Evolution at Two Levels in Fire Ants: The Relationship between Patterns of Gene Expression and Protein Sequence Evolution

Brendan G. Hunt,*1 Lino Ometto,2 Laurent Keller,3 and Michael A.D. Goodisman1

1School of Biology, Georgia Institute of Technology
2Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, San Michele all’Adige (TN), Italy
3Department of Ecology and Evolution, University of Lausanne, Biophore, Lausanne, Switzerland

*Corresponding author: E-mail: brendan.hunt@gatech.edu.
Associate editor: Douglas Crawford

Abstract

Variation in protein sequence and gene expression each contribute to phenotypic diversity, and may be subject to similar selective pressures. Eusocial insects are particularly useful for investigating the evolutionary link between protein sequence and condition-dependent patterns of gene expression because gene expression plays a central role in determining differences between eusocial insect sexes and castes. We investigated the relationship between protein coding sequence evolution and gene expression patterns in the fire ants Solenopsis invicta, S. richteri, and their hybrids to gain greater insight into how selection jointly operates on gene expression and coding sequence. We found that genes with high expression variability within castes and sexes were frequently differentially expressed between castes and sexes, as well as between species and hybrids. These results indicate that genes showing high variation in expression in one context also tend to show high variation in expression in other contexts. Our analyses further revealed that variation in both intra- and interspecific gene expression was positively associated with rate of protein sequence evolution in Solenopsis. This suggests that selective constraints on a gene operate both at the level of protein sequence and at the level of gene expression regulation. Overall, our study provides one of the strongest demonstrations that selective constraints mediate both protein sequence evolution and gene expression variability across different biological contexts and timescales.

Key words: caste polyphenism, hybrid incompatibility, selective constraint, sexual dimorphism, social insect.

Introduction

The rate at which a protein evolves reflects its functional importance. For example, essential proteins typically display relatively slow rates of sequence evolution, resulting from strong purifying selection, when compared with proteins of lesser importance to fitness (Hirsh and Fraser 2001; Wall et al. 2005; Waterhouse et al. 2011). Similarly, variation in the rate at which gene expression evolves may be linked to functional importance. Genes that less strongly affect fitness exhibit greater stochasticity in expression (Fraser et al. 2004) and may diverge in expression level at a higher rate between taxa than genes with larger fitness effects (Jordan et al. 2005).

The rate of gene expression evolution and protein sequence evolution may also be related to each other. Several studies suggest that rates of evolution for gene expression levels and protein sequences are indeed directly associated (Nuzhdin et al. 2004; Jordan et al. 2005; Khaitovich et al. 2005; Lemos et al. 2005), indicating that similar selective pressures may shape the evolutionary histories of protein sequence and gene expression. Furthermore, because gene expression and protein sequence each influence phenotype (Carroll 2005), a better understanding of the evolutionary relationship between gene expression and protein sequence may provide insights into processes that ultimately contribute to phenotypic diversification.

Eusocial insects are ideal models for studying the link between gene expression and phenotype because they offer dramatic cases of distinct phenotypes produced by variation in gene expression (Smith et al. 2008; Hunt et al. 2010; Gadau et al. 2012). For instance, female eusocial hymenopterans (ants, eusocial bees, and eusocial wasps) develop either into queens, which are primarily responsible for reproduction, or workers, which undertake tasks related to brood rearing and colony defense (Wilson 1971). In most species, these queen and worker castes derive from the differential expression of genes in otherwise equivalent genomes (Goodisman et al. 2008; Smith et al. 2008; Schwander et al. 2010). Moreover, male and female hymenopterans possess the same genes because sex is essentially determined by ploidy level (males are haploid and females diploid; Heimpel and de Boer 2008). Consequently, sexual dimorphism in hymenopteran insects also results from gene expression variation between the sexes.

These attributes make the eusocial insects particularly well suited to investigate associations between gene expression patterns among distinct morphs of an organism and constraints on protein coding sequence evolution.

Here, we investigate the relationship between the evolution of coding sequence and the evolution of gene expression in the fire ants Solenopsis invicta and S. richteri. We first test whether different forms of variation in gene expression, such as expression variation within and among distinct morphs of
a species (i.e., castes or sexes) or expression differentiation between taxa, are each mediated by similar evolutionary processes. If this is the case, we predict that different measures of variation in gene expression should be correlated with one another.

Second, we test how variation in gene expression is linked to rates of protein sequence evolution. We hypothesize that similar evolutionary processes affect variation at the level of gene expression and protein sequence. Based upon this hypothesis, we predict that protein evolutionary rate will be directly associated with variation in gene expression in diverse phenotypic contexts. To address this issue, we investigate whether measures of variation in gene expression within and between Solenopsis castes, sexes, and species are correlated with the rate of protein evolution.

Finally, we investigate the relationship between coding sequence evolution and patterns of gene expression in a hybrid population of fire ants (Ometto et al. 2012). *S. invicta* and *S. richteri* were introduced into the United States in the early 1900s and readily hybridize in their invasive range (Ross and Robertson 1990; Shoemaker et al. 1996; Goodisman et al. 1998). Hybridization can lead to gene expression differences between hybrids and parental species that may be associated with a decrease in fitness (Ranz et al. 2004; Haerty and Singh 2006). Based on the relationship between evolution occurring at the level of protein sequence and gene expression (Nuzhdin et al. 2004; Jordan et al. 2005; Khaitovich et al. 2005; Lemos et al. 2005), we predict that rapidly evolving genes will be more likely to exhibit disruption in expression levels in *S. invicta*–*S. richteri* hybrids (e.g., Artieri et al. 2007). To test this prediction, we determine whether genes that exhibit relatively high rates of sequence evolution are prone to differences in expression in hybrids relative to parental species. Overall, our study yields novel insights into evolutionary processes underlying phenotypic variation within and between species.

