populations will require methods that rigorously account for the confounding effects of long-range LD.

Alkes L. Price,[1,2] Michael E. Weale,[3] Nick Patterson,[2] Simon R. Myers,[2,4] Anna C. Need,[3] Kevin V. Shianna,[3] Dongliang Ge,[3] Jerome I. Rotter,[5] Esther Torres,[6] Kent D. Taylor,[5] David B. Goldstein,[3] and David Reich[1,2,*]
[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA; [2]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [3]Institute for Genome Sciences and Policy, Duke University, Durham, NC 27710, USA; [4]Department of Statistics, University of Oxford, Oxford OX1 3TG, UK; [5]Cedars-Sinai Medical Center, University of California Los Angeles, Los Angeles, CA 90048, USA; [6]University of Puerto Rico School of Medicine San Juan, PR 00936, USA
*Correspondence: reich@genetics.med.harvard.edu

Acknowledgments

A.L.P. is supported by a Ruth Kirschstein National Research Service Award from the NIH. N.P. is supported by a K-01 career development award from the NIH. D.R. is supported by a Burroughs Wellcome Career Development Award in the Biomedical Sciences. This research was also supported by U-01 award HG004168 from the NIH (D.R. and N.P.), by NIDDK grant PO1DK46763 (J.I.R.), and by the Board of Governor's Chair in Medical Genetics at Cedars-Sinai Medical Center (J.I.R.). Genotyping of the Puerto Rican samples was supported in part by grant M01-RR00425 to the Cedars-Sinai GCRC genotyping core (K.D.T.) and by NIH grant DK62413 (K.D.T).

References

1. Tang, H., Choudhry, S., Mei, R., Morgan, M., Rodriguez-Clinton, W., Burchard, E.G., and Risch, N.J. (2007). Recent genetic selection in the ancestral admixture of Puerto Ricans. Am. J. Hum. Genet. *81*, 626–633.
2. Tang, H., Coram, M., Wang, P., Zhu, X., and Risch, N. (2006). Reconstructing genetic ancestry blocks in admixed individuals. Am. J. Hum. Genet. *79*, 1–12.
3. Wall, J.D., and Pritchard, J.K. (2003). Linkage disequilibrium in the human genome. Nat. Rev. Genet. *4*, 587–597.
4. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.
5. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet *2*, e190.
6. Tian, C., Plenge, R.M., Ransom, M., Lee, A., Villoslada, P., Selmi, C., Klareskog, L., Pulver, A.E., Qi, L., Gregersen, P.K., et al. (2008). Analysis and application of European genetic substructure using 300 K SNP information. PLoS Genet *4*, e4.
7. Price, A.L., Patterson, N., Yu, F., Cox, D.R., Waliszewska, A., McDonald, G.J., Tandon, A., Schirmer, C., Neubauer, J., Bedoya, G., et al. (2007). A genomewide admixture map for Latino populations. Am. J. Hum. Genet. *80*, 1024–1036.
8. Smith, M.W., Patterson, N., Lautenberger, J.A., Truelove, A.L., McDonald, G.J., Waliszewska, A., Kessing, B.D., Malasky, M.J., Scafe, C., Le, E., et al. (2004). A high-density admixture map for disease gene discovery in African Americans. Am. J. Hum. Genet. *74*, 1001–1013.
9. Fan, J.B., Oliphant, A., Shen, R., Kermani, B.G., Garcia, F., Gunderson, K.L., Hansen, M., Steemers, F., Butler, S.L., Deloukas, P., et al. (2003). Highly parallel SNP genotyping. Cold Spring Harb. Symp. Quant. Biol. *68*, 69–78.
10. Reich, D., Patterson, N., De Jager, P.L., McDonald, G.J., Waliszewska, A., Tandon, A., Lincoln, R.R., DeLoa, C., Fruhan, S.A., Cabre, P., et al. (2005). A whole-genome admixture scan finds a candidate gene for multiple sclerosis susceptibility. Nat. Genet. *37*, 1113–1118.
11. Patterson, N., Hattangadi, N., Lane, B., Lohmueller, K.E., Hafler, D.A., Oksenberg, J.R., Hauser, S.L., Smith, M.W., O'Brien, S.J., Altshuler, D., et al. (2004). Methods for high-density admixture mapping of disease genes. Am. J. Hum. Genet. *74*, 979–1000.

