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

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

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 ($\theta = 2N_e\mu$).

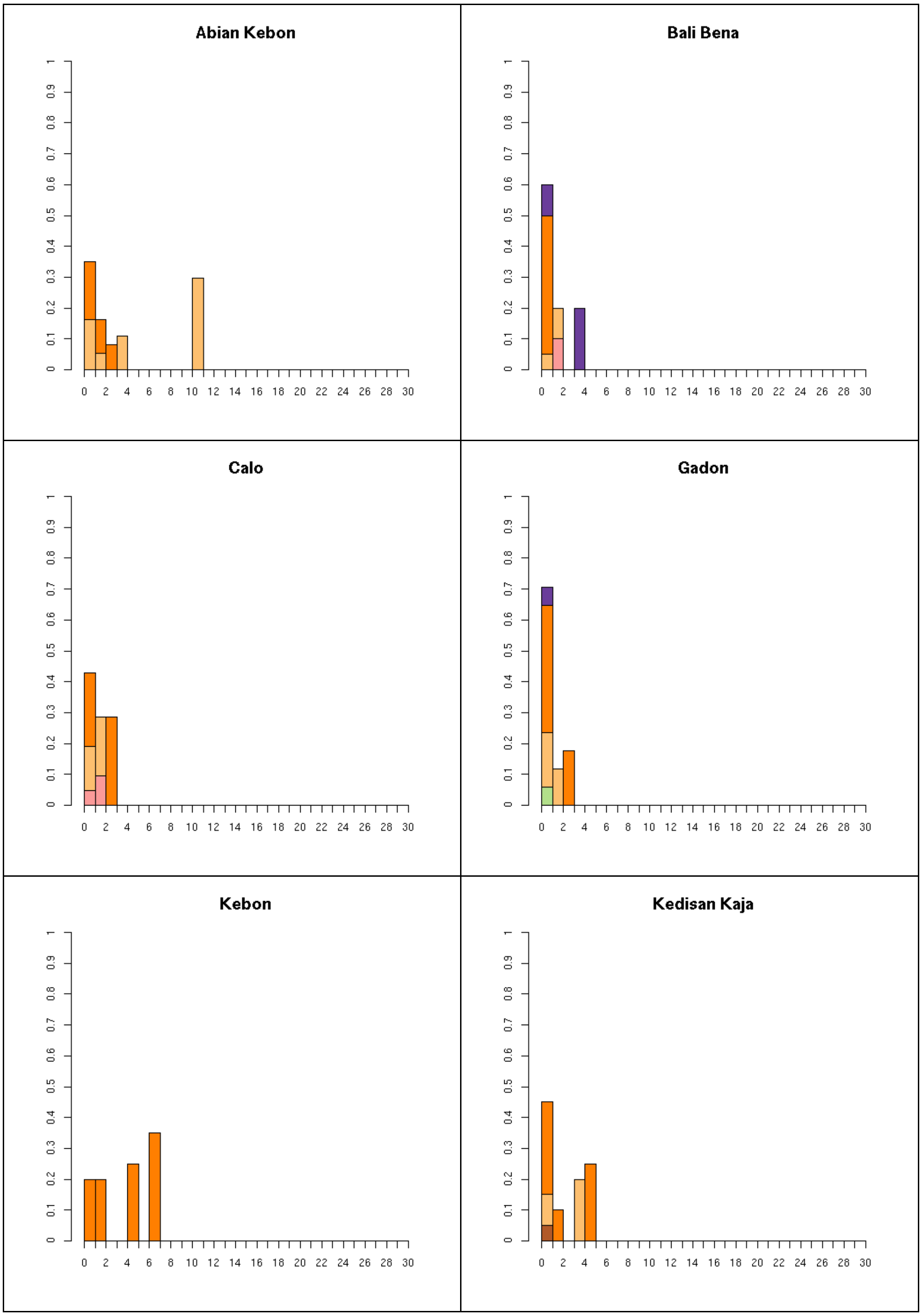
The case $p = 1$ falls under the general exchangeable models of Cannings (ref. 3 in text). Thus, samples under equilibrium will, under the same assumptions as before, satisfy the Ewens sampling formula and appear neutral under the Ewens exact test. In this case, however, the inbreeding effective population 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.

