Am. J. Hum. Genet. 67:1062–1066, 2000

INVITED EDITORIAL
Asymmetries in the Maternal and Paternal Genetic Histories of Colombian Populations

Mark Seielstad

Program for Population Genetics, Harvard School of Public Health, Boston

The mitochondrial genome was among the earliest molecules to be surveyed for variation at the DNA level in humans. Some of the advantages offered by mtDNA in the pre-PCR era were purely practical: it is abundant in most cells, and its circularity and small size allowed it to be purified from nuclear DNA by centrifugation in cesium chloride gradients. Even without cloning or PCR amplification, sufficient DNA could be obtained for surveys of restriction fragment–length polymorphisms. In addition to these features, which provide for its easy manipulation in the laboratory, the molecule possesses further advantages to the population geneticist. The most exploited feature has been its apparent inability to undergo genetic recombination; despite the large number of mitochondrial genomes in most cells (and even within single mitochondria) there is still no convincing evidence that portions of the genome are exchanged in a recombinational process between molecules (Jorde and Bamshad 2000; Kivisild and Villems 2000; Kumar et al. 2000; Parsons and Irwin 2000). This failure to recombine means that all modern mtDNA sequences descend from a single ancestral molecule at some point in the past (a feature known as "coalescence") and that a modern mtDNA sequence will differ from an ancestral molecule only by the accumulation of mutations—a process that appears to be fairly regular when viewed over appreciable evolutionary periods. Thus, it has been a relatively straightforward task to assess the extent of genetic variation in living humans and to estimate the time required to generate it. In practice, this is done by counting the number of mutations that have occurred between a reconstructed ancestral DNA sequence and a sample of modern sequences and then multiplying this by the mutation rate of mtDNA. Allan Wilson and his associates were among the first to exploit this feature of mtDNA and to recognize that our origins as a species should not antedate the coalescence time of mtDNA (barring a very complicated history of natural selection and migration) (Cann et al. 1987; Vigilant et al. 1991).

From these studies, we have learned that our species is not particularly old in evolutionary terms (probably <150,000 years, and perhaps substantially "younger") and that we, like our closest primate relatives, originated in Africa. This recent origin from a fairly small initial population, followed by a massive, culturally driven expansion within the last 10,000 years (Cavalli-Sforza et al. 1993), has produced a unique pattern of human variation: despite our towering modern population size (by far the greatest and most geographically extensive of any large mammal), we have very little genetic variation compared with other species—even those, like our primate relatives, with much smaller modern population sizes (Kaessmann et al. 1999). The implications of this unusual demographic history for the geographic distribution of disease-predisposing variation are just beginning to be considered by genetic epidemiologists, and whether our history will prove to aid or hinder the effort to locate disease-relevant genetic variation is a matter of considerable discussion (Collins et al. 1999; Kruglyak 1999). One way to map such disease-related variation is by detection of an indirect association via linkage disequilibrium (LD; the nonrandom association of alleles at adjacent loci) with markers in a genetic map that lie next to a disease-predisposing variant. LD is expected to be elevated in populations that began with a small number of founders in the recent past, so it is possible that our species will be reasonably well suited to gene mapping through LD, although this remains an area of some debate and empirical investigation (Collins et al. 1999; Kruglyak 1999; Pritchard et al. 1999; Taillon-Miller et al. 2000).

Regardless of the gross demographic history of our species and its influence on the distribution of genetic variation, the history of individual populations is undoubtedly quite variable, with examples of recently founded populations and very extreme genetic bottlenecks contrasting with populations that began to grow to their current large numbers in the more distant past (e.g., possibly parts of China). The hope that we will be

Received September 20, 2000; accepted for publication September 21, 2000; electronically published October 9, 2000.

Address for correspondence and reprints: Dr. Mark Seielstad, 665 Huntington Avenue, FXB-101, Boston, MA 02115-6096. E-mail: mark@ppg.harvard.edu

This article represents the opinion of the author and has not been peer reviewed.

0002-9297/2000/6705-0004$02.00

able to identify human populations whose demographic history endows them with just the right level and distribution of LD, or just the right amount of genetic variation for particular disease-mapping projects, may not be in vain. This, at any rate, is part of the logic motivating two articles in this month's *Journal* that explore the genetic history of several South American populations in detail (Carvajal-Carmona et al. 2000; Mesa et al. 2000 [both, in this issue]).

