Computational Anatolian Phylogenetics Using Maximum Parsimony

Oscar Billing & Erik Elgh

Uppsala Universitet

ICHL 26, Heidelberg

UPPSALA
UNIVERSITET

Introduction

The Anatolian language family

Subbranch of Indo-European

Consists of minimally 8 languages

Hittite	Palaic
Luwian	Lydian
Lycian	Carian
Sidetic	**Pisidian**

Extinct! (ca. 19th cent. BCE – 2nd cent. CE)

Anatolian phylogenetics

No universal agreement on topology in Anatolian phylogenetics

Problem mostly centers on position of Palaic and Lydian

Anatolian phylogenetics

No universal agreement on topology in Anatolian phylogenetics

Problem mostly centers on position of Palaic and Lydian

Clade [Luwian – Lycian – Palaic] (Oettinger 1979; Starke 1997; Yakubovich 2010; Kloekhorst 2022)

Clade [Luwian – Lycian – Lydian] (Rieken 2017; Sasseville 2020)

Clade [Hittite – Palaic] (tentatively, Carruba 1970)

Non-tree-like dialect group (Watkins 2001; Melchert 2003)

Anatolian phylogenetics

No universal agreement on topology in Anatolian phylogenetics

Problem mostly centers on position of Palaic and Lydian

Clade [Luwian – Lycian – Palaic] (Oettinger 1979; Starke 1997; Yakubovich 2010; Kloekhorst 2022)

Clade [Luwian – Lycian – Lydian] (Rieken 2017; Sasseville 2020)

Clade [Hittite – Palaic] (tentatively, Carruba 1970)

Non-tree-like dialect group (Watkins 2001; Melchert 2003)

Lack of consensus warrants use of **computer aided methods**

Anatolian phylogenetics

Central issue: **scarce attestation**

- **Lack** of material

- Meanings of many words **uncertain**

Anatolian phylogenetics

Central issue: **scarce attestation**

- **Lack** of material

- Meanings of many words **uncertain**

Knowledge, of **Carian**, **Pisidian**, and **Sidetic** too limited to include in study

Anatolian phylogenetics

Impossible to compile *exhaustive* and *reliable* **word lists**

Lexical data, often used in phylogenetics, is thus *not appropriate* for Anatolian

Phylogenetic signal centered on **phonological** and **morphological** developments

Anatolian phylogenetics

Impossible to compile *exhaustive* and *reliable* **word lists**

Lexical data, often used in phylogenetics, is thus *not appropriate* for Anatolian

Phylogenetic signal centered on **phonological** and **morphological** developments

Suitable model: *maximum parsimony*

Maximum Parsimony

Maximum parsimony is closely related to the *Principle of Economy*

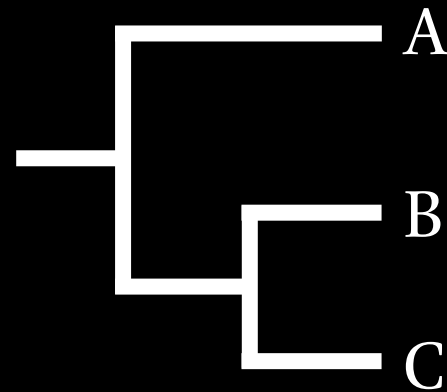The preferred solution is the one with the least amount of change

- Minimizes homoplasy (parallel innovations etc.)

