

Molecular Evolutionary Analyses of the Odorant-Binding Protein Gene *Gp-9* in Fire Ants and Other *Solenopsis* Species

Michael J. B. Krieger* and Kenneth G. Ross†

*Center for Physics and Biology, Rockefeller University; and †Department of Entomology, University of Georgia

The fire ant *Solenopsis invicta* exists in two social forms, one with colonies headed by a single reproductive queen (monogyne form) and the other with colonies containing multiple queens (polygyne form). This variation in social organization is associated with variation at the gene *Gp-9*, with monogyne colonies harboring only the *B* allelic variant and polygyne colonies containing *b*-like variants as well. We generated new *Gp-9* sequences from 15 *Solenopsis* species and combined these with previously published sequences to conduct a comprehensive, phylogenetically based study of the molecular evolution of this important gene. The exon/intron structure and the respective lengths of the five exons of *Gp-9* are identical across all species examined, and we detected no evidence for intragenic recombination. These data conform to a previous suggestion that *Gp-9* lies in a genomic region with low recombination, and they indicate that evolution of the coding region in *Solenopsis* has involved point substitutions only. Our results confirm a link between the presence of *b*-like alleles and the expression of polygyny in all South American fire ant species known to possess colonies of both social forms. Moreover, phylogenetic analyses show that *b*-like alleles comprise a derived clade of *Gp-9* sequences within the socially polymorphic species, lending further support to the hypothesis that monogyny preceded polygyny in this group of fire ants. Site-specific maximum likelihood tests identified several amino acids that have experienced positive selection, two of which are adjacent to the inferred binding-pocket residues in the GP-9 protein. Four other binding-pocket residues are variable among fire ant species, although selection is not implicated in this variation. Branch-specific tests revealed strong positive selection on the stem lineage of the *b*-like allele clade, as expected if selection drove the amino acid replacements crucial to the expression of polygyne social organization. Such selection may have operated via the ligand-binding properties of GP-9, as one of the two amino acids uniquely shared by all *b*-like alleles is predicted to be a binding-pocket residue.

Introduction

A major distinction in the social organization of ant societies is the number of queens that inhabit a colony (Hölldobler and Wilson 1977). In an important pest species of fire ant, *Solenopsis invicta*, two coexisting social forms differ in colony queen number and other fundamental reproductive traits. Colonies of the single-queen (monogyne) form are headed by a single reproductive queen, whereas colonies of the multiple-queen (polygyne) form contain several to hundreds of such queens (Ross and Fletcher 1985; Vargo and Fletcher 1987). This variation in social organization has been shown to be associated with genotypes at the gene *Gp-9*. The genotypic pattern associated with each social form is remarkably simple: monogyne colonies harbor only the *B* allelic variant of *Gp-9*, whereas polygyne colonies harbor both the *B* variant and *b*-like variants (Shoemaker and Ross 1996; Ross 1997; Krieger and Ross 2002). The consistent absence of *b*-like alleles from inhabitants of monogyne colonies has led to the hypothesis that such alleles are essential for the expression of polygyny in *S. invicta* (Ross 1997; Ross and Keller 1998; Krieger 2005).

Determination of the nucleotide sequence of *Gp-9* and the predicted amino acid sequence of its protein product (Krieger and Ross 2002) revealed that it shares the highest sequence similarity with genes encoding pheromone-binding proteins (PBPs), a subclass of the odorant-binding protein (OBP) family (Pelosi and Maida 1995). Insect OBPs are expressed mainly in the antennae and are typified by six absolutely conserved cysteine residues located in characteristic positions (Pikielny et al. 1994). The exact molecular role of OBPs in olfaction is not well established

currently, but several lines of evidence suggest that at least some OBPs are responsible for the selective transport of pheromones and other semiochemicals from the cuticle to dendritic olfactory receptors (Leal 2003; Vogt 2003). Upon entering the pores in the olfactory sensilla, a hydrophobic odorant molecule is taken up in the binding pocket of the OBP, allowing its transport through the aqueous phase of the sensillar lymph. After having reached the surface of the dendrites, the odorant is ejected to a receptor, a process induced in some cases by a pH-dependent conformational change of the OBP-odorant complex (Lee et al. 2002).

Fire ant workers regulate the number and identity of egg-laying queens in their colony by accepting queens that produce appropriate chemical signals and destroying those that do not (Fletcher and Blum 1983; Keller and Ross 1998; Ross and Keller 2002). Thus, the core feature of colony social organization, the number of egg-laying queens, is mediated by worker recognition of, and subsequent discrimination among, individual queens. It is therefore reasonable to assume that the difference in colony queen number between the monogyne and polygyne forms of *S. invicta* is at least partly caused by differences in workers' abilities to recognize queens and that such different chemoreceptive abilities may be induced by distinct binding affinities of the two allelic forms of GP-9 for queen-produced pheromones. The inferred phylogeny of *Gp-9* sequences from several fire ant species that are closely related to *S. invicta* and exhibit a similar polymorphism in social organization revealed that two allelic variants invariably coexist in each of these species, one resembling the *B* allele and the other resembling the polygyny-inducing *b*-like alleles of *S. invicta* (Krieger and Ross 2002). This same study indicated that queens from polygyne nests always seem to carry *b*-like alleles, establishing a broader link between the presence of these alleles and the expression of polygyne

Key words: fire ants, *Gp-9*, odorant-binding proteins, polygyny, *Solenopsis*.

E-mail: krieger@rockefeller.edu.

Mol. Biol. Evol. 22(10):2090–2103. 2005

doi:10.1093/molbev/msi203

Advance Access publication June 29, 2005

social organization in the clade of fire ants comprising the closest relatives of *S. invicta*. Furthermore, positive selection was detectable only on branches in the clade of *b*-like alleles (Krieger and Ross 2002), a pattern consistent with the view that selection at the molecular level has been important in the derivation of polygyny from monogyny in South American fire ants.

Given that *Gp-9* appears to represent a gene of major effect on important fitness-related social traits in fire ants, further information on the molecular changes that occurred over the history of the gene may be expected to shed light on some general features of adaptive social evolution (Robinson, Grozinger, and Whitfield 2005). Thus, we characterized 25 new *Gp-9* sequences from 15 *Solenopsis* species in order to infer the major features of molecular evolution of the gene in this ant clade. We combined our new data with previously published sequences (Krieger and Ross 2002) to generate a comprehensive phylogeny of *Gp-9* encompassing virtually all known fire ant species and other important exemplars from the genus. Tests for selection were applied to the expanded gene phylogeny to search for additional evidence of non-neutral evolution of *Gp-9*. The data also were used to further explore the link between the presence of *b*-like alleles and the occurrence of polygyny in all South American species known to display polygynous social organization. Finally, we linked the amino acid variation found among allelic GP-9 proteins to the inferred protein structure in order to consider the potential effects of such substitutions on the binding properties of the variant proteins.

Materials and Methods

Samples

Samples from which *Gp-9* was sequenced for this study were chosen based on the following criteria: (1) exemplars should represent the major lineages within *Solenopsis*, as reflected in the current classification of the genus (Ettershank 1966; Trager 1991; Pitts 2002), (2) as many species as possible among the 22 described species of fire ants and their social parasites (Pitts 2002) should be included, (3) multiple samples from throughout the ranges of the geographically widespread species *Solenopsis geminata* and *Solenopsis saevissima* should be included, and (4) any newly uncovered *b*-like alleles should be sequenced. New sequences were obtained from the following *Solenopsis* specimens collected from localities in North, Central, and South America (state/province and country in parentheses): *Solenopsis daguerrei* (Chaco, Argentina), *Solenopsis electra* (Santiago del Estero, Argentina), *S. geminata* (Espírito Santo, Pará, and Maranhão, Brazil; Chiapas, Mexico; five specimens), *Solenopsis nigella gensterblumi* (Mato Grosso do Sul, Brazil), *Solenopsis megergates* (Rio Grande do Sul, Brazil; two specimens), *Solenopsis pusillignis* (Mato Grosso do Sul, Brazil), *Solenopsis quinquecuspis* (Santa Fe, Argentina), *Solenopsis richteri* (Rio Grande do Sul, Brazil), *S. saevissima* (Ceará, Bahia, Mato Grosso do Sul, and Paraná, Brazil; five specimens), *Solenopsis substituta* (Bahia, Brazil), *Solenopsis tridens* (Maranhão, Brazil), and *Solenopsis xyloni* (Arizona). New sequences

from three other *Solenopsis* species also were included: a specimen of the undescribed fire ant *Solenopsis* species "A" (Pitts 2002) from Brazil (Santa Catarina), two specimens of the undescribed fire ant *Solenopsis* species "X" (Ross and Trager 1990) from Argentina (Santa Fe), and a specimen of an undetermined thief ant species from Brazil (Minas Gerais) (thief ants are species of *Solenopsis* that typically form small colonies and live in close proximity to other ants whose brood and stored food they raid, traits that distinguish them from fire ants).

To supplement these new sequences, previously obtained *Gp-9* sequences from the following *Solenopsis* specimens also were included in our analyses (Krieger and Ross 2002; GenBank accession numbers AF427889–AF427906 and AF459414): *Solenopsis amblychila* (Arizona), *Solenopsis aurea* (Arizona), *S. geminata* (Florida), *Solenopsis globularia littoralis* (Arizona), *Solenopsis interrupta* (Santiago del Estero, Argentina), *S. invicta* (California, Florida, Georgia, and Texas; Formosa, Argentina; 22 specimens), *Solenopsis macdonaghi* (Corrientes, Argentina), *S. quinquecuspis* (Santa Fe, Argentina), *S. richteri* (Santa Fe, Argentina), and *S. saevissima* (Minas Gerais, Brazil).

Social organization (monogyny or polygyny) of virtually every sampled colony of the socially polymorphic fire ants was inferred on the basis of well-established diagnostic criteria, including numbers of wingless (reproductive) queens, spacing of nests in the field, and worker size distributions (e.g., Greenberg, Fletcher, and Vinson 1985; Porter et al. 1991). When multiple specimens from a single species and locality were sequenced, these always originated from separate colonies. All specimens included in the study were identified to the species level by James P. Pitts using the keys in Pitts (2002).

