

**Comment on:
Bing Su et al. "Polynesian origins: Insights from the Y
chromosome," *PNAS* 97.15:8225-8228 (July 18, 2000)**

Lawrence A. Reid
University of Hawai'i

The appearance of another article by geneticists claiming to throw light on the homeland of a linguistic family,¹ this time on the basis of distribution patterns of haplotype frequencies identified in Y-chromosome data, points up once again the problems of attempting to use data from one discipline to support or challenge conclusions based on data from another discipline. These problems are compounded, at least in this case, by a serious misunderstanding of the claims made by linguists and of the significance of the methodology used by linguists to establish linguistic subgroups.

It is necessary to make a distinction between "the settlement of Polynesia," and "the origin of Polynesian," that is between claims that can be made about who the people were that have, over the last two millennia, moved into the island area known as Polynesia, and claims about the origin of the languages that have been called Polynesian, speakers of which live not only in the "Polynesian triangle" but in at least a dozen scattered enclaves in Melanesia, the so-called Polynesian Outliers (Pawley 1967). Answers to the first question may come from genetic studies of present-day populations, including Y-chromosome and mitochondria DNA studies, as well as from insights from physical anthropologists and archaeologists, among others. Answers to the second question can only come from the application of the well-established, comparative-historical methodology for the reconstruction of proto-languages, and a careful analysis of the patterns of linguistic innovation in phonology, morphology and syntax in order to establish subgrouping. To the extent that answers to the first question coincide with answers to the second, linguistic subgrouping decisions are confirmed, but where they do not coincide, linguistic subgroupings are not thereby invalidated. While it is obvious that language is spread by the movement of its speakers, especially in initial colonization, the reverse is not true, that is, evidence of movement of people does not necessarily imply that their language left any mark on the language that may already have been spoken by other populations at their destination.

¹ See, for example, J. Y. Chu et al. 1998.

That Bing Su et al. (henceforth B) are unaware of these distinctions is apparent throughout their article. Even the title of their article, "Polynesian origins", implies the possibility of multiple origins of Polynesian, regardless of the fact that no linguist would claim that Polynesian was anything other than an Austronesian language, having its ultimate source in Proto-Austronesian. The authors then claim that there are two competing hypotheses regarding the settlement of Polynesia. The first, they say, "claims that about 4,000 to 5,000 years B.P. a rapid eastward migration of humans began in Southern China spreading Austronesian language and the associated Lapita culture throughout the Pacific islands and culminating in the settlement of Polynesia," while the second "asserts a neighboring homeland of the Polynesians in Melanesia, in which the Polynesians evolved... among the already settled Pacific Islanders."

The first of these hypotheses only partly resembles that which most linguists and archaeologists would claim. First, the dates given are far too late. The earliest dates of neolithic settlement in Taiwan currently date to about 6,300 B.P. Secondly, the "rapid eastward migration" could hardly be said to start in mainland China, when it is clear from the archaeological record that the migrants to Taiwan were stuck there for almost a thousand years before moving south into the Northern Philippines [spelled Phillipines by B in their map]. Thirdly, the term 'Austronesian' labels the family whose origins most linguists believe developed during those thousand years in Taiwan. Regardless of the fact that archaeological evidence links some early western Formosan sites with sites in Central and Southeast China and the Pescadores Islands (Chang and Goodenough 1996, Bellwood 1997:205), there is no linguistic evidence that the homeland of Proto-Austronesian was anywhere other than in Taiwan. Neither is there evidence that the so-called Lapita cultural complex was located in "Southern China". It has been claimed that the precursors of Lapita can be found in island Southeast Asia, but not in Southern China.

The second of the Polynesian settlement hypotheses only partially resembles the claims made by Terrell (1988). Terrell is less interested in the origins of the Polynesians than in the causes of the migration that took them from the Melanesian area eastward to Polynesia. From the point of view of the linguist, Melanesia was a way-station for Austronesian-speaking peoples moving between island Southeast Asia and remote Oceania. They were not however speaking a Polynesian language, nor did Polynesian evolve in the area already populated by speakers of non-Austronesian languages. Polynesian languages originated in the Proto-Polynesian homeland, which has been shown by Geraghty (1983) and most recently by Kikusawa (in press) to have been in the area of Tonga-Niue, after they split off from the Proto-Central Pacific dialect chain spoken in Fiji. The only Polynesian languages spoken in Melanesia today are clearly the

result of back-migration from Polynesia, subsequent to the disintegration of Proto-Polynesian.

