A FORMAL ASSESSMENT OF GENE FLOW AND SELECTION IN THE FIRE ANT *SOLENOPSIS INVICTA*

Michael A. D. Goodisman,[1,2] Kenneth G. Ross,[3] and Marjorie A. Asmussen[1]
[1]*Department of Genetics, University of Georgia, Athens, Georgia 30602*
[3]*Department of Entomology, University of Georgia, Athens, Georgia 30602*

Abstract.—Recent studies of the introduced fire ant *Solenopsis invicta* suggest that introduced polygyne (with multiple queens per nest) populations are strongly influenced by male-mediated gene flow from neighboring monogyne (single queen per nest) populations and selection acting on a single locus, *general protein-9* (*Gp-9*). This investigation formally tests this hypothesis and determines if these processes can account for the genotypic structure of polygyne *S. invicta.* To increase the statistical power of this test, we considered the genotypes of polygyne queens and workers at both *Gp-9* and the closely linked, selectively neutral locus *Pgm-3.* We then constructed and analyzed a novel mathematical model to delimit the effects of monogyne male gene flow and selection on the joint genotypes at the *Pgm-3/Gp-9* superlocus. Using this framework, a hierarchical maximum-likelihood method was developed to estimate the best-fitting gene flow and selection parameters based on the fit of our model to data from both the current study and an earlier one of the same population. In each case, selection on polygyne queens and workers alone, with no monogyne male gene flow, provides the most parsimonious explanation for the observed genotype frequencies. The apparent discrepancy between this result and the empirical evidence for monogyne male gene flow indicates that undocumented factors, such as other forms of selection in polygyne males or workers, are operating in introduced polygyne *S. invicta.*

Key words.—Fire ants, gene flow, isozymes, migration-selection balance, population structure, selection, *Solenopsis invicta.*

Gene flow and selection are fundamental evolutionary processes that may act jointly to shape patterns of genetic variation within populations. For example, selection against deleterious alleles may reduce variation, whereas gene flow from nearby populations may act in opposition to such selection by reintroducing unfit alleles (May et al. 1975; Felsenstein 1976; Endler 1977; Barton and Hewitt 1985; Slatkin 1985; Barton and Clark 1990). Several studies have reported such a migration-selection balance within species (McNeilly 1968; Livingstone 1969; Jones et al. 1977; Turner 1977; Jones 1982; Riechert 1993; King and Lawson 1995; Chevillon et al. 1998). One well-studied case is found in introduced populations of the fire ant *Solenopsis invicta* in the United States, where it was thought that a recessive deleterious allele at the locus *phosphoglucomutase-3* (*Pgm-3*) was maintained at high frequency in one social form by gene flow from neighboring populations of the alternate social form (Ross 1992; Ross and Shoemaker 1993; Ross and Keller 1995).

The role of *Pgm-3,* however, was recently challenged by the discovery of the diallelic locus *general protein-9* (*Gp-9*), which is tightly linked ($r = 0.0016$) to *Pgm-3* and apparently under strong selection in *S. invicta.* In monogyne (with a single queen per nest) populations, all females are genotype *Gp-9BB,* and so the monogyne form is fixed for the *Gp-9B* allele. In contrast, in polygyne (multiple queens per nest) populations, queens are almost always (>99%) heterozygous, because *Gp-9BB* queens are executed by workers and *Gp-9bb* queens die as they mature. Consequently, both *Gp-9* alleles are common in polygyne *S. invicta.* Recent surveys also have failed to find significant numbers of *Gp-9bb* polygyne workers in natural populations, thus suggesting strong selection

against this genotype operates in workers as well as queens (Ross 1997).

Evidence from native and introduced populations initially suggested that divergent forms of selection act independently on *Pgm-3* and *Gp-9* (Ross 1997). However, more recent findings suggest that the effects of *Gp-9* alone could explain all of the observed variation in the behavioral and phenotypic traits of queens and workers originally attributed to *Pgm-3* (Keller and Ross 1999). Under this new scenario, the patterns of genetic variation observed at *Pgm-3* in polygyne *S. invicta* result from tight linkage to and selection acting on *Gp-9* (or a different closely linked gene), in conjunction with the presumed gene flow from neighboring monogyne populations.

The purpose of the present investigation is to formally test this new hypothesis. Specifically, our goal is to use rigorous mathematical and statistical frameworks to determine if selection in polygyne females and gene flow via monogyne males can explain the two-locus, *Pgm-3/Gp-9* genotype distributions in introduced polygyne *S. invicta.* In contrast to previous mathematical approaches (Ross and Keller 1995), it is now assumed that selection acts only on *Gp-9* and not on *Pgm-3.* We first report new data on the patterns of genetic variation at these two loci in polygyne *S. invicta* queens and workers. We then develop and analyze novel mathematical models that predict the patterns of genetic variation under the observed selection on polygyne queens and all possible combinations of gene flow mediated by monogyne males and selection against *Gp-9bb* workers. We next use these models to develop a hierarchical maximum-likelihood procedure for estimating the level of selection in polygyne workers and rates of interform matings between monogyne males and polygyne queens. Finally, these models are fit to the new data collected for this study, as well as to an earlier survey of this polygyne population, to determine the combination of gene flow and selection that best explains the genotypic structure of introduced polygyne *S. invicta.*

[2] Present address: Department of Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia; E-mail: mike.goodisman@jcu.edu.au.
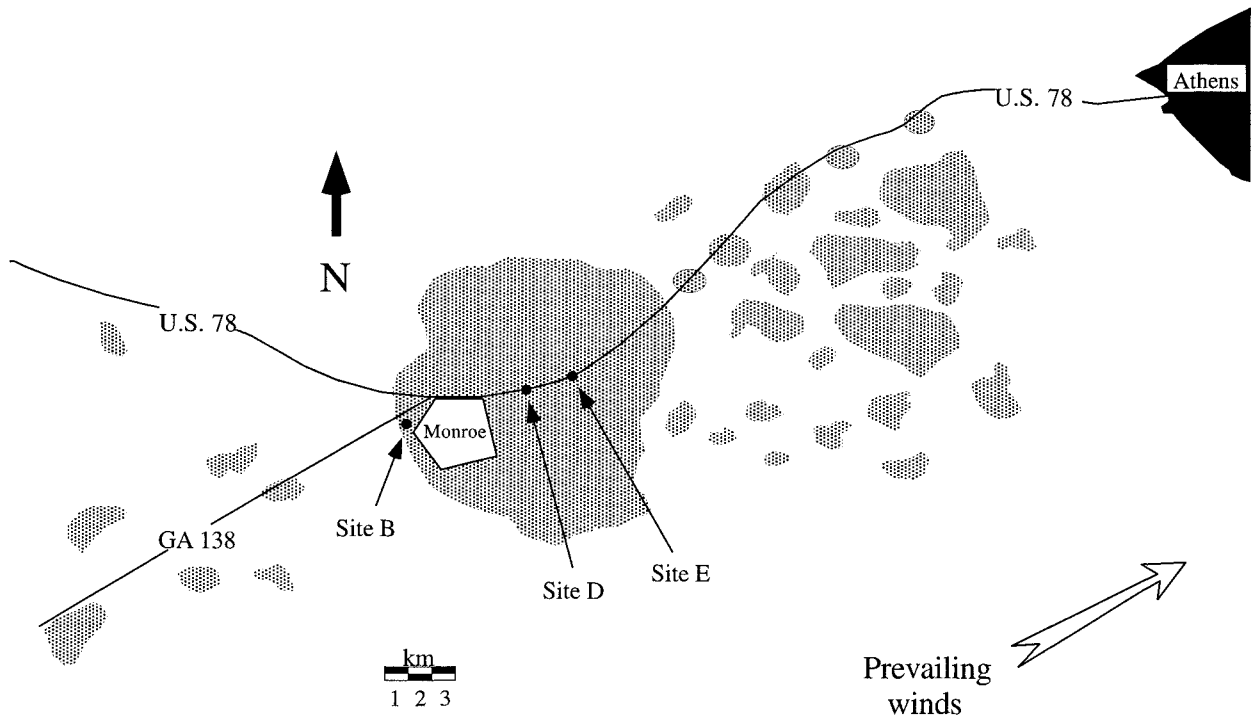
FIG. 1. Locations of three study sites in northern Georgia, United States, from which polygyne *Solenopsis invicta* queens and workers were collected. Areas containing only polygyne nests are indicated by stippling. All other areas are presumed to contain only monogyne nests. The open arrow indicates the direction of the prevailing winds during times of mating flights (see Ross and Keller 1995).