Laboratory Methods

DNA was extracted using the Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, Minn.), and the extracted DNA was cleaned using the DNA Clean & Concentrator-5 Kit (Zymo Research, Orange, Calif.). Filtered pipette tips were used for the isolation, cleaning, and amplification of DNA to avoid cross-contamination. Polymerase chain reactions (PCRs) were set up in a 20- μ l reaction mixture containing 1.5 U of proofreading DNA polymerase (Taq-Plus Precision, Stratagene, La Jolla, Calif.), PCR buffer, 0.5 mM deoxyribonucleotide triphosphates, 0.5 μ M of each primer, and 1 μ l of template DNA, and amplification was carried out using the following cycling profile: 30 cycles at 92°C (20 s), 58°C (30 s), and 72°C (1 min), with a final elongation step at 72°C for 5 min. Primer sequences used to amplify the *Gp-9* gene were identical to those used in our previous sequencing study (Krieger and Ross 2002). Initially, all *Gp-9* sequences were amplified with primers complementary to segments of the leader sequence and the 3' flanking region (*Gp-9*/33 forward: 5'-CATTCAAAGTACAGTAGAATAACTGCC-3' and *Gp-9*_2218 reverse: 5'-CAGGAGTTTGAGTTTGTACTGC-3', respectively), resulting in fragments of approximately 2,200 bp in length. DNA of species showing no clear amplification product was reamplified using a different reverse primer (*Gp-9*_490 reverse: 5'-GTATGCCAGCTGTTTTTAATTGC-3')

that anneals immediately downstream of the stop codon. The resulting PCR products were gel excised, cleaned (QIAquick Gel Extraction Kit, Qiagen, Valencia, Calif.), and cloned into a pCR 4-TOPO vector (Invitrogen, Carlsbad, Calif.), which was then used to transfect competent *Escherichia coli* cells (TOP10F, Invitrogen). A minimum of two clones was sequenced for each specimen. DNA sequences were obtained using the ABI PRISM 377 DNA Analyzer (Applied Biosystems) with the BigDye Terminator Cycle-Sequencing Kit (Applied Biosystems, Foster City, Calif.).

Most species of *Solenopsis* are assumed to form only monogyne colonies and thus are not expected to possess *b*-like *Gp-9* alleles. Nonetheless, the DNA of all newly sequenced specimens was tested for the presence of such alleles prior to sequencing by using an allele-specific PCR assay that was designed to detect alleles of this type by targeting the *b*-like isoleucine codon at position 139 (see Ross, Krieger, and Shoemaker 2003). In addition to the specimens that were sequenced, this allele-specific PCR assay was conducted on the following specimens in order to determine the taxonomic distribution of *b*-like alleles (single specimen/colony): *S. geminata* (11 specimens), *S. interrupta* (6 specimens), *S. megergates* (6 specimens), *S. pusillignis* (5 specimens), *S. saevissima* (64 specimens), *Solenopsis* species "X" (3 specimens), and *S. substituta* (20 specimens). DNA clones from the single specimen among these that was found to possess a *b*-like allele were sequenced repeatedly until the sequence of this allele was obtained (the individual was heterozygous for a *b*-like and a *B*-like allele).

Sequence and Phylogenetic Analyses

Gp-9 nucleotide sequences were aligned using the Genetics Computer Group program package 10.2 (2001; the alignment is available as Supplementary Material online). We checked the aligned sequences for evidence of recombination by employing two approaches (four methods). The first approach was undertaken to learn if recombination occurred in any ancestral *Solenopsis* lineages, possibly leading to mosaic sequences whose different parts have different evolutionary histories. The two methods employed to detect mosaic sequences were the difference of sums of squares (DSS) method that utilizes approximate distance-based phylogenetic methods (McGuire and Wright 2000), and the probabilistic divergence measure (PDM) method that uses a Markov chain Monte Carlo (MCMC) approach (Husmeier and Wright 2001). Both methods are implemented in TOPALi (Milne et al. 2004). For the DSS approach, which is suitable for a large number of sequences, we first removed all large gaps (>30 nt) and then treated the remaining gaps as missing values. Two different window sizes (250 and 500 bp), a step size of 10 bp, and the Felsenstein84 nucleotide substitution model were used. Because the PDM method is not suitable for a large number of sequences, we grouped our *Gp-9* sequences into 10 groups according to their sequence divergence and arbitrarily chose for analysis one sequence from each group (F. Wright and D. Husmeier, personal communication). Before analyzing the resulting reduced data set, sequences within each group were tested.

The second approach was undertaken to learn if recombination has occurred within the extant species of the socially polymorphic fire ant clade by using two population genetics methods. The first estimates the minimum number of recombination events based on the four-gamete test (Hudson and Kaplan 1985; Myers and Griffiths 2003), and the second uses a maximum likelihood (ML) method to estimate the recombination frequency (implemented in the LAMARC package; Kuhner et al. 2004). For the four-gamete test, we included four species from the socially polymorphic clade for which three or more sequences were available (*S. megergates*, *S. invicta*, *S. quinquecupis*, and *S. richteri*). For the ML analysis, we included only *S. invicta* (22 sequences) because too few sequences were available from other species to accurately assess recombination rates (Kuhner et al. 2004).

In order to choose an appropriate model of sequence evolution for the phylogenetic analyses, we used the procedure outlined by Huelsenbeck and Crandall (1997) and implemented in Modeltest 3.5 (Posada and Crandall 1998), which compares the likelihood scores from various models using a likelihood ratio test (LRT). An LRT is performed by comparing the LRT statistic ($2 \times [\log \text{likelihood}_2 - \log \text{likelihood}_1]$) to values of the chi-square distribution with degrees of freedom equal to the difference in the number of parameters between the two models. The Hasegawa-Kishino-Yano (HKY) model (Hasegawa, Kishino, and Yano 1985) with rate heterogeneity (HKY + Γ) was found to be the best fitting model of nucleotide substitution for our *Gp-9* sequences.

Maximum parsimony (MP), ML, minimum evolution (ME) (or neighbor-joining), and Bayesian inference (BI) were used for phylogenetic analyses of the nucleotide data. MP, ML, and ME analyses were performed with PAUP* 4.10 (Swofford 2002), while BI was performed with MrBayes 3.0 (Huelsenbeck and Ronquist 2001). Trees were rooted with the sequence from *S. globularia littoralis*, one of the study species believed to fall outside the clade including all fire ants (Trager 1991). Alternative rooting in which the thief ant sequence or both this sequence and that of *S. globularia littoralis* together were specified as out-groups yielded essentially identical topologies to those obtained using *S. globularia littoralis* alone as an out-group. For the MP, ML, and ME analyses, branch swapping was performed with the Tree Bisection-Reconnection algorithm. MP analysis was performed using heuristic searches with 100 random addition replicates, and gaps were treated as missing values. For the ML and ME trees, the HKY + Γ evolutionary model was applied, with Γ set to the value calculated with Modeltest 3.5. Confidence in the recovered nodes was assessed by nonparametric bootstrapping (Felsenstein 1985), with 100 replicates for the ML tree and 1,000 replicates for the MP and ME trees. BI was performed using the HKY model. Posterior probabilities of recovered nodes were calculated using the Metropolis-coupled MCMC approach. We ran four Markov chains simultaneously, three heated (temperature = 0.2) and one cold; each chain was started from a random tree and run for 1,000,000 generations, with sampling every 1,000th generation. Burn-in lengths were determined empirically from the likelihood values.

Selection Analyses

Several sequences differed only in their introns or flanking regions, so only a single representative of each unique coding-region sequence was used in the selection analyses. Examination of the data for evidence of selection was performed using the CODEML program in the PAML package 3.1.4 (Yang 1999), the ADAPTSITE program (Suzuki and Gojobori 1999; Suzuki 2004), and the method proposed by Zhang, Kumar, and Nei (1997).

The CODEML program uses an ML framework with codon-based models of sequence evolution to estimate the nonsynonymous (dN ; amino acid replacing) and synonymous (dS ; silent) substitution rate ratios (dN/dS or ω), which implicate negative selection when $\omega < 1$, neutral evolution when $\omega = 1$, or positive selection when $\omega > 1$ (Li 1997). We first applied site-specific models in which selection pressures can vary among codon sites, but the site-specific pattern is constrained to be identical across all lineages (Yang et al. 2000). Of the 14 models available (M0–M14), 6 (M0–M3, M7, and M8) have been recommended for practical use (Yang et al. 2000). These models have been described extensively in the literature (Yang et al. 2000), so we give only an abbreviated listing here: model M0—null model with no variation among sites; M1—“neutral” model, two site classes, with ω values fixed at 0 and 1; M2—“selection” model, three site classes, with two ω values fixed at 0 and 1 and the third estimated; M3—“discrete” model, three site classes, with all three ω values estimated; M7—“beta” model, eight site classes, with eight ω values ranging between 0 and 1 taken from a beta distribution; M8—“beta plus omega” model, eight site classes, with ω values taken from a beta distribution, as in model M7, plus an additional site class with an estimated ω value.

In addition to the site-specific selection models, we used models that assume the same dN/dS ratio for all sites but allow branch-specific variation in these ratios (Yang 1998). These likelihood models estimate two separate dN/dS ratios, one for a lineage of interest and one for all other lineages in the phylogeny, allowing a test of whether the dN/dS ratio for the focal lineage differs from the background ratio. We also included the “free-ratio” model, which estimates a different ω value for each branch in the tree. This analysis was used to estimate the numbers of nonsynonymous and synonymous substitutions and branch-specific ω values in the tree of coding-region sequences. Models that allow dN/dS to vary both among sites and branches (Yang et al. 2000) were not attempted because of their suspected unreliability (Zhang 2004). The statistical significance of instances of inferred selection was evaluated through an LRT that compared models that allow positive selection with those that do not.

