Male dominance rarely skews the frequency distribution of Y chromosome haplotypes in human populations

J. Stephen Lansing*†, Joseph C. Watkins‡§, Brian Hallmark*, Murray P. Cox†¶, Tatiana M. Karafet¶, Herawati Sudoyo‖, and Michael F. Hammer¶

Departments of *Anthropology and ‡Mathematics, and ¶Division of Biotechnology, Biological Sciences West, University of Arizona, Tucson, AZ 85721; †Santa Fe Institute, Santa Fe, NM 87521; and ‖The Eijkman Institute for Molecular Biology, Diponegoro 69, Jakarta 10430, Indonesia

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved June 3, 2008 (received for review October 25, 2007)

A central tenet of evolutionary social science holds that behaviors, such as those associated with social dominance, produce fitness effects that are subject to cultural selection. However, evidence for such selection is inconclusive because it is based on short-term statistical associations between behavior and fertility. Here, we show that the evolutionary effects of dominance at the population level can be detected using noncoding regions of DNA. Highly variable polymorphisms on the nonrecombining portion of the Y chromosome can be used to trace lines of descent from a common male ancestor. Thus, it is possible to test for the persistence of differential fertility among patrilines. We examine haplotype distributions defined by 12 short tandem repeats in a sample of 1269 men from 41 Indonesian communities and test for departures from neutral mutation-drift equilibrium based on the Ewens sampling formula. Our tests reject the neutral model in only 5 communities. Analysis and simulations show that we have sufficient power to detect such departures under varying demographic conditions, including founder effects, bottlenecks, and migration, and at varying levels of social dominance. We conclude that patrilines seldom are dominant for more than a few generations, and thus traits or behaviors that are strictly paternally inherited are unlikely to be under strong cultural selection.

cultural evolution | Indonesia | neutral theory | selection | Red Queen

Evolutionary social scientists analyze the fitness consequences of behavior, where the currency of fitness is reproductive success. Many studies have argued that reproductive skew biased toward dominant or high-ranking men is very common in human communities: "In more than one hundred well studied societies, clear formal reproductive rewards for men are associated with status: high-ranking men have the right to more wives" (1). Demographic statistics collected over short time scales support these claims (2). Although variation in male fitness is known to occur, an important unanswered question is whether such differences are heritable and persist long enough to have evolutionary consequences at the population level. In this report we show that it is possible to use genetic markers to test for the signature of heritable reproductive skew within population groups.

Several authors have modeled the effects of reproductive skew in different contexts (3, 4). Fig. 1 contrasts two models: the standard Wright-Fisher model in which all individuals produce offspring with equal probability (i.e., no reproductive skew) and a model of male dominance in which individuals whose parents had many offspring are more likely to have more children themselves. Investigations of genetic variation on the nonrecombining portion of the Y chromosome provide an opportunity to ask which of these models better characterizes a given population. Here, we genotype 12 short tandem repeats (Y-STRs) and a battery of SNPs on the Y chromosomes of 1269 men from 41 Indonesian communities. We then construct Y-chromosome haplotypes for each sample [supporting information (SI) Table S1]. Unlike human genetic studies that sample unrelated individuals from a broad geographical watershed and thus fail to capture community-level processes, we sample male residents of several indigenous and relatively isolated communities. These populations include neolocal Borneo hunter-gatherers and Central Javanese rice farmers, matrilocal horticulturalists on Flores, patrilocal Balinese wet-rice farmers, and patrilocal horticultural clans on Sumba, Nias, and Flores.

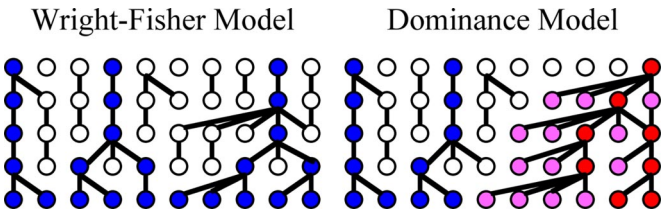


Fig. 1. Population models. Colored dots represent individuals who have descendants in the final generation. Red dots represent dominant individuals who are more likely to reproduce. Pink dots represent nondominant individuals having a dominant ancestor. Blue dots represent nondominant individuals having no dominant ancestor. Dominant individuals in 1 generation are chosen at random from the offspring of dominant individuals in the previous generation. (See *SI Text Population Models* for more details.)

To detect evidence of heritable reproductive skew, we use statistical procedures that are based on the Ewens sampling formula (5). In particular, we apply Slatkin's exact test of neutrality, which is based on the haplotype frequency distribution (6, 7). Communities experiencing reproductive skew among patrilines will tend to have haplotype frequency distributions that, over time, become unlikely under the neutral model, skewed toward an excess of common haplotypes (patrilines). Such skewed distributions would provide evidence for cultural selection, defined as the heritable nongenetic transmission of any kind of behavior that affects reproductive success (see ref. 8).

Results

Strikingly, we find that only 5 of 41 communities (12%) have haplotype frequency distributions that are unlikely under the neutral model (i.e., rejected the null hypothesis of neutrality). Fig. 2 shows the haplotype frequency distributions of 1 of these 5 communities (Fig. 2*B*), and a neutral deme (Fig. 2*A*). All 5

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

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

§To whom correspondence should be addressed. E-mail: jwatkins@math.arizona.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0710158105/DCSupplemental.

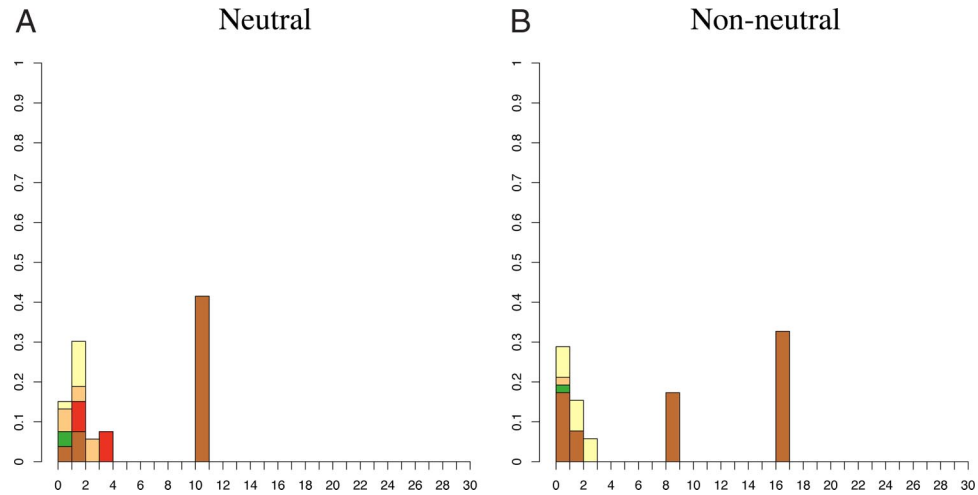


Fig. 2. Example haplotype distributions. The height of column *i* is the number of individuals whose haplotype is represented *i* times in the sample (normalized by sample size). (*A*) An example of a community haplotype distribution that does not show evidence of departure from neutrality based on the exact test: Sumba/Mamboro. (*B*) An example of a community showing departure from neutrality based on the exact test: Sumba/Wanokaka. Although nonneutral cases often appear to have mass shifted to the right, it generally is not possible to determine the outcome of the exact test by inspection. Colors indicate Y chromosome haplogroup membership (C, F, K, M, O, or S). All 41 distributions and the color key are given in Fig. S1. Additional haplogroup information is in Table S5.

cases that rejected neutrality show a similar pattern: 1 or more common haplotype(s) (i.e., a high-frequency patriline) paired with a large number of low-frequency haplotypes (Fig. S1). This contrasts with patterns produced under the neutral model with different population sizes. Observed haplotype frequency distributions typical of large communities (i.e., in which the genetic diversity of the sampled population is high) are characterized by many unique haplotypes and a few high-frequency lineages. Smaller communities with reduced neutral diversity tend to have fewer low-frequency haplotypes and fewer common patrilines.

Dominance Model and Power of the Exact Test. Given these findings, we wanted to know how often the Ewens exact test fails to reject neutrality when male dominance is in fact occurring. The Ewens test is nonparametric and is not based on any specific model. It is independent of both village effective population size and Y-STR mutation rate. We constructed an analytical model to determine the ability of this test to detect departures from neutral equilibrium as a result of heritable reproductive skew caused by social dominance. In this model, dominant males comprise a fraction δ of the total population and have a selective advantage σ over the nondominant males. Offspring in the next generation are chosen according to a haploid Wright-Fisher model with competition between the 2 classes, dominant and nondominant. To maintain the fraction of dominant males at δ for each subsequent generation, a proportion of sons of the dominant males is chosen at random to be dominant for the next iteration of the model. Thus, for this 2-parameter model, either $\delta = 0$ or $\sigma = 0$ yields neutral populations (see *SI Text Population Models* for additional details).

To calculate the power of the Ewens exact test under this model, we simulate nonneutral populations, sample them, and apply the exact test to observe departures from the expectation under neutrality. To condition the simulations on our Indonesian data, we simulate populations taking (*i*) samples of size 20 from smaller villages having effective population size 100 using 10 Y-STRs, and (*ii*) samples of size 35 from larger villages having effective population size 300 using 12 Y-STRs. Tables S2 and S3 show the results of these analyses for all parameters. For example, for a village of size 300 with only 4% of the men producing twice as many offspring ($\sigma = 1$), the power of our test is 0.39. Assuming this level of dominance, the probability that 5 (or fewer) of 41 villages rejects neutrality is ≈0.02%. Thus, even weak dominance would be detectable in a sample of 41 villages. Fig. 3 gives this probability for the power calculated from any model of selection.

Demographic Effects. Departures from neutrality can arise from both selective and demographic causes (e.g., changes in population size and structure). Consequently, inferring whether observed cases result from reproductive skew or from other factors is not straightforward. There are 2 possibilities. First, demographic processes might mask the effects of male dominance and make villages appear neutral when in fact they are experiencing heritable reproductive skew, leading to false negatives. Alternatively demographic processes might cause neutral populations to appear nonneutral (i.e., leading to false positives). In this section we discuss the effects of 2 classes of demographic process on the Ewens test: reductions in population size (including founder effects and bottlenecks) and migration.

In general, bottlenecks (or founder events) reduce genetic

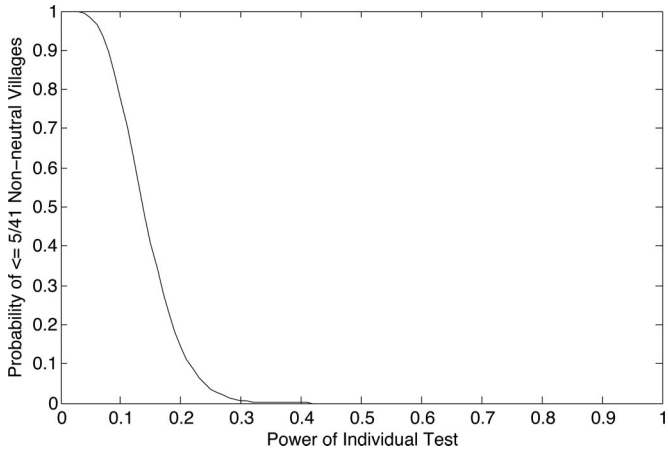


Fig. 3. Probability of seeing 5 or fewer (of 41) nonneutral villages for a given power of the individual tests to detect departure from neutrality. If the power of the individual tests is 0.2, the probability of observing ≤5 is ≈0.2. If the power of the individual tests is > 0.3, then this probability is vanishingly small.

