Estimation of Male Gene Flow from Measures of Nuclear and Female Genetic Differentiation

PHILIP W. HEDRICK, FRED W. ALLENDORF, AND C. SCOTT BAKER

From the School of Life Sciences, Arizona State University, Tempe, AZ 85287 (Hedrick); the Division of Biological Sciences, University of Montana, Missoula, MT (Allendorf); and the Marine Mammal Institute, Oregon State University, Newport, OR (Baker).

Address correspondence to Phil Hedrick at the address above, or e-mail: philip.hedrick@asu.edu.

Abstract

An approach is provided to estimate male gene flow and the ratio of male to female gene flow, given that there are estimates of diploid, nuclear gene flow and haploid, female gene flow. This approach can be applied to estimates of differentiation ($F_{ST}$) from biparentally and maternally inherited markers, assuming the equilibrium island model and equal effective numbers of males and females. Corrections to formulas used previously for California sea lions (González-Suárez M, Flatz R, Aurioles-Gamboa D, Hedrick PW, Gerber LR. 2009. Isolation by distance among California sea lion populations in Mexico: redefining management stocks. Mol Ecol. 18:1088–1099.) and American bison (Halbert ND, Gogan PJP, Hedrick PW, Wahl L, Derr JN. 2012. Genetic population substructure in bison in Yellowstone National Park. J Hered. 103:360–370.) are given and revised values for those species are calculated. The effect of unequal male and female effective population sizes, nonequilibrium conditions, and approximations of differentiation formulas are briefly discussed.

Key words: dioecious, island model, microsatellite loci, mtDNA, Y chromosome

The amount of overall genetic differentiation in populations (e.g., $F_{ST}$) can be estimated using biparentally inherited nuclear markers and the amount of genetic differentiation of populations resulting from genetic drift and female gene flow can be estimated using haploid, maternally inherited markers. The amount of genetic differentiation of populations resulting from male gene flow can also be estimated, given that there are paternally inherited haploid Y chromosomes (mammals, Drosophila, etc.) and there are Y chromosome markers.

Measuring the levels of female and male gene flow, particularly when they are quite different, is important for understanding the relative importance of gene flow and other sex-specific evolutionary factors (Hedrick 2007). However, in many organisms, Y chromosome markers are not available so that direct estimates of male gene flow are not possible. In this case, another approach is to utilize the combined estimates of overall differentiation (from nuclear DNA markers such as microsatellite loci or SNPs) and differentiation from maternally inherited markers (such as mtDNA in animals or W chromosome markers) to estimate male gene flow.

In plants, pollen dispersal is an indicator of male gene flow and seed dispersal is an indicator of both female and male gene flow. Ennos (1994) developed an approach to estimate the rates of seed and pollen gene flow. Subsequently, González-Suárez et al. (2009) and Halbert et al. (2012) used Equation 4 from Ennos (1994) to estimate relative rates of male and female gene flow in California sea lions and American bison. However, the equation they used from Ennos (1994) is for monoeocious (hermaphroditic) organisms, where single individuals produce both female and male gametes, whereas sea lions and bison are dioecious (gonochorous) organisms with distinct female and male individuals. In addition, the equation they derived was incorrect for monoeocious organisms. Below, we derive the appropriate equation for dioecious organisms, give the correct estimates for sea lions and bison, compare the dioecious and monoeocious estimates, and discuss some of the assumptions underpinning these estimates.

Derivation of Estimate

If an island model with populations of size $N$ and gene flow between them of rate $m$ is assumed, where $m$ is mean gene flow rate of females and males, the amount of genetic differentiation at equilibrium, assuming no mutation or selection, is approximately
\[ F_{ST} = \frac{1}{4Nm+1} \]  
(Wright 1940). Solving this expression for \( N \) gives

\[ N = \frac{1}{4m} \left( \frac{1}{F_{ST}} - 1 \right) \]  
(1b)

For a maternally inherited, haploid gene (mtDNA) in a dioecious organism, assuming that \( m_f \) is the rate of female gene flow, using the same approach as above, the equilibrium genetic differentiation is

\[ F_{ST(f)} = \frac{1}{Nm_f + 1} \]  
(2a)

and then

\[ N = \frac{1}{m_f} \left( \frac{1}{F_{ST(f)}} - 1 \right) \]  
(2b)

For a paternally inherited, haploid gene (Y chromosome) in a dioecious organism, assuming that \( m_m \) is the rate of male gene flow, the equilibrium genetic differentiation is

\[ F_{ST(m)} = \frac{1}{Nm_m + 1} \]  
(3a)

and then

\[ N = \frac{1}{m_m} \left( \frac{1}{F_{ST(m)}} - 1 \right) \]  
(3b)

Here it is assumed that these individuals have a binomial distribution of progeny so that the effective number of females is equal to the effective number of males. For simplicity, here \( N \) is assumed to be the effective number of individuals in the population as well as the census number. Below, we briefly discuss relaxing this assumption when the effective numbers of females and effective number of males differ.