DOI 10.1016/j.ajhg.2008.06.005.

Response to Price et al.

To the Editor: In 2006, Tang and colleagues[1] presented a novel statistical method for genetic admixture analysis based on high-density SNP arrays rather than conventional ancestry informative markers (AIMs). The chromosomes of an admixed individual represent a consecutive patchwork of ancestry blocks representing the ancestral populations contributing to the admixed individual. Their approach[1] is based on the probabilistic reconstruction of those chromosomal ancestry blocks within single individuals. From the block reconstructions, estimates of ancestry at any location in the genome can be derived. The authors recognized that high-density SNP arrays could include nearby markers that are in linkage disequilibrium (LD) in the ancestral population and that such LD could contribute noise to the block reconstructions and subsequent locus-specific ancestry estimation. Therefore, they proposed a Markov-Hidden Markov Model (MHMM) that allowed for pairwise dependency between adjacent markers in the ancestral populations in the estimation process and developed a computer program (SABER) to perform these calculations. They showed, through extensive simulations with data derived from the HapMap project,[2] that the method was robust in reconstructing ancestry blocks, even for very dense sets of markers and for an individual with three ancestral components, and when some of the model parameters were misspecified.[1] Subsequently, Tang et al.[3] used the MHMM to reconstruct ancestry blocks from Affymetrix 100K data in a sample of 192 Puerto Ricans from the Genetics of Asthma in Latino Americans (GALA) study[4] and examined the genome-wide distribution of African, European, and Native American ancestry in this sample.

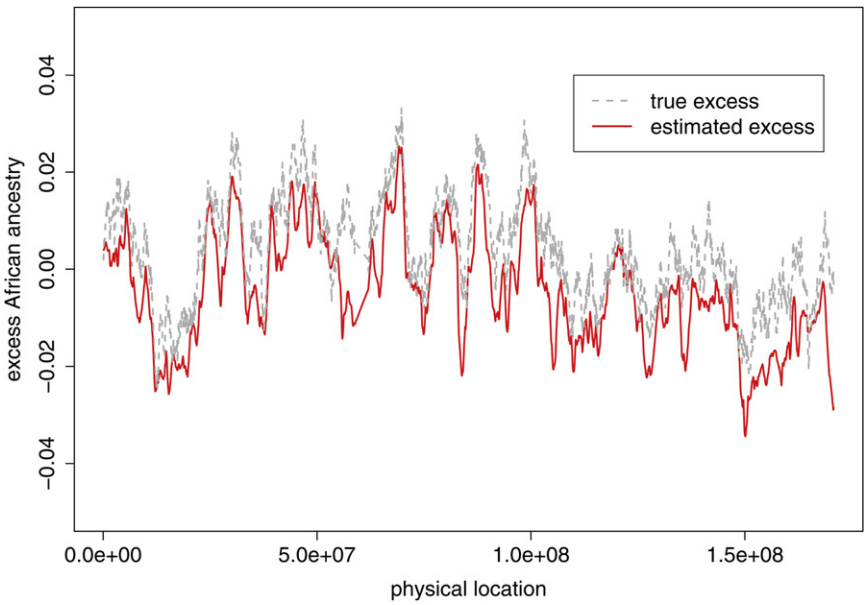

Figure 1. Comparison of Estimated and True Excess African Ancestry on Chromosome 6p

The authors found strong evidence for statistical deviation in ancestry at three chromosomal locations (chromosome 6p, 8q, and 11q), allowing for both statistical variation due to sample size and for ancestry genetic drift, which creates random ancestry variation around the genome.[5] In particular, the location on chromosome 6p overlaps with the HLA cluster of loci, and the authors replicated an observed excess of African ancestry and deficit of European ancestry in an independent sample of Puerto Ricans from the literature, by using also published HLA allele frequencies.

