Coevolution of languages and genes on the island of Sumba, eastern Indonesia

J. Stephen Lansing[†‡§], 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[¶]

[†]Department of Anthropology, University of Arizona, 1009 East South Campus Drive, Tucson, AZ 85721; [¶]Division of Biotechnology, Biosciences West, University of Arizona, Tucson, AZ 85721; [‡]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87521; [‖]Department of Mathematics, University of Arizona, 617 North Santa Rita Avenue, Tucson, AZ 85721; [††]Cotsen Institute of Archaeology, University of California, 308 Charles E. Young Drive North, Los Angeles, CA 90095; and [‡‡]Eijkman Institute for Molecular Biology, Diponegoro 69, Jakarta 10430, Indonesia

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved August 23, 2007 (received for review May 15, 2007)

Numerous studies indicate strong associations between languages and genes among human populations at the global scale, but all broader scale genetic and linguistic patterns must arise from processes originating at the community level. We examine linguistic and genetic variation in a contact zone on the eastern Indonesian island of Sumba, where Neolithic Austronesian farming communities settled and began interacting with aboriginal foraging societies ≈3,500 years ago. Phylogenetic reconstruction based on a 200-word Swadesh list sampled from 29 localities supports the hypothesis that Sumbanese languages derive from a single ancestral Austronesian language. However, the proportion of cognates (words with a common origin) traceable to Proto-Austronesian (PAn) varies among language subgroups distributed across the island. Interestingly, a positive correlation was found between the percentage of Y chromosome lineages that derive from Austronesian (as opposed to aboriginal) ancestors and the retention of PAn cognates. We also find a striking correlation between the percentage of PAn cognates and geographic distance from the site where many Sumbanese believe their ancestors arrived on the island. These language–gene–geography correlations, unprecedented at such a fine scale, imply that historical patterns of social interaction between expanding farmers and resident hunter-gatherers largely explain community-level language evolution on Sumba. We propose a model to explain linguistic and demographic coevolution at fine spatial and temporal scales.

Austronesian languages | cognate | contact zone | language evolution | Y chromosome haplogroups

Languages, like populations, change over time, but the rules governing language change are still not well understood. Because lexical and structural innovation, borrowing, and loss are difficult to observe and quantify over brief periods (1) and are impossible to witness over long periods, researchers are forced to undertake indirect approaches to infer the processes of language change. One such approach is to look for associations between linguistic and genetic classifications. Many well-known studies have identified associations between the languages and genes of human populations at continental and global geographic scales (2–6). A survey of these studies led Diamond and Bellwood (7) to hypothesize that many of these correlations are caused by the linked spread of prehistoric farmers and their languages outward from a number of widely dispersed agricultural homelands in Africa, the Near East/Europe, Asia, and the Americas. Under the simplest form of their hypothesis, genetic and linguistic variation evolves in parallel after the genes and languages of farmers replace those of hunter-gatherers in the path of expansion (Fig. 1*A*). According to Diamond and Bellwood (7), one of the best examples of the coevolution of language and genes was brought about by the Neolithic expansion of Austronesian-speaking farmers into previously uninhabited Polynesia and Micronesia.

Discrepancies between genetic and linguistic differentiation can arise through a number of processes (4, 8), perhaps the most important of which are genetic admixture (i.e., without language change) and language replacement (Fig. 1*A*) (7, 9, 10). These processes, which occur when migrating farmers meet resident hunter-gatherers face to face, likely characterize the expansion of Austronesian speakers into regions that were long occupied by indigenous populations in eastern Indonesia and New Guinea (7, 11, 12). However, most language–gene studies have sampled at a geographic scale, which is too coarse to permit any refined inference about the dynamics of language change in these contact zones. Information at a finer scale is essential to characterize the nature of contact relationships and infer mechanisms of linguistic and genetic transformation over recent temporal and fine spatial scales.