- Defines groups based on shared innovations
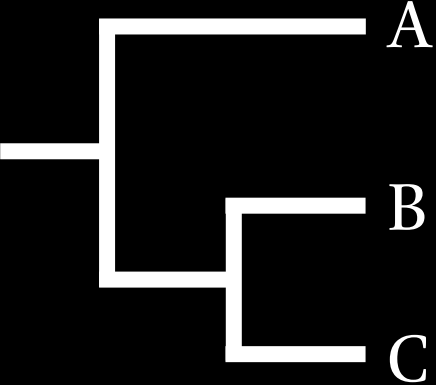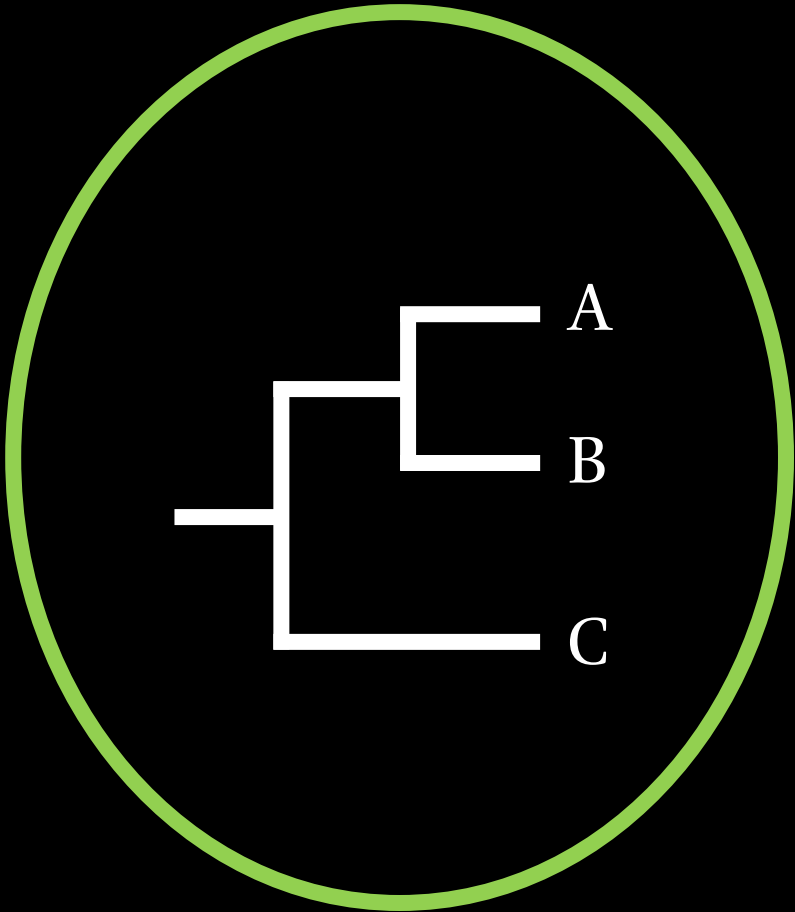
Maximum Parsimony

	Lang. A	Lang. B	Lang. C
Char. 1	√	√	X
Char. 2	√	√	X
Char. 3	X	√	√

Maximum Parsimony

	Lang. A	Lang. B	Lang. C
Char. 1	√	√	X
Char. 2	√	√	X
Char. 3	X	√	√

Maximum Parsimony

	Lang. A	Lang. B	Lang. C
Char. 1	√	√	X
Char. 2	√	√	X
Char. 3	X	√	√

Bootstrap Analysis

Bootstrap (Efron 1979, see Felsenstein 2003) is a method that can be used to assess the robustness of the resulting phylogeny

It consists of making new datasets by resampling with replacement, and assumes that the data points used in the analysis accurately represents the true distribution

