Gene flow and selection balance in haplodiploid social insects

JOEL D. PARKER* & PHILIP W. HEDRICK

Department of Biology, Arizona State University, Tempe AZ 85287–1501, U.S.A.

Understanding the joint effects of gene flow and selection is one of the more important but difficult areas of evolutionary genetics. A method is presented to describe cases of gene flow–selection balance while accounting for the life history of haplodiploid social insects (eusocial Hymenoptera). A continent–island model is used to estimate gene flow rates from allele frequency data and provide variance formulae. The occurrence of unequal yet stable allele frequencies among social insect castes is explained and the usefulness of models with discrete generations for eusocial insects is explained. Gene flow and selection balance has been demonstrated for the fire ant *Solenopsis invicta* at two allozyme loci. This continent–island model gives results that well describe the system. Issues fundamental to the differences between X-linked and haplodiploid genetic systems and diploid genetic systems are also discussed.

Keywords: gene flow, haplodiploid, Hymenoptera, migration, selection, *Solenopsis invicta*.

Introduction

A stable polymorphism can be maintained in a population by a balance between selection reducing the frequency of an allele, and gene flow from an outside source population with a high frequency of the allele (Barton & Clark, 1990; Barton, 1992). Despite progress made in describing clines maintained by gene flow and selection, it is difficult to detail either the extent or pattern of selection or the level of gene flow in natural populations (Owen, 1986; Mallet *et al.*, 1990; Goodisman *et al.*, 1998). The violation of assumptions needed to make these estimates from allele frequency data can have a significant impact on the estimates (Whitlock & McCauley, 1999). Furthermore, the effects of such evolutionary factors and assumptions in haplodiploids or X-linked genes is often more complicated than in diploids (Clegg & Cavener, 1982; Owen, 1986; Hedrick & Parker, 1997).

Most genetic models of haplodiploid social insects rely on assumptions leading rapidly to equilibrium (Avery, 1984; Nagylaki, 1996). Owen (1986) has shown that continuous allele frequency clines at haplodiploid loci correspond to diploid loci under three assumptions: no dominance, equal dispersal between the sexes, and average effect of alleles the same in each sex. These assumptions ensure Hardy–Weinberg proportions (HWP) that probably do not exist in cases of strong selection. Indeed, even under a neutral model, the effect of differential migration between the sexes can effect the amount of differentiation among populations (Berg *et al.*, 1998). Furthermore, in most eusocial Hymenoptera, the differences between the sexes in life history are so great that the assumption of alleles and their respective phenotypes having equal effect between the sexes is almost certainly violated.

In the south-eastern United States, the red imported fire ant, *Solenopsis invicta*, has two documented polymorphisms apparently maintained by high levels of selection and gene flow (Ross, 1992, 1997; Keller & Ross, 1993b, 1999a,b; Ross & Keller, 1995, 1998; Ross *et al.*, 1997). As a direct result, this population shows unusual patterns at these loci, including unequal but stable allele frequencies between queens and workers. Although this system of gene flow and selection balance has been well characterized and explained (Ross, 1997), the theoretical foundations that might be applied to similar, yet discovered systems, are not sufficiently developed. Indeed, the published estimations of migration rates are based on a diploid-based genetic model (Ross & Keller, 1995). Recently, general and complex derivations for gene flow and cyto-nuclear disequilibrium under various haplodiploid models of gene flow were presented in a series of papers (Goodisman & Asmussen, 1997; Goodisman *et al.*, 1998; Goodisman

*Correspondence and Present address: IE-ZEA, Bâtiment de Biologie, Université de Lausanne, CH-1015 Lausanne, Switzerland. E-mail: Joel.Parker@ie-zea.unil.ch

*et al.*, 2000). Here we develop and explain the foundations required for understanding these treatments.

To accomplish this goal, many assumptions above will be relaxed, and the haplodiploid case of gene flow without selection will be examined. The most general case of gene flow opposed by directional selection will be treated. One additional assumption, discrete generations, will be made, based upon the life history of eusocial insects. Reproductive females and workers within a monogynous (single queen) social insect colony always differ by one generation. When these systems are at equilibrium, this assumption usually holds for polygynous (multiple queen) colonies as well. Thus, even with the assumption of discrete generations, extremely robust predictions of observed allelic and genotypic frequencies are possible. Additionally, the general models must account for the oscillatory approach to HWP caused by differences in male and female allele frequencies arising from gene flow and selection. The most important characteristic of the models is the most different aspect of the genetics of haplodiploid social insects; the overlapping generations of males, reproductive females and sterile workers. Here, males are haploid, arising from unfertilized eggs. Thus, having no father, they have the allele frequencies of the previous generation's reproductive females.

Models

Gene flow

The response of an island population to gene flow is explained with a simplified continent–island model. Assume a single locus with two alleles, and discrete generations. The allele frequencies in resident, island males before gene flow are defined as p_m and q_m for alleles A and a while the genotypic frequencies of AA, Aa and aa in the female zygotes before gene flow are denoted by P_f, H_f and Q_f. The female allele frequencies are designated p_f and q_f and are not necessarily equal to the male allele frequencies. An asterisk designates the analogous allelic and genotypic frequencies in the continent population, and these genotypic frequencies are assumed to be in HWP. The proportion of migrant males mating with resident females is denoted m_m. Hence, the proportion of resident males mating with resident females is $1 - m_m$. Similarly, for female gene flow, m_f represents the proportion of migrant females reproducing in the resident population and the proportion of resident females reproducing is $1 - m_f$. These two groups of females are mating randomly with the two groups of males described above.