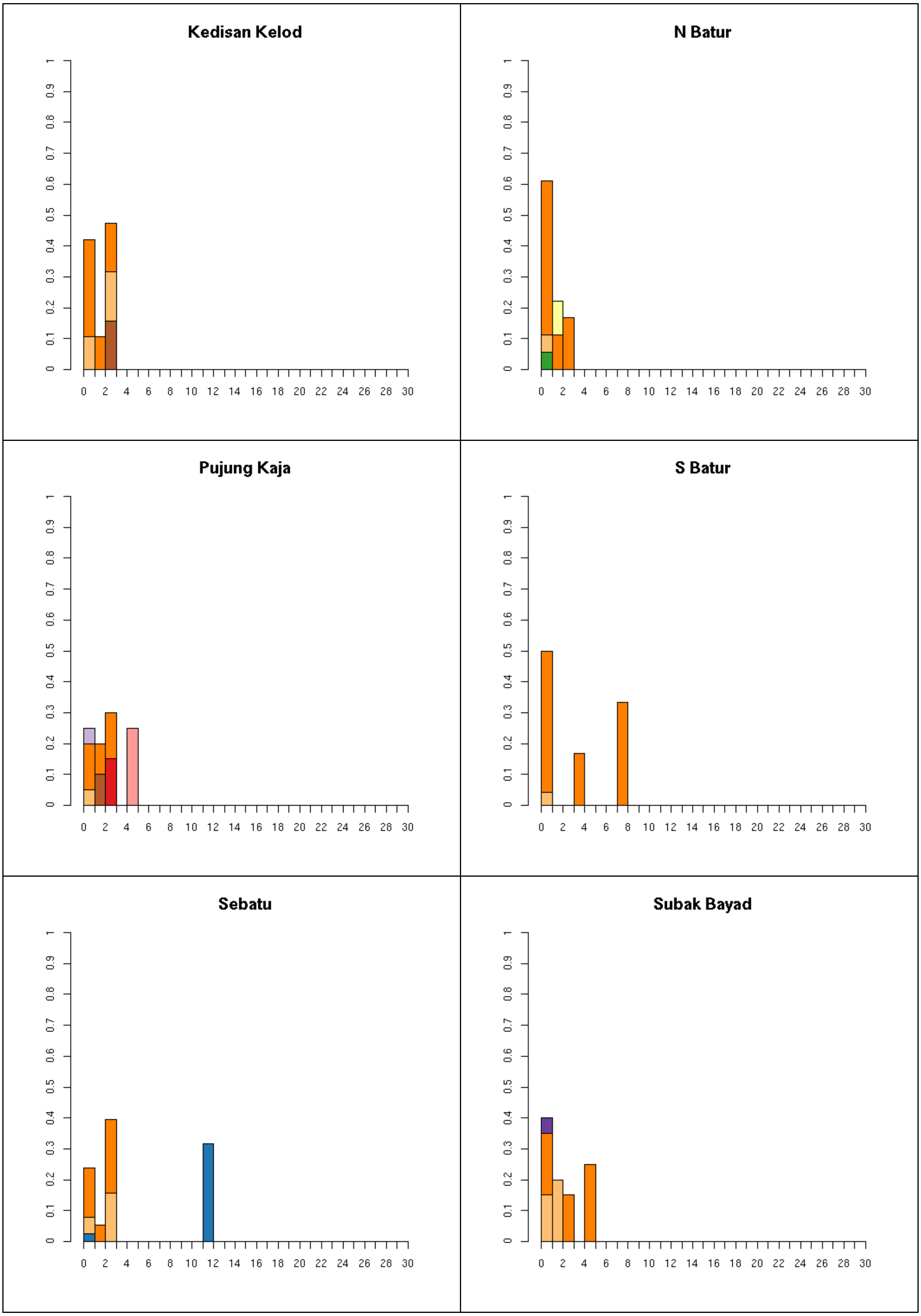
All simulations were performed in MATLAB, and copies of the code are available on request.

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.

