Population differentiation and migration: Coalescence times in a two-sex island model for autosomal and X-linked loci

Sohini Ramachandran [a,b,*], Noah A. Rosenberg [c,d,e], Marcus W. Feldman [f], John Wakeley [b]

[a] Society of Fellows, Harvard University, Cambridge, MA, USA
[b] Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA
[c] Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA
[d] Center for Computational Medicine and Biology, University Michigan, Ann Arbor, MI, USA
[e] The Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA
[f] Department of Biological Sciences, Stanford University, Stanford, CA, USA

ARTICLE INFO

ABSTRACT

Evolutionists have debated whether population-genetic parameters, such as effective population size and migration rate, differ between males and females. In humans, most analyses of this problem have focused on the Y chromosome and the mitochondrial genome, while the X chromosome has largely been omitted from the discussion. Past studies have compared F_{ST} values for the Y chromosome and mitochondrion under a model with migration rates that differ between the sexes but with equal male and female population sizes. In this study we investigate rates of coalescence for X-linked and autosomal lineages in an island model with different population sizes and migration rates for males and females, obtaining the mean time to coalescence for pairs of lineages from the same deme and for pairs of lineages from different demes. We apply our results to microsatellite data from the Human Genome Diversity Panel, and we examine the male and female migration rates implied by observed F_{ST} values.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

As sex-biased dispersal processes are common in a variety of species (Lawson Handley and Perrin, 2007), evolutionists have long been interested in how variables relating to demography and population structure differ between males and females.

Differences between human males and females in parameters such as migration rate and effective population size have generally been investigated using the uniparentally-inherited Y chromosome and mitochondrial genome. Past studies have observed differences in autosomal, Y-chromosomal and mitochondrial variation, and have typically explained these differences based on matrilocality or patrilocality (Wilkins and Marlowe, 2006; Wilkins, 2006).

In a patrilocal society, we expect to see more genetic differentiation across Y-chromosomal lineages than across mitochondrial lineages; such a pattern was observed using globally-distributed samples by Seielstad et al. (1998), while patterns consistent with matrilocality have been observed in Thailand (Oota et al., 2001) and

Melanesia (Kayser et al., 2008). Recent studies have questioned the spatial scale at which one can expect to infer a genetic signature of patrilocality or matrilocality, arguing that this signal may be observable within geographic regions, but likely not at a global level (Wilder et al., 2004a; Wilkins and Marlowe, 2006).

The X chromosome has contributed comparatively little to the inference of sex-specific human migration rates. Garrigan et al. (2007) compared genetic variation using resequence data at two X-linked loci totaling 8486 bp, 6650 bp encompassing 13 *Alu* elements on the Y chromosome, and 780 bp of the cytochrome oxidase subunit III on the mitochondrion. Their inference of migration rates among 10 human populations did not produce a consistent pattern of sex-biased gene flow across all the loci investigated, though different rates of male and female migration were inferred for many pairs of populations.

Although variation in the Y chromosome and the mitochondrion has generally been used in studies of sex-specific differences in human dispersal, comparisons between variation observed on the X chromosome and on autosomes also have the potential to shed light on evolutionarily interesting differences between males and females (Schaffner, 2004). In contrast with the Y chromosome and the mitochondrial genome, each of which is effectively a single absolutely-linked locus, the X chromosome and autosomes offer numerous independent markers. The availability of multiple

markers potentially adds power to the analysis, although recombination and the movement of the autosomes and X chromosome between males and females are expected to complicate the elucidation of sex-specific histories (Ramachandran et al., 2004; Wilkins and Marlowe, 2006).

Using 17 X-linked and 377 autosomal microsatellites genotyped in 52 globally-distributed populations in the Human Genome Diversity Panel (HGDP), Ramachandran et al. (2004) investigated differences in patterns of X-chromosomal and autosomal geographical variation around the world, as measured by F_{ST} among populations. These differences were studied by considering the different numbers of copies of X-linked and autosomal loci in a population, for a given female fraction of the total population size, and by deriving a formula for F_{ST} using a model of divergence from an ancestral population with subsequent isolation of descendant populations. Male and female effective population sizes were allowed to vary, but the model did not involve migration among subpopulations. Ramachandran et al. (2004) found that a ratio of the number of females to the total population size of 0.5 was sufficient to explain global differences in genetic variation between X-linked and autosomal microsatellites. However, the study could not explain differences in F_{ST} in some of the continental regions of the dataset where the divergence model might be less representative of population history (for example, Europe, where gene flow among populations post-divergence is likely to have been high).

Here we investigate the rates of coalescence for X-linked and autosomal loci in an island migration model with sex-specific population sizes and migration rates. Past theoretical studies have examined the effect of sex-specific gene flow and genetic drift on genetic differentiation and F-statistics (Wang, 1997; Rousset, 1999; Wang, 1999; Laporte and Charlesworth, 2002; Vitalis, 2002; Hedrick, 2007). We consider these issues from a coalescent perspective. We start with an exact discrete island model with migrating adults, and use a result due to Möhle (1998) to explicitly take the limit of the coalescent process as population size goes to infinity. We obtain simple expressions for F_{ST} at X-linked and autosomal loci in our model under the usual assumptions of the structured coalescent.

Applying the analytical results to the X-linked and autosomal microsatellite data from the HGDP (Cann et al., 2002; Ramachandran et al., 2004, 2005; Rosenberg et al., 2005), we find that global patterns of population differentiation as measured by F_{ST} can be explained without requiring different migration rates for males and females. Within geographic regions, however, the inferred sex-specific migration rates differ substantially, although the direction of the deviation is not always the same.

2. The migration model

Consider an island model with D demes and four sex-specific parameters, each of which has the same value for all demes: fixed numbers of males and females (N_m and N_f, respectively), and fixed numbers of male and female migrants per generation (M_m and M_f, respectively). The total population size is $DN = D(N_m + N_f)$ (each deme has the same number of individuals). Here we can write $N_f = Nr$, where r is the female fraction of the population size, assumed to be the same for each deme. It follows that $N_m = N(1 - r)$. Denote by m_f the backwards migration rate for females; that is, the probability that a female sampled from deme i has just migrated from some other deme in the generation during which sampling took place. The corresponding rate for males is m_m. Since M_m and M_f are fixed, $m_f = M_f/N_f$ and $m_m = M_m/N_m$. We shall assume throughout that m_f and m_m are of the order $1/N$. Migration takes place after reproduction within demes, and the probability that a male (for example) migrates to a specific deme is $m_m/(D-1)$.

Table 1
States in the migration model

Column number in Boxes I and II		Definition
Autosomal	X-linked	
1	1	In one female individual, not coalesced
2		In one male individual, not coalesced
3	2	In two female individuals, same deme
4	3	In two male individuals, same deme
5	4	In one male and one female, same deme
6	5	In two female individuals, different demes
7	6	In two male individuals, different demes
8	7	In one male and one female, different demes
9	8	In one female, coalesced
10	9	In one male, coalesced

Possible states in which two sampled lineages can be found in the island model with two sexes, and the columns of the autosomal and X-linked single-generation transition matrices that correspond to each state. Note that two sampled X-linked lineages cannot be found in the same male unless they have already coalesced.

We consider a single genetic locus. The resulting single-generation transition matrix for a sample of two autosomal lineages in this model has 10 states. For a sample of two X-linked lineages the model has 9 states, as listed in Table 1.

Let $\mathbf{P}_A$ be the 10×10 single-generation transition matrix for two lineages sampled from an autosomal locus, and let $(\mathbf{P}_A)_{ij}$ refer to the entry in the ith row and jth column of the matrix. Each matrix entry is the product of two terms: (a) a term involving migration among demes or lack of migration, and (b) a term describing inheritance.

For example, $(\mathbf{P}_A)_{56}$, according to Table 1, is the entry describing the probability that two lineages sampled from one male and one female in the same deme came from female parents in different demes in the previous generation. $(\mathbf{P}_A)_{56}$ is the product of (a) the probability that one male and one female lineage currently in the same deme were in different demes in the previous generation (either because one lineage was in a migrant or because both lineages were in migrants that arrived in the same deme), and (b) the probability that two autosomal lineages (one from a male and one from a female) both came from female parents. The latter probability is 1/4, since for each sampled individual we choose the maternal autosome with probability 1/2.