One generation in the life cycle of a haplodiploid eusocial insect is diagrammed in Fig. 1. After gene flow
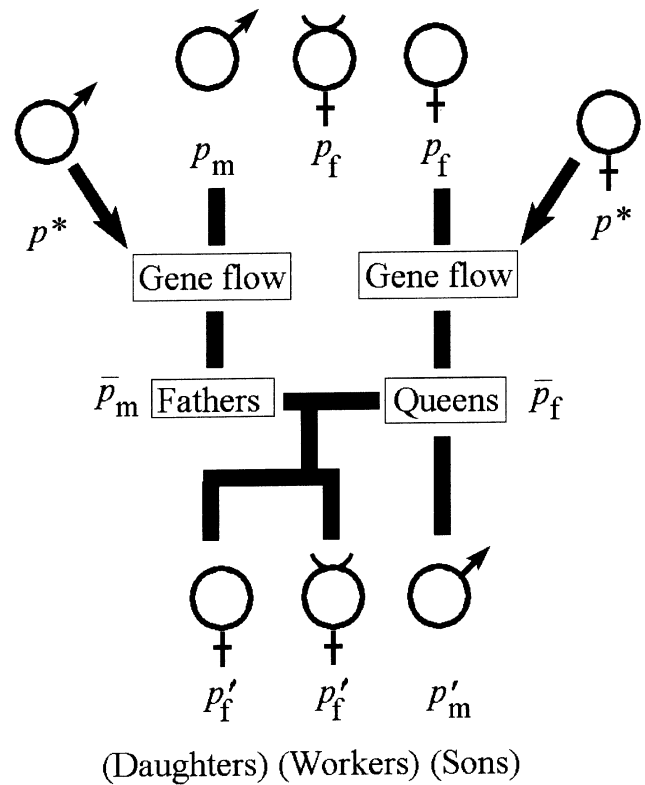
Fig. 1 Diagram illustrating one generation of the life cycle of typical eusocial insect receiving gene flow from an outside source. The zygotes destined to become males have allele frequency p_m and those destined to become queens and workers are allele frequency p_f. The mean male and female allele frequencies after gene flow are denoted $\bar{p}_m$ and $\bar{p}_f$. After mating, the next generation of queen and worker zygotes are formed with the allele frequency p'_f and males with allele frequency p'_m.

and selection queens and their sons have allele frequencies of p'_m. We describe the life cycle this way so that, except for queens, the allele frequencies in any given generation are those in the zygotes prior to gene flow and selection. The order of gene flow and selection is also important in whether the successful migrant individuals experience subsequent selection. For now, assuming no selection, let $\bar{p}_m$ be the average breeding male allele frequency in the island population weighted by migration ($(1 - m_m) p_m + m_m p^*$, where p_m is the allele frequency of island males before gene flow, and p^* the allele frequency of the migrant population). In the female case, the analogous notation is used for the genotypic averages, thus $\bar{p}_f = \bar{P}_f + \bar{H}_f/2$. Although most social insects have overlapping generations, the difference between reproductive queens and workers within a nest usually represents an average of exactly one generation change. Thus, despite the apparent

oversimplification of assuming discrete generations, this model will perform well at predicting these allele frequency differences between sex and caste.

Under these assumptions, the migrant's sex has an effect on an island population experiencing gene flow from a continent population. With no selection, the genotypic frequencies in females, both those destined to be reproductives or workers, should be equal. In contrast, males mating with these queens may not share the queen allele frequency, and instead have the allele frequency of the previous generation of queens after gene flow. This difference causes an excess of heterozygosity, which, left to itself, which would quickly approach HWP, (Robertson, 1965; Purser, 1966; Nagylaki, 1996). The difference is not just due to females carrying twice the number of alleles as males. Females have an impact on the allele frequencies of queens, workers and sons, whereas migrating males initially have an impact only on workers and the next generation of female reproductive frequencies.

Different responses over time to the influx of a new allele from a population fixed for the allele are shown in Fig. 2. Workers, males and females can have different allele frequencies as the island population approaches fixation, with female migrants causing a faster change in allele frequency than male migrants. Thus, not only might selection work differently on various castes, but the frequencies on which selection will work are initially different as well. For most realistic rates of gene flow, and differences in allele frequencies between the resident and migrants, these populations are predicted to show a slight, initial excess of heterozygotes that rapidly approaches HWP. However, these fluctuations can be accentuated by other unequal forces acting on the two sexes.

