Large-Scale Coding Sequence Change Underlies the Evolution of Postdevelopmental Novelty in Honey Bees

William Cameron Jasper,[1] Timothy A. Linksvayer,[2] Joel Atallah,[3] Daniel Friedman,[3] Joanna C. Chiu,[1] and Brian R. Johnson*[,1]

[1]Department of Entomology, University of California—Davis

[2]Department of Biology, University of Pennsylvania

[3]Department of Evolution and Ecology, University of California—Davis

***Corresponding author:** E-mail: brnjohnson@ucdavis.edu.

Associate editor: John H. McDonald

Abstract

Whether coding or regulatory sequence change is more important to the evolution of phenotypic novelty is one of biology's major unresolved questions. The field of evo–devo has shown that in early development changes to regulatory regions are the dominant mode of genetic change, but whether this extends to the evolution of novel phenotypes in the adult organism is unclear. Here, we conduct ten RNA-Seq experiments across both novel and conserved tissues in the honey bee to determine to what extent postdevelopmental novelty is based on changes to the coding regions of genes. We make several discoveries. First, we show that with respect to novel physiological functions in the adult animal, positively selected tissue-specific genes of high expression underlie novelty by conferring specialized cellular functions. Such genes are often, but not always taxonomically restricted genes (TRGs). We further show that positively selected genes, whether TRGs or conserved genes, are the least connected genes within gene expression networks. Overall, this work suggests that the evo–devo paradigm is limited, and that the evolution of novelty, postdevelopment, follows additional rules. Specifically, evo–devo stresses that high network connectedness (repeated use of the same gene in many contexts) constrains coding sequence change as it would lead to negative pleiotropic effects. Here, we show that in the adult animal, the converse is true: Genes with low network connectedness (TRGs and tissue-specific conserved genes) underlie novel phenotypes by rapidly changing coding sequence to perform new-specialized functions.

Key words: **novel traits, taxonomically restricted genes, evolution of novelty, honey, bees, RNA-Seq.**

Article

Introduction

The genetic basis of phenotypic novelty is a major unresolved question in evolutionary biology (Beldade and Brakefield 2002; Wray et al. 2003; Hahn et al. 2007; Arendt and Reznick 2008; Conant and Wolfe 2008; Stern and Orgogozo 2008). Understanding the processes that give rise to major phenotypic shifts has also generated considerable disagreement (Hoekstra and Coyne 2007; Mitchell-Olds et al. 2007; Halligan et al. 2013; Parker et al. 2014). Early work focused on the role played by coding sequence change underlying traits controlled by few genes (Daltry et al. 1996; Fry et al. 2003; Hoekstra 2006; Nadeau and Jiggins 2010), whereas later work focused on the role played by gene duplication (Lynch and Conery 2000; Lynch and Conery 2003; Conant and Wolfe 2008). The field of Evolutionary Developmental Biology (evo–devo) expanded our understanding of morphological innovation by highlighting the role played by regulatory shifts controlling the novel use of conserved genes during early development (Carroll 1995; Sucena and Stern 2000; Arthur 2002; Beldade and Brakefield 2002; Beldade et al. 2005; Toth and Robinson 2007; Wagner et al. 2007). More recently, work on taxonomically restricted genes (TRGs) has shifted attention to the role played by novel genes (Wilson et al. 2005, 2007; Khalturin et al. 2009; Toll-Riera et al. 2009; Johnson and Tsutsui 2011; Tautz and Domazet-Loso 2011; Ranz and Parsch 2012; Reinhardt et al. 2013; Shigenobu and Stern 2013; Wissler et al. 2013; Sumner 2014; Zhao et al. 2014). In general, a confounding issue is that multiple mechanisms can underlie phenotypic novelty in a nonmutually exclusive manner. A pressing concern is therefore to determine the respective genetic mechanisms that govern different evolutionary contexts of novelty.

The goal of the present study is to use the novel biology of honey bees to explore the role played by coding sequence change in the evolution of phenotypic novelty. Honey bees are an ideal model system for this question because they have evolved many novel traits for social functions missing from their solitary ancestors (Johnson and Linksvayer 2010). The mandibular and Nasonov glands, for example, make pheromones important for social communication, whereas the hypopharyngeal glands (HPG) produce brood food for young bees (Ueno et al. 2009; Wegener et al. 2009; Johnson 2010). These glands are either altogether missing in solitary bees or serve completely different purposes. Further, the honey bee has radically changed behavior relative to its use of some conserved structures. The sting gland produces

venom which is specialized for defense against vertebrates (Owen and Pfaff 1995; King and Spangfort 2000). Solitary bees, in contrast, defend against invertebrates. Honey bees also have highly derived social behaviors, exemplified by the waggle dance, which are encoded within the brain, and possibly the segmental ganglia (Brockmann and Robinson 2007). In sum, there are a variety of tissues in the honey bee that can be used to explore the genetic and evolutionary history of novelty. Patterns of gene expression found in these novel tissues can be contrasted with patterns found in conserved tissues to demonstrate that certain mechanisms (e.g., expression of particular classes of genes) are limited to tissues involved in novel functions.

A second feature of honey bee biology making it an ideal model for the study of phenotypic novelty relates to its system of division of labor, which makes identifying genes important for social behavior quite straightforward. Honey bees pass through several developmental phases as adults (each called a temporal caste) in which they specialize (both behaviorally and physiologically) for particular roles in the nest. These include cell builders, nurses, food processors/nest builders, and foragers (Seeley 1982; Johnson 2008; Johnson 2010). Here, we focus on nurses and foragers as they are the most amenable to study (Page and Robinson 1991; Zayed and Robinson 2012). Nurses and foragers have unique hormonal titers and it is known that the differences in behavior and physiology between them are the result of differences in the expression of many genes (Whitfield et al. 2003; Smith et al. 2008). We focus on tissues that are all either known to change in function when a bee transitions from nursing to foraging, or can be predicted to change in function based on the strong nutritional and activity level changes that occur at this transition. Supplementary table S1, Supplementary Material online, shows the changes in function that occur in each tissue explored in this study. The novel tissues have already been discussed, but conserved tissues are used in differential social ways as well. The flight muscles are little used in nurses, but are used to an extent unique in the animal kingdom in the foragers (who have the highest metabolic rates recorded for any organism) (Williams et al. 2008). The midgut also changes in function between nurses, who consume pollen to make brood food, and foragers who rely on nurse secretions as a protein source and do not digest pollen (Crailsheim 1991; Johnson 2010; Peters et al. 2010).

In this study, we focus on the evolution of novel physiological functions in the adult organism, as evo–devo has primarily focused on morphological innovations generated during early development. Based on previous work, we suspected that TRGs may be critically important for generating novel social functions (Johnson and Tsutsui 2011). A recent study also found that TRGs are important for fitness, as they are often positively selected (Harpur et al. 2014). We began by identifying all contexts in which TRGs potentially play major roles. We define "major role" as a gene being in the top 1% of all genes in terms of expression, or being significantly differentially expressed between life-history stages or tissues. In general, by focusing on TRGs, the most radical cases of coding sequence change, we expect a conservative estimate of the importance of coding sequence change in the evolution of novelty. However, we expect these results on TRGs to be informative for the general roles played by coding sequence change, as we hypothesize the TRG patterns to simply be the most extreme cases of broadly important evolutionary processes involving coding regions. Next, we test two hypotheses we generate based on our study of TRGs. First, we test the hypothesis that physiological novelty is based primarily on rapid coding sequence change in key genes with specialized (tissue specific) roles. Second, we predict and show such genes to be extremely highly expressed and to have low gene network connectivity. In essence, we test the hypothesis that genes central to the evolution of novel adult physiological phenotypes are distal branches of gene networks rather than hubs and are hence free to evolve coding sequence changes as needed without incurring negative pleiotropic effects.