From these equations (or using the approach of Ennos 1994), then

\[ \frac{1}{m_f} \left( \frac{1}{F_{ST}} - 1 \right) = \frac{2}{m_f} \left( \frac{1}{F_{ST(f)}} - 1 \right) + \frac{2}{m_m} \left( \frac{1}{F_{ST(m)}} - 1 \right) \]  
(4)

Solving this for \( F_{ST} \) gives

\[ F_{ST} = \frac{m_m F_{ST(m)} F_{ST(f)}}{2m_f F_{ST(f)} (1-F_{ST(m)})} + \frac{m_f F_{ST(f)} (1-F_{ST(m)})}{2m_m F_{ST(m)} F_{ST(f)}} \]  
(5a)

and then solving this equation for \( F_{ST(m)} \) gives

\[ F_{ST(m)} = \frac{(2m m_f F_{ST(f)} F_{ST(m)}) / (m_m F_{ST(f)} (1-F_{ST(m)}))}{2m_f F_{ST(f)} (1-F_{ST(m)})} \]  
(5b)

If it is assumed that the values of \( F_{ST} \) and \( F_{ST(f)} \) have been estimated from diploid nuclear and haploid mtDNA markers, respectively, then from the equilibrium values in Equations 1a and 2a

\[ m = \frac{1 - F_{ST}}{4NF_{ST}} \]  
(6a)

and

\[ m_\ell = \frac{1 - F_{ST(f)}}{NF_{ST(f)}} \]  
(6b)

If these values are substituted in Equation 5b, then

\[ F_{ST(m)} = \frac{2F_{ST} F_{ST(f)}}{F_{ST(f)} - 2F_{ST} + 3F_{ST} F_{ST(f)}} \]  
(7a)

Note that the equation used in both González-Suárez et al. (2009) and Halbert et al. (2012) is somewhat different from this and is incorrect.

With the assumptions as above, the male gene flow becomes

\[ m_m = \frac{1 - F_{ST(m)}}{NF_{ST(m)}} \]  

Then, with these values, the ratio of male to female gene flow can be estimated in terms of the female differentiation \( F_{ST(f)} \) and the estimated male differentiation \( F_{ST(m)} \) as

\[ \frac{m_m}{m_\ell} = \frac{F_{ST(m)} (1-F_{ST(m)})}{F_{ST(f)} (1-F_{ST(f))}} \]  
(7b)

Another approach to obtain an estimate of the ratio of male to female gene flow is to assume that

\[ \bar{m} = \frac{1}{2} (m_\ell + m_m) \]

so that

\[ m_m = 2\bar{m} - m_\ell \]

Using this expression for \( m_m \) and substituting the values for \( m_f \) and \( m_\ell \) from above, this gives the ratio of male to female gene flow in terms of the female differentiation \( F_{ST(f)} \) and the overall differentiation \( F_{ST} \) as

\[ \frac{m_m}{m_\ell} = \frac{(1 - F_{ST}) F_{ST(f)} - 2F_{ST} (1-F_{ST(f))}}{2F_{ST(f)} (1-F_{ST(f))}} \]  
(7c)

To show the relationship between \( F_{ST} \), \( F_{ST(f)} \), and \( m_m/m_\ell \), we used Equation 7c to calculate the male- to female–gene flow ratio for 2 levels of \( F_{ST} \) and \( F_{ST(f)} \) for values up to where \( m_m/m_\ell \) is equal to 10 (Figure 1). For example, when \( F_{ST} = 0.025 \) and \( F_{ST(f)} = 0.3 \), the value of \( m_m/m_\ell \) is 7.36. When \( m_m/m_\ell = 1 \) and \( F_{ST} = 0.025 \), then \( F_{ST(f)} = 0.093 \).

For comparison to monoecious organisms, using Equation 4 from Ennos (1994) and the same approach as above, the equation for male genetic differentiation is

\[ F_{ST(m)} = \frac{F_{ST(f)} F_{ST(f)}}{F_{ST(f)} - F_{ST} + F_{ST} F_{ST(f)}} \]  
(8)
Given the same values of $F_{ST}$ and $F_{ST(0)}$, the value of $F_{ST(m)}$ from Equation 7a for dioecious organisms is substantially higher than this one for monoecious organisms and the value of $m_m/m_f$ is substantially lower using the value of $F_{ST(m)}$ from Equation 8 in Equation 7b. Figure 1 (broken lines) gives the value for monoecious organisms for a range of $F_{ST}$ and $F_{ST(0)}$ values. For comparison, if $F_{ST} = 0.025$ and $F_{ST(0)} = 0.15$, then the values of $F_{ST(m)}$ are 0.067 and 0.029 for dioecious and monoecious organisms, respectively, and the values of the $m_m/m_f$ ratios are 2.44 and 5.88, respectively.