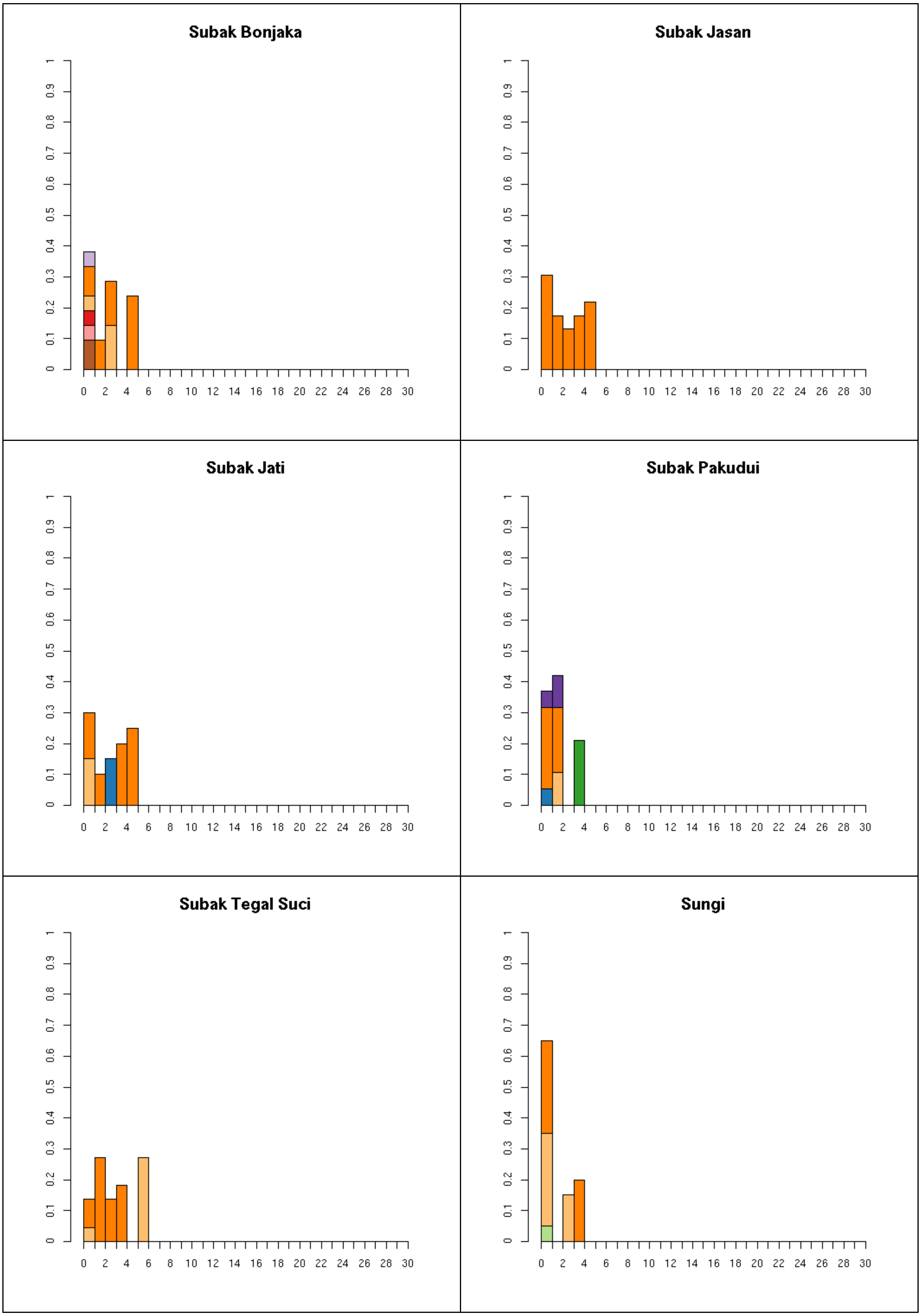


Haplotype distributions for sites on Bali

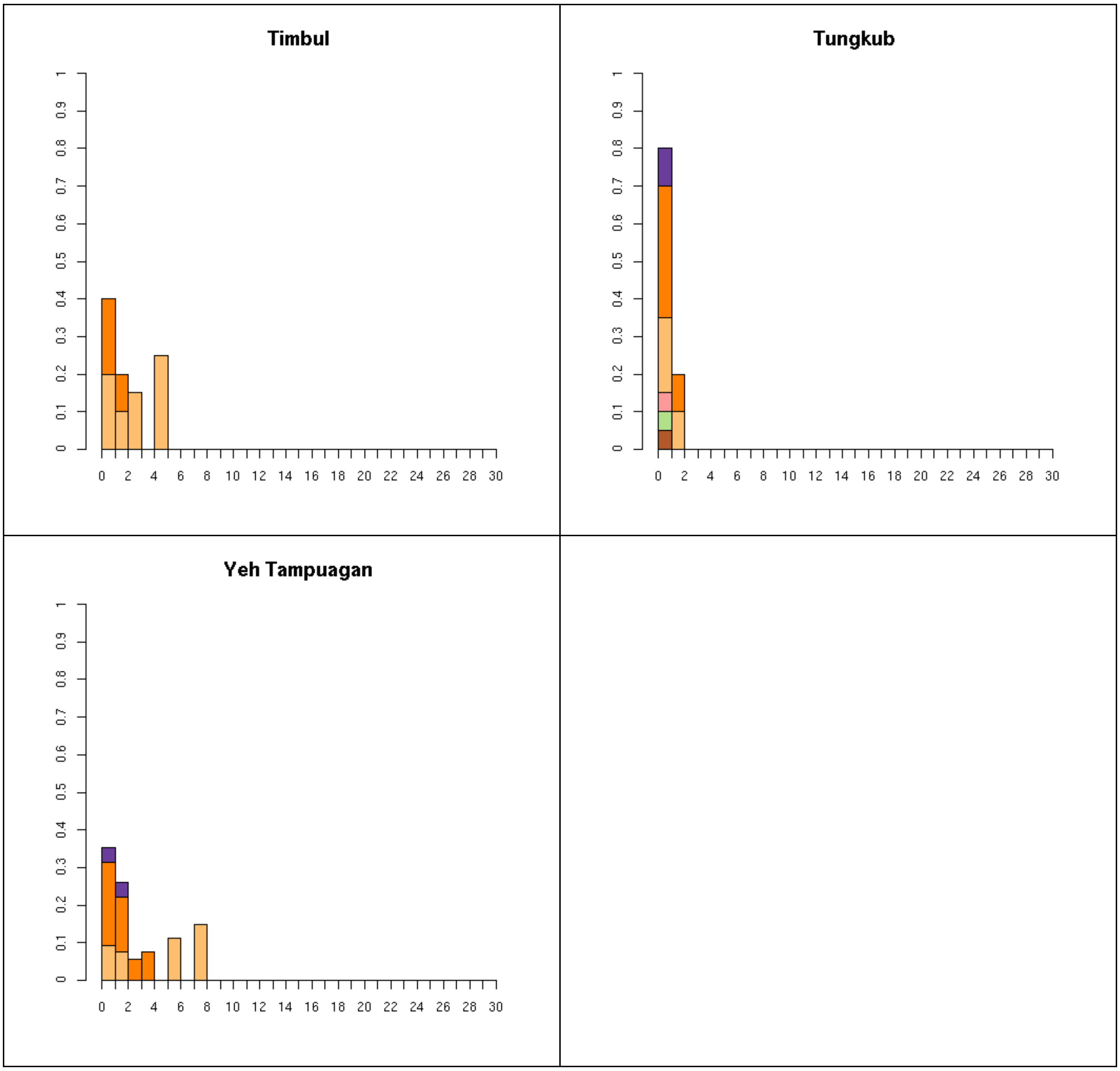
Fig. S1. All 41 haplotype distributions, normalized by number of samples for comparison. Colors represent haplogroup membership designated by the following colors: brown, C; light blue, F; dark blue, H; light green, J, dark green, K; pink, L; red, M; orange, O; dark orange, O-M95; light purple, Q; dark 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.