$\mathbf{P}_X$ denotes the 9×9 single-generation transition matrix for two lineages sampled from an X-linked locus. $(\mathbf{P}_X)_{45}$ is the probability that two X-linked lineages sampled from one male and one female in the same deme came from female parents in different demes in the previous generation (Table 1). The probability (a) above, that the lineages were in different demes in the previous generation, will not differ between an X-linked and autosomal locus. However, the analog to (b) above, the probability that two X-linked lineages (one from a male and one from a female) came from two female parents is 1/2. This is because the male allele would have had to come from the female parent in the previous generation, while we choose the female's allele from her maternal X with probability 1/2.

The matrices $\mathbf{P}_A$ and $\mathbf{P}_X$ are rather cumbersome due to their size. Since the terms describing migration among demes do not depend on whether the sampled locus is X-linked or autosomal, the matrices' entries can be written more simply by using the notation $g_{i,j}^{k,l}$ for terms of type (a) above in the following manner. Let us denote the state in which two lineages, regardless of sex, are in the same deme as state I; state II represents two lineages from different demes. Then $g_{I,II}^{\mathcal{M},\mathcal{F}}$ is the probability that a sample of one male and one female now in state I was in state II in the previous generation, which corresponds to (a) in the previous paragraph. The probabilities $g_{i,j}^{k,l}$ for all types of samples are given in Appendix A.

Using this notation, for example, $(\mathbf{P}_A)_{39}$ is equal to the product of (a) $g_{I,I}^{\mathcal{F},\mathcal{F}}$ (the probability two females currently in the same

$$\mathbf{P}_A = \begin{array}{cc|ccc|ccc|cc}
\overbrace{}^{\text{Same individual}} & & \overbrace{}^{\text{Same deme}} & & & \overbrace{}^{\text{Different demes}} & & & \overbrace{}^{\text{Coalesced}} & \\
\mathcal{F} & \mathcal{M} & \mathcal{F},\mathcal{F} & \mathcal{M},\mathcal{M} & \mathcal{M},\mathcal{F} & \mathcal{F},\mathcal{F} & \mathcal{M},\mathcal{M} & \mathcal{M},\mathcal{F} & \mathcal{F} & \mathcal{M} \\
\hline
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\dfrac{g^{\mathcal{F},\mathcal{F}}_{I,I}}{8N_f} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{I,I}}{8N_m} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{I,I}}{4}\!\left(1-\dfrac{1}{N_f}\right) & \dfrac{g^{\mathcal{F},\mathcal{F}}_{I,I}}{4}\!\left(1-\dfrac{1}{N_m}\right) & \dfrac{g^{\mathcal{F},\mathcal{F}}_{I,I}}{2} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{I,II}}{4} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{I,II}}{4} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{I,II}}{2} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{I,I}}{8N_f} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{I,I}}{8N_m} \\
\dfrac{g^{\mathcal{M},\mathcal{M}}_{I,I}}{8N_f} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{I,I}}{8N_m} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{I,I}}{4}\!\left(1-\dfrac{1}{N_f}\right) & \dfrac{g^{\mathcal{M},\mathcal{M}}_{I,I}}{4}\!\left(1-\dfrac{1}{N_m}\right) & \dfrac{g^{\mathcal{M},\mathcal{M}}_{I,I}}{2} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{I,II}}{4} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{I,II}}{4} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{I,II}}{2} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{I,I}}{8N_f} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{I,I}}{8N_m} \\
\dfrac{g^{\mathcal{M},\mathcal{F}}_{I,I}}{8N_f} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{I,I}}{8N_m} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{I,I}}{4}\!\left(1-\dfrac{1}{N_f}\right) & \dfrac{g^{\mathcal{M},\mathcal{F}}_{I,I}}{4}\!\left(1-\dfrac{1}{N_m}\right) & \dfrac{g^{\mathcal{M},\mathcal{F}}_{I,I}}{2} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{I,II}}{4} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{I,II}}{4} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{I,II}}{2} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{I,I}}{8N_f} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{I,I}}{8N_m} \\
\dfrac{g^{\mathcal{F},\mathcal{F}}_{II,I}}{8N_f} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{II,I}}{8N_m} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{II,I}}{4}\!\left(1-\dfrac{1}{N_f}\right) & \dfrac{g^{\mathcal{F},\mathcal{F}}_{II,I}}{4}\!\left(1-\dfrac{1}{N_m}\right) & \dfrac{g^{\mathcal{F},\mathcal{F}}_{II,I}}{2} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{II,II}}{4} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{II,II}}{4} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{II,II}}{2} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{II,I}}{8N_f} & \dfrac{g^{\mathcal{F},\mathcal{F}}_{II,I}}{8N_m} \\
\dfrac{g^{\mathcal{M},\mathcal{M}}_{II,I}}{8N_f} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{II,I}}{8N_m} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{II,I}}{4}\!\left(1-\dfrac{1}{N_f}\right) & \dfrac{g^{\mathcal{M},\mathcal{M}}_{II,I}}{4}\!\left(1-\dfrac{1}{N_m}\right) & \dfrac{g^{\mathcal{M},\mathcal{M}}_{II,I}}{2} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{II,II}}{4} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{II,II}}{4} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{II,II}}{2} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{II,I}}{8N_f} & \dfrac{g^{\mathcal{M},\mathcal{M}}_{II,I}}{8N_m} \\
\dfrac{g^{\mathcal{M},\mathcal{F}}_{II,I}}{8N_f} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{II,I}}{8N_m} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{II,I}}{4}\!\left(1-\dfrac{1}{N_f}\right) & \dfrac{g^{\mathcal{M},\mathcal{F}}_{II,I}}{4}\!\left(1-\dfrac{1}{N_m}\right) & \dfrac{g^{\mathcal{M},\mathcal{F}}_{II,I}}{2} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{II,II}}{4} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{II,II}}{4} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{II,II}}{2} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{II,I}}{8N_f} & \dfrac{g^{\mathcal{M},\mathcal{F}}_{II,I}}{8N_m} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} & \tfrac{1}{2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{2} & \tfrac{1}{2}
\end{array}$$

Box I.

deme were in the same deme in the previous generation) and (b) $1/(8N_f)$ (the probability two sampled lineages, one from each sampled female, coalesce in a female in the previous generation). $1/(8N_f)$ is the probability that in both females the maternal autosome is selected ($=1/2 \times 1/2$) times the probability the loci were inherited from the same maternal chromosome ($=1/(2N_f)$). $(\mathbf{P}_X)_{62}$ is equal to (a) $g^{\mathcal{M},\mathcal{F}}_{II,I}$ (the probability two sampled males are currently in different demes but were in the same deme in the previous generation) times (b) $1 - 1/N_f$ (the probability the sampled lineages come from two different females). Since a male's X chromosome must come from his mother, the probability that two male X chromosomes are found in two different females is simply the probability the chromosomes do not come from the same female.

Suppose the sampled lineages are currently in the same individual but that the lineages have not coalesced (columns 1 and 2 in $\mathbf{P}_A$ and column 1 in $\mathbf{P}_X$). Since migration occurs after reproduction within demes, the lineages had to be in a male and female (the individual's parents) in the same deme in the previous generation, regardless of whether or not the individual from whom the lineages were sampled had migrated (see rows 1 and 2 of the matrix in Box I and row 1 of the matrix in Box II).

Thus we can write down both the autosomal and X-linked single-generation transition matrices, $\mathbf{P}_A$ and $\mathbf{P}_X$, as in Boxes I and II. Above both matrices, we indicate the sex structure of the sample for each column (e.g., $\mathcal{M}, \mathcal{M}$ denotes lineages sampled from two males), and the physical locations associated with states (e.g., in the same individual but not coalesced, or from different demes).

3. Results

We can rewrite both transition matrices in Boxes I and II in the form

$$\mathbf{P} = \mathbf{D} + \mathbf{B}/N + \mathbf{E}_N. \tag{1}$$

Assuming that M_f and M_m do not depend on N (i.e., as N approaches infinity, the numbers of migrants per generation converge to some limiting constants, which are again denoted by M_f and M_m for convenience), then $\mathbf{D} = \lim_{N\to\infty} \mathbf{P}$ and $\mathbf{B} = \lim_{N\to\infty} N(\mathbf{P} - \mathbf{D})$ (which both do not depend on N). Note that, in Eq. (1), $\mathbf{E}_N = \mathbf{P} - \mathbf{D} - \mathbf{B}/N$ denotes some error matrix with terms of the order of m^2, $1/N^2$, and m/N. See Appendix A for an example of this decomposition.