Toward this goal, we examine linguistic and genetic variation in a contact zone on the Indonesian island of Sumba. Broader regional studies support the initial settlement of Southeast Asia/Oceania by foraging societies by 40,000 to 45,000 BP (13). Languages of the geographically expansive Austronesian family occupy much of the Indonesian archipelago, except in far eastern Indonesia, where diverse and unrelated Papuan languages dominate. Recent syntheses place the Neolithic transition, considered to mark the arrival of Austronesian colonists in the vicinity of Sumba, at between 4,000 and 3,500 years ago (14, 15). At that time, small numbers of farmers speaking an Austronesian language likely came into contact with an indigenous population of foragers speaking aboriginal languages. Several circumstances favor Sumba as a site to investigate the relationship between population incursion and language change. Sumba is remote and culturally conservative, the last island in the archipelago where the majority adhered to a tribal or pagan religion at the close of the 20th century. Today, nearly all Sumbanese live in traditional farming villages composed of patrilocal clans. Contact between villages is limited, and population density is low. Perhaps the most telling indicator of the extent of contact between villages is the large number of languages now spoken on the island despite its small size (220×75 km^2). In this report, we reconstruct Sumbanese language relationships using a 200-word Swadish list, and we examine Y chromosome SNP and short tandem repeat (STR) diversity in a sample of Sumbanese villages. We propose a model of language–gene coevolution to explain the striking associations we observe among linguistic,

Author contributions: J.S.L. designed research; J.S.L., S.S.D., B.H., T.M.K., and H.S. performed research; S.S.D., B.M.G., B.H., T.M.K., P.N., J.W.S., and J.C.W. analyzed data; and J.S.L., M.P.C., B.H., P.N., J.C.W., and M.F.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: PAn, Proto-Austronesian; PS, Proto-Sumba; STR, short tandem repeat.

[§]To whom all correspondence should be addressed. E-mail: lansing@santafe.edu.

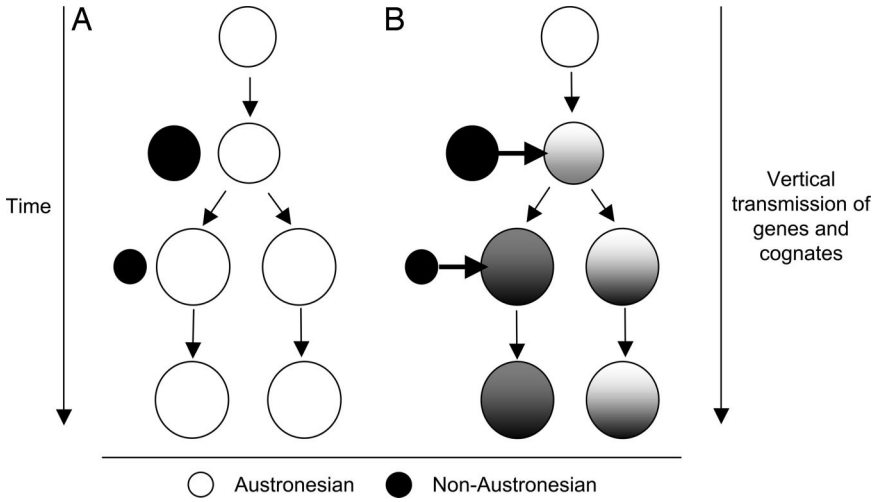

Fig. 1. Models for the evolution of languages and genes at two scales. Each circle represents a population of languages or genes, with parent populations shown on top and descendant populations shown below downward-facing arrows. Open circles, invading farming populations; filled circles, resident aboriginal populations. (*A*) Replacement models at larger geographic and temporal scales include two favored models in the literature: language replacement (i.e., the languages of an incoming population replace those of resident groups without gene flow) (10) and Diamond and Bellwood's (7) basic hypothesis (i.e., linguistic and genetic replacement by an incoming group with subsequent coevolution of descendant languages and genes). (*B*) An alternative model with codominant effects at smaller geographic and temporal scales involves both genetic admixture (e.g., demic diffusion) and the incursion of words that do not trace to PAn (horizontal arrows) in each descendant population after arrival of a founding Austronesian population (circle at center and top). A greater number of noncognates enters the population in the western part of Sumba where there are lower frequencies of Austronesian Y chromosome lineages (larger filled circles and thicker horizontal arrows) relative to the central part of Sumba.

genetic, and geographical data sets sampled at a fine geographic scale.

