

The background of the entire cover is a close-up photograph of a honeycomb. The hexagonal cells are filled with a golden-yellow substance, likely honey or bee larvae. Several bees are visible, some on the surface of the comb and others partially inside the cells. The top portion of the image is overlaid with a solid teal-colored band.

ADVANCES IN GENOMICS AND EPIGENOMICS OF SOCIAL INSECTS

EDITED BY : Greg J. Hunt and Juergen R. Gadau
PUBLISHED IN: Frontiers in Genetics



frontiers

Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-080-0

DOI 10.3389/978-2-88945-080-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

ADVANCES IN GENOMICS AND EPIGENOMICS OF SOCIAL INSECTS

Topic Editors:

Greg J. Hunt, Purdue University, USA

Juergen R. Gadau, Institute for Evolution and Biodiversity, Germany



Introductory paragraph figure: Illustration by Elizabeth Cash

Cover: Photo by Greg Hunt

an ancient clade of social insects, the termites (infraorder Isoptera) provide an opportunity to study alternative mechanisms of caste determination and lifestyles that are aided by an array of endosymbionts. This research topic explores the use of genome sequence data and genomic techniques to help us explore how sociality evolved in insects, how epigenetic processes enable phenotypic plasticity, and the mechanisms behind whether a female will become a queen or a worker.

Citation: Hunt, G. J., Gadau, J. R., eds. (2017). *Advances in Genomics and Epigenomics of Social Insects*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-080-0

Table of Contents

05 Editorial: Advances in Genomics and Epigenomics of Social Insects

Greg J. Hunt and Juergen R. Gadau

The Termites

07 A genomic comparison of two termites with different social complexity

Judith Korb, Michael Poulsen, Haofu Hu, Cai Li, Jacobus J. Boomsma, Guojie Zhang and Jürgen Liebig

19 Omic research in termites: an overview and a roadmap

Michael E. Scharf

Epigenetics in Social Evolution

38 Epigenetics as an answer to Darwin's "special difficulty"

Brian R. Herb

43 Epigenetics as an answer to Darwin's "special difficulty," Part 2: natural selection of metastable epialleles in honeybee castes

Douglas M. Ruden, Pablo E. Cingolani, Arko Sen, Wen Qu, Luan Wang, Marie-Claude Senut, Mark D. Garfinkel, Vincent E. Sollars and Xiangyi Lu

58 Cytosine modifications in the honey bee (*Apis mellifera*) worker genome

Erik M. K. Rasmussen and Gro V. Amdam

Comparing Social and Solitary Species and Castes

63 Social parasitism and the molecular basis of phenotypic evolution

Alessandro Cini, Solenn Patalano, Anne Segonds-Pichon, George B. J. Busby, Rita Cervo and Seirian Sumner

74 Function and evolution of microRNAs in eusocial Hymenoptera

Eirik Søvik, Guy Bloch and Yehuda Ben-Shahar

85 Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviors

Daria Molodtsova, Brock A. Harpur, Clement F. Kent, Kajendra Seevananthan and Amro Zayed

92 Neutral and adaptive explanations for an association between caste-biased gene expression and rate of sequence evolution

Heikki Helanterä and Tobias Uller

Mechanistic Studies of Behavioral and Developmental Plasticity

102 *Developmental regulation of ecdysone receptor (EcR) and EcR-controlled gene expression during pharate-adult development of honeybees (Apis mellifera)*

Tathiana R. P. Mello, Aline C. Aleixo, Daniel G. Pinheiro, Francis M. F. Nunes, Márcia M. G. Bitondi, Klaus Hartfelder, Angel R. Barchuk and Zilá L. P. Simões

121 *Biased Allele Expression and Aggression in Hybrid Honeybees may be Influenced by Inappropriate Nuclear-Cytoplasmic Signaling*

Joshua D. Gibson, Miguel E. Arechavaleta-Velasco, Jennifer M. Tsuruda and Greg J. Hunt

133 *The evolutionary dynamics of major regulators for sexual development among Hymenoptera species*

Matthias Biewer, Francisca Schlesinger and Martin Hasselmann

A Genomic-Geographic Survey of Honey Bee Disease and Microbiome

144 *Metatranscriptomic analyses of honey bee colonies*

Cansu Ö. Tozkar, Meral Kence, Aykut Kence, Qiang Huang and Jay D. Evans



Editorial: Advances in Genomics and Epigenomics of Social Insects

Greg J. Hunt^{1*} and Juergen R. Gadau²

¹ Department of Entomology, Purdue University, West Lafayette, IN, USA, ² School of Life Sciences, Arizona State University, Tempe, AZ, USA

Keywords: social evolution, phenotypic plasticity, caste determination, epigenetics, eusociality, social insect, reproductive caste, sterile caste

Editorial on the Research Topic

Advances in Genomics and Epigenomics of Social Insects

The adaptive advantage of the eusocial lifestyle is evident from the fact that social insects represent more than half of the world's arthropod biomass. This topic explores how the recent advances in genomics and epigenomics are helping researchers to ask and answer questions concerning the evolution of social behavior and the genetic and epigenetic mechanisms behind phenotypic plasticity, i.e., how environmental signals can morph the same genome in a reproductive or non-reproductive individual resulting in dramatically different phenotypes. The articles in this research topic deal broadly with the evolution of reproductive and sterile castes (workers), mechanisms of caste determination, and the role of epigenetic processes for division of labor. The termites were the first group of insects to evolve eusociality and a thorough review describes what is known about the development of subcastes from a mechanistic perspective (nymphs, workers, soldiers) and the genomic contributions of gut symbionts and their hosts in digestion of wood, and the role of symbionts in host fitness (Scharf). Korb et al. compares the genomes of two termites with contrasting social complexities and symbioses. One of the interesting findings was that gene families involved in chemical communication in other social insects are not expanded in termites with more complex social organization. But transposable elements are, suggesting a role for transposition in social evolution but perhaps also pointing toward other mechanisms.

Darwin had a "special difficulty" understanding how sterile worker castes arose in the social insects and the existence of morphological specializations in individuals that did not have progeny. Epigenetic processes could provide mechanisms to encode these specializations within a worker caste just as it does in clonal cells of developing tissues. For example, experimental manipulations that cause honeybee workers to switch task specializations are marked by specific methylation events (Herb). However, the function of gene body methylation in regards to behavioral plasticity of workers, although associated with alternative splicing remains uncertain. The less-studied, and less abundant 5hydroxymethylcytosine (5hmC) modifications are intriguingly enriched in germ cells and brain of honeybees just as they are in mammals (Rasmussen and Amdam). Ruden et al. continue Herb's answer to Darwin's dilemma by suggesting that solitary ancestors of social bees may have experienced nutrient limitations, leading to a de-facto sterile caste in communal nesting situations. Stresses such as this could also activate heat shock proteins such as those that are involved in multi-generational inheritance of bizarre phenotypes in *Drosophila* without a change in DNA sequence. For example, Hsp90 inactivation has been linked to *Ubx* expression and the formation of pollen baskets on the legs of bees. On the other hand, Cini et al. ask how it is that some eusocial species went the other way and lost the sterile caste? Some of these species that showed social reversals evolved into social parasites that still depend on workers, but they exploit workers of closely related eusocial species. It seems more data is needed to determine whether comparing expression levels of conserved genes such as *Ubx* in different castes and species will provide insight into this process.

OPEN ACCESS

Edited and reviewed by:

Samuel A. Cushman,
United States Forest Service Rocky
Mountain Research Station, USA

*Correspondence:

Greg J. Hunt
ghunt@purdue.edu

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 08 September 2016

Accepted: 31 October 2016

Published: 16 November 2016

Citation:

Hunt GJ and Gadau JR (2016)
Editorial: Advances in Genomics and
Epigenomics of Social Insects.
Front. Genet. 7:199.
doi: 10.3389/fgene.2016.00199

Comparative studies of social insects and their solitary relatives can be used to look for signatures of social evolution. Sovik et al. analyze the question of whether specific miRNAs may have predisposed bee species to evolve eusociality. One pattern that emerges is that taxonomically restricted genes apparently have the highest rates of adaptive evolution in the honeybee. Similarly, recent expansions of regulatory sequences are restricted to specific ant lineages. A population genomic study combined with a meta-analysis of microarray data in the honeybee suggest that both protein coding and regulatory sequences that are rapidly evolving tend to lie at the periphery of gene networks (Moldostova et al.). One question asked by Helanterä and Uller is whether genes that show biases in expression between morphological castes of ants and bees are under strong purifying selection or whether neutral processes allow genes to be co-opted for specific roles in castes. Similar differences in gene expression have been observed between morphs of plants and animals. More data comparing expression between and within castes is needed to answer these questions.

The final three chapters we will mention take a more mechanistic approach to understanding development and behavior of bees. It has been repeatedly shown that fundamental changes in gene expression during development of either the worker or queen phenotype are mediated by ecdysteroid hormones. An impressive series of experiments by Mello et al. characterize the interactions of ecdysone, juvenile hormone and ecdysone receptor expression, along with downstream gene regulation in the fat body of honeybees. Analysis of interacting miRNAs on differentially transcribed genes during development may provide even more insight into the making of a queen.