The entries in $\mathbf{D}$ represent a fast process, namely the movement of lineages between males and females according to Mendelian inheritance, while the entries in $\mathbf{B}$ represent rare processes of migration and coalescence which are assumed to occur once over a period on the order of N generations. Möhle's theorem (1998) states that if $\mathbf{R} = \lim_{t\to\infty} \mathbf{D}^t$ exists (letting the fast process run to its conclusion), then the rates of coalescence and migration among demes when time is scaled by N generations are given by the product matrix $\mathbf{G} = \mathbf{RBR}$. Specifically, $\lim_{N\to\infty} \mathbf{P}^{Nt} = \mathbf{R}e^{t\mathbf{G}}$ (Möhle, 1998).

We show $\mathbf{D}_X$ ($= \lim_{N\to\infty} \mathbf{P}_X$) and $\mathbf{R}_X$ in (2) and (3) below while the detailed derivations of the corresponding autosomal matrices and of $\mathbf{G}_X$ and $\mathbf{G}_A$ appear in Appendix B. In the case of $\mathbf{P}_X$ given by the matrix in Box II, $\mathbf{D}_X = \lim_{N\to\infty} \mathbf{P}_X$ is

$$\mathbf{D}_X = \begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/4 & 1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}. \tag{2}$$

The columns in matrix (2) can be interpreted using the definitions in Table 1. The terms in $\mathbf{D}_X$ are familiar terms based on the inheritance of X chromosomes, as are the entries of $\mathbf{R}_X = \lim_{t\to\infty} (\mathbf{D}_X)^t$:

$$\mathbf{R}_X = \begin{bmatrix}
0 & 4/9 & 1/9 & 4/9 & 0 & 0 & 0 & 0 & 0 \\
0 & 4/9 & 1/9 & 4/9 & 0 & 0 & 0 & 0 & 0 \\
0 & 4/9 & 1/9 & 4/9 & 0 & 0 & 0 & 0 & 0 \\
0 & 4/9 & 1/9 & 4/9 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 4/9 & 1/9 & 4/9 & 0 & 0 \\
0 & 0 & 0 & 0 & 4/9 & 1/9 & 4/9 & 0 & 0 \\
0 & 0 & 0 & 0 & 4/9 & 1/9 & 4/9 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 2/3 & 1/3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 2/3 & 1/3
\end{bmatrix}. \tag{3}$$

When applying Möhle's result to $\mathbf{P}_A$ and $\mathbf{P}_X$, a block structure emerges in the $\mathbf{R}$ and $\mathbf{G}$ matrices for both X-linked and autosomal loci, exemplified by the blocks seen in matrix (3). We can collapse some states together by summing the entries in their columns and by collapsing some rows, reducing the analysis to 3×3 matrices.

Box II.

$$\mathbf{P}_X =$$

Same female	Same deme			Different demes			Coalesced	
$\mathcal{F}$	$\mathcal{F},\mathcal{F}$	$\mathcal{M},\mathcal{M}$	$\mathcal{M},\mathcal{F}$	$\mathcal{F},\mathcal{F}$	$\mathcal{M},\mathcal{M}$	$\mathcal{M},\mathcal{F}$	$\mathcal{F}$	$\mathcal{M}$
0	0	0	1	0	0	0	0	0
$\frac{g_{I,I}^{\mathcal{F},\mathcal{F}}}{8N_f}$	$\frac{g_{I,I}^{\mathcal{F},\mathcal{F}}}{4}\left(1-\frac{1}{N_f}\right)$	$\frac{g_{I,I}^{\mathcal{F},\mathcal{F}}}{4}\left(1-\frac{1}{N_m}\right)$	$\frac{g_{I,I}^{\mathcal{F},\mathcal{F}}}{2}$	$\frac{g_{I,II}^{\mathcal{F},\mathcal{F}}}{4}$	$\frac{g_{I,II}^{\mathcal{F},\mathcal{F}}}{4}$	$\frac{g_{I,II}^{\mathcal{F},\mathcal{F}}}{2}$	$\frac{g_{I,I}^{\mathcal{F},\mathcal{F}}}{8N_f}$	$\frac{g_{I,I}^{\mathcal{F},\mathcal{F}}}{4N_m}$
$\frac{g_{I,I}^{\mathcal{M},\mathcal{M}}}{2N_f}$	$g_{I,I}^{\mathcal{M},\mathcal{M}}\left(1-\frac{1}{N_f}\right)$	0	0	$g_{I,II}^{\mathcal{M},\mathcal{M}}$	0	0	$\frac{g_{I,I}^{\mathcal{M},\mathcal{M}}}{2N_f}$	0
$\frac{g_{I,I}^{\mathcal{M},\mathcal{F}}}{4N_f}$	$\frac{g_{I,I}^{\mathcal{M},\mathcal{F}}}{2}\left(1-\frac{1}{N_f}\right)$	0	$\frac{g_{I,I}^{\mathcal{M},\mathcal{F}}}{2}$	$\frac{g_{I,II}^{\mathcal{M},\mathcal{F}}}{2}$	0	$\frac{g_{I,II}^{\mathcal{M},\mathcal{F}}}{2}$	$\frac{g_{I,I}^{\mathcal{M},\mathcal{F}}}{4N_f}$	0
$\frac{g_{II,I}^{\mathcal{F},\mathcal{F}}}{8N_f}$	$\frac{g_{II,I}^{\mathcal{F},\mathcal{F}}}{4}\left(1-\frac{1}{N_f}\right)$	$\frac{g_{II,I}^{\mathcal{F},\mathcal{F}}}{4}\left(1-\frac{1}{N_m}\right)$	$\frac{g_{II,I}^{\mathcal{F},\mathcal{F}}}{2}$	$\frac{g_{II,II}^{\mathcal{F},\mathcal{F}}}{4}$	$\frac{g_{II,II}^{\mathcal{F},\mathcal{F}}}{4}$	$\frac{g_{II,II}^{\mathcal{F},\mathcal{F}}}{2}$	$\frac{g_{II,I}^{\mathcal{F},\mathcal{F}}}{8N_f}$	$\frac{g_{II,I}^{\mathcal{F},\mathcal{F}}}{4N_m}$
$\frac{g_{II,I}^{\mathcal{M},\mathcal{M}}}{2N_f}$	$g_{II,I}^{\mathcal{M},\mathcal{M}}\left(1-\frac{1}{N_f}\right)$	0	0	$g_{II,II}^{\mathcal{M},\mathcal{M}}$	0	0	$\frac{g_{II,I}^{\mathcal{M},\mathcal{M}}}{2N_f}$	0
$\frac{g_{II,I}^{\mathcal{M},\mathcal{F}}}{4N_f}$	$\frac{g_{II,I}^{\mathcal{M},\mathcal{F}}}{2}\left(1-\frac{1}{N_f}\right)$	0	$\frac{g_{II,I}^{\mathcal{M},\mathcal{F}}}{2}$	$\frac{g_{II,II}^{\mathcal{M},\mathcal{F}}}{2}$	0	$\frac{g_{II,II}^{\mathcal{M},\mathcal{F}}}{2}$	$\frac{g_{II,I}^{\mathcal{M},\mathcal{F}}}{4N_f}$	0
0	0	0	0	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$
0	0	0	0	0	0	0	1	0

Box II.

$$\mathcal{g}_X = \begin{bmatrix} -\frac{2}{3}\left(\frac{2M_f}{r}+\frac{M_m}{1-r}\right)-\frac{2-r}{9r(1-r)} & \frac{2}{3}\left(\frac{2M_f}{r}+\frac{M_m}{1-r}\right) & \frac{2-r}{9r(1-r)} \\ \frac{2}{3(D-1)}\left(\frac{2M_f}{r}+\frac{M_m}{1-r}\right) & -\frac{2}{3(D-1)}\left(\frac{2M_f}{r}+\frac{M_m}{1-r}\right) & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(columns: Same deme, Different demes, Coalesced)

Box III.

$$\mathcal{g}_A = \begin{bmatrix} -\left(\frac{M_f}{r}+\frac{M_m}{1-r}\right)-\frac{1}{8r(1-r)} & \frac{M_f}{r}+\frac{M_m}{1-r} & \frac{1}{8r(1-r)} \\ \frac{1}{D-1}\left(\frac{M_f}{r}+\frac{M_m}{1-r}\right) & -\frac{1}{D-1}\left(\frac{M_f}{r}+\frac{M_m}{1-r}\right) & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(columns: Same deme, Different demes, Coalesced)

Box IV.