Results and Discussion

Linguistic Variation. We gathered twenty-nine 200-word Swadesh lists from diverse sites on the island (Fig. 2) and used traditional comparative linguistic approaches to identify cognates (i.e., words in two or more languages that can be traced to a common ancestor), sound correspondences, innovations, and loan words. On average, the Swadesh list for a Sumbanese language contains ≈70 cognates and 130 noncognates. In other words, ≈35% of the 200-word lexicon is directly descended from Proto-Austronesian

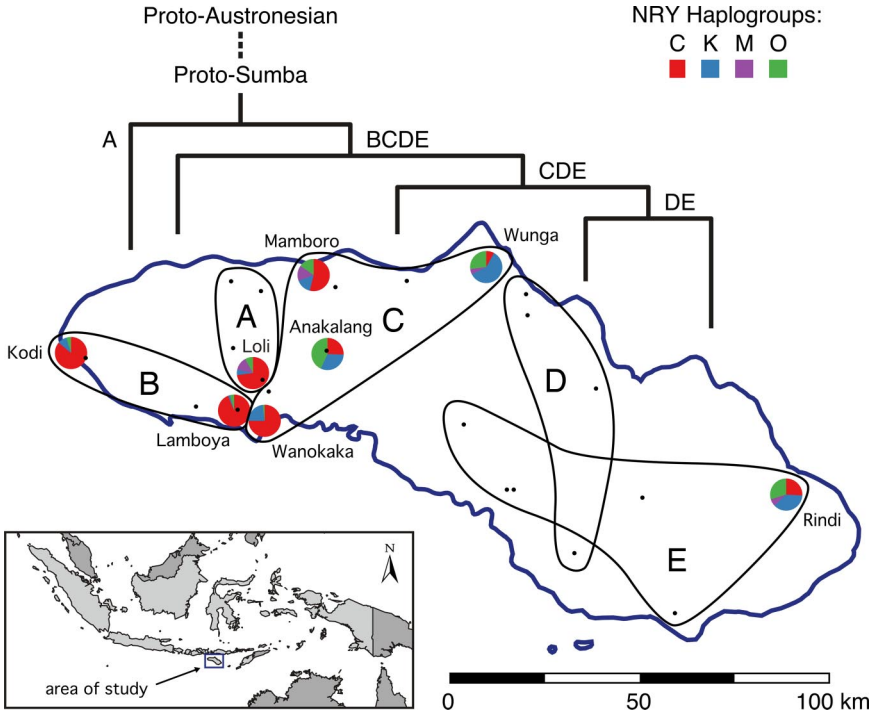

Fig. 2. Phylogenetic and geographic distributions of languages and Y chromosomes on Sumba. (*Upper*) Phylogenetic tree of Sumban language groups (A–E) (see *Materials and Methods*). (*Lower*) Map of Sumba showing geographic distribution of language groups (A–E) and Y chromosome haplogroups (C, K, M, and O). Pie charts represent frequencies of four Y chromosome haplogroups at eight locations sampled for both DNA and languages. Haplogroup O (green) is unevenly distributed, with lower frequencies in the western portion of the island. Small black dots indicate 20 additional language samples for which paired DNA samples were not available.