The ADAPTSITE program (Suzuki and Gojobori 1999; Suzuki 2004) allows the inference of selection separately at each amino acid site, similar to the CODEML program (Yang 1999), but it uses MP instead of ML. The *Gp-9* tree employed in this approach was constructed using the NJBOOT program included in the LINTREE package (Takezaki, Rzhetsky, and Nei 1995). All subsequent steps, including tests for neutrality, were performed with pro-

grams included in the ADAPTSITE package, as recommended by the authors.

Finally, Zhang, Kumar, and Nei (1997) proposed a method to infer positive selection with a Fisher’s exact test. The method specifically tests whether $n/N = s/S$ along each branch of the sequence phylogeny (as expected under the null hypothesis of neutral evolution), where n and s are the numbers of nonsynonymous and synonymous substitutions per sequence, respectively, and N and S are the numbers of potential nonsynonymous and synonymous sites, respectively. A significant deviation from neutral expectation along a branch ($n/N \neq s/S$) is interpreted as evidence for selection. To increase the statistical power of the tests, we pooled coding-region synonymous sites with both intron and 3’ untranslated region (3’ UTR) sites (Rooney and Zhang 1999), after first determining that there were no differences in substitution rates between synonymous and intron sites or between synonymous and 3’ UTR sites (Fisher’s exact tests, both $P > 0.09$).

Amino Acid Variation and Structure Prediction for GP-9

To measure the degree of coding-region variation, we introduce two diversity indices, codon diversity and amino acid diversity. These indices are adapted from nucleotide diversity, typically denoted as π , a measure used to assess polymorphism at the DNA level. Nucleotide diversity is defined as the average proportion of nucleotide differences between all pairs of sequences in a sample (Hartl and Clark 1997). Here, codon (or amino acid) diversity denotes the diversity at specific in-frame coding-region triplets or amino acid residues, with the proportion of codon (or amino acid) differences between all sequence pairs presented separately for each position. The codon and amino acid diversities for a given position range from zero to one; a value of zero denotes a site with identical codons (or amino acids) in all sequences, whereas a value of one indicates that every sequence displays a unique codon (or amino acid) at this position. Comparison of codon diversity with amino acid diversity reveals how much of the observed nucleotide variation translates into amino acid replacements.

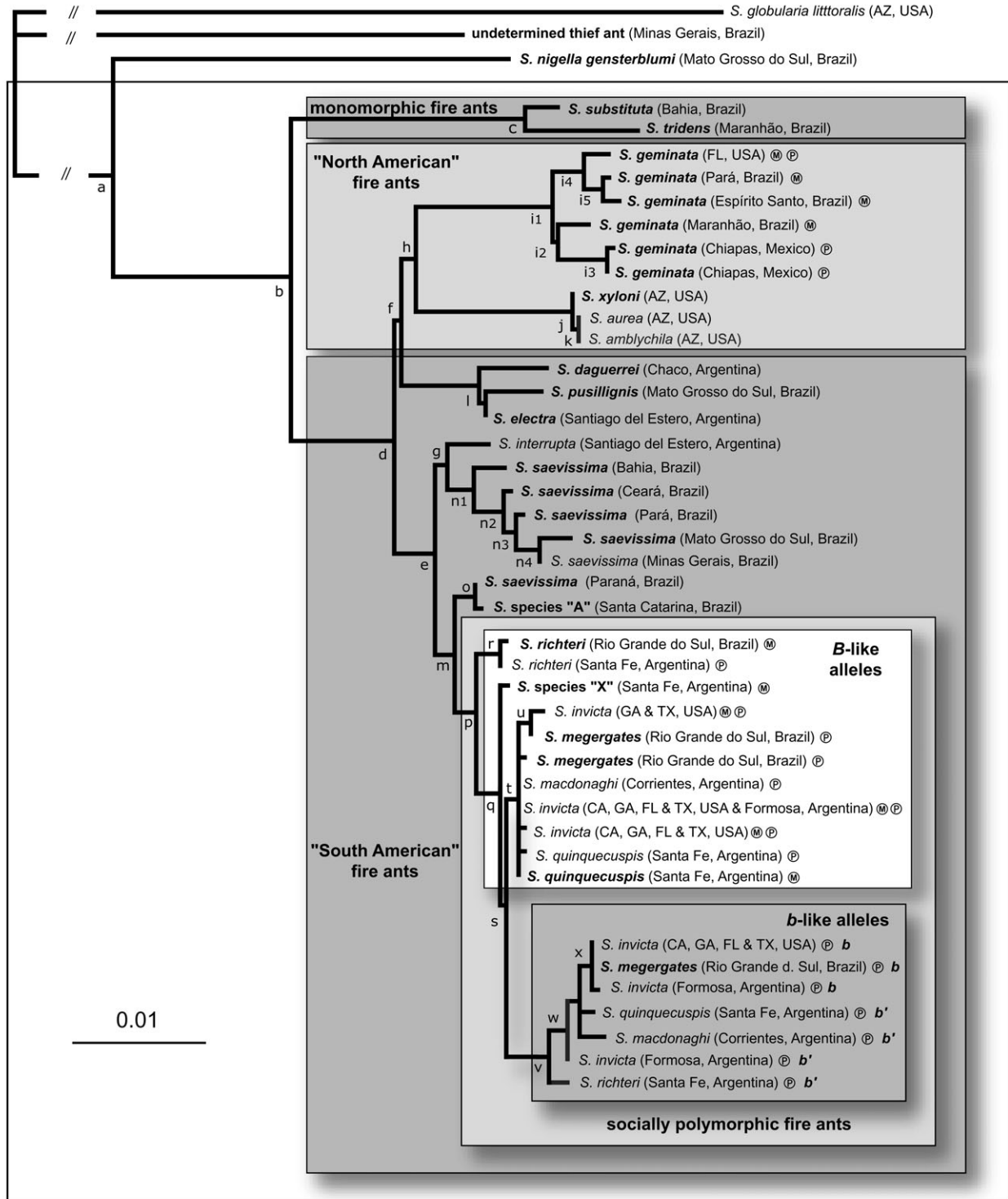
The potential functional importance of the amino acid variation observed among GP-9 proteins was investigated by mapping variable sites onto the predicted GP-9 protein structure obtained by GenTHREADER (McGuffin and Jones 2003) and the prediction software on the Robetta server (Chivian et al. 2003). We used the structure of a silkworm moth (*Bombyx mori*) PBP as a template for this analysis (Sandler et al. 2000; Protein Data Base code: 1DQE). Other insect PBPs/OBPs with structures that have been solved (from cockroach [Lartigue et al. 2003], fruit fly [Kruse et al. 2003], and honey bee [Lartigue et al. 2004]) also were evaluated initially, but these provided no information beyond that obtained from the moth template.

Results

Amplification of *Gp-9*

The *Gp-9* gene initially was amplified using primers located in the leader sequence and the 3’ flanking region. Amplification products approximately 2,200 bp in length were obtained from specimens of 11 of the 15 sampled

A



B

	a	b	c	d	e	f	g	h	i1	i2	i3	i4	i5	j	k	l	m	n1	n2	n3	n4	o	p	q	r	s	t	u	v	w	x
BI	100	100	100	100	100	75	100	75	100	77	100	100	100	100	95	100	100	100	100	94	100	100	92	100	100	57	100	100	100	100	100
ML	100	100	100	100	100	65	72	65	100	53	100	90	83	100	68	100	100	98	90	52	73	86	100	54	83	66	83	80	95	80	84
MP	100	100	100	100	96	67	66	69	100	55	100	87	85	100	65	99	77	95	90	-	71	90	-	-	89	-	91	70	89	65	85
ME	100	100	100	73	52	-	-	-	100	77	100	78	91	100	75	99	71	89	-	-	59	99	-	-	91	-	93	62	96	-	86

species by using these primers. The amplified fragments contained the complete gene (including the signal sequence, five exons, and four introns) as well as an additional 500 bp from the 3' flanking region (e.g., Krieger and Ross 2002). In the remaining four species, *S. nigella gensterblumi*, *S. substituta*, *S. tridens*, and an undetermined thief ant, we were unable to obtain amplification products using these primers. However, reamplification with a different reverse primer designed to anneal immediately downstream of the stop codon resulted in a single amplification product of 2,002–2,646 bp in each of these four species. This product consisted of the complete *Gp-9* gene; the five exons are identical in length to those observed in the other species and the four introns are, with one exception, comparable in length to those in the other species (the second intron is considerably longer in these four species than in the others: 786–1,414 vs. 470 bp). The apparent absence of a complementary binding site in the 3' flanking region of these four species, as well as in *S. globularia littoralis* (Krieger and Ross 2002), is reasonably interpreted as a plesiomorphy of *Solenopsis*, given the placement of these five *Gp-9* sequences at basal nodes within the phylogeny (see below).

Intragenic Recombination

The two tests employed to detect mosaic *Gp-9* sequences (DSS and PDM) failed to uncover evidence that historical recombination has generated such sequences. Our analyses of recombination within extant species using two population genetics methods similarly failed to detect evidence for significant recombination. The four-gamete test estimated the bound on the minimum number of recombination events at zero for all four species examined ($R_m = R_h = R_s = 0$), and the ML method estimated the recombination rate (r) in *S. invicta* at only 2.6×10^{-6} (95% consistency index: 2.6×10^{-11} –0.1). Because recombination evidently has not been a significant evolutionary force molding variation at *Gp-9*, each unique sequence can be presumed to have had a singular evolutionary history that potentially can be recovered using standard phylogenetic analyses (e.g., Posada and Crandall 2002).

Phylogeny of *Gp-9* Alleles

Results of the phylogenetic analyses of *Gp-9* sequences are summarized in figure 1. The phylogenetic trees from the ML and BI analyses are in perfect agreement, but with a generally higher branch support from BI than ML (fig. 1B). The trees obtained from MP and ME analyses are also

largely congruent with the ML/BI tree, as indicated by their high bootstrap support values for most nodes (fig. 1B).