**Discussion**

Using the approach above to estimate the amount of male gene flow could be useful in many situations because estimation of male gene flow is often difficult because of the lack of Y chromosome markers (or Y chromosomes). In addition, this approach estimates the cumulative effect of multiple generations of gene flow, not just the most recent one as from direct observation of gene flow. However, in using these estimates, one should be careful to consider the effect of the amount of variation on the estimates of $F_{ST}$. That is, the potential range of $F_{ST}$ is a function of the amount of variation (Hedrick 2005), and perhaps a standardized measure of genetic differentiation would be useful (Meirmans and Hedrick 2011), although it would be important to examine this effect theoretically and empirically before using such a measure.

González-Suárez et al. (2009) and Halbert et al. (2012) estimated the relative rate of male to female gene flow in California sea lions and American bison using this approach. However, they used an equation for monoecious organisms (which was incorrect, see below). Using Equations 7a and 7b for values of $F_{ST(m)}$ and $m_m/m_f$ for their examples from sea lions and bison, the correct values are calculated and given in Table 1. All the comparisons suggest that there is substantially more gene flow for males than females (although the values here are not as large as originally given by González-Suárez et al. 2009 and Halbert et al. 2012). For example, the $m_m/m_f$ value for northern and southern herds of Yellowstone bison is 5.25 and the $m_m/m_f$ value for the Lower Gulf of California and the Pacific Ocean populations of sea lions is 5.44. As comparisons, the uncorrected values of $F_{ST(m)}$ and $m_m/m_f$ from González-Suárez et al. (2009) and Halbert et al. (2012) are given in parentheses in Table 1.

The approach used here assumes equal effective population sizes for females and males. In some species, however, these values can differ substantially because of greater variance in male reproductive success so that the effective number of males ($N_m$) is much smaller than the effective number of females ($N_f$). In this case, the appropriate value of $N$ used in Equation 1a is the inbreeding effective population size of

$$N = \frac{4N_e N_m}{N_e + N_m}$$

(Crow and Denniston 1988). Differences in $F_{ST(0)}$ and $F_{ST(m)}$ can result either from differences in gene flow or differences in effective population size between males and females. Given that the differentiation differences are the result of unequal effective population sizes, then an approach such as suggested by Waples et al. (2011) to determine the sex-specific $N_e$ as a function of the variance in reproductive success might be appropriate.

**Figure 1.** The ratio of male to female gene flow $m_m/m_f$ using Equation 7c for dioecious organisms (solid lines) and using Equations 7b and 8 for monoecious organisms (broken lines) for 2 values of $F_{ST}$ as a function of $F_{ST(0)}$.

**Table 1** Estimation of the genetic differentiation in bison and California sea lions from male gene flow, $F_{ST(m)}$, using Equation 7a and the estimated ratio of male to female gene flow, $m_m/m_f$, using Equation 7b.

<table>
<thead>
<tr>
<th>Populations</th>
<th>$F_{ST}$</th>
<th>$F_{ST(0)}$</th>
<th>$F_{ST(m)}$</th>
<th>$m_m/m_f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yellowstone bison</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern</td>
<td>Southern</td>
<td>0.032</td>
<td>0.292</td>
<td>0.073 (0.032)</td>
</tr>
<tr>
<td>California sea lion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower gulf</td>
<td>Upper gulf</td>
<td>0.026</td>
<td>0.170</td>
<td>0.067 (0.028)</td>
</tr>
<tr>
<td>Pacific</td>
<td>Upper gulf</td>
<td>0.040</td>
<td>0.166</td>
<td>0.130 (0.046)</td>
</tr>
<tr>
<td>Lower gulf</td>
<td>Pacific</td>
<td>0.030</td>
<td>0.285</td>
<td>0.068 (0.030)</td>
</tr>
</tbody>
</table>

Data for $F_{ST}$ (using diploid nuclear microsatellite loci) and $F_{ST(0)}$ (using mtDNA markers) are from González-Suárez et al. (2009) and Halbert et al. (2012). The uncorrected estimates of $F_{ST(m)}$ and $m_m/m_f$ from González-Suárez et al. (2009) and Halbert et al. (2012) are given in parentheses for comparison.
When the effective number of males is much smaller than the effective number of females, this will increase the value of both \( F_{ST} \) and \( F_{ST(m)} \) from that expected with equal effective numbers of females and males. Therefore, in species such as bighorn sheep (\textit{Ovis canadensis}) (and potentially both California sea lions and American bison), Equations 7b and 7c will underestimate the actual dissimilarity in male and female gene flow. On the other hand, if gene flow is higher for males than females (Lawson Handley and Perrin 2007), then this will offset the effects on \( F_{CT} \) and \( F_{ST(m)} \) of a lower male than female effective population size.