Thousands of these new analyses are run, and the proportion of these that yield the clades on our tree are indicated
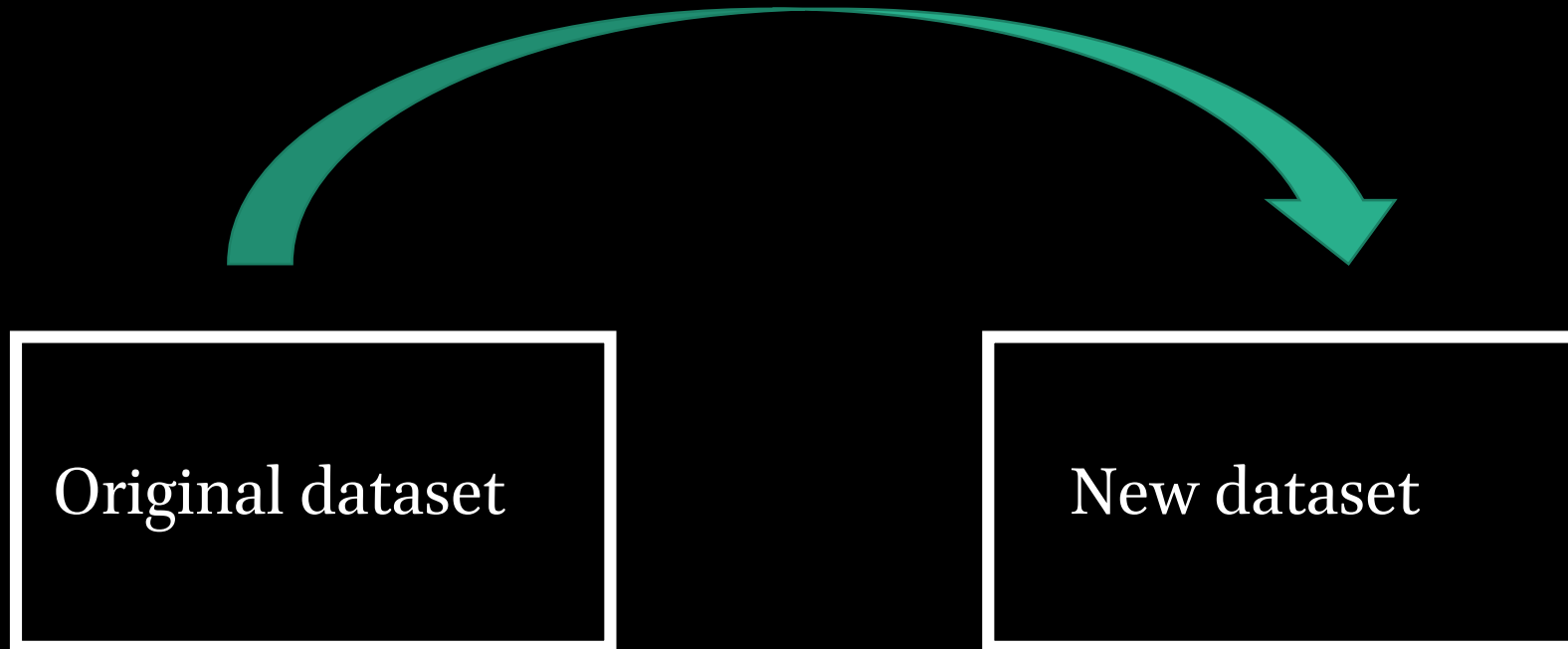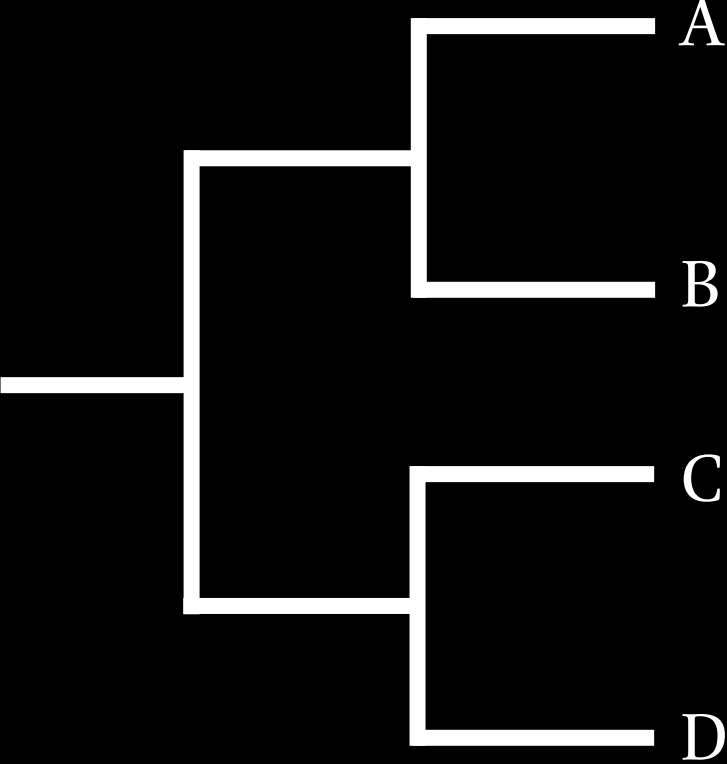
Bootstrap Analysis