The reconstructed *Gp-9* sequence relationships are in broad agreement with the proposed classification of the sampled species as derived from morphological character data (Ettershank 1966; Trager 1991; Pitts 2002). Sequences from *S. globularia littoralis* and the undetermined thief ant consistently are placed basally in the phylogeny, regardless of the specific choice of out-group. Among the remaining sequences, that of *S. nigella gensterblumi*, an unusual “fire ant” species with an atypical, morphologically dimorphic worker caste, is sister to a strongly supported clade comprising all sequences from species classified by Trager (1991) in the *S. geminata* species group (the “true” fire ants and their social parasites) (node a in fig. 1A). Two well-supported clades occur within the *S. geminata* group sequences (node b), one representing the morphologically monomorphic fire ant species (which form colonies with workers of essentially one size) and the other representing the morphologically polymorphic fire ants (which form colonies with workers with extreme size variation) together with their social parasite *S. daguerrei* (which lacks a worker caste).

The sequences from fire ant species with worker size variation fall into one clade corresponding to species whose ranges include or are confined to North America (the “North American” fire ants of Trager’s [1991] *S. geminata* species complex; node h) and a second, paraphyletic group of sequences from species whose ranges are exclusively South American (Trager’s [1991] *S. saevissima* species complex). This paraphyly stems from the clustering of a clade including the South American species *S. daguerrei*, *S. pusillignis*, and *S. electra* with the North American clade, although we note that support for this cluster (node f) is relatively weak. Within the North American clade are two strongly supported groups, one of which (node i1) contains all the *S. geminata* sequences from widely separated collection localities and the other (node j) comprising *S. amblychila*, *S. aurea*, and *S. xyloni* (Trager 1991). Interestingly, the two *S. geminata* sequences obtained from polygyne colonies in Mexico share a single amino acid replacement in the signal peptide that is not found in the sequences of any monogyne conspecifics or, indeed, in the sequences of any other *Solenopsis* specimens.

The remaining sequences from the South American fire ants group into two clades, one of which contains sequences from *S. interrupta* and all but one *S. saevissima* specimen (node g). The other, oddly divergent, *S. saevissima* sequence is sister to the *Solenopsis* species “A”

←

FIG. 1.—Evolutionary relationships of 43 unique *Gp-9* sequences from 21 *Solenopsis* species. (A) Tree constructed using ML analysis or BI, based on complete nucleotide sequences. Names of the species from which sequences were obtained are followed by names of localities (state/province, country) where the specimens were collected; new sequences generated for this study have the species of origin highlighted in bold. Letters in circles refer to the inferred social organization of the colonies of origin for specimens of the socially polymorphic South American fire ant species and *Solenopsis geminata* (M = monogyne, P = polygyne) (only coding-region sequence was obtained for the polygyne *S. geminata* from the United States). These are followed in the case of the *b*-like alleles by identification of the specific allele type. Codes shown on branches refer to bootstrap and posterior probability support values listed in figure 1B. The outermost box encompasses the sequences from species placed by Trager (1991) in the *S. geminata* species group (the “true” fire ants and their social parasites). CA = California, GA = Georgia, FL = Florida, TX = Texas. (B) Support values (>50%) for nodes depicted in figure 1A. BI = posterior probability values from BI, ML = bootstrap values from ML analysis, MP = bootstrap values from MP analysis, ME = bootstrap values from ME analysis.

sequence (node o), and these two in turn form the sister group to a substantial clade (node p) representing fire ant species assumed on morphological grounds to be the closest relatives of *S. invicta* (Pitts, McHugh, and Ross 2005). *Solenopsis invicta* and these closest relatives are notable for the fact that they are the only South American *Solenopsis* known to display intraspecific polymorphism in social organization, that is, the presence of both monogyne and polygyne colonies.

The *Gp-9* sequences of the socially polymorphic South American fire ants exhibit a remarkable evolutionary pattern in which the gene phylogeny departs radically from any conceivable species phylogeny, with essentially all species polyphyletic for alleles of the *B*-like and *b*-like classes. Specifically, the *B*-like alleles from these species, recognizable by two diagnostic amino acid residues at positions 42 and 139, constitute a basal paraphyletic assemblage in this part of the tree, while the *b*-like alleles, distinguishable by their set of alternate diagnostic amino acids at these same positions, form an exclusive apical clade of more recently derived *Gp-9* sequences.

We obtained new *Gp-9* sequences for this study from two previously uninvestigated species from the socially polymorphic group, *S. megergates* and *Solenopsis* species "X". Two specimens from a suspected polygyne *S. megergates* colony yielded both *B*-like and *b*-like alleles, as expected. The *Solenopsis* species "X" samples originated from presumed monogyne nests, and both specimens possessed only a *B*-like allele variant (confirmed by our diagnostic PCR assay). Surprisingly, this *B*-like variant encodes an isoleucine residue at position 95, which is characteristic of all *b*-like alleles, rather than a methionine residue, which is characteristic of the remaining *B*-like alleles. Assuming that the *Solenopsis* species "X" nests were in fact monogyne, this discovery makes it doubtful that the amino acid at position 95 is essential to the function of GP-9 with respect to expression of colony social organization in this group of species (as previously suggested by Krieger and Ross 2002; Ross, Krieger, and Shoemaker 2003; Krieger 2005).

We also obtained new sequences from two socially polymorphic species represented in our previous sequencing study (*S. quinquecupis* and *S. richteri*), but the new specimens originated from monogyne colonies rather than from polygyne colonies as in the previous study (Krieger and Ross 2002). Only *B*-like allelic variants were represented among these new sequences from monogyne colonies.

In sum, we now have sequence data for all the described socially polymorphic South American fire ants (multiple sequences for several species), and these data strongly support our earlier conclusion (Krieger and Ross 2002) that the expression of polygyny in species of this clade requires the presence in a colony of individuals bearing *b*-like alleles. Moreover, the new sequence data also support the hypothesis that *b*-like alleles do not occur outside the socially polymorphic clade, a pattern further implicating these alleles in the induction of polygyny. The inferred absence of *b*-like alleles outside the socially polymorphic group was further substantiated by the application of our diagnostic PCR assay to 106 colonies of five species

falling outside the clade; in none of these samples were *b*-like alleles detected.

Selection on *Gp-9*

Elimination of redundant sequences with identical coding-region nucleotide compositions resulted in 28 unique sequences available for the selection analyses, each with 153 codons. Estimation of nonsynonymous and synonymous substitution rates over the five exons was first performed by ML (CODEML program; Yang 1999). Many branches showed a numerical excess of nonsynonymous over synonymous substitutions (fig. 2). However, the estimate of ω (dN/dS) over all lineages (table 1; model M0) does not differ significantly from the value of one expected under neutral evolution, as determined by an LRT (table 2; model M0 vs. model M0').

To investigate whether selective pressures have differed among individual codons during the evolution of *Gp-9* in *Solenopsis*, we applied site-specific ML models that allow selection to vary among codon sites but impose identical site-specific patterns across all lineages. All three models that allow the presence of positively selected sites ($\omega > 1$) fit the data significantly better than corresponding models that do not (table 2; M1 vs. M2, M0 vs. M3, and M7 vs. M8). The dN/dS ratios at positions implied by these models to be under positive selection fall in the range 5.61–8.57. Using the empirical Bayesian approach implemented in CODEML, models M2 and M8 predict with >95% posterior probability that six positions (codons 45, 48, 117, 120, 134, and 145) have experienced positive selection (fig. 3), while model M3 predicts an additional nine such positions. (We consider below only the results of models M2 and M8 as model M3 may be prone to overestimate the number of positively selected sites [Yang et al. 2000; Anisimova, Bielawski, and Yang 2001].)

In an attempt to validate the findings of the site-specific ML analyses, we employed the parsimony-based method implemented in the ADAPTSITE program (Suzuki and Gojobori 1999; Suzuki 2004). This method failed to detect evidence of a single codon position under positive selection at the 5% significance level.

In addition to the site-specific models, we applied branch-specific models that allowed us to test whether dN/dS in a lineage of interest is different from the background ratio. In particular, we wished to learn whether there is evidence that the *b*-like alleles, which appear to be integral to the presumably derived polygyne social system of the socially polymorphic species, have been under different selective regimes than other *Gp-9* alleles. First, we employed the branch-specific model (Mb) implemented in CODEML, with all branches of the *b*-like clade considered as one group and all other branches in the phylogeny considered as a second group. The combined estimate of ω for the *b*-like branches (ω_b) was 3.80, whereas the background estimate was 0.94. Although this elevated dN/dS ratio suggests an apparent excess of replacement substitutions within the *b*-like clade, the model employing the two different ω values does not fit the data significantly better than the null model with a single ω value for all branches (table 2; M0 vs. Mb). To test whether the empirical estimate of