Price et al. now raise a number of concerns regarding both the accuracy and unbiased nature of our ancestry estimation with the MHMM method,[1] as well as our conclusion regarding historic selection as the cause for the significant local ancestry deviations we observed.[3] Their primary concern regarding ancestry estimation is that inclusion of markers that are not in linkage equilibrium (LE) in the ancestral populations can lead to both increased noise and bias. They provide an example of n consecutive SNPs that are in perfect LD (with identical allele frequencies) and show that one can get distorted ancestry estimates if the loci are assumed to be independent. They also provide an example from simulated data in which the inclusion of markers in LD in a set of 1852 markers leads to excess noise and bias in the ancestry estimates. However, all these analyses were performed with the program ANCESTRYMAP and the theory described therein.[6] As the authors have stated, ANCESTRYMAP requires the use of statistically independent markers (i.e., no LD) and furthermore only allows for two ancestral populations.[6] We agree that these requirements may create problems for high-density array data, or more generally for data with markers that are in LD, or for populations with three ancestral components. However, the examples they present are unrealistic and have little relevance for analyses with SABER.[1] SABER allows for LD between adjoining markers in ancestral populations. Therefore, for the example of n consecutive markers in perfect LD, SABER would effectively treat this collection as a single marker only, because no additional information is provided after accounting for the background LD, and produce an accurate ancestry estimate. The MHMM method in SABER also uses a great deal more information than just single SNP genotypes because it uses the empirical distribution of ancestry-block sizes in determining for a given individual the ancestry state at a specific location. In fact, in their extensive simulations with SABER, Tang et al.[1] clearly showed using real data (from HapMap) that the Markovian assumption of pairwise dependency in the ancestral populations was critical to obtain accurate ancestry-block reconstruction and locus-specific estimates. These simulations were performed with markers with an average spacing of 30 kb, 6 kb, and 3 kb (corresponding to the density of a 100K, 500K, and 1000K chip, respectively). Although the ancestry estimation became somewhat noisier with a higher density of SNPs when markers were assumed to be independent, the authors clearly showed robust reconstruction, even at the highest SNP density, when the Markov assumption of pairwise dependency was used via SABER.[1] For Price et al. to imply that their examples have relevance for SABER is incorrect and misleading.

Because Price et al. were particularly concerned about our results on chromosome 6p because of putative long-range linkage disequilibrium in this region in Europeans, we specifically re-examined the results on chromosome 6 from Simulation 2 in Tang et al.[3] According to those authors, "our simulated data incorporates a realistic level of high-order dependency among linked markers, and we have the opportunity to examine whether the MHMM is adequate."[3] Thus, LD between nearby but nonconsecutive SNPs in the real data in this region is featured in the simulated data as well. Figure 1 compares the estimated excess of African ancestry (that is, the estimated locus-specific African ancestry subtracting out the genome-wide average African ancestry) with the true excess African ancestry along chromosome 6. The red line provides the estimated values, and the gray line provides the true values. Overall, the estimated excess of African ancestry is within 2% of the true values, and in fact there is no evidence of any systematic bias near the MHC region located between 26.0 and 34.0 Mb. These results provide additional reassurance from real data that the methods employed in SABER provide unbiased results in the presence of possible background

LD in the ancestral populations on chromosome 6p. Furthermore, in our original paper,[3] we studied the even and odd subsets of markers and found comparable deviations in both subsets in all three regions reported. Thus, the concern raised by Price et al. of systematic distortion in our local ancestry estimates appears to be unwarranted.