Original dataset

Bootstrap Analysis

Original dataset

New dataset

Bootstrap Analysis

Bootstrap Analysis

Characters

Character set

27 characters

12 phonological   15 morphological
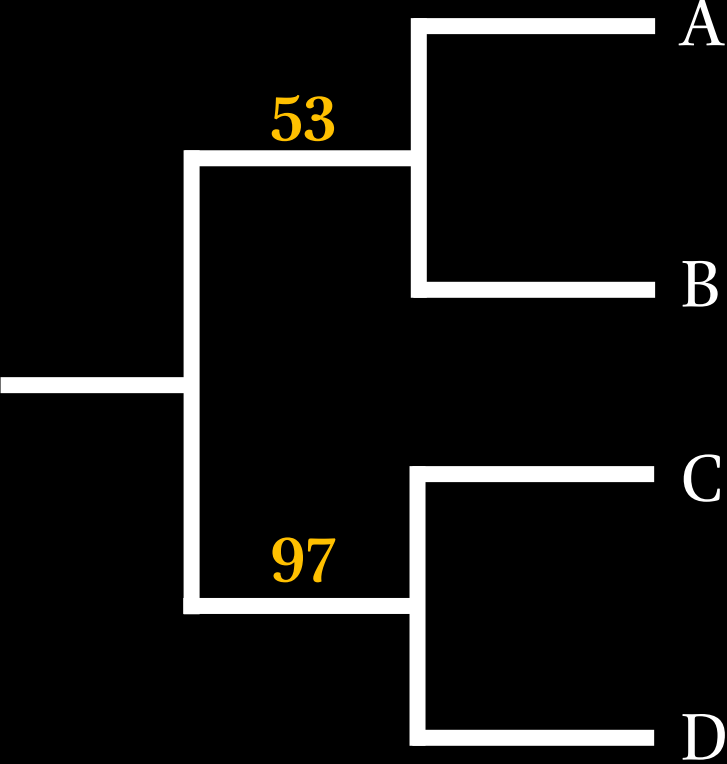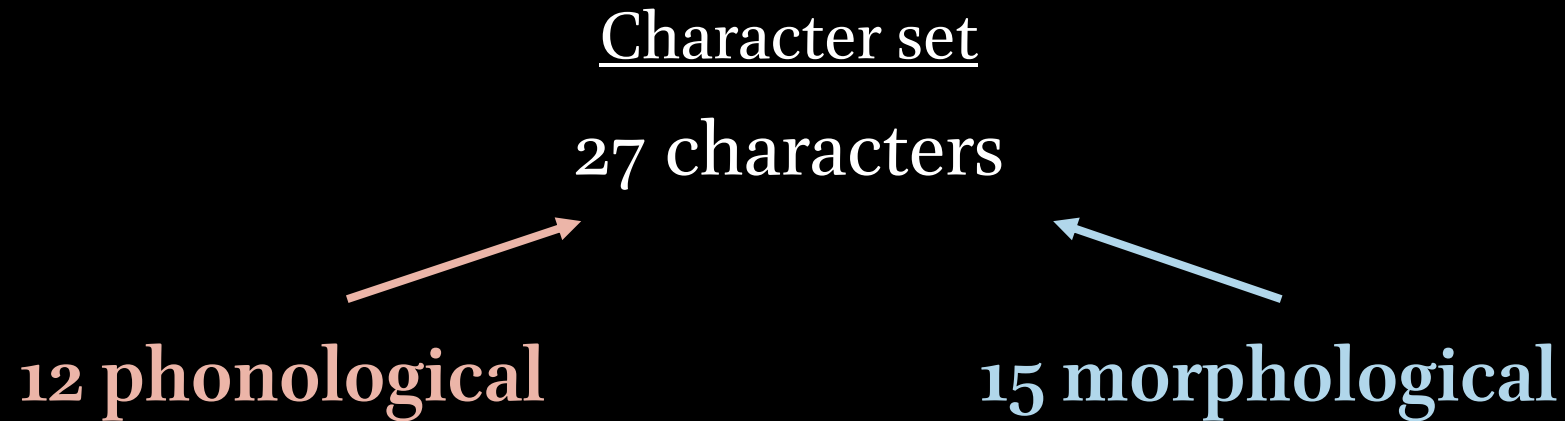
Gathered from **previous literature**