Results

Highly Expressed TRGs

To avoid analyzing TRGs that are spurious ORFs (Domazet-Loso and Tautz 2003; Chen et al. 2013), or simply genes of minor importance, we began by identifying and analyzing all those TRGs that are highly expressed in at least one of the ten tissues. Our hypothesis is that if TRGs are critical for novel functions, then highly expressed TRGs (HE-TRGs) should play a stronger role in tissues with novel functions relative to those with conserved functions. Figure 1 shows that this is the case. For each tissue, we plot the percentage of the top 1% of highly expressed genes (HEGs) that are conserved genes versus TRGs of various levels of restriction. We then plot the percentage of total expression in the top 1% that stems from each category. The categories of TRGs are Orphans, which are those found only in honey bee genome; arthropod, insect, hymenoptera, and bee, which are those found only in genomes of the taxonomic group, respectively; other TRG (abbreviated as O-TRG), which are those showing a complex presence/absence pattern; and conserved (see more details in Materials and Methods). For simple tissues with specialized functions such as the HPG and sting gland, the percentage of total expression stemming from TRGs was very high ($>90\%$) and was significantly biased with respect to the percentage of HEGs that fall into different TRG categories (chi square test: Sting gland up-regulated in nurses: $\chi^2 = 234.73$, $df = 1$, $P < 0.0001$, sting gland up-regulated in foragers: $\chi^2 = 50.47$, $df = 1$, $P < 0.0001$, HPG up-regulated in nurses: $\chi^2 = 175.0$, $df = 2$, $P < 0.0001$, HPG up-regulated in foragers: $\chi^2 = 75.66$, $df = 1$, $P < 0.0001$). For more complex tissues with novel functions such as the antenna, the Nasonov gland, and the mandibular gland, the same pattern holds (a statistically significant bias exists between percentage of HEGs that are TRGs and the percentage of total expression stemming from TRGs), but the bias is not so great and one out of six tests is only marginally significant (chi square test: Antenna up-regulated in nurses: $\chi^2 = 28.62$, $df = 2$, $P < 0.0001$, antenna up-regulated in foragers: $\chi^2 = 32.20$, $df = 2$, $P < 0.0001$; Nasonov gland up-regulated in

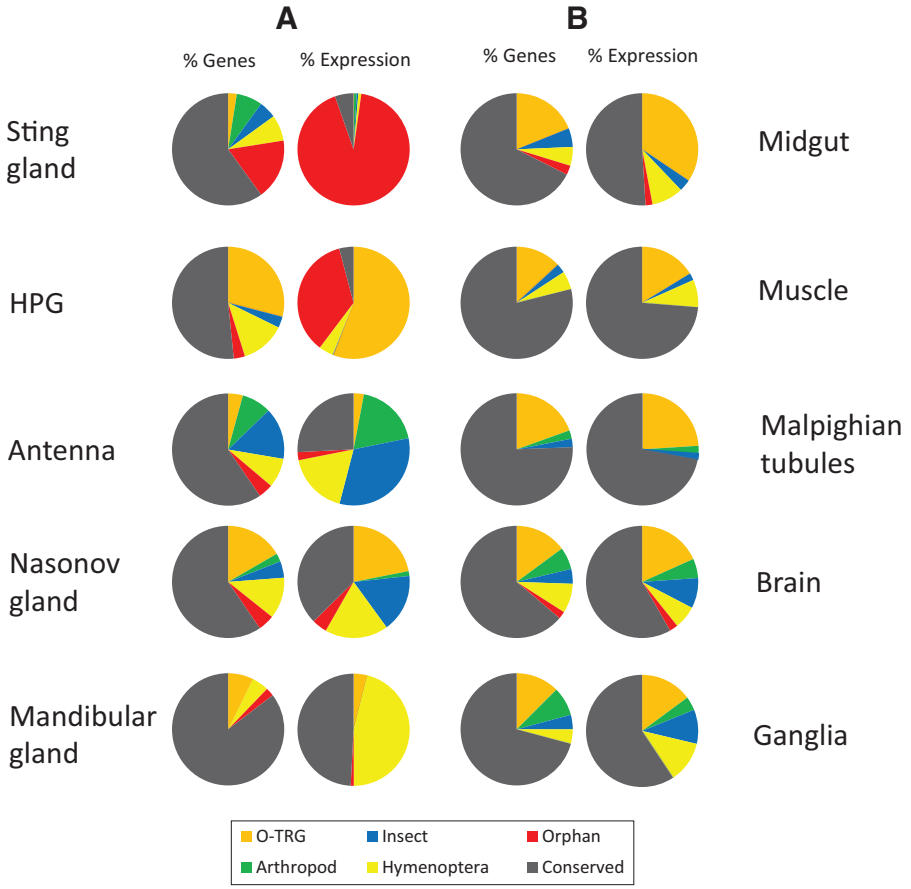

Fig. 1. Frequency of TRGs among the highest expressed genes in each tissue along with the total percentage of transcripts in the transcriptome stemming from genes in each TRG class. Although TRGs make up a minority of genes in the top 1% of expression in all tissues, in tissues with novel functions they account for a majority of expression. Tissues in the A column are those for which a statistically significant bias in expression of TRGs exists, whereas tissues in the B column do not show bias in expression with TRG status. Orphans are genes found only in *Apis mellifera*. Bee, insect, hymenoptera, and arthropod refer to TRG classes in which a gene is found only within this taxonomic group, whereas the O-TRG category refers to genes with a complex but highly restricted pattern of presence and absence in different clades. Conserved genes are genes not in any of the TRG categories (hence, widely found across organisms).

nurses: $\chi^2 = 8.96$, $df = 2$, $P = 0.01$, Nasonov gland up-regulated in foragers: $\chi^2 = 6.17$, $df = 2$, $P = 0.05$, mandibular gland up-regulated in nurses: $\chi^2 = 21.48$, $df = 1$, $P < 0.0001$, mandibular gland up-regulated in foragers: $\chi^2 = 2.88$, $df = 1$, $P = 0.091$). For highly conserved tissues (muscle, midgut, and malpighian tubules), in contrast, the percentage of transcripts from TRGs in the transcriptome equals the percentage of TRGs amongst HEGs (thoracic muscle up-regulated in nurses: $\chi^2 = 0.53$, $df = 1$, $P = 0.47$, thoracic muscle up-regulated in foragers: $\chi^2 = 0.09$, $df = 1$, $P = 0.76$, malpighian tubules up-regulated in nurses: $\chi^2 = 0.25$, $df = 1$, $P = 0.62$, malpighian tubules up-regulated in foragers: $\chi^2 = 0.06$, $df = 1$, $P = 0.81$, midgut up-regulated in nurses: $\chi^2 = 4.61$, $df = 2$, $P < 0.10$, midgut up-regulated in foragers: $\chi^2 = 11.48$, $df = 2$, $P < 0.0001$). Hence, TRG expression is not biased with respect to frequency in conserved tissues and TRGs do not appear to be disproportionately important to the function of conserved tissues (with one exception: The midgut in foragers in which there is a bias). Nervous tissue (brain and segmental ganglia) showed patterns similar to other conserved tissues, in that TRGs played a small role and were not biased in expression (chi square test: Brain up-regulated in nurses: $\chi^2 = 1.16$, $df = 3$, $P = 0.77$, brain up-regulated in foragers: $\chi^2 = 0.86$, $df = 2$, $P = 0.65$, ganglion up-regulated in nurses: $\chi^2 = 2.80$, $df = 2$, $P = 0.25$, ganglion up-regulated in foragers: $\chi^2 = 2.48$, $df = 2$, $P = 0.29$). Complete statistics are in supplementary table S2, Supplementary Material online and mean expression levels (mean RPKM values) for all expressed genes for each tissue are in supplementary table S3, Supplementary Material online.

DE-TRGs: Adult Social Function

To identify genes involved in adult social functions, we identified genes differentially expressed between nurses and foragers (for each tissue). A large body of work has used this approach to identify genes underlying social behavior in the brain (Robinson et al. 2005; Zayed and Robinson 2012), and we extend this approach to the whole body. Table 1 shows that an analysis of differentially expressed TRGs (DE-TRGs) underlying social functions leads to similar, though distinct, conclusions as found when examining HE-TRGs (differentially expressed genes [DEGs] in

Table 1. Bias in the Expression of DE-TRGs Relative to Conserved Genes.

	↑ Nurse		↑ Forager		↑ Nurse		↑ Forager	
	% Genes	% Exp.	% Genes	% Exp	% Genes	% Exp.	% Genes	% Exp.
	Sting Gland				Thoracic Muscle			
Conserved	60.3	4.7	67.4	29.4	64.3	61.3	80.8	78.5
O-TRG	13.4	0.1	10.1	3.4	11.4	19.4	7.8	19.2
Arthropod	4.3	0.0	3.5	24.7	3.8	1.1	1.1	0.1
Insect	7.3	0.2	11.0	6.6	11.7	5.6	4.8	1.3
Hymenoptera	6.0	0.1	4.4	33.7	4.8	12.0	4.4	0.6
Bee	0.4	0.0	0.9	0.0	0.5	0.0	0.2	0.1
Orphan	8.2	94.9	2.6	2.2	3.6	0.7	0.9	0.1
	HPG				Malpighian Tubules			
Conserved	82.0	9.4	70.2	10.3	64.9	83.6	68.3	77.7
O-TRG	9.2	89.5	11.4	0.2	11.9	12.4	14.6	4.0
Arthropod	1.2	0.0	4.3	0.0	4.4	1.1	4.9	13.5
Insect	3.1	0.1	6.1	0.7	7.5	1.1	6.8	1.8
Hymenoptera	2.7	0.9	4.5	2.5	6.9	1.3	2.9	2.8
Bee	0.6	0.0	0.5	0.0	1.4	0.0	0.3	0.0
Orphan	1.3	0.1	2.9	86.4	3.0	0.5	2.3	0.2
	Nasonov Gland				Midgut			
Conserved	63.5	37.5	61.5	38.6	76.8	65.4	60.5	65.7
O-TRG	12.8	20.1	13.0	19.2	10.3	27.8	14.0	17.5
Arthropod	5.3	1.7	4.0	2.3	4.3	0.1	4.7	0.4
Insect	7.5	23.1	9.8	9.3	4.3	6.0	7.8	5.7
Hymenoptera	5.3	11.9	8.2	20.2	3.2	0.7	8.5	10.5
Bee	0.8	0.0	0.4	0.0	0.0	0.0	2.3	0.1
Orphan	4.8	5.6	3.0	10.4	1.1	0.0	2.3	0.1
	Mandibular Gland				Segmental ganglion			
Conserved	79.8	56.8	61.3	65.1	52.5	44.7	69.0	54.2
O-TRG	8.6	3.6	9.7	15.3	12.3	13.8	12.6	40.1
Arthropod	2.3	0.9	3.8	6.5	12.3	25.9	2.3	1.3
Insect	4.4	0.5	8.6	3.3	11.5	12.9	9.2	3.2
Hymenoptera	2.9	37.0	5.8	11.1	4.9	2.6	2.3	0.2
Bee	0.1	0.0	0.2	0.0	1.6	0.0	0.0	0.0
Orphan	1.8	1.3	2.8	0.3	4.9	0.0	4.6	1.0