**Materials and Methods**

**Measures of Gene Expression**

Gene expression data for *S. invicta*, *S. richteri* (Ometto et al. 2011), and their hybrids (Ometto et al. 2012) were obtained using cDNA microarrays (Wang et al. 2007). Gene expression was measured for age-matched whole bodies of each of six different fire ant morphs: workers, queens, and males, each at both the pupal and adult stages (Ometto et al. 2011, 2012). Gene expression measurements were obtained from four to six biological replicates per morph type (as described in supplementary table S1 of Ometto et al. 2012). Relative gene expression levels were estimated using a Bayesian approach implemented in the program Bayesian Analysis of Gene Expression Levels (BAGEL; Townsend and Hartl 2002; Meiklejohn and Townsend 2005; Ometto et al. 2011, 2012). The BAGEL method is widely implemented in microarray comparisons of gene expression levels from multiple samples (Meiklejohn et al. 2003; Ranz et al. 2003; Whitfield et al. 2003; Meiklejohn and Townsend 2005; Gnad and Parsch 2006; Grozinger et al. 2007). For well-replicated microarray experiments, the results of BAGEL are largely in agreement with those produced by other methodologies (Meiklejohn and Townsend 2005), including analysis of variance (ANOVA) (Whitfield et al. 2003). However, BAGEL offers improvements over ANOVA by eliminating the need to correct for spot effects in the analysis of nonmodel microarray platforms, by not requiring balanced data, and by accepting variability in replication across expression nodes (Meiklejohn and Townsend 2005). A proxy for gene expression level in *S. invicta* was determined for each gene as the ratio between the net signal intensity of a microarray probe and the average net signal intensity across all microarray probes. This ratio, as first determined separately for each sample, was subsequently summed across all samples (from all morphs) to provide our measure of overall gene expression level (Hunt et al. 2011).

Expressed sequence tag (EST) sequences corresponding to cDNA microarray probes (Wang et al. 2007; Ometto et al. 2011) were mapped to official *S. invicta* gene models (Wurm et al. 2011) to facilitate molecular evolutionary analysis of protein coding genes (Hunt et al. 2011). When more than one EST sequence mapped to a given gene model, we used the mean expression value of the microarray probes representing these ESTs as a measure of expression for the given gene (Hunt et al. 2011).

We considered two measures of gene expression variability within *S. invicta*. First, for each gene, we calculated the coefficient of variation (standard deviation/mean) of expression values for each morph, as estimated by the ratio between the normalized intensities of each experimental sample to a common reference (Ometto et al. 2011). We then estimated the overall variability in expression for a given gene [Var(morph)] as the mean of coefficients of variation estimated for each of the six morphs (table 1). Thus, Var(morph) provided an estimate of variability of expression within morphs, with high values indicative of high levels of variation in gene expression among individuals within morphs in *S. invicta*. In contrast, low values of Var(morph) indicate that there is little variation in the expression of a gene within each morph.

We also measured a second set of metrics of within species expression variation corresponding to the degree of expression differentiation between morphs within *S. invicta*. These measures were calculated for adults and pupae separately as the absolute value of log2-transformed ratios between the BAGEL-estimated expression values of distinct morphs. The difference in expression between queens and workers is described by the measure Dif(caste) and the difference in expression between males and queens is described by the measure Dif(sex) (Hunt et al. 2011; Ometto et al. 2011). We further calculated an overall measure of gene expression differentiation among all morphs, Dif(morph), as the mean of Dif(caste adult), Dif(caste pupa), Dif(sex adult), and Dif(sex pupa) in *S. invicta* (table 1). Large values of Dif(morph), Dif(caste), or Dif(sex) indicated that a particular gene showed strong differences in expression among morphs, castes, or sexes, whereas low values indicated that a gene was expressed at similar levels among morphs, castes, or sexes, respectively.
Next, we calculated the difference in expression between *S. invicta* and *S. richteri* for each gene. This differentiation was calculated separately for each of the six morphs (described earlier) as the absolute value of the log2-transformed ratio between BAGEL-estimated expression values in *S. invicta* and *S. richteri* (Ometto et al. 2011). We then estimated the overall interspecific expression differentiation for a given gene, Dif(species), as the mean of gene expression differentiation measures calculated for each of the six morphs (Table 1). Therefore, genes with high values of Dif(species) are those with large differences in expression level between *S. invicta* and *S. richteri*, whereas genes with low values of Dif(species) show similar expression levels between species.

Finally, we investigated whether genes showed substantial differences in expression between hybrids and the hybridizing parental species, *S. invicta* and *S. richteri*. Gene expression differentiation between the hybrids and each parental species was calculated for each of the six morphs as the absolute value of the log2-transformed ratio between BAGEL-estimated expression in hybrids and each parental species separately. We then estimated measures of overall expression differentiation between the hybrids and each parental species as the mean of expression differentiation measures calculated for each of the six morphs. This produced two metrics representing how divergent the hybrid expression value was from *S. invicta* and *S. richteri*, respectively. We took the lesser of these two values as our measure of minimum hybrid gene expression differentiation, Dif(hybrid) (Table 1). Therefore, genes with high values of Dif(hybrid) are those with large differences in expression level between hybrids and both of the parental species. In contrast, genes with low values of Dif(hybrid) are expressed at similar levels in hybrids as in *S. invicta* and/or *S. richteri*.