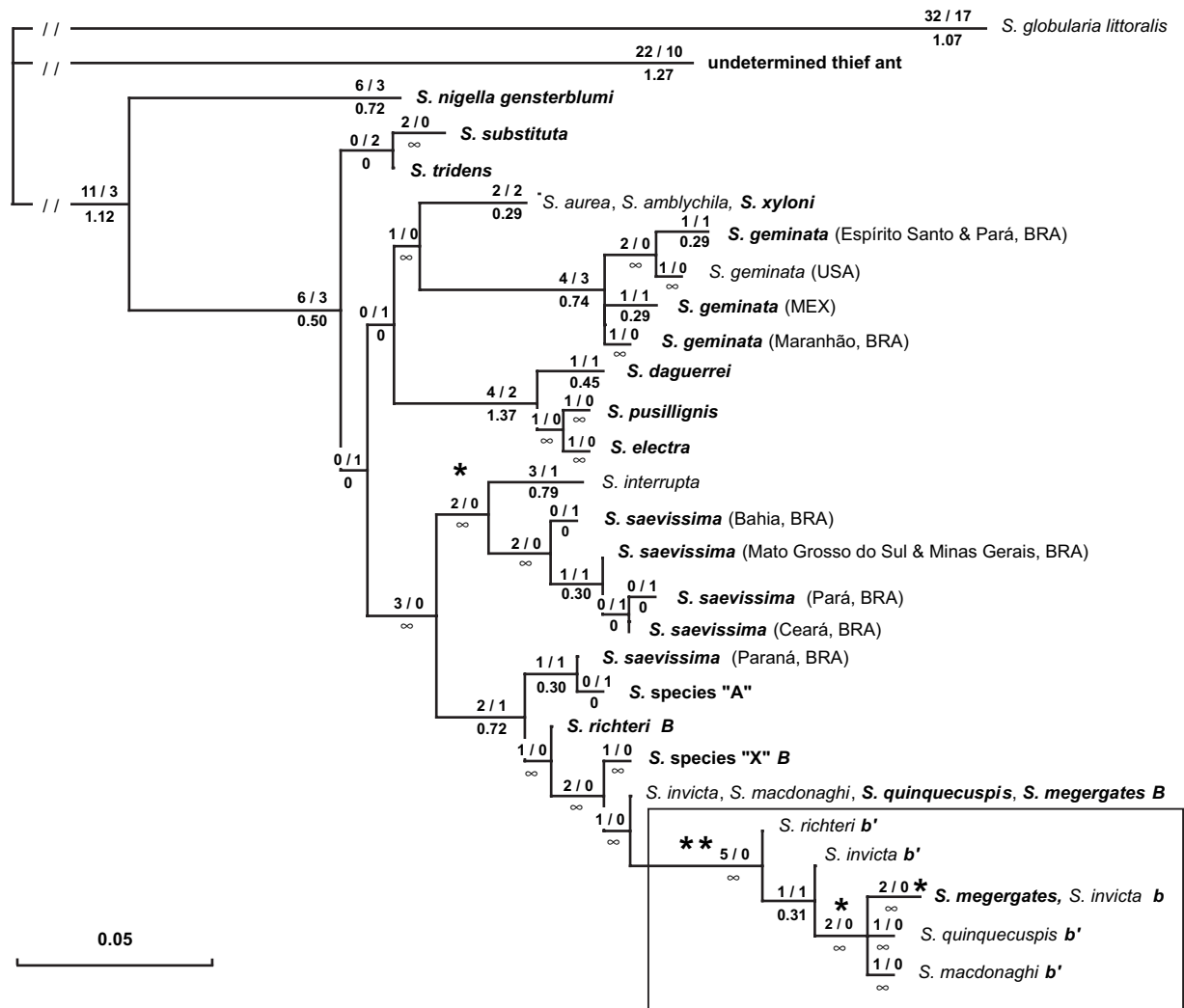


FIG. 2.—Inferred numbers of substitutions and dN/dS ratios (ω) shown on phylogeny of 28 unique *Gp-9* coding-region sequences from 21 *Solenopsis* species. Names of the species from which sequences were obtained are shown, with the new sequences generated for this study highlighted in bold. Branch lengths correspond to the expected numbers of nucleotide substitutions per codon and were computed as ML estimates under the assumption of independent ω for each branch (Yang 1999). The numbers above each branch represent nonsynonymous and synonymous substitutions, respectively, deduced from the ancestral sequences; the number below each branch represents ω . Branches for which episodes of significant positive selection were detected (using Fisher's exact test) are indicated by asterisks (* $P = 0.025$, ** $P = 0.002$). *B*-like and *b*-like sequences of the socially polymorphic South American species are labeled, and a box encloses the clade of polygyny-inducing *b*-like alleles. MEX = Mexico, BRA = Brazil.

ω_b is significantly greater than one, the log likelihood value was recalculated under the same branch-specific model but with ω_b constrained to equal one (Mb'), and an LRT was used to compare the two models. Again, this test showed no significant difference in the fit of the two models to the data (table 2; Mb vs. Mb'). Thus, there is no statistical justification on the basis of these analyses for concluding that the empirical estimate of ω_b exceeds unity.

To test each branch individually for evidence of selection, we used Fisher's exact tests to determine whether the number of nonsynonymous and synonymous substitutions along particular branches is in accordance with the null hypothesis of neutral evolution (Zhang, Kumar, and Nei 1997). To increase the power of the tests, we pooled coding-region synonymous sites with both intron and 3' UTR sites (Rooney and Zhang 1999). The numbers of potential

nonsynonymous and synonymous coding-region substitutions per sequence were estimated as 107 and 352, respectively; the potential substitutions in the synonymous category increased to 1,740 after inclusion of the intron and 3' UTR sites. Positive selection, signified by an excess of nonsynonymous substitutions, is statistically detectable in four branches of the *Gp-9* phylogeny, with all but one of these occurring in the *b*-like allele clade (fig. 2). Significance levels were $P = 0.025$ for all branches except the one at the base of the *b*-like clade, where $P = 0.002$. Considering that a single additional synonymous substitution in the branches at the $P = 0.025$ level would render them nonsignificant ($P = 0.17$), the evidence for positive selection acting on these lineages should be accepted with some caution. On the other hand, the significant result for the stem lineage of the *b*-like clade is very robust, as six additional

Table 1
Summary of Parameter Estimates and Likelihood Values Under Different Models of Selection During the Evolution of *Gp-9* in *Solenopsis*

Model	Estimates of Parameters	Log Likelihood
M0—"one ratio"	$\omega = 1.01$	-1,664.51
M0'—"one ratio"	$\omega^a = 1$	-1,664.51
M1—"neutral"	$\omega_0 = 0, \omega_1^a = 1,$ $p_0 = 0.33, p_1 = 0.67$	-1,650.18
M2—"selection"	$\omega_0 = 0, \omega_1^a = 1,$ $\omega_2 = \mathbf{7.36}, p_0 = 0.26,$ $p_1 = 0.67$	-1,631.51
M3—"discrete"	$\omega_1 = 0.32, \omega_2 = \mathbf{1.83},$ $\omega_2 = \mathbf{9.68}, p_0 = 0.61,$ $p_1 = 0.34$	-1,630.27
M7—"beta"	$p = 0.03, q = 0.02$	-1,649.72
M8—"beta + ω "	$\omega_2 = \mathbf{7.23}, p = 0.04,$ $q = 0.02, p_0 = 0.93,$ $p_1 = 0.07$	-1,631.28
Mb—"branch <i>b</i> "	$\omega_0 = 0.94, \omega_1 = \mathbf{3.80}$	-1,663.29
Mb'—"branch <i>b</i> "	$\omega_0 = 0.94, \omega_1^a = 1$	-1,664.46

NOTE.—Values of ω (dN/dS) in bold constitute evidence for positive selection. Variables p and q are parameters of the beta distribution (M7, M8). p_0 and p_1 denote the relative frequency of each ω category (M1, M2, M3, M8).

^a ω fixed at 1.

synonymous substitutions would be required to raise the significance level above 5%. Thus, this analysis supports the hypothesis that positive selection acted during the early evolution of the *b*-like allele lineage, presumably in association with the origin of polygyny (Krieger and Ross 2002).

Amino Acid Variation and Structural Considerations

A large number of codons are variable across the different *Gp-9* sequences of the *Solenopsis* specimens that we studied. Of the 153 triplets that make up the entire coding region, more than half (57.5%) are variable at the codon level, and 85.2% of these translate into variable amino acid residues. A great deal of this variation is attributable to the three basal species in our study, *S. globularia littoralis*, the thief ant, and *S. nigella gensterblumi*. To meaningfully characterize codon variability and link this variability to structural properties of the protein, we restrict further analyses to the clade of "true" fire ant species that excludes these three species. Furthermore, we generally include only one exemplar sequence from each species in these analyses; exceptions include *S. saevissima*, for which a sequence from each of the two divergent clades is included, and the socially polymorphic species, for which we selected exemplars of both *B*-like and *b*-like alleles from each species in which both were available.

Within the reduced set of 19 unique *Gp-9* sequences, 24.2% of the coding-region triplets show variation at the codon level, of which 78.4% are variable also at the amino acid level. Similar results are obtained when the signal peptide is excluded from consideration: the mature protein is 134 codons in length, with 25.4% of the codons variable across sequences and 79.4% of these variable codons generating variation also at the amino acid level. The locations of the variable codons in mature GP-9 protein from fire ants are depicted in figure 3 by means of two indices of the degree of variation, the codon diversity and amino acid diver-

Table 2
Likelihood Ratio Tests for Comparison of Different Models of Selection During the Evolution of *Gp-9* in *Solenopsis*

Models Compared	2 Δ lnL	Df	<i>P</i>
M0 versus M0'	0.08	1	0.95
M1 versus M2	30.94	2	<0.001
M0 versus M3	59.22	2	<0.001
M7 versus M8	30.41	2	<0.001
Branch-specific models			
M0 versus Mb	2.24	1	0.12
Mb versus Mb'	2.23	1	0.13

sity. Identical codon and amino acid diversities indicate that each unique codon at a given position codes for a unique amino acid, whereas higher codon diversity than amino acid diversity indicates that multiple codons code for the same amino acid. When comparing the two indices at all variable positions across mature GP-9 proteins, 85.7% of the positions have identical codon and amino acid diversities. Thus, the picture to emerge from these sets of analyses is that *Gp-9* has undergone considerable coding-region divergence during the radiation of fire ant species and that most of this divergence is apparent at the amino acid level (i.e., most coding-region base substitutions are replacement substitutions).

We next investigated the positions of the variable amino acid residues with respect to the inferred protein structure, with the goal of learning which substitutions occur at positions likely to be involved in binding odorant molecules (or other ligands). We used computerized sequence alignment algorithms (McGuffin and Jones 2003) in combination with threading techniques (Chivian et al. 2003) to approximate the structure of GP-9 protein, using as a template a silkworm moth PBP (Sandler et al. 2000). Of the 19 amino acid positions inferred to be involved in ligand binding, 4 have residues that vary among fire ant GP-9 proteins (positions 75, 119, 136, and 139; fig. 3). All nucleotide substitutions that have occurred at these four positions are nonsynonymous substitutions. Extending consideration to residues immediately adjoining these presumed binding residues (to accommodate uncertainties in the threading technique), the number of variable amino acid residues matching the implied ligand-binding sites increases to six (positions 120 and 134 are added). All nucleotide substitutions at the two additional positions also are nonsynonymous. More importantly, these additional positions are inferred by the ML models to have experienced positive selection (see above). The number of distinct amino acids occurring at each of the six variable positions that may be involved in ligand binding is either two or three across all the fire ant species that we surveyed.

