Genome-wide Analysis Indicates More Asian than Melanesian Ancestry of Polynesians

Manfred Kayser,[1,8,*] Oscar Lao,[1,8] Kathrin Saar,[2] Silke Brauer,[1,3] Xingyu Wang,[4] Peter Nürnberg,[5] Ronald J. Trent,[6] and Mark Stoneking[7]

Analyses of mitochondrial DNA (mtDNA) and nonrecombining Y chromosome (NRY) variation in the same populations are sometimes concordant but sometimes discordant. Perhaps the most dramatic example known of the latter concerns Polynesians, in which about 94% of Polynesian mtDNAs are of East Asian origin, while about 66% of Polynesian Y chromosomes are of Melanesian origin. Here we analyze on a genome-wide scale, to our knowledge for the first time, the origins of the autosomal gene pool of Polynesians by screening 377 autosomal short tandem repeat (STR) loci in 47 Pacific Islanders and compare the results with those obtained from 44 Chinese and 24 individuals from Papua New Guinea. Our data indicate that on average about 79% of the Polynesian autosomal gene pool is of East Asian origin and 21% is of Melanesian origin. The genetic data thus suggest a dual origin of Polynesians with a high East Asian but also considerable Melanesian component, reflecting sex-biased admixture in Polynesian history in agreement with the Slow Boat model. More generally, these results also demonstrate that conclusions based solely on uniparental markers, which are frequently used in population history studies, may not accurately reflect the history of the autosomal gene pool of a population.

Previous analyses of mtDNA and NRY variation in Polynesians have demonstrated a dramatic discrepancy in the origins of the Polynesian mtDNA and NRY gene pools:[1–4] about 94% of Polynesian mtDNAs are of East Asian origin and only 6% are of Melanesian origin (consisting of mainland New Guinea and surrounding islands, also known as Near Oceania), whereas 66% of Polynesian Y chromosomes are of Melanesian origin and only 28% are of East Asian origin.[4] To explain this discrepancy, the Slow Boat hypothesis for Polynesian origins was formulated, according to which the ancestors of Polynesians migrated from East Asia/Taiwan through New Guinea and Island Melanesia, mixing extensively with the local populations as they did so, before continuing eastward to colonize Remote Oceania/Polynesia.[2,4] This mixing was sex biased, involving many more Melanesian males than females admixing with East Asian newcomers, consistent with the view that ancestral Polynesian society may have been matrilocal in residence (men move to their wife's land) and matrilineal in descent (clans are inherited through the mothers line).[4–6] This large discrepancy between the mtDNA and NRY genetic contributions of East Asian and Melanesian ancestors to Polynesians raises the question as to the constitution of the autosomal gene pool of Polynesians. To address this question, we genotyped 377 genome-wide distributed autosomal short tandem repeat (STR) loci in 47 Polynesians (10 Cook Islanders, 10 Tongans, 18 Samoans, 5 Tokelau Islanders, and 4 Nuie Islanders), 44 Han Chinese from Beijing, and 24 Papua New Guineans from the interior highlands (15 from the Eastern Highlands, 9 from the Southern Highlands). Genotyping was performed essentially as described elsewhere.[7] Additional STR loci were chosen to reduce the average distance between markers to 8.8 cM; the genotype data are available from the authors.

Some basic statistics of genetic diversity in these three groups, based on the 377 STR loci, are presented in Table 1 and indicate that diversity in the Polynesians is intermediate between that of Chinese and New Guineans. Previous evidence from NRY and mtDNA data indicates that Polynesians have NRY and mtDNA haplogroups of both East Asian and Melanesian origins.[1–4,8–12] We therefore hypothesized that East Asians and Melanesians represent the parental populations of Polynesians, and we estimated the genome-wide autosomal proportion of the East Asian (represented by the Chinese samples) and Melanesian (represented by the PNG samples) contribution to Polynesians. However, the estimation of the East Asian and Melanesian components in the current Polynesian autosomal gene pool is potentially complicated by the severe bottlenecks that accompanied the migrations to and through the Pacific (Remote Oceania), as well as the relatively long period since the admixture occurred, approximately 3500 years ago.[13]

Several methods have been proposed for estimating the amount of admixture of a hybrid population,[14] of which two seem to be more robust to departures from model assumptions. These are the maximum likelihood method proposed by Wang[15] as implemented in the program LEADMIX and the coalescent approach proposed by