NOTE.—↑, refers to genes up-regulated in either nurses or foragers; %, genes refers to the percentage of DEGs in each taxonomic category, whereas % exp. refers to the total expression of genes in each taxonomic category.

supplementary table S4, Supplementary Material online). For novel tissues, expression is strongly biased toward DE-TRGs (chi square test: Sting gland up-regulated in nurses: $\chi^2 = 6{,}576.47$, $df = 2$, $P < 0.0001$, sting gland up-regulated in foragers: $\chi^2 = 247.84$, $df = 5$, $P < 0.0001$, HPG up-regulated in nurses: $\chi^2 = 3{,}564.32$, $df = 2$, $P < 0.0001$, HPG up-regulated in foragers: $\chi^2 = 3{,}423.57$, $df = 2$, $P < 0.0001$, Nasonov gland up-regulated in nurses: $\chi^2 = 112.17$, $df = 4$, $P < 0.0001$, Nasonov gland up-regulated in foragers: $\chi^2 = 143.26$, $df = 5$, $P < 0.0001$, mandibular gland up-regulated in nurses: $\chi^2 = 646.48$, $df = 5$, $P < 0.0001$, mandibular gland up-regulated in foragers: $\chi^2 = 347.61$, $df = 5$, $P < 0.0001$). As for HEGs, TRGs that are differentially expressed represent a much higher fraction of expression than would be expected based on their frequency. Essentially, the same TRGs found to be important in the analysis of HEGs are identified when searching for DE-TRGs in these tissues. These genes include those that encode venoms in the sting gland, royal jelly proteins in the HPG, and cuticle proteins in several tissues, to name a few characterized examples.

For conserved tissues, most of the comparisons of gene frequency to expression frequency also showed significant bias in the contribution of different taxonomic classes of genes to total expression (thoracic muscle up-regulated in nurses: $\chi^2 = 156.67$, $df = 5$, $P < 0.0001$, thoracic muscle up-regulated in foragers: $\chi^2 = 500.14$, $df = 5$, $P < 0.0001$, malpighian tubules up-regulated in nurses: $\chi^2 = 371.61$, $df = 5$, $P < 0.0001$, malpighian tubules up-regulated in foragers: $\chi^2 = 155.22$, $df = 4$, $P < 0.0001$, midgut up-regulated in nurses: $\chi^2 = 34.05$, $df = 2$, $P < 0.0001$, midgut up-regulated in foragers: $\chi^2 = 10.65$,

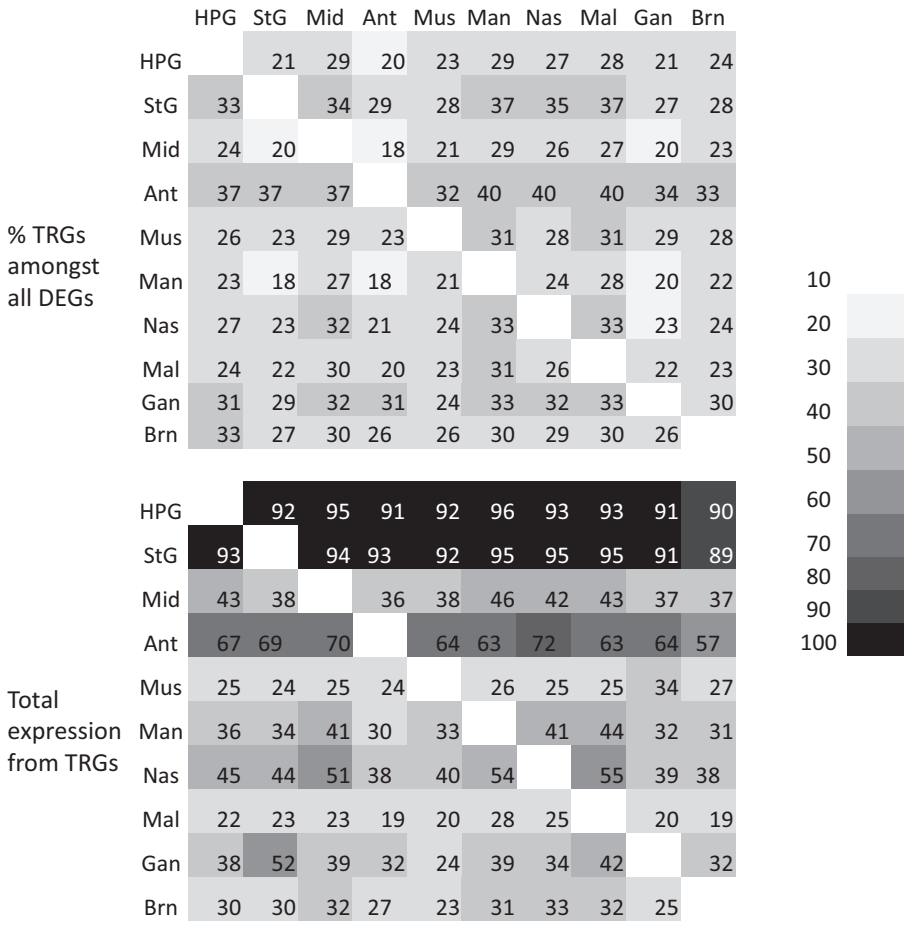

Fig. 2. Analysis of genes that show differential expression between tissues in nurse bees. The top panel shows the combined percentage of DEGs in all TRG categories (orphan, bee-specific, hymenopteran-specific, insect-specific, arthropod-specific, and O-TRGs) in each tissue by tissue comparison, whereas the bottom panel shows the total percentage of expression stemming from all TRGs in the same comparisons. Each row in both panels shows the results for the genes that are up-regulated in the tissue on the vertical column relative to that in the horizontal row. StG, sting gland; Mid, midgut; Ant, antenna; Mus, thoracic muscle; Man, mandibular gland; Nas, nasonov gland; Mal, malpighian tubules; Gan, second thoracic ganglia. Brn, brain.

$df = 3$, $P = 0.01$, ganglion up-regulated in nurses: $\chi^2 = 5.85$, $df = 4$, $P = 0.21$, ganglion up-regulated in foragers: $\chi^2 = 57.60$, $df = 3$, $P < 0.0001$). Hence, DE-TRGs in some cases also contribute to social functions in more conserved tissues. However, the results are nevertheless distinct from those with the more derived glandular tissues. Supplementary figure S1, Supplementary Material online, shows the results just for conserved genes in terms of their frequency amongst all DEGs and in their contribution to total expression from all DEGs. For three out of four of the glandular tissues, there is sharp bias between conserved genes in terms of frequency and expression, with much less expression than would be expected. For the more conserved tissues, however, there is not such a sharp drop, and some tissues show a positive bias in expression from conserved genes (statistics in supplementary information, Supplementary Material online). Hence, for the conserved tissues, there is still a much stronger contribution from conserved genes relative to TRGs than is the case for the more novel tissues. Brain and antennal comparisons between nurses and foragers produced relatively few DEGs (of any taxonomic class) and were excluded from the analysis.

DE-TRGs between Tissues

Our experimental design also allowed us to identify genes conferring tissue-specific functions by making comparisons between tissues (e.g., antenna to midgut). We made all 45 possible comparisons within the nurse bee samples to identify DEGs that confer tissue-specific functions. Figure 2 uses a rough heat map approach to depict the same analysis as was done for HEGs and DEGs in the context of social behavior. Each row of the figure shows either the percentage of DEGs in all of the TRG categories (top panel) or the percentage of transcripts that come from all TRGs (bottom panel). It is thus a pictorial representation of the same type of comparison shown in figure 1 for HEGs and in table 1 for DE-TRGs in the context of social behavior. The results are consistent with what we found in the analyses of HEGs and DE-TRGs between nurses and foragers, with notable exceptions. Essentially, what we found is that it does not matter what other tissue the novel tissues (such as HPG and sting gland) are compared with: The pattern that DE-TRGs are responsible for a large fraction of total gene expression holds. In addition, we found that the antenna is also a place of high overall expression of

Table 2. Extremely Highly Expressed TRGs of Known Function.