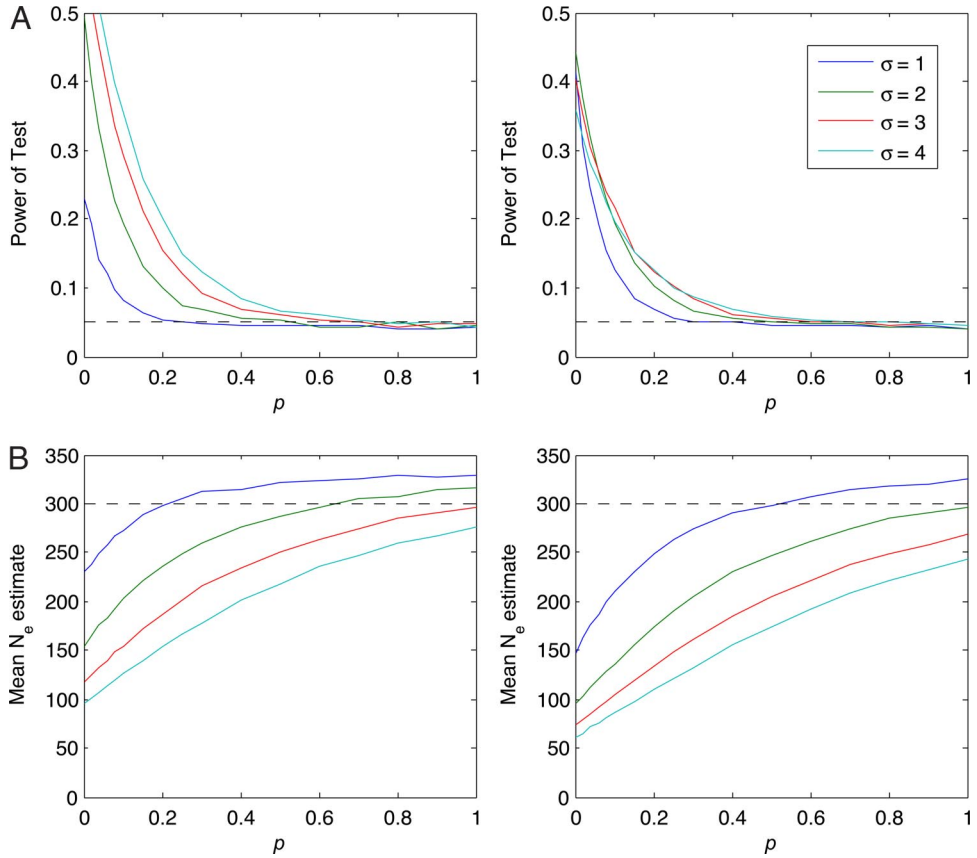


Fig. 4. Effects of adding nonheritable reproductive skew to a simulated population of size $n = 300$. The first column shows simulation results for $\delta = 0.02$, the second column for $\delta = 0.06$. (*A*) Reduction of power. As p increases, the power to detect skew is reduced. However, the length of time lineages are dominant also decreases, nullifying any long-term evolutionary advantage of dominance. (*B*) N_e reduction. Effective population size remains depressed for all values of p, providing an additional way to assess the strength of both heritable and nonheritable skew. Estimates eventually exceed the actual population size because the estimator is biased upwards.

variation (Fig. S2) by leading to the rapid loss of rare lineages. Eventually, if the bottleneck remains strong, more common lineages will be lost also. In the phase of population recovery after a bottleneck, new mutations tend to occur on different Y chromosomes. Thus, during the bottleneck phase we expect to find a reduction in the frequency of rare haplotypes, whereas in the recovery phase we expect to see an excess of rare haplotypes compared with the expectation for a population in neutral equilibrium. For false positives, the most extreme case is a neutral population undergoing a bottleneck in which all individuals possess the same haplotype. This is conceivable, for example, in the case of a founding event by a small kin group. Thus, neutral populations may test as nonneutral in the period immediately following a strong bottleneck (Fig. S3*A*) (see *Materials and Methods*).

In a population experiencing reproductive skew as a result of male dominance, a bottleneck will reduce genetic variation further. Because a reduction in haplotype diversity reduces the power of the exact test, dominance will be more difficult to detect during the recovery phase. Simulations indicate that populations return to the nonneutral condition in ~20 generations as long as dominance remains constant (Fig. S3*B*). This raises the question of whether we are underestimating the importance of male dominance as a result of low power to reject neutrality after a bottleneck. We believe this to be unlikely for several reasons. First, the mean village size, at 280 male household heads, is sufficiently large to indicate that frequent bottlenecks occurring in the past few generations are very unlikely (Table S4). Second, haplogroup diversity (i.e., the number and relative abundance of Y-chromosome lineages as defined by SNPs) is relatively high (Table S5). Finally, reductions in effective population size that one expects to accompany a bottleneck are not seen. The bottlenecks simulated here are quite extreme, with reductions in population size of 90% for 10 generations. Moreover, if bottlenecks do occur, they also independently will reduce the demographic effects of male dominance. For example, if a population is reduced from size 300 to 35 for 10 generations, then the chance that any individual has a descendant in a recovering population is $< 2\%$ (this becomes $< 1\%$ if the bottleneck lasts for 20 generations) (Fig. S2). Thus, it is unlikely that Y-chromosome lineages of dominant males would survive a severe population bottleneck.

Migration is another factor that affects patterns of diversity within populations. Increasing the number of new alleles entering the population each generation by migration is functionally equivalent to increasing the mutation rate, μ, in a Wright-Fisher model. Consequently, migration will not cause false positives because the exact test is not dependent on the mutation rate under this model. However, for villages experiencing reproductive skew, high migration reduces the power of the Ewens exact test. In our sample, the highest migration rates are expected in Balinese villages because they often are adjoining and speak the same language. We simulated an Isolation with Migration model to estimate migration rates (see *SI Text Estimation of Balinese Migration Rates*). To determine whether this could impact the power of our tests significantly, we simulated moderate and high migration ($m = 0.01$, 0.03/gen) as part of our power analyses (Tables S2 and S3). Even when sample sizes are small and

dominance is weak, the reduction in power with these levels of migration is minimal, amounting to only a few percent. For example, with 4% dominant males producing 3 times as many children as nondominant males, high migration reduces the power from 32% to 26% in small village simulations (Table S2).

Nonheritable Reproductive Skew. We also consider another model of reproductive skew in which the composition of the dominant group varies. The distinction between the heritable and nonheritable cases is important: any basis for cultural selection must involve some heritable behavior or trait that confers a reproductive advantage. In the context of social dominance, this distinguishes between situations in which, over multiple generations, males persistently pass dominance to their sons versus situations in which the membership of reproductively dominant groups is transient. To consider this case, we add a parameter p to the model of dominance. This parameter gives the fraction of the dominant class that enters from the nondominant class in a given generation. Consequently, for example, if $p = 0.05$, then a dominant lineage will span, on average, $20 = 1/0.05$ generations. If $p = 1$, then reproductive skew is uncorrelated between generations. In that limiting case, dominant individuals produce significantly more offspring than nondominants, but this tendency is attained randomly in each generation. Values of p significantly >0 could result in a "Red Queen" effect, in which even acute competition between patrilines does not translate into lasting dominance (8). The power analysis for this model shows a small reduction in power for the case $p = 0.05$. For $p = 0.15$, dominant lines persist on average only 6 or 7 generations, and, not surprisingly, we find a clear reduction in power for this case. The power of selection to structure the genetic composition of the community also is reduced as p increases and the composition of the dominant group becomes more fluid (Fig. 4A). Concomitant with this result, we also find in our simulations a clear reduction in the village effective population size, N_e, for small values of p and some reduction for all values (Fig. 4*B*). To look for reductions in N_e, we estimated the effective population sizes of sampled communities based on the number of haplotypes and compared these values with best-estimate demographic data made available by village records or the local Indonesian government. For the communities in which the neutrality test was not rejected, effective size estimates (compensating for migration) are similar to census demographic values (Table S4). In contrast, the nonneutral communities have N_e estimates that are much smaller than the number of male heads of households, in some cases dramatically smaller. Because we do not see a marked reduction in N_e for the villages that test neutral, and these villages maintain high levels of haplogroup diversity (Table S5), it appears that values of $p \gg 0$ are uncommon for the communities in our sample. This argues against significant reproductive skew, which would reduce both N_e and diversity quickly in small communities.

Cases That Depart from Neutral Expectations. Finally, we investigate possible explanations for departures from neutrality for each of the 5 nonneutral communities. Three communities come from Sumba, a remote island in which residence is patrilocal, descent is traced through the patriline, marriage is polygamous, and competition for status and resources among clans is endemic (9). Given these circumstances, it perhaps is remarkable that the remaining 5 Sumbanese communities we sampled failed to show statistically significant evidence of heritable reproductive skew. In Bali, where residence is patrilocal and there is competition among patrilineal descent groups, the only nonneutral community is South Batur. This village was part of the larger village of Batur, which fragmented in 1948 after a period of rivalry between factions that often were associated with descent groups. A very recent process of village fissioning led to resettlement of some households based on lineage affiliation (10). The remaining nonneutral community is a patrilineal and patrilocal community in central Flores. It once was the site of a minor princedom that became an administrative center during the Dutch colonial era and now serves as a district capital ("kecamatan"). Our sample includes a large proportion of civil servants born elsewhere, probably accounting for its diverse haplotype distribution.

In sum, most of the sampled populations do not show evidence of a departure from neutral stochastic equilibrium with respect to male lineages. If reproductive skew inherited between generations were a pervasive and ongoing process, we would expect to observe frequent rejections of the Ewens exact test of neutrality. We do not observe such rejection in 88% of our Indonesian communities. Even our nonneutral Indonesian communities may not necessarily reflect the action of male reproductive skew but rather the signature of a very recent or nonneutral founding group. For the reasons discussed earlier, recent bottlenecks are unlikely to be masking significant skew. We conclude that male reproductive skew is at best weak in most of our sampled population groups, despite their varied subsistence strategies and kinship practices.

Discussion

The implications of our results extend beyond the Indonesian cases described here to the broader question of the relationship between reproductive skew and the genetic structure of human communities. The genetic or cultural-evolutionary effects of dominance become apparent only by extending analyses from the inclusive fitness of individuals to population-level consequences over the course of multiple generations. Here, we take a population-genetic approach to infer selection by detecting departures from neutral drift-mutation equilibrium at the population level rather than from variation in the reproductive success of individuals (11–13). Our results indicate that dominance effects generally do not persist over multiple generations. The lack of evidence of reproductive skew in these communities means that heritable traits or behaviors that are passed paternally, be they genetic or cultural, are unlikely to be under strong selection.

The discovery that neutral processes can explain most haplotype frequency distributions in these communities parallels earlier results from the development of neutral theory in genetics and ecology. As Kimura (14) observed in his original article, the prevalent opinion in the 1960s held that almost all mutations are under selection. This opinion was slow to change. More recently, ecologists similarly have suggested that a neutral model, in which species in the same trophic level are functionally equivalent or neutral with respect to each other, might adequately explain species-abundance distributions in ecological communities (15–17). In anthropology, the recent availability of appropriately sampled community-level polymorphism data now enables us to distinguish both genetic and cultural selection from neutral demographic processes with surprising precision. In these Indonesian communities, male dominance seldom translates into increased fertility among descendents over evolutionary timescales.

Materials and Methods

Ewens Sampling Formula. Our statistical methods are based on the appropriateness of the Ewens sampling formula (5) as a null model for the process under study. This sampling formula applies to closed populations in situations that meet the following criteria for samples, genetic data, infinite alleles, and equilibrium:

1. Samples: The sample size n is small compared with the constant haploid population size, N.
2. Genetic Data: Each mutation is selectively neutral, taking place from 1 generation to the next with probability μ.

3. Infinite Alleles: Each mutation gives rise to a novel haplotype.
4. The population is in equilibrium.

The sampling formula gives the distribution of a sample taken from the assumed equilibrium frequency distribution or configuration of the population. A configuration is denoted by $\mathbf{b} = (b_1, b_2, \ldots)$ where b_i is the number of haplotypes represented i times in a sample of n individuals. Thus, possible sampling configurations satisfy $\Sigma_i i b_i = n$. The Ewens sampling formula states that this configuration has probability

$$P(\text{configuration } \mathbf{b}) = \frac{n!}{\theta(\theta+1)\ldots(\theta+\mathrm{n}-1)} \prod_i \left(\frac{\theta}{\mathrm{i}}\right)^{b_i} \frac{1}{b_i!}.$$

The choice of θ generally is approximately equal to $2N_e\mu$ where N_e is the effective population size.