To understand the importance of considering the possible difference in allele frequency between sexes on estimations of the migration rate, we first write the expressions for the new allele frequencies from Table 1 for females and males, respectively, as:
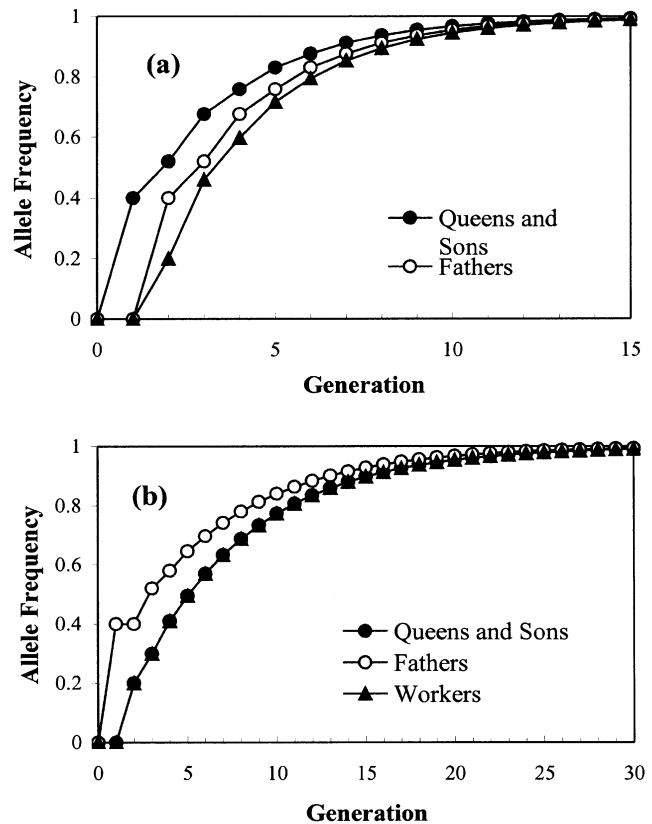
$$p_f' = \frac{p_f + p_m + m_m(p^* - p_m) + m_f(p^* - p_f)}{2} \quad (1a)$$

and

$$p_m' = p_f + m_f(p^* - p_f). \quad (1b)$$

These can then be rearranged to get the migration rates of:

$$m_m = \frac{(p_f' - p_f) + (p_f' - p_m)}{p^* - p_m} - m_f \frac{p^* - p_f}{p^* - p_m} \quad (2a)$$

Fig. 2 The response of allele frequencies in the various castes to gene flow and selection in an island population. Here male and female rates of gene flow are (a) $m_f = 0.4$, $m_m = 0$ and (b) $m_f = 0$, $m_m = 0.4$. The source population is fixed for the new allele. Note the difference between caste allele frequencies as the population approaches fixation and the more rapid approach to fixation in the female case. Note that the Queens and Sons, and the Workers curves are nearly superimposed.

for males, and

$$m_f = \frac{(p_f' - p_f) + (p_f' - p_m)}{p^* - p_f} + m_m \frac{p^* - p_m}{p_f - p^*}. \quad (2b)$$

and

$$m_f = \frac{p_f - p_m'}{p_f - p^*} \quad (2c)$$

for females.

The two migration eqns (2a–b) result from the equal contribution of alleles from migrating males and females to the next generation of females, whereas the last relation (2c) contain no male gene flow value because males are fatherless.

Table 1 The expected progeny frequencies from the different mating types

Mating		Progeny frequencies		
Type	Frequency	AA	Aa	aa
$A \times AA$	$\bar{p}_{\mathrm{m}}\bar{P}_{\mathrm{f}}$	$\bar{p}_{\mathrm{m}}\bar{P}_{\mathrm{f}}$	0	0
$A \times Aa$	$\bar{p}_{\mathrm{m}}\bar{H}_{\mathrm{f}}$	$\frac{\bar{p}_{\mathrm{m}}\bar{H}_{\mathrm{f}}}{2}$	$\frac{\bar{p}_{\mathrm{m}}\bar{H}_{\mathrm{f}}}{2}$	0
$A \times aa$	$\bar{p}_{\mathrm{m}}\bar{Q}_{\mathrm{f}}$	0	$\bar{q}_{\mathrm{m}}\bar{Q}_{\mathrm{f}}$	0
$a \times AA$	$\bar{q}_{\mathrm{m}}\bar{P}_{\mathrm{f}}$	0	$\bar{q}_{\mathrm{m}}\bar{P}_{\mathrm{f}}$	0
$a \times Aa$	$\bar{q}_{\mathrm{m}}\bar{H}_{\mathrm{f}}$	0	$\frac{\bar{q}_{\mathrm{m}}\bar{H}_{\mathrm{f}}}{2}$	$\frac{\bar{q}_{\mathrm{m}}\bar{H}_{\mathrm{f}}}{2}$
$a \times aa$	$\bar{q}_{\mathrm{m}}\bar{Q}_{\mathrm{f}}$	0	0	$\bar{q}_{\mathrm{m}}\bar{Q}_{\mathrm{f}}$
Expected frequencies (zygotes, workers)		$\bar{p}_{\mathrm{m}}\bar{p}_{\mathrm{f}}$	$\bar{p}_{\mathrm{m}}\bar{q}_{\mathrm{f}} + \bar{p}_{\mathrm{f}}\bar{q}_{\mathrm{m}}$	$\bar{q}_{\mathrm{m}}\bar{q}_{\mathrm{f}}$

Finally, substituting p_{r} in place of both (for resident allele frequency, assuming $p_{\mathrm{m}} = p_{\mathrm{f}}$) and calculating the weighted mean of the gene flow rates using eqns (2a) and (2c), we get the form analogous to the familiar diploid expression for gene flow,

$$\bar{m} = \frac{m_{\mathrm{m}}}{3} + \frac{2m_{\mathrm{f}}}{3} = \frac{p_{\mathrm{r}} \quad p_{\mathrm{r}}'}{p_{\mathrm{r}} \quad p^*} \tag{3a}$$

with variance (following the diploid treatment of Cavalli-Sforza & Bodmer (1971)),

$$V(\bar{m}) = \frac{1}{(p_{\mathrm{r}} \quad p^*)^2}\left[V(p_{\mathrm{r}}') + \bar{m}^2 V(p^*) + (1 \quad \bar{m})^2 V(p_{\mathrm{r}})\right] \tag{3b}$$

These last two relations do not hold for haplodiploid systems unless one can assume no, or a very small, difference in the allele frequencies between sexes, and an equal sex ratio. Implicit in eqn (3a) is the assumption of equal gene flow rates between the sexes to satisfy the assumption of equal allele frequencies between the sexes. If the sex ratio is also equal (not true for many Hymenopteran insects), then the gene flow rates of the sexes can be combined as a weighted mean of eqn (2). Otherwise these can be corrected to account for the sex ratio.