Reciprocal hybrids derived from European and Africanized honeybees exhibit both gene expression differences and aggressive behaviors that depend on the direction of the cross. In hybrids with European maternity (but not the reciprocal family), about 8% of genes tested were strongly biased toward expression of the maternal allele in European-maternity hybrids (Gibson et al.). The biased genes are enriched for mitochondrial proteins and genes of metabolic function. Most biased genes are dispersed in the genome but large tracts of them are localized to two quantitative trait loci reported to influence aggressive behavior and alarm pheromone production. The authors speculate that this phenomenon involves partial cytoplasmic incompatibility,

nuclear/mitochondrial signaling, heat-shock proteins and short interfering RNA.

The vast majority of social insects are in the order Hymenoptera—the bees, ants, and wasps, which exhibit male haploidy. In most of these species female development is determined by heterozygosity at a single locus but some wasp species rely on a process that signals fertilization of the egg. A common theme however is the involvement of the gene transformer. In honeybees, it appears that duplication of a putative ortholog of *tra*, called *fem*, followed by positive selection resulted in the single-locus, multi-allele complementary sex determiner (*csd*) gene. Biewer et al. present evidence that ancestral duplications of *fem* is restricted to specific bee lineages. They go on to discuss how the gene that sends the initial signal in sex determination could be re-purposed after duplication.

It has been 10 years since the honey bee genome was published. Currently (2016), we have about 50 social insect genomes published with an expected rapid increase in the rate of genome sequencing on the horizon. For example, a proposal to sequence all ant genera has just been put forward by a group of researchers (GAGA, Global Ant Genomics Alliance). Hence, in the near future comparative genomics will greatly increase our knowledge about the processes that shaped the genomes of social insects. For example, comparative studies of bees will be useful for understanding changes associated with the evolution of sociality because there were multiple gains and losses of the eusocial lifestyle in this clade (Kocher and Paxton, 2014). Sequencing of individuals from population studies, coupled with phenotypic data will help identify genes under selection during social evolution, including the genetic architecture of traits of primitively social species. Functional genomics of social insects will be greatly aided by gene editing using CRISPR/CAS methodologies, RNAi and physiological and behavioral assays that are informed by what is learned from metabolomics and transcriptomics will enable social insects to be models for understanding behavioral genetics in general and social evolution in particular.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Kocher, S. D., and Paxton, R. J. (2014). Comparative methods offer powerful insights into social evolution in bees. *Apidologie (Celle)*. 45, 289–305. doi:10.1007/s13592-014-0268-3

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Hunt and Gadau. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A genomic comparison of two termites with different social complexity

Judith Korb^{1*}, Michael Poulsen², Haofu Hu³, Cai Li^{3,4}, Jacobus J. Boomsma², Guojie Zhang^{2,3} and Jürgen Liebig⁵

¹ Department of Evolutionary Biology and Ecology, Institute of Biology I, University of Freiburg, Freiburg, Germany

² Section for Ecology and Evolution, Department of Biology, Centre for Social Evolution, University of Copenhagen, Copenhagen, Denmark

³ China National Genebank, BGI-Shenzhen, Shenzhen, China

⁴ Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

⁵ School of Life Sciences, Arizona State University, Tempe, AZ, USA

Edited by:

Juergen Rudolf Gadau, Arizona State University, USA

Reviewed by:

Seirian Sumner, University of Bristol, UK

Bart Pannebakker, Wageningen University, Netherlands

Michael E. Scharf, Purdue University, USA

*Correspondence:

Judith Korb, Department of Evolutionary Biology and Ecology, Institute of Biology I, University of Freiburg, Hauptstrasse 1, D-79104 Freiburg, Germany
e-mail: judith.korb@biologie.uni-freiburg.de

The termites evolved eusociality and complex societies before the ants, but have been studied much less. The recent publication of the first two termite genomes provides a unique comparative opportunity, particularly because the sequenced termites represent opposite ends of the social complexity spectrum. *Zootermopsis nevadensis* has simple colonies with totipotent workers that can develop into all castes (dispersing reproductives, nest-inheriting replacement reproductives, and soldiers). In contrast, the fungus-growing termite *Macrotermes natalensis* belongs to the higher termites and has very large and complex societies with morphologically distinct castes that are life-time sterile. Here we compare key characteristics of genomic architecture, focusing on genes involved in communication, immune defenses, mating biology and symbiosis that were likely important in termite social evolution. We discuss these in relation to what is known about these genes in the ants and outline hypothesis for further testing.

Keywords: chemical communication, genomes, immunity, social organization, social insects, symbiosis, termites, transposable elements

INTRODUCTION

The termites are “social cockroaches,” a monophyletic clade (Infraorder “Isoptera”) nested within the Blattodea (Inward et al., 2007a; Engel et al., 2009; Krishna et al., 2013). They superficially resemble the ants in having wingless worker foragers, but are fundamentally different in a series of ancestral traits that affect the organization of their eusocial colonies (Korb, 2008; Howard and Thorne, 2011). The (eu)social Hymenoptera are haplodiploid holometabolous insects whose males develop from haploid eggs and have transient roles in social life, because they survive only as sperm stored in the spermatheca of queens. Hymenopteran colonies thus consist of female adults that develop from fertilized eggs to differentiate into workers, virgin queens and occasionally soldiers of which only the former care for the helpless grub-like larvae. By contrast, termites are diploid hemimetabolous insects whose colonies usually have workers, soldiers, and reproductives of both sexes. Both have life-time monogamy upon colony founding as ancestral state (Hughes et al., 2008; Boomsma, 2013), but in contrast to the eusocial Hymenoptera, royal pairs regularly remate to produce immatures that increasingly come to resemble the workers, soldiers, and reproductives into which they differentiate. Hence, termite caste differentiation is based on phenotypic plasticity among immatures (Korb and Hartfelder, 2008; Miura and Scharf, 2011), while the eusocial Hymenoptera have castes of adults (Wilson, 1971).

Termites and ants also share many traits that convergently evolved in response to similar selective pressures (Thorne and

Traniello, 2003; Korb, 2008; Howard and Thorne, 2011). Both are mostly soil-dwelling and thus continuously exposed to high pathogen loads and their long-lived, populous and genetically homogenous colonies appear to be ideal targets for infections (Schmid-Hempel, 1998). However, both the ants and the termites also evolved impressive disease defense strategies, which have implied that very few pathogens have been able to specialize on infecting perennial ant and termite colonies over evolutionary time (Boomsma et al., 2005). In large part this appears to be due to immune defenses operating both at the individual and the collective (social immunity) level (Cremer et al., 2007; Rosengaus et al., 2011). Another common characteristic of the ants and termites is that both evolved complex communication systems that largely rely on chemical cues, such as cuticular hydrocarbons (CHCs), for nestmate recognition and within-colony communication (e.g., Liebig, 2010; Van Zweden and D’Ettorre, 2010). Strikingly, long-chained CHCs of queens often appear to function as fertility signals for workers of both lineages (Liebig et al., 2009; Weil et al., 2009; Liebig, 2010; van Oystaeyen et al., 2014). Here, we offer the first comparative exploration of the extent to which lineage ancestry has determined these convergent phenotypic similarities based on the first two termite genomes that became recently available (Poulsen et al., 2014; Terrapon et al., 2014).

The two termite genomes represent opposite ends of the social complexity spectrum within the Isoptera (Roisin, 2000) (Table 1) as they exemplify the two fundamental termite life types: the wood-dwelling one-piece nesters and the central place foraging

lineages that generally differ in social complexity, feeding ecology, gut symbionts, and developmental plasticity (Abe, 1987; Korb, 2007; Korb and Hartfelder, 2008) (**Figure 1**). *Zootermopsis nevadensis* belongs to the former type and *Macrotermes natalensis* to the latter. Wood-dwelling species (Abe, 1987; Shellman-Reeve, 1997) nest within a single piece of dead wood that serves both as food and nesting habitat so the termites never leave their nest to forage. This social syndrome is widely considered to be ancestral (e.g., Noirot and Pasteels, 1987, 1988; Inward et al., 2007b) and associated with high degrees of developmental plasticity for the individual termites (**Figure 2A**). Workers remain totipotent immatures throughout several instars that commonly develop further into sterile soldiers, winged sexuals (alates) that found new nests as primary reproductives, or neotenic reproductives that reproduce within the natal nest (**Figure 2A**).