One of the assumptions of the derivations above is that the populations are at genetic drift–gene flow equilibrium. To determine how the estimates are influenced if the populations are not at equilibrium, the recursion equations (Wright 1951)

\[
\begin{align*}
F^{t+1}_{ST} &= \frac{1}{2N} \left[ 1 - \frac{1}{N_m} \right] F^{t}_{ST} \left( 1 - m_t \right)^2 \\
F^{t+1}_{ST(m)} &= \frac{1}{N_m} \left[ 1 - \frac{1}{N_m} \right] F^{t}_{ST(m)} \left( 1 - m_t \right)^2 \\
F^{t+1}_{ST(f)} &= \frac{1}{N_f} \left[ 1 - \frac{1}{N_f} \right] F^{t}_{ST(f)} \left( 1 - m_f \right)^2
\end{align*}
\]

(10) can be used to determine the change in genetic differentiation over time, where \( N, N_m, \) and \( N_m \) are the total, female, and male effective population sizes.

As an instructive example of the effect of not being at equilibrium, assume that all 3 \( F_{ST} \) values were initiated at 0. Figure 2 gives the proportion observed of the equilibrium of the estimate of \( m_m/m_t \) using Equation 7b when \( m_m/m_t = 5.0 \). For the higher gene flow values \( (m = 0.15, m_m = 0.25, \) and \( m_t = 0.05 \)), the estimate of \( m_m/m_t \) reaches about 90% of the equilibrium by generation 20, whereas for the lower gene flow values \( (m = 0.0375, m_m = 0.0625, \) and \( m_t = 0.0125 \)), it takes 65 generations to reach 90% of the equilibrium. The potential effect of different population sizes \( (N = 100, 200, 400) \) were also investigated for the rate of approach to the equilibrium level. For a given gene flow value, the pattern of reaching the equilibrium was very similar for different population sizes and in fact the rates are indistinguishable when put in this figure. In other words, if a homogeneous population is divided into groups with more male gene flow than female gene flow, this difference is initially underestimated, particularly for lower gene flow, and is independent of the population size.

Another assumption of the derivations above is that the approximations in Equations 1a, 2a, and 3a are satisfactory. The equilibrium values for these equations are more accurately given by

\[
\begin{align*}
F_{ST} &= \frac{(1 - \bar{m})^2}{2N} - (2N - 1)(1 - \bar{m})^2 \\
F_{ST(f)} &= \frac{(1 - m_f)^2}{N_f} - (N_f - 1)(1 - m_f)^2 \\
F_{ST(m)} &= \frac{(1 - m_m)^2}{N_m} - (N_m - 1)(1 - m_m)^2
\end{align*}
\]

(11) as an instructive example of the effect of not being at equilibrium, assume that \( m, m_m, \) and \( m_t \) are 0.0375, 0.0625, and 0.0125, respectively, and \( N, N_m, \) and \( N_f \) are 400, 200, and 200, respectively. In this example, \( m_m/m_t \) is 5.0 from the known gene flow rates. The equilibrium values for \( F_{ST}, F_{ST(f)}, \) and \( F_{ST(m)} \) from Equation (11) are 0.015, 0.164, and 0.035 and the estimates of \( m_m/m_t \) from Equations 7b and 7c are 5.41 and 5.24, respectively. These are somewhat larger than the ratio of 5.0 from the known gene flow rates but they are in the direction that makes the approximations in the approach here conservative in evaluating the ratio of male to female gene flow.

Overall, the approach outlined here allows estimation of paternal differentiation and relative rate of male to female gene flow in populations where there are molecular data on nuclear diploid differentiation and data on maternal differentiation, even in the absence of \( Y \) chromosome markers. This approach needs to be evaluated in the context of the caveats above, that is, the assumption that differentiation is at gene flow–genetic drift equilibrium and the approximations used here do not bias the estimates significantly.

**Acknowledgments**

P.W.H. thanks the Ullman Professorship for partial support. F.W.A. thanks the Hatfield Marine Science Center, Oregon State University, for their hospitality and support.

**References**


Received April 2, 2013; First decision May 20, 2013; Accepted June 27, 2013

Corresponding Editor: Robin Waples