Observed Patterns of Genetic Variation

Sampling of Ants and Scoring of Genotypes

To allow a direct comparison of our conclusions based on joint *Pgm-3/Gp-9* genotype to those obtained from the previous analysis of *Pgm-3* genotype frequencies in this population (Ross and Keller 1995), we collected ants from three of the same localities used in that study. These sites, designated as sites B, D, and E (Ross and Keller 1995), lie along a transect through a well-studied polygyne population in northern Georgia, United States, downwind of a large monogyne population (Fig. 1). We expect that polygyne queens sampled from the windward side of the polygyne population (site B) should mate more frequently with monogyne males than do queens farther into the interior of the polygyne population (at sites D and E), because monogyne males presumably are blown into the polygyne population while on their mating flights (Ross and Keller 1995). Moreover, we anticipate that a higher frequency of queens sampled from site B should be mated because all monogyne males, but only a small fraction of polygyne males, are fertile (Ross and Fletcher 1985; Ross and Keller 1995).

We collected wingless (egg-laying) queens and workers from 85.00 ± 2.08 ($\bar{x} \pm$ SEM) nests in each of the three sites in the late spring and early summer of 1995. A mean of 2.24 ± 0.08 queens and 3.30 ± 0.07 workers were collected from each nest. (Males were not collected because they cannot be scored for *Gp-9*.) The mating status of queens was determined by examining their spermathecae; sperm in mated queens appeared as an opaque, whitish mass. Finally, the genotypes of all collected individuals were determined at *Pgm-3* and *Gp-9* using standard starch gel electrophoresis (Shoemaker et al. 1992; Ross 1997).

Estimation of population-level genotype frequencies from social insect colonies can pose difficulties when multiple individuals per nest are sampled because individuals within colonies may be related and thus not represent independent genetic information. To overcome this problem, we used a resampling technique that yields genotype distributions that are unbiased by colony-level genetic structure (Ross and Keller 1995; Ross 1997). Briefly, a computer program was written to select a single genotype from each of the sampled nests. This algorithm was repeated 1000 times to yield 1000 such datasets. For each genotype, the median number of such individuals observed was taken to be the unbiased estimate for the number of individuals of that genotype. This unbiased estimate was used in all further calculations. A similar resampling technique was used to obtain estimates for the proportion of mated queens within each of the three sites and in all sites combined. We conducted exact tests, as implemented by the program GENEPOP (Raymond and Rousset 1995), to detect genotypic differentiation among sites and deviations of genotype frequencies from Hardy-Weinberg equilibrium within sites.

Frequency of Mated Queens

As expected under the hypothesis of monogyne male-mediated gene flow, the frequency of mated queens in site B (0.798), which borders on the monogyne population, is higher than the frequencies from sites D (0.486) and E (0.474), which lie well within the polygyne population (Fig. 1). The

TABLE 1. Definition of the three *Pgm-3/Gp-9* superlocus alleles in terms of their allelic composition at the individual loci. The superlocus allele corresponding to the *Pgm-3*^{A}/*Gp-9*^{b} combination is absent from the study population.

Superlocus allele	*Pgm-3* allele	*Gp-9* allele
S_1	*Pgm-3*^{A}	*Gp-9*^{B}
S_2	*Pgm-3*^{a}	*Gp-9*^{B}
S_3	*Pgm-3*^{a}	*Gp-9*^{b}

overall proportion of mated queens across all sites is 0.597. The frequencies of mated queens are significantly lower than those observed in the previous study of Ross and Keller (1995) at sites B and E and for all sites combined (*G*-test of independence: site B, $G_1 = 5.83$, $P < 0.05$; site D, $G_1 = 0.31$, $P > 0.5$; site E, $G_1 = 4.43$, $P < 0.05$; all sites combined, $G_1 = 5.54$, $P < 0.05$). These differences may result from changes in the distribution of the polygyne form, which expanded somewhat since the original study was conducted (Ross and Shoemaker 1997).

Genotype Frequencies at the Pgm-3/Gp-9 Superlocus

In introduced populations, *Gp-9* genotype strongly influences the viability and reproductive success of polygyne females, whereas the genotype at the closely linked locus *Pgm-3* apparently does not affect these fitness components (Keller and Ross 1999). However, we may substantially increase the information obtained from our *Gp-9* data alone by considering the variation at *Pgm-3* as an extension of that at *Gp-9*. Treating the joint two-locus haplotypes as alleles at a single superlocus (Table 1) effectively adds a third ''allele'' to the simpler *Gp-9* locus. The potential fourth allele, represented by the *Pgm-3*^{A}/*Gp-9*^{b} haplotype, is absent from introduced populations (yielding maximal linkage disequilibrium of $D' = 1.0$ between the two loci; Ross 1997).

Table 2 presents the observed genotype frequencies at the composite superlocus. Both the queen and worker genotype frequencies differ significantly from expectations under Hardy-Weinberg equilibrium (exact test, $P < 0.001$ for each site and all sites combined in both castes), although they do not differ significantly across sites (exact test, $P > 0.5$ for workers and $P > 0.6$ for queens).

EXPECTED PATTERNS OF GENETIC VARIATION

Our goals are to determine if the superlocus frequencies in introduced polygyne populations of *S. invicta* can be accounted for by monogyne male gene flow in conjunction with female selection, as well as whether the data can be explained more parsimoniously by only one or neither of these processes. We begin a theoretical investigation of these issues by deriving and analyzing a general superlocus model to determine the effects of both monogyne male gene flow and selection in females. We then use these mathematical results to develop a hierarchical maximum-likelihood procedure for estimating the population parameters that best fit the observed genotypic counts. Our analyses assume that mutation and drift do not occur, alleles segregate according to Mendelian laws, and the genetic composition of monogyne males is constant over time. We also assume that generations are dis-
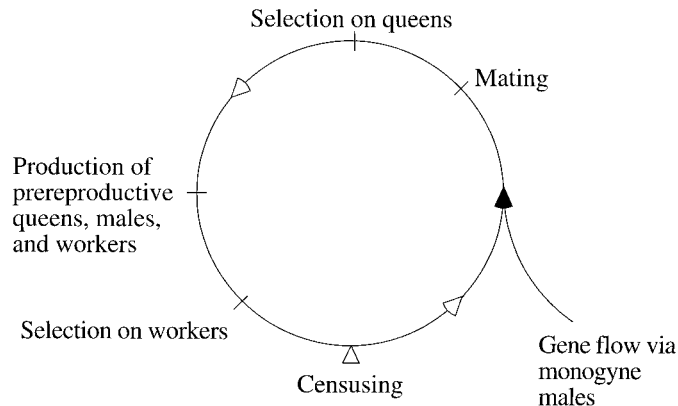

FIG. 2. Polygyne *Solenopsis invicta* generation cycle and census time with respect to gene flow via monogyne males and selection on queens and workers.

crete and nonoverlapping, a supposition justified by the fact that the majority of *S. invicta* males live for only a single season (Hölldobler and Wilson 1990) and the estimated life span of polygyne *S. invicta* queens is approximately one year (Goodisman and Ross 1999).