(PAn), whereas 65% are the result of innovations (i.e., word changes, losses, borrowings, etc.) that have occurred since the time of the common PAn ancestral language. Phylogenetic analysis was performed on the resulting database to reconstruct the relationships among these languages and to infer the scope and directionality of lexical borrowing and phonological change. The resulting tree clusters the languages of the 29 samples into five subgroups and strongly supports the hypothesis that Sumbanese languages derive from a single ancestral language, Proto-Sumba (PS), the identifiable lexicon of which is Austronesian in origin (Fig. 2). A sizable amount of the PAn vocabulary has been reconstructed (16, 17), and all of the languages in our sample retain PAn cognates. However, the proportion of these cognates in each language varies, and the lexicon of PS also contains many words that cannot be traced to PAn. Phonological and lexical variation between languages in our sample shows a clear geographic structure, with the main subgroups distributed across the island from west to east (Fig. 2). There is a pattern in the retention of PAn cognates, with the central languages being more conservative than those on the periphery, particularly the western languages. Sound change isoglosses (geographically bounded linguistic features) also are suggestive of higher heterogeneity in the western half of the island, where group A (and the northwest section of group C, apparently in heavy contact with group A) has been excluded by one of the most sweeping sound changes across the island, and where groups A and B (and again northwest C) have innovated another sound change that has not spread further east.

Y Chromosome Diversity. To investigate the paternal genetic relationships among Sumbanese villages, we obtained samples from 352 men inhabiting eight villages that also were sampled for languages and genotyped 71 SNPs. We identified 17 Y chromosome SNP haplogroups belonging to the C, K, M, and O lineages. These haplogroups have different histories: C-RPS4Y*, C-M38, K-M230, and M-P34 are believed to be associated with the colonization of eastern Indonesia by Paleolithic peoples. They are shared among eastern Indonesian, Papua New Guinean, and Melanesian populations and are absent or marginally present outside this region. An indigenous origin of K-M9* Y chromosomes also is suggested by its geographic distribution and the associated Y-STR diversity (data not shown). In contrast, haplogroup O likely entered this region much more recently (12, 18). Although there is controversy about the geographic origins of haplogroup O, there is general agreement that it is associated with the Late Holocene expansion of Austronesian-speaking farmers from southeast Asia to Indonesia and Oceania. Overall, only 16% of Sumbanese Y chromosomes belong to haplogroup O. The proportion varies among communities and in general exhibits a decreasing gradient from east to west, becoming rare in the western end of the island (Fig. 2).

Associations Among Linguistic, Genetic, and Geographic Distances. To test for language–gene associations, we performed matrix correlation tests (19) to evaluate the correspondence between linguistic distances determined from the reconstructed language tree and genetic distances among the eight sampled villages (Table 1). We observe a statistically significant positive correlation between linguistic and genetic distances ($r = 0.358$, $P = 0.023$). As pointed out by Nettle and Harriss (5), such a correlation could mean that languages and genes either coevolved from a single common ancestor (i.e., were shaped by the same events in population history) or evolved separately, but were conditioned by the same factors (i.e., were subjected to parallel but separate isolation by distance processes). One way to investigate this question is to control for geographic distance and test for a residual relationship between linguistic and genetic affiliations (5). Our finding of a slightly stronger correlation between linguistic and genetic variation when geography was held constant ($r = 0.474$, $P = 0.003$) provides evidence that languages and genetic variants on Sumba have actually evolved together. To verify that this association emerged within the time frame of the Austronesian expansion, we estimated the divergence time between the two most geographically distant communities in our sample, the villages of Rindi and Kodi, with an isolation-with-migration coalescent simulation model. The upper limit of the 95% confidence interval generated by the coalescent model is 4,875 years, which is consistent with the timing of the Austronesian expansion (20).