Price et al. also find fault in our analysis of HLA data in Puerto Ricans,[1] specifically regarding the appropriateness of populations we used to represent Native Americans in that analysis (Pima and Mayan). Although it is true that there is some genetic variation among Native American groups, and the Taino Indians were the Puerto Rican ancestors, the methods of ancestry estimation that we used and that were based on maximum likelihood (FRAPPE)[7] including the admixed subjects in the estimation of ancestral allele frequencies on the basis of the admixed subjects and not just the ancestral-population surrogates. We have shown previously that by allowing for re-estimation, we can accurately recapture the correct ancestral allele frequencies even when the surrogate-population allele frequencies are somewhat different.[7] Furthermore, a far more serious concern of bias in this type of analysis would arise from assuming that the Native American ancestry component in Puerto Ricans is 0, as Price et al. have done.

The admixture analysis of Price et al. of an independent sample of Puerto Rican Crohn's disease patients and controls, for which they claim no replication of our observed excess African and decreased European ancestry on chromosome 6p, also deserves comment. As they've stated, they reduced an initial marker set of 2459 SNPs to 1438 to eliminate markers that had allele frequency differences between Europeans and Native Americans as well as to "disallow LD between markers in the ancestral populations." This is because the ANCESTRYMAP program is not robust to background LD and also does not allow for more than two ancestral populations. Although this marker density corresponds to approximately one marker for every 2 Mb, because chromosome 6p putatively has an extended region of LD from 25.5 to 33.5 Mb,[8] we assume they allowed very few markers in this region, perhaps only one (rs451774 at 28.6 Mb). If so, the claim that "61% of maximum information about African vs. non-African ancestry at the chromosome 6p region" was obtained is difficult to imagine, especially because the allele frequency difference between Africans and Europeans for that marker is only approximately 0.40. Furthermore, the lack of allowance for Native American ancestry in their analysis makes their results difficult to interpret. They also show that their method is highly conservative because a simulated ancestry excess of .14 was reduced by more than a factor of two upon estimation. Despite the low power of their analysis, they still observed a modest increase in African ancestry at chromosome 6p and might have observed a greater increase with greater marker density and information.

Of course, we agree that all initial genetic observations, be they disease associations or arguments for ancestral selection, require independent replication. We therefore also conducted an independent replication study, this time with AIMs rather than high-density chip data. We examined a new sample of 383 Puerto Rican subjects from the GALA study,[4] approximately double in size of our original sample. We typed 104 AIMs from around the genome and obtained a genome-wide estimate of African, European, and Native American ancestry for each individual using FRAPPE.[7] For comparison, we estimated ancestry on chromosome 6p using five ancestry informative markers: rs393228, rs7773913, rs853693, rs6456883, and rs847851. These markers span from 25.07 to 35.01 Mb on chromosome 6p. The estimated average African ancestry outside of chromosome 6p was 25.5%; by contrast, the estimated African ancestry based on the five markers on chromosome 6p was 40.0%, an excess of 14.5%, comparable to the difference we observed in our original study.[3] To assess statistical significance of this difference, we estimated the African ancestry at chromosome 6p for each individual. To do this, we first performed a single-marker analysis, in which we computed the posterior probability that an allele is derived from an African ancestor, given the ancestral allele frequencies and the individual's genome-wide ancestry:

$$\hat{z} = P\Big(\text{African} \mid (t_{afr}, t_{eur}, t_{amr}), (p_{afr}, p_{eur}, p_{amr})\Big) = \frac{t_{afr} p_{afr}}{t_{afr} p_{afr} + t_{eur} p_{eur} + t_{amr} p_{amr}},$$

where (p_{afr}, p_{eur}, p_{amr}) are the allele frequencies in the three ancestral populations, respectively, and (t_{afr}, t_{eur}, t_{amr}) denote the genome-wide ancestry proportions for the individual. We then computed the location-specific ancestry of an individual by averaging over the five SNPs. This analysis is quite conservative because the ancestry estimate at chromosome 6p is shrunken significantly back toward the individual's genome-wide estimate by the Bayesian calculation. Thus, in this case we observed an average of 30.1% African ancestry at 6p, still greater than the 25.5% genome-wide estimate. We then calculated, for each individual, the difference between the estimated African ancestry at chromosome 6p (as derived above) and the genome-wide African ancestry. The mean of this difference was .0525, with a standard error of .0066. A t test to determine whether the mean is significantly different from 0 yielded a t value of 8.4, $p < 10^{-15}$. Thus, the conclusion of excess African ancestry on 6p compared with the rest of the genome in this sample is unequivocal and confirms our original observation.