[1]Department of Forensic Molecular Biology, Erasmus University Medical Center Rotterdam, PO Box 2040, 3000 CA Rotterdam, The Netherlands; [2]Gene Mapping Center, Max Delbrück Center for Molecular Medicine (MDC), Robert-Rössle-Strasse 10, 13092 Berlin, Germany; [3]Department of Biology, Netherlands Forensic Institute, PO Box 24044, 2490 AA Den Haag, The Netherlands; [4]Beijing Hypertension League Institute, Fu Wai Hospital, 167 Beilishi, 100037 Beijing, China; [5]Cologne Center for Genomics, University of Cologne, Zuelpicher Strasse 47, 50674 Köln, Germany; [6]Department of Molecular and Clinical Genetics, Royal Prince Alfred Hospital and Central Clinical School, The University of Sydney, Missenden Road, Camperdown, NSW 2050, Australia; [7]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany
[8]These authors contributed equally to this work.
*Correspondence: m.kayser@erasmusmc.nl
DOI 10.1016/j.ajhg.2007.09.010.

Table 1. Genetic Diversity in the Chinese, New Guinean, and Polynesian Samples, Based on 377 Autosomal STR Loci

Population	No. of Individuals	Average Gene Diversity over Loci (±1 SD)	Average Number of Alleles per Locus	Average Number of Nonshared Alleles per Locus	Mean Number of Pair-wise Differences (±1 SD)
Chinese	44	0.73 (0.35)	7.72	1.26	274.62 (118.53)
New Guineans	24	0.68 (0.33)	6.07	0.44	256.16 (111.48)
Polynesians	47	0.70 (0.33)	6.86	0.53	264.29 (114.01)

Bertorelle and Excoffier.[16] The latter approach has been recently incorporated into an Approximate Bayesian Computation (ABC) framework,[17] which seems to be especially suited for STR data.[18] The model underlying these two methods assumes that an ancestral population splits into the two parental populations, which diverge for some period of time and then meet once to form the admixed population (Figure 1).

We performed LEADMIX and ABC simulations on the total data set from 377 genome-wide autosomal STR loci. The ABC program was kindly provided to us by L. Excoffier. In the case of LEADMIX, we used the default parameters suggested previously;[15] results are presented in Table 2. For the ABC approach, we used the same prior distributions and summary statistics as suggested previously[18] with the exception of the statistic that computes the average extent of linkage disequilibrium D′ between independent markers in the admixed population. Because the time for computing this statistic increases to square with the number of loci, it turned out to considerably increase the computational time of each simulation with our current data set of 377 STRs. The proposed statistics try to capture different aspects of the demographic processes shaping the genetic diversity of the populations, including the presence of bottlenecks (i.e., the M statistics[19]) and the amount of admixture between the two parental populations (i.e., the my statistic). We performed 500,000 simulations with the program SIMCOAL 2.0.[20] The ABC program[18] was then used to estimate the parameters by means of local and weighted linear regression on the best 1000 simulations after transformation of the data.[21] We increased the number of simulations suggested by Excoffier et al.[18] by factor 5 to achieve reliable results given the relatively large number of genetic markers used and the expected complexity of the demographic history of the populations analyzed. Results of the ABC approach are presented in Table 3. An additional analysis considering 100,000 simulations with ascertaining the best 1000 simulations produced similar results although with larger credible intervals (data not shown).

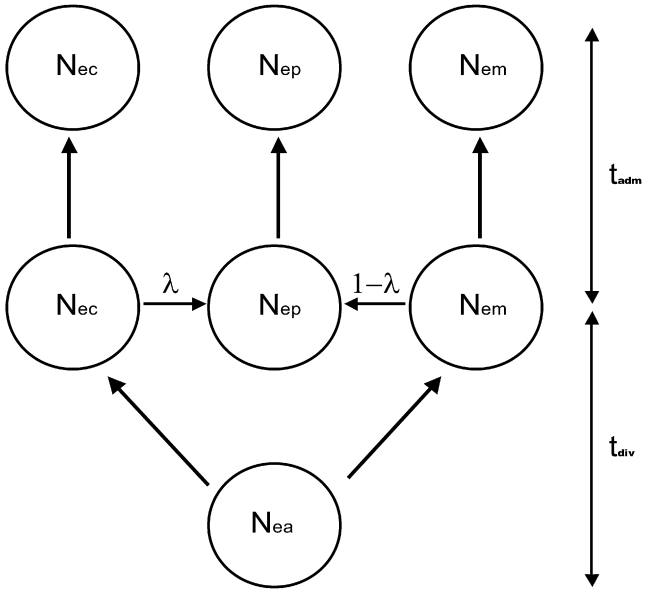

Figure 1. Model of Admixture Assumed in This Study, as Proposed Elsewhere[16]

In this model, the Polynesian population is considered to be the result of a single admixture event between the Asian and Melanesian parental populations, and after this event no further migration is observed between the three populations. Ne, effective population size; λ, the amount of parental ancestry of the Chinese population. The "c" index indicates the Chinese population, the "p" index the Polynesian population, the "m" index the Melanesian population, and the "a" index the ancestral population.