Table 1. Mantel test results for genetic, linguistic, and geographic distances

Correlation	*r*	*P*
Full correlation		
Genetics/geography	0.011	0.518
Genetics/language	0.358	0.023
Geography/language	0.673	0.000
Partial correlation		
Genetics/geography (language held constant)	−0.332	0.886
Genetics/language (geography held constant)	0.474	0.003
Language/geography (genetics held constant)	0.716	0.001

Genetic distance: Slatkin linearized $R_{ST}/1 - R_{ST}$. Language distance: ALINE distance (30).

We also found a strong positive correlation between linguistic and geographic distances, which becomes stronger when controlling for genetic variation ($r = 0.716$, $P = 0.001$), and no correlation between genetic and geographic distances (Table 1). The latter result is not surprising given the recent spread of farming on Sumba and the length of time typically needed for restricted gene flow between communities to produce a pattern of isolation by distance (21, 22). Moreover, because haplogroup O only accounts for ≈16% of Sumbanese Y chromosomes, we may have little power to detect a correlation between geographic and genetic differentiation associated with the expansion of Austronesian farmers. Focusing on the Austronesian component of genetic variation, we do find a positive correlation ($r = 0.627$, $P = 0.047$, by using a bootstrap approach; see *Materials and Methods*) between the percentages of haplogroup O and PAn cognates retained in the eight villages sampled for Y chromosomes (Fig. 3). This unprecedented correlation provides further

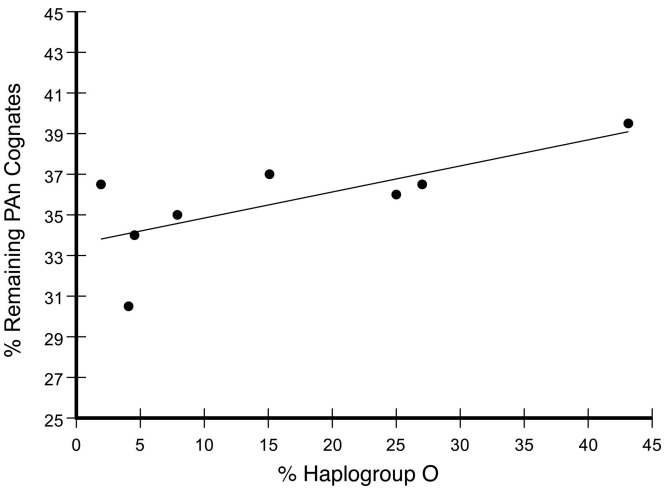

Fig. 3. Scatterplot of the percentage of PAn cognates versus the percentage of Austronesian Y chromosomes (haplogroup O) found at each location.

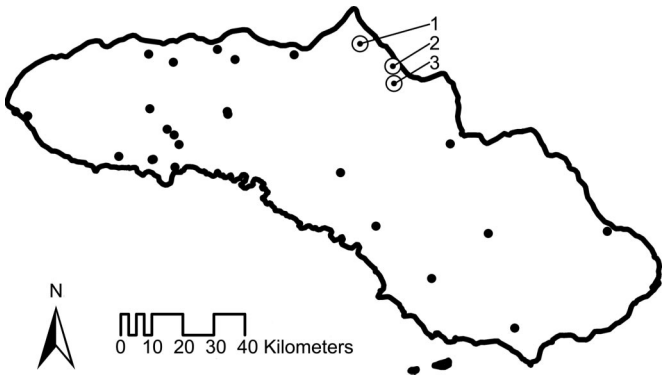

Fig. 4. Map showing approximate geographic locations of 29 language samples (black dots). Each of the three locations show a strongly significant correlation between their geographic distance from all other locations, and the percentage of PAn cognates retained are listed on the map (1, Wunga: *r*, −0.503; *P*, 0.006; 2, Rambangaru: *r*, −0.015; *P*, 0.005; 3, Kanatang: *r*, −0.507; *P*, 0.006).

evidence for the coevolution of linguistic and genetic variation on Sumba.