The specter of bias in our analysis was probably raised by Price et al. due to the fact that the three locations we identified as sites of ancestral selection mapped into three regions with long-range LD, as they have described in Table 1 of their letter. Ironically, long-range LD has been cited as evidence for historical selection, not by us but by others, including some of the authors of the current letter.[8] In fact, long-range LD was used as an argument for historical

selection on the lactase persistence SNP on chromosome 2q.[8] Interestingly, this region of chromosome 2q (134.5 to 138.0 Mb) is also on the list of extended LD in Table 1 of Price et al. It is puzzling that on the one hand long-range LD has been used as evidence for selection in one analysis[8] and on the other as evidence for bias and against selection in the current letter.

Furthermore, our initial distribution of genome-wide excess African ancestry was quite symmetric and fit a simulated null distribution quite well, with the exception of a very small number of outlier loci (Figure 2 in Tang et al.[3]). These outlier loci were on chromosome 6p. If polymorphic inversions in the European population and associated regions of extended LD were an important source of bias in our analyses, as suggested by Price et al., we would have expected to see more outlier points in this distribution, specifically at locations corresponding to the 24 regions identified in Table 1 of Price et al. Aside from the three regions already mentioned, and possibly another region at 8p, none of the remaining 20 regions showed any deviation of ancestry from background levels.[3]

Extended regions of LD in the human genome have been previously described. Huttley et al.[9] studied 5048 autosomal microsatellite markers in Europeans and identified ten regions with putative evidence of long-range LD. Price et al. have now extended these findings by examining 550K SNP markers. Although the two studies identified some overlapping regions (chromosomes 2p, 6p, and 7p), many other regions are distinct. Whereas numerous authors, including Huttley et al.[9] and Bersaglieri et al.,[8] have suggested these regions represent targets of historical selection, Price et al. now propose that regions of long-range LD they identified are due to polymorphic inversions but have provided no evidence to support this contention. We believe other evidence argues against this conclusion. Jorgenson et al.[10] compared genetic maps across four major racial-ethnic groups in a very large sample of sibships. They noted that polymorphic inversions impact genetic map distances, and when the frequencies of these inversions differ across groups, map distances between markers in and around the inversion will consequently also differ significantly between groups. They identified two regions, one on chromosome 8p and another on 12q, that displayed ethnic-specific map differences. The region on chromosome 8p coincided with a previously described polymorphic inversion.[11] However, they found no other genomic region (aside from 12q) with significant ethnic-specific map differences (including on chromosomes 6p, 8q, and 11q); in particular, none of the regions in Table 1 of Price et al. aside from chromosome 8p showed evidence of map differences. Thus, the suggestion that the regions of long-range LD identified by Price et al. (aside from 8p) are due to polymorphic inversions appears highly speculative, at best.

We do agree that the fact that our three regions on chromosomes 6p, 8q, and 11q coincided with three regions of extended LD in Table 1 of Price et al. is unlikely to be due to chance. However, it seems inconsistent to argue that long-range LD provides evidence of historical selection in one population but when similar evidence is found in a population derived from it, selection is deemed unlikely and artifact is invoked. For example, Price et al. and others have argued that the HLA region on chromosome 6p is particularly interesting because of its broad impact on disease. Again, it seems contradictory to argue that the HLA region on chromosome 6p has been a target of selection in Europeans and other populations but could not have been in Puerto Ricans, leading to a differential ancestry distribution. The region on chromosome 8p harboring a polymorphic inversion, which showed a suggestive but not significant ancestry deviation in our analysis, harbors an olfactory gene cluster and has shown phenotypic effects in other studies.[11] Furthermore, two of the other regions we identified (6p and 8q) also harbor olfactory gene clusters,[3] an observation that seems unlikely to be due purely to chance.