Finally, we mapped the locations of the two amino acid residues that are unique to all polygyny-inducing *b*-like alleles to the inferred protein structure. All *b*-like alleles share the amino acids Gly⁴² and Ile¹³⁹, of which Ile¹³⁹ was determined to be at a ligand-binding position (fig. 3). Gly⁴² appears to be located on a solvent-exposed loop-like structure that is not directly involved in ligand binding.

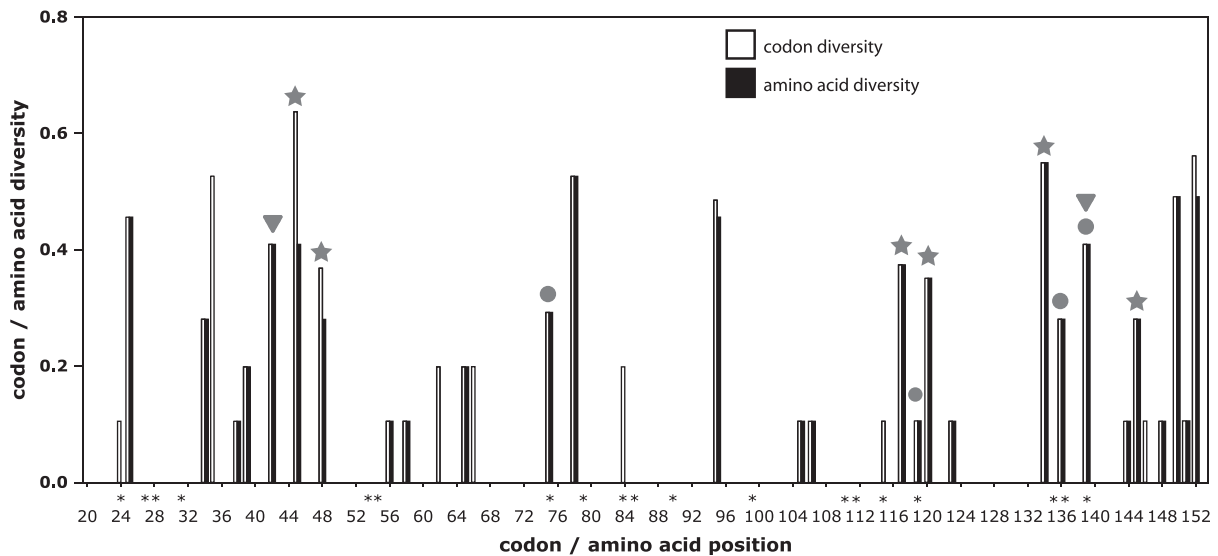


FIG. 3.—Codon and amino acid variation among mature proteins encoded by 19 unique *Gp-9* coding-region sequences from fire ants. Codon and amino acid diversity, calculated as the proportions of codon (or amino acid) differences between all pairs of sequences, are given for each codon position. Asterisks indicate positions of amino acids implied by comparative structural analysis to be involved in ligand binding. Gray circles represent locations of variable amino acids that map to positions involved in ligand binding, gray triangles represent locations of the two amino acids uniquely shared by all *b*-like alleles, and gray stars represent the codons determined to be under positive selection using the site-specific ML models (>95% posterior probability).

Discussion

Forty-three unique nucleotide sequences of *Gp-9*, a candidate gene believed to regulate important features of fire ant social organization, were analyzed in this study in order to infer the major features of its molecular evolution in the genus *Solenopsis*. The exon/intron structure and respective lengths of the five exons of *Gp-9* were found to be identical across all the 21 species examined. Moreover, tests for intragenic recombination revealed little or no evidence for its historical or recent occurrence at *Gp-9*. This latter result is consistent with previous speculation that *Gp-9* occurs in a genomic region with reduced recombination, which was based on a low measured frequency of recombination between *Gp-9* and the enzyme-encoding gene *Pgm-3* in *S. invicta* ($r = 0.0016$; Ross 1997). It is conceivable that both *Gp-9* and *Pgm-3* are contained in or flank an inversion (e.g., Keller and Ross 1999), which could account for the low effective intragenic and intergenic recombination frequencies (Griffiths et al. 2000). Regardless of its cause, reduced recombination suggests the potential profitability of a search for additional candidate genes in this region that could form with *Gp-9* an epistatic complex of genes that jointly governs the expression of social behavior in fire ants. Conservation of the exon/intron structure and exon lengths, together with a lack of recombination, indicates that variation in the coding region of *Gp-9* in *Solenopsis* has evolved primarily or solely by means of point substitutions.

The *Gp-9* sequence phylogeny depicted in figure 1 mirrors in its essential features the relationships reported earlier for a much smaller sample of *Solenopsis* sequences (Krieger and Ross 2002). In addition, the inferred phylogenetic relationships of *Gp-9* sequences reflect in many respects the presumed evolutionary history of the *Solenopsis* species from which they are derived, judging from the

existing classification (Ettershank 1966; Trager 1991; Pitts 2002) and from a cladistic analysis of the South American fire ants (Pitts, McHugh, and Ross 2005) (both based on morphological characters). Thus, for instance, sequences from *S. globularia littoralis*, *S. nigella gensterblumi*, and a thief ant are basal or sister to a clade comprising sequences from the “true” fire ants (*S. geminata* species group of Trager [1991]), sequences from the fire ants with a morphologically monomorphic worker caste are monophyletic (Trager 1991; Pitts 2002), and sequences from the North American fire ants (*S. geminata* species complex of Trager [1991]) are monophyletic, with *S. xyloni*, *S. aurea*, and *S. amblychila* sequences forming a sister clade to all *S. geminata* sequences (as implied by Trager [1991]). Most significant with respect to *Gp-9* and fire ant social evolution, the socially polymorphic South American species are inferred to be monophyletic on the basis of their *Gp-9* sequences as well as their morphology (Pitts, McHugh, and Ross 2005).

There are also points of disagreement between the phylogenetic hypotheses for fire ants derived from *Gp-9* sequences and from morphological characters, two of which bear particular mention. One is that worker size monomorphism is inferred from the morphological phylogeny to represent a secondary reversion in the common ancestor of *S. substituta* and *S. tridens* (Pitts, McHugh, and Ross 2005), whereas this characteristic of these species is inferred from the *Gp-9* phylogeny to represent the ancestral condition for the entire genus. A second important difference is the apparent paraphyly of *Gp-9* sequences from the South American fire ants with worker size polymorphism, which contradicts the inferred monophyly of the group based on morphological characters (Trager 1991; Pitts, McHugh, and Ross 2005). However, the sequence paraphyly stems from a single weakly supported node.

Within the socially polymorphic South American fire ant species, *Gp-9* sequences often exhibit insufficient phylogenetically informative variation to recover relationships with confidence. However, several biologically important patterns are well supported. The *B* alleles of *S. richteri* occur at the base of the clade of all *Gp-9* alleles found in this group of species, and the *b'* allele of *S. richteri* is basal within the *b*-like allele clade. These parallel patterns suggest that *S. richteri* is the earliest originating species of this group, as suggested also by some of the minimum-length MP trees of Pitts, McHugh, and Ross (2005). Second, both the *B*-like and *b*-like sequences of *S. invicta* appear polyphyletic, patterns consistent with mounting evidence from mitochondrial DNA sequences and allozyme markers that this nominal species may comprise multiple evolutionarily independent entities (Ross and Shoemaker 2005). Third, the *b* alleles of *S. invicta* and *S. megergates*, which differ from all other *b*-like alleles (those designated *b'*) by virtue of a charge-changing amino acid replacement at position 151, form a recently derived monophyletic group within the *b*-like clade. Finally, our analyses confirm that all *b*-like alleles form a derived clade occurring within the group of *Gp-9* sequences from the socially polymorphic South American species, lending further support to the hypothesis that monogyny preceded polygyny in these fire ant species (Krieger and Ross 2002).

Previous studies based on hundreds of samples of the important pest species *S. invicta* have shown that queens of the polygyne social form always bear a *b*-like allele at *Gp-9*, whereas all ants from monogyne colonies bear only *B*-like alleles (Ross 1997; Krieger and Ross 2002; Ross and Keller 2002). We also established earlier that polygyne queens of several South American fire ant species closely related to *S. invicta* always bear a *b*-like allele (Krieger and Ross 2002). These findings led us to suggest that *b*-like alleles are required for the expression of polygyny in the entire clade of socially polymorphic South American fire ants. In the present study, we included specimens of *S. megergates*, the only described species of this clade whose *Gp-9* is yet to be characterized. Two queens from a suspected polygyne nest carried a *b*-like allele, as expected if *b*-like alleles are required for the expression of polygyny in this species. We also PCR assayed and sequenced new specimens from two species included in our previous study (*S. quinquecuspis* and *S. richteri*), with these new samples originating from monogyne nests rather than polygyne nests as in the earlier study. These new samples harbored only the *B* allelic variant. The only species in the socially polymorphic clade in which a *b*-like allele has not been found is the undescribed *Solenopsis* species "X", but the five nests of this species that we studied were likely to have been monogyne. We predict that *b*-like alleles will be recovered when the *Gp-9* of queens from a confirmed polygyne nest of this species is sequenced.

South American fire ant species falling outside the socially polymorphic clade are not known to exhibit polygyne colony social organization. Accordingly, we were unable to detect *b*-like alleles in any of the large number of colonies surveyed from many such species (by means of sequencing and an allele-specific PCR assay). This result, taken together with the discovery of *b*-like alleles in all suspected

polygyne nests of the socially polymorphic species, strengthens the link between the presence of *b*-like *Gp-9* alleles and the expression of polygyny across the South American fire ant species.