It is not surprising that B was unable to arrive at clear results from the Y chromosome studies, although valiant attempts to interpret the data were made. Their conclusions, to be discussed below, were drawn on the basis of a set of data drawn from a very limited set of populations, and in many cases from only a few individuals in each population. Of the fourteen extant languages in Taiwan, constituting according to Blust (1999) nine proposed primary linguistic subgroups, only four (Atayal, Paiwan, Amis, and Bunun) were sampled. Yami (spoken on Botel Tobago, or Orchid Island, east of the southernmost tip of Taiwan) has never been classified as a Formosan language. Its closest relatives are spoken in the Batanes Islands, north of Luzon in the Philippines. There are no samples from any other of the more than one hundred Philippine ethnolinguistic groups. Of the Taiwan groups, three have fewer than ten samples each, one has eleven, and the other (Atayalic, with the greatest number of haplotypes represented), has only twenty-four. One wonders how many different families, were represented in those groups with very limited samples, and what the results might have been if a much larger set of samples from widely separated, and clearly unrelated families had been available.

The Southeast Asia group includes a mixed bag of individuals, representing Hmong-Mien, Tai-Kadai, Austroasiatic, and Austronesian (Malay, Batak, Javanese and Kota Kinabalu in northeastern Borneo--probably speakers of Central Dusun) languages. It is not clear which of the Aslian subgroups are represented by the "Orang Asli" sample. Of the various groups referred to in Malay by this term, one, the Semang are Negritos and speak an Austroasiatic language. Some of the other groups are related to Malay, and are therefore Austronesian (Parkin 1991).

In the Melanesian group, there are representatives of three Austronesian groups (Banks [spelled Bankes by B] and Torres, Maewo, and Santo), one clearly non-Austronesian Papuan group (Nasioi), and ninety individuals classed as "New Guinea". Are these representatives of Papuan or Austronesian populations in New Guinea? Of the forty Polynesian individuals sampled, only one is Tongan, twenty-nine are Samoan (one of whom carried the European specific haplotype H14) and the other ten are from Kapingamarangi, a remote Polynesian Outlier, spoken some 500 miles to the Southwest of Pohnpei in Micronesia, and showing precisely the same distribution of haplotypes (30% H1, 70% H5) as the ten individuals sampled from that island group, an interesting correlation of the contact that has obviously occurred between these groups in the past.

Their Micronesian sample, includes not only speakers of what are considered to be Micronesian languages (Truk, Majuro, Kiribati, and Pohnpei [spelled Phonpei by B]),

but also two clearly Western Austronesian languages (Guam, presumably Chamorro speakers, and Palauan).

B claims that because “Taiwanese aboriginal populations on the one hand and Micronesian and Polynesian on the other carry two different subsets of haplotypes found in the extant Southeast Asian populations...the data do not favor a Polynesian homeland in Taiwan.” They note that the Taiwanese aboriginal populations that were sampled (including the Philippine language, Yami) do not carry haplotypes H1 through H5, but a glance at the Table 1 which shows the frequency distribution of Y-chromosome haplotypes in their sampled populations, shows that the same five types are also missing from the eleven speakers of Li, a Kadai language on mainland Southeast Asia. They are also missing from eleven Kiribati speakers, as well as from the one Tongan speaker. In addition, the Paiwan, Amis, Bunun, and Yami speakers sampled are also like the Li sample in lacking H6 and H7.

If the Y-chromosome data is to be of any value in supporting what is undeniably a well-established language family, it would be useful to be able to identify any haplotype (not necessarily a pattern of different haplotype distributions), which seems to be distributed widely across all branches of the family, and which is generally missing in populations that do not constitute recognized parts of that family. Only thus is there any way of correlating the original population movement with a Y-chromosome, and of identifying the two possibly confounding factors, that is, the effects of contact with speakers of languages that did not carry that chromosome, and the possibility of the sporadic loss of that chromosome from the gene pool of native speakers of one or more of the languages.

Very few of the groups in any of the geographical areas are completely homogeneous. Among the Taiwanese populations, the Atayalic sample shows the greatest diversity (29.2% H6, 4.2% H7, 4.2% H8, 54.2 H9, 8.3% H10). While Amis (100% H9) and Yami (25% H9, 75% H11) show the least. Bunun (11.1% H9, 66.7% H10, 22.2% H12) and Paiwan (18.2% H8, 54.5% H9, 27.3% H10) with three haplotypes each do not completely overlap, sharing only H9 and H10. But H8 is also found in Truk (5.9%) as well as in Kiribati (9.1%), while H10 is also found in Majuro (presumably Marshallese speakers, 22.2%). 27.3% of the Majuro sample carried H12, also found in Bunun in Taiwan.

