

Squib

Languages 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. 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 Holocene

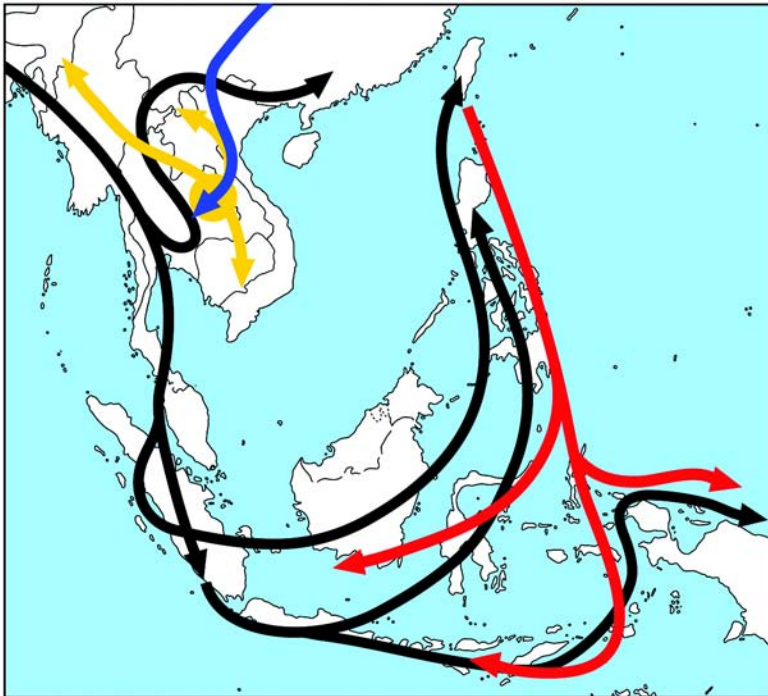
1. We thank the anonymous referees whose comments improved the clarity of this squib.

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.

(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 1 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 Austroasiatic (yellow, Sidwell 2009).³ The genetic composition of human populations in Southeast Asia is largely unchanged sin-

FIGURE 1. SCHEMATIC MAP SHOWING PRIMARY DIRECTIONS OF GENETIC SPREAD AND LANGUAGE SPREADS ON A CURRENT MAP OF SOUTHEAST ASIA*



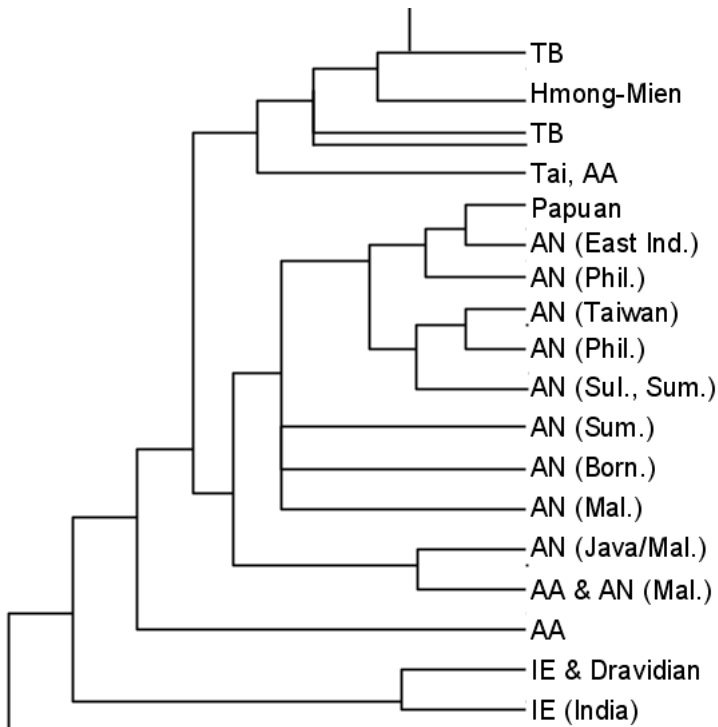
* Genetic spread is marked by the black arrows. Language spreads are marked by colored arrows: red for Austronesian, blue for Tai-Kadai and Sino-Tibetan, and yellow for Austroasiatic.

3. The black line was constructed as follows: for each subgroup in the data—drawing on figure 1 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.

cethe Pleistocene despite the large-scale—and in Island Southeast Asia nearly wholesale—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 Austroasiatic).

In the HUGO results (summarized in figure 2 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.⁴ Significantly the Austronesian language speakers closest to the top of the “Austronesian” clade in figure 2 are in Malaysia, where they are close (in human genetic terms) to the

FIGURE 2. OVERALL TOPOLOGY OF THE SOUTHEAST ASIA GENETIC TREE PRESENTED IN THE HUGO PAPER (MANY BRANCHES OMITTED)*



* Key: AA, Austroasiatic; An, Austronesian; IE, Indo-European; TB, Tibeto-Burman. Born., Borneo; East Ind., Eastern Indonesia; Mal., Malaysian peninsula; Phil., Philippines; Sul., Sulawesi; Sum., Sumatra.