We now examine the assumptions individually. Then we show how the Ewens sampling formula forms the basis for a test of neutrality and an estimator of θ based on the number of haplotypes. We apply the test and the estimator to the sample sites. Finally, we introduce a 2-parameter model to assess the impact of these departures from neutrality and the power of the test and compare our method with another approach.

1. Samples. In total, 1269 Y chromosomes were collected from 41 communities on 6 islands (Bali, Borneo, Flores, Java, Nias, Sumba) in Indonesia (Fig. S4). Sample sizes range from 16 to 54 individuals. All samples were collected from volunteer donors with written informed consent and appropriate permits from the Indonesian Government via the Eijkman Institute for Molecular Biology. The University of Arizona Human Subject Committee approved sampling protocols. In all cases, care was taken to exclude individuals related within the last 3 generations, as determined by detailed personal genealogies. This aspect of the sampling procedure effectively eliminates the effects of recent demographic events.

Samples were designed to be a small proportion of the village. The number of samples n in each community is given in Table S1. The assumption of constant population is a part of our null hypotheses. Departures from the constant population assumption will be considered in our power analysis.

2. Genetic data. The polymorphic sites from the nonrecombining part of human Y chromosome included a set of 74 previously published binary markers (18) and 4 additional polymorphisms: M208, M210, M346, and M356 (19–22). Binary markers were analyzed with a hierarchical strategy (23, 24) when additional sample genotyping was restricted to the appropriate downstream mutations along the haplogroup tree.

For the microsatellite analysis, 10 STRs (*DYS19*, *DYS388*, *DYS389I*, *DYS389II*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, *DYS426*, and *DYS439*) were typed as described by Redd *et al.* (25). Some sites also were scored for *DYS438* and *DYS457* loci. Table S1 indicates the total number of STRs typed for each sampling location.

A haplotype is defined by its sequence of repeat numbers. Consequently, a haplotype mutates if at least 1 STR mutates. If we let M denote the number of STRs, and let ν_i denote the mutation rate for the ith Y-STR, then we can calculate the mutation rate for Y-STR haplotypes as

$$\begin{aligned}\mu &= P(\text{at least one STR mutates}) = 1 - P(\text{no STR mutates})\\ &= 1 - \prod_i P(i\text{th STR does not mutate})\\ &= 1 - \prod_i (1-\nu_i)\\ &\approx 1 - \left(1 - \frac{1}{M}\sum \nu_i\right)^M.\end{aligned}$$

Consequently, μ depends only on the average mutation rate, not its variability. Using the average STR mutation rate of 0.0021 per generation (26, 27) for the average above, we calculate:

$$10 \text{ STRs: } \mu = 1 - (1-0.0021)^{10} = 0.0208,$$
$$12 \text{ STRs: } \mu = 1 - (1-0.0021)^{12} = 0.0249.$$

3. Infinite alleles. Y-STRs mutate by either increasing or decreasing their repeat number. The measure of the failure of the infinite alleles assumption is the mean frequency of STR mutations that do not result in a novel haplotype. This possibility of 2 individuals being identical in state but not identical by descent is called ''homoplasy.''

A variety of STR mutation models have been studied (28–30). Mutation models that permit multiple repeat changes are less likely to mutate to an existing haplotype than mutation models that permit only a mutation changing the repeat number by1. In addition, as we shall see, homoplasy affects tests for neutrality and estimates for θ only if it reduces the number of haplotypes in the sample. Irrespective of the model, the impact of homoplasy is to reduce the number of haplotypes and to increase the fraction of individuals that belong to common haplotypes.

4. The population is in equilibrium. Neutral models for the evolution of the configuration of individuals are Markov chains. For Markov chains that model this evolution, the distribution of the configuration of haplotypes converges to its equilibrium. The questions arise: What is the time needed for this convergence, and what is the nature of the convergence? Analytical results and simulations (data not shown here) confirm that these Markov chains possess a cutoff phenomenon. The cutoff occurs at approximately N_e (logθ)/θ generations. Here, we take a generation to be 31 years (31).

The analysis also shows that a neutrally evolving population has reached equilibrium if each individual in the population has experienced at least 1 mutation along the line of descent connecting that individual to a founding member of the population. If a community is neutrally evolving but the genetic data have not yet reached equilibrium, then the current population will have some signature of the founding population. Fig. S3*A* shows this phenomenon for a single neutral population that began with all individuals sharing the same haplotype.

Statistical Procedures. The Ewens sampling formula is a 1-parameter probability distribution function. It can be expressed as the product of two terms

$$\frac{n!}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_i \left(\frac{\theta}{i}\right)^{b_i} \frac{1}{b_i!}$$
$$= n! \prod_i \left(\frac{1}{i}\right)^{b_i} \frac{1}{b_i!} \cdot \frac{\theta^K}{\theta(\theta+1)\cdots(\theta+n-1)}.$$

For a fixed value of K, the first term gives the likelihood for a given configuration $\mathbf{b}$. The second shows that the number of haplotypes $K = \sum_i b_i$ is a sufficient statistic for the parameter θ. Note that the form of this product shows that, given the number of haplotypes, the distribution of haplotypes does not depend on the parameter θ and consequently does not depend on either the population size or the mutation rate.

Test for Neutrality. Based on the first term in the product of the Ewens sampling formula, Slatkin (6, 7) developed an exact test for significant departures from the neutral hypothesis. The exact test calculates the probability of all configurations with fixed sample size, n, and fixed haplotype number, K, and sums the probabilities of the configurations that are less likely than the observed configuration. That value, P_E, is reported in Table S1. The test is 2-tailed: values of $P_E < 0.025$ indicate a departure from neutrality in the direction of a distribution that is ''too even,'' often taken to indicate the presence of balancing selection. Values of $P_E > 0.975$ indicate departure in the direction of an ''overly uneven'' distribution, which can indicate either the presence of positive selection or demographic history. Table S1 shows the values of P_E for all of the sampled villages. Highlighted departures from neutrality have $P_E > 0.975$.

Power Analysis. To determine the power of the test to detect neutrality under the dominance model, populations were simulated under a range of parameters values for σ and δ under 2 choices for village population size and 2 levels of migration (Tables S2 and S3). For each set of parameters, 10,000 samples were simulated, the exact test was applied, and the number testing as nonneutral was counted. The estimated power of the test is the fraction testing as nonneutral. See *SI Text* for further details of the analysis

Estimate of θ. The maximum likelihood estimator of θ is found by solving (32):

$$K = \sum_{j=0}^{n-1} \frac{\theta}{\theta + j},$$

ANTHROPOLOGY

where n is the sample size. Using value of μ given earlier, we are able to estimate effective population size, N_e, using K. The values for K and the estimated values for θ for each community are given in Table S1. We compare these estimates with the number of male household heads known from census data in Table S4. Estimates for θ are biased upwards.

ACKNOWLEDGMENTS. We thank these individuals for their analysis of the genetic data at the Hammer Laboratory of the University of Arizona: Meryanne Tumonggor, Yuliana Ling, Daniel Teberg, Megan Dragony, Emily Landeen, Fernando Mendez, and Rachel Stewart. We also thank John Schoenfelder, who provided the sampling locations map, Jennifer Dunne, John Pepper, and two anonymous reviewers for comments, and the 1269 Indonesian men who made this study possible. Genetic samples were obtained by JSL and HS and by Golfiani Malik, Wuryantari Setiadi, Loa Helena Suryadi, and Meryanne Tumonggor of the Eijkman Institute, with the assistance of Indonesian Public Health clinic staff, following protocols for the protection of human subjects established by both the Eijkman Institute and the University of Arizona Institutional Review Boards. 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, Jakarta, Indonesia. Permission to conduct research in Indonesia was granted by the Indonesian Institute of Sciences.

1. Clarke AL, Low BS (2001) Testing evolutionary hypotheses with demographic data. *Popul Dev Rev* 27:633–660.
2. Winterhalder B, Smith E (2000) Analyzing adaptive strategies: Human behavioral ecology at twenty-five. *Evol Anthropol* 9:51–72.
3. Cannings C (1974) The latent roots of certain Markov chains arising in genetics: A new approach, I. Haploid models. *Adv Appl Probab* 6:260–290.
4. Sibert A, Austerlitz F, Heyer E (2002) Wright-Fisher revisited: The case of fertility correlation. *Theor Popul Biol* 62:181–197.
5. Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theor Popul Biol* 3:87–112.
6. Slatkin M (1996) A correction to the exact test based on the Ewens sampling distribution. *Genet Res* 68:259–260.
7. Heyer E, Sibert A, Austerlitz F (2005) Cultural transmission of fitness: Genes take the fast lane. *Trends Genet* 21(4):234–239.
8. Van Valen L (1973) A new evolutionary law. *Evol Theor* 1:1–30.
9. Lansing JS, *et al.* (2007) Coevolution of languages and genes on the island of Sumba, eastern Indonesia. *Proc Natl Acad Sci USA* 104:16022–16026.
10. Lansing JS (2006) *Perfect Order*: *Recognizing Complexity in Bali* (Princeton Univ Press, Princeton, New Jersey).
11. Fisher RA (1930) *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, UK).
12. Hofbauer J, Sigmund K (1991) *The Theory of Evolution and Dynamical Systems* (Cambridge Univ Press, Cambridge, UK).
13. Nowak, M. A (2006) *Evolutionary Dynamics*: *Exploring the Equations of Life* (Harvard Univ Press, Cambridge, Massachusetts).
14. Kimura M (1968) Evolutionary rate at the molecular level. *Nature* 217:624–626.
15. Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ Press, Princeton, New Jersey).
16. Hu X-S, He F, Hubbell SP (2006) Neutral theory in macroecology and population genetics. *Oikos* 113:548–556.
17. Alonso D, Etienne R, McKane A (2006) The merits of neutral theory. *Trends Ecol Evol* 21:451–457.
18. Karafet TM, *et al.* (2005) Balinese Y-chromosome perspective on the peopling of Indonesia: Genetic contributions from pre-Neolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Hum Biol* 77:93–114.
19. Kayser M, *et al.* (2003) Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet* 72:281–302.
20. Kayser M, *et al.* (2006) Melanesian and Asian origins of Polynesians: MtDNA and Y-chromosome gradients across the Pacific. *Mol Biol Evol* 23:2234–2244.
21. Hudjashov G, *et al.* (2007) Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc Natl Acad Sci USA* 104:8726–8730.
22. Sengupta S, *et al.* (2005) Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian Pastoralists. *Am J Hum Genet* 78:202–221.
23. Underhill PA, *et al.* (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358–361.
24. Hammer MF, *et al.* (2001) Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 18:1189–1203.
25. Redd AJ, *et al.* (2002) Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci Int* 130:97–111.
26. Dupuy BM, Stenersen M, Egeland T, Olaisen B (2004) Y-chromosomal microsatellite mutation rates: Differences in mutation rate between and within loci. *Hum Mutat* 23:117–124.
27. Gusmão L, *et al.* (2005) Mutation rates at Y chromosome specific microsatellites. *Hum Mutat* 26:520–528.
28. Di Rienzo A, *et al.* (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA* 91:3166–3170.
29. Sainudiin R, Durrett RT, Aquadro CF, Nielsen R (2004) Microsatellite mutation models: Insights from a comparison of humans and chimpanzees. *Genetics* 168:383–395.
30. Watkins JC (2007) Microsatellite evolution: Markov transition functions for a suite of models. *Theor Popul Biol* 71:147–159.
31. Fenner JN (2005) Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* 128:415–423.
32. Tavaré, S., Zeitouni, O (2004) *Lectures on Probability Theory and Statistics. Ecole d'Etés de Probabilité de Saint-Flour XXXI* (Springer-Verlag, New York).