Figure 2 illustrates the generation cycle within the polygyne population, with censusing of adults after selection. Each generation begins with mating and male gene flow such that a constant proportion of polygyne queens, $m(0 \leq m \leq 1)$, mate with monogyne males, whereas the remaining fraction, $1 - m$, mate with polygyne males (Ross and Shoemaker 1993; Ross and Keller 1995). Note that monogyne males do not physically join the polygyne population, they only contribute their alleles. After mating, all males die and polygyne queens attempt to enter existing polygyne nests. It is at this point that selection acts in the queen caste. Under the superlocus framework, only the queens with the superlocus genotypes S_1S_3 and S_2S_3 survive to reproduce (Ross 1997; Keller and Ross 1999). The surviving queens then produce the next generation of prereproductive queens, males, and workers, which comprise the three castes in this species. As with many social Hymenoptera, workers are obligately sterile, the haploid males are produced parthenogenetically by the queens, and the new diploid queens and workers are produced sexually by queens and males (Hölldobler and Wilson 1990). The final stage of the generation cycle involves viability selection on the workers, such that the recessive genotype S_3S_3 has constant viability $1 - s$, where $0 \leq s \leq 1$, relative to a viability of one for all other worker genotypes.

Superlocus Nomenclature

Under the superlocus framework, the genotype frequencies in polygyne queens, males, and workers are denoted by the letters *Q, M,* and *W,* respectively (Table 3). The frequencies for males in monogyne populations are distinguished by a superscripted asterisk (*M**). Polygyne queens have only two viable superlocus genotypes (S_1S_3 and S_2S_3), whereas polygyne males may display any of the three (S_1, S_2, or S_3) and polygyne workers any of the six (S_1S_1, . . . , S_3S_3) possible genotypes at the superlocus. In contrast, monogyne males have only two possible genotypes, S_1 and S_2, because the S_3

allele corresponding to the *Pgm-3ᵃ/Gp-9ᵇ* haplotype is absent from the monogyne population. Finally, the overall frequency of genotypes in the male mating pool is denoted by an overbar, so, for instance, the frequency of male genotype i in the pool of reproductive males is

$$\bar{M}_i = mM_i^* + (1 - m)M_i. \tag{1}$$

Note that equation (1) is simply a weighted average of the allele frequencies in monogyne (M_i^*) and polygyne (M_i) males, with the weights given by the frequencies with which polygyne queens mate with monogyne (m) or polygyne ($1 - m$) males.

Superlocus Model

By working through a mating table, we find that after one generation the polygyne queen genotype frequencies become

$$Q'_{13} = Q_{13}\bar{M}_3 + \bar{M}_1 \quad \text{and} \tag{2a}$$

$$Q'_{23} = Q_{23}\bar{M}_3 + \bar{M}_2, \tag{2b}$$

whereas the new male frequencies are simply

$$M'_1 = \frac{1}{2}Q_{13}, \tag{3a}$$

$$M'_2 = \frac{1}{2}Q_{23}, \quad \text{and} \tag{3b}$$

$$M'_3 = \frac{1}{2}. \tag{3c}$$

Note that $M_3^{(t)} = \frac{1}{2}$ for all $t \geq 1$. The frequencies of the six worker genotypes in the next generation are similarly found to be

$$W'_{11} = \frac{Q_{13}\bar{M}_1}{2 - s\bar{M}_3}, \tag{4a}$$

$$W'_{12} = \frac{Q_{13}\bar{M}_2 + Q_{23}\bar{M}_1}{2 - s\bar{M}_3}, \tag{4b}$$

$$W'_{13} = \frac{Q_{13}\bar{M}_3 + \bar{M}_1}{2 - s\bar{M}_3}, \tag{4c}$$

$$W'_{22} = \frac{Q_{23}\bar{M}_2}{2 - s\bar{M}_3}, \tag{4d}$$

$$W'_{23} = \frac{Q_{23}\bar{M}_3 + \bar{M}_2}{2 - s\bar{M}_3}, \quad \text{and} \tag{4e}$$

$$W'_{33} = \frac{(1 - s)\bar{M}_3}{2 - s\bar{M}_3}. \tag{4f}$$

These recursions apply for all combinations of gene flow and worker selection, where in the absence of gene flow ($m = 0$), the gene frequency $\bar{M}_i$ in the overall male gene pool is simply the corresponding frequency M_i in polygyne males.

Using equations (2)–(4), we may obtain explicit time-dependent solutions for the genotype frequencies in polygyne queens, males, and workers in every generation $t \geq 1$ (Appendices 1, 2). These show that both the dynamical and equi-

librium behavior vary substantially depending on whether polygyne queens mate with monogyne males. In the presence of such gene flow ($m > 0$, Appendix 1), the population converges to a single equilibrium that is uniquely determined by the allele frequencies in the monogyne male population (M_1^* and $M_2^* = 1 - M_1^*$). In this case, the final genotypic frequencies for the three polygyne castes are

$$\hat{Q}_{13} = M_1^* \quad \text{and} \tag{5a}$$

$$\hat{Q}_{23} = 1 - M_1^* \tag{5b}$$

for queens,

$$\hat{M}_1 = \frac{1}{2}M_1^*, \tag{6a}$$

$$\hat{M}_2 = \frac{1}{2}(1 - M_1^*), \quad \text{and} \tag{6b}$$

$$\hat{M}_3 = \frac{1}{2} \tag{6c}$$

for males, and

$$\hat{W}_{11} = \frac{(1 + m)(M_1^*)^2}{4 - s(1 - m)}, \tag{7a}$$

$$\hat{W}_{12} = \frac{2(1 + m)M_1^*(1 - M_1^*)}{4 - s(1 - m)}, \tag{7b}$$

$$\hat{W}_{13} = \frac{2M_1^*}{4 - s(1 - m)}, \tag{7c}$$

$$\hat{W}_{22} = \frac{(1 + m)(1 - M_1^*)^2}{4 - s(1 - m)}, \tag{7d}$$

$$\hat{W}_{23} = \frac{2(1 - M_1^*)}{4 - s(1 - m)}, \quad \text{and} \tag{7e}$$

$$\hat{W}_{33} = \frac{(1 - s)(1 - m)}{4 - s(1 - m)} \tag{7f}$$

for workers, where the equilibrium formulas rely on the final frequencies $\hat{\bar{M}}_1 = \frac{1}{2}(1 + m)M_1^*$, $\hat{\bar{M}}_2 = \frac{1}{2}(1 + m)(1 - M_1^*)$, and $\hat{\bar{M}}_3 = \frac{1}{2}(1 - m)$ in the overall pool of males.

If polygyne queens do not mate with monogyne males ($m = 0$, Appendix 2), however, the equilibrium frequencies in the polygyne population are determined by the initial genotype frequencies in polygyne queens and males, with

$$\hat{Q}_{13} = \frac{2}{3}Q_{13}^{(0)}M_3^{(0)} + \frac{1}{3}Q_{13}^{(0)} + \frac{2}{3}M_1^{(0)} \quad \text{and} \tag{8a}$$

$$\hat{Q}_{23} = 1 - \hat{Q}_{13} \tag{8b}$$

in queens,

$$\hat{M}_1 = \frac{1}{2}\hat{Q}_{13}, \tag{9a}$$

$$\hat{M}_2 = \frac{1}{2}(1 - \hat{Q}_{13}), \quad \text{and} \tag{9b}$$

$$\hat{M}_3 = \frac{1}{2} \tag{9c}$$

in males, and

TABLE 2. Genotype frequencies of queens and workers at the joint *Pgm-3/Gp-9* superlocus for three sites in a single polygyne population of *Solenopsis invicta* in northern Georgia, United States.

Site	Queen genotype frequencies						Worker genotype frequencies					
	S_1S_1	S_1S_2	S_1S_3	S_2S_2	S_2S_3	S_3S_3	S_1S_1	S_1S_2	S_1S_3	S_2S_2	S_2S_3	S_3S_3
B	0	0	0.735	0	0.265	0	0.195	0.115	0.540	0.011	0.138	0
D	0	0	0.747	0	0.253	0	0.160	0.049	0.667	0.012	0.111	0
E	0	0	0.795	0	0.205	0	0.214	0.083	0.571	0.012	0.119	0
All sites	0	0	0.758	0	0.242	0	0.190	0.083	0.591	0.012	0.123	0

$$\hat{W}_{11} = \frac{(\hat{Q}_{13})^2}{4 - s}, \tag{10a}$$

$$\hat{W}_{12} = \frac{2\hat{Q}_{13}(1 - \hat{Q}_{13})}{4 - s}, \tag{10b}$$

$$\hat{W}_{13} = \frac{2\hat{Q}_{13}}{4 - s}, \tag{10c}$$

$$\hat{W}_{22} = \frac{(1 - \hat{Q}_{13})^2}{4 - s}, \tag{10d}$$

$$\hat{W}_{23} = \frac{2(1 - \hat{Q}_{13})}{4 - s}, \quad \text{and} \tag{10e}$$

$$\hat{W}_{33} = \frac{1 - s}{4 - s} \tag{10f}$$

in workers.