The genetic story of colonization is itself a fascinating one, and, as the authors of these articles indicate, the complex and highly variable patterns of migration and gene flow that it produced provide some of the best natural experiments available to human geneticists. With their focus on sexual asymmetries in the direction and extent of gene flow among Native Colombian populations and more-recent immigrants from Europe and Africa, these researchers working in Colombia and Britain have added a novel dimension to the genetic story of colonization. As shown in table 1, it is not only their lack of recombination that distinguishes mtDNA and the Y chromosome from the autosomes. A possibly more interesting feature is their sex-specific mode of inheritance. Through its involvement in sex determination, the Y chromosome is paternally transmitted by necessity. In contrast, mtDNA is maternally transmitted, because the mitochondrion is a cytoplasmically inherited organelle, with the vast bulk of the cytoplasm provided by the mother's egg cell.

The sex-specific features of mtDNA and the Y chromosome are just beginning to be exploited (Poloni et al. 1997; Seielstad et al. 1998; Perez-Lezaun et al. 1999; Jorde et al. 2000). Early worldwide analyses indicated that populations exhibit a high degree of population substructure (a high "F_{ST}") for the Y chromosome (Poloni et al. 1997; Seielstad et al. 1998). When compared with a relatively lower degree of population substructure for mtDNA, gene flow (or migration) among most human populations, it was suggested, was generally mediated by women (Seielstad et al. 1998). Women, in other words, appear to have a higher migration rate than men in most societies; populations are generally more similar to one another with respect to mtDNA than they are with respect to the Y-chromosome. However surprising this result may initially seem, it is easily understood in terms of the rules of postmarital residence that operate in the majority of traditional societies. Most human populations are patrilocal, meaning that, soon after marriage, a wife will typically move from her own family into the natal household of her husband. Men, whatever their movements over a lifetime, will generally return to and rear their children at the place of their birth. This is especially pronounced when land and other forms of nonportable wealth are paternally inherited.

In the first of two articles, Mesa et al. (2000) set out to test this hypothesis of a higher female-migration rate on a more local scale, or at least to see whether any sexual asymmetries in migration rates can be ascertained by examination of autosomal, mitochondrial, and Y-chromosomal variation in five Native Colombian populations. Migration rates among populations are typically assessed by a measure of genetic differentiation, such as F_{ST}, among the populations (in this case, the authors use an essentially equivalent measure known as "G_{ST}"). The larger the F_{ST} value, the more distinct are the populations, implying less gene flow to homogenize them. The actual formula relating F_{ST} to migration rates under a simplified island model of migration is $F_{ST} = 1/(1 + 4Nm)$ for the autosomes, where N is the effective population size and m is the migration rate among populations. In fact, F_{ST} also depends on the mutation rate (m) of the genetic markers used to calculate it, although m has traditionally been ignored, since it is assumed to be several orders of magnitude lower than the migration rate among populations. Comparisons of autosomal and Y-chromosomal F_{ST} values calculated from single-nucleotide polymorphisms (SNPs) and more-rapidly mutating microsatellites call this assumption into question, however, and it may be useful to incorporate mutation rate estimates into future analyses. The authors examine

Table 1

Variable	mtDNA	Autosomes	Y Chromosome	X Chromosome
Transmission	Maternal	Biparental	Paternal[a]	Biparental[b]
Copy number	Sexes equal	M: 2; F: 2	M: 1; F: 0	M: 1; F: 2
Ideal N_e relative to autosomes	1/4	1	1/4	3/4
Recombination	None	In both sexes	None[c]	In F only[c]

NOTE.—M = male; F = female; N_e = effective population size.

[a] Paternal transmission of the Y chromosome is limited to male offspring.

[b] Paternal transmission of the X chromosome is limited to female offspring.

[c] Recombination does occur between the X and Y chromosomes in males but is limited to two small pseudoautosomal regions at the Y-chromosomal telomeres.

this question by calculating R_{ST}, a measure of population differentiation that specifically accounts for the stepwise mutational behavior of microsatellites. They find little difference in the values calculated by G_{ST} and R_{ST} in one restricted data set, and they infer from this that mutation rates are not making an important contribution to G_{ST} values. This is not a completely convincing argument, however, and it does not explain the discrepancy between F_{ST} values that has sometimes been observed between SNPs and microsatellites. When F_{ST} values from genetic markers with differing mutation rates are compared, it may be necessary to incorporate mutation-rate estimates into the analyses—when they are known.

For uniparentally inherited molecules such as mtDNA and the Y chromosome, $F_{ST} = 1/(1 + Nm)$. Because the Y chromosome and mtDNA have fewer copies in a population, the F_{ST} among populations increases more rapidly, as a result of drift for these markers, than it does for the autosomes. This is clear in the study by Mesa et al. (2000), where G_{ST} values are 0.17, 0.18, and 0.068 for mtDNA, the Y chromosome, and autosomal markers, respectively.

