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. Goodisman[1]

[1]School of Biology, Georgia Institute of Technology

[2]Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige (TN), Italy

[3]Department 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

Article

Fast Track

Downloaded from http://mbe.oxfordjournals.org/ at UniversitÃ© & EPFL Lausanne on January 30, 2013

© The Author 2012. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

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 $\log_2$-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.

Downloaded from http://mbe.oxfordjournals.org/ at UniversitÃ© & EPFL Lausanne on January 30, 2013

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

Abbreviation	Measure of Gene Expression	Calculation[a]
Var(morph)	Variation among samples within morphs in *S. invicta*	$\sum_{morph=1}^{n} \frac{SD\left(Y_{morph}^{invicta}\right)}{Y_{morph}^{invicta}} / n$
Dif(caste)	Differentiation between castes in *S. invicta*	$\left\lvert \log_2 \frac{X_{queen}^{invicta}}{X_{worker}^{invicta}} \right\rvert$
Dif(sex)	Differentiation between sexes in *S. invicta*	$\left\lvert \log_2 \frac{X_{queen}^{invicta}}{X_{male}^{invicta}} \right\rvert$
Dif(morph)	Differentiation among morphs in *S. invicta*	$\sum_{stage=1}^{n} \left(\text{Dif(caste)}_{stage}^{invicta} + \text{Dif(sex)}_{stage}^{invicta}\right) / n$
Dif(species)	Differentiation between *S. invicta* and *S. richteri*	$\sum_{morph=1}^{n} \left\lvert \log_2 \frac{X_{morph}^{invicta}}{X_{morph}^{richteri}} \right\rvert / n$
Dif(hybrid)	Minimum differentiation between hybrids and *S. invicta* or *S. richteri*	$\min\left(\sum_{morph=1}^{n} \left\lvert \log_2 \frac{X_{morph}^{hybrid}}{X_{morph}^{invicta}} \right\rvert / n, \sum_{morph=1}^{n} \left\lvert \log_2 \frac{X_{morph}^{hybrid}}{X_{morph}^{richteri}} \right\rvert / n \right)$

[a]*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.

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 $\log_2$-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 $\log_2$-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 (d*S*) 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 d*S* 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

Downloaded from http://mbe.oxfordjournals.org/ at UniversitÃ© & EPFL Lausanne on January 30, 2013

Table 2. Spearman's Rank Correlations (ρ) between Measures of Protein Coding Sequence Evolution (d*N*/d*S*, d*N*, and d*S*) or Gene Expression Level, and Gene Expression Variability within *Solenopsis invicta* [Var(morph)] or Gene Expression Differentiation between *S. invicta* and *S. richteri* [Dif(species)].

X	$\rho_{X,\ \mathrm{Var(morph)}}$	$\rho_{X,\ \mathrm{Dif(species)}}$
dN/dS	**0.228******	**0.257******
dN	**0.168******	**0.235******
dS	**−0.130******	**-0.032^{NS}**
Gene expression level	**-0.027^{NS}**	**-0.049^{NS}**

$^{NS}P > 0.05$, ****$P < 0.0001$.

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 Dif(species) remained positively correlated with d*N*/d*S* with a Spearman's rank correlation of 0.226 ($P < 0.0001$), suggesting that the correlation between Dif(species) and d*N*/d*S* 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 (d*N* and d*S*, 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 d*N*/d*S* as a measure of selective constraint on protein sequence, with larger values of d*N*/d*S* generally corresponding to weaker selective constraint (Hunt et al. 2011). Substitution rates were averaged across all aligned codons for a given protein, with free d*N*/d*S* 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 d*N*/d*S* 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 Dif(species), were indeed highly correlated (Spearman's $\rho = 0.523$, $P < 0.0001$; 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 Dif(hybrid), was also highly correlated with Var(morph) and Dif(species) (table 3). The positive correlation between Dif(species) and Dif(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 [Dif(caste)] and sex bias [Dif(sex)] in both the adult and pupal stages were each positively correlated with Var(morph) (table 4). Furthermore, overall morph bias [Dif(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

Downloaded from http://mbe.oxfordjournals.org/ at UniversitÃ© & EPFL Lausanne on January 30, 2013

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 (d*N*, d*S*, and d*N*/d*S*).