Supporting Information

Lansing *et al.* 10.1073/pnas.0710158105

SI Text

Population Models. We take as our neutral model the standard haploid Wright-Fisher model with constant population size, N, and discrete generations (1). In each generation, the N sons choose their fathers from the previous generation with equal probability $1/N$ and inherit the label of their father. This is equivalent to choosing sons according to a multinomial $Multi(N;(1/N,\ldots,1/N))$ distribution.

With reproductive skew, some subset of the population is more likely to reproduce. In the Wright-Fisher model this corresponds to non-uniform probabilities of individuals being chosen as fathers. For each of the N individuals, let ρ_i, $1 \leq i \leq 1$, $\Sigma_i \rho_i = 1$, be the probability that the i^{th} individual is chosen as a father. This defines the "propensity to reproduce" vector:

$$\rho = (\rho_1, \ldots, \rho_N)$$

For the neutral case this is simply:

$$\rho = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right).$$

To investigate the power of the test, we must set an alternative to neutrality. For this purpose, we introduce a 3-parameter model that captures the central concept of a dominant group reproducing more on average. Let δ denote the fraction of *dominant* individuals, σ denote the selective advantage for the *dominant* individuals, and p denote the fraction of the dominant class that enters from the non-dominant class in a given generation.

Write

$$\rho = (1+\sigma, 1+\sigma, \ldots, 1+\sigma, 1, 1, \ldots 1)/(\mathrm{N}(1+\sigma\delta)),$$

where $N\delta$ entries take the value $1+\sigma$ and the remaining $N(1-\delta)$ entries are 1. Equivalently, in this model, sons are chosen according to a $Multi(N; 1+\sigma, \ldots, 1+\sigma, 1, \ldots, 1)/(N(1+\sigma\delta))$ distribution.

To include migration into these models, a fraction m of the men are replaced with migrants. These migrants are modeled as males having a new haplotype.

$N\delta$ males are chosen to become the dominant males in the next generation. For our model of heritable reproductive skew ($p = 0$), the dominant males in 1 generation are chosen at random from the offspring of dominant males in the previous generation. For the model of non-heritable reproductive skew, these males are chosen at random from the entire population. For $p > 0$, each individual in the dominant group is removed from this group with probability p and replaced with a previously non-dominant individual who now becomes dominant. Thus, under these models of reproductive skew, the fraction of the dominant group remains fixed at δ.

Robustness and Power Analysis. All of the communities that reject neutrality do so by having too few haplotypes and by having common haplotypes too commonly represented. This coincides with the concerns about homoplasy affecting the infinite alleles assumption and too recent a founding of communities affecting the assumption of equilibrium. This makes the rejection of neutrality more likely and leads us to conclude that the claims of neutrality are conservative based on the assumptions needed to use the Ewens sampling formula.

The power of the Ewens exact test depends on the parameters δ, σ, p, and $\theta = 2N_e(\mu + m)$. Values for μ have been determined in the *Materials and Methods* section. We now look to establish representative values for the migration rate, m and the effective population size, N_e.

Estimation of Balinese Migration Rates from Y-Chromosome Microsatellites. Migration rates were estimated for 6 population pairs located in the central Mount Batur region of Bali. Gene flow was inferred under a generalized, non-equilibrium demographic model, isolation-with-migration (IM). This model describes a constant-sized ancestral population that splits into 2 daughter populations linked by ongoing migration (2). Current implementations allow these daughter populations to grow or contract (3). Unlike traditional equilibrium models (e.g., island and divergence models), the IM model allows migration rates to be inferred against a backdrop of shared polymorphism derived through common ancestry.

The computer software IM (http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm#IM; version 16/8/07), implements the IM model using Markov chain Monte Carlo and a Bayesian approach to parameter estimation. Uniform prior distributions are established for 7 parameters of interest: the effective sizes of the ancestral deme and its 2 daughter demes, the proportion of the ancestral population that founded the first daughter deme, the time at which the ancestral population split, and unidirectional migration rates between the 2 descendent demes. Only marginal posterior densities for the migration rate parameters are of immediate interest here.

Analyses were undertaken on completely linked Y-chromosome STRs under a single-step mutation model of microsatellite evolution. Three independent runs with multiple geometrically heated Markov chains ensured complete chain mixing via Metropolis-Hastings coupling, thereby leading to convergence of the posterior density to its true stationary distribution.

Averaged across the 6 population pairs, unidirectional migration rates (i.e., considering only the men entering a community) approach $m = 0.036$ per generation. Note that these values probably are toward the high end for migration rates between Indonesian populations. These particular communities are located only a few kilometers apart, usually on the same river system, and were chosen specifically because they are close both geographically and culturally. For instance, 1 community in each population pair probably budded from the other within the last $\approx$30 generations, and all communities speak the same language, Balinese. These conditions are not representative of most of our samples. We conclude that the migration rates inferred above for central Balinese communities provide upper limits on migration rates between small human communities in Indonesia.

Village Size. For each of the 41 villages, the number of households was determined. These values and the source of the information is given in Table S4. In addition, estimates of effective population size N_e are based on the maximum likelihood estimate of θ assuming no migration. Consequently, villages experiencing historically significant migration will have increased estimates for N_e. This is particularly noticeable for villages in Bali, as suggested by the analysis of migration rates given previously.

Power Analysis. Based on the information on migration and village size, we shall consider 2 circumstances for the power analysis:

1. Villages population $n = 100$, samples of size $n = 20$, and mutation rate $\mu = 0.0208$.
2. Villages population $n = 300$, samples of size $n = 35$, and mutation rate $\mu = 0.0249$.

To compute an empirical power function, we fix a value for δ and σ and simulate a non-neutral population of size N until it reaches equilibrium and then draw at random a sample of size n for the next 10,000 generations. At each generation, compute the statistics P_E from Slatkin's exact text. Under the neutral model, P_E is uniformly distributed from 0 to 1 independent of the value of K, the number of haplotypes.

To estimate the power for a 5%-level test, we take the 2-sided rejection region consisting of values of P_E either below 0.025 or above 0.975 and tabulate the fraction of instances in which the test is rejected. This fraction is the estimate of the power of the test.

Impact of Non-Heritable Reproductive Skew. The impact of non-heritable reproductive skew is assessed by inclusion of the parameter p described previously. Simulations to determine power, displayed in Fig. 4*A*, are calculated in the same way as those in Tables S2 and S3 for σ and δ. Values for the effective population size (Fig. 4*B*) were determined by taking the maximum likelihood estimate for θ and dividing by twice the mutation rate ($\theta = 2N_e\mu$).

The case $p = 1$ falls under the general exchangeable models of Cannings (ref. 3 in text). Thus, samples under equilibrium will, under the same assumptions as before, satisfy the Ewens sampling formula and appear neutral under the Ewens exact test. In this case, however, the inbreeding effective population size is reduced by a factor $(\sigma\delta + 1)^2/((\sigma + 1)^2\delta + 1 - \delta)$.

Comparison with Another Approach. As described in the text, Sibert *et al.* (4) present a classification for the cultural inheritance of fitness. Their inferential techniques are based on statistics that measure the imbalance of genealogical trees. In Blum, *et al.* (5), these techniques are applied to examine mtDNA data and make inferences concerning maternal fertility inheritance. They apply their techniques to compare fertility inheritance between food-producing populations and hunter-gatherer populations.

Direct comparison of methods is difficult. Their studies use publicly available mtDNA data. In this study, we collect Y-chromosome data under a sampling strategy designed to investigate the history of communities. In addition, we choose a model for departure from neutrality that is more suitable for our study. The exact test has an advantage over the tree imbalance test because it is non-parametric.

All simulations were performed in MATLAB, and copies of the code are available on request.

1. Tavaré S, Zeitouni O (2004) *Lectures on Probability Theory and Statistics. Ecole d'Etés de Probabilité de Saint-Flour XXXI* (Springer-Verlag, New York).
2. Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158:885–896.
3. Hey J (2005) On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biol* 3:e193.
4. Sibert A, Austerlitz F, Heyer E (2002) Wright-Fisher revisited: The case of fertility correlation. *Theor Popul Biol* 62:181–197.
5. Blum MGB, Heyer E, François O, Austerlitz F (2006) Matrilineal fertility inheritance detected in hunter-gatherer populations using the imbalance of gene genealogies. *PLoS Genet* 2:1138–1146.