For example, we can sum the entries $\sum_{j=1}^4 (\mathbf{G}_X)_{1j}$ (see Appendix B) and get a single rate of staying in the same deme for two lineages sampled in the same female individual, but not coalesced (the state described by row and column 1 of $\mathbf{P}_X$). The sum $\sum_{j=1}^4 (\mathbf{G}_X)_{ij}$ has the same value for each $i = 1, 2, 3, 4$. This is because in the fast process occurring according to $\mathbf{D}_X$ and $\mathbf{D}_A$, lineages move quickly between males and females, so the current sex structure of the sample becomes unimportant and instead we need only follow whether sampled lineages are in the same deme or in different demes. Thus, the product matrices $\mathbf{G}_X$ and $\mathbf{G}_A$ for X-linked and autosomal lineages in this process simplify to $\mathcal{g}_X$ and $\mathcal{g}_A$ (equations in Boxes III and IV, respectively; see Appendix B for the derivation).

Using first-step analysis, we can calculate the expected times to coalescence for a pair of lineages, sampled from the same deme ($E[T^{\text{same}}]$) or sampled from different demes ($E[T^{\text{diff}}]$). In the discrete-time processes studied here, the expected time to arrive in state j given that the current state is i equals the time to make the jump from state i to another state plus the expected time it takes to reach state j after the jump is made. We are interested in time to coalescence, so we need to solve the following equations to get,

for example, expected times to coalescence for a pair of autosomal lineages, $E[T_A^{\text{same}}]$ and $E[T_A^{\text{diff}}]$:

$$E[T_A^{\text{same}}] = \frac{1}{(\mathcal{g}_A)_{12}+(\mathcal{g}_A)_{13}} + \frac{(\mathcal{g}_A)_{12}}{(\mathcal{g}_A)_{12}+(\mathcal{g}_A)_{13}}E[T_A^{\text{diff}}]$$

$$E[T_A^{\text{diff}}] = \frac{1}{(\mathcal{g}_A)_{21}} + E[T_A^{\text{same}}].$$

Solving these and the analogous equations for X-linked loci gives Eqs. (4)–(7), measured in units of N generations.

$$E[T_A^{\text{same}}] = 8Dr(1-r) \tag{4}$$

$$E[T_A^{\text{diff}}] = 8Dr(1-r) + \frac{(D-1)(1-r)r}{M_f(1-r)+M_m r} \tag{5}$$

$$E[T_X^{\text{same}}] = \frac{9Dr(1-r)}{2-r} \tag{6}$$

$$E[T_X^{\text{diff}}] = \frac{9Dr(1-r)}{2-r} + \frac{3}{2}\left(\frac{(D-1)(1-r)r}{2M_f(1-r)+M_m r}\right). \tag{7}$$

Using our notation, Slatkin's (1991) formulation of F_{ST} at an autosomal locus in a set of D demes is $F_{ST,A} = 1 -$
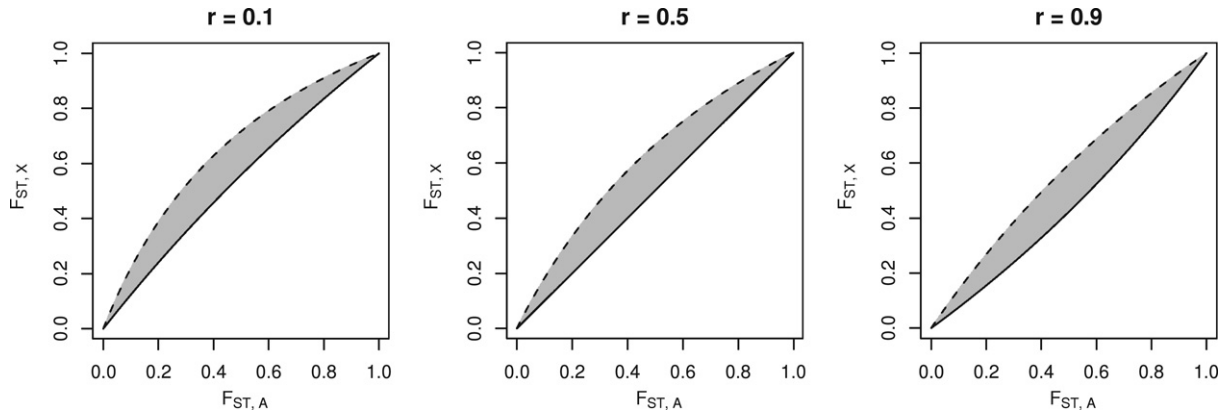
Fig. 1. The region in which the ratio M_f/M_m is positive, as computed from Eqs. (10) and (11) for fixed values of r, with $F_{ST,X}$ and $F_{ST,A}$ varying on the interval [0, 1]. The region is shaded in grey. The solid line is $2F_{ST,A}(2-r)/[3+F_{ST,A}(1-2r)]$, which $F_{ST,X}$ must exceed for M_m to be greater than zero. The dashed line is $4F_{ST,A}(2-r)/[3+F_{ST,A}(5-r)]$, which $F_{ST,X}$ must be less than for M_f to exceed zero.

$E[T_A^{\text{same}}]/\{E[T_A^{\text{same}}]/D + (D - 1)E[T_A^{\text{diff}}]/D\}$. The relationship between coalescence times and F_{ST} in this formulation depends on the mutation rate being very small. As D approaches infinity, we get

$$F_{ST,A} = \frac{1}{1 + 8[M_f(1-r) + M_m r]} \tag{8}$$

$$F_{ST,X} = \frac{1}{1 + \frac{6}{2-r}[2M_f(1-r) + M_m r]}. \tag{9}$$

Given estimates of F_{ST} at X-linked and autosomal loci, and assuming some value on the interval (0, 1) for r, we can estimate M_f and M_m from (8) and (9) as:

$$M_f = \frac{1}{(1-r)}\left[\frac{2-r}{6F_{ST,X}} - \frac{1}{8}\left(\frac{1}{F_{ST,A}} + \frac{1}{3}\right) - \frac{1-r}{6}\right] \tag{10}$$

$$M_m = \frac{1}{r}\left[-\frac{2-r}{6F_{ST,X}} + \frac{1}{4}\left(\frac{1}{F_{ST,A}} + \frac{1}{3}\right) - \frac{r}{6}\right]. \tag{11}$$

Application to HGDP-CEPH data

A total of 783 autosomal microsatellites from Marshfield Screening Sets #10 and #52 have been reported in the HGDP individuals from 52 populations. Screening Set #10 also contained the 17 non-pseudoautosomal X-linked microsatellites studied by Ramachandran et al. (2004), and Screening Set #52 provided 19 additional non-pseudoautosomal X-linked microsatellites studied here. The data files used in this analysis are available from the authors.

We inferred the sex of individuals from their X-chromosomal genotypes at the 36 loci examined, and we verified the inferences against the corresponding inferences made using the X-chromosomal data of Conrad et al. (2006). With one exception, individuals treated as males in our analysis all had <15% heterozygous loci and females all had >19% heterozygous loci on the X chromosome, among loci with no missing data. The exception, individual #139, was verified to be male on the basis of the data of Conrad et al. (2006), which included a larger number of X-chromosomal loci. Males were treated as hemizygous for calculations. Some males were reported as heterozygous at non-pseudoautosomal X-linked loci; in such cases males were coded as having missing data at these loci.

Since the initial announcement of the HGDP (Cann et al., 2002), subsequent analyses have called attention to individuals who appear to be duplicated or closely related. Here we calculate F_{ST} for two sets of HGDP individuals (Tables 2 and 3): 1048 individuals, where one individual from each pair of putatively duplicated

individuals (Mountain and Ramakrishnan, 2005; Rosenberg, 2006) is excluded; and 952 individuals, a proper subset of the set of 1048, where individuals with first- and second-degree relationships are excluded (Rosenberg, 2006).

