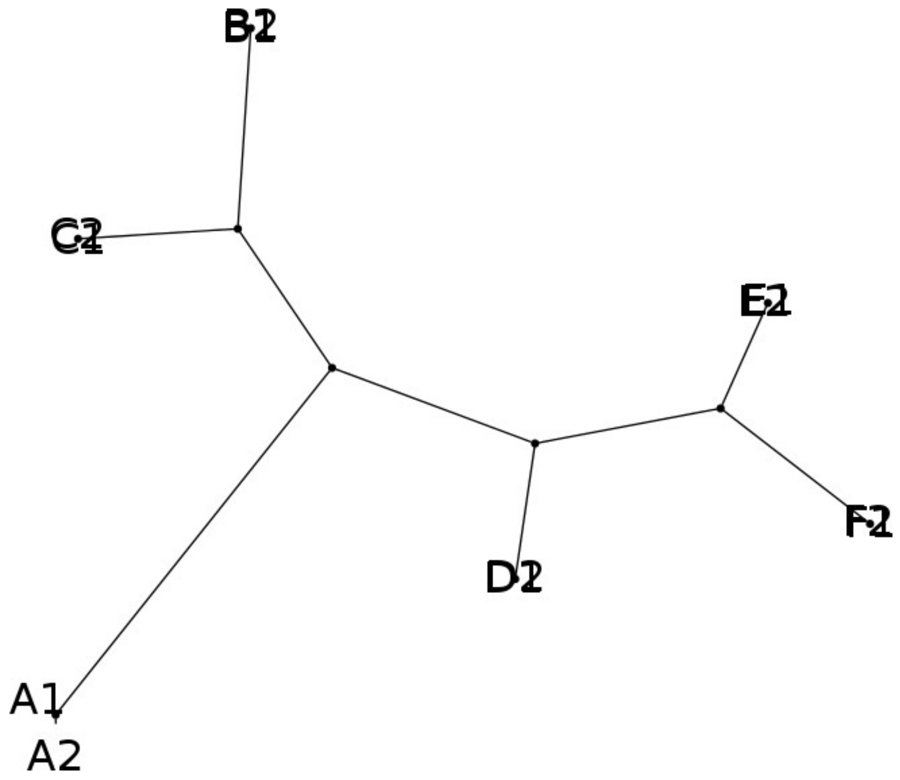
A-S Chronicle tradition: PS



PAUP, PS

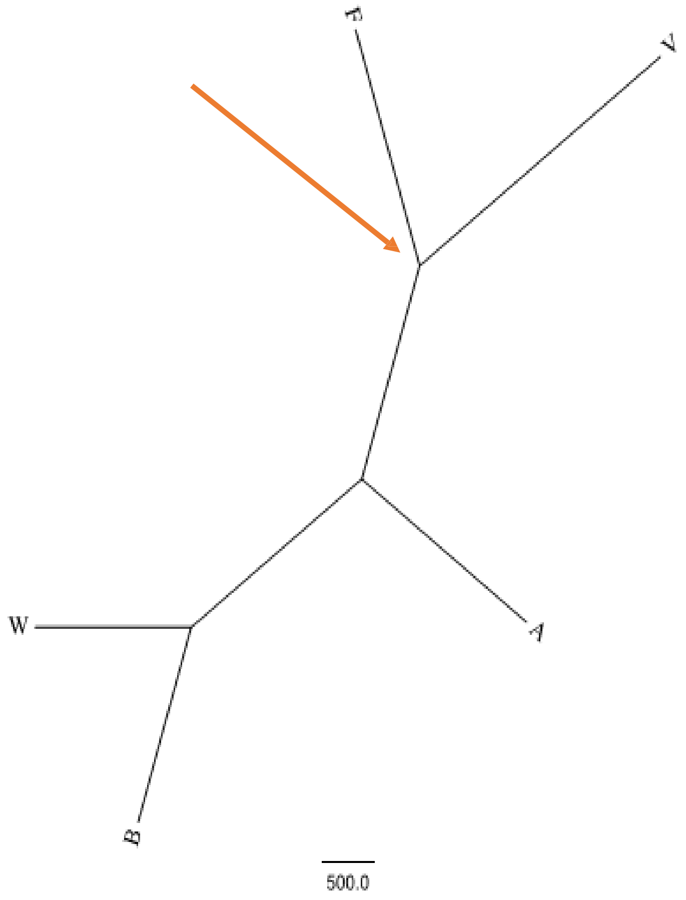


Plummer’s claim confirmed

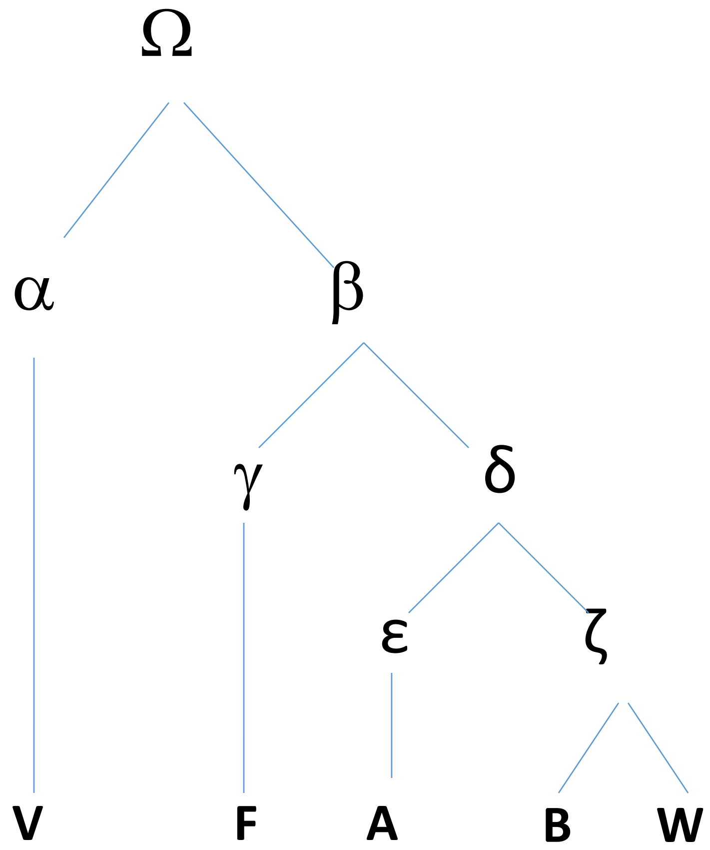


The phylogram provides clear confirmation of Plummer’s claim that «A, C(/B), D and E have every right to be considered distinct Chronicles.» (Plummer 1972, II.xxiii)