Expanded by **original research**

Ancestral state specified in **25/27 characters**

(NB missing data: Luw. 2, Lyd. 5, Pal. 2)

Weighting

Weighting of characters = crucial for resulting tree

For consistency: **categorization procedure**

	Weight 1	Weight 2	Weight 3	Weight 4
Phonology	**Trivial sound change**	**Sound change**	**Non-trivial sound change**	**Irregular sound change**
Morphology	**Allomorph generalization** **Morpheme loss**	**Analogy (e.g. proportional)**	**Non-trivial analogy**	**Highly non-trivial analogy**

Based on **linguistic experience** and **comparison** to changes elsewhere
(NB some subjectivity is unavoidable!)

Weighting

Example (phonology):   **Raising** *e* > *i* /*j*_

Sound law, by which an *e* is raised to *i* when preceded by yod

Conditioned sound change, thus given **weight 2**

Weighting

Example (phonology):   **Raising *_e_ > _i_ / _j__**

Sound law, by which an *_e_ is raised to _i_ when preceded by yod

Conditioned sound change, thus given **weight 2**

Example (morphology):   **Generalize 1SG.PRET.ACT. ending *-_Ha_**

Generalization of a verbal ending, from certain conjugation type

Simply levelling of verbal system, given **weight 1**

Directionality

Most characters are unidirectional (cannot be reversed)

- Phonological mergers   (original distribution irrecoverable)

- Morphological levelling   (loss of model)

Directionality

Most characters are unidirectional (cannot be reversed)

- Phonological mergers   (original distribution irrecoverable)

- Morphological levelling   (loss of model)

Example (phonology):   **Raising** $*e > i \, / \, j\_$ ($0 > 1$)

New instances of i merge with inherited $*i$

Original distribution is irrecoverable

Thus, $0 > 1$ is *directional*, as $1 > 0$ is *impossible*

Multistate characters

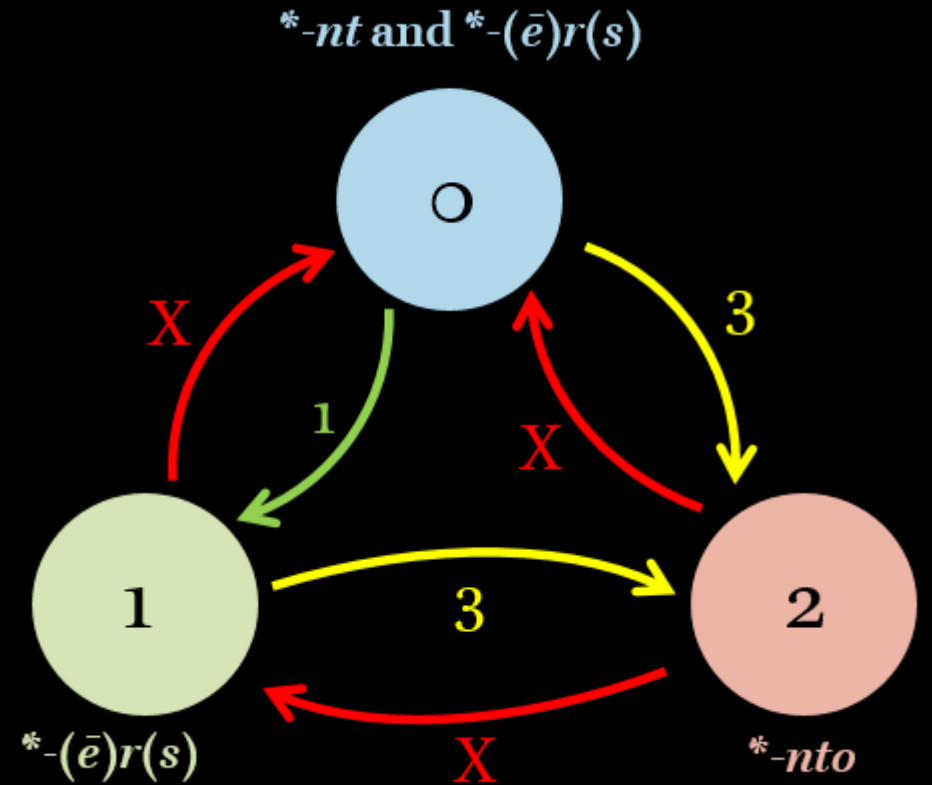Some characters have **more than 2 possible states**, requiring special coding