We calculated F_{ST} based on the 36 X-linked and 783 autosomal microsatellites typed in the Human Genome Diversity Panel, using Weir's estimator (Weir, 1996) for the proportion of genetic variation distributed among populations. F_{ST} was calculated among all populations, as well as among populations within the same continental region, as defined previously by Rosenberg et al. (2002); the estimator was obtained separately for X-linked loci and for autosomal loci, following equation (5.3) on page 174 of Weir (1996). For the computation we grouped all Bantu individuals into one population with a sample size of 20 individuals. We obtained confidence intervals for X-linked and autosomal F_{ST} values by bootstrapping separately over each set of loci 1000 times (see intervals in Tables 2 and 3).

We employ Eqs. (10) and (11) to estimate the ratio of female migrants to male migrants using observed F_{ST} values from the data, for a given assumed proportion of females in the population. Note that in order for M_f and M_m to be interpretable they must be positive, which may not be the case for certain combinations of F_{ST} and r values. In order for both M_f and M_m to be greater than zero, the condition $2F_{ST,A}(2-r)/[3+F_{ST,A}(1-2r)] < F_{ST,X} < 4F_{ST,A}(2-r)/[3+F_{ST,A}(5-r)]$ must be satisfied. The region in which M_f/M_m is positive for various fixed values of r, as $F_{ST,X}$ and $F_{ST,A}$ vary on the interval [0, 1], is shown in Fig. 1.

We obtained intervals for M_f/M_m (Tables 2 and 3) by taking the 1000 bootstrapped $F_{ST,X}$ and 1000 bootstrapped $F_{ST,A}$ values, and computing M_f/M_m for all 10^6 possible pairs of bootstrapped F_{ST} values. We disregarded those estimates of M_f/M_m which were negative, choosing to interpret negative estimates of M_f and M_m as providing little support for the assumed r-value or for our migration model. The number of values used to generate the intervals in Tables 2 and 3 after the exclusion of negative estimates is also given.

4. Discussion

In this paper, we apply Möhle's theorem (1998) to transition matrices for X-linked and autosomal loci sampled in an island model of D demes with sex-specific population sizes and migration rates, and we obtain simple expressions under the model for expected times to coalescence for two sampled alleles and for F_{ST} at X-linked and autosomal loci. Möhle's result is useful because it gives us a continuous-time limit of a discrete-time process where events are occurring on two time scales: in this case, the fast

Table 2
Estimated ratio of M_f/M_m, using data from 1048 individuals

Sample	Number of populations	F_{ST}, autosomal (95% C.I.)			F_{ST}, X-linked (95% C.I.)			M_f/M_m at $r = 0.5$ (95% C.I.)			Number of ratios ≥ 0
World	52	0.0561	(0.0543,	0.0579)	0.0718	(0.0620,	0.0839)	1.1524	(0.4149,	4.4257)	999535
Africa	6	0.0300	(0.0286,	0.0314)	0.0539	(0.0401,	0.0691)	0.0936	(0.0078,	1.1324)	716775
Eurasia	21	0.0158	(0.0152,	0.0165)	0.0226	(0.0183,	0.0276)	0.6345	(0.1519,	2.7184)	998081
Europe	8	0.0079	(0.0071,	0.0087)	0.0122	(0.0069,	0.0180)	0.4127	(0.0208,	9.1672)	818855
Middle East	4	0.0137	(0.0130,	0.0145)	0.0162	(0.0121,	0.0208)	2.1807	(0.4004,	32.2857)	852251
C/S Asia	9	0.0137	(0.0130,	0.0145)	0.0149	(0.0095,	0.0208)	4.9120	(0.3772,	59.5790)	636089
East Asia	18	0.0125	(0.0117,	0.0134)	0.0156	(0.0102,	0.0215)	1.4720	(0.1563,	23.0857)	865762
Oceania	2	0.0635	(0.0577,	0.0692)	0.0847	(0.0544,	0.1220)	0.8746	(0.0542,	18.3075)	863700
America	5	0.1174	(0.1127,	0.1219)	0.1367	(0.1166,	0.1567)	2.1349	(0.7401,	19.5850)	964990

Estimates of the among-population component of genetic variation based on 783 autosomal and 36 X-linked microsatellites in the HGDP-CEPH individuals, for global and various regional subsets of the data. Also reported are the estimated ratios of M_f/M_m when $r = 0.5$, calculated using 1048 individuals (Rosenberg, 2006). "C/S Asia" refers to Central/South Asian populations from the panel (see Rosenberg et al. (2002)). Confidence intervals for F_{ST} were obtained by bootstrapping over loci 1000 times. The number of values out of 10^6 used to generate intervals for M_f/M_m, after the exclusion of negative estimates, is given in the last column.

Table 3
Estimated ratio of M_f/M_m, using data from 952 individuals

Sample	Number of populations	F_{ST}, autosomal (95% C.I.)			F_{ST}, X-linked (95% C.I.)			M_f/M_m at $r = 0.5$ (95% C.I.)			Number of ratios ≥ 0
World	52	0.0455	(0.0438,	0.0472)	0.0586	(0.0491,	0.0702)	1.1354	(0.3530,	5.2649)	992554
Africa	6	0.0260	(0.0245,	0.0274)	0.0465	(0.0338,	0.0611)	0.1026	(0.0087,	1.2628)	720323
Eurasia	21	0.0150	(0.0144,	0.0156)	0.0218	(0.0172,	0.0266)	0.5896	(0.1265,	2.8729)	994778
Europe	8	0.0076	(0.0069,	0.0084)	0.0112	(0.0061,	0.0167)	0.5722	(0.0262,	15.7893)	808661
Middle East	4	0.0130	(0.0121,	0.0137)	0.0150	(0.0111,	0.0194)	2.5593	(0.4367,	40.5666)	817864
C/S Asia	9	0.0127	(0.0119,	0.0134)	0.0146	(0.0089,	0.0205)	2.7676	(0.2370,	34.5307)	719969
East Asia	18	0.0113	(0.0105,	0.0121)	0.0134	(0.0081,	0.0190)	2.1811	(0.1713,	35.2105)	745435
Oceania	2	0.0552	(0.0493,	0.0616)	0.0753	(0.0410,	0.1155)	0.7702	(0.0320,	19.1807)	766059
America	5	0.0836	(0.0799,	0.0876)	0.0942	(0.0789,	0.1087)	3.0884	(0.9013,	32.3565)	909730

This table reports a similar analysis to that reported in Table 2, but using 952 individuals (Rosenberg, 2006). This set of HGDP individuals contains no two individuals with a second-degree relationship (half siblings, avuncular, or grandparent/grandchild).

process of movement of lineages between males and females, and the slow processes of movement of individuals among demes and of coalescence.

The entries in the matrices in Boxes III and IV give us the rates at which, when time is measured in units of N generations, two sampled lineages move among three states: being in the same deme, being in different demes, or coalescing. $(\mathcal{G}_A)_{12}$ gives the rate (over N generations) at which autosomal lineages move out of the same deme into different demes, $(\mathcal{G}_A)_{21}$ gives the rate of movements of lineages into the same deme from different demes, and $(\mathcal{G}_A)_{13}$ gives the rate of coalescence, which can only happen in the model when lineages are in the same deme. For both $\mathcal{G}_A$ and $\mathcal{G}_X$ the last row contains only zeros because coalescence is an absorbing state.

The rates of coalescence given by $(\mathcal{G}_A)_{13}$ and $(\mathcal{G}_X)_{13}$ are familiar: they are half the reciprocals of the variance effective population size of autosomal and X-linked genes in a sexual population with an unequal sex ratio (e.g., Nordborg and Krone (2002) and Hartl and Clark (2007)). The expected times to coalescence given in Eqs. (4)–(7) also reflect that two lineages sampled from different demes must enter the same deme to coalesce, and then coalesce at a rate expected for an X-linked or autosomal locus in a population with an unequal sex ratio.

Using the expected times to coalescence for loci sampled in the same deme or in different demes, we can calculate F_{ST} at autosomal and X-linked loci in our model, as in Eqs. (8) and (9). The forms of (8) and (9) are $1/(1 + 4N_e m_e)$, where $N_e = 4N_m N_f/(N_m + N_f) = 4Nr(1 - r)$ and $m_e = (m_m + m_f)/2$ for autosomal loci, and $N_e = 9N_m N_f/(4N_m + 2N_f) = 9Nr(1 - r)/[2(2 - r)]$ and $m_e = (2m_f + m_m)/3$ for X-linked loci.