Dif(hybrid) (*X*)	Genes (*n*)	$\rho_{X,\mathrm{Dif(species)}}$	$\rho_{X,\mathrm{Var(morph)}}$	$\rho_{X,dN}$	$\rho_{X,dS}$	$\rho_{X,dN/dS}$
Over-expressed	222	0.727****	0.791****	0.266****	-0.074^{NS}	0.336****
Intermediately expressed	512	0.715****	0.602****	0.161***	-0.074^{NS}	0.187****
Under-expressed	367	0.589****	0.604****	0.143**	-0.047^{NS}	0.165**
All genes	1,101	0.660****	0.651****	0.184****	−0.070*	0.218****

$^{NS}P > 0.05$, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

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 (d*N*, d*S*, and d*N*/d*S*).

Expression Bias (*X*)	$\rho_{X,\mathrm{Var(morph)}}$	$\rho_{X,\mathrm{Dif(species)}}$	$\rho_{X,dN}$	$\rho_{X,dS}$	$\rho_{X,dN/dS}$
Dif(caste adult)	0.280****	0.210****	0.138****	0.016^{NS}	0.143****
Dif(caste pupa)	0.307****	0.259****	0.065*	−0.085**	0.103**
Dif(sex adult)	0.330****	0.366****	0.157****	-0.038^{NS}	0.175****
Dif(sex pupa)	0.233****	0.228****	0.137****	-0.007^{NS}	0.143****
Dif(morph)[a]	0.473****	0.442****	0.213****	-0.046^{NS}	0.239****

$^{NS}P > 0.05$, $*P < 0.05$, $**P < 0.01$, $****P < 0.0001$.
[a]Mean 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; Leichty 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), Dif(hybrid), 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. 1*a*). 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. 1*a* 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 (d*N*) and selective constraint (d*N*/d*S*) 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 d*N* 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 d*N* and d*N*/d*S* (Artieri et al. 2007).

Downloaded from http://mbe.oxfordjournals.org/ at UniversitÃ© & EPFL Lausanne on January 30, 2013

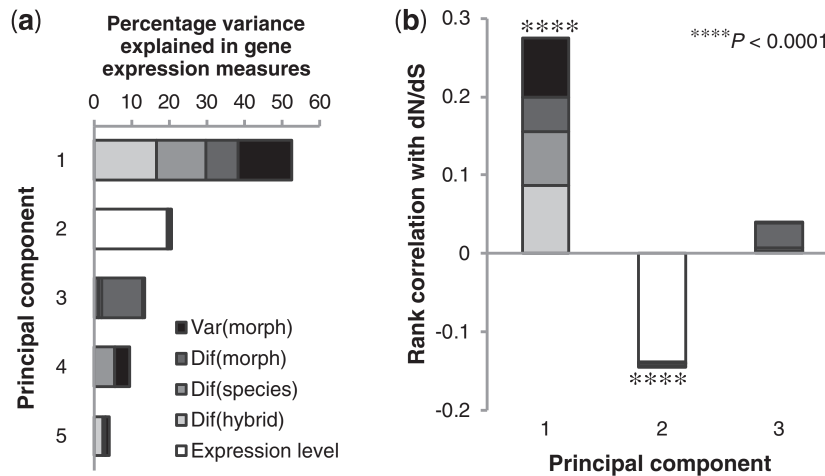

Fig. 1. Principal component analysis of *Solenopsis* gene expression measures. (*a*) Percentage variance explained by five PCs generated from gene expression variation within morphs [Var(morph)], morph bias in gene expression [Dif(morph)], gene expression divergence between species [Dif(species)], gene 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*).