Gene	TRG level	Name	Function	Tissues Expressed in
NM_001011582.1	Orphan	apisimin	Microbe defense	Man HPG
NM_001011589.1	Orphan	odorant_binding_protein_4	Olfaction	Ant
NM_001011607.1	Orphan	melittin	Venom	Nas Sting
NM_001011611.2	Orphan	mast_cell-degranulating_peptide	Venom	Sting
NM_001011612.1	Orphan	apamin_protein	Venom	Sting
NM_001011613.1	Orphan	apidaecin_1	Microbe defense	Nas
NM_001040220.1	Orphan	odorant_binding_protein_7	Olfaction	Nas
NM_001040270.1	Orphan	venom_allergen_Api_m_6	Venom	Sting
NM_001085344.1	Orphan	apidermin_3	Cuticle formation	Sting
NM_001085345.1	Orphan	apidermin_1	Cuticle formation	Man Nas Sting
XM_006557893.1	Orphan	secapin	Venom	Sting
NM_001011591.1	Hymenoptera	odorant_binding_protein_2	Olfaction	Ant
NM_001011616.2	Hymenoptera	defensin_1	Microbe defense	Man HPG
NM_001040206.1	Hymenoptera	odorant_binding_protein_21	Olfaction	Ant Nas Sting
NM_001040223.1	Hymenoptera	odorant_binding_protein_14	Olfaction	Nas
NM_001114198.1	Hymenoptera	apolipophorin-III-like_protein	Lipid transport	Man Nas Sting
XM_006558297.1	Hymenoptera	odorant_binding_protein_15	Olfaction	Ant Sting
XM_006563358.1	Hymenoptera	chymotrypsin_inhibitor-like	Protein metabolism	HPG
NM_001011588.1	Insect	odorant_binding_protein_5	Olfaction	Ant
NM_001011590.1	Insect	odorant_binding_protein_1	Olfaction	Ant
NM_001040205.1	Insect	odorant_binding_protein_16	Olfaction	Ant
NM_001040226.1	Insect	odorant_binding_protein_11	Olfaction	Ant
NM_001144839.1	Insect	C1q-like_venom_protein	n/a	Nas Sting
XM_001123076.3	Insect	glucose_dehydrogenase_B_acceptor	Glucose metabolism	HPG
XM_006563359.1	Insect	chymotrypsin_inhibitor	Protein metabolism	Nas
NM_001011583.2	Arthropod	chemosensory_protein_3	Olfaction	Ant Man Nas Sting
NM_001077820.1	Arthropod	chemosensory_protein_1	Olfaction	Ant Sting
NM_001270813.1	Arthropod	cuticular_protein_14	Cuticle formation	Ant Nas
XM_001122696.3	Arthropod	protein_takeout-like	Juvenile hormone binding	Ant
XM_006561656.1	Arthropod	protein_takeout-like	Juvenile hormone binding	Ant
XM_393105.5	Arthropod	circadian_clock-controlled_protein	Circadian rhythm	Ant Nas Sting

NOTE.—Ant, antenna; Man, mandibular gland; Nas, Nasonov gland.

DE-TRGs (odorant binding proteins, of which insects have taxonomically restricted forms [Leal 2013]).

What Do TRG Expression Patterns Tell Us of Coding Sequence Evolution?

TRGs represent the most extreme form of coding sequence evolution (as they are completely unique coding sequences), yet there is no reason to suspect that their patterns of expression should not be informative for the role played by coding sequence change in general. Table 2 shows all TRGs identified in this study that are both extremely highly expressed (using the modENCODE definition of RPKM > 1,000) and are of known function. These genes should provide a clue as to where coding sequence change is most pronounced in evolution (if TRGs are merely extreme cases of general patterns). The TRGs in table 2, particularly Orphan, are biased toward secreted proteins that occur in tissues that have radically changed in function. The venoms, for example, are proteins produced in the sting gland and secreted to form the bee's venom (Owen and Bridges 1976; Roat et al. 2004; Peiren et al. 2008). The antimicrobial peptides are produced by many tissues, but are secreted into the hemolymph to fight pathogens (Bulet et al. 1999; Evans et al. 2006). The cuticle proteins are secreted for use in forming the exoskeleton (Andersen et al. 1995; Kucharski et al. 2007). Many, but not all, odorant-binding proteins and chemosensory proteins are also secreted in order to facilitate molecular transport (reviewed in Leal 2013). Because proteins in all of these classes are secreted (and then function in a relatively independent manner; that is, they are not part of protein complexes or complex signal transduction pathways); they are amongst the most downstream of genes. Hence, they are likely not hubs that would be expected to be used repeatedly in many functionally distinct contexts. The evo–devo paradigm of needing to avoid negative pleiotropic effects resulting from strong coding sequence change in gene network hubs would therefore not apply to such genes (Beldade and Brakefield 2002; Wagner et al. 2007; Carroll 2008; Stern and Orgogozo 2008).

A series of testable hypotheses and predictions that should extend beyond TRGs to all genes can be produced based on the preceding argument made in reference to table 2.

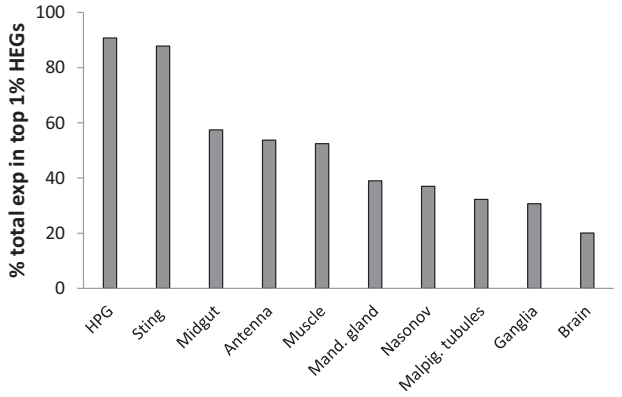

FIG. 3. Percentage of total expression in the each whole transcriptome that stems from just the top 1% of HEGs.

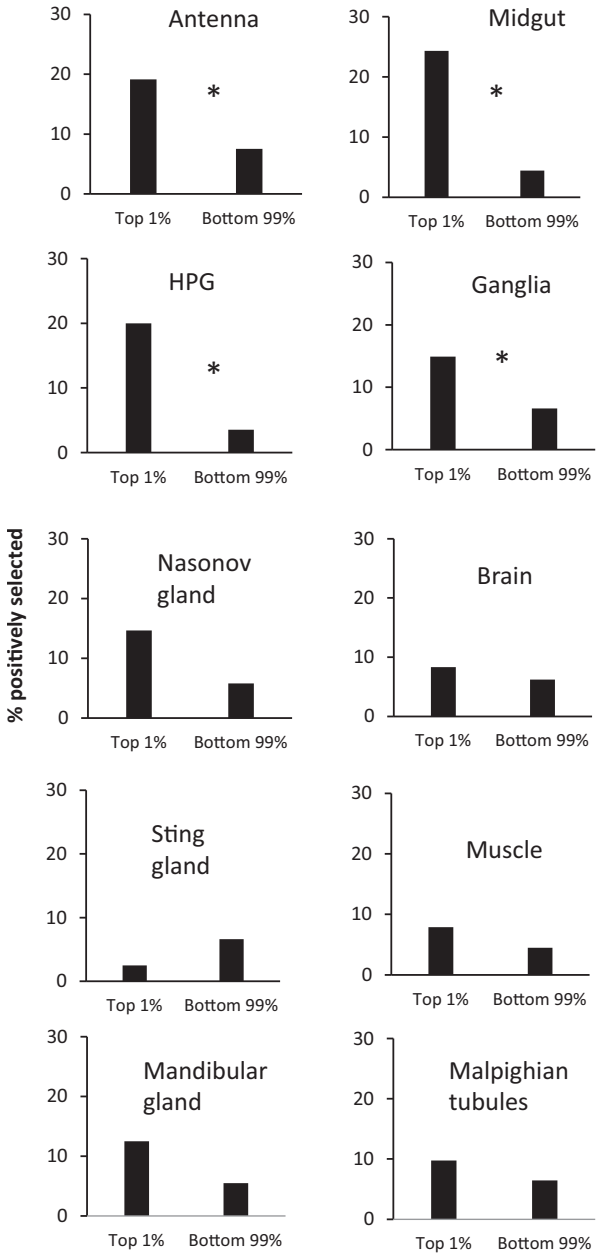

FIG. 4. Percentage of genes that are positively selected in the top 1% of HEGs versus the bottom 99% for each tissue. An asterisk signifies significance at the 0.05 level using the chi square test.