The resulting admixture estimates from the two methods were highly concordant (Tables 2 and 3). Also, the posterior distribution of the amount of admixture as estimated with the ABC approach showed a clear improvement of the uncertainty about the parameters when compared with the uniform prior distribution (see Figure 2). Both methods indicate that 79% of the autosomal gene pool of the Polynesian sample is of East Asian origin and about 21% is of Melanesian origin, which is more similar

Table 2. Maximum Likelihood Estimations of the Parameters of Interest from LEADMIX

Parameter	Mode	+95%CI	−95%CI
t1	0.029	0.033	0.025
t2	0.086	0.094	0.077
T1	0.004	0.006	0.002
T2	0.002	0.008	<0.0001
TH	0.031	0.032	0.0230
λ	0.792	0.837	0.761

t1 is the time scaled by the effective population size (Ne) of the Chinese population after the split of the ancestral population, t2 is the time scaled by Ne of the Melanesian population after the split of the ancestral population, T1 is the time scaled by Ne of the Chinese population after the admixture, T2 is the time scaled by Ne of the Melanesian population after the admixture, TH is the time scaled by Ne of the Polynesian population, and λ is the proportion of admixture of the Chinese population in the Polynesian population (last row).

Table 3. Results from the ABC Simulations

Parameter	Mode	+95%CI	−95%CI
t1	0.89168814	1.50137272	0.43075782
t2	22.65190747	36.38889742	12.1004948
T1	0.00073835	0.02132547	0.00032565
T2	0.51937706	6.64472347	0.1695372
TH	6.93398065	8.19012123	0.43595085
λ	0.791	0.977	0.274

Abbreviations as in Table 2.

to the mtDNA estimates than to the NRY estimates of admixture in these very same samples (Figure 3).

Our admixture estimates could be distorted by European admixture, which is prevalent in some Polynesian groups[22] but is much less in the sample set from which the 47 Polynesians used here were ascertained.[4] We therefore had selected Polynesian samples for autosomal STR genotyping that do not carry European mtDNA or NRY haplogroups and whose self-described ancestry does not include any Europeans for at least two generations. Nonetheless, we cannot exclude some small amount of European admixture in these samples.

The admixture estimates also depend critically on the choice of "parental" populations. Linguistic evidence strongly suggests that Taiwan was the ancestral homeland of the proto-Austronesians[23] who ultimately colonized Polynesia, which in turn suggests that Taiwan Aborigines may be a more appropriate "Asian parental" population than Han Chinese. However, mtDNA and NRY evidence suggests further bottlenecks in aboriginal Taiwanese[1,24] and that Han Chinese and other East Asian groups are highly similar to one another and are as similar (or even more similar) to Polynesians than are aboriginal Taiwanese.[4] Thus, the use of a different East Asian group is unlikely to change the admixture estimate. With regard to the "Melanesian parental" population, the admixture between Polynesian ancestors and Melanesians took most likely place somewhere in coastal/island New Guinea (probably the Bismarck Archipelago[9,10]) because Austronesians arrived in Melanesia by boat, so it may be argued that a coastal/island New Guinea population would be more appropriate than a highland New Guinea population as the parental Melanesian population. However, coastal/island New Guinea populations usually exhibit some proportion of Asian mtDNA and/or NRY types[4,9] as result of genetic admixture, whereas non-Austronesian-speaking highland New Guinea groups usually lack Asian-specific markers resulting from a lack of such admixture.[4,25] Therefore, Highland New Guineans provide the best available estimate of a nonadmixed Melanesian population.

To investigate any possible influence of selection on these admixture estimates obtained from the autosomal STR data, we repeated the simulations for a subset of 107 putatively neutral loci that were selected to be at least 100 kb away from any known genes. The admixture and other demographic parameter estimates did not differ significantly between this subset of putatively neutral loci and the full data set (results not shown), indicating that selection on genes is not influencing the admixture estimates obtained here.