Melione tradition



PAUP, PS



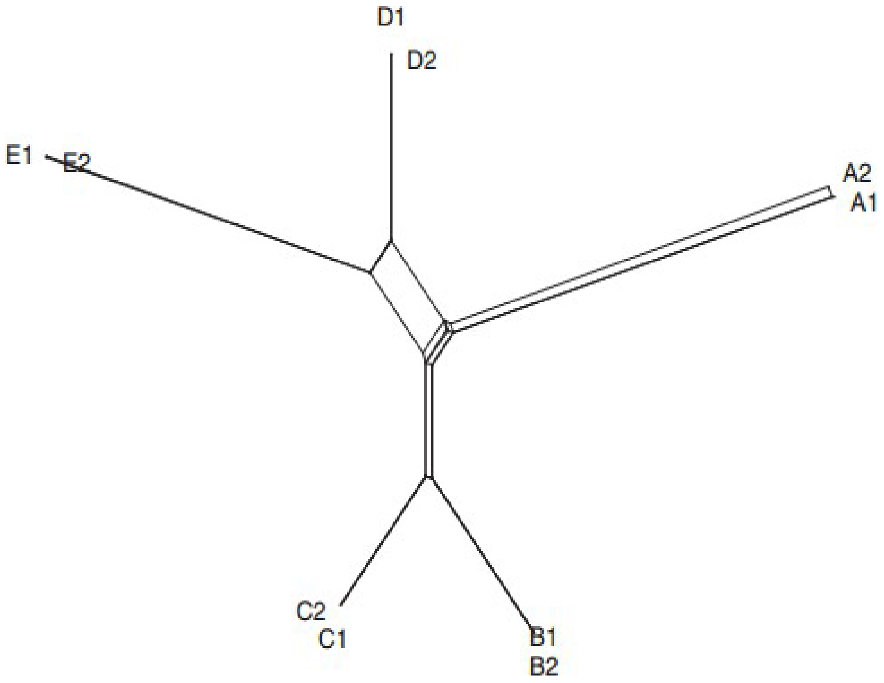
Spiderwebs

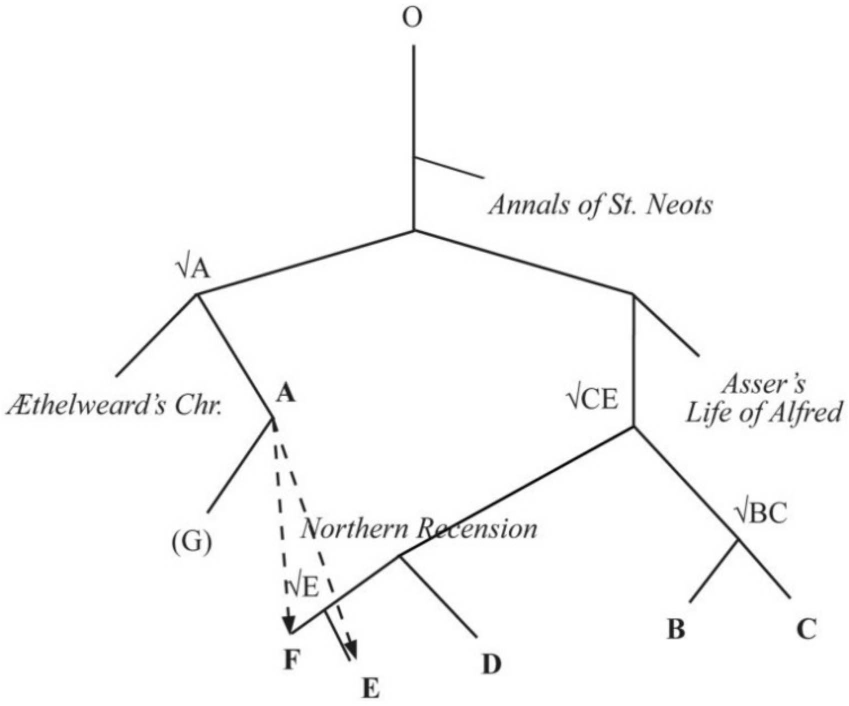
NeighborNet

- The advantage of this method for stemmatologists is basically that it allows to construct **phylogenetic networks**, rather than simply tree-like glyphs.

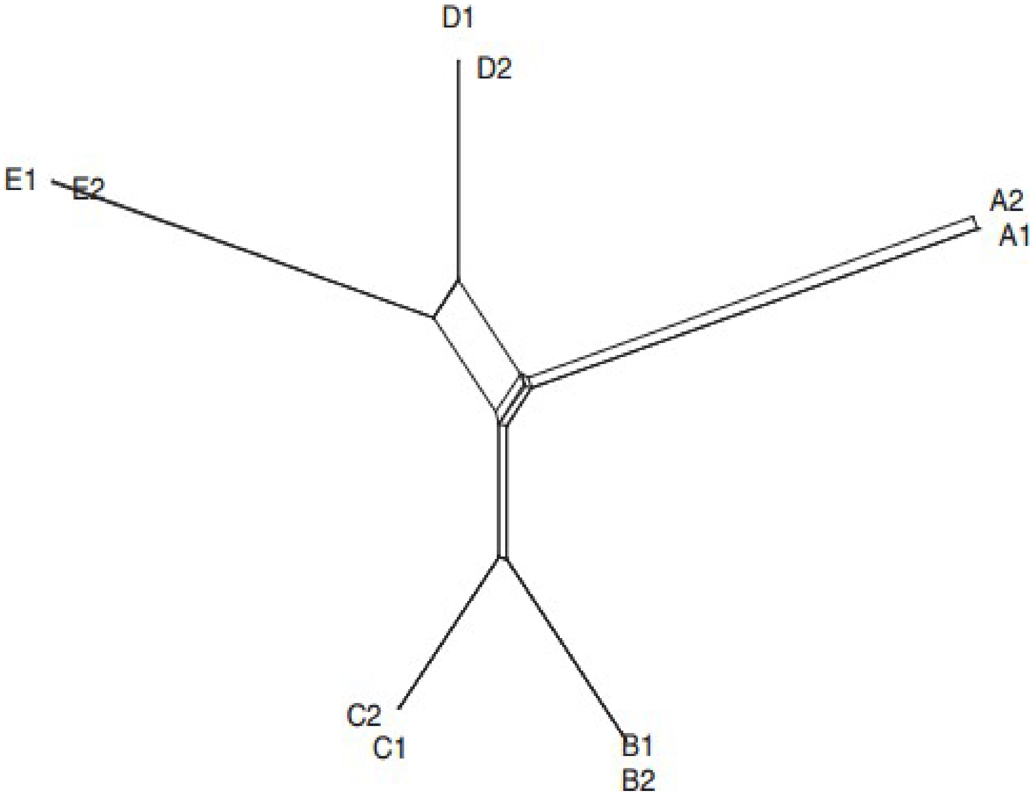
- As a consequence, it can be useful to detect and represent "contamination" between the witnesses.

A-S Chronicle tradition: NN





Good job!