The foraging termite species (also called “multiple piece nesters”; Abe, 1987; Shellman-Reeve, 1997) forage for food outside the nest at some point after colony foundation and bring it back to the colony to feed nestmates. They represent more than 85% of the extant termite species (Kambhampati and Eggleton, 2000). They have true workers and an early separation into distinct developmental pathways (Roisin, 2000; Korb and Hartfelder, 2008) (**Figure 2B**). In the apterous line, individuals are unable to develop wings and can thus never disperse as reproductives. They become workers and soldiers, but can in some species also advance to become neotenic reproductives in their own nest. In the nymphal line, however, individuals develop wings and dispersing phenotypes that found new colonies elsewhere (**Figure 2B**). The Macrotermitinae to which *Macrotermes natalensis* belongs are special examples of foraging termites because their colonies are dependent on nutrition provided by a *Termitomyces* symbiont (Basidiomycota: Agaricales) (Wood and Thomas, 1989; Nobre et al., 2011). This fungal symbiosis is evolutionarily derived and comes in addition to more fundamental protist (lower termites) and bacterial gut symbionts (all termites), which have played major roles throughout termite evolution. *Macrotermes* species have two (major/minor) worker castes and two (major/minor) soldier castes (Ruelle, 1970) that may be determined as early as the egg stage (suggested for *Macrotermes*

michaelseni by Okot-Kotber, 1985). *Macrotermes* colonies often build conspicuous mounds that may harbor several millions of individuals (Noirot and Darlington, 2000; Korb, 2011).

We compare the genomes of these divergent species (**Table 1**) with those of other insects and outline first hypotheses how sociality and ecological factors left their footprints in the genomes.

MATERIALS AND METHODS

CONSTRUCTION OF GENE FAMILIES

To gain insight into the evolution of gene families in termites, we clustered genes from 12 insect genomes (pea aphid: *Acyrtosiphon pisum*: The International Pea Aphid Genomics Consortium, 2010; body louse: *Pediculus humanus*: Kirkness et al., 2010; flour beetle: *Tribolium castaneum*: Richards et al., 2008; fruitfly: *Drosophila melanogaster*: Adams et al., 2000; jewel wasp: *Nasonia vitripennis*: Werren et al., 2010; honeybee: *Apis mellifera*: The Honeybee Genome Sequencing Consortium, 2006; ants: *Acromyrmex echinatior*: Nygaard et al., 2011, *Atta cephalotes*: Suen et al., 2011, *Camponotus floridanus*, *Harpegnathos saltator*: Bonasio et al., 2010; termites: *Z. nevadensis*, *M. natalensis*), the water flea *Daphnia pulex* (Colbourne et al., 2011), and the round worm *Caenorhabditis elegans* (Coulson and C. elegans Genome Consortium, 1996). The gene sets of the species that we chose were downloaded from the Ensembl database (Flicek et al., 2014), except for ants and termites which were downloaded from their own reference databases. Then we used Treefam (Li et al., 2006) to construct gene families. For more information see also Terrapon et al. (2014) and Poulsen et al. (2014) (Table S1).

FUNCTIONAL ANNOTATION OF TERMITE GENES

InterproScan v4.8 (Zdobnov and Apweiler, 2001) was used to annotate motifs and domains of translated proteins in two termites. Protein sequences were searched against SUPERFAMILY, Pfam, PRINTS, PROSITE, ProDom, Gene3D, PANTHER, and SMART databases in Interpro with default parameter settings. GO (gene ontology) terms for each gene were obtained from the Interpro database according to the relationship of GO and Interpro terms. The KEGG annotation (Kanehisa and Goto, 2000)

Table 1 | Summary of traits that differ between the two study species.

Traits	<i>Z. nevadensis</i>	<i>M. natalensis</i>
Social complexity	Less complex	Highly complex
Life type	Wood-dwelling single-piece nester	Foraging multiple-piece nester
Developmental plasticity	Totipotent workers and a single linear developmental pathway	Restricted developmental options for both workers and reproductives; bifurcated development
Food and digestion	Decaying wood, digested with the help of protists and bacterial gut symbionts	Dead plant material (incl. wood), which is primarily decomposed by symbiotic <i>Termitomyces</i> fungi, with additional roles of gut bacteria
Potential pathogen load	Predicted to be high, mainly because the logs inhabited by dampwood termites also harbor many wood-decaying fungi	Predicted to be high, with sources being mainly soil microbes and wood-decaying fungi carried to the nest with the substrate particles
Geographic distribution	Temperate	Tropical and sub-tropical

Traits 1–3 co-vary in termites in that wood-dwelling termites with totipotent workers are always less socially complex, while foraging termites are more socially complex with workers having restricted developmental options. However, huge trait variability exists within foraging species, see also **Figure 1**.

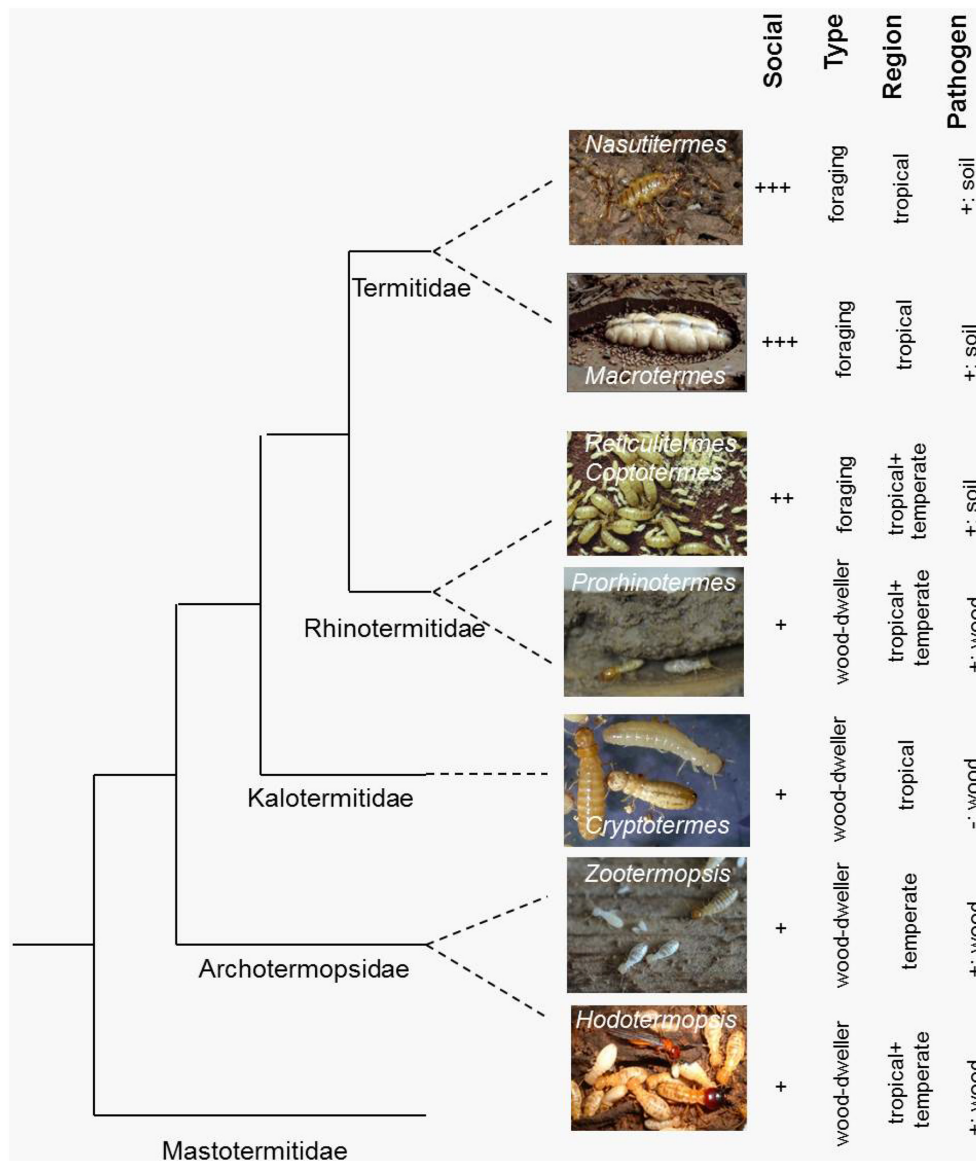


FIGURE 1 | Simplified phylogeny of the main termite study species with their key traits. Shown is a cladogram of termite genera on which some genomic/molecular genetic research has been done. Added to the right are characteristic social and ecological traits. Social: increasing social complexity from + to +++ (e.g., increasing colony size, division of labor, morphological differentiation between castes); Type: life type, foraging vs. wood dwelling; Region: temperate vs. tropical; Pathogens:

soil pathogens vs. wood-decaying fungi, +, present; -, absent. Study species (photo credits): *Nasutitermes takasagoensis* (Kenji Matsuura), *Macrotermes natalensis* (Judith Korb), *Reticulitermes speratus* (Kenji Matsuura), *Reticulitermes flavipes* (not shown), *Coptotermes formosanus* (not shown), *Prorhinotermes simplex* (Judith Korb), *Cryptotermes secundus* (Judith Korb), *Zootermopsis nevadensis* (Judith Korb), *Hodotermes sjostedti* (Toru Miura).

was done via the KAAS online server (Moriya et al., 2007) using the SBH method against the eukaryotic species set.