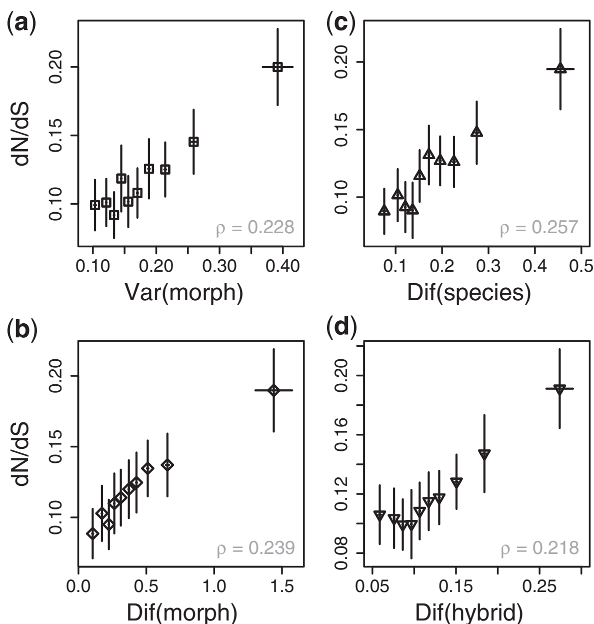

Fig. 2. Protein sequence evolution is correlated with gene expression variation and divergence. dN/dS is associated with (*a*) gene expression variability within *S. invicta* [Var(morph)], (*b*) gene expression differentiation among morphs in *S. invicta* [Dif(morph)], (*c*) gene expression differentiation 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 d*N* and d*N*/d*S*. We found that within-morph variability in gene expression [Var(morph)] was positively correlated with d*N* and d*N*/d*S* (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 d*N* and d*N*/d*S* (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

Downloaded from http://mbe.oxfordjournals.org/ at UniversitÃ© & EPFL Lausanne on January 30, 2013

expression level, intraspecific variation in expression [Var(morph)], morph bias in expression [Dif(morph)], interspecific divergence in expression [Dif(species)], and expression divergence between hybrids and parental species [Dif(hybrid)]. This analysis revealed that a single PC (PC1, fig. 1), which reflected overall variation in gene expression, was significantly positively correlated with d*N*/d*S* (fig. 1*b*). Furthermore, gene expression level loaded primarily on PC2, suggesting that the association between PC1 and d*N*/d*S* was independent from the well-described association between gene expression level and protein evolutionary rate (Drummond and Wilke 2008) (fig. 1*b*). Thus, overall gene expression variability is linked to the rate of protein evolution (supplementary fig. S1, Supplementary Material online).

To further explore the nature of selection shaping variation in protein evolution in our data, we compared sequence polymorphism within *S. invicta* to sequence divergence between the *S. invicta* lineage and the outgroup taxa *P. barbatus* and *L. humile* (Hunt et al. 2011). We detected strongly significant ($P = 6.0 \times 10^{-9}$) 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 ($P = 0.366$; 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 (Sella 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 (Munsky 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. barbatus* and *S. invicta*, which diverged from a common ancestor approximately 110 Ma (Moreau et al. 2006). In contrast, gene expression divergence was estimated between *S. invicta* and *S. richteri*, which are sister species within the *S. 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.

Downloaded from http://mbe.oxfordjournals.org/ at UniversitÃ© & EPFL Lausanne on January 30, 2013

Table 5. Polymorphism and Divergence for Genes with Low and High Expression Variability in *Solenopsis invicta*.

Gene Expression Category	No. Genes	Direction of Selection[a]	D_n[b]	D_s[b]	P_n[b]	P_s[b]	D_n/D_s	P_n/P_s	P value[c]
Low variability: PC1[d] lower quartile	90	−0.19	1,591	9,754	50	102	0.16	0.49	6.0e−09
High variability: PC1[d] upper quartile	93	−0.03	4,102	9,837	62	129	0.42	0.48	0.366

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

[b]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.

[c]The *P* value denotes the results of the McDonald-Kreitman (1991) test according to a *G*-test of independence with the Williams correction for continuity.

[d]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

Artieri CG, Haerty W, Singh RS. 2007. Association between levels of coding sequence divergence and gene misregulation in *Drosophila* male hybrids. *J Mol Evol.* 65:697–704.

Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biol.* 3: 1159–1166.