The equivalence of G_{ST} values calculated from mitochondrial and Y-chromosomal markers suggests that there is no significant difference between male and female migration rates and effective population sizes among these populations. However, it is possible that European colonization has disrupted the patterns of social structure that existed in these populations before contact. Since many of the Y chromosomes in these Native Colombian populations have recognizably European or African origins, these can be excluded from the analysis to get an idea of the pre-Columbian situation. When this is done, the Y-chromosomal G_{ST} value does indeed increase to 0.23, although it still fails to differ significantly from the mitochondrial value of 0.17 with the sample size available here. The simplest interpretation of this result is that migration rates between the various populations are essentially equal for males and females, although Mesa et al. (2000) note that there are roughly equal numbers of matrilocal and patrilocal populations in South America. Is it possible that the average picture presented by a G_{ST} analysis fails to indicate a strong bias in migration rates because the proportion of male and female migrants over the entire region is roughly equal, even though biases in the dispersing sex may exist between individual pairs of populations? An additional question, which even ethnographers seem to have ignored, is how the rules of postmarital residence are applied to marriages between members of different cultures. It is unfortunate that more-detailed demographic and ethnographic information is not available for the exact populations in this study; having now corrected the deficiencies of earlier studies by typing the same samples and populations for mitochondrial, autosomal, and Y-chromosomal markers, the next step is to combine the genetic data with good demographic data on the dispersing sex, extent of polygyny, and migration rates among populations.

Finally, it needs to be acknowledged that differences in levels of differentiation among populations need not indicate differences in migration rates among the populations. As shown in the equations above, F_{ST} depends on the product of effective population size and migration rate. Without independent data on one of these parameters, it is impossible with current methods to disentangle the relative contribution of N or m to differences in F_{ST}. So, without additional data, it is possible that the equivalence of G_{ST} values calculated from mtDNA and the Y chromosome result from differing male-to-female migration rates that are offset by an inverse reduction in population size. By the same token, it is possible to argue that polygyny—which would lower the male and Y-chromosomal effective population sizes—and not a higher female migration rate has led to the greater extent of differentiation for Y-chromosome versus mitochondrial markers observed by others (Seielstad et al. 1998). However, polygyny also reduces the autosomal effective population size, and the ratio of the Y-chromosomal to autosomal effective population sizes is never much less than ¼ (Nunney 1991, 2000).

In addition to examining potential differences in migration rates, Mesa et al. (2000) were able to get some insight into the asymmetries of gene flow between native, European, and African populations. Data from three autosomal systems (microsatellites, β-globin RFLPs, and DQA1 typing), four mitochondrial RFLPs and a 9-bp deletion, and seven Y-chromosomal microsatellites have allowed them to estimate the geographic origins of variation at each genetic system. Nearly all the mitochondrial haplotypes observed are restricted to Native American populations, so there is little evidence for the entry of immigrant females into Native Colombian populations. However, there are signs of gene flow from European and African populations in the autosomes and Y chromosomes of four of the five populations. This pattern implies that most of the gene flow has occurred via males. There are discrepancies in the results as summarized in their table 6, however. For example, autosomal estimates of nonnative contributions for the Ingano are lower than the estimates from both the Y chromosome and mtDNA. Similarly, there are instances of strikingly different estimates of African versus European admixture for the autosomes versus the Y chromosome in some of the populations. Again, when very little of the admixture appears to have entered via immigrant females, these discrepancies are hard to explain. One limitation is the use of STRs (microsatellites) for the Y-chromosomal estimates. Y chromosome microsatellite haplotypes show less geographic structure and more homoplasy (convergent mutation) than do hap-

lotypes based on SNPs, so the geographic origins of these haplotypes may not be so easy to recognize.

A similar approach was taken in the accompanying article by Carvajal-Carmona et al. (2000), to estimate the geographic origins of the mtDNA and Y-chromosomal variation in the admixed Colombian population of Antioquia. What the authors find is a striking discordance between the geographic origins of the population's Y chromosomes and mtDNA. Where autosomal data had indicated a European contribution of ~70%, the present authors' closer examination using paternally and maternally transmitted regions of the genome indicates that the proportion of European males founding the population is likely to have been much higher, with the majority of female founders of local Amerindian derivation. Some 90% of the mitochondrial lineages in the Antioquian population belonged to one of four haplotypes unique to Native Americans, whereas only one of the 80 Y chromosomes examined was typical of Native American populations. Three Y chromosome haplotypes were of West African origin, whereas the remaining 90% were most likely from European and North African populations. These results, although striking, agree with historical data that show that many men immigrated from Spain without an accompanying spouse and that women comprised as few as 10% of European immigrants to the western hemisphere.