Selection before migration and gene flow balance

There is no simple and general way to deal with selection and gene flow simultaneously in haplodiploid systems. At least two variables are required for the gene flow rates of the two sexes, one variable for the allele frequency in the migrant population assuming HWP, three variables to describe the allele frequencies in the resident population (which may not be in HWP) and the variables to describe the various selection models where

selection can differ between the sexes. In addition, the timing of the assay relative to gene flow doubles the possibilities (Goodisman & Asmussen, 1997). However, any specific cases can be modelled by using this approach. Furthermore, the assumptions for specific cases will usually greatly simplify the derivations.

For example, extending the above neutral model to directional selection without dominance and selection not acting on the migrants is accomplished by replacing p_{m} with $p_{\mathrm{m}}(1 - s/2)$, P_{f} with $P_{\mathrm{f}}(1 - s)$, and H_{f} with $H_{\mathrm{f}}(1 - s/2)$ and normalizing to one before applying migration. These values are then used in Table 1 (into $\bar{p}_{\mathrm{m}}$, $\bar{p}_{\mathrm{f}}$, and $\bar{H}_{\mathrm{f}}$, respectively) to write lengthy expressions for the new male and female allele frequencies.

By setting the change in these allele frequencies equal to zero for both sexes and solving for m_{m} and m_{f}, it can be shown that

$$m_{\mathrm{m}} = \frac{4(\hat{p}_{\mathrm{m}} \quad \hat{p}_{\mathrm{f}}) + s\hat{p}_{\mathrm{m}}(\hat{p}_{\mathrm{f}} \quad \hat{q}_{\mathrm{f}} + \hat{p}_{\mathrm{m}})}{2(\hat{p}_{\mathrm{m}} \quad p^*) \quad s\hat{p}_{\mathrm{m}}q^*}, \tag{4a}$$

and

$$m_{\mathrm{f}} = \frac{4(\hat{p}_{\mathrm{f}} \quad \hat{p}_{\mathrm{m}} \quad s\hat{p}_{\mathrm{f}}\hat{q}_{\mathrm{m}}) + s\hat{H}_{\mathrm{f}}}{4(\hat{p}_{\mathrm{f}} \quad p^* \quad s\hat{p}_{\mathrm{f}}q^*) + s\hat{H}_{\mathrm{f}}}. \tag{4b}$$

The hats indicate that the equations are the allele frequencies at equilibrium. The algebraic expressions leading to these results can be quite long and involved, however, computer software packages such as Maple™ and Mathematica™ allow solutions to be found in a straightforward way which can then be verified by numerical iteration. Variances can also be derived using the delta method as is shown in the neutral case, although these quite long expressions are not shown here for space considerations.

Other sources of complication are dosage compensation and differing selective forces on the sexes and castes.

If there is dosage compensation, such as random X chromosome silencing, then the male would be equivalent to a homozygous female rather than a heterozygous female as assumed here. Another layer of complexity arises in social insects where males and females experience very different life histories and are probably under extremely different selective regimes.

Solenopsis invicta cases

Selection after migration and gene flow balance

An excellent example of how a specific case can greatly simplify the analysis is found in the case of the introduced fire ant, *Solenopsis invicta* (Ross, 1992, 1997; Ross & Keller, 1995). These data describe six polygynous (multiple queen) populations of *S. invicta* that occur downwind with increasing distance from a large monogynous (single queen) population of conspecifics. The population of the monogyne form has an estimated allele frequency of 0.815 for the allozyme variant *PGM-3*[A] and is in HWP with respect to the other allozyme allele *PGM-3*[a] across all castes (Ross & Keller, 1995). Homozygous queens for the *PGM-3*[A] allele do not live to become queens in the polygynous populations; they are either killed prior to flight (Keller & Ross, 1995) or after entering established colonies (Keller & Ross, 1993a,b). This is effectively selection after migration because only males are successfully migrating. Thus no homozygous *PGM-3*[A] queens were observed in any of the polygynous populations, but there was an excess of heterozygotes and a difference in allele frequencies between reproducing queens and workers. Furthermore, these frequencies appear constant over many generations suggesting an equilibrium in the polygyne population has been reached between selection through queen killing and male gene flow from the monogyne population. In addition, both polygynous and monogynous colonies occur sympatrically in native populations in Argentina, suggesting this phenomenon is not a result of its recent introduction into the United States (Ross *et al.*, 1996; Ross, 1997).

Later work revealed that selection maintaining this pattern does not occur at the *PGM-3* locus but at another tightly linked locus, *GP-9* (Ross, 1997; Ross & Keller, 1998; Keller & Ross, 1999a). The result is elimination of all *PGM-3*[A] homozygous queens and both types of homozygous *GP-9* allozyme queens in the polygynous population. Additionally, selection at the *GP-9* locus acts on workers, but we assume for now that only the *PGM-3* locus is affecting the system. A *PGM-3*-like case is more likely encountered, given that most researchers screen with neutral markers, with the true loci responsible for perturbations from HWP being detected only through linkage. Later, we deal with the *GP-9* locus and show an improved but not significantly better fit to the observed patterns. The necessary implication is that simply matching the observed pattern of frequencies does not prove a causal link. In the *PGM-3* case, we will see very good agreement between predictions from observed selection behaviours and the resulting allele and genotype frequencies, even though this locus is not directly responsible for the selection.