Castillo-Davis CI, Hartl DL, Achaz G. 2004. cis-Regulatory and protein evolution in orthologous and duplicate genes. *Genome Res.* 14: 1530–1536.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization, and analysis in functional genomics research. *Bioinformatics* 21:3674–3676.

Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134:341–352.

Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol.* 17:68–74.

Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. 2004. Noise minimization in eukaryotic gene expression. *PLoS Biol.* 2:e137.

Gadau J, Helmkampf M, Nygaard S, Roux J, Simola DF, Smith CR, Suen G, Wurm Y, Smith CD. 2012. The genomic impact of 100 million years of social evolution in seven ant species. *Trends Genet.* 28:14–21.

Gnad F, Parsch J. 2006. Sebida: a database for the functional and evolutionary analysis of genes with sex-biased expression. *Bioinformatics* 22:2577–2579.

Goodisman MAD, Kovacs JL, Hunt BG. 2008. Functional genetics and genomics in ants (Hymenoptera: Formicidae): the interplay of genes and social life. *Myrmecol News.* 11:107–117.

Goodisman MAD, Shoemaker DD, Asmussen MA. 1998. Cytonuclear theory for haplodiploid species and X-linked genes. II. Stepping-stone models of gene flow and application to a fire ant hybrid zone. *Evolution* 52:1423–1440.

Grozinger CM, Fan YL, Hoover SER, Winston ML. 2007. Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol Ecol.* 16:4837–4848.

Haerty W, Singh RS. 2006. Gene regulation divergence is a major contributor to the evolution of Dobzhansky-Muller incompatibilities between species of *Drosophila*. *Mol Biol Evol.* 23:1707–1714.

Heimpel GE, de Boer JG. 2008. Sex determination in the Hymenoptera. *Annu Rev Entomol.* 53:209–230.

Hirsh AE, Fraser HB. 2001. Protein dispensability and rate of evolution. *Nature* 411:1046–1049.

Hunt BG, Goodisman MAD. 2010. Evolutionary variation in gene expression is associated with dimorphism in eusocial vespid wasps. *Insect Mol Biol.* 19:641–652.

Hunt BG, Ometto L, Wurm Y, Shoemaker D, Yi SV, Keller L, Goodisman MAD. 2011. Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proc Natl Acad Sci U S A.* 108: 15936–15941.

Hunt BG, Wyder S, Elango N, Werren JH, Zdobnov EM, Yi SV, Goodisman MAD. 2010. Sociality is linked to rates of protein evolution in a highly social insect. *Mol Biol Evol.* 27:497–500.

Jordan IK, Marino-Ramirez L, Koonin EV. 2005. Evolutionary significance of gene expression divergence. *Gene* 345:119–126.

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Pääbo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* 309:1850–1854.

Leichty AR, Pfennig DW, Jones CD, Pfennig KS. 2012. Relaxed genetic constraint is ancestral to the evolution of phenotypic plasticity. *Integr Compar Biol.* 52:16–30.

Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. 2005. Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. *Mol Biol Evol.* 22:1345–1354.

Lindblad-Toh K, Garber M, Zuk O, et al. (85 co-authors). 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478:476–482.

Mank JE, Ellegren H. 2009. Are sex-biased genes more dispensable? *Biol Lett.* 5:409–412.

Mank JE, Hultin-Rosenberg L, Webster MT, Ellegren H. 2008. The unique genomic properties of sex-biased genes: insights from avian microarray data. *BMC Genomics* 9:14.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. *Proc Natl Acad Sci U S A.* 100:9894–9899.

Meiklejohn CD, Townsend JP. 2005. A Bayesian method for analysing spotted microarray data. *Brief Bioinform.* 6:318–330.

Meisel RP. 2011. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. *Mol Biol Evol.* 28:1893–1900.

Moreau CS, Bell CD, Vila R, Archibald SB, Pierce NE. 2006. Phylogeny of the ants: diversification in the age of angiosperms. *Science* 312: 101–104.

Munsky B, Neuert G, van Oudenaarden A. 2012. Using gene expression noise to understand gene regulation. *Science* 336:183–187.

Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM. 2004. Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol Biol Evol.* 21:1308–1317.

Ometto L, Ross KG, Shoemaker D, Keller L. 2012. Disruption of gene expression in hybrids of the fire ants *Solenopsis invicta* and *S. richteri*. *Mol Ecol.* 21:2488–2501.

Ometto L, Shoemaker D, Ross KG, Keller L. 2011. Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. *Mol Biol Evol.* 28:1381–1392.

Pitts JP, McHugh JV, Ross KG. 2005. Cladistic analysis of the fire ants of the *Solenopsis saevissima* species-group (Hymenoptera: Formicidae). *Zool Scripta.* 34:493–505.

Downloaded from http://mbe.oxfordjournals.org/ at UniversitÃ© & EPFL Lausanne on January 30, 2013

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* 300:1742–1745.

Ranz JM, Namgyal K, Gibson G, Hartl DL. 2004. Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. *Genome Res.* 14:373–379.

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.

Schwander T, Lo N, Beekman M, Oldroyd BP, Keller L. 2010. Nature versus nurture in social insect caste differentiation. *Trends Ecol Evol.* 25:275–282.

Sella G, Petrov DA, Przeworski M, Andolfatto P. 2009. Pervasive natural selection in the *Drosophila* genome? *PLoS Genet.* 5:e1000495.

Shoemaker DD, Ross KG, Arnold ML. 1996. Genetic structure and evolution of a fire ant hybrid zone. *Evolution* 50:1958–1976.

Smith CD, Zimin A, Holt C, et al. (50 co-authors). 2011. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proc Natl Acad Sci U S A.* 108:5673–5678.

Smith CR, Smith CD, Robertson HM, et al. (45 co-authors). 2011. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc Natl Acad Sci U S A.* 108:5667–5672.

Smith CR, Toth AL, Suarez AV, Robinson GE. 2008. Genetic and genomic analyses of the division of labour in insect societies. *Nat Rev Genet.* 9: 735–748.

Snell-Rood EC, Cash A, Han MV, Kijimoto T, Andrews J, Moczek AP. 2011. Developmental decoupling of alternative phenotypes: insights from the transcriptomes of horn-polyphenic beetles. *Evolution* 65: 231–245.

Snell-Rood EC, Van Dyken JD, Cruickshank T, Wade MJ, Moczek AP. 2010. Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity. *Bioessays* 32:71–81.

Stoletzki N, Eyre-Walker A. 2011. Estimation of the neutrality index. *Mol Biol Evol.* 28:63–70.

Townsend J, Hartl D. 2002. Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments. *Genome Biol.* 3:research0071–research0071.16.

Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G, Eisen MB, Feldman MW. 2005. Functional genomic analysis of the rates of protein evolution. *Proc Natl Acad Sci U S A.* 102:5483–5488.

Wang J, Jemielity S, Uva P, Wurm Y, Graff J, Keller L. 2007. An annotated cDNA library and microarray for large-scale gene-expression studies in the ant *Solenopsis invicta*. *Genome Biol.* 8:R9.

Waterhouse RM, Zdobnov EM, Kriventseva EV. 2011. Correlating traits of gene retention, sequence divergence, duplicability, and essentiality in vertebrates, arthropods, and fungi. *Genome Biol Evol.* 3:75–86.

Whitehead A, Crawford DL. 2006. Variation within and among species in gene expression: raw material for evolution. *Mol Ecol.* 15:1197–1211.

Whitfield CW, Cziko AM, Robinson GE. 2003. Gene expression profiles in the brain predict behavior in individual honey bees. *Science* 302: 296–299.

Wilson EO. 1971. The insect societies. Cambridge (MA): Harvard University Press.

Wurm Y, Wang J, Riba-Grognuz O, et al. (38 co-authors). 2011. The genome of the fire ant *Solenopsis invicta*. *Proc Natl Acad Sci U S A.* 108:5679–5684.

Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. 2007. Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature* 450:233–238.

Downloaded from http://mbe.oxfordjournals.org/ at UniversitÃ© & EPFL Lausanne on January 30, 2013