The availability of a detailed gene phylogeny for *Gp-9* has made it possible to thoroughly investigate the role of selection during the evolutionary history of this gene in *Solenopsis*. Many branches in the phylogeny show a numerical excess of replacement over synonymous substitutions; however, the *dN/dS* ratio (ω) estimated over the entire assemblage of lineages is not significantly different from the neutral expectation of one. To learn if selection has acted differently at various amino acid positions, we applied site-specific models that allow selection pressures to vary among codons. All three ML models allowing the presence of some positively selected amino acids fitted the sequence data significantly better than corresponding models that do not, implying a history of heterogeneous selection pressures along the *Gp-9* gene, and ML analyses identified a handful of specific positions that apparently have been under positive selection. None of these positively selected codons map to positions inferred to be involved in formation of the binding pocket of OBPs (the family of proteins to which GP-9 belongs); however, two of them (at positions 120 and 134) adjoin binding-pocket residues. Moreover, amino acids at four additional positions that do correspond to presumed binding-pocket locations are variable among the GP-9 proteins of fire ants (see below). These findings raise the possibility that selection has driven changes in the amino acid composition at positions influencing binding-pocket formation in GP-9, with a likely result being the alteration of the ligand-binding properties of the protein during the course of fire ant evolution.

We were unable to validate with a parsimony-based method the presence of any of the positively selected GP-9 residues inferred by ML analyses. The parsimony method tends to be conservative under some circumstances, making it difficult to reject neutrality. One such circumstance is the presence of long branches in the phylogenetic tree (Suzuki and Nei 2002), a factor unlikely to be important in our study because long branches were confined to just a few basal lineages. A second such circumstance is of more concern. The power of parsimony-based methods is reported to be low unless a large number of sequences is used (Suzuki and Nei 2002, 2004), and the 28 sequences included in our analysis may not have been sufficient to detect modest levels of positive selection. On the other hand, the ML methods tend to err toward false positives under the same conditions that cause the parsimony method to be conservative (Suzuki and Nei 2002). This probably occurs because the posterior probabilities inferred using the ML methods are computed under the assumption that the estimates of *dN/dS* are the true values, whereas they are subject to sampling error (Suzuki and Nei 2002).

We also applied branch-specific selection models that allowed us to test whether the *dN/dS* ratio for a lineage of interest differs from the background ratio. Given that the alleles in the *b*-like clade invariably are linked with the expression of polygyny, it is reasonable to hypothesize that positive selection was involved in honing GP-9 protein to bring about the changes in chemoreceptive (or other)

abilities of ants necessary for this alternate social system to function properly. When we applied the branch-specific ML model, the combined dN/dS ratio estimated for branches in the *b*-like clade showed an excess of replacement substitutions ($\omega = 3.80$), whereas the background ratio was close to neutral expectation ($\omega = 0.94$). Despite this stark difference, a model with two distinct dN/dS ratios did not fit the data significantly better than a null model with a single ratio for all branches in the *Gp-9* tree. The reason for this result is not a lack of replacement substitutions within the *b*-like clade but rather the occurrence of a fair number of replacement substitutions in other parts of the tree. A different approach that avoids the insensitivity caused by pooling branches in the ML methods is based on comparison of the numbers of realized nonsynonymous and synonymous substitutions with the numbers of potential nonsynonymous and synonymous sites along each branch. Applying this method, a statistically significant excess of nonsynonymous substitutions was detected in four branches, three of which are in the *b*-like allele clade. Among these four detectable signals of positive selection, only the result for the stem lineage of the *b*-like clade is highly significant. Evidently, positive selection was involved in driving the amino acid substitutions that were crucial to the origin of polygyny in the South American fire ants.

The use of protein alignment algorithms in combination with threading techniques allowed us to examine the positions of the variable amino acids in GP-9 proteins in relation to the inferred protein structure. Of the 28 positions with variable residues in the mature protein of the surveyed fire ant species, 4 map to 1 of the 19 positions determined to be involved in ligand binding in the silkworm moth PBP (Sandler et al. 2000). Thus, changes in any of these four residues potentially could have induced changes in the binding properties and function of GP-9 protein during the radiation of fire ant species. One of these residues, Ile¹³⁹, is one of the two amino acids uniquely shared by all *b*-like alleles and is especially interesting because it is predicted to extend its side chain into the binding pocket. These findings constitute compelling evidence that the substitution of isoleucine for valine at position 139 in the stem lineage of the *b*-like allele clade altered the ligand-binding properties of GP-9, thereby inducing or facilitating the evolutionary switch from polygyny to monogyny colony social organization.

Allelic induction of polygyny by *b*-like *Gp-9* variants now has been established in all of the described South American fire ant species known to exhibit polymorphism in colony social organization. Outside of this clade of socially polymorphic species, polygyny in fire ants is well documented only in *S. geminata* (Adams, Banks, and Plumley 1976; Mackay et al. 1990), a member of the North American clade of fire ants. We showed recently that *Gp-9* sequences from a polygyny population of this species in Florida do not have the set of amino acid replacements characteristic of *b*-like alleles (Ross, Krieger, and Shoemaker 2003). Thus, these replacements cannot universally underlie the expression of polygyny in fire ants. (We speculated that polygyny in the *S. geminata* population from Florida resulted instead from a bottleneck-induced reduction of variation at genes encoding queen pheromones,

which presumably diminishes workers' abilities to recognize individual queens and to regulate their numbers in a colony.) In the present study, we included additional specimens of polygyny *S. geminata* from a population in Chiapas, Mexico. Again, *Gp-9* sequences from this population lacked the *b*-like set of amino acid replacements. However, in this case, sequences of both polygyny specimens featured a single amino acid replacement in the signal peptide, a substitution not found in any other *Solenopsis* sequences. It is conceivable that this substitution, located in the portion of the protein that initiates its translocation through cellular membranes, hinders export of GP-9 to its appropriate target. If GP-9 mediates worker recognition and discrimination of individual queens, a reduced amount of the protein in the lymph of chemosensory sensilla could interfere with this process and perhaps lead to worker acceptance of supernumerary queens (a defining feature of polygyny). Our research on the molecular mechanisms of social evolution in fire ants thus has now revealed three possible routes to polygyny, each involving distinct changes in the molecular components of the chemoreception systems involved in regulation of colony queen number.

Insect OBPs are characterized by several criteria, including their unique expression in the antennae, their ability to bind odorant molecules, and the presence of six absolutely conserved cysteine residues located in characteristic positions (Vogt 2003). GP-9 protein has the greatest sequence similarity to moth PBPs (a subclass of OBPs) and clearly belongs structurally to the OBP gene family, as indicated by the six conserved cysteines present in all sequences. On the other hand, the presence of GP-9 protein in the thorax of adult fire ants (Krieger and Ross 2002) (K. G. Ross and M. J. B. Friege, unpublished data) appears to be at variance with the first criterion for classification as an OBP (specific sites of expression of GP-9 currently are unknown). However, as two recent studies show, expression of many OBPs is not restricted to the antennae. A survey of the genome of the mosquito *Anopheles gambiae* identified 29 putative OBP-encoding genes (Vogt 2002), yet an extensive screening of an antennal cDNA library isolated only eight expressed transcripts (Justice et al. 2003). Similarly, the *Drosophila* genome contains some 50 OBP-encoding genes (Hekmat-Scafe et al. 2002), of which only 9 are expressed exclusively in sensilla of the antennae (Galindo and Smith 2001). Four *Drosophila* OBP genes were found to be expressed exclusively in gustatory sensilla of the mouthparts, legs, and wings, and five showed expression in broad regions that include areas devoid of any chemosensory organs (Galindo and Smith 2001). These diverse expression patterns may suggest that ancestral insect OBPs originally served as carrier proteins that bound and solubilized hydrophobic molecules in a variety of tissues and roles and that some of these proteins only secondarily evolved the more specific role of transporting odorant molecules in chemosensilla. Alternatively, insect OBPs may originally have functioned as odorant-carrying proteins, with the members of this class not expressed in chemosensilla subsequently co-opting their effective binding properties to serve other transport functions. In either case, current characterization of OBPs as strictly antennal proteins seems too restrictive in light of the observed expression patterns.

Also, a more appropriate name for this protein family may be in order if roles other than the binding and transport of odorant molecules are confirmed for some members (see also Leal 2003).

Supplementary Material

Alignment is available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>). Previously unpublished *Gp-9* sequences are deposited in GenBank (accession numbers AY818614–AY818640).

Acknowledgments

We thank L. Cruz Lopez, L. Keller, M. Mescher, J. Pitts, and D. Shoemaker for assistance in obtaining specimens, S. Edwards, D. Gotzek, D. Shoemaker, and two anonymous reviewers for comments on the manuscript, and M. A. Moran for use of laboratory resources. This work was funded in part by the Georgia Agricultural Experiment Stations (University of Georgia).