To investigate how the settlement history of the island may have influenced language evolution, we tested for correlations between the percentage of retained PAn cognates and geographic distances from a putative source population. We measured the geographic distances between each (putative source) population and all other populations and tested for a correlation with the percentage of PAn cognates retained. Only 3 of the 29 tested populations show a strongly significant correlation ($P < 0.01$), all of which are located on the central northern coast (Fig. 4). This result is concordant with oral history suggesting an origin near the village of Wunga (23) and implies that population history had a strong influence on patterns of language variation. Indeed, the language analysis (Fig. 2) is consistent with a splitting model in which a common founding population gave rise to daughter populations that subsequently diverged and produced several language subgroups. We hypothesize that, after settling near Wunga, the ancestral Austronesian population on Sumba expanded southward toward the center of the island. The first population split, which resulted in group A in the northwest, may have occurred before the Austronesians expanded south. By the time group B split and moved into the southwest, the main population must have expanded at least to the center of the island. This finding implies a later expansion to the east, with groups D and E splitting after the initial western expansion and probably after the main population (represented by group C) overtook the center of the island. Given the linguistic and genetic picture detailed previously, we hypothesize that this expansion must have involved a high degree of contact and intermarriage with the aboriginal population in successive stages, which would explain the generally low distribution of haplogroup O, its uneven distribution across the island, and the correlation with the percentage of retained Austronesian vocabulary in the subgroups.

Models of Language Change. This investigation of language and genetic variation on the island of Sumba produced results that are not easily explained by models of language evolution formulated on the basis of large-scale patterns of language variation (Fig. 1*A*). For example, simple models of language replacement without gene flow (e.g., elite dominance) or complete replacement of genes and languages (7) are not appropriate given the evidence for genetic admixture between Austronesian farmers and indigenous Papuan populations. Indeed, the frequency of indigenous Y chromosomes surpasses that of Austronesian Y chromosomes on Sumba (84% vs. 16%), with haplogroup O varying from 25–45% in the central and eastern parts of the island to <5% in the west. This distribution is consistent with a pattern of demic diffusion, whereby the incremental spread of farmers from their point of entry on the island was accompanied by frequent intermarriage with resident hunter-gatherers (9). It is unlikely that indigenous languages were fully replaced during the initial expansion of Austronesian on Sumba because we observe a high proportion of words (65%) that cannot be traced to PAn and loan words shared between different language groups that may have been absorbed from a now extinct indigenous source. Evidence for the latter hypothesis comes from the presence of non-Austronesian words (in particular, culturally significant words such as *husband*, *animal*, *dog*, and *sea*) in groups A and B (which do not form a subgroup) and their absence in subgroups C, D, and E to the east. Given the phylogenetic relationships in Fig. 2, this pattern is more easily explained by loans of these vocabulary items from a common non-Austronesian source, rather than by losses of ancestral Austronesian words in PS and later recovery in groups C, D, and E.