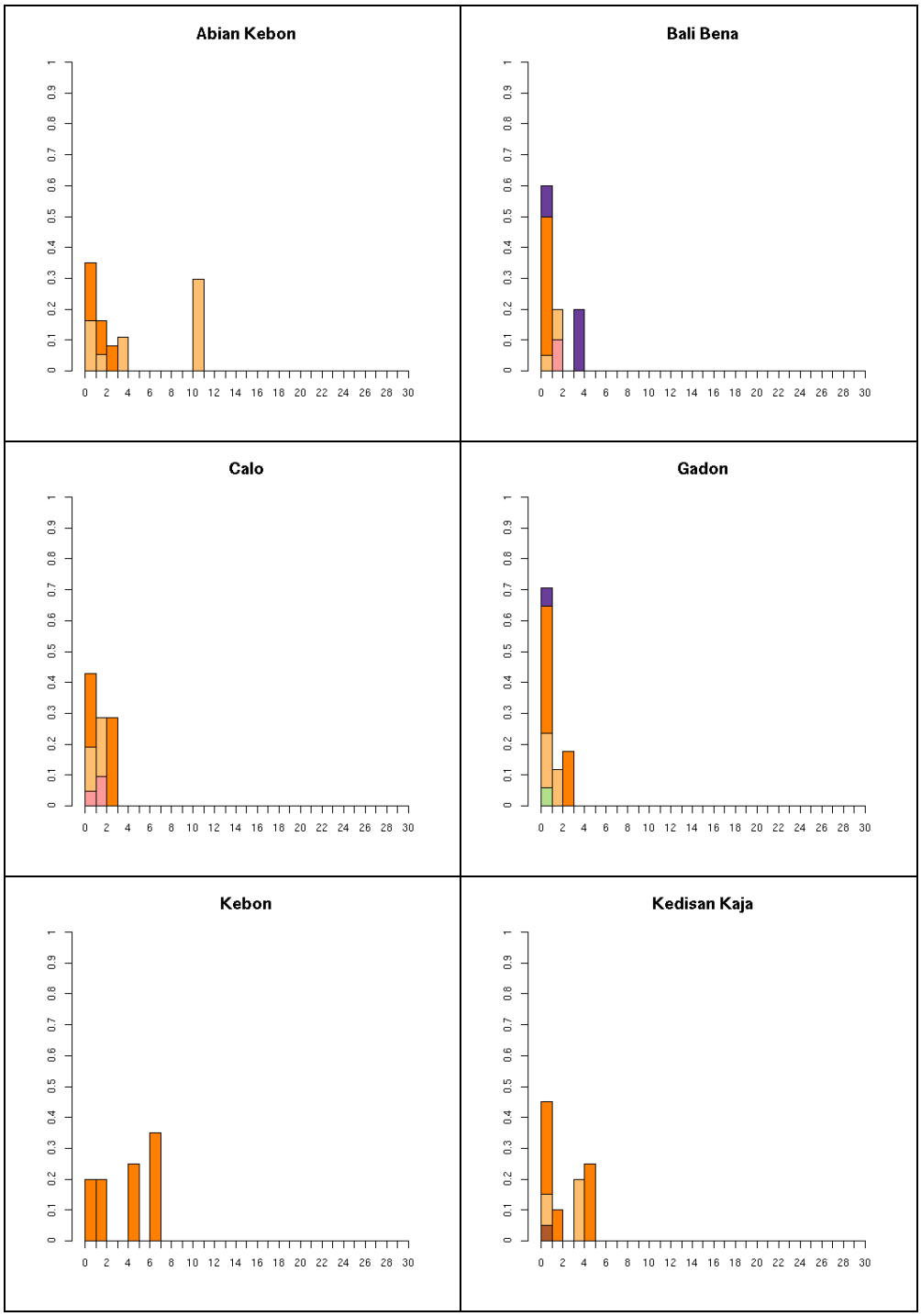
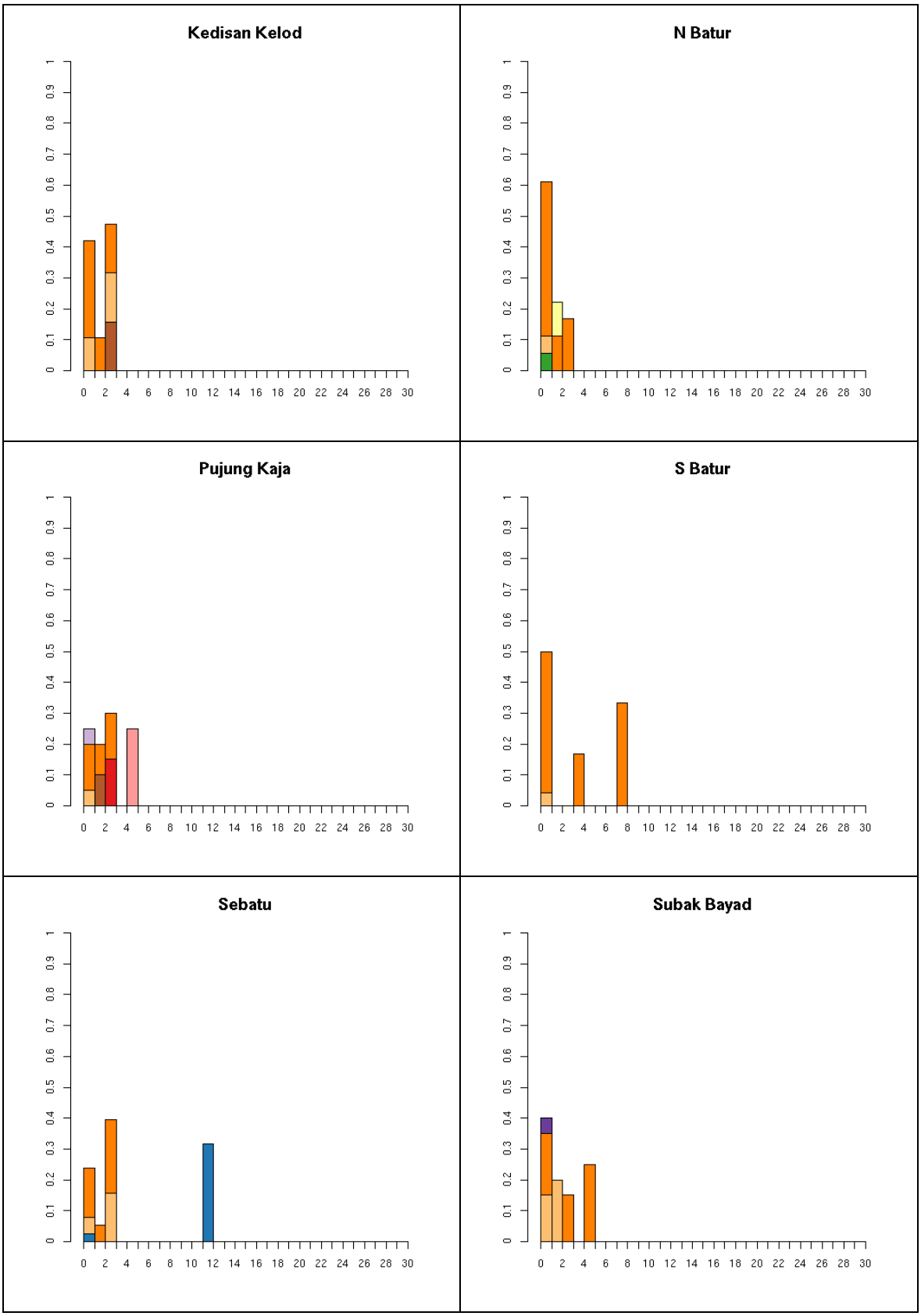
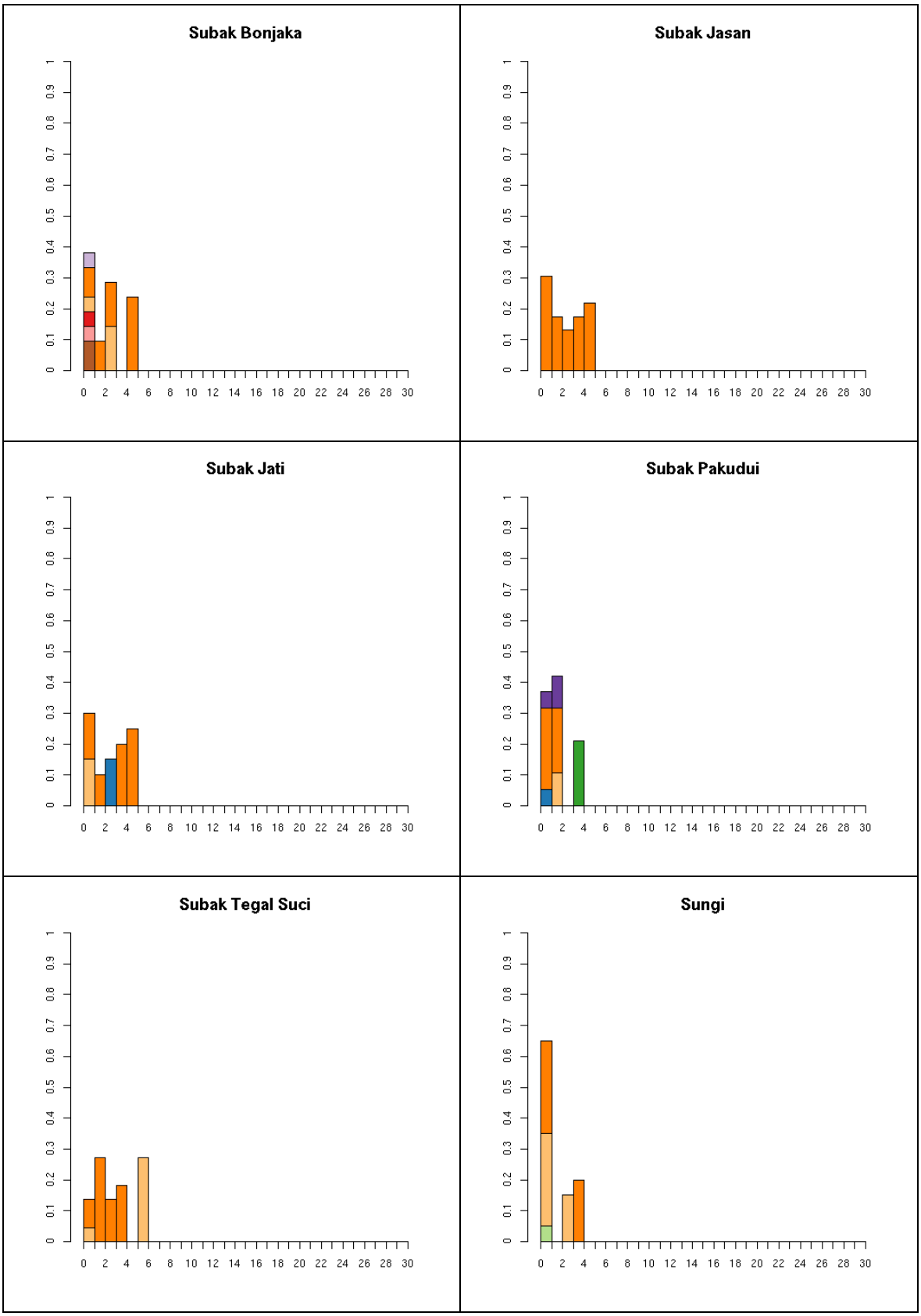


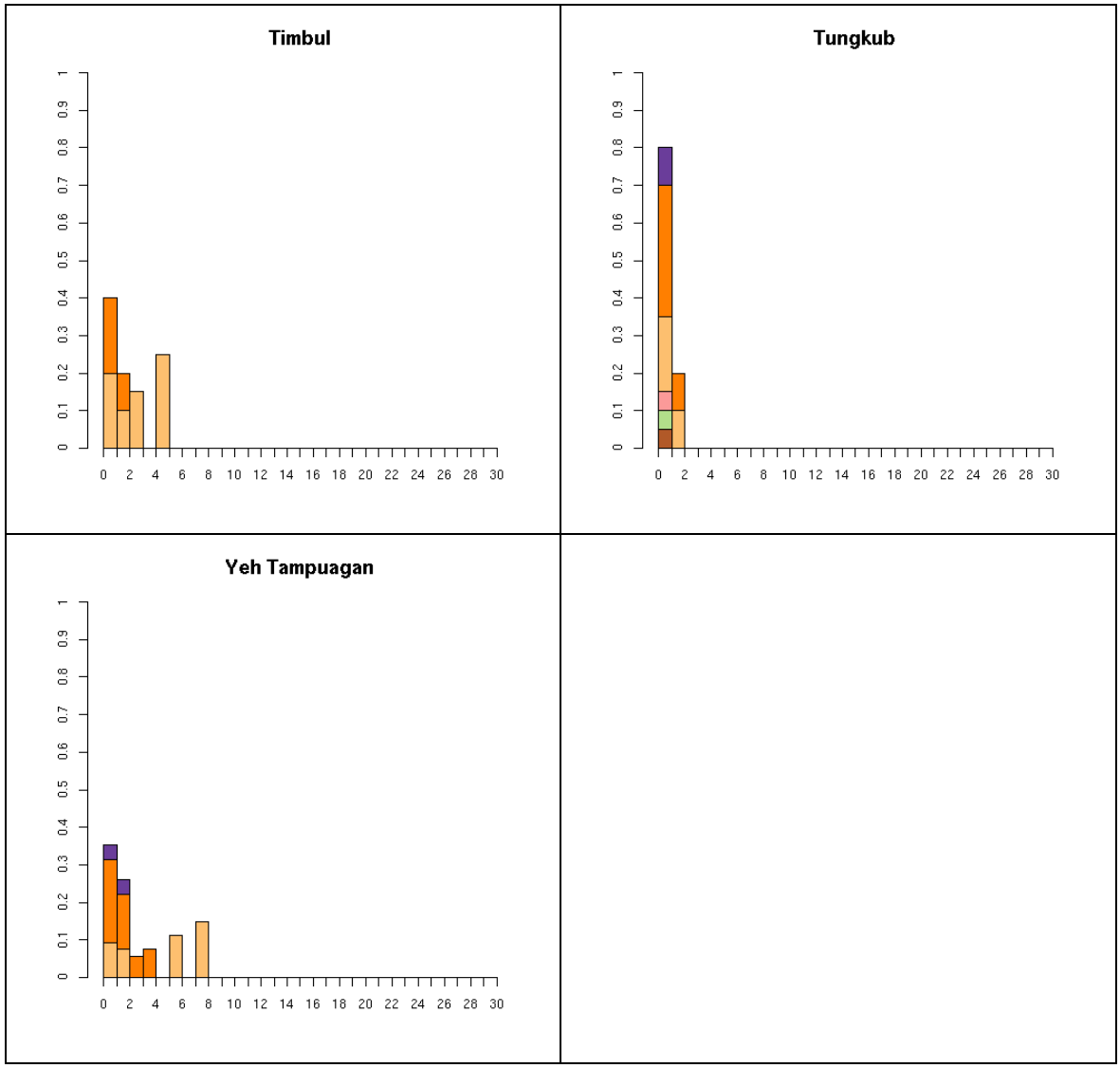
Fig. S1. All 41 haplotype distributions, normalized by number of samples for comparison. Colors represent haplogroup membership designated by the following colors: brown, C; light blue, F; dark blue, H; light green, J, dark green, K; pink, L; red, M; orange, O; dark orange, O-M95; light purple, Q; dark purple, R; yellow, S. In some cases multiple haplogroups have been grouped together (e.g., different O lineages) for clarity so the total number of colors may be smaller than the number of haplogroups in Table S5. In general, it is not possible to tell whether the exact test will reject neutrality by inspection. This is partially because under neutrality the shape of the distribution will vary depending on the underlying population size. The distributions are presented in the same order as the tables for comparison.