Price et al. argue that because they observed no evidence of ancestry distortions in African Americans, there must not be any in Puerto Ricans either. We do not see the relevance of this observation because these are populations with distinct and nonoverlapping social, demographic, and genetic histories.

In summary, we have shown that the MHMM approach, as implemented in the program SABER, is robust to putative regions of extended LD in real data. This method should be particularly useful for investigators studying admixed populations with high-density chips. Furthermore, we have shown a convincing replication of our prior results of excess African ancestry on chromosome 6p in Puerto Ricans.

Hua Tang,[1] Shweta Choudhry,[2] Rui Mei,[5]
Martin Morgan,[6] William Rodriguez-Cintron,[7,8]
Esteban Gonzalez Burchard,[2,4] and Neil J. Risch[3,4,9,*]
[1]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA; [2]Departments of Biopharmaceutical Sciences and Medicine, [3]Department of Epidemiology and Biostatistics, [4]Institute for Human Genetics, University of California San Francisco, San Francisco, CA 94143, USA; [5]Affymetrix, Santa Clara, CA 95051, USA; [6]Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [7]Veterans Caribbean Health Care System, San Juan, PR 00921, USA; [8]University of Puerto Rico School of Medicine San Juan, PR 00936-5067, USA; [9]Division of Research, Kaiser Permanente, Oakland, CA 94611, USA
*Correspondence: rischn@humgen.ucsf.edu

References

1. Tang, H., Coram, M., Wang, P., Zhu, X., and Risch, N. (2006). Reconstructing genetic ancestry blocks in admixed individuals. Am. J. Hum. Genet. *79*, 1–12.
2. International HapMap Consortium (2005). A haplotype map of the human genome. Nature *437*, 1299–1320.
3. Tang, H., Choudhry, S., Mei, R., Morgan, M., Rodriguez-Cintron, W., Burchard, E.G., and Risch, N.J. (2007). Recent genetic

selection in the ancestral admixture of Puerto Ricans. Am. J. Hum. Genet. *81*, 626–633.
4. Burchard, E.G., Avila, P.C., Nazario, S., Casal, J., Torres, A., Rodriguez-Santana, J.R., Syliva, J.S., Fagan, J.K., Salas, J., Lilly, C.M., et al. (2004). Lower bronchodilator responsiveness in Puerto Rican than in Mexican asthmatic subjects. Am. J. Respir. Crit. Care Med. *169*, 386–392.
5. Long, J. (1991). The genetic structure of admixed populations. Genetics *127*, 417–428.
6. Patterson, N., Hattangadi, N., Lane, B., Lohmueller, K., Hafler, D., Oksenberg, J., Hauser, S., Smith, M., O'Brien, S., Altshuler, D., et al. (2004). Methods for high-density admixture mapping of disease genes. Am. J. Hum. Genet. *74*, 979–1000.
7. Tang, H., Peng, J., Wang, P., and Risch, N. (2005). Estimation of individual admixture: Analytical and study design considerations. Genet. Epidemiol. *28*, 289–301.
8. Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A., Rhodes, M., Reich, D.E., and Hirschhorn, J.N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. Am. J. Hum. Genet. *74*, 1111–1120.
9. Huttley, G.A., Smith, M.W., Carrington, M., and O'Brien, S.J. (1999). A scan for linkage disequilibrium across the human genome. Genetics *152*, 1711–1722.
10. Jorgenson, E., Tang, H., Gadde, M., Province, M., Leppert, M., Kardia, S., Schork, N., Cooper, R., Rao, D.C., Boerwinkle, E., and Risch, N. (2005). Ethnicity and human genetic linkage maps. Am. J. Hum. Genet. *76*, 276–290.
11. Giglio, S., Broman, K.W., Matsumoto, N., Calvari, V., Gimelli, G., Neumann, T., Ohashi, H., Voullaire, L., Larizza, D., Giorda, R., et al. (2001). Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. Am. J. Hum. Genet. *68*, 874–883.