We classified hybrid genes as either “intermediately expressed,” “underexpressed,” or “overexpressed” as follows. First, we calculated the differentiation in expression of hybrids relative to 1) *S. invicta* and 2) *S. richteri* as discussed earlier, but did not take the absolute value of these measures. If the two hybrid differentiation measures were positive, the gene was considered “overexpressed” in hybrids. That is, the hybrid expression level was greater than that of the two parental taxa. If the two measures differed in sign, the gene was considered “intermediately expressed” in hybrids. That is, the hybrid expression value fell between that of the two parental species. Finally, if the two measures of differentiation were negative, the gene was considered “underexpressed” in hybrids. In this final case, the hybrid expression value was below that of both *S. invicta* and *S. richteri*.

**Interspecific Array Hybridization**

*S. richteri* expression levels were determined using *S. invicta* cDNA microarrays. If *S. richteri* target sequences exhibited diminished annealing to *S. invicta* microarray probe sequences due to sequence mismatches, an ortholog with identical expression levels in each species may display decreased expression in *S. richteri* as a technical artifact. Consequently, the use of *S. invicta* microarrays for determining expression in *S. richteri* may have affected the measures of expression if *S. richteri* gene sequences differed from those in *S. invicta*. However, two distinct analyses suggest that this was not a major issue in our study.

First, we investigated if the rate of synonymous substitution (dS) was positively correlated with Dif(species), as expected if probe mismatches in *S. richteri* were driving the correlation between Dif(species) and protein coding sequence divergence (Nuzhdin et al. 2004). This analysis revealed that Dif(species) and dS were not significantly correlated (*P* = 0.284; Table 2).

As a second control, we limited our analysis of the correlation between rates of molecular evolution and gene expression to only those 540 genes that exhibited a higher mean BAGEL-estimated expression value across morphs in *S. richteri* than in *S. invicta*. We reasoned that such genes were unlikely

---

**Table 1. Measures of Gene Expression Variation in Solenopsis Fire Ants.**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Measure of Gene Expression</th>
<th>Calculation*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var(morph)</td>
<td>Variation among samples within morphs in <em>S. invicta</em></td>
<td>$\sum_{n=1}^{n} \frac{SD(Y_{\text{invicta}}/X_{\text{morph}})}{n}$</td>
</tr>
<tr>
<td>Dif(caste)</td>
<td>Differentiation between castes in <em>S. invicta</em></td>
<td>$\log_{Y_{\text{invicta}}/X_{\text{worker}}} - \log_{Y_{\text{invicta}}/X_{\text{queen}}}$</td>
</tr>
<tr>
<td>Dif(sex)</td>
<td>Differentiation between sexes in <em>S. invicta</em></td>
<td>$\log_{Y_{\text{invicta}}/X_{\text{male}}} - \log_{Y_{\text{invicta}}/X_{\text{female}}}$</td>
</tr>
<tr>
<td>Dif(morph)</td>
<td>Differentiation among morphs in <em>S. invicta</em></td>
<td>$\sum_{n=1}^{n} \left( \frac{\text{Dif(caste)}}{\text{stage}} + \frac{\text{Dif(sex)}}{\text{stage}} \right) / n$</td>
</tr>
<tr>
<td>Dif(species)</td>
<td>Differentiation between <em>S. invicta</em> and <em>S. richteri</em></td>
<td>$\min\left( \sum_{n=1}^{n} \log_{Y_{\text{invicta}}/X_{\text{morph}}} - \sum_{n=1}^{n} \log_{Y_{\text{richteri}}/X_{\text{morph}}} / n \right)$</td>
</tr>
</tbody>
</table>

$X$ represents the BAGEL gene expression value of a given morph and $Y$ represents the ratio between the normalized dye intensities of each experimental sample and a common reference for a given morph; “morph” represents each of the six fire ant morphs (queens, workers, and males, each at the adult and pupal stages); “stage” represents the pupal and adult developmental stages; SD, standard deviation; min, the minimum value; superscripts are used to denote the taxon.
to have been affected by putative array mishybridization, which would instead be expected to result in lower levels of expression in *S. richteri* than in *S. invicta*. Our analysis of this subset of genes revealed that Diff(species) remained positively correlated with dN/dS with a Spearman’s rank correlation of 0.226 (P < 0.0001), suggesting that the correlation between Diff(species) and dN/dS in our entire data set did not result from biases arising from *S. richteri* mishybridization to *S. invicta* microarrays.

### Measures of Sequence Evolution

To estimate relative rates of protein coding sequence evolution in *S. invicta*, we used previously published (Hunt et al. 2011) measures of nonsynonymous and synonymous substitutions (dN and dS, respectively) for *S. invicta* sequences (Wurm et al. 2011) compared with orthologous sequences from the ants *Pogonomyrmex barbatus* (Smith, Smith, et al. 2011) and *Linepithema humile* (Smith, Zimin, et al. 2011). We also calculated dN/dS as a measure of selective constraint on protein sequence, with larger values of dN/dS generally corresponding to weaker selective constraint (Hunt et al. 2011). Substitution rates were averaged across all aligned codons for a given protein, with free dN/dS ratios for each branch, using PAML (Yang 2007). After filtering our data based on the mapping of array sequences to the *S. invicta* official gene set, high-confidence aligned blocks of orthologous sequence, sequence quality, and the presence of infinite dN/dS values (see Hunt et al. 2011 for detailed methods), we assessed paired measures of protein coding sequence evolution and measures of gene expression for a total of 1,101 *S. invicta* genes.