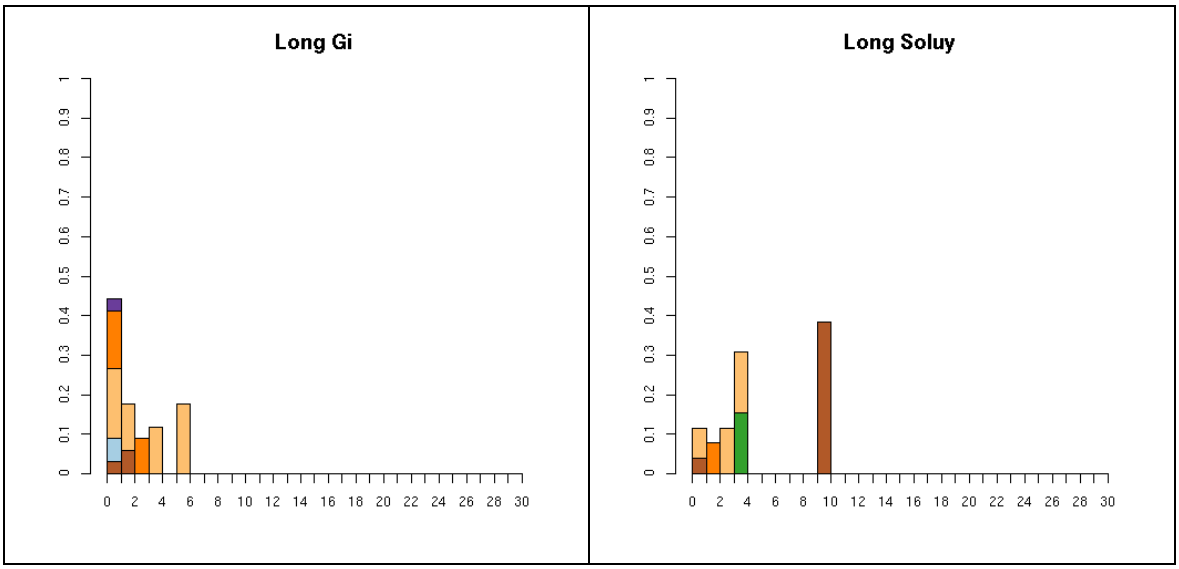
Haplotype distributions for sites on Bali



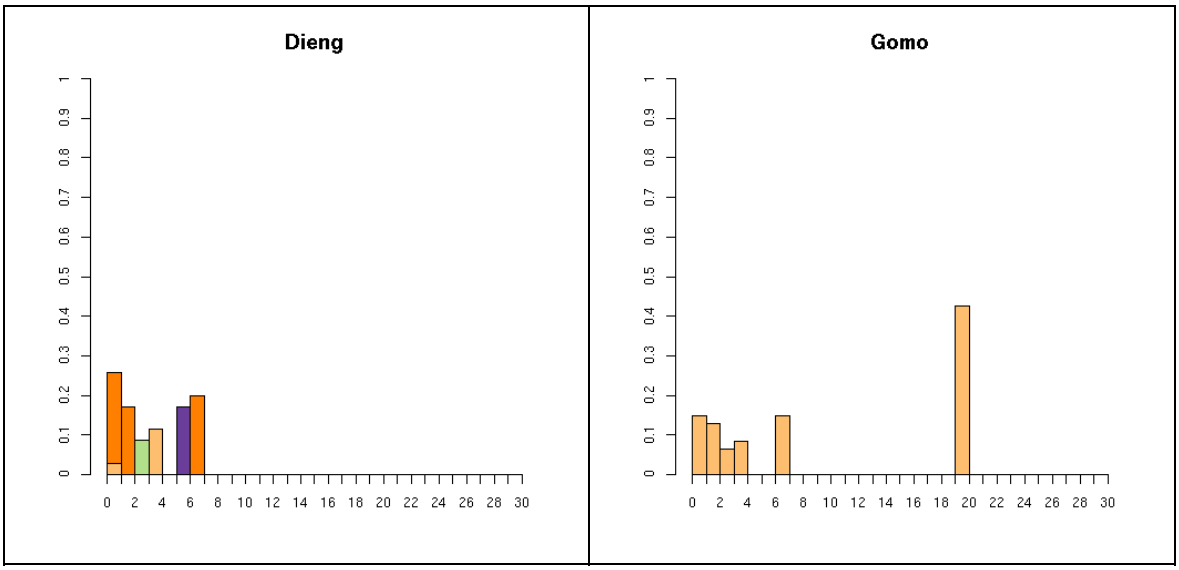
Haplotype distributions for sites on Bali



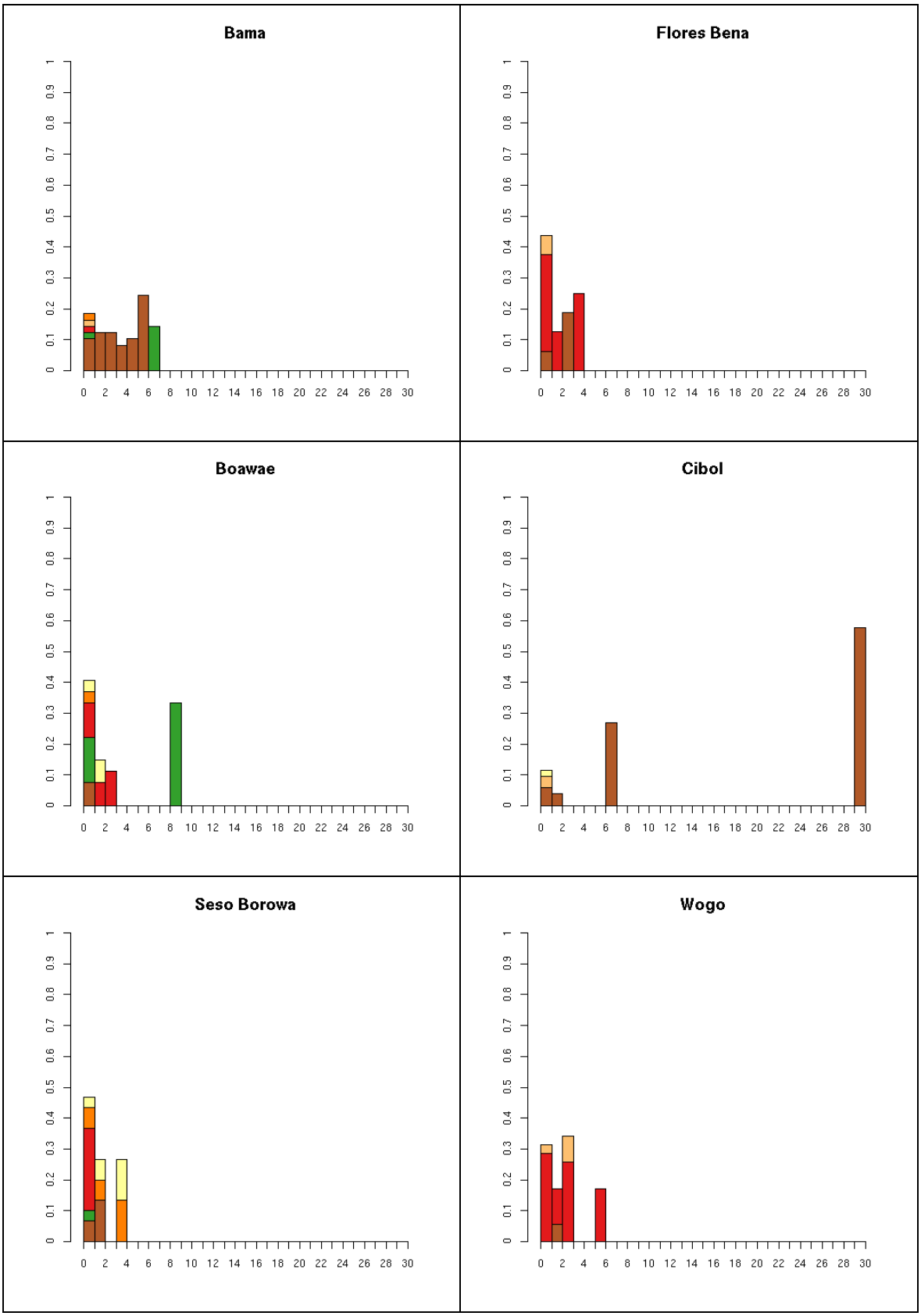
Haplotype distributions for sites on Bali



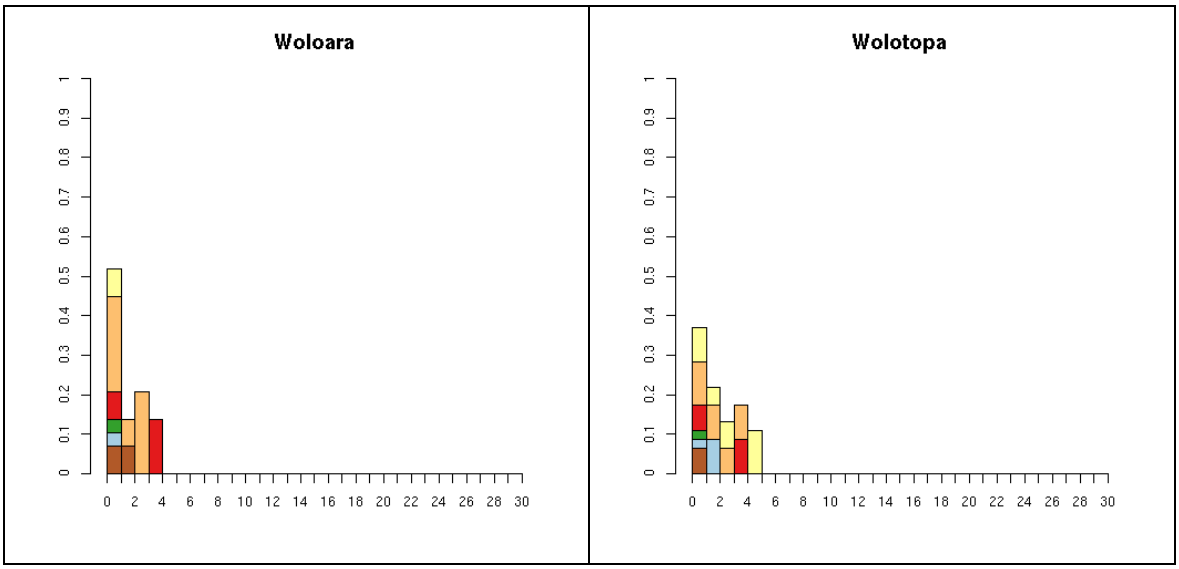
Haplotype distributions for sites on Borneo