- NN makes visible a feature that traditional stemmata usually conceal;
- A as a source of contamination of the BC-branch (in addition to the DE-branch)

Historical data corroborate the result

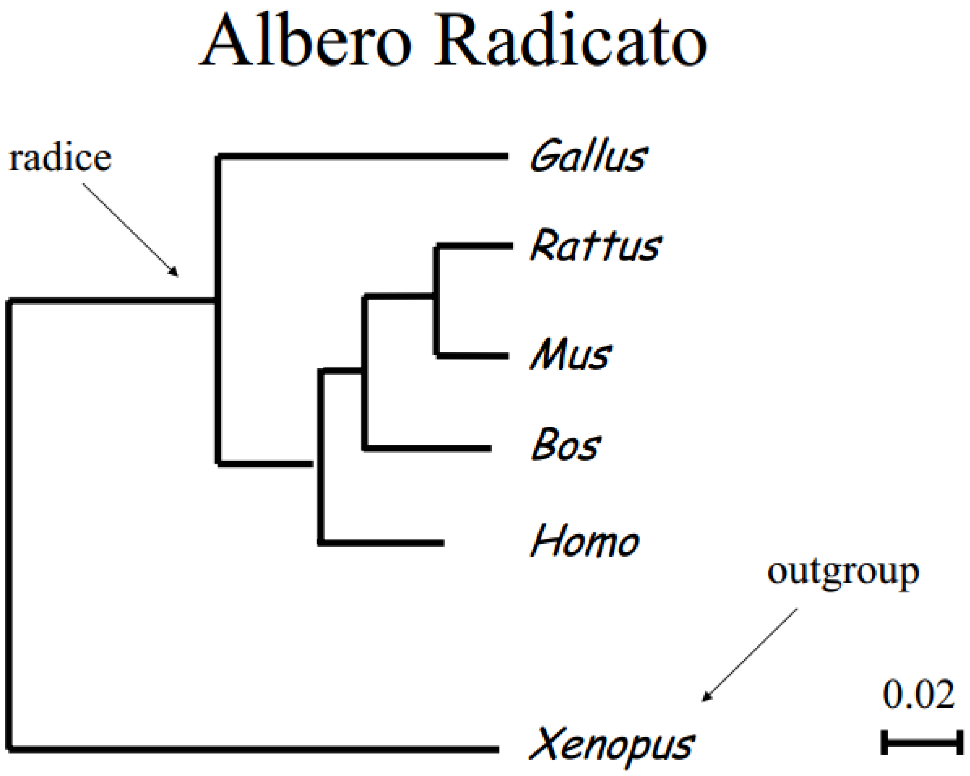
- Susan Taylor (1983, xxxv) observes that {A} e {B C} share «a few supplementary notes», namely the annals 957, 959, 971, 976 (only in C), 977;
- Taylor’s hypothesis is that a group of annals of A (known as «Continuation 2») was sent to Abingdon, where an extended version of the *Chronicle* was compiled. Her conclusion is that «of this Abingdon Chronicle, BC were copies».
- NN is (much) better than manually drawn stammata in detecting this line of contamination (from A to B,C).

(Phylogenetic) graphs vs. traditional stemmata, I

- Graphs are **unrooted** (more like bushes than trees)
- Graphs are **unoriented** (though they can be both rooted and oriented subsequently, on the basis e.g. of historical data)

How to root a phylogram

An **outgroup** is a related group of organisms that serves as a reference group when determining the evolutionary relationships of the in-group.



Outgroup and texts

«In the case of Gregory’s homilies, it was possible to use what in biology is called an **outgroup**. This is rather unusual in philology, but it occasionally happens that an indirect witness (a translation or another recension of the work) can be proved to be independent from the archetype of the direct tradition and can therefore be used as an outgroup (see 3.2.8). In this instance, we observed that **translations into Latin (ca. 400, made by Rufinus of Aquileia) and Armenian (ca. 500) had preserved shared readings that must be seen as primary** and which are absent from all the Greek manuscripts (Macé 2011). This discovery has enormous consequences for the stemma and for the history of the tradition. It means that all Greek manuscripts (and the Syriac translation) share secondary readings and therefore depend upon a hyparchetype φ. This hyparchetype must be dated before the Syriac translation, which was reworked and completed around 625 in Cyprus. »

(*Handbook of Stemmatology*, pp. 285-287)

