Squib

Language and Genes Attest Different Histories in Island Southeast Asia

Mark Donohue and Tim Denham

AUSTRALIAN NATIONAL UNIVERSITY AND MONASH UNIVERSITY

Recent work on autosomal DNA genetic variation across Southeast Asia suggests that genetic diversity largely reflects Pleistocene colonization by modern humans, and was not influenced to any significant degree by major cultural and linguistic changes during the mid-to-late Holocene (roughly, from ~5,000 years ago to the present). These results seemingly show that the spread of Austronesian languages across Island Southeast Asia was not associated with population movements that were significant enough to affect the overall phylogeny of the autosomal DNA tree. Consequently, the spread of genes is not significantly linked to the spread of languages in Island Southeast Asia; each represents different processes of different antiquity.

Much work on “the Austronesians” has, either explicitly or implicitly, taken the view that the history reflected in the linguistic data largely corresponds to the histories revealed by the archaeological record and by the human genetic record, respectively (Bellwood 1995, 1997; Diamond 2000, 2001; see Donohue and Denham 2010 for review). Given that the linguistic record clearly indicates an out-of-Taiwan history, the archaeological and biological records are similarly assumed to favor an out-of-Taiwan model.¹

Recent work by the HUGO Pan-Asian single-nucleotide polymorphism (SNP) consortium (2009) reveals a fascinating human genetic history for large parts of mainland and Island Southeast Asia, including modern Taiwan, the Philippines, Indonesia, and Malaysia (see figure 1). Autosomal DNA variation indicates primary directions of human movement from west to east and from south to north, movements that reflect the original colonization of Southeast Asia by modern humans during the Late Pleistocene (see, for example, Macaulay et al. 2005).² The genetics of human populations have undergone “substantial recent admixture” at a local level (HUGO 2009), but the overall pattern has seemingly remained largely undisrupted, despite major cultural and linguistic changes across Southeast Asia during the Holocene (cf., for example, Bellwood 1997). The mapping of other genetic traits for populations in Island Southeast Asia has produced variable results in terms of understanding the history of population movements during the

¹. We thank the anonymous referees whose comments improved the clarity of this note.
². Note that we make no comment about the settlement of Remote Oceania, which was (to judge from the archaeological data) a very different settlement pattern for Austronesians than was the settlement of Near Oceania and Island Southeast Asia.

Oceanic Linguistics, Volume 50, no. 2 (December 2011) © by University of Hawai‘i Press. All rights reserved.
Holocene (Oppenheimer and Richards 2001; Hill et al. 2007; Kayser et al. 2008; Soares et al. 2008, 2011), and also in terms of correspondences between genes and languages (Lansing et al. 2007; Donohue and Denham 2010).

Significantly, the directionality of language spread inferred from historical linguistics is almost consistently opposite to that inferred from the human genetic phylogeography (after Avise 2000) for Southeast Asia. Figure 2 shows a geographic reconstruction of the HUGO phylogeny (black), and the inferred spread of three major language families: Austronesian (red, Blust 2009); Tai-Kadai (Diller, Edmonson, and Luo 2005) and Sino-Tibetan (Matisoff 2003), both blue; and Austro-Asiatic (yellow, Sidwell 2009). The genetic composition of human populations in Southeast Asia is largely unchanged since the Pleistocene despite the large-scale—and in Island Southeast Asia nearly whole-sale—replacement of previous languages during the mid-late Holocene. In sum, the spread of contemporary language families was not associated with a significant transformation in the genetic composition of human populations across Island Southeast Asia, as has often been claimed (for example, Bellwood 1995; Diamond 2000, 2001). The generally contiguous distribution of linguistic subgroups in space correlates with the prehistoric population divergence of genetic markers (HUGO 2009:1544), and does not match the subgrouping structures predicted by the known language families in the region (for example, Blust 2009 on Austronesian and Sidwell 2009 on Austro-Asiatic).

In the HUGO results (summarized in figure 1 from HUGO 2009:1542), speakers of Austronesian languages on Taiwan are, in terms of human genetics, subgroups within a larger clade that contains some (but not all) speakers of other Austronesian languages. Significantly the Austronesian language speakers closest to the top of the “Austronesian” clade in figure 1 are in Malaysia, where they are close (in human genetic terms) to the neighboring “Austro-Asiatic” speaking samples. Elsewhere, we see Tai-Kadai not forming a distinct clade separate from other northern mainland groups (Chinese, Hmong, Japanese, Korean), and linguistically Dravidian genetic samples subgrouping with those from (Indo-European) Indic populations, but not with Indo-European populations from outside South Asia.

The HUGO results are noteworthy because, although they are based on specific scans, they were undertaken on large and robust sample populations. They clearly demonstrate a lack of historical consilience between human genetic and linguistic phylogeographies, a discrepancy that confirms previous studies of mitochondrial DNA (mtDNA) and Y-chromosome data (for example, Capelli et al. 2001; Hurles et al. 2005; Kayser et al. 2008; Li et al. 2008; Soares et al. 2011; and many others) whenever a large sample of Asian populations is included. It would be wrong to suppose that there is no genetic signature from Taiwan, but the proportion does not represent the wholesale replacement or even absorption of preexisting languages.

---

3. The black line was constructed as follows: for each subgroup in the data—drawing on figure 2 of HUGO 2009:1543), and the references listed above for the language families—we determined a mean location based on the averaged centroid locations for each member in that subgroup. A line of best fit was drawn through members of the same hierarchical order, with directionality running from higher (intermediary) nodes down to the tips (following Avise 2000 and others). Not all low-level lines are shown for the biological phylogeography or for the linguistic phylogeography.