To investigate whether variation in protein evolutionary rate was mostly influenced by purifying or positive selection, we compared measures of coding sequence polymorphism within *S. invicta* to measures of protein substitutions between species (McDonald and Kreitman 1991). Sequence polymorphism was assessed in *S. invicta* as described previously (Hunt et al. 2011). In total, we identified 381 genes with single nucleotide polymorphisms (SNPs) in *S. invicta*, sequence divergence data among taxa, and gene expression information from *Solenopsis*. Because few SNPs per locus were detected from available data, we performed polymorphism analyses on groups of pooled genes.

### Gene Ontology

*S. invicta* gene ontology (GO) annotations were assigned using Blast2GO (Conesa et al. 2005). Blast2GO’s inbuilt “gossip” package was used to test for enrichment using a Fisher’s exact test, correcting for multiple testing using a Benjamini–Hochberg false discovery rate (FDR). The “generic GO slim” subset of GO terms was used to assess significantly enriched terms (FDR, P < 0.05), which were reduced to the most specific enriched terms for presentation.

### Statistics

The JMP statistical software package (SAS Institute Inc., Cary, NC) was used to calculate Spearman’s rank correlations, to calculate 95% confidence intervals of means, and to perform principal component analysis.

### Results and Discussion

#### Measures of Gene Expression Variation in Different Biological Contexts Are Correlated

Gene expression variation within species may be subject to the same evolutionary processes as gene expression divergence between species (Whitehead and Crawford 2006). If this is the case, gene expression variability within species should correlate with differentiation between species. We found that intraspecific variation in gene expression, as measured by Var(morph), and interspecific differentiation in gene expression, as measured by Diff(species), were indeed highly correlated (Spearman’s $p = 0.523$, $P < 0.0001$; supplementary table S1, Supplementary Material online) (Meiklejohn et al. 2003). Furthermore, the differentiation of gene expression in hybrids of *S. invicta* and *S. richteri* relative to parental species, as measured by Diff(hybrid), was also highly correlated (Spearman’s $p = 0.523$, $P < 0.0001$; supplementary table S1, Supplementary Material online) (Meiklejohn et al. 2003). The positive correlation between Diff(species) and Diff(hybrid) suggests that regulatory incompatibilities between genomes may increase in line with gene expression divergence between species (Ranz et al. 2004; Haerty and Singh 2006). Overall, these data support the idea that gene expression variability within species and differentiation between taxa are subject to similar selective processes.

These results led us to hypothesize that gene expression variability would be associated with gene expression differentiation, not only between species but also between phenotypic morphs, such as sexes or castes, within species. Accordingly, we compared measures of morph-biased (i.e., caste- or sex-biased) gene expression to gene expression variability within each caste or sex in *S. invicta* [Var(morph)]. We found that caste bias [Diff(caste)] and sex bias [Diff(sex)] in both the adult and pupal stages were each positively correlated with Var(morph) (table 4). Furthermore, overall morph bias [Diff(morph)], measured as the mean of sex and caste bias measures, was even more strongly associated with Var(morph) than individual measures of caste or sex bias (table 4). Thus, genes with high within-morph expression variability were also more likely to exhibit high expression differences between morphs (Meiklejohn et al. 2003; Mank et al. 2008; Leichty et al. 2012). This finding is consistent with...
Table 3. Spearman's Rank Correlations (ρ) between the Differentiation of Gene Expression in Hybrids Relative to Parental Taxa [Dif(hybrid)] for Genes Showing Different Hybrid Expression Patterns and Gene Expression Differentiation between Solenopsis invicta and S. richteri [Dif(species)], Gene Expression Variability within S. invicta [Var(morph)], or Measures of Protein Coding Sequence Evolution (dN, dS, and dN/dS).

<table>
<thead>
<tr>
<th>Expression Bias (X)</th>
<th>Genes (n)</th>
<th>ρX,Var(morph)</th>
<th>ρX,Dif(species)</th>
<th>ρX,Dif(hybrid)</th>
<th>ρX,dN</th>
<th>ρX,dS</th>
<th>ρX,dN/dS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over-expressed</td>
<td>222</td>
<td>0.727***</td>
<td>0.791***</td>
<td>0.266***</td>
<td>-0.074N</td>
<td>0.336***</td>
<td></td>
</tr>
<tr>
<td>Intermediately</td>
<td>512</td>
<td>0.715***</td>
<td>0.602***</td>
<td>0.161***</td>
<td>-0.074N</td>
<td>0.187***</td>
<td></td>
</tr>
<tr>
<td>Under-expressed</td>
<td>367</td>
<td>0.589***</td>
<td>0.604***</td>
<td>0.143**</td>
<td>-0.047N</td>
<td>0.165***</td>
<td></td>
</tr>
<tr>
<td>All genes</td>
<td>1,101</td>
<td>0.660****</td>
<td>0.651****</td>
<td>0.184****</td>
<td>-0.070*</td>
<td>0.218****</td>
<td></td>
</tr>
</tbody>
</table>

*N < 0.05, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

Table 4. Spearman's Rank Correlations (ρ) between Differentiation of Expression between Adult or Pupal Solenopsis invicta Castes or Sexes, and within-Morph Variability in Gene Expression [Var(morph)], Gene Expression Differentiation between Taxa [Dif(species)], or Measures of Protein Coding Sequence Evolution (dN, dS, and dN/dS).