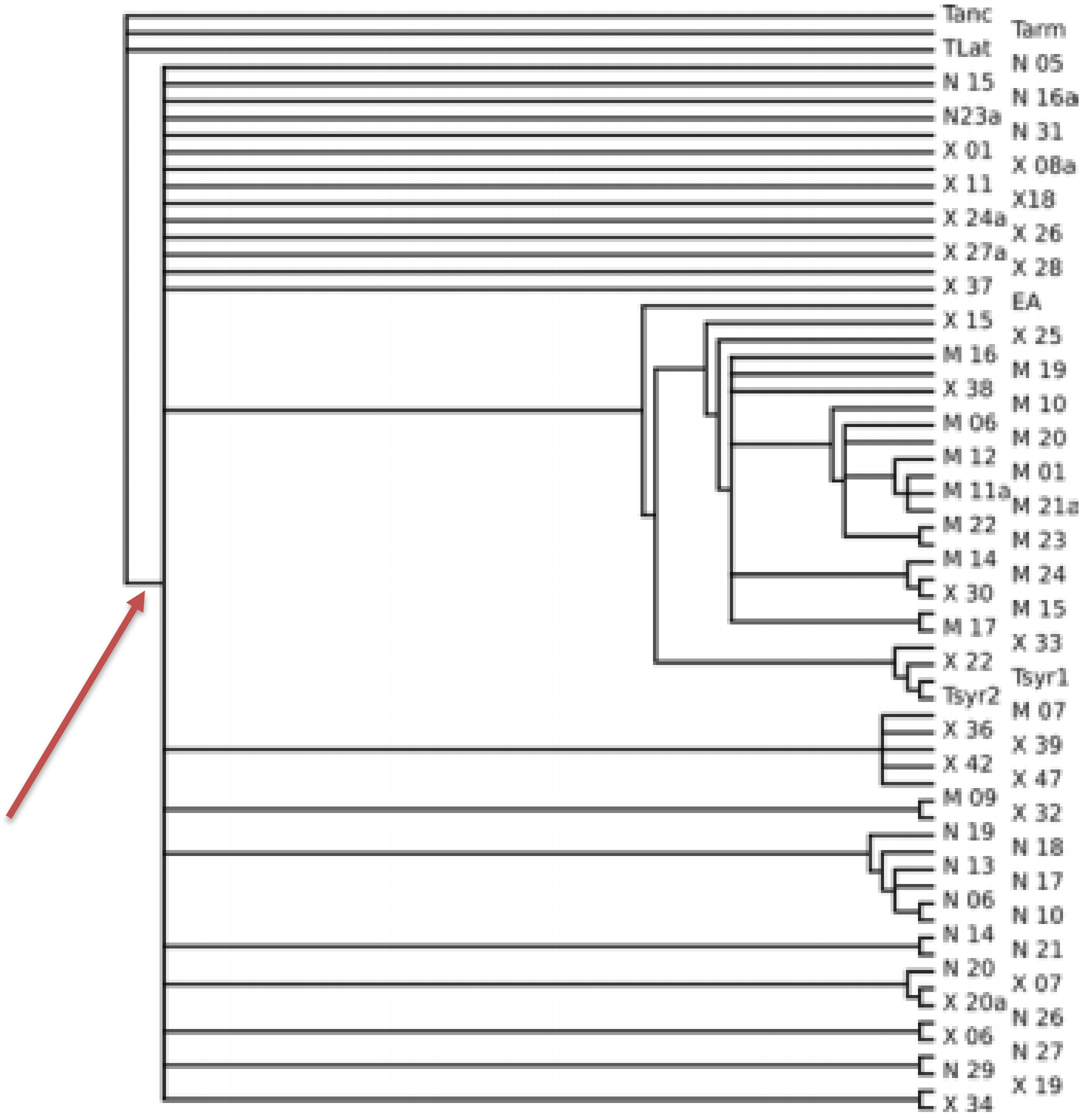


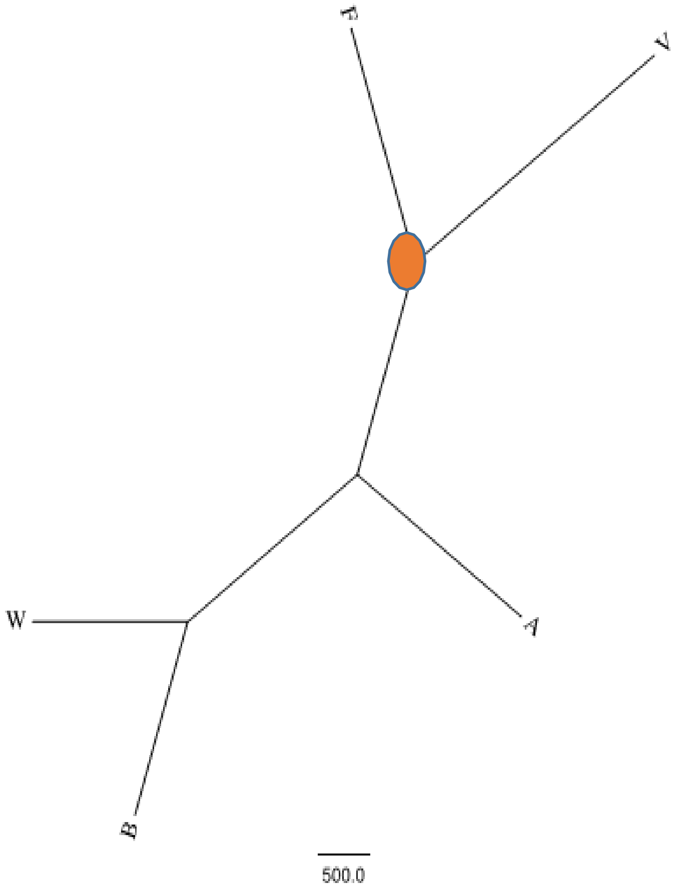
Fig. 4.5-14: Phylogenetic tree of manuscripts with a complete collection (fifty-nine manuscripts) containing homily 27 (branch length is not represented in the plot), rooted on *Tarm*/*Tlat*.