The distribution of the H6 haplotype, however, is of most interest. It is found in all of the mainland and island Southeast Asian populations (except for Li, a Tai-Kadai language, and So, a Mon-Khmer language). In Taiwan, as noted above, it is found in 29.2% of the Atayal sample. It is found in Melanesia in the Banks and Torres, and Maewo samples. In Micronesia, it is found in the samples from both Guam and Palau, as well as in Truk, and in the majority (63.6%) of the Kiribati sample. It is also found in

71.4% of the Nauru sample. In Polynesia, it is carried by the only Tongan speaker sampled, as well as by 41.4% of the twenty-nine Samoans. The H6 haplotype is missing from the two mainland populations mentioned above, the Taiwanese populations other than Atayal, and the Marshallese, Pohnpeian and Kapingamarangi samples in Micronesia. And it is missing from the three East Papuan Nasioi Melanesian samples, all of the New Guinea sample, and from the four Santo samples. What is one to make of this distribution?

It would seem that H6 alone among the various haplotypes identified is distributed from one end of the Austronesian family to the other, and in each of the subgroups sampled. The haplotype has apparently been sporadically lost in a few of the mainland populations whose languages are probably ultimately genetically related to Austronesian, as it has in the Formosan populations, a scenario allowed by B, "our data could be viewed as a spread of Y-chromosome haplotypes out of China to Taiwan and then to Micronesia/Polynesia; and in this model, the haplotypes became distinguished on account of random extinction of Y-chromosome lineages in different geographic regions."

The effects of contact with speakers of languages that did not carry the H6 chromosome has also been noted. B identifies H17 as "Melanesian specific". It is found in all of the Nasioi samples and in 38.9% of the New Guinea sample. Although it is found in two (11.8%) of the Truk sample, as noted by B, two of the Melanesian populations in which it is found are not speakers of Papuan languages but of Austronesian languages, Maewo (20%) and Banks and Torres (16.7%), who have clearly had long and intensive contact with speakers of Papuan languages. But, interestingly, H17 does not appear at all in the Santo (Austronesian-speaking) sample of only four speakers, which is 100% H5.

The wide distribution of H5 across the family, also implies that it was probably carried by the early Austronesian migrants as they moved through to Polynesia. Its absence from the Taiwan sample could again be the result of random extinction as noted by B, but I would think is much more likely to be the result of the very limited number of both language groups and individuals that were sampled. But whatever the reason, it is impossible to conclude that the data presented give any evidence that Polynesian somehow developed from some group that didn't begin their migration south and east from their island home we now call Taiwan.

References

- Bellwood, Peter. 1997. *The Prehistory of the Indo-Malaysian Archipelago*. (2nd edition). Honolulu: University of Hawai'i Press.
- Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. *Selected Papers for the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jenkuei Li, 31-94. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Chang, Kwang-chih, and Ward H. Goodenough. 1996. Archaeology of southeastern coastal China and its bearing on the Austronesian homeland. *Prehistoric Settlement of the Pacific*, ed. by Ward H. Goodenough, 36-56. Transactions of the American Philosophical Society 86.4. Philadelphia: American Philosophical Society.
- Chu, J. Y. et al. 1998. Genetic relationship of populations in China. *Proceedings of the National Academy of Sciences* 95:11763-11768.
- Geraghty, Paul. 1983. *The History of the Fijian Languages*. Oceanic Linguistics Special Publication 19. Honolulu: University of Hawai'i Press.
- Kikusawa, Ritsuko. (in press). *Reconstructing the Actancy Systems of Proto Central Pacific and Its Daughter Languages: Ergativity, Accusativity and Their Historical Development*. Canberra: Pacific Linguistics.
- Parkin, Robert. 1991. *A Guide to Austroasiatic Speakers and Their Languages*. Oceanic Linguistics Special Publication 23. Honolulu: University of Hawai'i Press.
- Pawley, Andrew. 1967. The relationships of Polynesian outlier languages. *Journal of the Polynesian Society* 76.3:259-296.
- Terrell, John. 1988. History as a family tree, history as an entangled bank: Constructing images and interpretations of prehistory in the South Pacific. *Antiquity* 62:642-657.

Linguistics Department
University of Hawai'i at Manoa
1890 East-West Road, Moore Hall 569
Honolulu, HI 96822, U.S.A.
reid@hawaii.edu