Although reasonably close to the values we will later derive, the original estimates of the gene flow rate were based on three equations for autosomal rather than X-linked genes (equations 1–3 of Ross & Keller (1995)) and may be replaced with equations 2.8a, p. 60 of Hedrick (2000). One other minor problem with this study was that gene flow rates were estimated by averaging three nonindependent estimates of genotypic frequencies when only two degrees of freedom were justified. The use of a bootstrap technique rather than an analytical derivation of the variance ameliorated this degree-of-freedom problem. The original treatment also assumed discrete generations and worked well because the average relations between queen and worker castes satisfy this assumption. Most recently, an attempt has been made at understanding the system by considering the two loci as one superlocus (Goodisman *et al.*, 2000).

PGM-3 model, complete elimination of PGM-3[AA] queens

The assumptions are that only males are migrating (Ross & Keller, 1995), the system is at equilibrium (not HWP, but the allele frequencies are constant over time), and the only selection is complete elimination of *PGM-3*[A] homozygous queens in the polygynous population after migration and before they reproduce. Again note that this form of selection is caused by another closely linked locus, not *PGM-3*, however, the resulting *PGM-3* allelic and genotypic frequencies are statistically consistent with selection at *PGM-3*. Values from Table 1 in Ross & Keller (1995) are used for all populations (pooled frequencies across all six populations with hats are used to symbolize equilibrium (observed unchanging frequencies) frequencies, $\hat{p}_f = 0.552$, $\hat{p}_m = 0.390$, and $p^* = 0.815$). We arrived at the first value by assuming the zygotic allele frequency of resident males ($\hat{p}_m$) is equal to the allele frequency of the resident mated reproductive queens (queens in our Fig. 1). The allele frequency of resident worker pupae is used as an estimate of the zygotic frequency of queens prior to selection. This is probably not a valid assumption for this system as there is some selection at the *GP-9* locus at this stage, but we will still find very good agreement between expected and empirical frequencies. We have assumed no selection on

workers and that genotypic frequencies of worker- and queen-destined zygotes are initially equal. This female frequency ($\hat{p}_f$) can not be measured directly because most PGM-3^A homozygous female alates (virgin winged reproductive females) are killed before they mature (Keller & Ross, 1995). In addition, the few that may mate are thought to be killed after re-entering a polygynous nest, as are all females from the migrant population (Keller & Ross, 1993a,b). Either way, this mortality can be modelled using Table 2 by setting the fitness of PGM-3^{AA} queens equal to zero.

With these assumptions and solving for the proportion of matings, by setting the change in allele frequencies equal to zero, the rate of male gene flow is just

$$m_m = \frac{2(\hat{p}_m - \hat{p}_f)}{\hat{p}_m - p^*}, \tag{6a}$$

with variance, from the delta method of

$$V(m_m) = \frac{1}{(\hat{p}_m - p^*)^2}\left[4V(\hat{p}_f) + 4\frac{(\hat{p}_f - p^*)^2}{(\hat{p}_m - p^*)^2}V(\hat{p}_m)\right.$$
$$\left. + m_m^2 V(p^*)\right] \tag{6b}$$

Using the values above gives $m_m = 0.762$ vs. 0.80 from the previous study (Ross, 1997). The variances of the allele frequencies $V(p'_f = p'_f(1 - p'_f)/2N$ and $V(p'_m = p'_m(1 - p'_m)/N)$ have not been reported so we can not calculate $V(m_m)$. Assuming queen killing as described, and using numerical iteration, a polygynous population initially fixed for the A allele will respond as shown in Fig. 3 (a) to male gene flow at $m_m = 0.762$ from a population with $p_m = 0.815$. The predicted equilibrium allele frequencies are extremely close to those observed ($\hat{p}_f = 0.54$ vs. 0.55 observed, and $\hat{p}_m = 0.38$ vs. 0.39 observed) showing little effect, if any, of using pupae as surrogates for female sexuals. Only knowledge of the variances or numerical resamp-

ling of the original data can resolve whether the system significantly deviates from this model, but a significant difference seems highly unlikely, with the possible exception of the male migration rate. If significant, the difference is likely to be the result of selection acting on worker pupae at the GP-9 locus (Ross, 1997). The original gene flow rates reported for this system were 0.77, 0.86 and 0.89, based on the three nonindependent genotypic frequencies used, and 0.91 from another unspecified method (Ross & Keller, 1995). More recently, m_m was given as 0.8 in Fig. 4 of Ross (1997). By setting the change in allele frequency equal to zero and solving for the female zygotic or worker heterozygosity, $\hat{H}_f$ it can be shown that the expected heterozygote frequency before selection (in the worker caste) is

$$\hat{H}_f = \frac{2\hat{p}_m\hat{q}_f}{\hat{q}_m}. \tag{7a}$$

with variance

$$V(\hat{H}_f) = \frac{4}{\hat{q}_f^2}\left(\hat{p}_m^2 V(\hat{q}_f) + \frac{\hat{q}_f^2}{\hat{q}_m^2}V(\hat{q}_m)\right). \tag{7b}$$

The result is quite simple considering that it does not depend on any assumptions of weak selection, HWP, or equal allele frequencies between sexes. However, this relation holds only for queen killing countered by male gene flow. Virtually all of the observed excess heterozygosity can be explained by eqn (7a). Using pooled data from the original study, the reported heterozygosity in worker pupae is 0.585, whereas eqn (7a) with the same values used above predicts that the heterozygosity would be 0.573, less than 3% below the observed. Although this difference may be of limited biological importance, it is possible that it may be statistically significant given the large sample sizes in the original study (multiple individuals from more than 400 colonies).