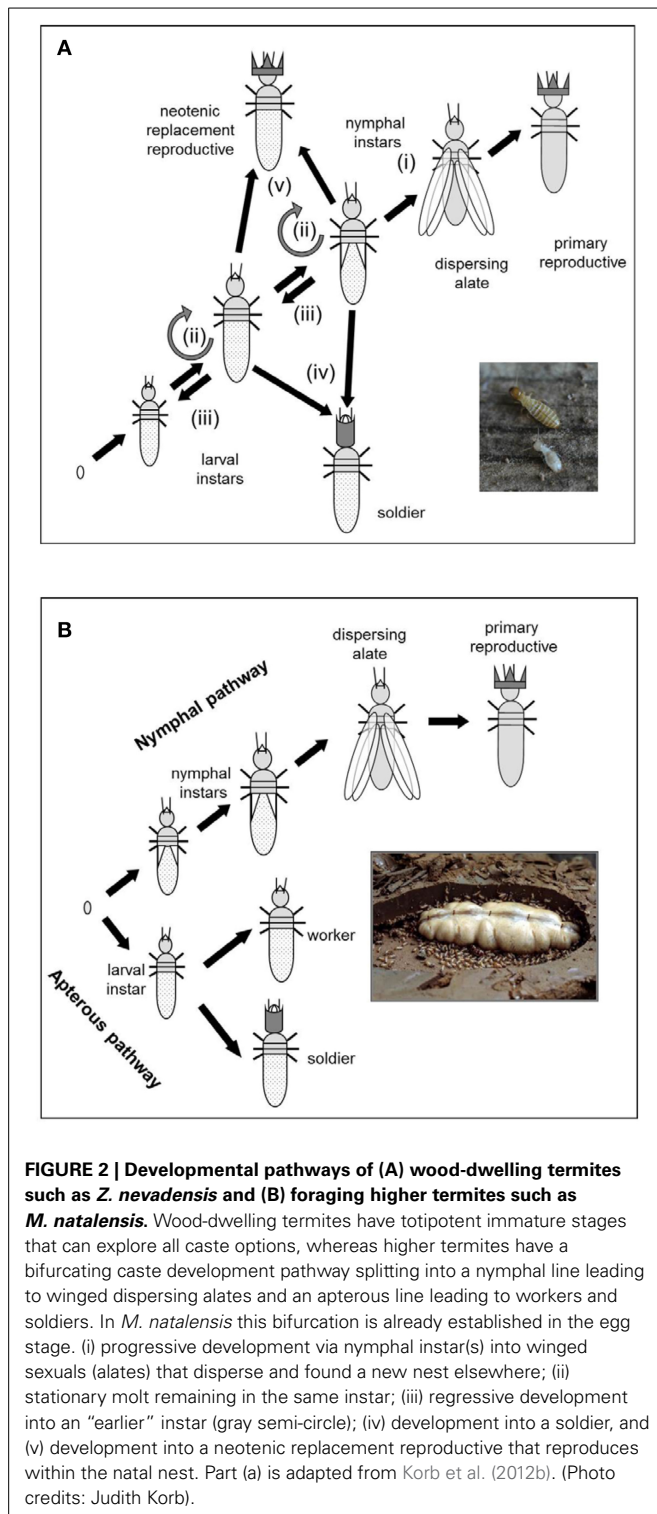
TERMITE-SPECIFIC GENES

Some gene families were termite-specific and absent from the other investigated genomes. For these genes we performed functional enrichment analyses of GO and IPR (Interpro domain) annotation. *P*-values for significant difference were obtained by χ^2 -tests adjusted by FDR (false discovery rate). Similarly, we analyzed differences between the gene sets of *Z. nevadensis* und *M. natalensis* by comparing IPR annotation, KEGG pathways, and

gene families. We constructed gene families for both genomes using Treefam (Li et al., 2006) and tested for differences in gene numbers using χ^2 -tests (or Fisher's exact test for small sample sizes). For gene families that were specific to *M. natalensis* and/or *Z. nevadensis*, we performed IPR enrichment analyses to obtain information on the putative functions of these genes.

REPEAT ANALYSES

We used the *M. natalensis* and *Z. nevadensis* genome assemblies to perform repetitive sequence annotation. First, we did homologous repeat family annotation to identify transposable



elements (TEs) using the TE database Repbase v17.06 (Jurka and Kapitonov, 2005) and the programs RepeatMasker (parameter `-norna`) and RepeatProteinMask v4.0.1 (<http://www.RepeatMasker.org>) (parameter `-p 0.0001`) (Smit et al., 1996–2010). *De-novo* repeat family annotation was done with PILER v1.0 (Edgar and Myers, 2005), LTRfinder v1.05 (Zhao and Wang,

2007) and RepeatModeler v1.05, (<http://www.RepeatMasker.org>) (Smit et al., 1996–2010) using default parameters. TEs identified by PILER were converted into TE families and aligned with Muscle v3.28 (Edgar, 2004) to obtain consensus sequences from the alignments. In order to reduce redundancy in the results of LTRfinder and PILER, an “all against all” BLASTn (*e*-value $1e-5$) was performed. If sequences overlapped for more than 80% we kept the longer TE.

We combined the TE families with the consensus sequences of LTRfinder and PILER together with those identified using RepeatModeler to obtain the final TE sequence library for the two termites. All TE sequences were classified with RepeatClassifier in the RepeatModeler package against Repbase v17.06 (Jurka and Kapitonov, 2005) (Dataset S1). Finally, we used the *de novo* TE library to annotate all TEs in the two genomes and combined the results of homologous TE annotation and the *de novo* annotation. If there were overlapping annotations we kept the longer TE. In addition, we predicted tandem repeats using TRF finder (parameters settings: match = 2, mismatch = 7, delta = 7, PM = 80, PI = 10, Minscore = 50, and MaxPeriod = 12) (Benson, 1999). In total, the non-redundant repetitive sequences accounted for 27.8 and 45.9% of the *Z. nevadensis* and *M. natalensis* genome, respectively (Table 2, Dataset S1).

We also checked for Talua elements in both termite species, SINE elements that were first identified in termites (Luchetti, 2005; Luchetti and Mantovani, 2009). Talua reference sequences (Dataset S1) were mapped to the TE annotations using BLASTn (*e*-value $1e-5$). If the alignment contained more than 50% of the Talua domain, the TE was considered to be a Talua containing TE. In total, we found 1575 and 4385 Talua containing TEs in the *Z. nevadensis* and *M. natalensis* genome, respectively.

RESULTS AND DISCUSSION

GENOME ARCHITECTURE AND REPETITIVE SEQUENCES

A striking difference between ants and termites is that termite genomes are about three times larger (Table S1), which appears to be an ancestral cockroach characteristic (always several Gbs; Koshikawa et al., 2008). Termites actually have smaller genomes than cockroaches and it has been hypothesized that sociality was in fact associated with a reduction in genome size (Koshikawa et al., 2008). Yet the socially more complex *M. natalensis* has a genome size that is more than twice the genome size of *Z. nevadensis* (1.31 Gb vs. 562 Mb), which has the smallest genome known for any termite so far (Koshikawa et al., 2008). On the other hand, ant genome size appears to vary relatively little around an average of 300 Mb, with the largest ant genome published so far being 352 Mb (the red fire ant *Solenopsis invicta*) and smallest genome being 219 Mb (the Argentine ant *Linepithema humile*) (Table S2).

The two termite assemblies covered over 85% of the genomes, so any differences observed are unlikely to be related to the slightly fewer protein coding genes in *Z. nevadensis* (15,876 vs. 16,310 in *M. natalensis*). However, the *M. natalensis* genome contained a much higher proportion of repeat sequences (67.1 vs. 26.0% in *Z. nevadensis*) (Table 2). Subtracting these repeat sequences leads to comparable respective genome sizes of 367 and 365 Mb. Further genomic data will be needed to find out whether these

Table 2 | The number and length of each type of repetitive sequence.

Type	<i>Macrotermes natalensis</i>			<i>Zootermopsis nevadensis</i>		
	Number of repeats	Repeat length (bp)	Percentage of Genome (%)	Number of repeats	Repeat length (bp)	Percentage of Genome (%)
TEs	525,847	118,593,042	10.12	307,278	53,444,656	10.83
LINE	1,027,017	237,020,224	20.22	171,545	32,495,416	6.59
LTR	33,435	6,864,870	0.59	10,625	1,980,023	0.40
Rolling Circle	12,725	3,630,172	0.31	2427	384,875	0.08
SINE	13,624	2,671,925	0.23	109,498	17,763,792	3.60
Unknown	535,062	121,413,841	10.36	115,074	22,629,266	4.59
Other	64	10,006	<0.001	3	185	<0.001
Simple repeat	390,741	40,059,393	3.42	88,333	9,086,992	1.84
Simple repeats	164,090	6504,930	0.55	113,670	4,338,842	0.88
Satellite and tandem repeats	221,634	74677,411	6.37	34,394	11,591,981	2.35
Non-redundant total	2,924,239	537,702,043	45.87	952,847	137,154,152	27.79

Simple repeats are 2–5 bp repetitive units while longer satellite and tandem repeats have 6–40 bp. “Other” includes repeats that do not belong to any of the listed types, such as DNA-viruses or centromeric regions (listed in Table S1).

ca. 365 Mbs represent a kind of “core genome” for termites and whether additional variation in genome size would then only be due to variation in repeat sequences. It will also be interesting to evaluate the first cockroach genomes to see whether their huge genomes (multiple Gbs) are associated with a higher number of coding or repeat sequences. In ants, genome-wide repeat content so far varies between 11.5 and 28.0% (Gadau et al., 2012) and no overall correlation with genome size appears to exist.