(Phylogenetic) graphs vs. traditional stemmata, II

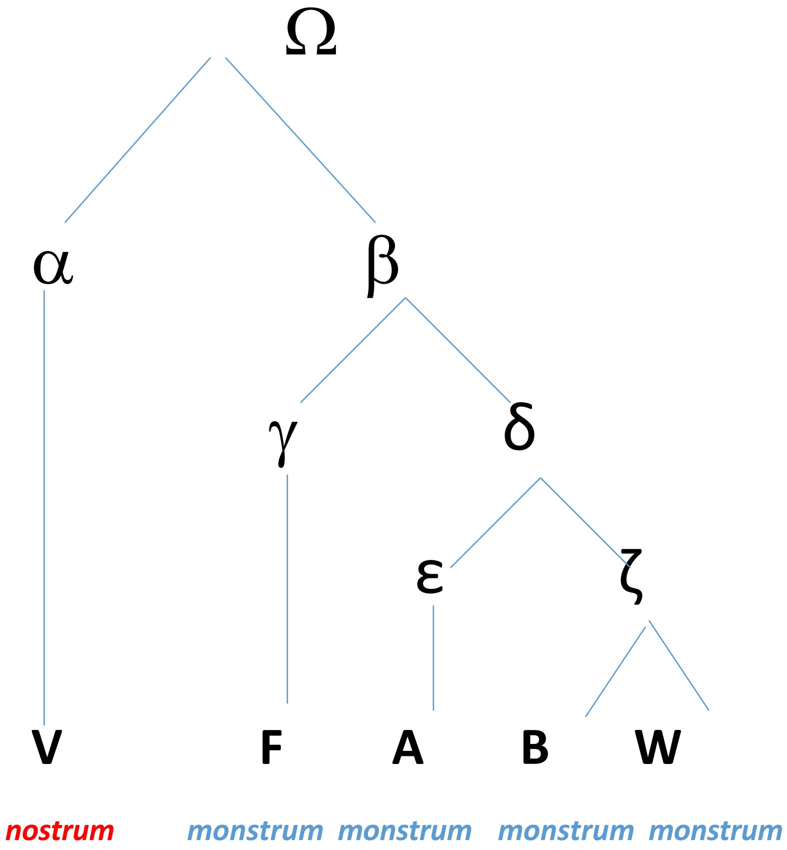
- **Internal nodes issue**: in a traditional stemma internal nodes correspond to *codices interpositi* (reconstructed ancestors); in phylogenetic glyphs internal nodes are generated by the mathematical model(s)

- A question remains open (the grounding question): can they be of any use in a reconstructive perspective?

A 28 (f. 17r)	quia inde incipiemus monstrum dicere
B 29 (f. 5d)	q. i. incipiemus dicere monstrum
W 29 (f. 126b)	q. i. incipiemus dicere monstrum
V 31 (f. 13r)	quia inde incipiam dicere nostrum [iter]
F 31 (f. 5bis)	quia inde incipiam dictum monstrum



