

***The Covariation of Genes and Languages:
Theoretical and Methodological Background with Examples from the Americas
and Siberia***

Stephen L. Zegura

Department of Anthropology, University of Arizona

Different referents for the concept of similarity underlie competing classificatory paradigms in both evolutionary biology and linguistics. Today, cladistics principles govern historical reconstruction in both fields with shared, derived, homologous characters employed as the only arbiter of genealogical relationship. In addition, the major causes of homoplasy in biological data have analogues in linguistics. After a short review of the methodological parallels in evolutionary biology and historical linguistics, the topic of different genomic compartments will be addressed within the context of the framework provided by the position paper.

Literally hundreds of publications have investigated the relationship of language and human biology. Over the last 20 years much of this research has concentrated on comparisons of genetic and linguistic data sets at global, regional, continental, and local levels. Statistically significant associations between genes and languages have been found at all of these geographical divisions. Unfortunately, the converse is also true. Numerous studies have failed to detect concordance between these two types of data. Deciphering the causes of these different outcomes would constitute an important research program for future collaboration among the various disciplines interested in reconstructing the biocultural history of our species. For instance, specific details about a population system's history and structure can sometimes provide an explanation for these contradictory results. In other cases, differing transmission patterns, inadequate sampling regimes, poor data quality, and inappropriate statistical methodology have been invoked when expectations were not realized. Moreover, differing *a priori* expectations have often been articulated by human evolutionary geneticists and mainstream historical linguists regarding the association of genes and language.

The Americas have proved to be an especially contentious arena for investigating the relationship between genes and language. As hypothesis testing replaced hypothesis compatibility arguments, Greenberg, Turner, and Zegura's postulated tri-partite association of genes and language families has frequently (but not always) been refuted. Contradictory results have emerged both between and within genomic compartments using the same or different statistical techniques, different levels of population, and alternative linguistic classifications. Findings based on Native American and Siberian populations from the Hammer lab will be compared with those of other investigators primarily (but not exclusively) within the context of Y chromosome haplogroup-language associations. Interestingly, Siberian paternal genetic systems exhibit a much better fit with language than do those of Native Americans.