We also attempted to obtain individual-based estimates of admixture by using the program STRUCTURE,[26] but the output results were highly sensitive to the model assumptions in that different results were obtained in different runs of the program. We speculate that this failure to obtain consistent results with STRUCTURE could be due to the high mutation rate of the STR loci and/or the impact of bottlenecks on the autosomal allele frequencies. This is supported by the fact that the parental populations and the Polynesians carry a relatively large number of nonshared alleles per locus (Table 1), as a result of new mutations and/or alleles lost because of drift.

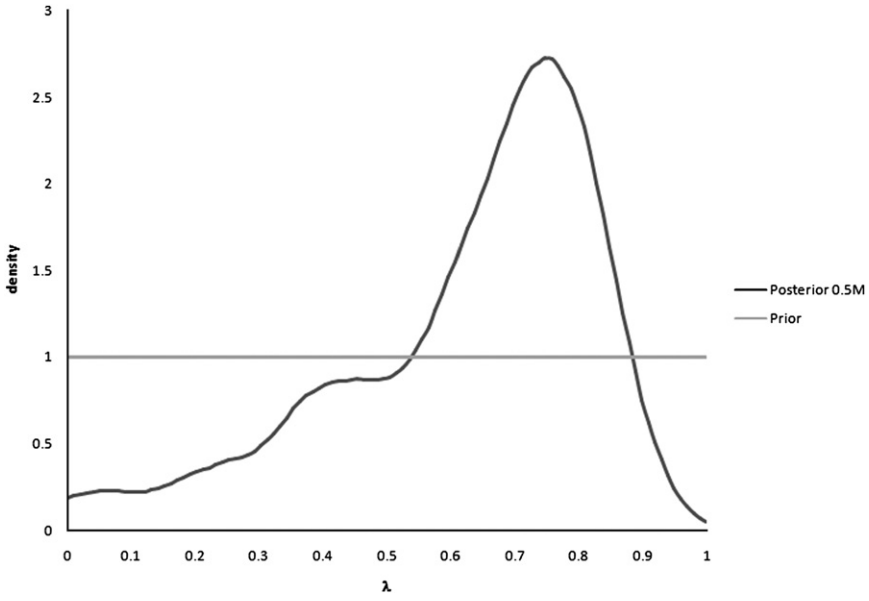

Figure 2. Prior and Posterior Distribution of the Amount of Asian Admixture (Expressed as λ) in Polynesians.

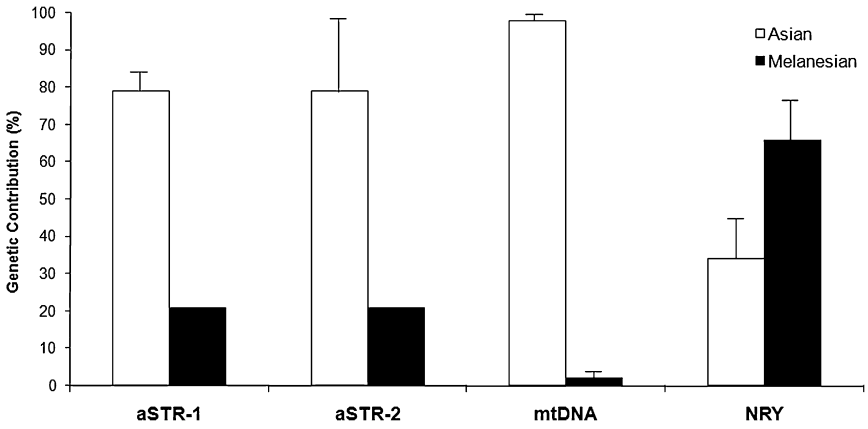

Figure 3. Estimated Proportion of Asian and Melanesian Ancestry in 47 Polynesians

Based on the 377 autosomal STR loci (aSTR-1, based on LEADMIX; aSTR-2, based on the ABC approach). The estimated contribution of Asian and Melanesian mtDNA and NRY haplogroups for this sample of 47 Polynesians does not differ significantly from a larger sample of 321 Polynesians[4] (data not shown). For the mtDNA and NRY data, standard errors were calculated from frequency estimates of haplogroups of Asian and Melanasian origin among those Polynesian individuals also analyzed for the aSTRs. For the aSTR data, standard errors were estimated from the 95% CI values of the amount of Asian admixture (expressed as λ) in the Polynesian individuals analyzed by the two simulation approaches. Only the positive errors are shown because they are symmetrical.