MAXIMUM-LIKELIHOOD ESTIMATION

The superlocus model provides the necessary mathematical foundation to construct a maximum likelihood procedure (Edwards 1992; Weir 1996) to formally estimate the rate of male-mediated gene flow ($\check{m}$) and the strength of selection acting on S_3S_3 workers ($\check{s}$) in polygyne populations. Because models with additional parameters generally provide a better fit to data, we proceed hierarchically and allow for both, one, or neither of the two processes to be present. In each instance, estimation is based on the parameter values whose equilibrium genotypic frequencies are closest to the observed genotypic counts. This approach assumes the polygyne population is at equilibrium, a supposition that is justified here by the close agreement found between the data collected for this study and those obtained in previous surveys (Ross 1992, 1997; Ross and Shoemaker 1993; Ross and Keller 1995).

The final step is to formally determine which, if any, of these scenarios can account parsimoniously for the superlocus structure of introduced *S. invicta*. This is achieved by comparing the *G*-tests for goodness-of-fit between the ob-served and expected genotypic counts under each of the four models. In general, model α provides superior fit over model β, which has a greater number of degrees of freedom than model α, if the difference in the *G*-statistics, $G_{Dif} = G_\beta - G_\alpha$, with degrees of freedom $df_{Dif} = df_\beta - df_\alpha$, is significant (Fienberg 1977).

Monogyne Male Gene Flow and Worker Selection

We first develop the estimation procedure for the full model with both monogyne male gene flow ($m > 0$) and worker selection ($s > 0$). In the presence of gene flow, the equilibrium frequencies for polygyne queens and workers, given by equations (5) and (7), depend on the allele frequency in the monogyne male population (M_1^*), as well as m and s. The likelihood equation thus takes the form

$$L(M_1^*, m, s) = C \prod_{i=1}^{2} \hat{Q}_i^{N_i^Q} \prod_{j=1}^{6} \hat{W}_j^{N_j^W} \prod_{k=1}^{2} (M_k^*)^{N_k^{M*}}, \tag{11}$$

where C is the constant multinomial coefficient, the first product is over the two polygyne queen classes (S_1S_3 and S_2S_3), the second product over the six polygyne worker classes ($S_1S_1, \ldots, S_3S_3$), and the third product over the two monogyne male classes (S_1 and S_2). The genotypic counts for the queens, workers, and monogyne males are denoted by N_i^Q, N_j^W, and N_k^M, respectively (Table 3).

Differentiation of the corresponding log-likelihood function immediately yields the maximum-likelihood estimate (MLE) for the frequency of monogyne males with the S_1 genotype,

$$\check{M}_1^* = \frac{N_{13}^Q + N_1^{M*} + 2N_{11}^W + N_{12}^W + N_{13}^W}{N^Q + N^{M*} + 2A + B}, \tag{12}$$

where N^Q and N^{M*} are the total number of polygyne queens and monogyne males in the sample, $A = N_{11}^W + N_{12}^W + N_{22}^W$, and $B = N_{13}^W + N_{23}^W$. The corresponding MLEs for the rate of monogyne male gene flow and selection coefficient against S_3S_3 workers are

TABLE 3. Genotype frequencies, counts, and viabilities for polygyne queens, males, and workers at the *Pgm-3/Gp-9* superlocus.

	Queen genotype						Male genotype[1]			Worker genotype					
	S_1S_1	S_1S_2	S_1S_3	S_2S_2	S_2S_3	S_3S_3	S_1	S_2	S_3	S_1S_1	S_1S_2	S_1S_3	S_2S_2	S_2S_3	S_3S_3
Frequency	Q_{11}	Q_{12}	Q_{13}	Q_{22}	Q_{23}	Q_{33}	M_1	M_2	M_3	W_{11}	W_{12}	W_{13}	W_{22}	W_{23}	W_{33}
Count	N_{11}^Q	N_{12}^Q	N_{13}^Q	N_{22}^Q	N_{23}^Q	N_{33}^Q	—[2]	—[2]	—[2]	N_{11}^W	N_{12}^W	N_{13}^W	N_{22}^W	N_{23}^W	N_{33}^W
Viability	0	0	1	0	1	0	1	1	1	1	1	1	1	1	$1 - s$

[1] Corresponding values in monogyne males are denoted by a superscripted asterisk.
[2] Males cannot be scored at *Gp-9*.

$$\check{m} = \frac{2A - B}{B} \quad \text{and} \quad (13)$$

$$\check{s} = \frac{A - B + N_{33}^W}{A - B}. \quad (14)$$

Equations (12)–(14) determine admissible parameter estimates in the interval (0,1) if and only if $N_{33}^W > 0$ and $\frac{1}{2}B < A < B - N_{33}^W$.

The approximate large sample variances for the three parameter estimates (M_1^*, m, and s) can be obtained by inverting the 3×3 information matrix for the system. We find that the resulting variance for the monogyne male frequency,

$$\text{Var}(\check{M}_1^*) \cong \frac{M_1^*(1 - M_1^*)[4 - s(1 - m)]}{[4 - s(1 - m)](N^Q + N^{M*}) + 2(2 + m)N^W}, \quad (15)$$

involves m and s, as well as M_1^*, whereas those for m and s,

$$\text{Var}(\check{m}) \cong \frac{(1 + m)(3 + m)[4 - s(1 - m)]}{2N^W} \quad \text{and} \quad (16)$$

$$\text{Var}(\check{s}) \cong \frac{(1 - s)[4 - s(1 - m)][4 - s(3 + m)]}{(1 - m)^2 N^W} \quad (17)$$

are independent of M_1^*. All three variance estimates in equations (15)–(17) are inversely related to N^W, the total number of workers in the sample.

Monogyne Male Gene Flow Only

If the polygyne population is subject only to monogyne male gene flow, with no selection against S_3S_3 workers ($0 < m \leq 1$, $s = 0$), the equilibrium in equations (5) and (7) is determined by the frequency of S_1 monogyne males (M_1^*) and rate of monogyne male gene flow (m). The estimate and variance for M_1^* are the same as those in equations (12) and (15) for the full model, whereas those for m are now

$$\check{m} = \frac{A - N_{33}^W}{A + N_{33}^W} \quad \text{and} \quad (18)$$

$$\text{Var}(\check{m}) \cong \frac{2(1 - m^2)}{N^W}. \quad (19)$$

Equation (18) determines an admissible MLE for m in (0, 1) if and only if $0 < N_{33}^W < A$; the MLE for m is zero if $N_{33}^W \geq A$ and one if $N_{33}^W = 0$.

Worker Selection Only

In the presence of worker selection alone ($m = 0$, $0 < s \leq 1$), the polygyne equilibrium in equations (8) and (10) is a function of the selection coefficient against S_3S_3 workers (s) and the initial genotypic frequencies in queens ($Q_{13}^{(0)}$) and males ($M_1^{(0)}$, $M_3^{(0)}$). However, because the final frequencies are fully determined by s and the equilibrium frequency of S_1S_3 queens ($\hat{Q}_{13}$), it suffices to estimate only these latter two quantities. The likelihood equation with only worker selection then takes the form

$$L(\hat{Q}_{13}, s) = C \prod_{i=1}^{2} \hat{Q}_i^{N_i^Q} \prod_{j=1}^{6} \hat{W}_j^{N_j^W}, \quad (20)$$

where C is the constant multinomial coefficient, the first product is over the two queen classes, and the second product is over the six worker classes.