The *M. natalensis* genome had almost twice as many TEs (transposable elements) than the *Z. nevadensis* genome (45.9 vs. 27.8%; Table 2) and most of these were LINEs (long interspersed nuclear elements), which accounted for 20% of the *M. natalensis* genome (Table 2). According to the Rebase classification, most LINEs in *M. natalensis* resemble BovB retrotransposons, accounting for 16% of the genome, while LINEs contribute only ca. 3% in *Z. nevadensis* (Table 2). BovBs are relatively well known from vertebrates where they have a patchy distribution in squamates, monotremes, marsupials, ruminants, and several African mammals (Afrotheria), possibly as a consequence of horizontal gene transfer via reptile ticks (Walsh et al., 2013). In ruminants, part of one BovB LINE seems to have been recruited into a functional gene after duplication (Iwashita et al., 2006), but whether similar cooption processes may have occurred in termites remains to be explored.

The *M. natalensis* genome appears to have fewer SINEs (short interspersed nuclear elements) than the *Z. nevadensis* genome (3.6 vs. 0.2%). A new SINE retrotransposon, *Talua*, has recently been described for termites (Luchetti, 2005; Luchetti and Mantovani, 2009). It belongs to a new family of tRNA-derived elements that are very G+C-rich (55–60%) but makes up only a small proportion of the termite genomes (0.25 and 0.19% in *Z. nevadensis* and *M. natalensis*, respectively; Table S3). There are multi-copy TEs that are present in both termite genomes that do not resemble any known TEs. They may thus be novel termite-specific TEs, but additional termite and non-termite genomes will be needed to test this against a null hypothesis of being more general TEs that also occur in other hemimetabolous insects.

TE sequence divergence (i.e., percentage of different base pairs) relative to TE consensus sequences showed a peak at about 25% for both *M. natalensis* and *Z. nevadensis* (Figure 3), but *M. natalensis* had an additional divergence rate peak at ca. 7–8% (Figure 3). This might indicate that the lineage leading to *M. natalensis* has undergone a genome expansion that multiplied TE copies and BovB retrotransposons, which could then explain why the *M. natalensis* genome is so much larger than the *Z. nevadensis* genome.

Consistent with the high prevalence of repeat sequences, IPR annotation results showed a functional enrichment of DNA/RNA cutting genes in termites (Ribonuclease H domain: 22 genes, Ribonuclease H-like domain: 26 genes, endonuclease/exonuclease/phosphatase: 26 genes) compared to other insects (Table S4). Strikingly, *M. natalensis* had at least twice as many of such transposon-related genes than *Z. nevadensis*, supporting the idea that selfish replicating elements played a major role in the evolution of termite genome architecture and size (Tables S5, S6).

Cluster analyses of caste-specific transcriptomes in *Z. nevadensis* revealed that several of these DNA/RNA-cutting genes are overexpressed in the nymphal stages (i.e., instars with wing buds) compared to all other stages and castes (Terrapon et al., 2014). Nymphs are individuals destined to develop into winged dispersing reproductives, suggesting that TE activity might be linked to maturation processes such as gonad development. Such functions remain speculative at this point, but would be consistent with TEs having been coopted to fulfill host functions and to play fundamental roles in epigenetic regulation in organisms as different as *Arabidopsis thaliana* plants, *Caenorhabditis elegans* worms, *Drosophila melanogaster* flies and *Mus musculus* house mice (e.g., Lippman et al., 2004; Slotkin and Martienssen, 2007; Fedoroff, 2012). Silenced TEs are often activated through stressful environmental conditions (Slotkin and Martienssen, 2007; Fedoroff, 2012). In wood-dwelling termites, such conditions may arise by reduced food availability or possibly parasite pressure inducing higher rates of nymphal (sexual dispersing) development (Lenz, 1994; Korb and Schmidinger, 2004; Korb and Fuchs,

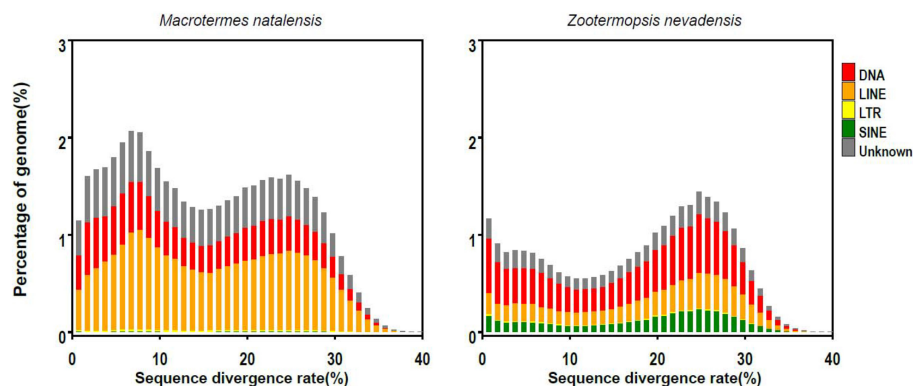


FIGURE 3 | The distribution of sequence divergence rates of transposable elements (TEs) as percentages of the genome size of *M. natalensis* (left) and *Z. nevadensis* (right).

2006). Hence, it may be interesting to test whether a similar link between TE activity and stressful conditions exists during nymphal development.

Whether TEs can also be linked with epigenetic regulation of gene expression through DNA methylation (Lippman et al., 2004; Slotkin and Martienssen, 2007; Fedoroff, 2012) remains to be seen. DNA methylation has been proposed to regulate caste differentiation (Kucharski et al., 2008; Elango et al., 2009; Gadau et al., 2012; Terrapon et al., 2014) and the complete epigenetic toolbox was indeed identified in *Z. nevadensis* with orthologs of *DNMT1* and *DNMT3* (Terrapon et al., 2014). However, in *M. natalensis* only *DNMT1* (and possibly *DNMT2*) could be confirmed, but not *DNMT3*.

COMMUNICATION

Termite-specific expansions for gene families were also found among chemoperception genes that are important for communication (Table S4). Given the disparate social systems of *Z. nevadensis* and *M. natalensis*, differences in expansions of such genes may be related to divergent communication systems. Chemoperception genes mainly comprise four families: Odorant receptors (ORs), gustatory receptors (GRs), ionotropic receptors (IRs), and odorant binding proteins. ORs mostly control for the specificity and sensitivity of insect olfaction. GRs are primarily involved in contact chemoperception and IRs belong to a recently discovered gene family for olfaction and gustation in *Drosophila* (Benton et al., 2009; Grosjean et al., 2011; Rytz et al., 2013). Odorant binding proteins primarily shuttle such compounds through the hydrophilic environment of the sensory lymph to the receptors.

The IR family is most consistently expanded in *Z. nevadensis*, representing the highest known value in insects (Terrapon et al., 2014). This IR number was between 4 and 10-fold higher in *Z. nevadensis* than in eusocial Hymenopterans, but the 80 intact GR genes remained within the overall range of 10–97 known from ants and honeybees (Zhou et al., 2012). The number of OR genes in *Z. nevadensis* was between one third and one half of the numbers normally found in the ants (Zhou et al., 2012), consistent with the lifestyle of wood-dwelling termites likely requiring lower levels of olfactory communication.

Overall, we found termite-specific enrichment in all four major gene families relating to olfaction (Table S4). Most IPR enrichment occurred in the ionotropic glutamate receptors that include IR genes (21). Significant enrichment was also found in ORs (7), GRs (7 TM chemoreceptor: 7), and various odorant-binding proteins (9, 7, 5). Direct comparison between *Z. nevadensis* and *M. natalensis* (Table S6) showed that *Z. nevadensis* had significantly more genes related to chemical communication than *M. natalensis* (Table S7). However, chemoperception genes are notoriously difficult to assemble and annotate (Terrapon et al., 2014), so this difference should be considered with caution, also because these genes were manually annotated in *Z. nevadensis* (with support from antennal RNAseq data), but automatically in *M. natalensis*. More work will therefore be needed before solid conclusions on the relative role of ORs in different termite species can be drawn.

IMMUNE DEFENSES

Both termite species live in potentially pathogen-rich habitats. *Z. nevadensis* nests in decaying wood with abundant fungal growth that has probably selected for intensive allogrooming behaviors (Korb et al., 2012a). Also *M. natalensis* is potentially exposed to many pathogens both from its soil-nesting habitat and across its foraging range. *Macrotermes* species are known to protect their *Termitomyces* fungal symbiont from being overgrown by other fungi (Nobre et al., 2011) and termite-specific antimicrobial peptides (AMPs) have been described in another genus of fungus-growing termites (Lamberty et al., 2001).