Ex.: Generalization of 3PL.PRET.ACT. allomorph

Anc. state	Hittite	Palaic	Lydian	Luwian	Lycian
0	1	{02}	1	2	2

0 = **both *-*nt* and *-($\bar{e}$)*r*(*s*)**

1 = **generalize *-($\bar{e}$)*r*(*s*)**

2 = **introduce *-*nto* (from middle voice)**

Results

Results

Best tree (score 66)

Proto-Anatolian

Hittite

Palaic

Lydian

Luwian

Proto-Luwic

Lycian
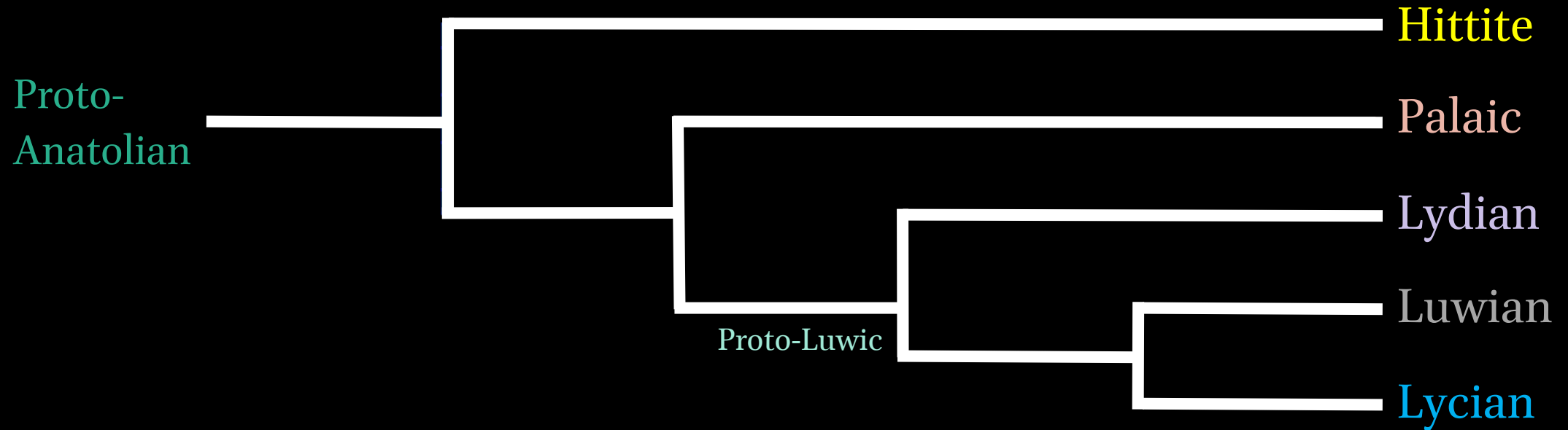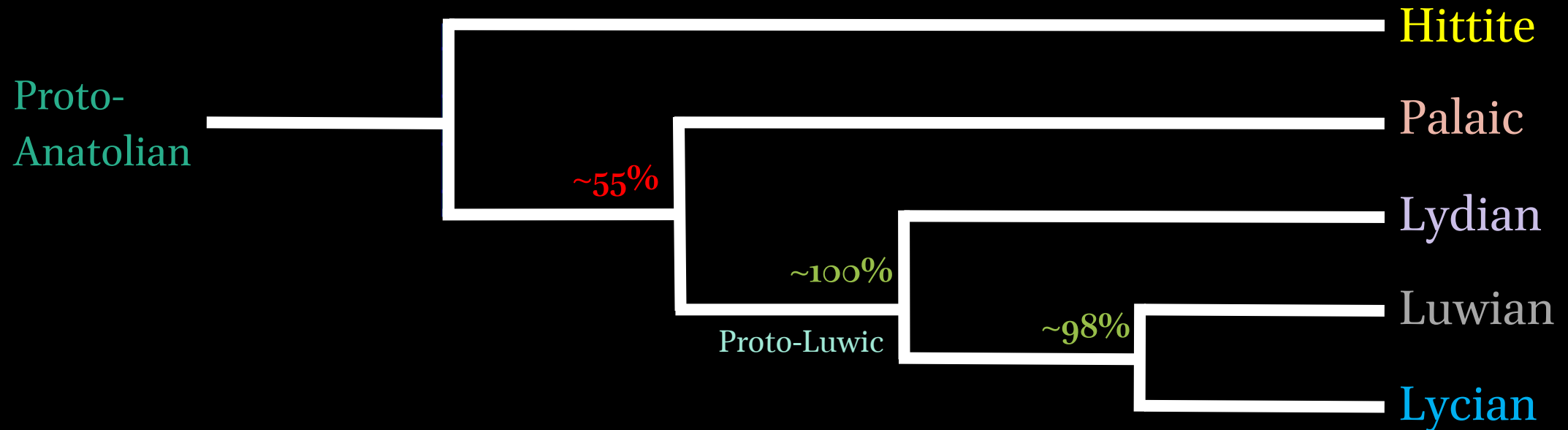
NB differs from abstract! (no [Hittite - Palaic] clade)