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.

neighboring “Austroasiatic” 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; Hurler 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 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.

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

- Addison, David, and Elizabeth Matisoo-Smith. 2010. Rethinking Polynesian origins: A West Polynesia Triple-I model. *Archaeology in Oceania* 45:1–12.
- Avise, Paul. 2000. *Phylogeography: The history and formation of species*. Cambridge MA: Harvard University Press.
- Barbujani, Guido, and Robert R. Sokal. 1990. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Science of the USA* 87:1816–19.
- Bellwood, Peter. 1995. Austronesian prehistory in Southeast Asia: Homeland, expansion and transformation. In *The Austronesians: Historical and comparative perspectives*, ed. by Peter Bellwood, Jim Fox, and Darrell Tryon, 103–18. Canberra: Australian National University.
- . 1997. *Prehistory of the Indo-Malaysian archipelago*. Honolulu: University of Hawai‘i Press.
- Blust, Robert. 2009. *The Austronesian languages*. Canberra: Pacific Linguistics.
- Capelli, Cristian, James F. Wilson, Martin Richards, Michael P. H. Stumpf, Fiona Gratrix, Stephen Oppenheimer, Peter Underhill, Vincenzo L. Pascali, Tsang-Ming Ko, and David B. Goldstein. 2001. A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *American Journal of Human Genetics* 68(2):432–43.
- Cavalli-Sforza, Luigi Luca, Alberto Piazza, Paolo Menozzi, and Joanna Mountain. 1988. Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proceedings of the National Academy of Science of the USA* 85:6002–6.
- Diamond, Jared. 2000. Taiwan’s gift to the world. *Nature* 403:709–10.
- . 2001. Slow boat to Melanesia: Reply. *Nature* 410:167.
- Diamond, Jared, and Peter Bellwood. 2003. Farmers and their languages: The first expansions. *Science* 300:597–603.
- Diller, Anthony, Jerrold Edmonson, and Yongxian Luo, eds. 2005. *The Tai-Kadai languages*. London: Routledge.
- Donohue, Mark, and Tim Denham. 2010. Farming and language in Island Southeast Asia: Reframing Austronesian history. *Current Anthropology* 51:223–56.
- Gayden, Tenzin, Alicia M. Cadenas, Maria Regueiro, Nanda B. Singh, Lev A. Zhivotovsky, Peter A. Underhill, Luigi L. Cavalli-Sforza, and Rene J. Herrera. 2007. The Himalayas

- as a directional barrier to gene flow. *The American Journal of Human Genetics* 80:884–94.
- Hill, Adrian V. S., and Susan W. Serjeantson, eds. 1989. *The colonization of the Pacific: A genetic trail*. Oxford: Clarendon.
- Hill, Catherine, Pedro Soares, Maru Mormina, Vincent Macaulay, D. Clarke, Petya B. Blumbach, Matthieu Vizuete-Forster, Peter Forster, David Bulbeck, Stephen Oppenheimer, and Martin Richards. 2007. A mitochondrial stratigraphy for Island Southeast Asia. *American Journal of Human Genetics* 80(1):29–43.
- HUGO Pan-Asian SNP Consortium. 2009. Mapping human genetic diversity in Asia. *Science* 326:1541–45.
- Hurles, Matthew E., Bryan C. Sykes, Mark A. Jobling, and Peter Forster. 2005. The dual origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from maternal and paternal lineages. *American Journal of Human Genetics* 76(5):894–901.
- Kayser, Manfred, Silke Brauer, Gunter Weiss, Wulf Schiefenhövel, Peter A. Underhill, and Mark Stoneking. 2001. Independent histories of human Y chromosomes from Melanesia and Australia. *American Journal of Human Genetics* 68(1):173–90.
- Kayser, Manfred, Ying Choi, Mannis van Oven, Stefano Mona, Silke Brauer, Ronald J. Trent, Dagwin Suarkia, Wulf Schiefenhövel, and Mark Stoneking. 2008. The impact of the Austronesian expansion: Evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Molecular Biology and Evolution* 25:1362–74.
- Lansing, J. Stephen, Murray P. Cox, Sean S. Downey, Brandon M. Gabler, Brian Hallmark, Tatiana M. Karafet, Peter Norquest, John W. Schoenfelder, Herawati Sudoyo, Joseph C. Watkins, and Michael F. Hammer. 2007. Coevolution of languages and genes on the island of Sumba, eastern Indonesia. *Proceedings of the National Academy of Science of the USA* 104:16022–26.
- 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 Stoneking. 2009. Genetic variation and recent positive selection in worldwide human populations: Evidence from nearly 1 million SNPs. *PLoS-ONE* 4(11):e7888. doi:10.1371/journal.pone.0007888.
- Macaulay, Vincent, Catherine Hill, Alessandro Achilli, Chiara Rengo, Douglas Clark, William Meehan, James Blackburn, Ornella Semino, Rosaria Scozzari, Fulvio Cruciani, Adi Taha, Norazila Kassim Shaari, Joseph Maripa Raja, Patimah Ismail, Zafarina Zainuddin, William Goodwin, David Bulbeck, Hans-Jürgen Bandelt, Stephen Oppenheimer, Antonio Torroni, and Martin Richards. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034–36.
- Matisoff, James. 2003. *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction*. Berkeley: University of California Press.
- Mona, Stefano, Katharina E. Grunz, Silke Brauer, Brigitte Pakendorf, Loredana Castri, Herawati Sudoyo, Sangkot Marzuki, Robert H. Barnes, Jörg Schmidtke, Mark Stoneking, and Manfred Kayser. 2009. Genetic admixture history of eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Molecular Biology and Evolution* 26(8):1865–77.

