

Computational Anatolian phylogeny using maximum parsimony

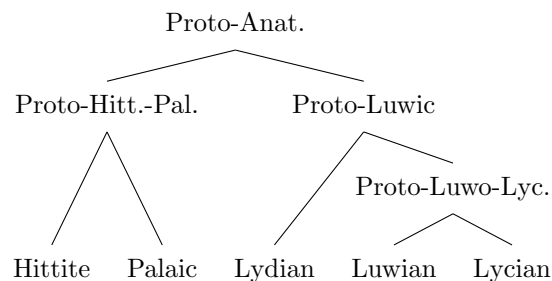
The Anatolian languages constitute an extinct branch of the Indo-European language family, attested across modern day Turkey from ca. the 19th cent. BCE to the 2nd cent. CE (Zinko 2017). Prominent members include Hittite, Luwian, Lydian, and Lycian.

Previous traditionally oriented work on the internal phylogeny of Anatolian has not reached a consensus. For example, in some studies (Oettinger 1979; Kloekhorst 2022), Palaic, Luwian, and Lycian form a clade, whereas other treatments (Melchert 2003; Rieken 2017) tend to assume a closer relationship between Lydian, Luwian, and Lycian to the exclusion of Palaic. It is consequently warranted to explore alternative methods for determining the topology of the Anatolian tree.

Recent studies applying computational methods to linguistic phylogenies have mostly operated with lexical cognate data as the sole input (e.g. Bouckaert et al. 2012; Chang et al. 2015; Ringe et al. 2002; Ringe et al. use some morphological characters but still principally base their analysis on lexical data). This MO is not viable for Anatolian, as the languages involved are too scarcely attested to allow for the compilation of a reliable cognate data set. Indeed, the material available for Carian, Sidetic, and Pisidian is too scanty for any analysis. Rather, Anatolian phylogeny must operate primarily with phonological and morphological data.

A solid candidate method for conducting computational phylogeny using non-cognate based data sets is *Maximum Parsimony*. This study employs PAUP* (Swofford 2003) to infer an Anatolian tree on the basis of a data set consisting of 27 characters (12 phonological and 15 morphological) gathered from existing literature and additional research. The taxa involved are Hittite, Palaic, Lydian, Luwian, and Lycian. Considering that the characters used here are predominantly the result of the historical-comparative method, a root state is often possible to assign confidently. Consequently, our tree is rooted. Characters are assigned a weight from 1–4 on the basis of pre-established parameters (e.g. *unconditioned sound changes* are weighted 1 and *sporadic sound changes* are weighted 4). It should be noted that these parameters are unavoidably to some extent subjective, but we do not expect any strong objections from specialists.

Our analysis gives the following most parsimonious tree:



A bootstrap analysis (Felsenstein 1985) indicates that our tree is highly robust. However, considering that the innovations assumed for Proto-Hittite-Palaic under this topology are rather trivial, it may be most prudent to assume a polytomy between Hittite, Palaic, and Proto-Luwic, pending further evidence.

The method used in this study could be exploited for other scarcely attested extinct language families (e.g. Sabellic). An advantage over alternative methods is furnished by the transparency in the grounds on which trees are evaluated (cf. Hammarström et al. 2019: 236). Accordingly, a classically trained historical linguist and/or specialist on the language family at hand is given the opportunity to qualitatively assess the validity of the developments postulated by the analysis.

References

- Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., ... Atkinson, Q. D. (2012). Mapping the origins and expansion of the Indo-European language family. *Science*, *337*(6097), 957–960.
- Chang, W., Hall, D., Cathcart, C., & Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*, *91*, 194–244.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*(4), 783–791.
- Hammarström, H., Rönchen, P., Elgh, E., & Wiklund, T. (2019). On computational historical linguistics in the 21st century. *Theoretical Linguistics*, *45*(3-4), 233–245.
- 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.
- Ringe, D., Warnow, T., & Taylor, A. (2002). Indo-European and computational cladistics. *Transactions of the Philological Society*, *100*(1), 59–129.
- Swofford, D. L. (2003). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. <http://paup.phylosolutions.com/>. Sunderland, Massachusetts: Sinauer Associates.
- Zinko, C. (2017). The documentation of Anatolian. In J. Klein, B. Joseph, & M. Fritz (Eds.), *Handbook of Comparative and Historical Indo-European Linguistics. Vol. 1* (pp. 239–249). Berlin – Boston: De Gruyter Mouton.