Literature Cited

- Adams, C. T., W. A. Banks, and J. K. Plumley. 1976. Polygyny in the tropical fire ant, *Solenopsis geminata* with notes on the imported fire ant, *Solenopsis invicta*. *Fla. Entomol.* **59**:411–415.
- Anisimova, M., J. P. Bielawski, and Z. H. Yang. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**:1585–1592.
- Chivian, D., D. E. Kim, L. Malmstrom, P. Bradley, T. Robertson, P. Murphy, C. E. M. Strauss, R. Bonneau, C. A. Rohl, and D. Baker. 2003. Automated prediction of CASP-5 structures using the Robetta server. *Proteins* **53**:524–353.
- Ettershank, G. 1966. A generic revision of the world Myrmicinae related to *Solenopsis* and *Pheidologeton*. *Aust. J. Zool.* **14**:73–171.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- Fletcher, D. J. C., and M. S. Blum. 1983. Regulation of queen number by workers in colonies of social insects. *Science* **219**:312–314.
- Galindo, K., and D. P. Smith. 2001. A large family of divergent *Drosophila* odorant-binding proteins expressed in gustatory and olfactory sensilla. *Genetics* **159**:1059–1072.
- Greenberg, L., D. J. C. Fletcher, and S. B. Vinson. 1985. Differences in worker size and mound distribution in monogynous and polygynous colonies of the fire ant *Solenopsis invicta* Buren. *J. Kans. Entomol. Soc.* **58**:9–18.
- Griffiths, A. J. F., J. H. Miller, D. T. Suzuki, R. C. Lewontin, and W. M. Gelbart. 2000. An introduction to genetic analysis. 7th edition. W. H. Freeman, New York.
- Genetics Computer Group (GCG). 2001. Wisconsin Package. Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.
- Hartl, D. L., and A. G. Clark. 1997. Principles of population genetics. Sinauer Associates, Sunderland, Mass.
- Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**:160–174.
- Hekmat-Scafe, D. S., C. R. Scafe, A. J. McKinney, and M. A. Tanouye. 2002. Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster*. *Genome Res.* **12**:1357–1369.
- Hölldobler, B., and E. O. Wilson. 1977. The number of queens: an important trait in ant evolution. *Naturwissenschaften* **64**:8–15.
- Hudson, R. R., and N. L. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA-sequences. *Genetics* **111**:147–164.
- Huelsenbeck, J. P., and K. A. Crandall. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annu. Rev. Ecol. Syst.* **28**:437–466.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**:754–755.
- Husmeier, D., and F. Wright. 2001. Probabilistic divergence measures for detecting interspecies recombination. *Bioinformatics* **17**(Suppl. 1):S123–S131.
- Justice, R. W., S. Dimitratos, M. F. Walter, D. F. Woods, and H. Biessmann. 2003. Sexual dimorphic expression of putative antennal carrier protein genes in the malaria vector *Anopheles gambiae*. *Insect Mol. Biol.* **12**:581–594.
- Keller, L., and K. G. Ross. 1998. Selfish genes: a green beard in the red fire ant. *Nature* **394**:573–575.
- . 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.
- Krieger, M. J. B. 2005. To b or not to b: a pheromone-binding protein regulates colony social organization in fire ants. *Bioessays* **27**:91–99.
- Krieger, M. J. B., and K. G. Ross. 2002. Identification of a major gene regulating complex social behavior. *Science* **295**:328–332.
- Kruse, S. W., R. Zhao, D. P. Smith, and D. N. M. Jones. 2003. Structure of a specific alcohol-binding site defined by the odorant binding protein LUSH from *Drosophila melanogaster*. *Nat. Struct. Biol.* **10**:694–700.
- Kuhner, M. K., J. Yamato, P. Beerli, L. P. Smith, E. Rynes, E. Walkup, C. Li, J. Sloan, P. Colacurcio, and J. Felsenstein. 2004. LAMARC v 1.2.1. University of Washington.
- Lartigue, A., A. Gruez, L. Briand, F. Blon, V. Bezirard, M. Walsh, J. C. Pernollet, M. Tegoni, and C. Cambillau. 2004. Sulfur single-wavelength anomalous diffraction crystal structure of a pheromone-binding protein from the honeybee *Apis mellifera* L. *J. Biol. Chem.* **279**:4459–4464.
- Lartigue, A., A. Gruez, S. Spinelli, S. Riviere, R. Brossut, M. Tegoni, and C. Cambillau. 2003. The crystal structure of a cockroach pheromone-binding protein suggests a new ligand binding and release mechanism. *J. Biol. Chem.* **278**:30213–30218.
- Leal, W. S. 2003. Proteins that make sense. Pp. 447–476 in G. J. Blomquist and R. G. Vogt, eds. *Insect pheromone biochemistry and molecular biology: the biosynthesis and detection of pheromones and plant volatiles*. Elsevier Academic Press, London.
- Lee, D., F. F. Damberger, G. H. Peng, R. Horst, P. Guntert, L. Nikonova, W. S. Leal, and K. Wuthrich. 2002. NMR structure of the unliganded *Bombyx mori* pheromone-binding protein at physiological pH. *FEBS Lett.* **531**:314–318.
- Li, W.-H. 1997. *Molecular evolution*. Sinauer Associates, Sunderland, Mass.
- Mackay, W. P., S. Porter, D. Gonzalez, A. Rodriguez, H. Armendado, A. Rebeles, and S. B. Vinson. 1990. A comparison of monogyne and polygyne populations of the tropical fire ant, *Solenopsis geminata* (Hymenoptera: Formicidae), in Mexico. *J. Kans. Entomol. Soc.* **63**:611–615.
- McGuffin, L. J., and D. T. Jones. 2003. Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics* **19**:874–881.
- McGuire, G., and F. Wright. 2000. TOPAL 2.0: improved detection of mosaic sequences within multiple alignments. *Bioinformatics* **16**:130–134.
- Milne, I., F. Wright, G. Rowe, D. F. Marshal, D. Husmeier, and G. McGuire. 2004. TOPALi: software for automatic identification

- of recombinant sequences within DNA multiple alignments. *Bioinformatics* **20**:1806–1807.
- Myers, S. R., and R. C. Griffiths. 2003. Bounds on the minimum number of recombination events in a sample history. *Genetics* **163**:375–394.
- Nielsen, R., and Z. H. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**:929–936.
- Pelosi, P., and R. Maida. 1995. Odorant-binding proteins in insects. *Comp. Biochem. Physiol. B* **111**:503–514.
- Pikielny, C. W., G. Hasan, F. Rouyer, and M. Rosbash. 1994. Members of a family of *Drosophila* putative odorant-binding proteins are expressed in different subsets of olfactory hairs. *Neuron* **12**:35–49.
- Pitts, J. P. 2002. A cladistic analysis of the *Solenopsis saevissima* species group (Hymenoptera: Formicidae). Doctoral dissertation, University of Georgia, Athens.
- Pitts, J. P., J. V. McHugh, and K. G. Ross. 2005. A cladistic analysis of the fire ants of the *Solenopsis saevissima* species-group (Hymenoptera: Formicidae). *Zool. Scr.* (in press).
- Porter, S. D., A. Bhatkar, R. Mulder, S. B. Vinson, and D. J. Clair. 1991. Distribution and density of polygyne fire ants (Hymenoptera: Formicidae) in Texas. *J. Econ. Entomol.* **84**:866–874.
- Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- . 2002. The effect of recombination on the accuracy of phylogeny estimation. *J. Mol. Evol.* **54**:396–402.
- Robinson, G. E., C. M. Grozinger, and C. W. Whitfield. 2005. Sociogenomics: social life in molecular terms. *Nat. Rev. Genet.* **6**:257–270.
- Rooney, A. P., and J. Z. Zhang. 1999. Rapid evolution of a primate sperm protein: relaxation of functional constraint or positive Darwinian selection? *Mol. Biol. Evol.* **16**:706–710.
- Ross, K. G. 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. Comparative study of genetic and social structure in two forms of the fire ant, *Solenopsis invicta* (Hymenoptera: Formicidae). *Behav. Ecol. Sociobiol.* **17**:349–356.
- Ross, K. G., and L. Keller. 1998. Genetic control of social organization in an ant. *Proc. Natl. Acad. Sci. USA* **95**:14232–14237.
- . 2002. Experimental conversion of colony social organization by manipulation of worker genotype composition in fire ants (*Solenopsis invicta*). *Behav. Ecol. Sociobiol.* **51**:287–295.
- Ross, K. G., M. J. B. Krieger, and D. D. Shoemaker. 2003. Alternative genetic foundations for a key social polymorphism in fire ants. *Genetics* **165**:1853–1867.
- Ross, K. G., and D. D. Shoemaker. 2005. Species delimitation in native South American fire ants. *Mol. Ecol.* (in press).
- Ross, K. G., and J. C. Trager. 1990. Systematics and population genetics of fire ants (*Solenopsis saevissima* complex) from Argentina. *Evolution* **44**:2113–2134.
- Sandler, B. H., L. Nikonova, W. S. Leal, and J. Clardy. 2000. Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex. *Chem. Biol.* **7**:143–151.
- Shoemaker, D. D., and K. G. Ross. 1996. Effects of social organization on gene flow in the fire ant *Solenopsis invicta*. *Nature* **383**:613–616.
- Suzuki, Y. 2004. New methods for detecting positive selection at single amino acid sites. *J. Mol. Evol.* **59**:11–19.
- Suzuki, Y., and T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* **16**:1315–1328.
- Suzuki, Y., and M. Nei. 2002. Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* **19**:1865–1869.
- . 2004. False-positive selection identified by ML-based methods: examples from the *Sig1* gene of the diatom *Thalassiosira weissflogii* and the *tax* gene of a human T-cell lymphotropic virus. *Mol. Biol. Evol.* **21**:914–921.
- Swofford, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.10. Sinauer Associates, Sunderland, Mass.
- Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**:823–833.
- Trager, J. C. 1991. A revision of the fire ants, *Solenopsis geminata* group (Hymenoptera: Formicidae: Myrmicinae). *J. NY Entomol. Soc.* **99**:141–198.
- Vargo, E. L., and D. J. C. Fletcher. 1987. Effect of queen number on the production of sexuals in natural-populations of the fire ant, *Solenopsis invicta*. *Physiol. Entomol.* **12**:109–116.
- Vogt, R. G. 2002. Odorant binding protein homologues of the malaria mosquito *Anopheles gambiae*; possible orthologues of the OS-E and OS-F OBPs of *Drosophila melanogaster*. *J. Chem. Ecol.* **28**:2371–2376.
- . 2003. Biochemical diversity of odor detection: OBPs, ODEs and SNMPs. Pp. 391–445 in G. J. Blomquist and R. G. Vogt, eds. *Insect pheromone biochemistry and molecular biology: the biosynthesis and detection of pheromones and plant volatiles*. Elsevier Academic Press, London.
- Yang, Z. H. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**:568–573.
- . 1999. Phylogenetic analysis by maximum likelihood (PAML). (<http://abacus.gene.ucl.ac.uk/software/paml.html>). University College London, London.
- Yang, Z. H., R. Nielsen, N. Goldman, and A. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino-acid sites. *Genetics* **155**:431–449.
- Zhang, J. Z. 2004. Frequent false detection of positive selection by the likelihood method with branch-site models. *Mol. Biol. Evol.* **21**:1332–1339.
- Zhang, J. Z., S. Kumar, and M. Nei. 1997. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. *Mol. Biol. Evol.* **14**:1335–1338.

Scott Edwards, Associate Editor

Accepted June 21, 2005