First, novel tissues (and specialized tissues in general) should function such that the highest expressed genes represent the majority of overall expression (including all genes, not just the highest 1% of expressed genes as in the previous analyses in this study). Further, strength of selection on the genes with the highest expression should be higher than on the genes with lower expression. In other words, cells have to divide their efforts between housekeeping and specialized functions related to their differentiated role in the organism. For many tissues, this balance should be far from parity such that cells exert most of their effort on tissue-specific functions. Should this be true, then when cells change in function, it should be a minority of HEGs that change coding sequences as they provide most of the specialized function.

Figure 3 shows that genes in the top 1% in terms of expression do in fact make up a strongly disproportionate fraction of total expression in many tissues (and the majority of expression in half the tissues). Nervous tissue is an interesting outlier that we will return to in the discussion section. Figure 4 further shows that for four of the tissues (antenna, midgut, HPG, and ganglia) genes in the top 1% of expression have a higher probability of being positively selected relative to genes in the lower 99% of expression. We show this by determining the percentage of genes in each category (top 1% or bottom 99% in terms of expression) that are experiencing positive selection based on a recent population genomic study using the McDonald–Kreitman (MK) test (Harpur et al. 2014). Interestingly, the midgut is the tissue in which the highest percentage of genes in the top 1% of expression are positively selected. The midgut is not a place of strong novel function; however, it is a place in which conserved genes with the properties of the TRGs in table 2 dominate in overall function. Digestive enzymes are secreted proteins that can sometimes work independently of other genes (Terra and Ferreira 1994). The midgut is also a tissue which is known to be experiencing positive selection in honey bees, as bees eat a novel substance (bee bread produced from pollen with the addition of microbes and likely other bee derived enzymes) (Woodard et al. 2011). Hence, the predictions stemming from TRGs extend to conserved genes with the same properties (secreted proteins that operate in a somewhat mechanistically independent manner). The sting gland does not show this pattern, but as the most HEGs are Orphans in this tissue, this still supports our basic hypothesis, which is that the highest 1% of genes in terms of expression are genes with novel or positively selected coding sequence. The Nasonov and mandibular glands were not significant for this effect, but the data trended in this direction. In general, these two tissues, which produce social pheromones (and not simple protein products), are not as simple in function as some of the other novel tissues.

A second major prediction stemming from table 2 is that genes that show tissue-specific expression (and are hence likely not used repeatedly) should be the most likely to be

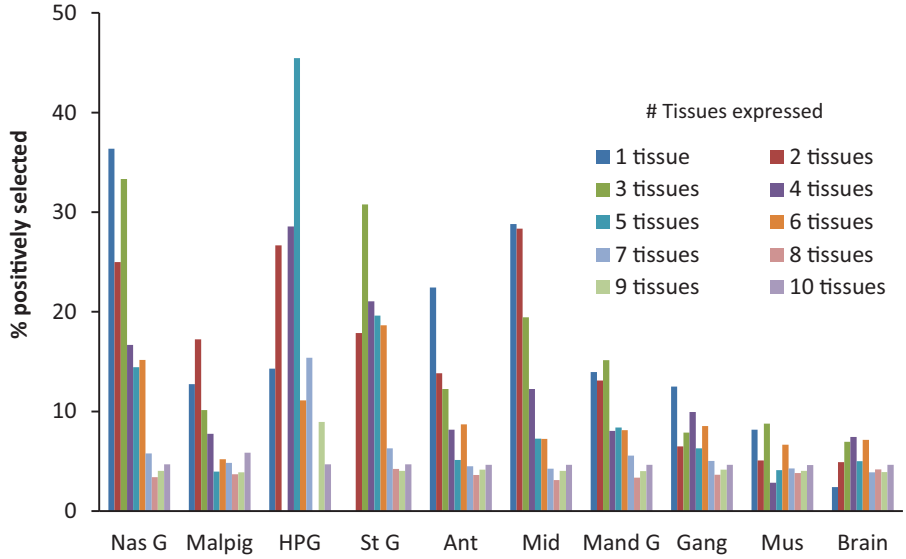

Fig. 5. The percentage of positively selected genes in genes expressed in variable numbers of tissues. Genes expressed in fewer tissues have a higher probability of being positively selected for 8 out of 10 tissues.

positively selected. Figure 5 shows that this is the case. Genes expressed in fewer tissues have a higher probability of being positively selected for all tissues with the exception of the thoracic muscles and the brain (χ^2 test: $df = 9$: Nasonov gland: $P < 0.0001$; malpighian tubules: $P = 0.003$; HPG: $P < 0.0001$; sting gland: $P < 0.0001$; antenna: $P < 0.0001$; midgut: $P < 0.0001$; mandibular gland: $P = 0.02$; ganglia: $P = 0.32$; muscle: $P = 0.67$; brain: $P = 0.85$). The pattern is particularly strong in the glands with novel social communication functions (the Nasonov and mandibular glands). The pattern is also evident in the tissues with the most novel functions, the sting gland and HPG, but is surprisingly weaker. There is a simple explanation for this exception, however, given that the highest expressed genes in the sting gland are strongly biased toward Orphans, and Orphans (and other narrowly restricted TRGs) cannot be tested for rates of positive selection across long time scales as they have few to no homologs. Essentially, although the Orphans and narrowly restricted TRGs of the HPG and sting gland may not show strong selection within the genus *Apis*, they are strikingly novel sequences when compared over longer taxonomic scales.

For the HPG, the highest proportion of positively selected genes by the MK test was seen for genes that are expressed in four to five tissues. This likely represents a problem of contamination leading to an overestimate of the number of tissues where key genes are expressed, as the positively selected genes (genes encoding the royal jelly proteins) are produced at such high levels and in such a large filamentous structure that contamination from this gland occurs in all other tissues taken from the head capsule (Whitfield et al. 2002). This problem has been long known (Whitfield et al. 2002) and hence, it is likely that the number of tissues for which the royal jelly proteins are expressed should be lowered by two (as expression in the brain and mandibular gland likely represents contamination from the HPG).

The final major prediction regarding our hypothesis for the evolution of novelty in adult organisms is that connectedness within a gene network should predict both probability of being a TRG and being positively selected. To test these predictions, we estimated the network connectedness of each gene based on coexpression patterns across all samples in our data set (Langfelder and Horvath 2008). As expected, total connectivity in transcriptional networks depended on TRG category (Mood median test: $\chi^2 = 54.25$, $df = 6$, $P < 0.001$), with the most restricted groups generally having lower connectivity than conserved genes (fig. 6*A*; statistical details in supplementary information, Supplementary Material online). The same general pattern was also true for those genes showing evidence of positive selection (Mood median Test: $\chi^2 = 13.61$, $df = 5$, $P = 0.018$; fig. 6*B*). The genes in table 2 have a connectivity level consistent with that found for other highly restricted genes (mean = 30.03, s.d. = 19.49). There was no difference in gene connectivity depending on whether genes had been identified as being in the class of positively selected genes, but there was a trend in that direction (Mann–Whitney *U*: Positively selected genes: $N = 744$, median = 32.7, nonpositively selected genes: $N = 7{,}389$, median = 32.7, $P = 0.06$). Among the class of positively selected genes, the selection coefficient and connectivity were negatively correlated (Spearman's rank correlation rho = −0.106, $P = 0.00394$), indicating that less highly connected genes experience elevated rates of molecular evolution. All connectivity values, along with whether a gene is positively selected in *Apis* are in supplementary table S5, Supplementary Material online.

Discussion

This study makes a number of important discoveries regarding the genetic mechanisms underlying the evolution of novelty. Previous work has also shown that novel genes are associated with novel tissues (Neme and Tautz 2013;