<table>
<thead>
<tr>
<th>Expression Bias (X)</th>
<th>Genes (n)</th>
<th>ρX,Var(morph)</th>
<th>ρX,Dif(species)</th>
<th>ρX,Dif(hybrid)</th>
<th>ρX,dN</th>
<th>ρX,dS</th>
<th>ρX,dN/dS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dif(caste adult)</td>
<td>0.280****</td>
<td>0.210****</td>
<td>0.138****</td>
<td>0.016**</td>
<td>0.143***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dif(caste pupa)</td>
<td>0.307****</td>
<td>0.259****</td>
<td>0.065*</td>
<td>-0.085**</td>
<td>0.103**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dif(sex adult)</td>
<td>0.330****</td>
<td>0.360****</td>
<td>0.157****</td>
<td>-0.038**</td>
<td>0.175***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dif(sex pupa)</td>
<td>0.233****</td>
<td>0.222****</td>
<td>0.137****</td>
<td>-0.007NS</td>
<td>0.143***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dif(morph)*</td>
<td>0.473****</td>
<td>0.442****</td>
<td>0.213****</td>
<td>-0.046NS</td>
<td>0.239***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*N < 0.05, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

*aMean of four other sex and caste “Dif” measures.

the hypothesis that genes exhibiting higher differential expression among morphs tend to be subject to diminished selective constraint on expression relative to genes which are expressed similarly among morphs (Mank and Ellegren 2009; Hunt et al. 2011; Leichthy et al. 2012).

We also found that the degree of caste bias or sex bias in gene expression, at both the adult and pupal stages in S. invicta, was positively correlated with the difference in expression between S. invicta and S. richteri [Dif(species)] (table 4). Furthermore, overall morph bias [Dif(morph)] was even more strongly associated with Dif(species) than individual measures of caste bias or sex bias (table 4). This suggests that genes with a higher degree of morph-biased expression diverge more rapidly in their expression levels between taxa than genes with less-biased expression among morphs (Meiklejohn et al. 2003; Ranz et al. 2003; Zhang et al. 2007; Hunt and Goodisman 2010; Ometto et al. 2011).

We next performed a principal component analysis of Var(morph), Dif(morph), Dif(species), and gene expression level to examine the extent to which these measures were interdependent. This analysis revealed that over 50% of the variation in these five measures was explained by a single principal component (PC1, fig. 1a). Var(morph), Dif(morph), Dif(species), and Dif(hybrid) each contributed heavily to PC1, indicating that PC1 is representative of overall variation in gene expression. In contrast, gene expression level did not contribute substantially to PC1 but was the sole dominant contributor to PC2, indicating that gene expression level was largely independent of variation in gene expression in our data (fig. 1a and table 2). These results demonstrate that gene expression variation occurring within a morph, between morphs, and between taxa all covary in Solenopsis (supplementary fig. S1, Supplementary Material online).

This covariation of gene expression may be explained by similar evolutionary processes operating on gene expression in each of these contexts.

Gene Expression Evolution Is Associated with Protein Sequence Evolution

Protein sequence evolution and gene expression evolution may be affected by similar selective processes. If this is the case, the rate of divergence in expression levels between taxa [Dif(species)] should be correlated with the rate of protein sequence evolution. In support of this idea, we found that the rate of protein sequence evolution (dN) and selective constraint (dN/dS) were each positively correlated with the differentiation of expression between S. invicta and S. richteri (table 2, supplementary table S2, Supplementary Material online, and fig. 2). Similar results have been found in flies (Nuzhdin et al. 2004; Lemos et al. 2005) and mammals (Jordan et al. 2005; Khaitovich et al. 2005). Thus, genes showing larger differences in expression levels between species tend to evolve more rapidly at the sequence level than genes with similar expression between species.

We also investigated the relationship between coding sequence evolution and gene expression differentiation in S. invicta–S. richteri hybrids relative to the parental species [Dif(hybrid)]. We hypothesized that divergence in a gene’s regulatory machinery would be associated with divergence in protein sequences among hybridizing taxa (Castillo-Davis et al. 2004), similar to the observed correlation between Dif(species) and dN in S. invicta (table 2). Consequently, we predicted that the degree of differentiation of hybrid gene expression relative to parental taxa [Dif(hybrid)] would be positively correlated with dN and dN/dS (Artieri et al. 2007).
Fig. 1. Principal component analysis of Solenopsis gene expression measures. (a) Percentage variance explained by five PCs generated from gene expression variation within morphs [Var(morph)], morph bias in gene expression [Dif(morph)], gene expression divergence between species [Dif(species)], gene expression divergence between hybrids and parental species [Dif(hybrid)], and gene expression level. Shading depicts the proportional contribution of each gene expression measure to each PC. PC1 is representative of overall gene expression variation, whereas PC2 is representative of gene expression level. (b) Spearman’s rank correlations between PCs and dN/dS, with relative contribution of each gene expression measure to each PC shown. Only those PCs explaining at least 10% of the total variation in gene expression measures are presented in (b).