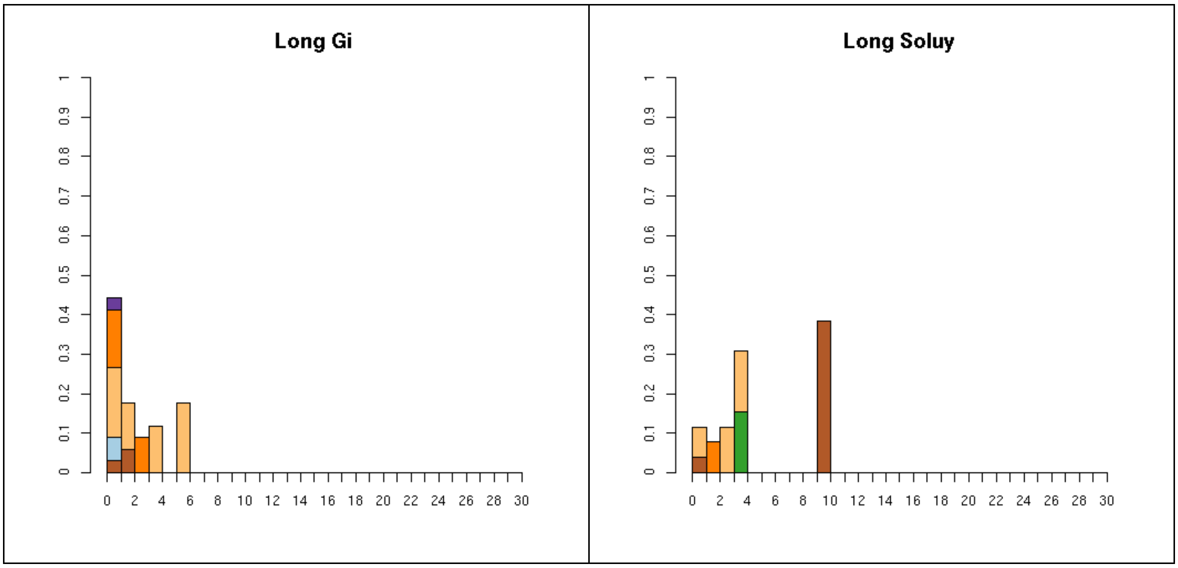
Haplotype distributions for sites on Bali



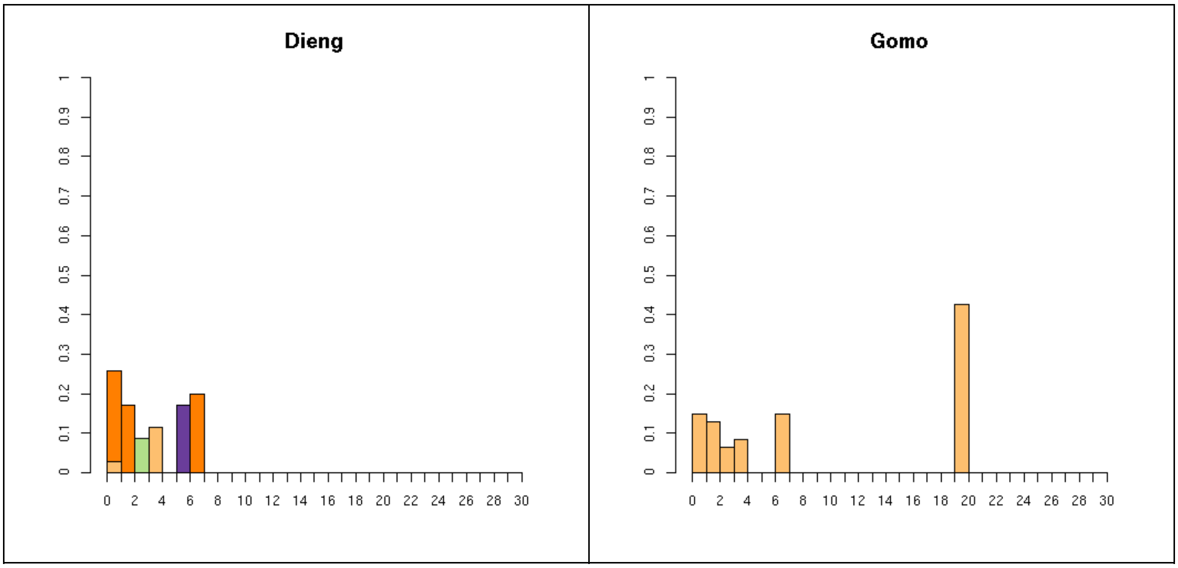
Haplotype distributions for sites on Bali