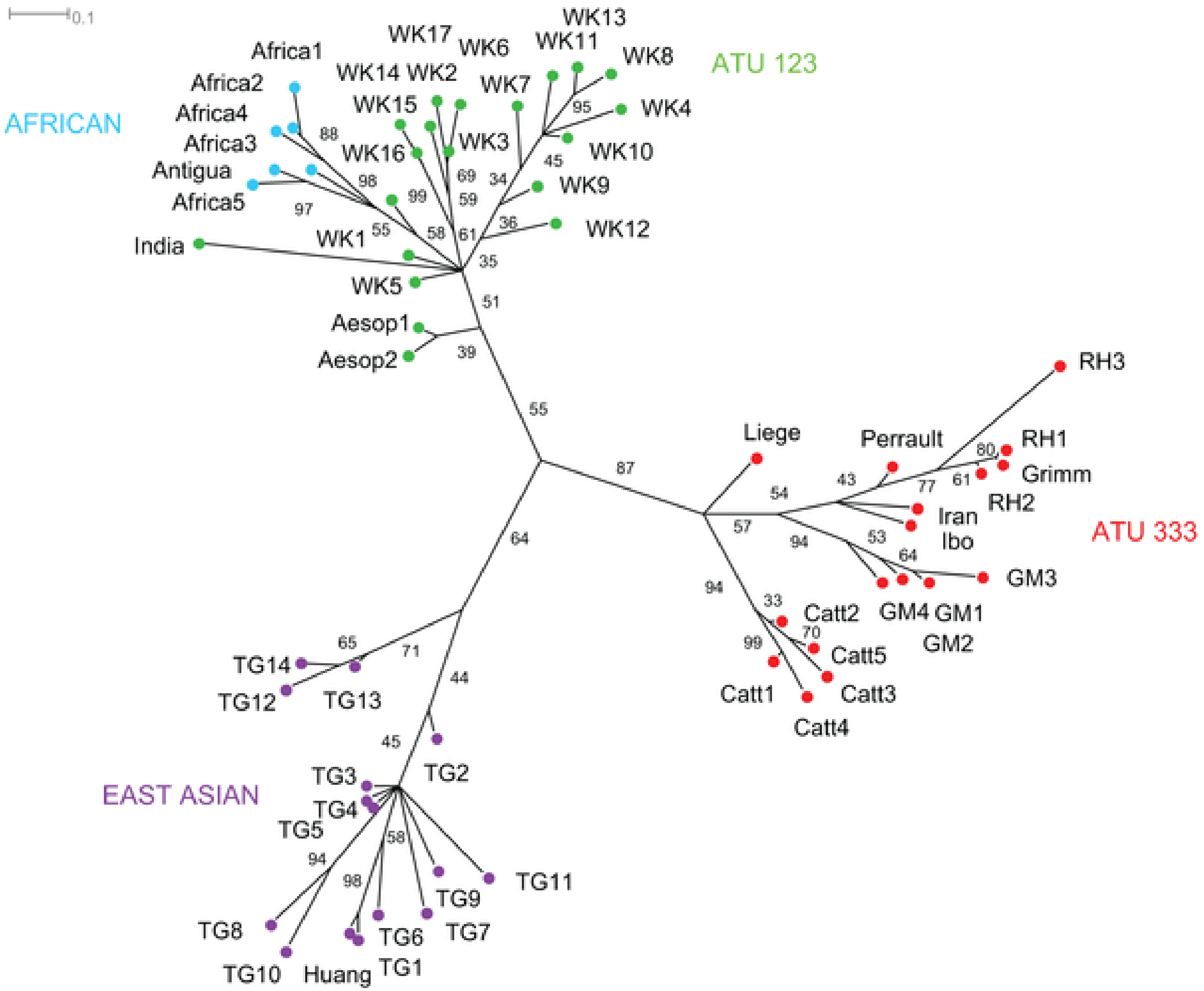
PAUP, PS



What’s next