Results

Best tree with **bootstrap frequencies** (100 000 iterations)

Note **low frequency** for [Palaic – Luwic] clade, **robustness** elsewhere
(freq. ~45% for [Hittite – Palaic] clade, parsimony score 67 vs. 66)

Results

Most reliable current tree

Proto-Anatolian

Proto-Luwic

Hittite

Palaic

Lydian
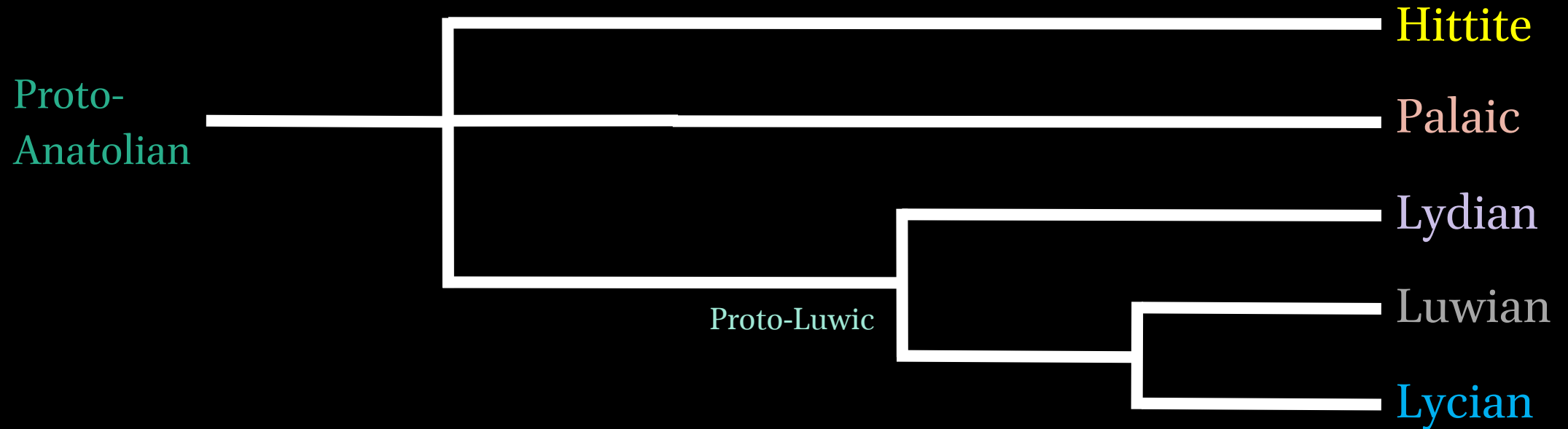
Luwian

Lycian

Polytomy indicates *uncertainty*!