4. We note that Li et al.’s Y-chromosome study (2008) similarly fails to position Taiwanese aboriginal populations at the top of any genetic tree that includes extra-Formosan Austronesian speakers.
populations (Oppenheimer and Richards 2001). Maximally 20 percent of Y chromosome variation (Capelli et al. 2001) and at most 20 percent of the mtDNA variation (Hill et al. 2007) in Island Southeast Asian populations can be explained by an out-of-Taiwan hypothesis. By implication, these findings are relevant beyond just Southeast Asia, and call into question the assumptions about correspondence between human genetic and language distributions elsewhere in the world (for example, Diamond and Bellwood 2003; Cavalli-Sforza et al. 1988; Sokal et al. 1990; Barbujani and Sokal 1990; Mona et al. 2007). These findings imply in the main a model that is dependent on social geographic factors rather than linguistic affiliations (Hill and Serjeantson 1989; Su et al. 1999; Rosenberg et al. 2006; Gayden et al. 2007; Wang et al. 2007; López Herráez et al. 2009; Mona et al. 2009; Simonson et al. 2011). Nichols and Peterson (1998:612) note that “biological genetic features have geographically based rather than linguistically based distributions,” and we add that, to judge from the studies quoted above, there is a correlation with linguistic classifications only in those circumstances in which the phylogeny of the linguistic family spreads out across contiguous regions in space, thus making geography the common factor that links genetics and linguistics.

FIGURE 1. OVERALL TOPOLOGY OF THE SOUTHEAST ASIA GENETIC TREE PRESENTED IN THE HUGO PAPER (MANY BRANCHES OMITTED) *

* Key: AA, Austro-Asiatic; An, Austronesian; IE, Indo-European; TB, Tibeto-Burman. Born., Borneo; East Ind., Eastern Indonesia; Mal., Malaysian peninsula; Phil., Philippines; Sul., Sulawesi; Sum., Sumatra.
The HUGO authors suggest several reasons to account for the discrepancies between genetic clades and linguistic affiliation (HUGO 2009:1543): “These patterns are consistent either with substantial recent admixture among the populations . . . , a history of language replacement . . . , or uncertainties in the linguistic classifications themselves.” Even taking these potential factors into account, the lack of consilience is still unaccounted for and undermines presumptions of any significant correspondence between biological genetics and linguistics in Island Southeast Asia during the Holocene. How is it possible to have near-complete language replacement over wide regions with only minor changes in the genetic composition of resident populations? How is it that more and more Austronesian-speaking populations are found to have genetic affinities that do not link closely to Taiwan (for example, Simonson et al. 2011; Soares et al. 2011)? Such questions are important not only in Island Southeast Asia, where the initial dispersal of Austronesian languages outside Taiwan occurred, but also in understanding the history of Austronesian languages in Melanesia and, ultimately, the “origins” of the Polynesians (see Addison and Matisoo-Smith 2010). Recent work on Polynesian genetic origins indicates that the “Polynesian motif” gene mutation can be dated to island Melanesia at a time depth that precludes its introduction by any human gene flow in the last 6,000 years, indicating that those biological elements that make Polynesians distinctive cannot be associated with their distinct linguistic history (see also Soares et al. 2011). These lines of enquiry force us to refocus our attention upon a greater engagement with the historical and social processes that led to the formation of present-day genetic and linguistic distributions in “Austronesia” (necessarily including the non-Austronesian languages of the relevant regions: Donohue and Denham 2010) and beyond. Whatever the underlying social events (and it is certain that there is more than a single contact story underlying the dispersal of a language family as large and geographically distributed as Austronesian), it is clear that demic diffusion did not play a major role in the Austronesianization of Island Southeast Asia and Near Oceania.

REFERENCES


Li, Hui, Bo Wen, Shu-Juo Chen, Bing Su, Patcharin Pramoonjago, Yangfan Liu, Shangling Pan, Zhendong Qin, Wenhong Liu, Xu Cheng, Ningning Yang, Xin Li, Dinhbinh Tran, Daru Lu, Mu-Tsu Hsu, Ranjan Deka, Sangkot Marzuki, Chia-Chen Tan, and Li Jin. 2008. Paternal genetic affinity between western Austronesians and Daic populations. BMC Evolutionary Biology 8:146.

López Herráez, David, Marc Bauchet, Kun Tang, Christoph Theunert, Irina Pugach, Jing Li, Madhusudan R. Nandineni, Arnd Gross, Markus Scholz, and Mark


Su, Bing, Junhua Xiao, Peter Underhill, Ranjan Deka, Weiling Zhang, Joshua Akey, Wei Huang, Di Shen, Daru Lu, Jingchun Luo, Jiayou Chu, Jiazhen Tan, Peidong


mark@donohue.cc  
tim.denham@arts.monash.edu.au
FIGURE 2. SCHEMATIC MAP SHOWING PRIMARY DIRECTIONS OF GENETIC SPREAD AND LANGUAGE SPREADS ON A CURRENT MAP OF SOUTHEAST ASIA*

* Genetic spread is marked by the black arrows. Language spreads are marked by colored arrows: red for Austronesian, blue for Tai-Kadai and Sino-Tibetan, and yellow for Austro-Asiatic.