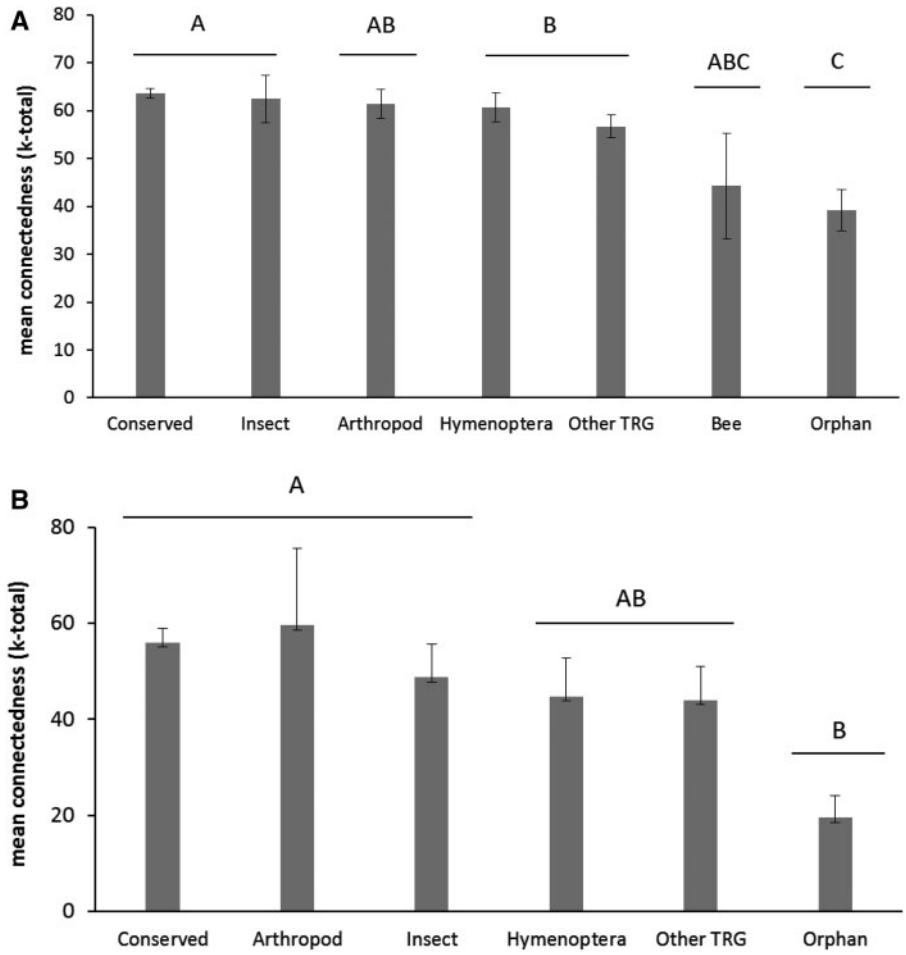

Fig. 6. TRGs, and orphans in particular, are less well connected within transcriptional networks than more highly conserved genes. This is true for all genes (panel *A*) and also just for those that are positively selected (panel *B*). Error bars are standard deviations, gene classes sharing the same letter (*A*, *B*, or *C* in panel *A* and *A* or *B* in panel *B*) above the bar are not statistically different, whereas those not sharing the same letter are according to the Mood median test. Hence, for example, Orphans are significantly lower than all other categories except bee in panel *A*.

Reinhardt et al. 2013; Shigenobu and Stern 2013; Light et al. 2014; Zhao et al. 2014). However, earlier studies focused on single cases of novelty, whereas the present study focuses on a species for which multiple case histories of the evolution of novelty can be contrasted with evolution for more moderate levels of change in conserved tissues. Because of this, first, we were able to show that TRGs are strongly associated with novel functions and tissues, but much less so with conserved tissues and functions. Second, we show that the function of cells having undergone cell line differentiation into specialized tissues is associated with extremely high expression of genes conferring tissue-specific functions. Essentially, more effort often goes into specialized functions rather than housekeeping functions in cells in the adult organism. We find that genes conferring specialized functions are often TRGs or conserved genes with high rates of coding sequence change. Hence, the evolution of novelty postdevelopment is strongly associated with coding sequence change either in the form of novel genes or positive selection in the coding regions of conserved genes. Finally, we show that genes with more tissue-specific expression have higher probabilities of being under positive selection than genes with expression across many tissues. As tissue-specific genes are not repeatedly used, they likely represent the most distal branches of gene networks. As such, changes to their coding sequence (and even their complete loss or gain) would not incur strong pleiotropic changes to other unrelated systems (Clark et al. 2007). Our demonstration that connectedness within a gene network is associated with the probability of a gene being positively selected further supports this argument. Both of these results (tissue-specific genes being positively selected and connectedness being associated with the probability of being positively selected) further support our hypothesis that the evolution of novelty is associated with changes to the coding sequence of genes that facilitate the specialized roles of cells post cell line differentiation.

The evo–devo toolkit model posits that although key genes change function, they do so via changes in their regulation, not in their coding sequence (Beldade et al. 2002; Carroll 2008; Stern and Orgogozo 2008). This is because changes to the coding sequence caused by selection in one context would likely cause strongly negative pleiotropic effects in those other contexts in which the gene is used. It is likely that this paradigm extends to the novel phenotypes studied here in part. This is because the transcription factors that control the expression of HEGs conferring tissue-specific

functions likely obey these rules. Essentially, it is probable that many transcription factors have taken on new roles controlling gene expression in the novel tissues and they have done so via changes to their regulatory rather than their coding sequence. This hypothesis, of course, will require future experimental testing. However, it suggests that the model proposed in this study based on coding sequence change to HEGs conferring specialized functions and the classic evo–devo model of changes to regulatory genes are complementary mechanisms for how novelty evolves postdevelopment. In short, after development the cells in an organism specialize via epigenetic reprogramming to play a variety of limited roles. As we show, these specialized and limited roles are the result of massive levels of expression of a relatively small number of genes that confer tissue-specific functions. We show that such genes have coding sequences that can change radically in response to selection for novel functions. However, the regulatory genes upstream from these HEGs conferring specialized functions are repeatedly used throughout the different tissues of the adult organism, and even earlier in development. They may take on their new regulatory roles for physiological novelty in the adult animal via the classic evo–devo paradigm in that their coding sequence does not changes, but new regulatory elements evolve to facilitate their new roles. This is a speculative conceptual framework for how the work conducted here might interact with earlier work in evo–devo and it awaits experimental verification.

In this study, we used the expression patterns of TRGs in order to explore the role of coding sequence change in general in the evolution of novelty. This approach differs from the norm, in that most studies assume the generation of novel genes and coding sequence change in conserved genes to be separate phenomena. Our data suggest a simpler hypothesis. It is possible that coding sequence change can be so strong in tissues selected for novel functions that conserved genes at distal points in gene networks diverge so much that they lose all resemblance to their homologs and become TRGs (Neme and Tautz 2013). This conceptual framework would provide an explanation for why patterns of TRG expression should be informative for the general nature of coding sequence change.

Lynch and Conery (2000, 2003) stressed the role played by gene duplication in the formation of new genes for new purposes. This idea has since been strongly supported by many studies (Zhang 2003; Taylor and Raes 2004; Ding et al. 2012; Kondrashov 2012). More recent studies have shown how novel genes can arise de novo from previously noncoding sequence or from noncoding functional RNA (Toll-Riera et al. 2009; Zhao et al. 2014). In this study, our focus was on the roles played by novel genes and not on their origin. However, it is worth noting that many of the key genes on which we focus (genes encoding odorant binding proteins in the antenna, various digestive enzymes in the midgut, and the royal jelly proteins in the HPG) are the result of gene duplications leading to large gene families with novel functions for different paralogs. This would support an important role for gene duplications in the evolution of novel genes for new functions. However, several of the genes we identify encode very small uncharacterized proteins that are not part of large gene families and likely arose via de novo processes. This would support a role for de novo gene formation leading to important new genes. Perhaps the simplest interpretation of these data is that evolution may be opportunistic and depend on gene duplications in some cases and de novo formation in others, but the end result is the same in that new genes for new functions emerge.

The nervous tissues studied here, the brain and the second thoracic segmental ganglion, did not obey the conceptual model proposed in which distal branches of gene networks are key to specialized functions postdevelopment. This is a case of the exception proving the rule. For behavior, novel function is not the result of key genes with high expression conferring specialized tissue-specific functions. In contrast, novel behavior is the result of the rearrangement of nerve cells into new circuits (Yao and Shafer 2014). In this context, the evo–devo paradigm does hold as the same genes (underlying neurotransmitter function and nerve cell growth) are reused to create novel nervous system modules (Winslow et al. 1993; Insel et al. 1994; Carter et al. 1995; Fitzpatrick et al. 2005; Turner et al. 2010). Changes to the regulatory elements controlling the expression of these nervous system building blocks would be expected to be the loci of evolutionary change (Harpur et al. 2014).

In summary, the adult animal is likely a hodgepodge of qualitatively different tissue types that have specialized behavior encoded by different evolutionary genetic mechanisms. Tissues such as the digestive tract, most glands, and the antenna are highly specialized such that a relatively small number of key genes (that are often TRGs or rapidly evolving) make up most expression and contribute disproportionately to novel function, whereas tissues such as the brain are characterized by more complex transcriptomes in which many highly conserved genes are differentially regulated to create novelty. Further, when other adult phenotypes such as the immune system, the endocrine system, and various sensory systems are examined, it is likely that some may make use of still other mechanisms for the evolution of novelty. Hence, the conceptual framework emphasized by evo–devo can be seen as a special case in which selection operates on gene networks in which a variety of topological features (relating to location in a gene network (upstream or downstream) and redundancy of function with close paralogs) can determine overall cellular and tissue function.

Materials and Methods

Colonies and Collection Methods

Honey bees were kept in apiaries at the Laidlaw beekeeping facility at UC Davis. Bees were managed according to standard beekeeping practices. Three full size (two stories) colonies were used in the study. All colonies were healthy and populous at the time of collection. All collections were made in August–September of 2012. Nurse bees were identified by observing bees with their head in a larval cell for at least 3 s, whereas foragers were collected at the entrance with pollen on their legs. All bees were collected onto dry ice for

transport to the lab where they were stored at −80 °C until use.