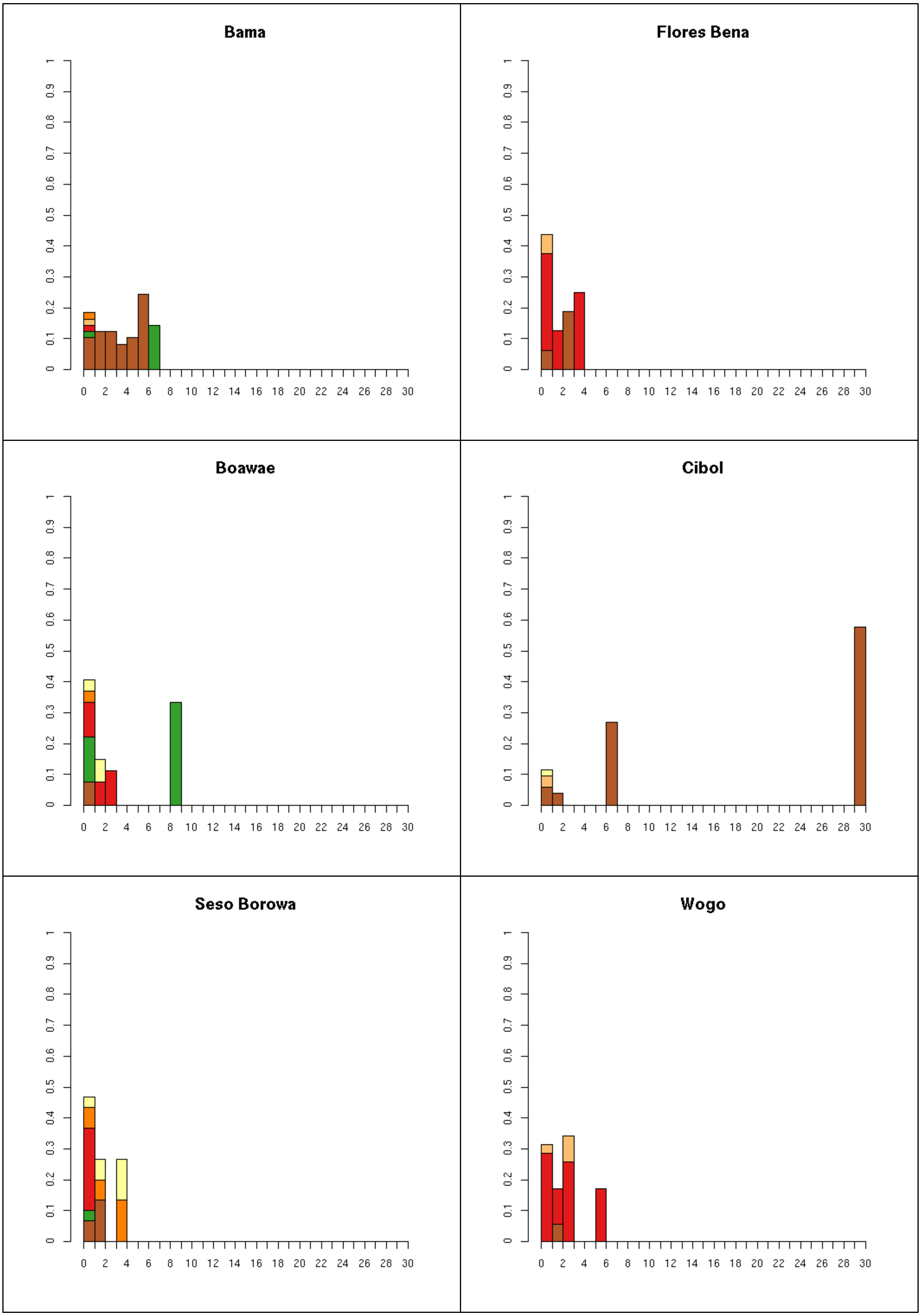
Haplotype distributions for sites on Bali



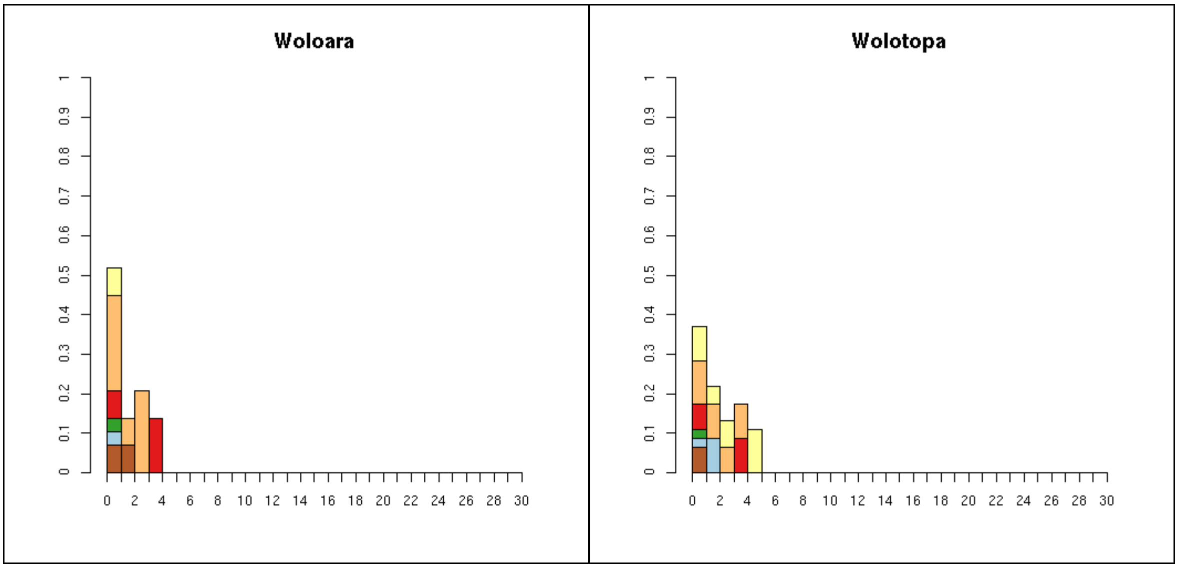
Haplotype distributions for sites on Borneo