Under this approach, the MLE and approximate variance for the equilibrium frequency of S_1S_3 queens are readily found to be

$$\check{Q}_{13} = \frac{N_{13}^Q + 2N_{11}^W + N_{12}^W + N_{13}^W}{N^Q + 2A + B} \quad \text{and} \quad (21)$$

$$\text{Var}(\check{Q}_{13}) \cong \frac{(4 - s)Q_{13}(1 - Q_{13})}{(4 - s)N^Q + 4N^W}, \quad (22)$$

whereas those for the selection coefficient, s, against S_3S_3 polygyne workers are

$$\check{s} = \frac{N^W - 4N_{33}^W}{N^W - N_{33}^W} \quad \text{and} \quad (23)$$

$$\text{Var}(\check{s}) \cong \frac{(1 - s)(4 - s)^2}{3N^W}. \quad (24)$$

Equation (23) determines a valid MLE for s in (0, 1) if and only if $0 < N_{33}^W < \frac{1}{4}N^W$; the MLE for s is zero if $N_{33}^W \geq \frac{1}{4}N^W$ and one if $N_{33}^W = 0$.

Neither Monogyne Male Gene Flow nor Worker Selection

Lastly, under the simplest model, corresponding to the null hypothesis of neither monogyne male gene flow nor worker selection ($m = 0$, $s = 0$), the final frequency of polygyne S_1S_3 queens ($\hat{Q}_{13}$) alone determines the polygyne equilibrium in equations (8) and (10). The MLE and approximate variance for $\hat{Q}_{13}$ are as in equations (21) and (22), with $s = 0$.

ESTIMATES OF GENE FLOW AND SELECTION

We now use the superlocus framework to obtain the MLEs for the rate of monogyne male gene flow ($\check{m}$) and the strength of selection operating on polygyne workers ($\check{s}$) from the observed genotypic counts (Table 3). Our hierarchical estimation procedure is applied to data collected from both this (current) study and an earlier (1997) investigation of this polygyne population (Ross 1997). In both cases, the counts for monogyne males that we use are derived from a different, previous study, which reported that, of 108 males sampled, 88 were genotype *Pgm-3^A* (N_1^{M*}) and 20 were genotype *Pgm-3^a* (N_2^{M*}) (Ross 1992). Because the monogyne population is fixed for the *Gp-9^B* allele, these *Pgm-3* genotypes correspond directly to the superlocus genotypes S_1 and S_2, respectively.

Current Study

Table 4 displays the MLEs for the rate of monogyne male gene flow ($\check{m}$) and the strength of worker selection ($\check{s}$), the expected genotype frequencies in the polygyne population under the estimates, and the fit of our models to the data under all four possible combinations of gene flow and selection. We find that the simplest model (Table 4, first panel), with no monogyne male gene flow or worker selection ($m = 0$, $s = 0$), fails to fit the data ($P < 0.001$). This poor fit does not result from deviations in the polygyne queen frequencies;

TABLE 4. Summary of the fit of the *Pgm-3/Gp-9* superlocus model with and without gene flow via monogyne males (gene flow) and with and without selection against S_3S_3 workers (selection). The maximum-likelihood estimates for the male-mediated gene flow rate ($\check{m}$) and worker selection coefficient ($\check{s}$) are given, where appropriate, along with the expected equilibrium queen and worker genotype frequencies. The significance of the deviation of the observed frequencies from those expected (fit) was determined via a *G*-test for goodness-of-fit.

Gene flow	Selection	Site	$\check{m}$	$\check{s}$	$\hat{Q}_{13}$	$\hat{Q}_{23}$	$\hat{W}_{11}$	$\hat{W}_{12}$	$\hat{W}_{13}$	$\hat{W}_{22}$	$\hat{W}_{23}$	$\hat{W}_{33}$	Fit
no	no	B	—[1]	—[1]	0.768	0.232	0.147	0.089	0.384	0.013	0.116	0.250	$G_4 = 45.16$***
		D			0.805	0.195	0.162	0.078	0.402	0.010	0.098	0.250	$G_4 = 38.79$***
		E			0.814	0.186	0.166	0.076	0.407	0.009	0.093	0.250	$G_4 = 42.16$***
		all sites			0.795	0.205	0.158	0.081	0.397	0.011	0.103	0.250	$G_4 = 131.8$***
no	yes	B	—[1]	1.0	0.768	0.232	0.197	0.119	0.512	0.018	0.155	0.000	$G_3 = 1.04$
		D			0.805	0.195	0.216	0.105	0.536	0.013	0.130	0.000	$G_3 = 8.30$*
		E			0.814	0.186	0.221	0.101	0.543	0.012	0.124	0.000	$G_3 = 0.64$
		all sites			0.795	0.205	0.211	0.109	0.530	0.014	0.137	0.000	$G_3 = 6.14$
yes	no	B	1.0	—[1]	0.735	0.265	0.307	0.169	0.392	0.023	0.108	0.000	$G_3 = 12.59$**
		D			0.747	0.253	0.326	0.155	0.404	0.018	0.096	0.000	$G_3 = 30.01$***
		E			0.814	0.186	0.331	0.151	0.407	0.017	0.093	0.000	$G_3 = 12.91$**
		all sites			0.798	0.202	0.318	0.161	0.399	0.020	0.101	0.000	$G_3 = 51.34$***
yes	yes	B	0.0	1.0	0.768	0.232	0.197	0.119	0.512	0.018	0.155	0.000	$G_2 = 1.04$
		D			0.805	0.195	0.216	0.105	0.536	0.013	0.130	0.000	$G_2 = 8.30$*
		E			0.814	0.186	0.221	0.101	0.543	0.012	0.124	0.000	$G_2 = 0.64$
		all sites			0.795	0.205	0.211	0.109	0.530	0.014	0.137	0.000	$G_2 = 6.14$

[1] Not applicable.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

the MLEs for the equilibrium queen frequencies fit the observed frequencies (Table 2) quite closely. Rather, the lack of fit arises because S_3S_3 workers are expected to be present at a frequency of 0.25 at equilibrium, but no workers of this genotype are present in our data set.

Proceeding to the second tier of models with only one of the evolutionary processes, we find that selection against polygyne S_3S_3 workers and no monogyne male gene flow ($m = 0$, $0 < s \leq 1$) fits both the queen and worker data well (Table 4, second panel). (The observed distribution of genotypes at site D differs only slightly from that expected [$P = 0.04$].) The MLEs $\pm$ SE are $\check{s} = 1.0 \pm 0.0$ at all sites, corresponding to the complete inviability of S_3S_3 workers. In contrast, the model with gene flow only ($0 < m \leq 1$, $s = 0$) gives poor fit to the data (Table 4, third panel), even though the MLEs of $\check{m} = 1.0 \pm 0.0$ at all sites predict the observed lack of polygyne S_3S_3 workers. Although only the selection model fits the data, both of the second tier models provide significant improvement in fit over the simplest model with neither monogyne male gene flow nor worker selection. The improvement in fit with worker selection only is $G_1 = 44.12$, 30.49, 41.52, and 125.76 for sites B, D, E, and all sites combined, with $P < 0.001$; the improvement in fit with male gene flow only is $G_1 = 32.57$, 8.78, 29.25, and 80.56 for sites B, D, E, and all sites combined, with $P < 0.01$.

When we attempt to fit the full model on the third tier of our hierarchy with both monogyne male gene flow and worker selection ($0 < m \leq 1$, $0 < s \leq 1$) to our data, we find that equation (13) yields values of $\check{m} < 0.0$ at all sites. Therefore, the MLEs for the male gene flow rate and the worker selection coefficient must lie on the boundary of their admissible parameter space, $[0, 1] \times [0, 1]$. (The MLE for M_1^* is independent of those for m and s and so is not affected.) This complication was addressed by mathematically analyzing the series of nested models that correspond to the four sides of the (m, s) region, and then calculating the 95% confidence

intervals for the resulting estimates numerically, by bootstrapping the original data set 1000 times. The results show that the best-fitting parameters for the full model (Table 4, fourth panel) correspond to those for the best-fitting simpler model, with worker selection only; the MLEs are $\check{m} = 0.0$ and $\check{s} = 1.0$ at all sites, with no variance in $\check{s}$ and 95% confidence intervals for $\check{m}$ of [0.0, 0.41], [0.0, 0.0], [0.0, 0.36], and [0.0, 0.05] for sites B, D, E, and all sites combined, respectively. We therefore conclude that complete inviability of S_3S_3 workers, in conjunction with selection in queens, not only fits the data well without any gene flow, but also provides the most parsimonious explanation for the distribution of polygyne genotypes in this population.