DOI 10.1016/j.ajhg.2008.06.009.

East Asian and Melanesian Ancestry in Polynesians

To the Editor: Kayser et al.[1] estimated the ancestry of Polynesians by using 377 autosomal microsatellite loci and concluded that 0.79 of the ancestry was from East Asians (95% CI, 0.76–0.84) and 0.21 from Melanesians. In contrast, maternally inherited mtDNA ancestry was previously estimated to be 0.94 East Asian and 0.06 Melanesian and paternally inherited Y chromosome ancestry was estimated to be 0.28 East Asian and 0.66 Melanesian.[2] One might guess that the East Asian autosomal ancestry would be approximately the arithmetic average of the mtDNA and Y ancestry, 0.61, but the estimated autosomal ancestry of 0.79 is substantially higher. To account for this difference and the different estimates in ancestry from different chromosomes, strong sex differences in gene flow, occurring in a particular chronological order, are necessary. Here I present a simple two-phase scenario to explain the different observed ancestries for autosomal, mtDNA, and Y markers and then discuss how this scenario could be modified and still result in the observed patterns.

First, assume that a population of East Asian ancestry, which eventually became the Polynesians, settled in Melanesia and that subsequently there was male gene flow from Melanesians into this population. This pattern is consistent with both matrilocality and matrilinearity in this population.[1] The effect of this male gene flow at a rate of m_m per generation over t generations on Y ancestry can be given[3] as

$$q_t = (1 - m_m)^t q_0 + \left[1 - (1 - m_m)^t\right] q_{Mel}$$

where q_0 and q_t are the initial and t generation East Asian ancestry in the population and q_{Mel} is the East Asian ancestry in the Melanesian migrants. Assuming that $q_{Mel} = 0$, $q_0 = 1$, and $q_t = 0.28$, then

$$0.28 = (1 - m_m)^t \text{ or } m_m = 1 - e^{[\ln(0.280)/t]}.$$

For example, if $t = 50$, then $m_m = 0.0251$.

For autosomal loci in this population, the East Asian ancestry is

$$q_t = \left[1 - \frac{1}{2}(m_f + m_m)\right]^t q_0 + \left\{1 - \left[1 - \frac{1}{2}(m_f + m_m)\right]^t\right\} q_{Mel}$$

where m_f is the per-generation rate of female gene flow. Again, assume that $q_{Mel} = 0$, $q_0 = 1$, $m_f = 0$, and with the estimated m_m of 0.0251 used, $q_t = (0.987)^t$. For example, if $t = 50$, then $q_t = 0.532$.

Second, assume that subsequently there was female gene flow from the East Asians into this population for x generations so that the autosomal East Asian ancestry can be expressed as

$$q_{t+x} = \left[1 - \frac{1}{2}(m_f + m_m)\right]^x q_t + \left\{1 - \left[1 - \frac{1}{2}(m_f + m_m)\right]^x\right\} q_{EA}$$

where q_{EA} is the East Asian ancestry in the East Asian female migrants. Assuming that $q_{EA} = 1$, $q_t = 0.532$, $q_{t+x} = 0.79$, and $m_m = 0$,

$$0.79 = 1 - 0.468\left(1 - \frac{1}{2}m_f\right)^x \text{ or } m_f = 2\left(1 - e^{[\ln(0.449)/x]}\right).$$

For example, if $x = 50$, then $m_f = 0.0318$.

This two-phase scenario is presented in Figure 1, which shows a decline in Y and autosomal East Asian