

## Reading social structure from the genome: gene flow and word flow in Eastern Indonesia

J. Stephen Lansing, Michael F. Hammer, Tatiana Karafet, Murray Cox, Sean Downey, Peter Norquest, Brian Hallmark, Joseph Watkins, Herawati Sudoyo, Sangkot Marzuki

The association between genetic and linguistic patterns is conventionally analyzed with comprehensive statistical measures such as the Mantel test. Here we present several new mathematical tools designed to disentangle the effects of horizontal (borrowing) and vertical (drift) transmission of genetic markers (microsatellites) and linguistic features. We apply these tools to the analysis of patterns of genetic and linguistic diversity on the Indonesian islands of Flores, Sumba, Lembata, Bali, Java, Mentawai and Nias.

With regard to word flow, our procedure involves coding features from 62 200-word samples of Austronesian languages and dialects spoken in Flores, Sumba, and Lembata; as well as comparable samples from Bali, Java, Mentawai and Nias. We sort the resulting feature lists by stripping apparent loanwords, in order to analyze rates of drift in the residual lists. Analysis of gene flow begins with genetic samples of 20-60 male villagers from ~45 villages located on the same islands. Samples include NRY, mtDNA and autosomal markers. A new model of microsatellite evolution by Watkins (in press) is used to calculate TMRCA within and between these village communities. Standard techniques of phylogenetic analysis are used to analyze migration. In addition, sequence cluster analysis (Plotkin et al 2002) is used to characterize patterns of relatedness within communities, and to explore their robustness. Because these patterns involve nonlinear relationships a forward-in-time simulation model, SAIL (“Simulated Agents in Love”) was developed as a tool to characterize typical sequence cluster patterns, and compare with backward-in-time coalescent simulations.

The overall goal of the project is to relate processes of genetic and linguistic change, beginning with an attempt to improve our understanding of rates of drift in both domains for typical social systems. The second goal is to investigate their interactive effects, the thresholds at which language change affects gene flow. In this way we hope to begin to characterize the processes which produce the cumulative patterns revealed by Mantel tests.

### References:

Plotkin, J.D., Dushoff S. and S.A. Levin, “Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus”, *PNAS* 99:9 (April 2002):6263-6268.

<http://www.ic.arizona.edu/~lansing/SAIL/index.htm>