Where Bacteria and Languages Concur

Colin Renfrew

Two articles in this issue mark a substantial advance in our understanding of human population history in the Pacific area. On page 479, Gray et al. (1) report a computational linguistic analysis that offers a detailed and precise scenario for the dispersal and development of the Austronesian languages, and by implication of human populations among the Pacific islands. The authors come down decisively in favor of one of the two major models for the peopling of the Pacific. On page 527, Moodley et al. (2) come to the same conclusion as Gray et al. about the source and trajectory of spread of the human populations in question, based on results from a seemingly unrelated field: the archaegenetics of human gastric bacterial parasites. In the rapidly developing field of computational historical linguistics (3), this impressive reassessment of the Pacific languages and its corroboration from a very different source are likely to have an impact in linguistic studies far beyond the Pacific area.

The reconstruction of Pacific population history, especially in Polynesia, has been a focus of archaeological interest for many years. The recognition of a characteristically decorated pottery style as a marker left by the first human inhabitants of western Polynesia is one of the contributions made by prehistoric archaeology (4). Because this pottery is associated with the first crop cultivators in the area, agricultural dispersal is often seen as a vehicle for language dispersal.

The languages of Polynesia are part of the widely distributed Austronesian language family, one of the largest language families in the world (5). Its more than 1000 constituent languages include the Micronesian and Polynesian subfamilies as well as the languages of Malaya, much of Indonesia, the Philippines, Taiwan, and Madagascar. The origin of this family has been disputed. One theory, favored by many linguists, places the homeland of the Austronesian languages in Taiwan (6), where languages of several Austronesian subfamilies are located and where farming communities existed as early as 5000 years ago. This theory envisages a farming-language dispersal from Taiwan to the Philippines and then to West Polynesia, starting around 5000 years ago. The alternate, gradualist model sees the process starting very much earlier in island Southeast...

ANTHROPOLOGY

References
10.1126/science.1169409
Phylogenetic trees for Pacific human populations. (Top) Tree derived from linguistic data by Gray et al. (Bottom) Tree based on DNA analysis of the bacterium H. pylori by Moodley et al.

Asia (7). Genetic studies have given conflicting results, and human mitochondrial DNA data do not seem to point to a Taiwanese origin for the populations that now speak Austronesian languages (8). The archaeogenetic evidence is, however, not easy to interpret, and there may have been substantial gene flow in recent colonial times (9).

Gray et al. now apply computer-based phylogenetic methods to this problem. Language trees have been a tool in historical linguistics since the 19th century (see the first figure) (10), but the computational analysis enables a more systematic investigation, which also offers a chronology for the various stages. Gray and Atkinson previously used the same method to study the Indo-European language family (11), but that analysis has not yet found favor with most historical linguists. The present analysis of the Pacific languages is, however, based on a very much larger database of more than 400 languages [compared with 87 languages in (11)]. Moreover, it relies for its lexical data on the work of Blust (12) and other linguists generally regarded as the leading authorities on the Austronesian languages.

A remarkably clear scenario emerges (see the second figure, top panel). The dating rests on 10 externally dated calibration points, of which the more ancient are based on the archaeological data for the Austronesian entry into the Philippines, Micronesia, and Eastern Polynesia (13). The overall scenario, however, derives from the topology of the tree, which does not depend on the archaeology. In the scenario, an Austronesian origin in Taiwan ~5200 years ago was followed by a first pause, and then a major pulse or migration dispersal reaching across the Pacific as far as Micronesia ~3000 years ago. A second pause occurred after the settlement of Western Polynesia around 2800 years ago. A second migration after 1500 years ago led to the peopling of Central and Eastern Polynesia. The level of detail offered by the analysis is impressive, and because the method relies on archaeologically or historically established calibration points, the nodes in the tree—that is, the splitting points resulting from human dispersals—can be dated to within a few centuries.

Support for this picture comes from Moodley et al.’s genetic analysis of samples for the bacterial parasite Helicobacter pylori, taken from the genetic tracts of Pacific human populations. The data also strongly favor a Taiwanese origin, producing a tree (see the second figure, bottom panel) that is similar in many ways to the linguistic tree of Gray et al. The analysis relies on the observation that, although most human populations share a gastric flora of H. pylori, at a molecular genetic level these bacteria differ from continent to continent. These differences are likely to be the product of genetic drift following the splitting and separation of populations. These processes enable reconstruction of a phylogenetic tree similar to that derived from human mitochondrial or Y-chromosome DNA (14). The dates in the bacterial analysis have large error margins and are again derived from archaeologically dependent calibration points. So the fact that both papers date the dispersal from Taiwan to ~5000 years ago is not so much a corroboration but a result of using the same archaeological data. But the topology and detail of the two trees are genuinely independent.

It will be interesting to see how well the topologies of the two trees correlate at a more detailed level, as clearly they do in general structure. The use of modern genetic data to reconstruct phylogenetic trees shows that the past is still “within us” (15) today. Our past is within us in a different sense when the vocabularies of specific modern languages are the basis for historical analysis. And the past is within us in a very literal way when the early history of humankind is reconstructed based on the bacterial flora in our guts. The convergence between the approaches suggests that a synthesis between linguistic and genetic interpretations of human history may soon be possible on a worldwide basis.

References