Fig. 2. Protein sequence evolution is correlated with gene expression variation and divergence. dN/dS is associated with (a) gene expression variability within S. invicta [Var(morph)], (b) gene expression differentiation among morphs in S. invicta [Dif(morph)], (c) gene expression differentiation between Solenopsis invicta and S. richteri [Dif(species)], and (d) gene expression differentiation between Solenopsis hybrids and parental taxa [Dif(hybrid)]. Means and 95% confidence intervals are plotted on each axis, with genes grouped according to Var(morph), Dif(morph), Dif(species), or Dif(hybrid) deciles (n = 110 or 111 genes in each decile). Spearman’s rank correlation coefficients (ρ) are shown (all P < 0.0001, n = 1,101).

We found that Dif(hybrid) was positively correlated with dN and dN/dS in S. invicta (table 3 and fig. 2). These positive correlations were present in genes with intermediate expression levels in hybrids relative to parental species, as well as genes that exhibited under- or overexpression in hybrids relative to parental species (table 3). Thus, genes that exhibit divergent expression in hybrids relative to the parental species tend to evolve more rapidly at the sequence level, irrespective of whether hybrid gene expression is intermediate to, or outlying, parental gene expression.

To assess whether gene expression variation within a species was also associated with the rate of protein evolution, we determined whether variation in expression within S. invicta morphs [Var(morph)] and differentiation between S. invicta morphs [Dif(morph)] were each correlated with dN and dN/dS. We found that within-morph variability in gene expression [Var(morph)] was positively correlated with dN and dN/dS (table 2, supplementary table S3, Supplementary Material online, and fig. 2) (Fraser et al. 2004; Leichty et al. 2012). Furthermore, the degree of differentiation in gene expression among morphs [Dif(morph)] was positively correlated with dN and dN/dS (table 4 and fig. 2) (Gnad and Parsch 2006; Hunt et al. 2011; Snell-Rood et al. 2011). Thus, gene expression variation, both within and between morphs of a species, is positively correlated with the rate of protein sequence evolution. Notably, the positive correlation we observed between the rate of sequence evolution and expression bias among whole-organism morphs mirrors the widely observed positive correlation between the rate of sequence evolution and expression bias among tissues within an organism (e.g., Duret and Mouchiroud 2000). These similarities may reflect a general link between context-dependent expression and reduced selective constraint (Snell-Rood et al. 2010) or the importance of tissue-specific genes to differentiating morphs (e.g., Meisel 2011).

We next sought to test whether the correlations observed between distinct measures of gene expression variation and protein evolution were independent of one another. To do so, we correlated dN/dS with PCs from our analysis of gene
expression level, intraspecific variation in expression \([\text{Var(morph)}]\), morph bias in expression \([\text{Dif(morph)}]\), interspecific divergence in expression \([\text{Dif(species)}]\), and expression divergence between hybrids and parental species \([\text{Dif(hybrid)}]\). This analysis revealed that a single PC (PC1, fig. 1), which reflected overall variation in gene expression, was significantly positively correlated with \(dN/dS\) (fig. 1b). Furthermore, gene expression level loaded primarily on PC2, suggesting that the association between PC1 and \(dN/dS\) was independent from the well-described association between gene expression level and protein evolutionary rate (Drummond and Wilke 2008) (fig. 1).

To further explore the nature of selection shaping variation in protein evolution in our data, we compared sequence polymorphism within \(S. \text{invicta}\) to sequence divergence between the \(S. \text{invicta}\) lineage and the outgroup taxa \(P. \text{barbatus}\) and \(L. \text{humile}\) (Hunt et al. 2011). We detected strongly significant \((P = 6.0 \times 10^{-5})\) purifying selection in a pooled analysis of genes with relatively invariable expression according to PC1 from our principal component analysis (PC1 lower quartile, table 5). In contrast, genes with relatively variable expression (PC1 upper quartile) did not deviate significantly from neutrality \((P = 0.366; \text{table 5})\). Thus, variation in protein evolutionary rate appears to have been driven largely by variation in the strength of purifying selection (Hunt et al. 2011). This finding is consistent with a recent analysis of protein evolution in 29 mammalian genomes (Lindblad-Toh et al. 2011), which revealed that 85% of genes were subject to uniformly high purifying selection throughout their lengths. Nevertheless, our data do not allow us to rule out a possible role of positive selection in variation in the rate of protein evolution (Sela et al. 2009).

Our results suggest that selective constraints on a protein’s function act on both coding sequence and expression (Nuzhdin et al. 2004; Jordan et al. 2005; Khaitovich et al. 2005; Lemos et al. 2005). This, in turn, suggests that gene expression differentiation may arise more frequently at loci that are subject to weak selective constraint (Hunt et al. 2011; Leichty et al. 2012). Consistent with this idea, we hypothesized that genes exhibiting low variability in expression would tend to be associated with “housekeeping” functions essential to all cells (Fraser et al. 2004). To test this hypothesis, we grouped genes according to whether they were in the upper or lower 50th percentile for PC1 values (fig. 1) and tested for enrichment of GO functional annotations. We found that genes with relatively low expression variability (lower 50th percentile of PC1) exhibited enrichment of several GO annotations related to core cellular processes, including approximately 3-fold enrichment of annotations associated with “translation” (supplementary table S4, Supplementary Material online). In contrast, genes with relatively variable expression (upper 50th percentile of PC1) did not exhibit significant enrichment of GO annotations. These results are consistent with a scenario in which genes that exhibit low variability of expression are constitutively expressed (Minsky et al. 2012) and preferentially perform functions integral to all cells of an organism. This suggests that genes with relatively low expression variation may be preferentially associated with cellular “housekeeping” functions.