Dissection, Extractions, and Sequencing

Bees were removed from the freezer and immediately dissected in 60% ethanol over dry ice. Each tissue was dissected within 5 min of thawing and was immediately homogenized in Trizol. Tissue from 5 to 20 individuals (depending on the size of the structure) was pooled for each biological replicate. Three biological replicates were produced for each tissue for both castes (nurses and foragers).

RNA was confirmed to be of high quality with the Nanodrop 1000 and Bioanalyzer 2100. Libraries were made with the NEBNext Illumina RNA-Seq library kit according to the manufacturer's instructions. Sequencing was performed on the HiSeq 2000 (100 bp paired end). The raw data are available at the NCBI SRA archive (SRP027395, SRP020361, SRP041189). The number of reads produced for each replicate in each tissue is given in supplementary table S6, Supplementary Material online.

RNA-Seq Analyses

Initial quality control was conducted with the fastx toolkit and the cutadapt software packages. Reads with average quality scores less than 25 were removed, and the ends of reads were clipped such that the mean quality of the last five bases was greater than 25. Illumina adapter contamination was also removed. Tophat (v2.04), with bowtie2, and default parameters were used for alignment of reads to the most recent build (4.5) of the *Apis mellifera* genome (Elsik et al. 2014). We used the HTSeq package (with default parameters) to generate counts of reads aligning to each gene (with intersection union setting). EdgeR was used to call DEGs with default parameters and tag-wise dispersion option (with false discovery rate < 0.05).

Determination of TRG Status

We used the same approach we used in a previous study (Johnson and Tsutsui 2011) to identify TRGs. Essentially, *A. mellifera* genes without a blast hit in another genome at an *E* value of less than 10^{-4} were considered to not be present in that genome (Domazet-Loso and Tautz 2003; Zhang et al. 2007; Toll-Riera et al. 2009b). We used the list of all mRNA transcripts in the most recent NCBI honey bee genome build (4.5) as our set of honey bee genes and blasted (with BLASTx) each transcript against all the proteins available in 71 published genomes. These genomes covered a wide taxonomic range given in supplementary table S7, Supplementary Material online. From each set of blasts, we acquired a list of species that contain or are missing each honey bee transcript. We then made a MySQL database from which we identified genes falling into several TRG classes. Orphans were genes found only in *A. mellifera*. Bee-specific genes were found in at least one other species of bee but nowhere else. Hymenoptera genes were found in at least one other hymenopteran but nowhere else. Insect-specific genes were found in Hymenoptera and at least one other clade of insects, but nowhere else. Arthropod-specific genes were found in at least one of the arthropod genomes examined but not anywhere else. The class O-TRG was used for genes that were clearly restricted, but which had a complex pattern of restriction. O-TRGs range from genes that were found in only one other genome (but not a closely related species to honey bees) to genes that occurred in less than 50% of the major taxonomic groups we canvassed (arthropods, annelids, platyhelminthes, nematodes, porifera, cnidaria, echinoderms, mollusks, primitive chordates, vertebrates, all plants, all fungi, all protists, and all bacteria). Conserved genes were all genes not falling into any of the TRG categories.

Our main goal in identifying TRGs was not to definitively label particular genes as Orphans, or hymenopteran specific, for example, as this depends on the actively growing database of available genomes. As new genomes are sequenced, new homologs of many genes are found to be present but rare in various taxonomic groups. Our categories of TRGs are rather meant to reflect increasing levels of taxonomic restriction. Given the large number of genomes we included in our analysis, the Orphans we identify can be reliably considered to be more restricted than hymenoptera genes, which are more restricted than the insect-specific genes and so forth, even they are not true Orphans, for example.

Determination of Tissue Specificity

To explore tissue specificity of gene expression, we focused on genes that are at least moderately expressed (RPKM > 25). We first identified all genes in each tissue with average RPKM values (across the three replicates) greater than 25. We then determined in how many other tissues each gene in the focal tissue showed at least moderate expression. We thus obtained for each tissue a data set of how many other tissues each gene expressed in that tissue was expressed. This ranged from genes only expressed in the focal tissue, to genes expressed in all tissues. We focused on genes of at least moderate expression because we are primarily interested in HEGs and genes of very high expression in one tissue that are only lowly expressed in another tissue are candidates for being false positives for expression in the second tissue. Several venom genes, for example, which show extremely high expression in the sting gland, are also lowly expressed in the gut and the Nasonov gland (the gut and Nasonov gland are anatomically very close to the sting gland). It is more likely that contamination during dissection causes the low levels of expression in the neighboring structures then actual expression because the venom genes are so specialized in function. Similar problems have been noted in the past with respect to the HPG and the brain (Whitfield et al. 2002).

Rates of Molecular Evolution

A recently published population genomic study has estimated population size-adjusted selection coefficients for all *A. mellifera* genes, by comparing sequence divergence between *A. mellifera* and *A. cerana* to segregating variation within an *A. mellifera* population, using the MK test (Harpur et al. 2014).

We use the list of genes identified as having experienced positive selection for the testing of many hypotheses in this article. This list is binary such that genes are categorized as being either positively selected or not positively selected. Hence, our analyses focus on the frequency of each class (high or low levels of positively selected genes) between lists of genes identified in different contexts (e.g., genes that show tissue-specific expression vs. genes expressed everywhere).

Gene Network Connectivity

We hypothesized that TRGs associated with novel functions were located more peripherally within transcriptional networks relative to more highly conserved genes. We used the Weighted Gene Coexpression Network Analysis in R package to identify modules of genes which displayed similar patterns of coexpression (Langfelder and Horvath 2008). We used default settings together with an empirically estimated soft threshold of eight, which corresponded to the asymptote for scale free topology model fit and mean connectivity (supplementary fig. S2, Supplementary Material online). Subsequently, for each gene we estimated the within-module and total connectivity, which is meant to reflect the number of genes with which a focal gene interacts within a gene regulatory network.

Supplementary Material

Supplementary information, tables S1–S7, and figures S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Gerard Smith and Billy Synk for assistance with collecting bees. The authors also thank two anonymous referees and an anonymous editor for comments that improved the manuscript. This work was funded by the University of California, Davis and a Hatch grant to B.R.J. (CA-D-ENM 2161-H).

References

Andersen SO, Hojrup P, Roepstorff P. 1995. Insect cuticular proteins. *Insect Biochem Molec Biol.* 25:153–176.

Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol.* 23:26–32.

Arthur W. 2002. The emerging conceptual framework of evolutionary developmental biology. *Nature* 415:757–764.

Beldade P, Brakefield PM. 2002. The genetics and evo-devo of butterfly wing patterns. *Nat Rev Genet.* 3:442–452.

Beldade P, Brakefield PM, Long AD. 2005. Generating phenotypic variation: prospects from "evo-devo" research on *Bicyclus anynana* wing patterns. *Evol Dev.* 7:101–107.

Beldade P, Koops K, Brakefield PM. 2002. Modularity, individuality, and evo-devo in butterfly wings. *Proc Natl Acad Sci U S A.* 99: 14262–14267.

Brockmann A, Robinson GE. 2007. Central projections of sensory systems involved in honey bee dance language communication. *Brain Behav Evol.* 70:125–136.

Bulet P, Hetru C, Dimarcq JL, Hoffmann D. 1999. Antimicrobial peptides in insects; structure and function. *Dev Comp Immunol.* 23:329–344.

Carroll SB. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature* 376:479–485.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134:25–36.

Carter CS, Devries AC, Getz LL. 1995. Physiological substrates of mammalian monogamy: the prairie vole model. *Neurosci Biobehav Rev.* 19:303–314.

Chen SD, Krinsky BH, Long MY. 2013. New genes as drivers of phenotypic evolution. *Nat Rev Genet.* 14:645–660.

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203 218.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet.* 9:938–950.

Crailsheim K. 1991. Interadult feeding of jelly in honeybee (*Apis mellifera*) colonies. *J Comp Physiol B.* 161:55–60.

Daltry JC, Wuster W, Thorpe RS. 1996. Diet and snake venom evolution. *Nature* 379:537–540.

Ding Y, Zhou Q, Wang W. 2012. Origins of new genes and evolution of their novel functions. *Ann Rev Ecol Evol Syst.* 43:345–363.

Domazet-Loso T, Tautz D. 2003. An evolutionary analysis of orphan genes in *Drosophila*. *Genome Res.* 13:2213–2219.

Elsik CG, Worley KC, Bennett AK, Beye M, Camara F, Childers CP, de Graaf DC, Debyser G, Devreese B, Elhaik E, et al. 2014. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* 15:86.

Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, Kanost M, Thompson GJ, Zou Z, Hultmark D. 2006. Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Molec Biol.* 15:645–656.

Fitzpatrick MJ, Ben-Shahar Y, Smid HM, Vet LEM, Robinson GE, Sokolowski MB. 2005. Candidate genes for behavioural ecology. *Trends Ecol Evol.* 20:96–104.

