Incomplete lineage sorting as a complicating factor in phylogenetic reconstruction: some Indo-European examples

In evolutionary biology, it is a well-known fact that the family trees of individual replicators (e.g. of an individual gene and its alleles) may appear to contradict the topology of the species tree. This is because a polymorphism present in the common ancestor of three or more species may pass unresolved through multiple speciation events, provided that they occur in sufficiently quick succession. In other words, competing variants present in an ancestral species may continue to co-occur in its descendants, rather than being neatly segregated between the daughter lineages. The same allele may eventually undergo fixation in two species independently, producing an illusion of a sister relationship. The examination of a large number of allele trees shows that they collectively support a certain phylogeny on an average, while the few discordant cases point to complications caused by imperfect lineage sorting.

This phenomenon has an analogue in language evolution. Apart from spurious relationships suggested by lateral transfer (linguistic borrowing and other forms of contact phenomena), we should also pay attention to cases where an ancestral variation passes through consecutive splits in a family tree of languages. If A_1 and A_2 stand for competing variant forms of the same linguistic unit, present already in a language ancestral to languages X, Y and Z, the later distribution of their reflexes – say, A_1 in X and Y, but A_2 in Z – does not guarantee that X and Y are more closely related to each other than either is to Z. It is thinkable that, for example, X was the first lineage to branch off, but the $A_1 \sim A_2$ polymorphism was present throughout the common history of Y and Z, and the replacement of A_2 by A_1 took place independently in X and Y, while A_2 was generalised in Z. Given the fact than extensive variation is normal in all living languages, and that the resolution of a conflict between competing variants requires considerable time, we must expect incomplete lineage sorting to be occasionally observable in language evolution.

Since the history of the Indo-European (IE) family has been studied very thoroughly and the topology of at least some parts of the IE family tree is securely established, we can see if the evidence collected by Indo-Europeanists throws any light on the more general problem of incomplete lineage sorting. We shall try to identify a few characteristic examples of the phenomenon from different strata of language structure (phonology, morphology and vocabulary) in a variety of IE languages. It will be argued that variation and the mechanisms of its resolution have to be taken into account in the evaluation of phylogenetic hypotheses.