The evolutionary coupling we observed between gene expression and protein sequence appears to persist over vast timescales, as indicated by differences in the divergence times between species used to address protein sequence evolution and gene expression evolution. Rates of protein sequence evolution were estimated from sequence divergence between the ants \(P. \text{barbatus}\) and \(S. \text{invicta}\), which diverged from a common ancestor approximately 110 Ma (Moreau et al. 2006). In contrast, gene expression divergence was estimated between \(S. \text{invicta}\) and \(S. \text{richteri}\), which are sister species within the \(S. \text{saevissima}\) species group (Pitts et al. 2005). Thus, long-standing evolutionary constraints on proteins are apparently associated with evolution in gene expression on a much smaller timescale (Mank and Ellegren 2009; Hunt et al. 2010, 2011; Leichty et al. 2012).

**Conclusion**

We have shown that distinct measures of gene expression variation are correlated with one another, indicating that genes that show substantial variation in expression in one biological context tend to show substantial variation in other contexts as well. Furthermore, similar selective processes may operate on gene expression and protein sequence evolution, as suggested by the association between rates of protein evolution and several correlated measures of gene expression variability. Overall, these results suggest that variation in the rate of gene expression evolution, like protein sequence evolution, is mediated by variation in the level of selective constraint acting on a gene.

**Table 5.** Polymorphism and Divergence for Genes with Low and High Expression Variability in \(Solenopsis \text{invicta}\).

<table>
<thead>
<tr>
<th>Gene Expression Category</th>
<th>No. Genes</th>
<th>Direction of Selection*</th>
<th>(D_n^b)</th>
<th>(D_s^b)</th>
<th>(P_n^b)</th>
<th>(P_s^b)</th>
<th>(D_n/D_s)</th>
<th>(P_n/P_s)</th>
<th>(P\text{ value})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low variability: PC1&lt;sup&gt;4&lt;/sup&gt; lower quartile</td>
<td>90</td>
<td>–0.19</td>
<td>1,591</td>
<td>9,754</td>
<td>50</td>
<td>102</td>
<td>0.16</td>
<td>0.49</td>
<td>6.0e–09</td>
</tr>
<tr>
<td>High variability: PC1&lt;sup&gt;4&lt;/sup&gt; upper quartile</td>
<td>93</td>
<td>–0.03</td>
<td>4,102</td>
<td>9,837</td>
<td>62</td>
<td>129</td>
<td>0.42</td>
<td>0.48</td>
<td>0.366</td>
</tr>
</tbody>
</table>

*Direction of selection \((D_n/D_s + D_n) \rightarrow P_n/(P_n + P_s)\) is calculated according to Stoletzki and Eyre-Walker (2011), where a negative value indicates purifying selection and a positive value indicates positive selection.

<sup>b</sup>\(D_s\) represents the total number of synonymous fixed differences, \(P_s\) represents the total number of synonymous polymorphisms, \(D_n\) represents the total number of nonsynonymous fixed differences, and \(P_n\) represents the total number of nonsynonymous polymorphisms.

<sup>c</sup>The \(P\text{ value}\) denotes the results of the McDonald-Kreitman (1991) test according to a \(\chi^2\)-test of independence with the Williams correction for continuity.

<sup>d</sup>PC1 values are indicative of overall gene expression variability. See figure 1 and text for details.
Supplementary Material

Supplementary discussion, literature cited, figure S1, and tables S1–S4 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Kenneth G. Ross and D. DeWayne Shoemaker for sample collection, Karl M. Glastad for GO annotation, and anonymous reviewers for helpful comments. This work was supported by the US National Science Foundation grant (DEB 0640690), the Swiss National Science Foundation, and a European Research Council advanced grant.

References


Ross KG, Robertson JL. 1990. Developmental stability, heterozygosity, and fitness in two introduced fire ants (Solenopsis invicta and Solenopsis richteri) and their hybrid. Heredity 64:93–103.


**Supplementary Material**

Evolution at two levels in fire ants: the relationship between patterns of gene expression and protein sequence evolution

Brendan G. Hunt, Lino Ometto, Laurent Keller, and Michael A. D. Goodisman

**SUPPLEMENTARY DISCUSSION**

Mechanisms of gene expression evolution

Gene regulation is achieved by a complex, multi-layered regulatory landscape, which is influenced by cis-regulatory elements and chromatin-mediated genome accessibility to transcription factors (Negre et al. 2011; Lenhard et al. 2012; Wittkopp and Kalay 2012). Accordingly, the gene expression differences we observed between individuals, morphs, and species in Solenopsis could be driven, in part, by variation in cis-regulatory element sequence, transcription factor sequence, or chromatin states associated with diverse epigenetic modifications. For example, genes with low variation in expression may preferentially exhibit core promoter architecture typical of genes with high expression breadth among tissues (Lenhard et al. 2012; Munsky et al. 2012). However, gene expression variation can also arise independent of regulatory evolution, based on differences in the abiotic or biotic environment (Gibson 2008). Thus, gene expression variation may result from a diverse range of molecular mechanisms and environmental factors.