Relative to ants and other insects, we did not find enrichments for immune defense genes in the two termite genomes and neither were there substantial differences between the two termite genomes (Tables S4, S6). All of the immune-related pathways, including pattern recognition, signaling, and gene regulation (as described for *Drosophila melanogaster* and other insects; Hoffmann, 2003; Hultmark, 2003; Schmid-Hempel, 2005) are present in both termite genomes (Table S8). Only two differences are noteworthy (Table 3). First, *Z. nevadensis* has 6 gram-negative binding proteins (GNBPs), whereas only four of these were recovered in *M. natalensis*. These four GNBPs are all termite-specific (Figure 4) and some of them were previously shown

to be under positive selection in several *Nasutitermes* species, especially in species with arboreal nests (Bulmer and Crozier, 2006). The *Macrotermes* genome seems to lack the insect-typical GGBP duplicate and one GGBP gene that has so far only been found in *Z. nevadensis* (Figure 4). Second, while AMPs were not enriched in either termite genome (Table S4), their identities were completely different with *Z. nevadensis* having 2 AMPs and *M. natalensis* having 3 other AMPs (Table 3, Table S8). *M. natalensis* has a termite-specific defensin-like gene termicin, a category of genes that seem to have duplicated repeatedly during the radiation of *Nasutitermes* termites (Bulmer and Crozier, 2004). After duplication, one copy seems to often be under strong selection, while the other evolves toward neutrality (Bulmer and Crozier, 2004; Bulmer et al., 2010). Also in the soil-foraging *Reticulitermes* species these genes seem to be under positive selection (Bulmer et al., 2010).

In contrast to other insects where AMP production is normally induced, these genes seem to be constitutively expressed in fungus-growing termites, as has been shown for *Pseudacanthotermes spiniger* (Lamberty et al., 2001), which might be an adaptation to protect the symbiont against competing fungi. Termicin and other defensins (Table S8) were absent in *Z. nevadensis* but this species has GNBPs that are differentially expressed between castes (Terrapon et al., 2014) and may thus serve a similar function in protecting the nest from fungal infections. For the arboreal nesting termite *Nasutitermes corniger* it has been shown that GGBP2 has (1,3)-glucanase effector activity and functions as an antifungal agent (Bulmer et al., 2009). It is incorporated in the nest building material, where it cleaves and releases pathogenic components while priming termites for improved antimicrobial defense (Bulmer et al., 2009). Such a defensive strategy is likely to be most effective for termites with closed nests, consistent with positive selection on GGBP being most pronounced in *Nasutitermes* that live in arboreal nests (Bulmer and Crozier, 2006). Hence, antifungal strategies might differ in termites with different habitats; with GNBPs and termicin possibly playing complementary roles. This is supported by the fact that GNBPs in subterranean, foraging *Reticulitermes* species evolve neutrally while termicin was shown to have been under strong positive selection in these species (Table 3).

We can reject the possible alternative hypothesis that different defense strategies are linked to the gut symbionts that need different defense strategies to protect the symbiotic partner. As lower termites harbor protists as well as bacteria, while higher termites

only have bacteria, we would then have expected higher termites having more AMPs and lower termites more GNBPs, but this is not the case because lower *Reticulitermes* termites have positively selected termicins. If there is an association between nesting habit and defense strategy, we expect that GNBPs are under positive selection in other wood-dwelling termites, and termicins are selected in soil-foraging termites. Additional genomic data, particularly for wood-dwelling termites, would be needed to validate this hypothesis.

Reduced numbers of immune defense genes were found in ants and the honeybee (Evans et al., 2006; Gadau et al., 2012) but also here there seems to be selection on some of the AMP genes. Similar to termicin, positive selection was detected on defensin in ants (Viljakainen and Pamilo, 2008), but this gene was not overexpressed after experimental fungal infections of leaf-cutting ant colonies, whereas two other AMPs were (Yek et al., 2013). This contrasts with dipterans (*Drosophila* and *Anopheles*) for which no evidence was found for positive selection on any AMPs (Sackton et al., 2007; Simard et al., 2007), but instead for immune recognition and signaling proteins (Schlenke and Begun, 2003; Jiggins and Kim, 2005; Sackton et al., 2007). This provides further support for the hypothesis that social insects have responded differently to selection pressure caused by microbial pathogens than solitary insects (Viljakainen and Pamilo, 2008).

MATING BIOLOGY

Compared to *M. natalensis*, the *Z. nevadensis* genome is enriched in genes that are related to male fertility/spermatogenesis (e.g., *KLHL10*) (Table 4, Table S7). This suggests that the co-expansion (and co-expression) of these genes in *Z. nevadensis* is not typical for termite sociality but rather taxon-specific. It might be linked to the seasonal reproduction of this temperate zone species where spermatogenesis is cyclically switched on and off, which contrasts with tropical *Macrotermes* males that produce offspring all year round. However, some members of two spermatogenesis-related gene families, seven-in-absentia (SINA) proteins and α -tubulins, do not show *Z. nevadensis*-specific expansions.

An alternative evolutionary explanation could be that males of wood-dwelling termites have low but consistent probabilities to face sperm competition when neighboring colonies merge after colony foundation. Such mergers are impossible in foraging termites where unrelated males never compete for inseminating the

Table 3 | Gram-negative binding proteins (GNBPs) and anti-microbial-peptide (termicin) genes known from different termites.

Species	GGBP	Termicin	References
<i>Z. nevadensis</i>	6 copies	0 copies	Terrapon et al., 2014
<i>Reticulitermes</i> sp.	Neutral	Positive selection	Bulmer et al., 2010
<i>Nasutitermes</i> spp. (Australia)	Positive selection in some species	Positive selection	Bulmer and Crozier, 2004, 2006
<i>Nasutitermes corniger</i>	Antifungal	?	Bulmer et al., 2009
<i>Pseudacanthotermes spiniger</i>	?	Yes	Lamberty et al., 2001
<i>M. natalensis</i>	4 copies	1 copy	Poulsen et al., 2014, this study

GNBPs and termicins might serve complementary roles in fungal defense in termites. GNBPs might be more important in species with closed nests, whereas termicins seem to be under strong positive selection in foraging termites with subterranean nests. ?, unknown.

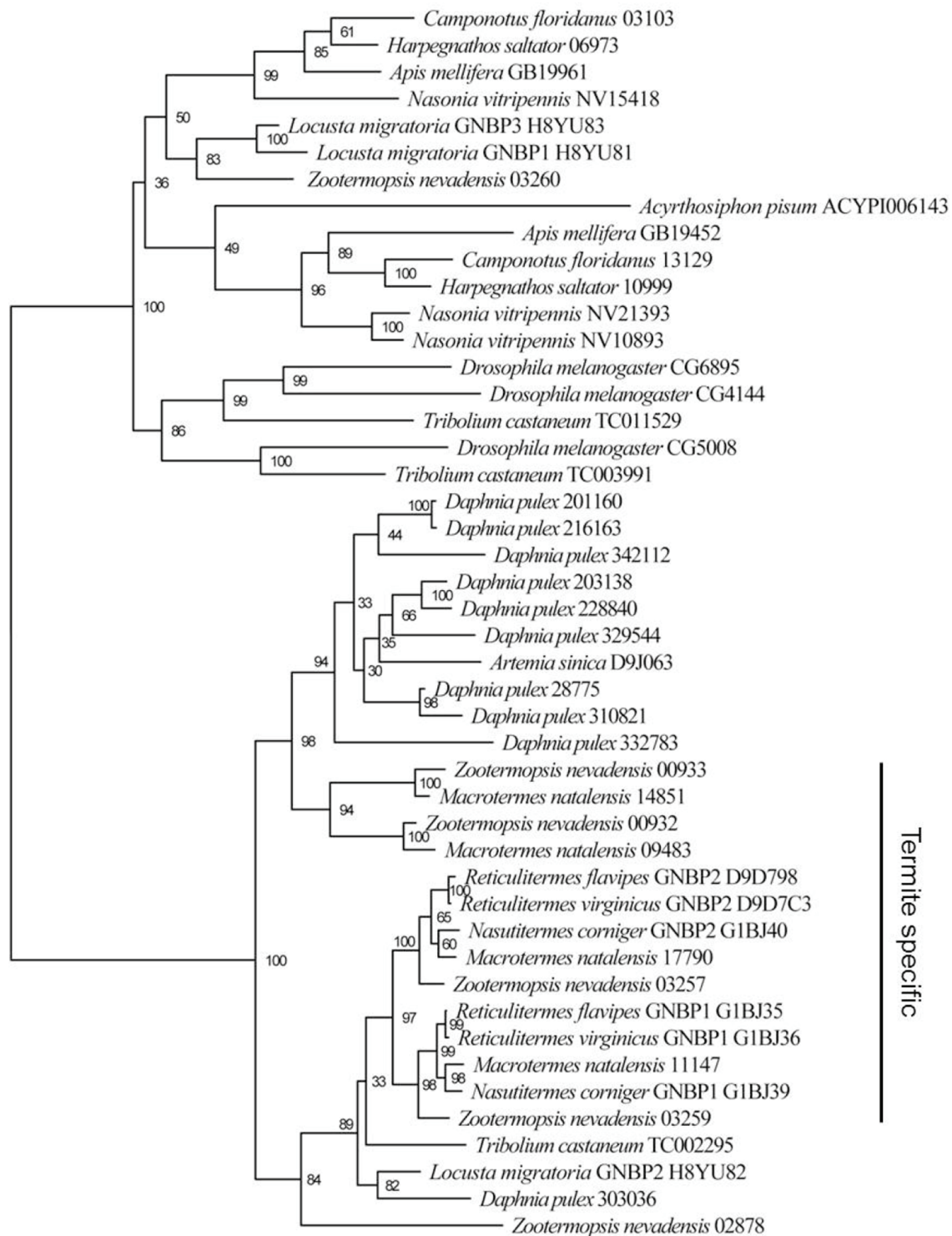


FIGURE 4 | Phylogeny of gram-negative binding proteins (GNBPs) constructed with PhylML v3.0 (LG substitution model with 100 bootstrap replicates) after alignment of the peptide sequences in ClustalW2.

same queen (Boomsma, 2013). This hypothesis would predict no difference between temperate and tropical wood-dwelling termites, but a series of termite genomes will be needed to test these contentions.