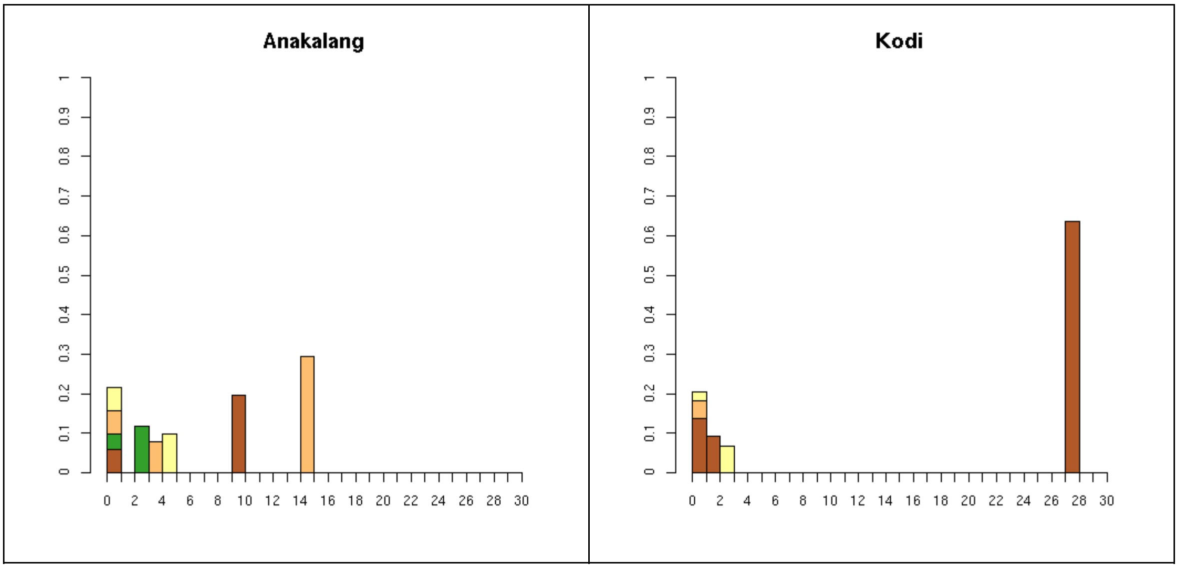
Haplotype distributions for sites on Java and Nias