Overall, the genetic data suggest a dual origin of Polynesians, with a mostly East Asian but also considerable Melanesian genetic component. This is in keeping with models of Polynesian origins, such as the Triple-I[27] and Slow Boat[2,4] models, which propose substantial admixture between Polynesian ancestors and Melanesians before the colonization of the Pacific (Remote Oceania). Models that do not allow for both a primarily East Asian source of Polynesians as well as considerable Melanesian genetic input, such as the Express Train[28] and Entangled Bank[29] models, can be rejected based on the genetic evidence. Models that propose an older origin of Polynesians, such as the Slow Boat model of Oppenheimer and Richards,[30] which postulates an origin in island Southeast Asia some 17,000 years ago, are also not supported by the genetic evidence.

In addition and more generally, our results suggest that single-locus studies of population histories (in particular, mtDNA and/or the Y chromosome) should be interpreted cautiously, because they may not be representative of the entire genome. In particular, in the present case, estimates of the Melanesian contribution to Polynesians based on NRY data differ dramatically from estimates based on genome-wide STR data (Figure 3) and are thus not representative of the total (autosomal) Polynesian gene pool. However, although the autosomal STR data provide more reliable admixture estimates than can be obtained from mtDNA or NRY data (because they are based on many loci and are representative of the entire genome), the mtDNA and NRY data nonetheless provide insights into the admixture process that the autosomal data cannot provide. In particular, the large discrepancy in the estimated Asian and Melanesian contributions to Polynesians for mtDNA versus the Y chromosome suggests sex-biased genetic admixture during Polynesian history. This admixture mostly involved Melanesian men (as evidenced by the high proportion of Melanesian Y chromosomes in Polynesia) and Asian women (as evidenced by the high proportion of Asian mtDNAs in Polynesia), an interesting finding that cannot be observed from the autosomal data. This scenario is supported by suggestions of matrilineal descent and matrilocal residence in the ancient Polynesian society,[5,6] which would therefore favor incorporation of Melanesian men rather than women into the ancestral Polynesian groups. Thus, although it is entirely appropriate (and necessary) that genetic studies of human population history should turn to autosomal data, in particular to take full advantage of the increasingly sophisticated demographic models as well as of increasingly automated genotyping technologies that are now becoming available, studies of mtDNA and NRY variation can still provide invaluable insights into the history of human populations.

Acknowledgments

We are grateful to all individuals who provided samples for this study. We thank Inka Szangolies for excellent technical assistance, Gunter Weiss for his contribution to a previous version of this work, and Guillaume Laval for help with the program SIMCOAL2. Financial support was provided by the Max Planck Society.

Received: June 22, 2007
Revised: September 4, 2007
Accepted: September 6, 2007
Published online: January 10, 2008

Web Resources

The URLs for data presented herein are as follows:

ABC, http://cmpg.unibe.ch/people/Excoffier-perso.htm
LEADMIX, http://www.zoo.cam.ac.uk/ioz/software.htm
SIMCOAL2, http://cmpg.unibe.ch/software/simcoal2/

References

1. Melton, T., Peterson, R., Redd, A.J., Saha, N., Sofro, A.S., Martinson, J., and Stoneking, M. (1995). Polynesian genetic