Table 2 Expected progeny frequencies for locus PGM-3 where the matings with AA are not given because they are eliminated by selection

Mating		Progeny frequencies		
Type	Frequency	AA	Aa	aa
$A \times Aa$	$\bar{p}_m H_f$	$\frac{\bar{p}_m H_f}{2}$	$\frac{\bar{p}_m H_f}{2}$	0
$A \times aa$	$\bar{p}_m Q_f$	0	$\bar{p}_m Q_f$	0
$a \times Aa$	$\bar{q}_m H_f$	0	$\frac{\bar{q}_m H_f}{2}$	$\frac{\bar{q}_m H_f}{2}$
$a \times aa$	$\bar{q}_m Q_f$	0	0	$\bar{q}_m Q_f$
Expected frequencies (zygotes, workers)		$\frac{\bar{p}_m H_f}{2(1-P_f)}$	$\frac{2\bar{p}_m q_f + \bar{q}_m H_f}{2(1-P_f)}$	$\frac{\bar{q}_m q_f}{(1-P_f)}$
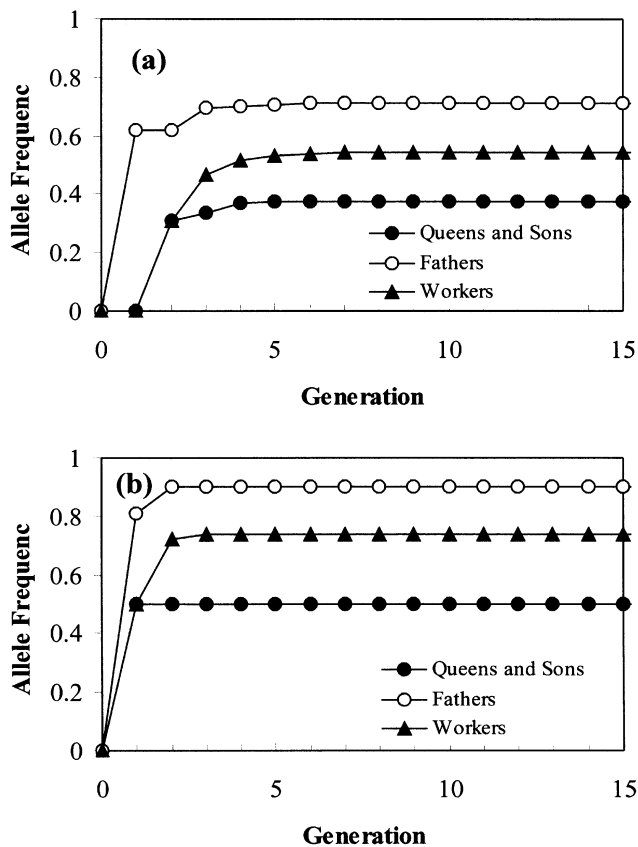
Fig. 3 Stable yet different allele frequencies due to gene flow–selection balance. The differences between caste are predicted as island populations respond to male gene flow balanced by selection. (a) Complete removal of $PGM-3^{AA}$ queens. The frequencies are chosen to model the *S. invicta* populations where the source population (p^*) is 0.815 and the rate of male gene flow (m_m) is 0.762. This plot was generated using numerical iteration starting with an initial allele frequency of zero for $PGM-3^A$ in the island population. (b) Complete removal of $GP-9^{BB}$ and $GP-9^{bb}$ queens where the source population (p^*) is fixed for the $GP-9^B$ allele and the rate of male gene flow (m_m) is 0.796. Again the rate of male gene flow and selection regime is chosen to model the *S. invicta* populations. The worker frequencies were corrected to account for elimination of $GP-9^{bb}$ homozygotes in that caste. These predicted stable frequencies were calculated by iteration starting with a small initial frequency (0.001) for $GP-9^b$.

GP-9 model, complete elimination of both homozygous queens

After discovery of the gene flow–selection balance at the *PGM-3* locus, another locus, *GP-9*, was found to either be directly responsible or more closely linked to the cause of this gene flow–selection balance (Ross, 1997; 1999b; Keller & Ross, 1999a). Complete elimination of both homozygotes at *GP-9* (consisting of two alleles, designated *B* and *b*) combined with gametic disequilib-

rium probably contributes to the difference between the expected and observed allele frequencies noted in the last section (Ross, 1997). Following the same procedure as before to make Table 3, the fitnesses of *BB*, *Bb*, *bb*, *B*, and *b* are 0, 1, 0, 1, and 1, respectively. Selection also acts on the workers to eliminate $GP-9^b$ homozygotes so one more line is added to Table 3 for workers. In this case, the migrant population is fixed for the *B* allele, thus no migrant females can survive and $m_f = 0$. Also, $\hat{p}_m$ is always equal to 0.5 because all surviving queens are heterozygous. Again simply setting the change in allele frequencies to zero for both sexes and solving for the male gene flow rate gives

$$m_m = \frac{2(\hat{q}_f \quad \hat{p}_f)}{(q^* \quad p^*)}, \tag{8a}$$

and variance, by the delta method, is

$$V(m_m) = \frac{4}{(q^* \quad p^*)^2}(4V(\hat{p}_f) + m_m^2 V(p^*)). \tag{8b}$$

Using the data reported for *GP-9* ($\hat{p}_f = 0.699$ from zygotes and $p^* = 1$ from Ross (1997)) and eqn (8a) gives a migration rate estimate of $m_m = 0.796$. Using numerical iteration, the response of a population with initial allele frequency of 0.001 receiving gene flow at $m_m = 0.796$ under these assumptions is shown in Fig. 3(b). This was calculated with correction for the selection against $GP-9^b$ homozygotes in the workers. The predicted worker steady state allele frequency falls inside the 95% confidence limits (0.737 predicted vs. 0.702 observed with reported 95% CL 0.616–0.867) consistent with selection at this locus as described (Ross, 1997).