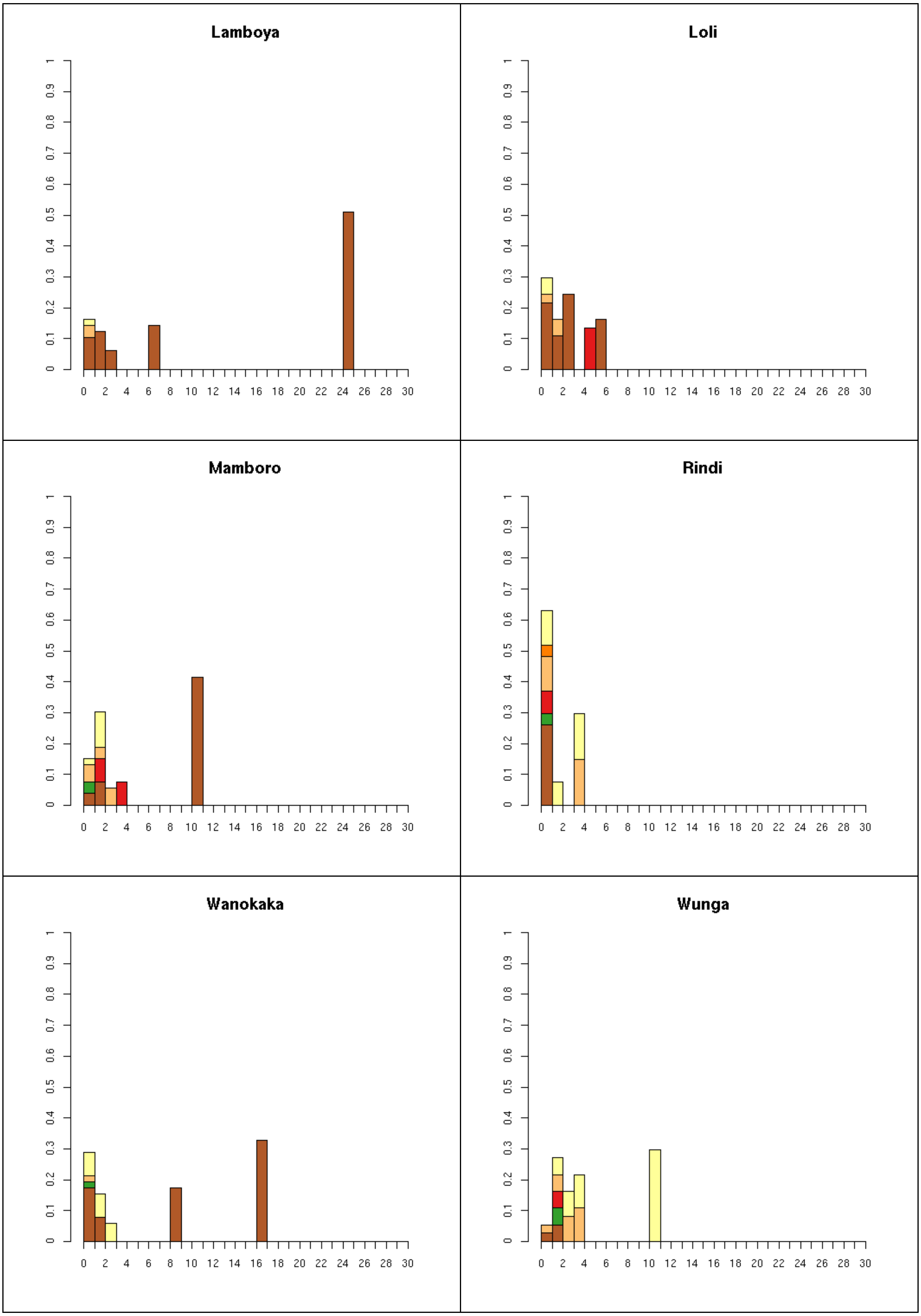
Haplotype distributions for sites on Flores



Haplotype distributions for sites on Flores



Haplotype distributions for sites on Sumba



Haplotype distributions for sites on Sumba

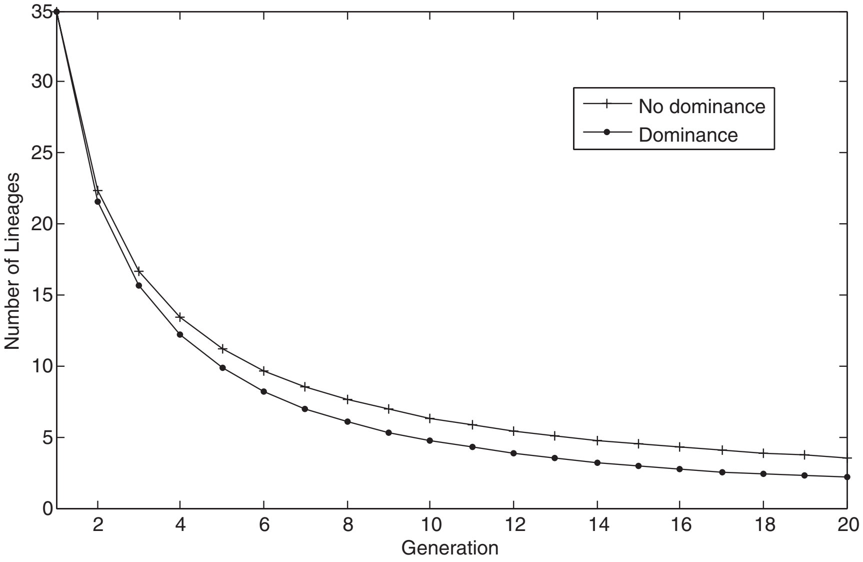


Fig. S2. Reduction of the number of lineages during a bottleneck of size 35. The graph shows the average number of lineages each generation for 5000 simulated populations. For the dominance case, the parameters are $\delta = 0.06$ and $\sigma = 2$.