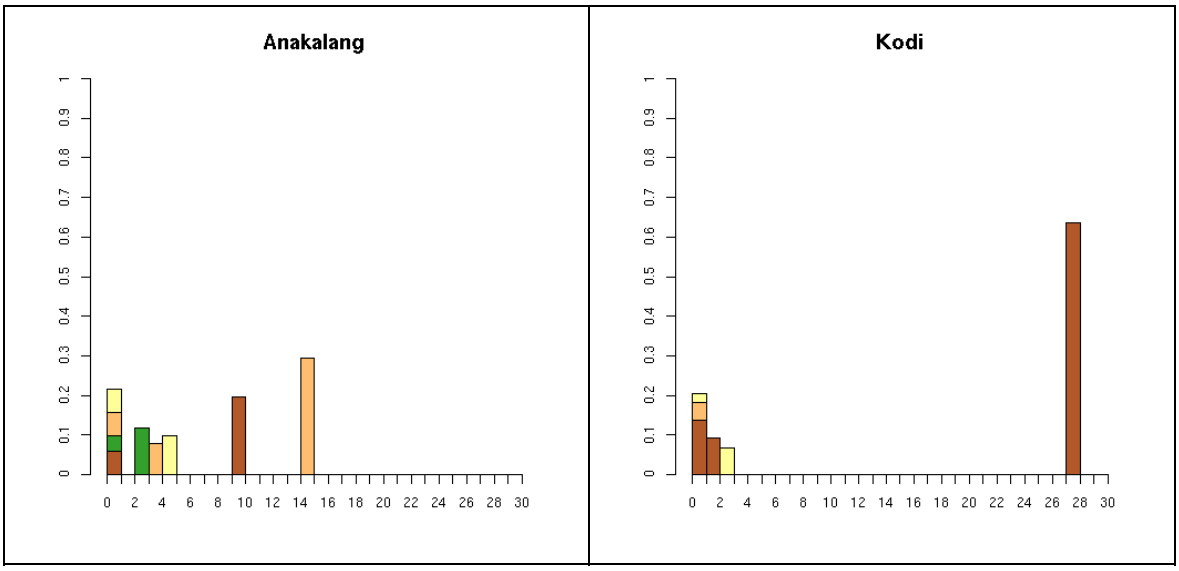
Haplotype distributions for sites on Java and Nias



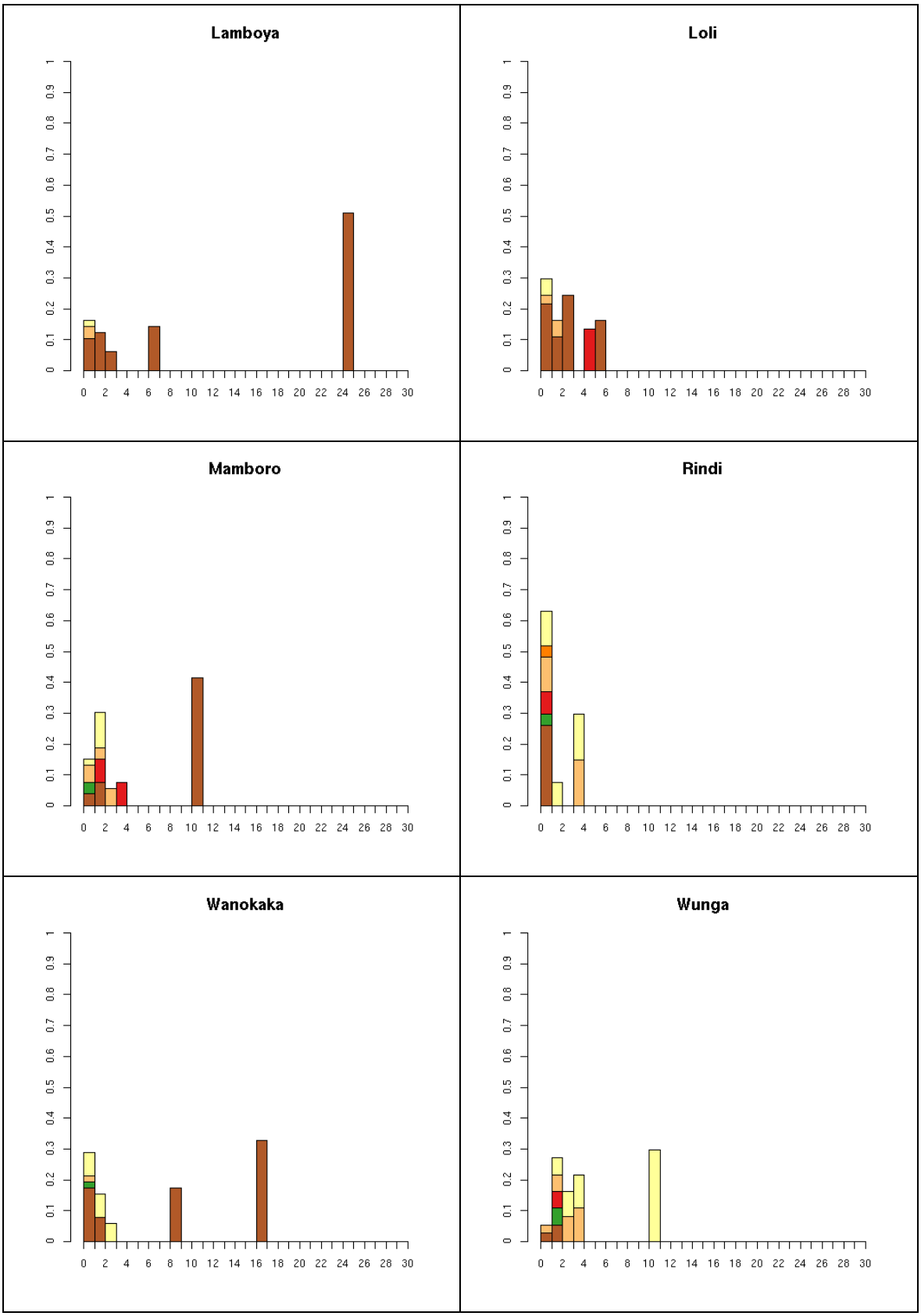
Haplotype distributions for sites on Flores



Haplotype distributions for sites on Flores



Haplotype distributions for sites on Sumba



Haplotype distributions for sites on Sumba

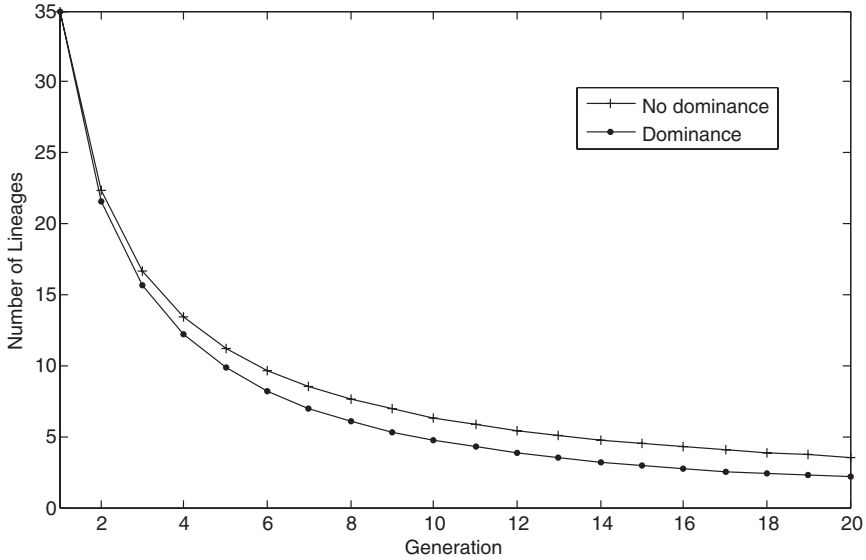


Fig. S2. Reduction of the number of lineages during a bottleneck of size 35. The graph shows the average number of lineages each generation for 5000 simulated populations. For the dominance case, the parameters are $\delta = 0.06$ and $\sigma = 2$.

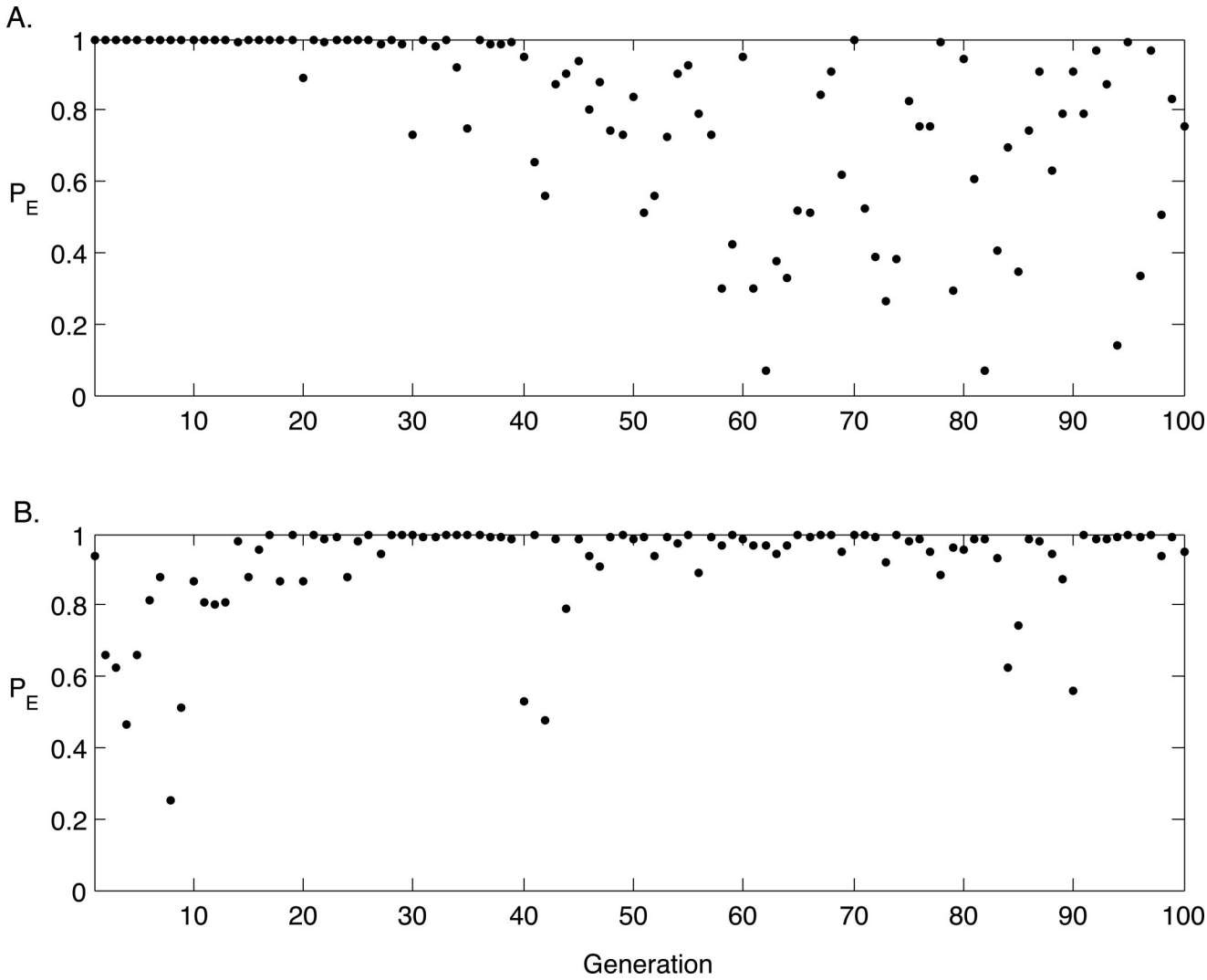


Fig. S3. Examples of performance of exact test after recovery from bottleneck/founder effect. In both simulations the initial population size is 35 and the population grows to a final size of 300 over ≈25 generations (≈750 years). The graphs show the value of P_E from the exact test for each generation. (*A*) Neutral population. The initial state is all individuals sharing a single haplotype, the case yielding the most false positives for the exact test. (*B*) Non-neutral population with $\delta = 0.05$ and $\sigma = 2$ with the initial state a sample of 35 from a non-neutral population of size 300 with the same dominance parameters. This choice of initial state is more likely to yield false negatives than starting the reduced haplotype diversity expected to emerge from a bottleneck. This indicates that bottlenecks are more likely to produce false positive results rather than to mask dominance. For this simulation, 0.54 samples tested as non-neutral, similar the rate expected based on the power of the test with constant population size (Tables S2 and S3).