Discussion

The high gene flow-strong selection example for two allozyme loci in fire ants is one of the most important case studies of this phenomenon currently known. However, understanding the details of this situation is a complicated task, as haplodiploidy at such zones causes many deviations from analogous diploid systems. The task can be greatly simplified by the details of a particular situation, as seen by the brief expressions obtained in the fire ant cases.

We have shown how differences in allele frequency between the sexes can occur and manifest themselves in differing allele frequencies in social insect castes. In the case of fire ants, an excess of heterozygosity is expected, given a difference in allele frequency between the sexes. This is somewhat analogous to the gamete vs. zygote models of gene flow for plants where different effects of

Table 3 Expected progeny frequencies for locus *GP-9* where the two homozygous mating types are not given because they are eliminated before reproducing

Mating		Progeny frequencies		
Type	Frequency	*BB*	*Bb*	*bb*
$B \times Bb$	$\bar{p}_{\mathrm{m}}H_{\mathrm{f}}$	$\frac{\bar{p}_{\mathrm{m}}H_{\mathrm{f}}}{2}$	$\frac{\bar{p}_{\mathrm{m}}H_{\mathrm{f}}}{2}$	0
$b \times Bb$	$\bar{q}_{\mathrm{m}}H_{\mathrm{f}}$	0	$\frac{\bar{q}_{\mathrm{m}}H_{\mathrm{f}}}{2}$	$\frac{\bar{q}_{\mathrm{m}}H_{\mathrm{f}}}{2}$
Expected frequencies (zygotes)		$\frac{\bar{p}_{\mathrm{m}}}{2}$	$\frac{1}{2}$	$\frac{\bar{q}_{\mathrm{m}}}{2}$
Expected frequencies after selection on workers (applies to workers only)		$\frac{\bar{p}_{\mathrm{m}}}{\bar{p}_{\mathrm{m}}+1}$	$\frac{1}{\bar{p}_{\mathrm{m}}+1}$	0

pollen vs. seed dispersal must be taken into account (Ennos, 1994; Hedrick, 2000). Even with equal gene flow from each sex, the system will subtly oscillate to equilibrium, with the heterozygous excess only occurring in the initial generations. Some of these effects may be obscured when the assumption of discrete generations is broken; however, the average difference between the queen and worker caste will always differ by exactly one generation of gene flow, yielding a difference in allelic and genotypic frequencies between castes before equilibrium is reached. Hence, there is close agreement between predicted and observed frequencies in the fire ant example (by us and the original study) despite overlapping generations in natural nests that violate the assumption of discrete generations. The timing and extent of differential selection acting on reproductives, and differential gene flow between the sexes, are expected to create these types of differences. Thus, the assumption of equal allele frequencies between sexes in social insects should be made only with supporting empirical data. These situations can be modelled using the approach given here. The analysis of social insect genetic systems has a great advantage with the availability of castes and can add much to our understanding, as shown in *S. invicta*.

Our second purpose here is to estimate gene flow and to better explain the gene flow–selection balance in *S. invicta*. Compared to the original analysis (Ross & Keller, 1995; Ross, 1997), we find a minor difference in the estimated gene flow rate and a need to test for significant deviation from expected in the case of selection at the *PGM-3* locus. Additionally, estimates of *PGM-3* heterozygosity and steady state frequencies appear very close to those observed, showing the absence of *GP-9bb* reproductive queens has a negligible impact on these *PGM-3* allele frequencies (see Ross, 1997). Although the data may be statistically consistent with selection at the *PGM-3* locus, the selection at *GP-9* better explains both loci simultaneously (Ross, 1997). Also, the traits associated with queen killing were recently shown to not be caused by the *PGM-3* locus (Keller & Ross, 1999a, b). The difference between observed and expected gene flow rate estimated from our model is more likely due to the nature of the six populations sampled. It is based on the mean of six downwind polygynous populations, whereas each population is experiencing a rate decreasing in distance from the monogyne population. A superlocus model incorporating both loci simultaneously revealed an equilibrium without gene flow, although more direct measures of gene flow agree with the models presented here (Goodisman *et al.*, 2000).

The fire ant example also demonstrates the usefulness of discrete generation models for social insect genetics. The difference of one generation between castes in monogynous colonies allows measurement of one generation of selection and gene flow through assaying the appropriate caste. Furthermore, this applies to polygynous populations at equilibrium although in cases of queen re-adoption, it may not strictly hold during certain periods of the life cycle. In the case of the fire ants, each nest contains overlapping generations of queens and workers yet genotypic frequencies of these two castes only differ by the effects of one average generation of gene flow and selection. The same is even more directly true between queens and males from a given nest. Overlapping generations will probably buffer the slight oscillatory approach to equilibrium observed in Figs 2 and 3, however, the final stable frequencies and the mean level of migration should be predicted correctly.