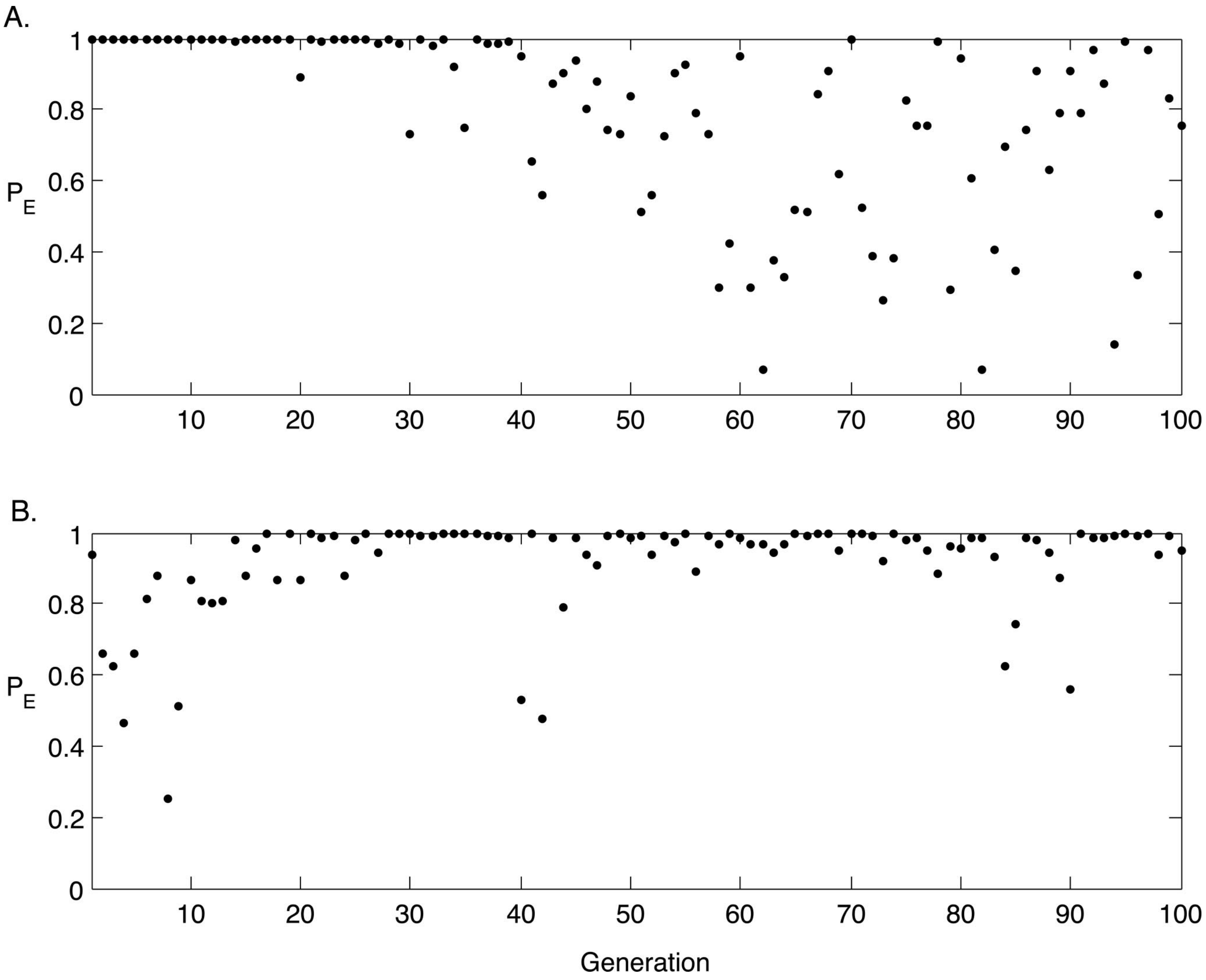


Fig. S3. Examples of performance of exact test after recovery from bottleneck/founder effect. In both simulations the initial population size is 35 and the population grows to a final size of 300 over ≈25 generations (≈750 years). The graphs show the value of P_E from the exact test for each generation. (*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).

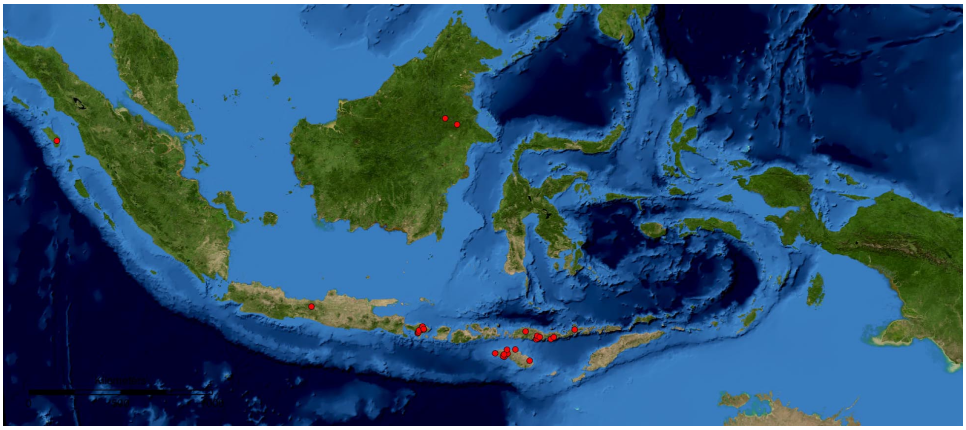


Fig. S4. Map showing community sampling locations (*red*) in Indonesia. Sampled islands (west to east) are Nias, Java, Bali, Borneo, Sumba and Flores.

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	ş = 1	s = 2	ş = 3	ş = 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	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		3				1		8	1						
	S Batur	24	2															1		23							
	Sebatu	38	4				13								6			2		17							
	Subak Bayad	20	5												1	2		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	3	27			4				
Borneo	Long Gi	34	8	3		2									5	13		1	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						8							7
	Wogo	35	5	1	1								4	25				4									
	Woloara	29	7	1	3	1					1			6				15									2
	Wolotopa	46	7		3	5					1		1	6				16									14
Java	Dieng	35	5						3							1		4		21				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