When $M_m = M_f = M$, then $F_{ST,A} = 1/(1 + 8M)$ and $F_{ST,X} = 1/(1 + 6M)$, with $F_{ST,X}$ being greater than $F_{ST,A}$. If $M_f/M_m > 1$, then $F_{ST,X}$ can become less than $F_{ST,A}$ for some values of r on $(0, 1)$, as shown in Fig. 2(A), where $F_{ST,X}$ crosses $F_{ST,A}$ at $r = 0.5076$ and $r = 0.9949$. If the number of male migrants exceeds the number

of female migrants per generation, then $F_{ST,X}$ can exceed $F_{ST,A}$; in Fig. 2(B) these values become closer for larger values of r.
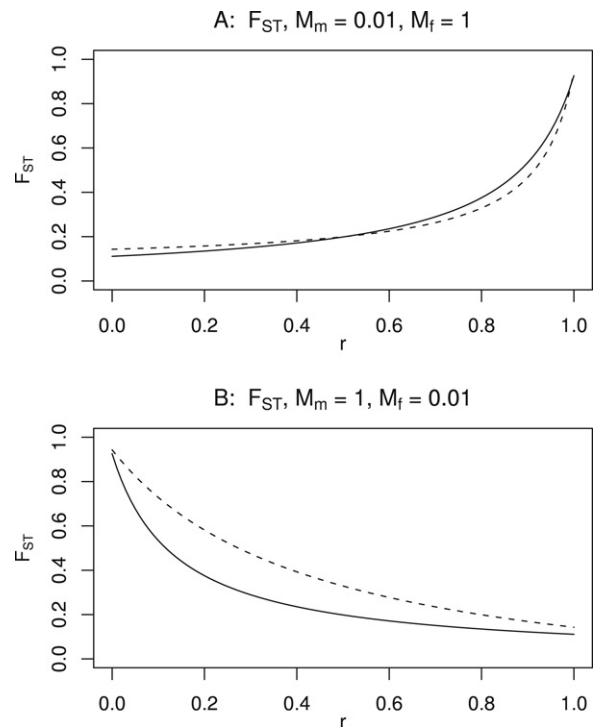

Fig. 2. F_{ST} at X-linked and autosomal loci (Eqs. (8) and (9)) as r, the female fraction of the population, varies on $(0, 1)$. The dashed line is $F_{ST,X}$ and the solid line is $F_{ST,A}$. A: $M_f = N_f m_f = 1$ migrant per generation, while the number of male migrants per generation is 0.01. B: M_m, the number of male migrants per generation, is equal to 1, while $M_f = 0.01$.
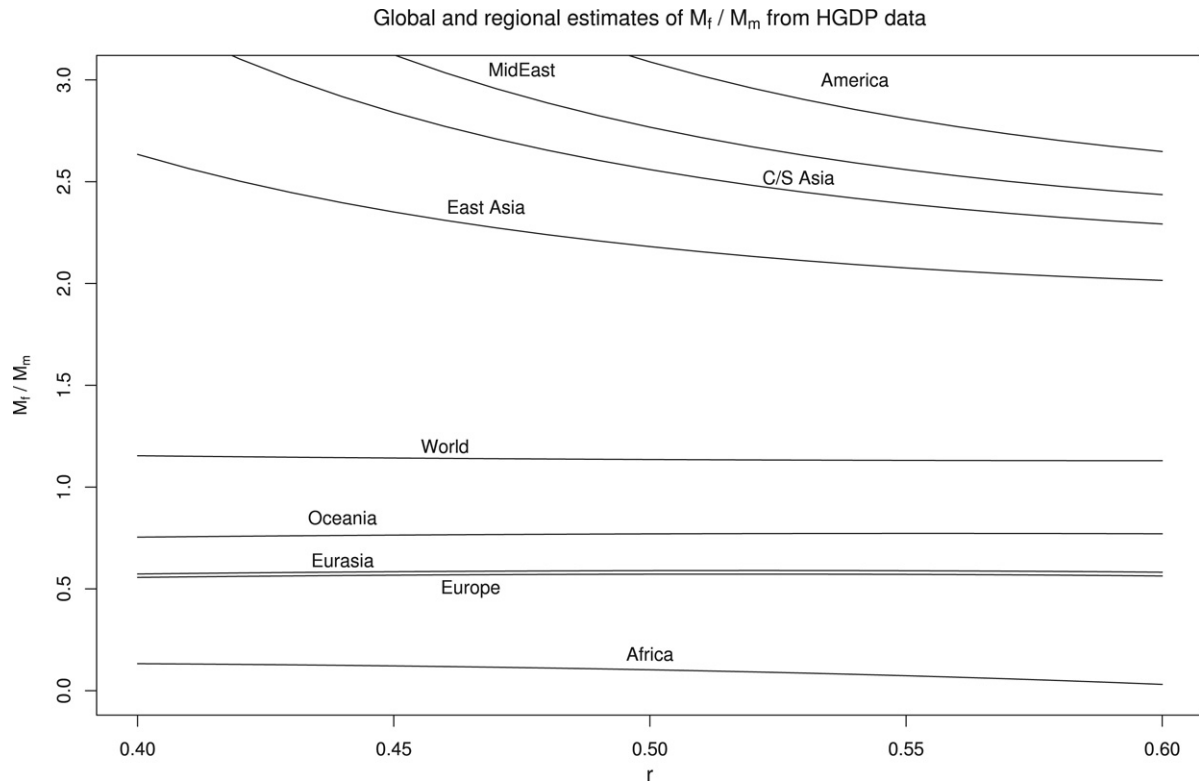
Global and regional estimates of M_f / M_m from HGDP data

Fig. 3. Global and regional estimates of the ratio of female to male migrants (the ratio of Eq. (10) to Eq. (11)) as r varies over [0.4, 0.6], based on F_{ST} values calculated using 952 individuals from the Human Genome Diversity Panel (see Table 3). "MidEast" denotes the Middle East; "C/S Asia" denotes "Central/South Asia".

Using observed values of F_{ST} from the Human Genome Diversity Panel at X-linked and autosomal loci, we then use Eqs. (10) and (11) to estimate the ratio of female to male migrants M_f/M_m. We assume equal numbers of males and females when calculating the estimates in Tables 2 and 3. In this model, there are no differences between the rates of reproductive success in males and females. However, the consequence of differences between reproductive success in males and females is an important question for further investigation (see, for example, Helgason et al. (2003) and Wilder et al. (2004b)).

When $r = 0.5$, global F_{ST} values across HGDP populations can be explained by requiring a ratio of female to male migrants only slightly larger than 1. Regional values vary a great deal, both when $r = 0.5$ and when r varies over [0.4, 0.6] (Fig. 3). In the Middle East, Central/South Asia, East Asia, and the Americas, assuming $r = 0.5$, observed F_{ST} values require a greater number of female migrants than male migrants to be explained in our model, while in Africa, Europe, and Oceania, the analysis finds support for more male migrants than female migrants. This could be due to differences in reproductive success for males and females in these regions, or to some other assumption made in our model.

Although we are not able to draw strong empirical conclusions from these data, we have rigorously derived F_{ST} and expected times to coalescence under this new model, making explicit how population differentiation, as measured by F_{ST}, depends on the number of males and females in a population and on the migration rates of the sexes. A scenario with males migrating more than females (Fig. 2(B)) creates a bigger discrepancy between $F_{ST,A}$ and $F_{ST,X}$

than the reverse situation, producing differences that are much larger than those observed between F_{ST} values for the autosomes and for the X chromosome in the HGDP dataset. In combination with other tools, our results may assist in further investigations of the contributions of males and females to the history of human migration.