1997 Study

The first study of *Gp-9* reported *Pgm-3/Gp-9* genotypes in females from a site east of site E (Fig. 1) within the polygyne population examined in this investigation (Ross 1997). We can test the robustness of our conclusions by applying our theoretical framework to these earlier data to see if they are also best explained by female selection alone. Queens and workers were collected from 106 and 146 nests, respectively. After correcting for the nest-level structure, we find that the 1997 superlocus frequencies (S_1S_1, . . . , S_3S_3) were 0.0, 0.0, 0.783, 0.0, 0.217, 0.0 in queens and 0.226, 0.158, 0.459, 0.014, 0.144, 0.0 in workers.

We again find that the simplest model, with no gene flow and no worker selection, fits the data poorly ($G_4 = 75.51$, $P < 0.001$), and that both second-tier models provide significant improvement in fit (improvement in fit with gene flow only: $G_1 = 68.18$, $P < 0.001$; improvement in fit with worker selection only: $G_1 = 71.98$, $P < 0.001$). Although the 1997 data are well accounted for by either monogyne male gene flow ($G_3 = 7.33$, $P > 0.05$, $\check{m} = 1.0 \pm 0.0$) or selection on S_3S_3 workers ($G_3 = 3.53$, $P > 0.3$, $\check{s} = 1.0 \pm 0.0$), the

selection model still fits better. When we move up to the third tier, we find that the combination of gene flow and worker selection yields good fit to the 1997 superlocus frequencies ($G_2 = 1.13$, $P > 0.5$), with $\check{m} = 0.32 \pm 0.22$ and $\check{s} = 1.0 \pm 0.0$. However, the full model does not provide a significant improvement in fit over the model of worker selection alone ($G_1 = 2.4$, $P > 0.1$), and the estimate of $\check{m}$ does not differ significantly from zero. Overall, the 1997 data thus support the conclusion of the current study, that selection in the queen caste and inviability of polygyne S_3S_3 workers is sufficient to account for the genotypic distribution in this polygyne population. There is no need to assume any monogyne male gene flow.

DISCUSSION

Extensive studies of the introduced fire ant *S. invicta* suggest that selection and gene flow strongly affect the patterns of genetic variation within polygyne populations (Ross and Keller 1995; Ross 1997). Most recently, it has been hypothesized that genotypic variation at the loci *Pgm-3* and *Gp-9* can be explained by selection acting solely on *Gp-9* in females and gene flow occurring via monogyne males mating with polygyne queens. We have formally tested this current hypothesis by fitting mathematical models that incorporate these two evolutionary processes to two sets of data collected from a polygyne population.

Selection in Workers

Our maximum-likelihood results strongly affirm that selection operates on *Gp-9* genotype in the worker caste. In particular, both data sets indicate that selection against *Gp-9bb* workers, in conjunction with the observed inviability of *Gp-9BB* and *Gp-9bb* queens, provides a parsimonious explanation of the observed polygyne genotype frequencies. The importance of worker selection found here agrees with previous laboratory and field investigations that have consistently found a deficit of *Gp-9bb* (S_3S_3) workers (Ross 1997; Goodisman et al. 1999). The nature of this selection is unclear, although recent studies suggest that it may be due, in part, to a significant correlation between *Gp-9* genotype and polygyne worker mass, with *Gp-9BB* > *Gp-9Bb* > *Gp-9bb* (Goodisman et al. 1999). These size differences may lead to differences in viability or longevity of workers, with larger (*Gp-9BB* and *Gp-9Bb*) workers living longer than smaller (*Gp-9bb*) workers (Porter and Tschinkel 1985, 1986; Tschinkel 1993).

The size effect from *Gp-9* genotype could compromise the interpretation of our data if worker genotype frequencies were obtained in a biased fashion. This is because it is more difficult to assay the genotype of smaller workers, which would more often be *Gp-9bb* (or *Gp-9Bb*); therefore, it is possible that the frequency of *Gp-9Bb* and *Gp-9bb* workers is greater than has been reported. However, we do not believe that any such bias can completely account for the deficit of polygyne *Gp-9bb* workers because the distributions of masses of workers of all three *Gp-9* genotypes show a large degree of overlap (Goodisman et al. 1999), and genotype *Gp-9bb* also is strongly underrepresented among winged queens, all of which are readily genotyped (Ross 1997). Thus, we conclude that the viabilities of workers probably differ strongly across different *Gp-9* genotypes.

Gene Flow via Monogyne Males

Our analyses indicate that selection in workers and queens in the absence of male gene flow from monogyne populations can parsimoniously explain the distribution of female genotypes in polygyne populations. Nevertheless, indirect evidence for matings between polygyne queens and monogyne males does exist. For instance, the allele frequencies at more than 25 presumably neutral nuclear makers show a high degree of similarity between neighboring monogyne and polygyne populations, thus suggesting the two social forms are linked by gene flow (Ross and Shoemaker 1993; Ross et al. 1999). Also, our analyses of the distribution of queen and worker genotypes from the 1997 study of this polygyne population indicate that some gene flow (with $\check{m} = 0.32 \pm 0.22$) may occur in conjunction with selection in workers. The 1997 study also directly examined progeny arrays from polygyne queens and found that 90% of the males that mated with these queens had genotype *Gp-9B*. However, only 50% of polygyne males are expected to have this genotype, whereas 100% of monogyne males do. Therefore, if males of genotype *Gp-9B* and *Gp-9b* had equal viability and mating success, 80% of the polygyne queens examined must have mated with monogyne males (Ross 1997).

Another indication that polygyne queens mate with monogyne males comes from the cline in the frequency of mated queens at our three study sites. The proportion of mated queens was highest at site B, which lies on the border between the monogyne and polygyne populations, whereas mated queens were far less frequent at sites D and E, which lie well within the polygyne population. Since a large proportion of polygyne males are sterile, this pattern is consistent with queens close to the monogyne population having greater access to fertile (monogyne) male mates (Ross and Fletcher 1985, 1986; Ross and Keller 1995).

A final piece of evidence that supports the existence of some gene flow via monogyne males stems from the theoretical findings for the superlocus presented in this study. Our analytical results suggest that the frequencies of the S_1S_3 and S_2S_3 queens should ultimately equal the frequencies of monogyne males of genotype S_1 and S_2 if any monogyne male-mediated gene flow occurs. We find that the polygyne queen genotype frequencies do not differ significantly at any of the three sites or at all sites combined from the corresponding monogyne male frequencies ($z = 1.32$, 1.11, 0.33, and 1.08 for sites B, D, E, and all sites, respectively; $P > 0.1$ for all tests). The equilibrium queen genotype frequencies under the models without gene flow depend on the initial conditions, and it is thus possible that the male and corresponding queen genotype frequencies equal one another coincidentally. However, given the other lines of evidence, we feel that this result is not due to chance, but is instead explained by some level of monogyne male-mediated gene flow into the polygyne study population.

Selection, Gene Flow, and Genotypic Structure of Polygyne Solenopsis invicta

This study was motivated by previous work in polygyne *S. invicta* that attempted to explain patterns of genetic variation at the loci *Pgm-3* and *Gp-9* (Ross 1992, 1997; Ross and Shoemaker 1993; Ross and Keller 1995; Keller and Ross 1999). The most recent investigations suggested that variation at these loci was influenced by selection acting on *Gp-9* in combination with gene flow via conspecific monogyne males. However, our application of new theoretical models to both earlier and newly collected data indicate that the distribution of worker genotypes can be explained parsimoniously by selection on *Gp-9* in the queen and worker castes alone, without the need to invoke any gene flow via monogyne males. However, several other lines of evidence indicate that some interform matings do occur.