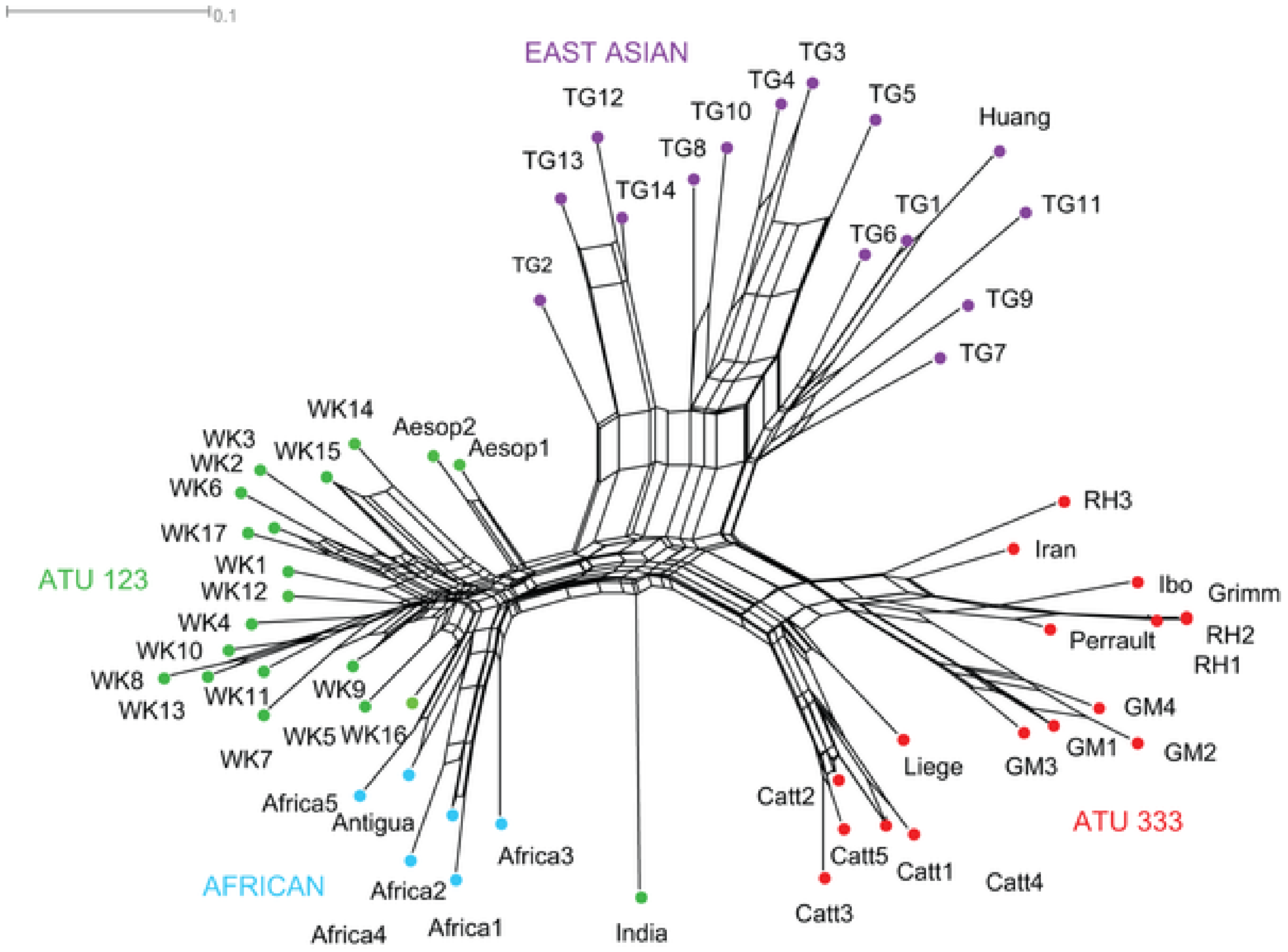
Bayesian Method?
LLMs?