- Mona, Stefano, Mila Tommaseo-Ponzetta, Silke Brauer, Herawati Sudoyo, Sangkot Marzuki, and Manfred Kayser. 2007. Patterns of Y-chromosome diversity intersect with the Trans-New Guinea hypothesis. *Molecular Biology and Evolution* 24(11):2546–55.
- Nichols, Johanna, and David A. Peterson. 1998. Amerind personal pronouns: A reply to Campbell. *Language* 74:605–14.
- Oppenheimer, Stephen, and Martin Richards. 2001. Polynesian origins: Slow boat to Melanesia? *Nature* 410:166–67.
- Rosenberg, Noah, Saurabh Mahajan, Catalina Gonzalez-Quevedo, Michael G. B. Blum, Laura Nino-Rosales, Vasiliki Nini, Parimal Das, Madhuri Hegde, Laura Molinari, Gladys Zapata, James L. Weber, John W. Belmont, and Pragna I. Patel. 2006. Low levels of genetic divergence across geographically and linguistically diverse populations from India. *PLoS-Genetics* 2 (12):e215. doi:10.1371/journal.pgen.0020215.
- Sidwell, Paul. 2009. *Classifying the Austroasiatic languages: History and state of the art*. Munich: Lincom Europa.
- Simonson, Tatum S., Jinchuan Xing, Robert Barrett, Edward Jerah, Peter Loa, Yuhua Zhang, W. Scott Watkins, David Witherspoon, Chad D. Huff, Scott Woodward, Bryan Mowry, and Lynn B. Jorde. 2011. Ancestry of the Iban is predominantly Southeast Asian: Genetic evidence from autosomal, mitochondrial, and Y chromosomes. *PLoS-ONE* 6 (1):e16338. doi:10.1371/journal.pone.0016338.
- Soares, Pedro, Teresa Rito, Jean Trejaut, Maru Mormina, Catherine Hill, Emma Tinkler-Hundai, Michelle Braid, Douglas J. Clarke, Jun-Hun Loo, Noel Thomson, Tim Denham, Mark Donohue, Vincent Macaulay, Marie Lin, Stephen Oppenheimer, and Martin Richards. 2011. Ancient voyaging and Polynesian origins. *American Journal of Human Genetics* 88:239–47.
- Soares, Pedro, Jean Alain Trejaut, Jun-Hun Loo, Catherine Hill, Maru Mormina, Chien-Liang Lee, Yao-Ming Chen, Georgi Hudjashov, Peter Forster, Vincent Macaulay, David Bulbeck, Stephen Oppenheimer, Marie Lin, and Martin Richards. 2008. Climate change and postglacial human dispersals in Southeast Asia. *Molecular Biology and Evolution* 25:1209–18.
- Sokal, Robert R., Neil L. Oden, Pierre Legendre, Marie-Josée Fortin, Junhyong Kim, Barbara A. Thomson, Alain Vaudor, Rosalind M. Harding, and Guido Barbaranji. 1990. Genetics and language in European populations. *American Naturalist* 135:157–75.
- 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 Shen, Ron Davis, Luca Cavalli-Sforza, Ranajit Chakraborty, Momiao Xiong, Ruofu Du, Peter Oefner, Zhu Chen, and Li Jin. 1999. Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last Ice Age. *American Journal of Human Genetics* 65(6):1718–23.
- Wang, Sijia, Cecil M. Lewis Jr., Mattias Jakobsson, Sohini Ramachandran, Nicolas Ray, Gabriel Bedoya, Winston Rojas, Maria V. Parra, Julio A. Molina, Carla Gallo, Guido Mazzotti, Giovanni Poletti, Kim Hill, Ana M. Hurtado, Damian Labuda, William Klitz, Ramiro Barrantes, Maria Cátira Bortolini, Francisco M. Salzano, Maria Luiza Petzl-Erler, Luiza T. Tsuneto, Elena Llop, Francisco Rothhammer, Laurent Excoffier, Marcus W. Feldman, Noah A. Rosenberg, and Andrés Ruiz-Linera. 2007. Genetic variation and population structure in Native Americans. *PLoS-Genetics* 3(11):e185. doi:10.1371/journal.pgen.0030185.