affinities with Southeast Asian populations as identified by mtDNA analysis. Am. J. Hum. Genet. *57*, 403–414.
2. Kayser, M., Brauer, S., Weiss, G., Underhill, P.A., Roewer, L., Schiefenhovel, W., and Stoneking, M. (2000). Melanesian origin of Polynesian Y chromosomes. Curr. Biol. *10*, 1237–1246.
3. Capelli, C., Wilson, J.F., Richards, M., Stumpf, M.P., Gratrix, F., Oppenheimer, S., Underhill, P., Pascali, V.L., Ko, T.M., and Goldstein, D.B. (2001). A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. Am. J. Hum. Genet. *68*, 432–443.
4. Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovsky, L.A., Moyse-Faurie, C., Rutledge, R.B., Schiefenhoevel, W., Gil, D., et al. (2006). Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. Mol. Biol. Evol. *23*, 2234–2244.
5. Hage, P. (1998). Was Proto-Oceanic society matrilineal? J. Polynesian Soc. *107*, 365–379.
6. Hage, P., and Marck, J. (2003). Matrilineality and the Melanesian origin of Polynesian Y chromosomes. Curr. Anthropol. *44*, 121–127.
7. Saar, K., Geller, F., Ruschendorf, F., Reis, A., Friedel, S., Schauble, N., Nurnberg, P., Siegfried, W., Goldschmidt, H.P., Schafer, H., et al. (2003). Genome scan for childhood and adolescent obesity in German families. Pediatrics *111*, 321–327.
8. Trejaut, J.A., Kivisild, T., Loo, J.H., Lee, C.L., He, C.L., Hsu, C.J., Li, Z.Y., and Lin, M. (2005). Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. PLoS Biol. *3*, e247. 10.1371/journal.pbio.0030247.
9. Scheinfeldt, L., Friedlaender, F., Friedlaender, J., Latham, K., Koki, G., Karafet, T., Hammer, M., and Lorenz, J. (2006). Unexpected NRY chromosome variation in Northern Island Melanesia. Mol. Biol. Evol. *23*, 1628–1641.
10. Friedlaender, J.S., Friedlaender, F.R., Hodgson, J.A., Stoltz, M., Koki, G., Horvat, G., Zhadanov, S., Schurr, T.G., and Merriwether, D.A. (2007). Melanesian mtDNA complexity. PLoS ONE *2*, e248. 10.1371/journal.pone.0000248.
11. Pierson, M.J., Martinez-Arias, R., Holland, B.R., Gemmell, N.J., Hurles, M.E., and Penny, D. (2006). Deciphering past human population movements in Oceania: provably optimal trees of 127 mtDNA genomes. Mol. Biol. Evol. *23*, 1966–1975.
12. Hurles, M.E., Nicholson, J., Bosch, E., Renfrew, C., Sykes, B.C., and Jobling, M.A. (2002). Y chromosomal evidence for the origins of oceanic-speaking peoples. Genetics *160*, 289–303.
13. Kirch, P.V. (1997). The Lapita Peoples: Ancestors of the Oceanic World (Oxford: Blackwell).
14. Choisy, M., Franck, P., and Cornuet, J.M. (2004). Estimating admixture proportions with microsatellites: comparison of methods based on simulated data. Mol. Ecol. *13*, 955–968.
15. Wang, J. (2003). Maximum-likelihood estimation of admixture proportions from genetic data. Genetics *164*, 747–765.
16. Bertorelle, G., and Excoffier, L. (1998). Inferring admixture proportions from molecular data. Mol. Biol. Evol. *15*, 1298–1311.
17. Beaumont, M.A., Zhang, W., and Balding, D.J. (2002). Approximate Bayesian computation in population genetics. Genetics *162*, 2025–2035.
18. Excoffier, L., Estoup, A., and Cornuet, J.M. (2005). Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. Genetics *169*, 1727–1738.
19. Garza, J.C., and Williamson, E.G. (2001). Detection of reduction in population size using data from microsatellite loci. Mol. Ecol. *10*, 305–318.
20. Laval, G., and Excoffier, L. (2004). SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. Bioinformatics *20*, 2485–2487.
21. Hamilton, G., Stoneking, M., and Excoffier, L. (2005). Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. Proc. Natl. Acad. Sci. USA *102*, 7476–7480.
22. Hurles, M.E., Irven, C., Nicholson, J., Taylor, P.G., Santos, F.R., Loughlin, J., Jobling, M.A., and Sykes, B.C. (1998). European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. Am. J. Hum. Genet. *63*, 1793–1806.
23. Blust, R. (1999). Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. Symposium Series Inst. Ling. Acad. Sinica *1*, 31–94.
24. Kayser, M., Brauer, S., Weiss, G., Schiefenhovel, W., Underhill, P.A., and Stoneking, M. (2001). Independent histories of human Y chromosomes from Melanesia and Australia. Am. J. Hum. Genet. *68*, 173–190.
25. Kayser, M., Brauer, S., Weiss, G., Schiefenhovel, W., Underhill, P., Shen, P., Oefner, P., Tommaseo-Ponzetta, M., and Stoneking, M. (2003). Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. Am. J. Hum. Genet. *72*, 281–302.
26. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959.
27. Green, R.C. (1991). The Lapita cultural complex: current evidence and proposed models. Indo-Pacific Prehist. Assoc. Bull. *11*, 295–305.
28. Diamond, J.M. (1988). Express train to Polynesia. Nature *336*, 307–308.
29. Terrell, J.E. (1988). History as a family tree, history as an entangled bank: constructing images and interpretations of prehistory in the South Pacific. Antiquity *62*, 642–657.
30. Oppenheimer, S.J., and Richards, M. (2001). Polynesian origins. Slow boat to Melanesia? Nature *410*, 166–167.