

## Problems in Computational Phylogenetics: The Germano-Italo-Celtic Clade

**Authors :** Laura Mclean

**Abstract :** A recurring point of interest in computational phylogenetic analysis of Indo-European family trees is the inference of a Germano-Italo-Celtic clade in some versions of the trees produced. The presence of this clade in the models is intriguing as there is little evidence for innovations shared among Germanic, Italic, and Celtic, the evidence generally used in the traditional method to construct a subgroup. One source of this unexpected outcome could be the input to the models. The datasets in the various models used so far, for the most part, take as their basis the Swadesh list, a list compiled by Morris Swadesh and then revised several times, containing up to 207 words that he believed were resistant to change among languages. The judgments made by Swadesh for this list, however, were subjective and based on his intuition rather than rigorous analysis. Some scholars used the Swadesh 200 list as the basis for their Indo-European dataset and made cognacy judgements for each of the words on the list. Another dataset is largely based on the Swadesh 207 list as well although the authors include additional lexical and non-lexical data, and they implement 'split coding' to deal with cases of polymorphic characters. A different team of scholars uses a different dataset, IECOR, which combines several different lists, one of which is the Swadesh 200 list. In fact, the Swadesh list is used in some form in every study surveyed and each dataset has three words that, when they are coded as cognates, seemingly contribute to the inference of a Germano-Italo-Celtic clade which could happen due to these clades sharing three words among only themselves. These three words are 'fish', 'flower', and 'man' (in the case of 'man', one dataset includes Lithuanian in the cognacy coding and removes the word 'man' from the screened data). This collection of cognates shared among Germanic, Italic, and Celtic that were deemed important enough to be included on the Swadesh list, without the ability to account for possible reasons for shared cognates that are not shared innovations, gives an impression of affinity between the Germanic, Celtic, and Italic branches without adequate methodological support. However, by changing how cognacy is defined (ie. root cognates, borrowings vs inherited cognates etc.), we will be able to identify whether these three cognates are significant enough to infer a clade for Germanic, Celtic, and Italic. This paper examines the question of what definition of cognacy should be used for phylogenetic datasets by examining the Germano-Italo-Celtic clade as a case study and offers insights into the reconstruction of a Germano-Italo-Celtic clade.

**Keywords :** historical, computational, Italo-Celtic, Germanic

**Conference Title :** ICHLCL 2025 : International Conference on Historical Linguistics and Comparative Linguistics

**Conference Location :** Paris, France

**Conference Dates :** July 19-20, 2025