SYMBIOSIS

The ancestral termite gut microbiota was derived from a cockroach ancestor, but major subsequent changes occurred, most notably when the higher termites evolved (Dietrich et al., 2014).

The guts of the wood-dwelling termites are dominated by protists that appear to be primarily adapted to break down wood (Cleveland, 1923; Brugerolle and Radek, 2006), with complementary roles of bacteria that are often symbiotic with the gut-flagellates (Dietrich et al., 2014). The common ancestor of the evolutionarily derived Termitidae lost these flagellate symbionts so their gut microbiotas became dominated by bacteria, which may have facilitated their dietary diversification (Brune and Ohkuma, 2011; Dietrich et al., 2014). The single origin of fungiculture by the Macrotermitinae led to *Termitomyces* taking over primary plant decomposition and the gut microbiota shifting phylogenetically and functionally to perform complementary roles (Liu et al., 2013; Dietrich et al., 2014; Otani et al., 2014; Poulsen et al., 2014).

Changes in symbiont associations are tightly associated with termite life styles (for a recent review on termite gut symbionts, see Brune, 2014), but this may hardly induce structural genomic changes in the termite hosts, consistent with the similar gene repertoires for plant biomass decomposition found in the two termite genomes (Poulsen et al., 2014). A comparison of carbohydrate-active enzyme (CAZy) profiles of the two termite species showed a reduction in the absolute number of glycoside hydrolase enzymes (85) in *M. natalensis* compared to *Z. nevadensis* (97) (Table 5), but very similar relative abundances of specific enzyme families (Poulsen et al., 2014). Profile similarities suggest that plant-biomass decomposition genes may be ancestrally conserved across the termites, but additional termite genomes are needed to shed light on this. Such additional genomic work will need to be accompanied by enzyme function validations to test whether differences in absolute numbers reflect changes in the relative importance of termite-derived enzymes.

CONCLUSION

Despite the striking differences in social complexity between *Z. nevadensis* and *M. natalensis* we did not find major differences in gene composition. The gene families underlying chemical communication seem not to be expanded in the more complex fungus-growing termite compared to *Z. nevadensis*. The major differences between the two termite genomes are related to genome architecture and the presence of transposons that can explain the much larger genome size of *M. natalensis*. Whether these ancestrally selfish elements have been domesticated for functions related to the increased social complexity of *M. natalensis* needs further work.

Table 4 | Number of genes related to spermatogenesis in *Z. nevadensis* and *M. natalensis* based on Pfam domains.

Protein families	<i>Z. nevadensis</i>	<i>M. natalensis</i>
BTB-BACK-Kelch (KLHL10)	37	10
Kelch (KLHL1)	20	2
BTB+KELCH	6	1
BACK+KELCH	4	0
SINA (Seven-in-absentia)	33	17
Alpha tubulin	13	8
PKD (polycystin)	10	1

Our comparison allowed us to generate hypotheses that can be tested with functional genomic studies and with more advanced comparative analyses as more termite genomes become available.

We have highlighted the contours of further testable predictions concerning TE number and genome size, male fertility, and habitat-specific disease pressure. For any next termite genome to be sequenced (Figure 1), authors should ask questions like: (1) Is the habitat of this (e.g., drywood) termite more disease-ridden than the habitat of a comparable dampwood termite such as *Z. nevadensis*? (2) Would this tropical new wood-dwelling termite have similar gene family expansions for male fertility as *Z. nevadensis*? (3) Has this arboreal higher (e.g., *Nasutitermes*) termite lost specific immune defenses that match the disease pressure of its habitat and is it equally burdened by TEs as *Macrotermes natalensis*?

While two genomes are a major achievement in some sense, these genomes also leave us with insufficient resolution to move much beyond the crude comparisons that we offer in this paper, because *Z. nevadensis* and *M. natalensis* differ in too many evolutionary and ecological factors (Table 1). It has also become clear

Table 5 | Number of glycoside hydrolases of different GH families identified in *Z. nevadensis* and *M. natalensis* (from Table S28; Poulsen et al., 2014).

CAZy family	<i>M. natalensis</i>	<i>Z. nevadensis</i>
GH1	11	7
GH2	5	4
GH9	4	6
GH13	8	9
GH15	1	1
GH16	4	5
GH18	12	14
GH20	6	8
GH22	3	3
GH27	1	2
GH29	1	2
GH30	2	2
GH31	4	6
GH35	2	1
GH37	3	3
GH38	3	3
GH39	1	1
GH47	4	5
GH56	1	1
GH63	1	1
GH74	1	1
GH79	1	2
GH84		1
GH85	1	1
GH89	1	1
GH99	1	1
GH109	2	5
GH119	1	1
Total	85	97

from comparative ant genomics that gene expression mechanisms may be more informative than structural gene differences (Simola et al., 2013). Finally, apart from obtaining more termite genomes and population genomic studies on gene expression and signatures of selection, it will also be crucially important to obtain a *Cryptocercus* cockroach sister lineage genome and more distant outgroup genomes for non-social hemimetabolous insects. Many surprises will likely be waiting in the wings, as both the pea aphid and the body louse genomes turned out to be unusual because of the specialized feeding habits of these insects with or without symbionts.

ACKNOWLEDGMENTS

Judith Korb was supported by research grant from the German Science Foundation (DFG; KO1895/6), Michael Poulsen by a STENO grant from The Danish Council for Independent Research Natural Sciences, Jacobus J. Boomsma by a Danish National Research Foundation grant (DNRF57), Guojie Zhang by a Marie Curie International Incoming Fellowship (300837), and Jürgen Liebig by the Agriculture and Food Research Initiative (2007-35302-18172 to Jürgen Liebig and Colin S. Brent). We thank the three referees for helpful comments and the editors Jürgen Gadau and Greg Hunt for inviting us to contribute to this special journal issue.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2015.00009/abstract>