To account for these patterns of linguistic and genetic variation, we propose an alternative model of language evolution appropriate for the spatial scale of Sumba (Fig. 1*B*). In this model, intermarriage between expanding farmers and resident hunter-gatherers leads to progressively lower frequencies of haplogroup O Y chromosomes at increasing distances from the source population. What factors could lead to an association between Austronesian male lineages and the retention of PAn vocabulary across Sumba? Climate and population density data suggest that eastern Sumba remained sparsely populated during this expansion and new agricultural communities were relatively isolated. The north coast of East Sumba is the driest region in Indonesia, whereas West Sumba averages nearly three times more annual rainfall. This climatic variation is reflected in contemporary population densities: 28/km^2 in East Sumba and 97/ km^2 in West Sumba (24). We infer that, in preagricultural times, Sumba probably resembled aboriginal Australia, where human population density scaled with rainfall (25). In the wetter and more fertile region of West Sumba, expanding farmers likely came into contact with a larger indigenous population speaking non-Austronesian languages. This theory is attested to by lower frequencies of Austronesian lexical items (presumed to be due, at least in part, to loan words), as well as certain prominent phonological patterns in the west. As new farming villages proliferated in the populous west, the proportion of settlers of Austronesian descent would decrease, whereas the opportunities for linguistic contact would increase. Over time, these community-level processes gave rise to differential rates of language divergence/lexical borrowing and the association between languages and genes on Sumba. This scenario suggests a mechanism for language change: Rather than elite dominance, where a few individuals of an invading culture impose their language on a resident population, the extent of retention of PAn items is governed by the proportion of men in the population with Austronesian paternal ancestry. This codominant model also differs from the basic hypothesis of Diamond and Bellwood (7), in that a linguistic–genetic association evolves despite ongoing processes of demic diffusion and language shift. Whether the processes integrated in this model can explain patterns observed at continental scales remains an open question. However, a link can be postulated because large-scale patterns are contingent on processes occurring at local scales. This finding may be particularly true in the many cases of languages and genes spread by the recent dispersal of farmers (7, 26). More local-scale studies in contact zones with variable degrees of interaction among groups speaking different languages (e.g., Bantu and Khoisan in southern Africa, Indo-Iranian and Tibeto-Burman in south Asia, etc.) would be particularly helpful for determining the generality

of the model presented here. By incorporating lists of culturally appropriate words reflecting functional differences between farming and indigenous populations, future studies also may reveal more about the social dynamics favoring the retention of borrowed words. For now, the evidence provided here strongly suggests that language change in contact zones is scalar, language change can potentially vary in magnitude and character depending on factors inherent in the individual contact situation, and genetic analysis is a powerful tool that can be used to help formulate hypotheses of incipient language speciation.

Materials and Methods

Linguistic Classification. Lexical samples from 18 Sumbanese languages were obtained from lists collected and published by the National Language Center of the Indonesian National Department of Education (27) and cross-checked with word lists videotaped by J.S.L. and H.S at the sample locations. Videotaped recordings of 11 additional languages also were recorded at this time. These materials were organized and analyzed according to the principles of the traditional comparative method. Languages were first organized into rough groups according to shared lexical items. Where PAn or its descendent protoforms were known, these formed a backdrop with which to compare individual words, so that lexical innovations could be tagged as such and grouped together when shared. The second step was to organize these larger groups into subgroups based on shared phonological features. General reconstructions of lexical items were often possible at this point, and individual words could be compared with these reconstructions and phonological innovations noted. Languages that showed the same innovations were grouped together where appropriate. Finally, a search for loan words was conducted, the criteria being lexical and/or phonological innovations or retentions, which were unexpected within a certain subgroup. When potential loan words were identified, donor languages were sought out and, in many cases, identified based on geographic proximity to the borrowing language(s).

Y Chromosome Analysis. Two classes of markers on the Y chromosome, including 71 SNPs and 12 STRs, were genotyped as described elsewhere (18, 28).

Statistical Analyses. Geographic distances were calculated as a great circle distance based on GPS coordinates taken at the sample locations. Slatkin's linearized R_{ST} distances based on 12 Y chromosome microsatellites calculated with ARLEQUIN 3.0 (29) was used as the genetic distance. Quantitative measurement of distance between all pairs of languages was made by using ALINE distance based on the ALINE algorithm (30). This algorithm generates a score reflecting the phonetic similarity between words, which is converted to a distance by using a methodology currently in review (S.S.D., B.H, P.N., M.P.C, and J.S.L., unpublished data). The distance between two languages is calculated as the average distance between shared-meaning word pairs for those languages. To evaluate the correlation among linguistic, genetic, and geographic distances, we performed Mantel tests with ARLEQUIN 3.0 (27). A bootstrap analysis was used to estimate confidence intervals in the correlation between the percentages of PAn cognates and haplogroup O (see Fig. 3).