**SUPPLEMENTARY LITERATURE CITED**


Figure S1. Genes with highly variable expression patterns in *Solenopsis* exhibit higher rates of protein evolution than genes with limited expression variation. (a) A heatmap generated from gene expression variation measures determined for each morph (adult and pupal queens, workers, and males). Two primary clusters emerge, corresponding to relatively low (cluster 1) and high (cluster 2) gene expression variation in diverse contexts. The heatmap was produced by the Ward hierarchical clustering method with dendrograms proportional to distance. (b) Means and 95% confidence intervals, with genes grouped by cluster in panel a, of measures of gene expression variation used throughout the manuscript and measures of protein sequence evolution. For each measure presented in panel b, genes belonging to cluster 2 are significantly higher than those belonging to cluster 1 (*P* < 0.0001, Mann–Whitney *U* test).
**Table S1.** Spearman’s rank correlation coefficients between Dif(species) and Var(morph) of distinct *Solenopsis* morphs (red is calibrated to highest correlation coefficient, white to lowest).

<table>
<thead>
<tr>
<th>Var(X adult)</th>
<th>X</th>
<th>Dif(species X adult)</th>
<th></th>
<th>Dif(species X pupa)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Queen</td>
<td></td>
<td>Worker</td>
<td></td>
</tr>
<tr>
<td>Var(X adult)</td>
<td>Queen</td>
<td>0.26</td>
<td>0.27</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Worker</td>
<td>0.26</td>
<td>0.17</td>
<td>0.17</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.25</td>
<td>0.14</td>
<td>0.26</td>
<td>0.19</td>
</tr>
<tr>
<td>Var(X pupa)</td>
<td>Queen</td>
<td>0.20</td>
<td>0.16</td>
<td>0.15</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>Worker</td>
<td>0.24</td>
<td>0.12</td>
<td>0.18</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.15</td>
<td>0.14</td>
<td>0.13</td>
<td>0.18</td>
</tr>
</tbody>
</table>
Table S2. Spearman’s rank correlation coefficients between Dif(species) of distinct *Solenopsis* morphs and dN, dS, and dN/dS (red is calibrated to highest correlation coefficient, white to lowest).

<table>
<thead>
<tr>
<th></th>
<th>dN</th>
<th>dS</th>
<th>dN/dS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dif(species queen adult)</td>
<td>0.18</td>
<td>-0.01</td>
<td>0.20</td>
</tr>
<tr>
<td>Dif(species worker adult)</td>
<td>0.09</td>
<td>-0.03</td>
<td>0.10</td>
</tr>
<tr>
<td>Dif(species male adult)</td>
<td>0.15</td>
<td>0.01</td>
<td>0.16</td>
</tr>
<tr>
<td>Dif(species queen pupa)</td>
<td>0.08</td>
<td>-0.04</td>
<td>0.10</td>
</tr>
<tr>
<td>Dif(species worker pupa)</td>
<td>0.13</td>
<td>-0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>Dif(species male pupa)</td>
<td>0.13</td>
<td>0.02</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>dN</td>
<td>dS</td>
<td>dN/dS</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>Var(queen adult)</td>
<td>0.14</td>
<td>-0.06</td>
<td>0.16</td>
</tr>
<tr>
<td>Var(worker adult)</td>
<td>0.08</td>
<td>-0.13</td>
<td>0.14</td>
</tr>
<tr>
<td>Var(male adult)</td>
<td>0.10</td>
<td>-0.02</td>
<td>0.12</td>
</tr>
<tr>
<td>Var(queen pupa)</td>
<td>0.10</td>
<td>-0.10</td>
<td>0.14</td>
</tr>
<tr>
<td>Var(worker pupa)</td>
<td>0.10</td>
<td>-0.12</td>
<td>0.15</td>
</tr>
<tr>
<td>Var(male pupa)</td>
<td>0.18</td>
<td>-0.07</td>
<td>0.21</td>
</tr>
</tbody>
</table>
Table S4. Gene ontology annotation enrichment of genes with low expression variability\(^{a}\) relative to all genes in our analyses.

<table>
<thead>
<tr>
<th>Term</th>
<th>Category</th>
<th>Accession</th>
<th>No. genes</th>
<th>Fold enrichment</th>
<th>FDR P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>P</td>
<td>GO:0006412</td>
<td>59</td>
<td>2.79</td>
<td>0.0001</td>
</tr>
<tr>
<td>Regulation of biological process</td>
<td>P</td>
<td>GO:0050789</td>
<td>104</td>
<td>1.72</td>
<td>0.0013</td>
</tr>
<tr>
<td>Protein complex</td>
<td>C</td>
<td>GO:0043234</td>
<td>97</td>
<td>1.61</td>
<td>0.0085</td>
</tr>
<tr>
<td>Translation factor activity, nucleic acid binding</td>
<td>F</td>
<td>GO:0008135</td>
<td>23</td>
<td>3.62</td>
<td>0.0122</td>
</tr>
<tr>
<td>Ribosome</td>
<td>C</td>
<td>GO:0005840</td>
<td>33</td>
<td>2.60</td>
<td>0.0125</td>
</tr>
<tr>
<td>Nuclear lumen</td>
<td>C</td>
<td>GO:0031981</td>
<td>34</td>
<td>2.29</td>
<td>0.0219</td>
</tr>
<tr>
<td>Cellular component organization</td>
<td>P</td>
<td>GO:0016043</td>
<td>75</td>
<td>1.57</td>
<td>0.0301</td>
</tr>
</tbody>
</table>

\(^{a}\) A gene was considered to exhibit low expression variability if it exhibited a principal component 1 (PC1) value within the lower 50\(^{th}\) percentile (n = 551) of all PC1 values. PC1 values are indicative of overall gene expression variability (see main text for details).

\(^{b}\) P, biological process; F, molecular function; C, cellular component