Figure 3. Maximum clade credibility tree returned by the Bayesian phylogenetic analysis of the tales.



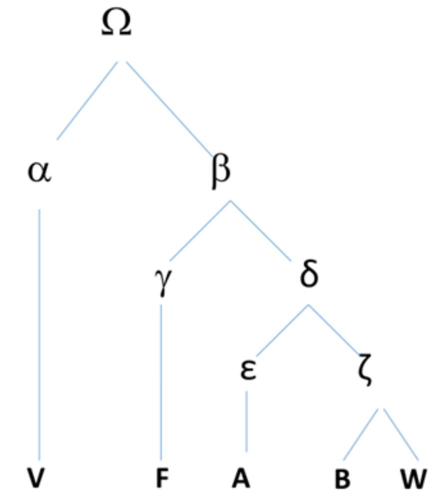
Tehrani JJ (2013) The Phylogeny of Little Red Riding Hood. PLOS ONE 8(11): e78871.
https://doi.org/10.1371/journal.pone.0078871
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078871

Figure 4. Split graph returned by the NeighbourNet analysis of the tales.

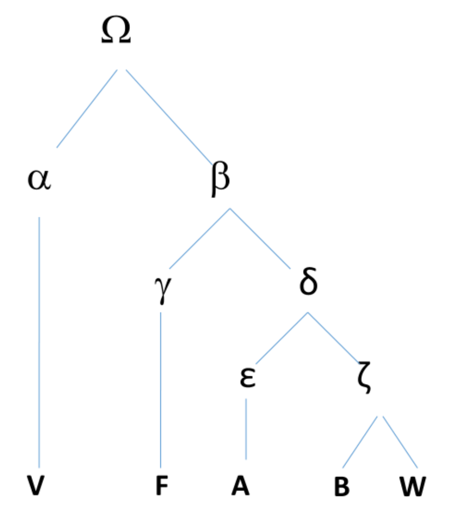


Tehrani JJ (2013) The Phylogeny of Little Red Riding Hood. PLOS ONE 8(11): e78871. https://doi.org/10.1371/journal.pone.0078871
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078871

```
      Z
      |
  ----|----
 |          |
 V          |
          --|--
         |     |
       Fr2     |
         |     |
  -------|     |
 |       |     |
Fr1      F     R
 |             |
     ---------|-------
    |         |       |
   VB         VA      P
                       \
                        L
                         \
                        TB
                          \
                          TA
```



```
                   /---- Fr1
          /-------|
         |         \---- F
         |
   /-----|                 /---- L
  |      |         /-------|
  |      |        |        |---- TB
  |      |        |
--|      |        |        /---- TA
  |      \--------|-------|
  |               |       \---- R
  |               |
  |               |        /---- VA
  |                \-------|
  |                        \---- P
  |
   \------------------------- V
```




thank
you:-)