Swadesh word lists for Sumbanese languages were provided by the National Language Center of the Indonesian Department of Education. The Indonesian Institute of Science assisted with data collection. This work was supported by the National Science Foundation, the James McDonnell Foundation Robustness program at the Santa Fe Institute, and the Eijkman Institute for Molecular Biology.

1. Labov W (1994) *Principles of Linguistic Change*: *Internal Factors* (Blackwell, Oxford).
2. Barbujani G, Sokal RR (1990) *Proc Natl Acad Sci USA* 87:1816–1819.
3. Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J (1988) *Proc Natl Acad Sci USA* 85:6002–6006.
4. Chen JT, Sokal RR, Ruhlen M (1995) *Hum Biol* 67:595–612.
5. Nettle D, Harriss L (2003) *Hum Biol* 75:331–444.
6. Sokal RR (1988) *Proc Natl Acad Sci USA* 85:1722–1726.
7. Diamond J, Bellwood P (2003) *Science* 300:597–603.
8. Barbujani G (1997) *Am J Hum Genet* 61:1011–1014.
9. Ammerman AJ, Cavalli-Sforza LL (1984) *Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ Press, Princeton, NJ).
10. Renfrew C (1987) *Archaeology and Language* (Jonathan Cape, London).
11. Cox MP, Lahr MM (2006) *Am J Hum Biol* 18:35–50.
12. Kayser M, Brauer S, Cordaux R, Casto A, Lao O, Zhivotovsky LA, Moyse-Faurie C, Rutledge RB, Schiefenhoevel W, Gil D, *et al.* (2006) *Mol Biol Evol* 23:2234–2244.
13. O'Connell JF, Allen J (2004) *J Archaeol Sci* 31:835–853.
14. Bellwood P (1997) *Prehistory of the Indo-Malaysian Archipelago* (Univ of Hawaii Press, Honolulu).
15. Spriggs M (2003) *Rev Archaeol* 24:57–80.
16. Blust R (1995) *J World Prehist* 9:453–510.
17. Tryon DT (1995) *The Austronesian Languages* (Mouton de Gruyter, Berlin).
18. Karafet TM, Lansing JS, Redd AJ, Reznikova S, Watkins JC, Surata SP, Arthawiguna WA, Mayer L, Bamshad M, Jorde LB, *et al.* (2005) *Hum Biol* 77:93–114.
19. Mantel N (1967) *Cancer Res* 27:209–220.
20. Bellwood P (2005) *The First Farmers*: *The Origins of Agricultural Societies* (Blackwell, Oxford).
21. Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, *et al.* (2000) *Proc Natl Acad Sci USA* 97:6769–6774.
22. Slatkin M (1993) *Evolution (Lawrence, Kans.)* 47:264–279.
23. Hoskins J (1993) *The Play of Time*: *Kodi Perspectives on Calendars*, *History*, *and Exchange* (Univ of California Press, Berkeley, CA).
24. Badan PS, Kabupaten ST (2004). *Sumba Timur in Figures 2003* (Percetakan Usaha Mulia, Waikabubak, Sumba).
25. Yengoyan A (1972) *Oceania* 43:85–95.
26. Renfrew C (2000) *Cambridge Archaeol J* 10:7–34.
27. Bahasa P (2002) *Kosakata Dasar Swadesh di Kabupaten Belu*, *Ngada*, *Sumba Barat*, *Sumba Timur*, *dan Timor Tengah Utara* (Departemen Pendidikan Nasional, Rawamangun, Jakarta).
28. Redd AJ, Agellon AB, Kearney VA, Contreras VA, Karafet T, Park H, de Knijff P, Butler JM, Hammer MF (2002) *Forensic Sci Int* 130:97–111.
29. Excoffier L, Laval G, Schneider S (2005) *Evol Bioinf Online* 1:47–50.
30. Downey SS, Hallmark B, Cox MP, Norquest P, Lansing JS (2007) *Working Paper of the Santa Fe Institute* 07-08-021 (Santa Fe Inst, Santa Fe, NM).