We suggest that the apparent discrepancy among these studies results from basing our mathematical approach exclusively on those processes for which strong empirical evidence already exists. The fact that our most parsimonious explanation appears to be at odds with empirical evidence thus implies that other evolutionary processes are operating within introduced polygyne *S. invicta.* Two prominent candidates are other forms of viability selection on polygyne workers or selection on polygyne males. Such selection is plausible, given that *Gp-9* genotype influences the mass of polygyne workers and males (Goodisman et al. 1999), and mass is known to be associated with worker viability and male reproductive strategies in ants (Hölldobler and Wilson 1990; Heinze and Tsuji 1995). Moreover, we cannot rule out the possibility that selection on other linked loci may influence the genotype frequencies within the population. Therefore, this investigation strongly indicates that the factors influencing the genotypic structure of introduced polygyne *S. invicta* are even more complex than previously suggested and calls for further empirical and theoretical studies to identify these additional evolutionary forces.

LITERATURE CITED

Barton, N. H., and A. G. Clark. 1990. Population structure and processes in evolution. Pp. 115–173 *in* K. Wöhrmann and S. K. Jain, eds. Population biology: ecological and evolutionary viewpoints. Springer, Berlin.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113–148.

Chevillon, C., Y. Rivet, M. Raymond, F. Rousset, P. E. Smouse, and N. Pasteur. 1998. Migration/selection balance and ecotypic differentiation in the mosquito *Culex pipiens.* Mol. Ecol. 7: 197–208.

Edwards, A. W. F. 1992. Likelihood. Johns Hopkins Univ. Press, Baltimore, MD.

Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton Univ. Press, Princeton, NJ.

Felsenstein, J. 1976. The theoretical population genetics of variable selection and migration. Annu. Rev. Ecol. Syst. 10:253–280.

Fienberg, S. E. 1977. The analysis of cross-classified categorical data. MIT Press, Cambridge, U.K.

Goodisman, M. A. D., and K. G. Ross. 1999. Queen recruitment in a multiple-queen population of the introduced fire ant *Solenopsis invicta.* Behav. Ecol. 10:428–435.

Goodisman, M. A. D., P. D. Mack, D. E. Pearse, and K. G. Ross. 1999. Effects of a single gene on worker and male mass in the fire ant *Solenopsis invicta* (Hymenoptera: Formicidae). Ann. Entomol. Soc. Am. 92:563–570.

Heinze, J., and K. Tsuji. 1995. Ant reproductive strategies. Res. Ecol. 7:135–149.

Hölldobler, B., and E. O. Wilson. 1990. The ants. The Belknap Press of Harvard University Press, Cambridge, MA.

Jones, J. S. 1982. More to melanism than meets the eye. Nature 300:109–110.

Jones, J. S., B. H. Leith, and P. Rawlings. 1977. Polymorphism in *Cepaea*: a problem with too many solutions? Annu. Rev. Ecol. Syst. 8:109–143.

Keller, L., and K. G. Ross. 1999. 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.

King, R. B., and R. Lawson. 1995. Color-pattern variation in Lake Erie water snakes: the role of gene flow. Evolution 49:885–896.

Livingstone, F. B. 1969. Gene frequency clines of the hemoglobin locus in various human populations and their simulation by models involving differential selection. Human Biol. 41:223–236.

May, R. M., J. A. Endler, and R. E. McMurtrie. 1975. Gene frequency clines in the presence of selection opposed by gene flow. Am. Nat. 109:659–676.

McNeilly, T. 1968. Evolution in closely adjacent plant populations. III. *Agrostis tenuis* on a small copper mine. Heredity 23:99–108.

Porter, S. D., and W. R. Tschinkel. 1985. Fire ant polymorphism (Hymenoptera: Formicidae): factors affecting worker size. Ann. Entomol. Soc. Am. 78:381–386.

———. 1986. Adaptive value of nanitic workers in newly founded red imported fire ant colonies (Hymenoptera: Formicidae). Ann. Entomol. Soc. Am. 79:723–726.

Raymond, M., and F. Rousset. 1995. GENEPOP (vers. 1.2): population genetics software for exact tests and ecumenicism. J. Hered. 86:248–249.

Riechert, S. E. 1993. Investigation of potential gene flow limitation of behavioral adaptation in an aridlands spider. Behav. Ecol. Sociobiol. 32:355–363.

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

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

Ross, K. G., and D. J. C. Fletcher. 1985. Genetic origin of male diploidy in the fire ant, *Solenopsis invicta* (Hymenoptera: Formicidae), and its evolutionary significance. Evolution 39: 888–903.

———. 1986. Diploid male production—a significant colony mortality factor in the fire ant *Solenopsis invicta* (Hymenoptera: Formicidae). Behav. Ecol. Sociobiol. 19:283–291.

Ross, K. G., and L. Keller. 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 D. D. Shoemaker. 1993. An unusual pattern of

gene flow between the two social forms of the fire ant *Solenopsis invicta.* Evolution 47:1595–1605.

———. 1997. Nuclear and mitochondrial genetic structure in two social forms of the fire ant *Solenopsis invicta*: insights into transitions to an alternate social organization. Heredity 78:590–602.

Ross, K. G., D. D. Shoemaker, M. J. B. Krieger, C. J. DeHeer, and L. Keller. 1999. Assessing genetic structure with multiple classes of molecular markers: a case study involving the introduced fire ant *Solenopsis invicta.* Mol. Biol. Evol. 16:625–543.

Shoemaker, D. D., J. T. Costa, and K. G. Ross. 1992. Estimates of heterozygosity in two social insects using a large number of electrophoretic markers. Heredity 69:573–582.

Slatkin, M. 1985. Gene flow in natural populations. Annu. Rev. Ecol. Syst. 16:393–430.

Tschinkel, W. R. 1993. Sociometry and sociogenesis of colonies of the fire ant *Solenopsis invicta* during one annual cycle. Ecol. Mono. 63:425–457.

Turner, J. R. G. 1977. Butterfly mimicry: the genetical evolution of an adaptation. Evol. Biol. 10:163–206.

Weir, B. S. 1996. Genetic data analysis. Vol. II. Sinauer Associates, Sunderland, MA.

Corresponding Editor: D. Wheeler

APPENDIX 1

Time-Dependent Solutions for Superlocus Frequencies with Gene Flow

The time-dependent solutions under the full model with gene flow via monogyne males and selection in workers are fully determined by the dynamics of polygyne S_1S_3 queens ($Q_{13}^{(t)}$) and S_1 males ($M_1^{(t)}$). Using the fact that $\bar{M}_3^{(t)} \equiv \frac{1}{2}(1 - m)$ after one generation of gene flow and mating, the queen and male recursions in equations (2) and (3) become

$$Q_{13}^{(t+1)} = \frac{1}{2}(1 - m)Q_{13}^{(t)} + (1 - m)M_1^{(t)} + mM_1^* \quad \text{and} \quad (A1a)$$

$$M_1^{(t+1)} = \frac{1}{2}Q_{13}^{(t)} \quad (A1b)$$

for $t \geq 1$. We may rewrite (A1a,b) as a matrix equation of the form, $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{b}$ for $t \geq 1$, where $\mathbf{x}_t = (Q_{13}^{(t)}, M_1^{(t)})^T$, $\mathbf{b} = (mM_1^*, 0)^T$, T denotes transpose, and the coefficient matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{pmatrix} \frac{1}{2}(1 - m) & 1 - m \\ \frac{1}{2} & 0 \end{pmatrix}. \quad (A2)$$

Iterating the matrix recursion now shows that

$$\mathbf{x}_t = \mathbf{A}^{t-1}\mathbf{x}_1 + \left(\sum_{i=0}^{t-2} \mathbf{A}^i\right)\mathbf{b} \quad \text{for } t \geq 2, \quad (A3)$$

where $\mathbf{A}^0 = \mathbf{I}$ is the 2×2 identity matrix. Equation (A3) may be expanded using the spectral decomposition, $\mathbf{A}^i = \mathbf{P}\Lambda^i\mathbf{P}^{-1}$. Here,

$$\Lambda^i = \begin{pmatrix} \lambda_1^i & 0 \\ 0 & \lambda_2^i \end{pmatrix}, \quad (A4)$$

where

$$\lambda_1, \lambda_2 = \frac{1 - m \pm \sqrt{(1 - m)(9 - m)}}{4} \quad (A5)$$

are the two eigenvalues of matrix $\mathbf{A}$, with $-1 < \lambda_2 < 0 < \lambda_1 < 1$, and the other matrix factors are

$$\mathbf{P} = \begin{pmatrix} 2\lambda_1 & 2\lambda_2 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad (A6)$$

$$\mathbf{P}^{-1} = \frac{1}{2(\lambda_1 - \lambda_2)} \begin{pmatrix} 1 & -2\lambda_2 \\ -1 & 2\lambda_1 \end{pmatrix}. \quad (A7)$$