Conclusions and prospects

Parsimony analysis **strongly** supports

        Luwo-Lycian clade [Luwian – Lycian]

        Luwic clade   [Lydian – [Luwian – Lycian]]

Conclusions and prospects

Parsimony analysis **strongly** supports

        Luwo-Lycian clade   [Luwian – Lycian]

        Luwic clade   [Lydian – [Luwian – Lycian]]

  and **weakly** supports

        Palao-Luwic clade   [Palaic [Lydian – [Luwian – Lycian]]]

Conclusions and prospects

Parsimony analysis **strongly** supports

       Luwo-Lycian clade   [Luwian – Lycian]

       Luwic clade   [Lydian – [Luwian – Lycian]]

  and **weakly** supports

       Palao-Luwic clade   [Palaic [Lydian – [Luwian – Lycian]]]

  with Hittite as outgroup

Conclusions and prospects

Parsimony analysis **strongly** supports

Luwo-Lycian clade   [Luwian – Lycian]

Luwic clade   [Lydian – [Luwian – Lycian]]

and **weakly** supports

Palao-Luwic clade   [Palaic [Lydian – [Luwian – Lycian]]]

with Hittite as outgroup

A **polytomy**, indicating *uncertainty*, between Hittite, Palaic, and Luwic
seems **most prudent** at the present level of knowledge

Conclusions and prospects

- **Further research** into each individual language (particularly Palaic) could **alter/improve** the best tree

Conclusions and prospects

- **Further research** into each individual language (particularly Palaic) could **alter/improve** the best tree

- **Transparency** of parsimony analysis allows **close inspection by specialists**

- Analysis could be rerun with:

    Additional characters

    Different **character coding**

    Modified weights

Conclusions and prospects

- **Further research** into each individual language (particularly Palaic) could **alter/improve** the best tree

- **Transparency** of parsimony analysis allows **close inspection by specialists**

- Analysis could be rerun with:

    Additional characters

    Different **character coding**

    Modified weights

- Tree here reflects *best tree given current knowledge and assumptions*

Thank you!
Danke für Ihre Aufmerksamkeit!

References

Carruba, O. (1970). *Das Palaische. Texte, Grammatik, Lexikon.* Wiesbaden: Otto Harrassowitz.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.

Felsenstein, J. (2003). *Inferring phylogenies*. Sinauer Associates.

Kloekhorst, A. (2022). Anatolian. In T. Olander (Ed.), *The Indo-European Language Family. A Phylogenetic Perspective* (pp. 63–82). Cambridge: Cambridge University Press.

Melchert, H. C. (2003). The dialectal position of Lycian and Lydian within Anatolian. In M. Giorgieri, M. Salvini, M.-C. Trémouille, & P. Vannicelli (Eds.), *Licia e Lidia prima dell'Ellenizzazione. Atti del Convegno internazionale. Roma, 11-12 ottobre 1999* (pp. 265–272). Roma: Consiglio Nazionale delle Ricerche.

Oettinger, N. (1979). Die Gliederung des anatolischen Sprachgebietes. *Zeitschrift für vergleichende Sprachforschung*, 92, 74–92.

Rieken, E. (2017). The dialectology of Anatolian. In J. Klein, B. Joseph, & M. Fritz (Eds.), *Handbook of Comparative and Historical Indo-European Linguistics. Vol. 1* (pp. 298–308). Berlin – Boston: De Gruyter Mouton.

Sasseville, D. (2020). *Anatolian Verbal Stem Formation. Luwian, Lycian and Lydian.* Leiden – Boston: Brill.

Starke, F. (1997). Troia im Kontext des historisch-politischen und sprachlichen Umfeldes Kleinasiens im 2. Jahrtausend. *Studia Troica, 7*, 447–487.

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Yakubovich, I. (2010). *Sociolinguistics of the Luvian Language.* Leiden – Boston: Brill.

Watkins, C. (2001). An Indo-European Linguistic Area and its Characteristics: Ancient Anatolia. Areal Diffusion as a Challenge to the Comparative Method? In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics* (pp. 43–63). Oxford: Oxford University Press.