Another implication of this result is that more men than women entered the Antioquian population, since the autosomal admixture estimates—which are essentially an average of the male and female contribution—are closer to the estimates from the Y chromosome in showing a nearly 70% European contribution to Antioquia, than they are to estimates from mtDNA. As the authors indicate, this need not reflect a larger initial male founding population. Instead, it is possible that males continued to arrive in Antioquia from Spain, with declining numbers of native women entering the population—partly the result of catastrophic native mortality rates. This could also be viewed as a higher level of inbreeding for mtDNA versus the Y chromosome, within the same population.

Carvajal-Carmona et al. (2000) go on to estimate the more precise origin of the European contribution to Antioquia. The possibly significant Basque and Sephardic contributions to the population of Antioquia are interesting, but this analysis is a bit speculative and depends heavily on which populations are available for comparison. Furthermore, haplotype frequencies for mtDNA and the Y chromosome are subject to considerable levels of drift, making the comparison of haplotype frequencies among populations somewhat complicated. Unfortunately, within the Iberian Peninsula, only data from the Basques and Catalans are available to the authors, so it is possible that they have overstated a Basque contribution, and they have less information about the rest of Spain (particularly its southern populations) than one would like. In fact, most of the inferences regarding a genetic contribution from southern Spain are based on haplotype frequencies from North African populations. This is not entirely unreasonable, given the close historical connections between southern Spain and North Africa, although it is clearly an imperfect solution. Also, because the Basques and Catalans have similar haplotype profiles, it seems almost equally likely that Catalans (or even another unsampled [Iberian] population) contributed to the Colombian population. The estimation of the Sephardic contribution is probably less than definitive as well. Here again, the effects of drift and significant geographic gaps in sampling might affect the confidence with which we attribute these Y chromosomes to Sephardic Jews. Nevertheless these results are intriguing, and many of them are likely to be substantiated by further research.

It remains to be seen whether this complex demographic history will produce favorable circumstances for the mapping of disease-predisposing allelic variation. The founding size of Antioquia appears to have been reasonably large, so one may not expect much LD to have been generated by a founder effect. On the other hand, LD is also created by admixture between populations with divergent allele frequencies (Chakraborty and Weiss 1988; Wilson and Goldstein 2000). In such a circumstance, LD initially results from allele frequency differences alone, although this source of disequilibrium dissipates quickly if mating is random (which may not have been the case in Colombia), leaving a residual disequilibrium that depends primarily on the recombinational distance between markers.

The most exciting aspect of these articles is their implicit suggestion of just how much variety we might expect in the demographic history of populations—especially in colonial situations. As the authors suggest, the complex and sexually asymmetrical events that shaped the genetics of modern Colombian populations provide a strong rationale for continuing studies of this type: the likelihood of mapping complex disease genes may vary significantly as a function of a population's history, and knowledge of this history may allow us to more rationally choose populations for genetic epidemiological studies.

References

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Carvajal-Carmona LG, Soto ID, Pineda N, Ortíz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, Ruiz-Linares A (2000) Strong Amerind/white sex bias and a possible Sephardic contribution

among the founders of a population in northwest Colombia. Am J Hum Genet 67:1287–1295 (in this issue)
Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human evolution. Science 259:639–646
Chakraborty R, Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci USA 85: 9119–9123
Collins A, Lonjou C, Morton NE (1999) Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci USA 96:15173–15177
Jorde LB, Bamshad M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288: 1931
Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 66:979–988
Kaessmann H, Wiebe V, Paabo S (1999) Extensive nuclear DNA sequence diversity among chimpanzees. Science 286: 1159–1162
Kivisild T, Villems R (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931
Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139–144
Kumar S, Hedrick P, Dowling T, Stoneking M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931
Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortíz-Barrientos D, García LF, Velez ID, Bravo ML, Múnera JG, Bedoya G, Bortolini M-C, Ruiz-Linares A (2000) Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. Am J Hum Genet 67:1277-1286 (in this issue)
Nunney L (1991) The influence of age structure and fecundity on effective population size. Proc R Soc Lond B 246:71–76
Nunney L (2000) Social structures and effective population sizes. Presented at Social Systems and Population Genetics Conference; La Sage; September 2000
Parsons TJ, Irwin JA (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931
Perez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martinez-Arias R, Clarimon J, Fiori G, Luiselli D, Facchini F, Pettener D, Bertranpetit J (1999) Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. Am J Hum Genet 65:208–219
Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L (1997) Human genetic affinities for Y-chromosome P49a,f/*Taq*I haplotypes show strong correspondence with linguistics. Am J Hum Genet 61:1015–1035
Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16: 1791–1798
Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278–280
Taillon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen T, Cao A, Kere J, Pilia G, Rice JP, Kwok PY (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nat Genet 25:324–328
Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507
Wilson JF, Goldstein DB (2000) Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. Am J Hum Genet 67:926–935