Finally, we have shown how stable yet different allele frequencies between castes and sex can be maintained in social Hymenoptera under gene flow–selection balance. The approach taken here should work in general for any future examples encountered. Effort should be made in future genetic work on eusocial Hymenoptera to measure all castes whenever possible to detect such polymorphism.

Acknowledgements

References

AVERY, P. J. 1984. The population genetics of haplodiploids and X-linked genes. *Genet Res.*, **44**, 321–341.

BARTON, N. H. 1992. On the spread of new gene combinations in the third phase of Wright's shifting-balance. *Evolution*, **46**, 551–557.

BARTON, N. H. AND CLARK, A. 1990. Population structure and process in evolution. In: Worhrmann, K. and Jain, S. K. (eds) *Population Biology. Ecological and Evolutionary Viewpoints*, pp. 115–173. Springer, Berlin.

BERG, L. M., LASCOUX, M. AND PAMILO, P. 1998. The infinite island model with sex-differentiated gene flow. *Heredity*, **81**, 63–68.

CAVALLI-SFORZA, L. L. AND BODMER, W. F. 1971. *The Genetics of Human Populations*. Freeman, San Francisco.

CLEGG, M. T. AND CAVENER, D. R. 1982. Dynamics of correlated genetic systems. VII. Demographic aspects of sex-linked transmission. *Am. Nat.*, **120**, 108–118.

ENNOS, R. A. 1994. Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.

GOODISMAN, M. A. D. AND ASMUSSEN, M. A. 1997. Cytonuclear theory for haplodiploid and X-linked genes. I. Hardy–Weinberg dynamics and continent–island, hybrid zone models. *Genetics*, **147**, 321–338.

GOODISMAN, M. A. D., SHOEMAKER, D. D. AND ASMUSSEN, M. A. 1998. Cytonuclear theory for haplodiploid species and X-linked genes. II. Stepping stone models of gene flow and application to a fire ant hybrid zone. *Evolution*, **52**, 1423–1440.

GOODISMAN, M. A. D., ROSS, K. G. AND ASMUSSEN, M. A. 2000. A formal assessment of gene flow and selection in the fire ant *Solenopsis invicta. Evolution*, **54**, 606–616.

HEDRICK, P. W. 2000. *Genetics of Populations*. Jones and Bartlett, Boston.

HEDRICK, P. W. AND PARKER, J. D. 1997. Evolutionary genetics and genetic variation of haplodiploids and X-linked genes. *Ann. Rev. Ecol. Syst.*, **28**, 55–83.

KELLER, L. AND ROSS, K. G. 1993a. Phenotypic plasticity and 'cultural transmission' of alternative social organization in the fire ant *Solenopsis invicta. Behav. Ecol. Sociobiol.*, **33**, 121–129.

KELLER, L. AND ROSS, K. G. 1993b. Phenotypic basis of reproductive success in a social insect: genetic and social determinants. *Science*, **260**, 1107–1110.

KELLER, L. AND ROSS, K. G. 1995. Gene by environment interaction: effects of a single gene and social environment on reproductive phenotypes of the fire ant queens. *Funct. Ecol.*, **9**, 667–676.

KELLER, L. AND ROSS, K. G. 1999a. Major gene effects on phenotype and fitness: the relative roles of *Pgm-3* and *GP-9* in introduced populations of the fire ant *Solenopsis invicta. J. Evol. Biol.*, **12**, 672–680.

KELLER, L. AND ROSS, K. G. 1999b. Selfish genes: a green beard in the red fire ant. *Nature*, **394**, 573–575.

MALLET, J., BARTON, N., LAMAS, G. M., SANTISTEBAN, J. C., ET AL., 1990. Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics*, **124**, 921–936.

NAGYLAKI, T. 1996. The diffusion model for migration and selection in a dioecious population. *J. Math. Biol.*, **34**, 334–360.

OWEN, R. E. 1986. Gene frequency clines at X-linked or haplodiploid loci. *Heredity*, **57**, 209–219.

PURSER, A. F. 1966. Increase in heterozygote frequency with differential fertility. *Heredity*, **21**, 322–327.

ROBERTSON, A. 1965. The interpretation of genotypic ratios in domestic animal populations. *Anim. Prod.*, **7**, 319–324.

ROSS, K. G. 1992. Strong selection on a gene that influences reproductive competition in a social insect. *Nature*, **355**, 347–349.

ROSS, K. G. 1997. Multilocus evolution in fire ants: effects of selection, gene flow, and recombination. *Genetics*, **145**, 961–974.

ROSS, K. G. AND KELLER, L. 1995. Joint influence of gene flow and selection on a reproductively important genetic polymorphism in the fire ant *Solenopsis invicta. Am. Nat.*, **146**, 325–348.

ROSS, K. G. AND KELLER, L. 1998. Genetic control of social organization in an ant. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 14232–14237.

ROSS, K. G., VARGO, E. L. AND KELLER, L. 1996. Simple genetic basis for important social traits in the fire ant *Solenopsis invicta. Evolution*, **50**, 2387–2399.

ROSS, K. G., KRIEGER, M. J. B., SHOEMAKER, D. D., VARGO, E. L. ET AL. 1997. Hierarchical analysis of genetic structure in native fire ant populations: results from three classes of molecular markers. *Genetics*, **147**, 643–655.

WHITLOCK, M. C. AND MCCAULEY, D. E. 1999. Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm + 1)$. *Heredity*, **82**, 117–125.