Equation (A3) then becomes

$$\mathbf{x}_t = \mathbf{P}\Lambda^{t-1}\mathbf{P}^{-1}\mathbf{x}_1 + \mathbf{P}\left(\sum_{i=0}^{t-2} \Lambda^i\right)\mathbf{P}^{-1}\mathbf{b}, \quad (A8)$$

where

$$\sum_{i=0}^{t-2} \Lambda^i = \begin{pmatrix} \sum_{i=0}^{t-2} \lambda_1^i & 0 \\ 0 & \sum_{i=0}^{t-2} \lambda_2^i \end{pmatrix} = \begin{pmatrix} \frac{1 - \lambda_1^{t-1}}{1 - \lambda_1} & 0 \\ 0 & \frac{1 - \lambda_2^{t-1}}{1 - \lambda_2} \end{pmatrix}. \quad (A9)$$

After substantial algebra and simplification, we find that the genotype frequencies in polygyne queens and males in any generation $t \geq 2$ reduce to

$$Q_{13}^{(t)} = M_1^* + k_1\lambda_1^t + k_2\lambda_2^t \rightarrow \hat{Q}_{13} = M_1^* \quad (A10)$$

as $t \rightarrow \infty$ and

$$M_1^{(t)} = \frac{1}{2}(M_1^* + k_1\lambda_1^{t-1} + k_2\lambda_2^{t-1}) \rightarrow \hat{M}_1 = \frac{1}{2}M_1^* \quad (A11)$$

as $t \rightarrow \infty$, where

$$k_1 = \frac{Q_{13}^{(1)} - 2\lambda_2M_1^{(1)}}{\lambda_1 - \lambda_2} - \frac{mM_1^*}{(1 - \lambda_1)(\lambda_1 - \lambda_2)} \quad \text{and} \quad (A12)$$

$$k_2 = \frac{Q_{13}^{(1)} - 2\lambda_1M_1^{(1)}}{\lambda_2 - \lambda_1} - \frac{mM_1^*}{(1 - \lambda_2)(\lambda_2 - \lambda_1)}. \quad (A13)$$

From the recursions given by equations (2) and (3), we may write the frequencies of polygyne S_1S_3 queens and S_1 males in generation 1 as $Q_{13}^{(1)} = Q_{13}^{(0)}\bar{M}_3^{(0)} + \bar{M}^{(0)}$ and $M_1^{(1)} = \frac{1}{2}Q_{13}^{(0)}$, thereby expressing the time-dependent solutions given in (A10) and (A11) as functions of the initial genotype frequencies in queens and males. The exact time-dependent solutions for the genotype frequencies in polygyne workers may be obtained by substituting the dynamical solutions from equations (A10) and (A11) into the frequencies in the male mating pool in equation (1) and the worker recursions in (4). For example, the frequency of S_1S_1 workers with gene flow is given by

$$W_{11}^{(t)} = \frac{Q_{13}^{(t-1)}\bar{M}_1^{(t-1)}}{2 - s\bar{M}_3^{(t-1)}} = \frac{Q_{13}^{(t-1)}[mM_1^* + (1 - m)M_1^{(t-1)}]}{2 - s(1 - m)M_3^{(t-1)}} \quad (A14)$$

for all $t \geq 1$, where $M_3^{(t-1)} \equiv \frac{1}{2}$ for $t \geq 2$.

APPENDIX 2

Time-Dependent Solutions for Superlocus Frequencies with No Gene Flow

As in Appendix 1, the dynamics under the model with no gene flow are determined by those for the frequency of polygyne S_1S_3 queens ($Q_1^{(t)}$) and S_1 males ($M_1^{(t)}$). By utilizing the fact that $M_3^{(t)} \equiv \frac{1}{2}$ for all $t \geq 1$, we may write their recursions, given by (2) and (3), in the time-dependent forms,

$$Q_{13}^{(t+1)} = \frac{1}{2}Q_{13}^{(t)} + M_1^{(t)} \quad \text{and} \quad (A15a)$$

$$M_1^{(t+1)} = \frac{1}{2}Q_{13}^{(t)} \quad (A15b)$$

for all $t \geq 1$. These can be written as the matrix equation, $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t$ for $t \geq 1$, where $\mathbf{x}_t = (Q_{13}^{(t)}, M_1^{(t)})^T$ and the coefficient matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{pmatrix}. \quad (A16)$$

Iterating the matrix recursion yields $\mathbf{x}_t = \mathbf{A}^{t-1}\mathbf{x}_1$ for $t \geq 1$. We may use the spectral decomposition of $\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}^{-1}$ to rewrite this time-dependent solution as $\mathbf{x}_t = \mathbf{P}\Lambda^{t-1}\mathbf{P}^{-1}\mathbf{x}_1$, where

$$P = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}, \tag{A17a}$$

$$P^{-1} = \frac{1}{3} \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}, \quad \text{and} \tag{A17b}$$

$$\Lambda^{t-1} = \begin{pmatrix} \lambda_1^{t-1} & 0 \\ 0 & \lambda_2^{t-1} \end{pmatrix}. \tag{A17c}$$

Here, $\lambda_1 = 1$ and $\lambda_2 = -\frac{1}{2}$ are the two eigenvalues of matrix A, and the columns of matrix P are the corresponding right eigenvectors. After simplification, we find that

$$Q_{13}^{(t)} = \frac{2}{3}(Q_{13}^{(1)} + M_1^{(1)}) + \frac{1}{3}(Q_{13}^{(1)} - 2M_1^{(1)})\left(-\frac{1}{2}\right)^{t-1} \quad \text{and} \tag{A18}$$

$$M_1^{(t)} = \frac{1}{3}(Q_{13}^{(1)} + M_1^{(1)}) - \frac{1}{3}(Q_{13}^{(1)} - 2M_1^{(1)})\left(-\frac{1}{2}\right)^{t-1} \tag{A19}$$

for every generation $t \geq 1$.

Equations (A18) and (A19) yield time-dependent solutions in terms of the queen and male frequencies in generation $t = 1$. These may be rewritten in terms of the initial genotype frequencies by using recursions (2) and (3), which yield $Q_{13}^{(1)} = Q_{13}^{(0)}M_3^{(0)} + M_1^{(0)}$ and $M_1^{(1)} = \frac{1}{2}Q_{13}^{(0)}$. The equilibria for the system are found by letting $t \to \infty$ and are dependent on the initial genotype frequencies of queens and males in the polygyne population, with $Q_{13}^{(t)} \to \hat{Q}_{13} = \frac{2}{3}Q_{13}^{(0)}M_3^{(0)} + \frac{1}{3}Q_{13}^{(0)} + \frac{2}{3}M^{(0)}$ as $t \to \infty$ and $M_1^{(t)} \to \hat{M}_1 = \frac{1}{3}Q_{13}^{(0)}M_3^{(0)} + \frac{1}{6}Q_{13}^{(0)} + \frac{1}{3}M_1^{(0)}$ as $t \to \infty$.

The genotype frequencies in polygyne workers may be obtained by substituting the dynamical solutions (A18) and (A19) into the worker recursions in equation (4). For example, the time-dependent solution for the frequency of S_1S_1 workers with no gene flow is given by

$$W_{11}^{(t)} = \frac{Q_{13}^{(t-1)}M_1^{(t-1)}}{2 - sM_3^{(t-1)}} \tag{A20}$$

for $t \geq 1$, where $M_3^{(t-1)} \equiv \frac{1}{2}$ for $t \geq 2$.