Fry BG, Wuster W, Kini RM, Brusic V, Khan A, Venkataraman D, Rooney AP. 2003. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *J Mol Evol.* 57:110–129.

Hahn MW, Han MV, Han S-G. 2007. Gene family evolution across 12 Drosophila genomes. *PLoS Genet.* 3:2135–2146.

Halligan DL, Kousathanas A, Ness RW, Harr B, Eory L, Keane TM, Adams DJ, Keightley PD. 2013. Contributions of protein-coding and regulatory change to adaptive molecular evolution in Murid Rodents. *PLoS Genet.* 9:e1003995.

Harpur BA, Kent CF, Molodtsova D, Lebon JMD, Alqarni AS, Owayss AA, Zayed A. 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci U S A.* 111:2614–2619.

Hoekstra HE. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* 97:222–234.

Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61:995–1016.

Insel TR, Wang ZX, Ferris CF. 1994. Patterns of brain vasopressin receptor distribution associated with social organization in microtine rodents. *J Neurosci.* 14:5381–5392.

Johnson BR. 2008. Within-nest temporal polyethism in the honey bee. *Behav Ecol Sociobiol.* 62:777–784.

Johnson BR. 2010. Division of labor in honeybees: form, function, and proximate mechanisms. *Behav Ecol Sociobiol.* 64:305–316.

Johnson BR, Linksvayer TA. 2010. Deconstructing the superorganism: social physiology, groundplans, and sociogenomics. *Q Rev Biol.* 85:57–79.

Johnson BR, Tsutsui ND. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* 12:164.

Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. 2009. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* 25:404–413.

King TP, Spangfort MD. 2000. Structure and biology of stinging insect venom allergens. *Int Arch Allergy Immunol.* 123:99–106.

Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc R Soc B-Biol Sci.* 279: 5048–5057.

Kucharski R, Maleszka J, Maleszka R. 2007. Novel cuticular proteins revealed by the honey bee genome. *Insect Biochem Molec Biol.* 37: 128–134.
Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.
Leal WS. 2013. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annu Rev Entomol.* 58: 373–391.
Light S, Basile W, Elofsson A. 2014. Orphans and new gene origination, a structural and evolutionary perspective. *Curr Opin Struct Biol.* 26: 73–83.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302:1401–1404.
Mitchell-Olds T, Willis JH, Goldstein DB. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nat Rev Genet.* 8:845–856.
Nadeau NJ, Jiggins CD. 2010. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet.* 26:484–492.
Neme R, Tautz D. 2013. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. *BMC Genomics* 14:117.
Owen MD, Bridges AR, Maleszka R. 1976. Aging in venom glands of queen and worker honey bee (Apis mellifera-L)-some morphological and chemical observations. *Toxicon* 14:1–5.
Owen MD, Pfaff LA. 1995. Melittin synthesis in the venom system of the honey bee (*Apis mellifera* L.). *Toxicon* 33:1181–1188.
Page RE, Robinson GE. 1991. The genetics of division of labor in honey bee colonies. *Adv Insect Physiol.* 23:117–169.
Parker DJ, Gardiner A, Neville MC, Ritchie MG, Goodwin SF. 2014. The evolution of novelty in conserved genes; evidence of positive selection in the Drosophila fruitless gene is localised to alternatively spliced exons. *Heredity* 112:300–306.
Peiren N, de Graaf DC, Vanrobaeys F, Danneels EL, Devreese B, Van Beeurnen FJ, Jacobs J. 2008. Proteomic analysis of the honey bee worker venom gland focusing on the mechanisms of protection against tissue damage. *Toxicon* 52:72–83.
Peters L, Zhu-Salzman K, Pankiw T. 2010. Effect of primer pheromones and pollen diet on the food producing glands of worker honey bees (*Apis mellifera* L.). *J Insect Physiol.* 56:132–137.
Ranz JM, Parsch J. 2012. Newly evolved genes: moving from comparative genomics to functional studies in model systems. *Bioessays* 34: 477–483.
Reinhardt JA, Wanjiru BM, Brant AT, Saelao P, Begun DJ, Jones CD. 2013. De Novo ORFs in *Drosophila* are important to organismal fitness and evolved rapidly from previously non-coding sequences. *Plos Genet.* 9:e1003860.
Roat TC, Nocelli RCF, de Moraes R, Cruz-Landim C. 2004. Juvenile hormone effect on the venom gland secretory cycle in workers of *Apis mellifera* (Hymenoptera, apidae). *Sociobiology* 43:193–199.
Robinson GE, Grozinger CM, Whitfield CW. 2005. Sociogenomics: social life in molecular terms. *Nat Rev Genet.* 6:257–270.
Seeley TD. 1982. Adaptive significance of the age polyethism schedule in honeybee colonies. *Behav Ecol Sociobiol.* 11:287–293.
Shigenobu S, Stern DL. 2013. Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. *P R Soc B.* 280:20121952.
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.
Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* 62:2155–2177.
Sucena E, Stern DL. 2000. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby. *Proc Natl Acad Sci U S A.* 97: 4530–4534.
Sumner S. 2014. The importance of genomic novelty in social evolution. *Mol Ecol.* 23:26–28.
Tautz D, Domazet-Loso T. 2011. The evolutionary origin of orphan genes. *Nat Rev Genet.* 12:692–702.
Taylor JS, Raes J. 2004. Duplication and divergence: The evolution of new genes and old ideas. *Ann Rev Genet.* 38:615–643.
Terra WR, Ferreira C. 1994. Insect digestive enzymes: properties, compartmentalization and function. *Comp Biochem Physiol B-Biochem Mol Biol.* 109:1–62.
Toll-Riera M, Bosch N, Bellora N, Castelo R, Armengol L, Estivill X, Mar Alba M. 2009. Origin of primate orphan genes: a comparative genomics approach. *Mol Biol Evol.* 26:603–612.
Toth AL, Robinson GE. 2007. Evo-devo and the evolution of social behavior. *Trends Genet.* 23:334–341.
Turner LM, Young AR, Roempler H, Schoeneberg T, Phelps SM, Hoekstra HE. 2010. Monogamy evolves through multiple mechanisms: evidence from V1aR in Deer Mice. *Mol Biol Evol.* 27: 1269–1278.
Ueno T, Nakaoka T, Takeuchi H, Kubo T. 2009. Differential gene expression in the hypopharyngeal glands of worker honeybees (*Apis mellifera* L.) Associated with an age-dependent role change. *Zool Sci.* 26: 557–563.
Wagner GP, Pavlicev M, Cheverud JM. 2007. The road to modularity. *Nat Rev Genet.* 8:921–931.
Wegener J, Huang ZY, Lorenz MW, Bienefeld K. 2009. Regulation of hypopharyngeal gland activity and oogenesis in honey bee (*Apis mellifera*) workers. *J Insect Physiol.* 55:716–725.
Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson HM, Soares MB, Robinson GE. 2002. Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Res.* 12: 555–566.
Whitfield CW, Cziko AM, Robinson GE. 2003. Gene expression profiles in the brain predict behavior in individual honey bees. *Science* 302: 296–299.
Williams JB, Roberts SP, Elekonich MM. 2008. Age and natural metabolically-intensive behavior affect oxidative stress and antioxidant mechanisms. *Exp Gerontol.* 43:538–549.
Wilson GA, Bertrand N, Patel Y, Hughes JB, Feil EJ, Field D. 2005. Orphans as taxonomically restricted and ecologically important genes. *Microbiol-Sgm.* 151:2499–2501.
Wilson GA, Feil EJ, Lilley AK, Field D. 2007. Large-scale comparative genomic ranking of taxonomically restricted genes (TRGs) in bacterial and archaeal genomes. *PLoS One* 2:e324.
Winslow JT, Hastings N, Carter CS, Harbaugh CR, Insel TR. 1993. A role for central vasopressin in pair bonding in monogamous prairie voles. *Nature* 365:545–548.
Wissler L, Gadau J, Simola DF, Helmkampf M, Bornberg-Bauer E. 2013. Mechanisms and Dynamics of Orphan Gene Emergence in Insect Genomes. *Genome Biol Evol.* 5:439–455.
Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG, Robinson GE. 2011. Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci U S A.* 108: 7472–7477.
Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol.* 20:1377–1419.
Yao Z, Shafer OT. 2014. The Drosophila circadian clock is a variably coupled network of multiple peptidergic units. *Science* 343: 1516–1520.
Zayed A, Robinson GE. 2012. Understanding the relationship between brain gene expression and social behavior: lessons from the honey bee. *Annu Rev Genet.* 46:591–615.
Zhang JZ. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18:292–298.
Zhang G, Wang H, Shi J, Wang X, Zheng H, Wong GKS, Clark T, Wang W, Wang J, Kang L. 2007. Identification and characterization of insect-specific proteins by genome data analysis. *BMC Genomics* 8:93.
Zhao L, Saelao P, Jones CD, Begun DJ. 2014. Origin and spread of de novo genes in *Drosophila melanogaster* populations. *Science* 343:769–772.