Acknowledgments

Appendix A. The migration components of transition matrix entries

Recall $g_{i,j}^{\mathcal{M},\mathcal{M}}$ is the probability that a sample of two males are currently found in i number of demes (i = I, II) and were found in j number of demes (j = I, II) in the previous generation; states i and j refer to whether the sampled lineages were in the same deme (denoted as I) or different demes (II). For any sample, $g_{i,j}$ will depend on the backwards migration rate and the population sizes of males (and/or females, depending on the individuals from which lineages were sampled), but will not depend on whether the sampled locus is X-linked or autosomal. Note that sampling of individuals is done without replacement. Thus, $g_{i,j}$ for a sample of two males is given below:

$$g_{\mathrm{I,I}}^{\mathcal{M},\mathcal{M}} = \overbrace{\frac{N_m(1-m_m)}{N_m}\left(\frac{N_m(1-m_m)-1}{N_m-1}\right)}^{\text{neither lineage was in a migrant}}$$

$$+ \overbrace{\left(\frac{N_m m_m}{N_m}\right)\frac{N_m m_m - 1}{N_m - 1}\left(\frac{1}{D-1}\right)}^{\substack{\text{both lineages were in}\\\text{migrants from the same deme}}}$$

$$g_{\mathrm{I,II}}^{\mathcal{M},\mathcal{M}} = \overbrace{\frac{2N_m(1-m_m)}{N_m}\left(\frac{N_m m_m}{N_m-1}\right)}^{\text{exactly one lineage was in a migrant}}$$

$$+ \overbrace{\left(\frac{N_m m_m}{N_m}\right)\frac{N_m m_m - 1}{N_m - 1}\left(\frac{D-2}{D-1}\right)}^{\substack{\text{both lineages were in}\\\text{migrants from different demes}}}$$

$$g_{\mathrm{II,I}}^{\mathcal{M},\mathcal{M}} = \frac{2m_m(1-m_m)}{D-1} + \frac{m_m^2(D-2)}{(D-1)^2}$$

$$g_{\mathrm{II,II}}^{\mathcal{M},\mathcal{M}} = (1-m_m)^2 + \frac{2m_m(1-m_m)(D-2)}{D-1}$$
$$+ m_m^2\left[\frac{1}{D-1} + \left(\frac{D-2}{D-1}\right)^2\right].$$

The corresponding probabilities for a sample of two females are obtained by substituting m_f and N_f for m_m and N_m, respectively, in the equations above.

For a sample with one male lineage and one female lineage:

$$g_{\mathrm{I,I}}^{\mathcal{M},\mathcal{F}} = (1-m_m)(1-m_f) + \frac{m_f m_m}{D-1}$$

$$g_{\mathrm{I,II}}^{\mathcal{M},\mathcal{F}} = m_f(1-m_m) + m_m(1-m_f) + m_f m_m\frac{D-2}{D-1}$$

$$g_{\mathrm{II,I}}^{\mathcal{M},\mathcal{F}} = \frac{1}{D-1}\left[m_f(1-m_m) + m_m(1-m_f)\right] + \frac{m_f m_m}{D-1}\frac{D-2}{D-1}$$

$$g_{\mathrm{II,II}}^{\mathcal{M},\mathcal{F}} = (1-m_m)(1-m_f)$$
$$+ \frac{D-2}{D-1}\left[m_f(1-m_m) + m_m(1-m_f)\right]$$
$$+ m_f m_m\left[\frac{1}{D-1} + \left(\frac{D-2}{D-1}\right)^2\right].$$

Note that $g_{i,\mathrm{I}}^{k,l} + g_{i,\mathrm{II}}^{k,l}$ (the probability alleles sampled from two individuals with sexes k and l in state i in the present were in individuals in either the same or different demes in the previous generation) is 1 for all i, k, l.

To give an example of the decomposition of terms in the matrices in Boxes I and II according to Eq. (1), let us examine $(\mathbf{P}_A)_{35}$ more closely. $(\mathbf{P}_A)_{35}$ is the probability that two autosomal alleles sampled from two females in the same deme in the present were in one male and one female in the same deme one generation ago.

$$(\mathbf{P}_A)_{35} = \frac{g_{\mathrm{I,I}}^{\mathcal{F},\mathcal{F}}}{2}$$
$$= \frac{1}{2}\left[\frac{N_f(1-m_f)}{N_f}\left(\frac{N_f(1-m_f)-1}{N_f-1}\right)\right.$$
$$\left. + \left(\frac{N_f m_f}{N_f}\right)\frac{N_f m_f - 1}{N_f - 1}\left(\frac{1}{D-1}\right)\right].$$

Substituting $m_f = M_f/N_f = M_f/(Nr)$,

$$(\mathbf{D}_A)_{35} = \lim_{N\to\infty}(\mathbf{P}_A)_{35}$$
$$= \lim_{N\to\infty}\frac{g_{\mathrm{I,I}}^{\mathcal{F},\mathcal{F}}}{2}$$
$$= \lim_{N\to\infty}\frac{1}{2}\left[\left(1-\frac{M_f}{Nr}\right)\left(\frac{Nr[1-\frac{M_f}{Nr}]-1}{Nr-1}\right)\right]$$
$$+ \lim_{N\to\infty}\overbrace{\frac{M_f}{2Nr}\left(\frac{M_f-1}{Nr-1}\right)\frac{1}{D-1}}^{\text{goes to 0 as } N\to\infty}$$
$$= \lim_{N\to\infty}\frac{1}{2}\left(1-\frac{M_f}{Nr}\right)\left(\frac{Nr-M_f-1}{Nr-1}\right) = \frac{1}{2}$$

(see $(\mathbf{D}_A)_{35}$ in Appendix B).

Using the definition of $\mathbf{B}_A$ from Eq. (1), we get

$$(\mathbf{B}_A)_{35} = \lim_{N\to\infty}N\left(\frac{g_{\mathrm{I,I}}^{\mathcal{F},\mathcal{F}}}{2} - (\mathbf{D}_A)_{35}\right)$$
$$= \lim_{N\to\infty}N\left(\frac{g_{\mathrm{I,I}}^{\mathcal{F},\mathcal{F}}}{2} - \frac{1}{2}\right)$$
$$= \lim_{N\to\infty}\frac{N}{2}\left[\left(1-\frac{M_f}{Nr}\right)\left[\frac{Nr(1-\frac{M_f}{Nr})-1}{Nr-1}\right]\right.$$
$$\left. + \frac{M_f}{Nr}\left(\frac{M_f-1}{Nr-1}\right)\frac{1}{D-1} - 1\right]$$
$$= \lim_{N\to\infty}\left(\frac{Nr-M_f}{2r}\left[\frac{Nr-M_f-1}{Nr-1}\right]\right.$$
$$+ \overbrace{\frac{M_f}{2r}\left(\frac{M_f-1}{Nr-1}\right)\frac{1}{D-1}}^{\text{goes to 0 as } N\to\infty} - \frac{N}{2}\Bigg)$$
$$= \lim_{N\to\infty}\frac{(Nr-M_f)^2 - (Nr-M_f) - Nr(Nr-1)}{2r(Nr-1)}$$
$$= \lim_{N\to\infty}\frac{(Nr)^2 - 2M_f Nr + M_f^2 - \cancel{Nr} + M_f - \cancel{(Nr)^2} + \cancel{Nr}}{2r(Nr-1)}$$
$$= \lim_{N\to\infty}\frac{-2M_f Nr}{2r(Nr-1)} + \lim_{N\to\infty}\frac{M_f^2 + M_f}{2r(Nr-1)}. \tag{12}$$

The second term of the right-hand side of Eq. (12) above goes to 0 as $N\to\infty$; using L'Hospital's Rule on the first term, then

$$(\mathbf{B}_A)_{35} = \lim_{N\to\infty}\frac{-2M_f Nr}{2r(Nr-1)}$$
$$= \lim_{N\to\infty}\frac{-2M_f r}{2r^2} = \frac{-M_f}{r} \quad \text{(see } (\mathbf{B}_A)_{35} \text{ in Appendix B).}$$

Let $\mathbf{E}_{N,A}$ denote the error matrix $\mathbf{E}_N$ from Eq. (1) of the autosomal transition matrix $\mathbf{P}_A$. Then

$$(\mathbf{E}_{N,A})_{35} = (\mathbf{P}_A)_{35} - (\mathbf{D}_A)_{35} - \frac{(\mathbf{B}_A)_{35}}{N}$$
$$= \frac{g_{\mathrm{I,I}}^{\mathcal{F},\mathcal{F}}}{2} - \frac{1}{2} + \frac{M_f}{Nr}$$
$$= \frac{D(M_f-1)M_f}{2(D-1)Nr(Nr-1)} \xrightarrow[N\to\infty]{} 0.$$

Appendix B. The derivation of G_X and G_A using Möhle's (1998) result

From Eq. (1), as N approaches infinity, $\mathbf{B}_X$ can be written as in Box V. Using the equation in Box V and $\mathbf{R}_X$ given in matrix (3), the product matrix $\mathbf{G}_X$ can be written as in Box VI.