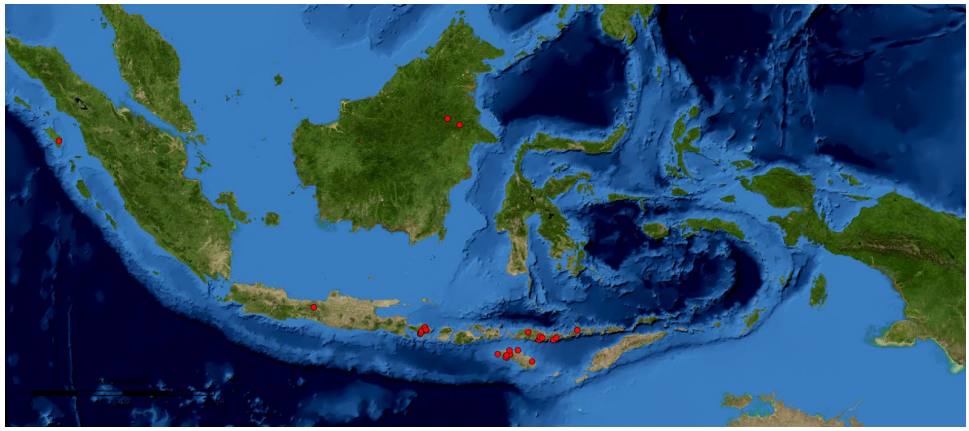


Fig. S4. Map showing community sampling locations (*red*) in Indonesia. Sampled islands (west to east) are Nias, Java, Bali, Borneo, Sumba and Flores.

Table S1. Full list of sampling locations, sample sizes, *n*, number of STRs typed, number of haplotypes *K* in the sample, estimate of θ, and P_E from the Ewens' exact test

Island	Site	*n*	# STRs	*K*	Estimated θ	P_E
Bali	Abian Kebon	37	12	19	14.98	0.968
	Bena	20	10	15	25.59	0.922
	Calo	21	10	14	17.19	0.382
	Gadon	17	10	14	34.68	0.956
	Kebon	20	10	8	4.44	0.701
	Kedisan Kaja	20	10	12	11.75	0.904
	Kedisan Kelod	19	10	12	12.96	0.469
	N Batur	18	10	14	27.05	0.812
	Pujung Kaja	20	10	10	7.29	0.409
	S Batur	**24**	**10**	**14**	**13.19**	**0.996**
	Sebatu	38	10	16	9.88	0.803
	Subak Bayad	20	10	12	11.75	0.787
	Subak Bonjaka	21	10	12	10.80	0.728
	Subak Jasan	23	10	12	9.41	0.582
	Subak Jati	20	10	10	7.29	0.637
	Subak Pakudui	19	10	12	12.96	0.406
	Subak Tegal Suci	22	10	9	5.17	0.225
	Sungi	20	10	15	25.59	0.955
	Timbul	20	10	12	11.75	0.787
	Tungkub	20	10	18	82.23	0.864
	Yeh Tampuagan	54	12	30	27.00	0.819
Borneo	Long Gi	34	12	21	22.44	0.865
	Long Soluy	26	12	8	3.56	0.455
Flores	Bama	49	12	19	10.91	0.377
	Bena	16	12	10	10.41	0.778
	Boawae	**27**	**12**	**15**	**13.10**	**0.979**
	Cibol	52	12	10	3.41	0.974
	Seso Borowa	30	12	20	25.06	0.638
	Wogo	35	12	19	16.23	0.415
	Woloara	29	12	20	27.20	0.747
	Wolotopa	46	12	27	26.53	0.494
Java	Dieng	35	12	16	10.81	0.679
Nias	Gomo	47	12	14	6.38	0.926
Sumba	Anakalang	51	12	17	8.52	0.954
	Kodi	**44**	**12**	**13**	**5.86**	**0.999**
	Lamboya	**49**	**12**	**14**	**6.19**	**0.988**
	Loli	37	12	19	14.98	0.531
	Mamboro	53	12	20	11.23	0.656
	Rindi	27	12	20	33.31	0.952
	Wanokaka	**52**	**12**	**22**	**13.86**	**0.997**
	Wunga	37	12	12	5.77	0.093

Communities in which the test of neutrality was rejected (P_E > 0.975) are highlighted in blue.

Table S2. Power of exact test under different model parameters*

Power for different parameters

N = 100 n = 20 m = 0.0208 10000 samples

m = 0

% dom	s = 1	s = 2	s = 3	s = 4
0	0.06	0.06	0.05	0.06
2	0.05	0.06	0.13	0.35
4	0.05	0.32	0.43	0.47
6	0.18	0.37	0.42	0.43
8	0.27	0.35	0.40	0.45
10	0.28	0.36	0.36	0.39

m = 0.01

0	0.05	0.05	0.05	0.05
2	0.05	0.05	0.17	0.08
4	0.09	0.29	0.35	0.41
6	0.07	0.32	0.36	0.40
8	0.24	0.32	0.32	0.36
10	0.22	0.30	0.28	0.31

m = 0.03

0	0.05	0.05	0.05	0.04
2	0.04	0.04	0.11	0.16
4	0.04	0.26	0.34	0.39
6	0.06	0.27	0.34	0.35
8	0.10	0.29	0.30	0.29
10	0.23	0.26	0.26	0.24

Power Cutoff	Prob ≤ 5/41
0.19	0.20
0.22	0.10
0.24	0.05
0.29	0.01

The population and samples sizes were chosen to correspond with samples from Bali, where in general there were fewer samples and the villages are smaller. Power of the test with moderate and high migration is indicated by the values under the headings *m* = 0.01 and *m* = 0.03. The orange colors correspond to the cutoff values for different levels of power needed to make seeing 5/41 or fewer non-neutral villages unlikely at different levels (Figure 3).

Table S3. Power of exact test under different model parameters

Power for different parameters

N = 300 n = 35 m = 0.0249 10000 samples

m = 0

% dom	s = 1	s = 2	s = 3	s = 4
0	0.04	0.04	0.05	0.04
2	0.23	0.51	0.61	0.64
4	0.39	0.52	0.54	0.52
6	0.41	0.42	0.40	0.37
8	0.40	0.38	0.28	0.25
10	0.33	0.28	0.23	0.19

m = 0.01

0	0.03	0.04	0.04	0.04
2	0.21	0.47	0.64	0.65
4	0.36	0.54	0.52	0.54
6	0.41	0.47	0.42	0.37
8	0.35	0.40	0.32	0.28
10	0.30	0.32	0.24	0.18

m = 0.03

0	0.04	0.03	0.04	0.04
2	0.17	0.40	0.54	0.61
4	0.28	0.50	0.56	0.57
6	0.30	0.44	0.41	0.40
8	0.30	0.35	0.31	0.30
10	0.27	0.28	0.25	0.22

Power Cutoff	Prob ≤ 5/41
0.19	0.20
0.22	0.10
0.24	0.05
0.29	0.01

The population and samples sizes were chosen to correspond with the samples where in general we have more samples and the villages are larger. Power of the test with moderate and high migration is indicated by the values under the headings $m = 0.01$ and $m = 0.03$. The orange colors correspond to the cutoff values for different levels of power needed to make seeing 5/41 or fewer non-neutral villages unlikely at different levels (Figure 3).

Table S4. Sites, sample size n, estimates of N_e, number of households and source of data for household numbers

Island	Site	*n*	Estimated N_e	# of households	Source
Bali	Abian Kebon	37	300.87	223	1
	Bena	20	615.24	190	1
	Calo	21	413.16	68	1
	Gadon	17	833.69	545	1
	Kebon	20	106.73	70	1
	Kedisan Kaja	20	282.38	93	1
	Kedisan Kelod	19	311.50	89	1
	N Batur	18	650.34	431	1
	Pujung Kaja	20	175.24	140	1
	S Batur	**24**	**317.02**	**2027**	**1**
	Sebatu	38	237.58	216	1
	Subak Bayad	20	282.38	80	1
	Subak Bonjaka	21	259.61	70	1
	Subak Jasan	23	226.19	158	1
	Subak Jati	20	175.24	53	1
	Subak Pakudui	19	311.50	65	1
	Subak Tegal Suci	22	124.32	69	1
	Sungi	20	615.24	230	1
	Timbul	20	282.38	210	1
	Tungkub	20	1976.62	617	1
	Yeh Tampuagan	54	542.23	256	1
Borneo	Long Gi	34	450.51		NA
	Long Soluy	26	71.52		NA
Flores	Bama	49	219.13	169	1
	Bena	16	208.96		NA
	Boawae	**27**	**263.14**	**734**	**1**
	Cibol	52	68.47	75	1
	Seso Borowa	30	503.12	299	1
	Wogo	35	325.82	400	2
	Woloara	29	546.24	240	3
	Wolotopa	46	532.64	227	1
Java	Dieng	35	217.09	170	1
Nias	Gomo	47	128.02		NA
Sumba	Anakalang	51	171.08		NA
	Kodi	**44**	**117.64**	**450**	**2**
	Lamboya	**49**	**124.30**	**352**	**2**
	Loli	37	300.87	143	1
	Mamboro	53	225.42	320	1
	Rindi	27	668.91	237	4
	Wanokaka	**52**	**278.27**		**NA**
	Wunga	37	115.78	93	2

Orange indicates sampling locations where N_e < # households. Bold sites tested as non-neutral. Key: 1, village records (male household heads); 2, village informant estimate; 3, regression fit; 4, Government Statistics Office, East Sumba 2003. The regression fit estimated the number of household heads from the number of males in the population based on the regression line for other sites that had both pieces of information.

Table S5. Haplogroup counts for all sites

Island	Site	*n*	# HG	C-M216*	C-M38*	F-P14*	H-M69*	J-12f2*	J-M172*	J-M12	K-M9*	L-M20	M-M5*	M-P34	O-M122*	O-L1Ya	O-M134	O-M119*	O-M110	O-M95*	Q-M346	R-M207*	R-M124	R-M17	R-P25*	R-M269	S-M230
Bali	Abian Kebon	37	3												1			22	14							
	Bena	20	6									2						3	9		1	3				2
	Calo	21	4									3			3			4	11							
	Gadon	17	6							1					3			1	10				1			
	Kebon	20	1																20							
	Kedisan Kaja	20	3	1														6	13							
	Kedisan Kelod	19	3	3														5	11							
	N Batur	18	4								1							1	14							2
	Pujung Kaja	20	6	2								5						1	8	1						
	S Batur	24	2															1	23							
	Sebatu	38	4				13											2	17							
	Subak Bayad	20	5												6			4	12				1			
	Subak Bonjaka	21	6	2								1				1		4	12	1						
	Subak Jasan	23	1																23							
	Subak Jati	20	3				3											3	14							
	Subak Pakudui	19	6				1				4				2				9				1	2		
	Subak Tegal Suci	22	3												6			1	15							
	Sungi	20	5					1							2	1		6	10							
	Timbul	20	3														1	13	6							
	Tungkub	20	6	1						1		1						6	9			2				
	Yeh Tampuagan	54	7												8	1	1	10	27			4				
Borneo	Long Gi	34	8	3		2									5	13		1	8				1			
	Long Soluy	26	4	11							4				9				2							
Flores	Bama	49	6	3	35						8	1						1	1							
	Bena	16	4		4							2	9					1								
	Boawae	27	6		2						13	1	7						1							3
	Cibol	52	4	48	1										2											1
	Seso Borowa	30	6	1	5						1		8					4	8							7
	Wogo	35	5	1	1								25													
	Woloara	29	7	1	3	1					1		6					15								2
	Wolotopa	46	7		3	5					1	1	6					16	21							14
Java	Dieng	35	5						3					1				4					6			
Nias	Gomo	47	1															47								
Sumba	Anakalang	51	5		13						8							1	21							8
	Kodi	44	4		38										1			1								4
	Lamboya	49	4		46										1			1								1
	Loli	37	5		27								5		1			2								2
	Mamboro	53	7		28						2		8		1			5	2							7
	Rindi	27	8	1	6						1		2		2			5		1						9
	Wanokaka	52	4		39						1							1								11
	Wunga	37	6		3						2		2		4			6								20
Total		1269	24	78	254	8	17	1	3	2	47	12	9	82	58	19	1	204	28	329	2	1	9	10	2	2	91