To obtain the terms $(\mathscr{G}_X)_{ij}$ given in the matrix in Box III in the main text,

$$(\mathscr{G}_X)_{11} = \sum_{j=1}^{4}(\mathbf{G}_X)_{1j};$$

$$(\mathscr{G}_X)_{12} = \sum_{j=5}^{7}(\mathbf{G}_X)_{1j}; \qquad (\mathscr{G}_X)_{13} = \sum_{j=8}^{9}(\mathbf{G}_X)_{1j}$$

$$(\mathscr{G}_X)_{21} = \sum_{j=1}^{4}(\mathbf{G}_X)_{5j};$$

$$(\mathscr{G}_X)_{22} = \sum_{j=5}^{7}(\mathbf{G}_X)_{5j}; \qquad (\mathscr{G}_X)_{23} = \sum_{j=8}^{9}(\mathbf{G}_X)_{5j}.$$

The autosomal matrices $\mathbf{D}_A$, $\mathbf{R}_A$, $\mathbf{B}_A$, and $\mathbf{G}_A$ are all 10×10 matrices, with states as in Table 1. Using $\mathbf{P}_A$ given by the matrix in Box I, from Eq. (1) $\mathbf{D}_A = \lim_{N \to \infty} \mathbf{P}_A$ is

$$\mathbf{D}_A = \begin{bmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2
\end{bmatrix}.$$

$$\mathbf{R}_A = \lim_{t \to \infty}(\mathbf{D}_A)^t$$

$$= \begin{bmatrix}
0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2
\end{bmatrix}.$$

For B_A and G_A, see Boxes VII and VIII.

To obtain the terms $(\mathscr{G}_A)_{ij}$ given in the matrix in Box IV in the main text,

$$(\mathscr{G}_A)_{11} = \sum_{j=1}^{5}(\mathbf{G}_A)_{1j};$$

$$(\mathscr{G}_A)_{12} = \sum_{j=6}^{8}(\mathbf{G}_A)_{1j}; \qquad (\mathscr{G}_A)_{13} = \sum_{j=9}^{10}(\mathbf{G}_A)_{1j}$$

$$(\mathscr{G}_A)_{21} = \sum_{j=1}^{5}(\mathbf{G}_A)_{6j};$$

$$(\mathscr{G}_A)_{22} = \sum_{j=6}^{8}(\mathbf{G}_A)_{6j}; \qquad (\mathscr{G}_A)_{23} = \sum_{j=9}^{10}(\mathbf{G}_A)_{6j}.$$

$$\mathbf{B}_X = \lim_{N \to \infty} N(\mathbf{P}_X - \mathbf{D}_X)$$

Box V.

Box VI.

$$\mathbf{G}_X = \mathbf{R}_X \mathbf{B}_X \mathbf{R}_X =$$

where $g_X^* = M_m/(1-r) + 2M_f/r$.

Box VII.

$$\mathbf{B}_A = \lim_{N\to\infty} N(\mathbf{P}_A - \mathbf{D}_A) =$$

where $b_A^* = M_m/(1-r) + M_f/r$.

The product matrix $\mathbf{G}_A = \mathbf{R}_A \mathbf{B}_A \mathbf{R}_A$

$$\mathbf{G}_A =$$

Box VIII.

where $g_A^* = b_A^* = M_m/(1-r) + M_f/r$ (see the matrix in Box VII).

References

Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-Tamir, B., Cambon-Thomsen, A., et al., 2002. A human genome diversity cell line panel. Science 296, 261–262.

Conrad, D.F., Jakobsson, M., Coop, G., Wen, X., Wall, J.D., Rosenberg, N.A., Pritchard, J.K., 2006. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat. Genet. 38, 1251–1260.

Garrigan, D., Kingan, S.B., Pilkington, M.M., Wilder, J.A., Cox, M.P., Soodyall, H., Strassmann, B., Destro-Bisol, G., de Knijff, P., Novelletto, A., Friedlaender, J., Hammer, M.F., 2007. Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. Genetics 177, 2195–2207.

Hartl, D.L., Clark, A.G., 2007. Principles of Population Genetics. Sinauer, Sunderland, MA, pp. 124–125.

Hedrick, P.W., 2007. Sex: Differences in mutation, recombination, selection, gene flow, and genetic drift. Evolution 61, 2750–2771.

Helgason, A., Hrafnkelsson, B., Gulcher, J.R., Ward, R., Stefánsson, K., 2003. A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: Evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. Am. J. Hum. Genet. 72, 1370–1388.

Kayser, M., Choi, Y., van Oven, M., Mona, S., Brauer, S., Trent, R.J., Suarkia, D., Schiefenhövel, W., Stoneking, M., 2008. The impact of the Austronesian expansion: Evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. Mol. Biol. Evol. 25, 1362–1374.

Laporte, V., Charlesworth, B., 2002. Effective population size and population subdivision in demographically structured populations. Genetics 162, 501–519.

Lawson Handley, L.J., Perrin, N., 2007. Advances in our understanding of mammalian sex-biased dispersal. Mol. Ecol. 16, 1559–1578.

Möhle, M., 1998. A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. Adv. Appl. Probab. 30, 493–512.

Mountain, J.L., Ramakrishnan, U., 2005. Impact of human population history on distributions of individual-level genetic distance. Hum. Genomics 2, 4–19.

Nordborg, M., Krone, S.M., 2002. Separation of time scales and convergence to the coalescent in structured populations. In: Slatkin, M., Veuille, M. (Eds.), Modern Developments in Theoretical Population Genetics. Oxford University Press, Oxford, pp. 194–232.

Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., Stoneking, M., 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat. Genet. 29, 20–21.

Ramachandran, S., Rosenberg, N.A., Zhivotovsky, L.A., Feldman, M.W., 2004. Robustness of the inference of human population structure: A comparison of X-chromosomal and autosomal microsatellites. Hum. Genomics 1, 87–97.

Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., Cavalli-Sforza, L.L., 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. USA 102, 15942–15947.

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., Feldman, M.W., 2002. Genetic structure of human populations. Science 298, 2381–2385.

Rosenberg, N.A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J.K., Feldman, M.W., 2005. Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet. 1, e70.

Rosenberg, N.A., 2006. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. Ann. Hum. Genet. 70, 841–847.

Rousset, F., 1999. Genetic differentiation in populations with different classes of individuals. Theor. Popul. Biol. 55, 297–308.

Schaffner, S.F., 2004. The X chromosome in population genetics. Nat. Rev. Genet. 5, 43–51.

Seielstad, M.T., Minch, E., Cavalli-Sforza, L.L., 1998. Genetic evidence for a higher female migration rate in humans. Nat. Genet. 20, 278–280.

Slatkin, M., 1991. Inbreeding coefficients and coalescence times. Genet. Res. 58, 167–175.

Vitalis, R., 2002. Sex-specific genetic differentiation and coalescence times: Estimating sex-biased dispersal rates. Mol. Ecol. 11, 125–138.

Wang, J., 1997. Effective size and F-statistics of subdivided populations. II. dioecious species. Genetics 146, 1465–1474.

Wang, J., 1999. Effective size and F-statistics of subdivided populations for sex-linked loci. Theor. Popul. Biol. 55, 176–188.

Weir, B.S., 1996. Genetic Data Analysis II. Sinauer, Sunderland, MA.

Wilder, J.A., Kingan, S.B., Mobasher, Z., Pilkington, M.M., Hammer, M.F., 2004a. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. Nat. Genet. 36, 1122–1125.

Wilder, J.A., Mobasher, Z., Hammer, M.F., 2004b. Genetic evidence for unequal effective population sizes of human females and males. Mol. Biol. Evol. 21, 2047–2057.

Wilkins, J.F., 2006. Unraveling male and female histories from human genetic data. Curr. Opin. Genet. Dev. 16, 611–617.

Wilkins, J.F., Marlowe, F.W., 2006. Sex-biased migration in humans: What should we expect from genetic data? BioEssays 28, 290–300.