REFERENCES

- Abe, T. (1987). "Evolution of life types in termites," in *Evolution and Coadaptation in Biotic Communities*, eds S. Kawano, J. H. Connell, and T. Hidaka (Tokyo: University of Tokyo Press).
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195. doi: 10.1126/science.287.5461.2185
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573
- Benton, R., Vannice, K. S., Gomez-Diaz, C., and Vosshall, L. B. (2009). Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* 136, 149–162. doi: 10.1016/j.cell.2008.12.001
- Bonasio, R., Zhang, G., Ye, C., Mutti, N. S., Fang, X., Qin, N., et al. (2010). Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* 329, 1068–1071. doi: 10.1126/science.1192428
- Boomsma, J. J. (2013). Beyond promiscuity: mate choice commitments in social breeding. *Philos. Trans. R. Soc. B* 368:20120050. doi: 10.1098/rstb.2012.0050
- Boomsma, J. J., Schmid-Hempel, P., and Hughes, W. O. H. (2005). "Life histories and parasite pressure across the major groups of social insects," in *Insect Evolutionary Ecology*, eds M. D. E. Fellowes, G. J. Holloway, and J. Rolff (Wallingford, CT: CABI), 139–175.
- Brugerolle, G., and Radek, R. (2006). "Symbiotic protozoa of termites," in *Intestinal Microorganisms of Termites and Other Invertebrates*, eds H. König and A. Varma (Berlin, Heidelberg: Springer), 243–269. doi: 10.1007/3-540-28185-1_10
- Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12, 168–180. doi: 10.1038/nrmicro3182
- Brune, A., and Ohkuma, M. (2011). "Role of the termite gut microbiota in symbiotic digestion," in *Biology of Termites: A Modern Synthesis*, eds D. E. Bignell, Y. Roisin, and N. Lo (Dordrecht; Heidelberg; London; New York: Springer), 439–476.
- Bulmer, M. S., Bachelet, I., Raman, R., Rosengaus, R. B., and Sasisekharan, R. (2009). Targeting an antimicrobial effector function in insect immunity as a pest control strategy. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12652–12657. doi: 10.1073/pnas.0904063106
- Bulmer, M. S., and Crozier, R. H. (2004). Duplication and diversifying selection among termite antifungal peptides. *Mol. Biol. Evol.* 21, 2256–2264. doi: 10.1093/molbev/msh236
- Bulmer, M. S., and Crozier, R. H. (2006). Variation in positive selection in termite GNBPs and Relish. *Mol. Biol. Evol.* 23, 317–326. doi: 10.1093/molbev/msj037
- Bulmer, M. S., Lay, F., and Hamilton, C. (2010). Adaptive evolution in subterranean termite antifungal peptides. *Insect Mol. Biol.* 19, 669–674. doi: 10.1111/j.1365-2583.2010.01023.x
- Cleveland, L. R. (1923). Symbiosis between termites and their intestinal protozoa. *Proc. Natl. Acad. Sci. U.S.A.* 9, 424–428. doi: 10.1073/pnas.9.12.424
- Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., et al. (2011). The ecoresponsive genome of *Daphnia pulex*. *Science* 331, 555–561. doi: 10.1126/science.1197761
- Coulson, A., and C. elegans Genome Consortium (1996). The *Caenorhabditis elegans* genome project. *Biochem. Soc. Trans.* 24, 289.
- Cremer, S., Armitage, S. A. O., and Schmid-Hempel, P. (2007). Social immunity. *Curr. Biol.* 17, R693–R702. doi: 10.1016/j.cub.2007.06.008
- Dietrich, C., Köhler, T., and Brune, A. (2014). The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl. Environ. Microb.* 80, 2261–2269. doi: 10.1128/AEM.04206-13
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Edgar, R. C., and Myers, E. W. (2005). PILER: identification and classification of genomic repeats. *Bioinformatics* 21, i152–i158. doi: 10.1093/bioinformatics/bti1003
- Elango, N., Hunt, B. G., Goodisman, M. A. D., and Yi, S. V. (2009). DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11206–11211. doi: 10.1073/pnas.0900301106
- Engel, M. S., Grimaldi, D. A., and Krishna, K. (2009). Termites (Isoptera): their phylogeny, classification, and rise to ecological dominance. *Am. Museum Novitates* 3650, 1–27. doi: 10.1206/651.1
- Evans, J. D., Aronstein, K., Chen, Y. P., Hetru, C., Imler, J. L., Jiang, H., et al. (2006). Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Mol. Biol.* 15, 645–656. doi: 10.1111/j.1365-2583.2006.00682.x
- Fedoroff, N. V. (2012). Transposable elements, epigenetics and genome evolution. *Science* 338, 758–767. doi: 10.1126/science.338.6108.758
- Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., et al. (2014). Ensembl 2014. *Nucleic Acids Res.* 42, D749–D755. doi: 10.1093/nar/gkt1196
- Gadau, J., Helmkampf, M., Nygaard, S., Roux, J., Simola, D. F., Smith, C. R., et al. (2012). The genomic impact of 100 million years of social evolution in seven ant species. *Trends Genet.* 28, 14–21. doi: 10.1016/j.tig.2011.08.005
- Grosjean, Y., Rytz, R., Farine, J. P., Abuin, L., Cortot, J., Jefferis, G., et al. (2011). An olfactory receptor for food-derived odours promotes male courtship in *Drosophila*. *Nature* 478, 236–240. doi: 10.1038/nature10428
- Hoffmann, J. A. (2003). The immune response of *Drosophila*. *Nature* 426, 33–38. doi: 10.1038/nature02021
- Howard, K. J., and Thorne, B. L. (2011). "Eusocial evolution in termites and hymenoptera," in *Biology of Termites: A Modern Synthesis*, eds D. E. Bignell, Y. Roisin, and N. Lo (Dordrecht; Heidelberg; London; New York: Springer), 97–132.
- Hughes, W. O. H., Oldroyd, B. P., Beekman, M., and Ratnieks, F. L. W. (2008). Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* 320, 1212–1216. doi: 10.1126/science.1156108
- Hultmark, D. (2003). *Drosophila* immunity: paths and patterns. *Curr. Opin. Immunol.* 15, 12–19. doi: 10.1016/S0952-7915(02)00005-5
- Inward, D., Beccaloni, G., and Eggleton, P. (2007a). Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett.* 3, 331–335. doi: 10.1098/rsbl.2007.0102
- Inward, D. J. G., Vogler, A. P., and Eggleton, P. (2007b). A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. *Mol. Phylogenet. Evol.* 44, 953–967. doi: 10.1016/j.ympev.2007.05.014
- Iwashita, S., Ueno, S., Nakashima, K., Song, S. Y., Ohshima, K., Tanaka, K., et al. (2006). A tandem gene duplication followed by recruitment of a

- retrotransposon created the paralogous bucentaur gene (bcnt) in the ancestral ruminant *Mol. Biol. Evol.* 23, 798–806. doi: 10.1093/molbev/msj088
- Jiggins, F. M., and Kim, K.-W. (2005). The evolution of antifungal peptides in *Drosophila*. *Genetics* 171, 1847–1859. doi: 10.1534/genetics.105.045435
- Jurka, J., and Kapitonov, V. V. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110, 462–467. doi: 10.1159/000084979
- Kambhampati, S., and Eggleton, P. (2000). “Taxonomy and phylogeny of termites,” in *Termites: Evolution, Sociality, Symbiosis and Ecology*, eds T. Abe, D. E. Bignell, and M. Higashi (Dordrecht: Kluwer Academic Publishers), 1–23. doi: 10.1007/978-94-017-3223-9_1
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kirkness, E. F., Haas, B. J., Sun, W. L., Braig, H. R., Perotti, M. A., Clark, J. M., et al. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12168–12173. doi: 10.1073/pnas.1003379107
- Korb, J. (2007). Termites. *Curr. Biol.* 17, R995–R999. doi: 10.1016/j.cub.2007.10.033
- Korb, J. (2008). Termites, hemimetabolous diploid white ants? *Front. Zool.* 5:e15. doi: 10.1186/1742-9994-5-15
- Korb, J. (2011). “Termite mound architecture, from function to construction,” in *Biology of Termites: A Modern Synthesis*, eds D. E. Bignell, Y. Roisin, and N. Lo (Dordrecht: Heidelberg; London; New York: Springer), 349–376.
- Korb, J., Buschmann, M., Schafberg, S., Liebig, J., and Bagnères, A. G. (2012a). Brood care and social evolution in termites. *Proc. R. Soc. B* 279, 2662–2671. doi: 10.1098/rspb.2011.2639
- Korb, J., and Fuchs, A. (2006). Termites and mites - adaptive behavioural responses to infestation? *Behaviour* 143, 891–907. doi: 10.1163/156853906778017917
- Korb, J., and Hartfelder, K. (2008). Life history and development - a framework for understanding the ample developmental plasticity in lower termites. *Biol. Rev.* 83, 295–313. doi: 10.1111/j.1469-185X.2008.00044.x
- Korb, J., Hoffmann, K., and Hartfelder, K. (2012b). Molting dynamics and juvenile hormone titer profiles in the nymphal stages of a lower termite, *Cryptotermes secundus* (Kalotermitidae) - signatures of developmental plasticity. *J. Insect Physiol.* 58, 376–383. doi: 10.1016/j.jinsphys.2011.12.016
- Korb, J., and Schmidinger, S. (2004). Help or disperse? Cooperation in termites influenced by food conditions. *Behav. Ecol. Sociobiol.* 56, 89–95. doi: 10.1007/s00265-004-0757-x
- Koshikawa, S., Miyazaki, S., Cornette, R., Matsumoto, T., and Miura, T. (2008). Genome size of termites (Insecta, Dictyoptera, Isoptera) and wood roaches (Insecta, Dictyoptera, Cryptocercidae). *Naturwissenschaften* 95, 859–867. doi: 10.1007/s00114-008-0395-7
- Krishna, K., Grimaldi, D. A., Krishna, V., and Engel, M. S. (2013). Treatise on the Isoptera of the world. *Bull. Am. Mus. Nat. Hist.* 377, 1–2704. doi: 10.1206/377.1
- Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319, 1827–1830. doi: 10.1126/science.1153069
- Lamberty, M., Zachary, D., Lanot, R., Bordereau, C., Robert, A., Hoffmann, J. A., et al. (2001). Insect immunity - Constitutive expression of a cysteine-rich antifungal and a linear antibacterial peptide in a termite insect. *J. Biol. Chem.* 276, 4085–4092. doi: 10.1074/jbc.M002998200
- Lenz, M. (1994). “Food resources, colony growth and caste development in wood-feeding termites,” in *Nourishment and Evolution in Insect Societies*, eds J. Hunt and C. A. Nalepa (New Delhi: Oxford and IBH Publishing Co. Pvt. Ltd), 159–209.
- Li, H., Coghlan, A., Ruan, J., Coin, L. J., Heriche, J. K., Osmotherly, L., et al. (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* 34, 572–580. doi: 10.1093/nar/gkj118
- Liebig, J. (2010). “Hydrocarbon profiles indicate fertility and dominance status in ant, bee, and wasp colonies,” in *Insect Hydrocarbons. Biology, Biochemistry, and Chemical Ecology*, eds G. J. Blomquist and A. G. Bagnères (Cambridge: Cambridge UP), 254–281. doi: 10.1017/CBO9780511711909.014
- Liebig, J., Eliyahu, D., and Brent, C. S. (2009). Cuticular hydrocarbon profiles indicate reproductive status in the termite *Zootermopsis nevadensis*. *Behav. Ecol. Sociobiol.* 63, 1799–1807. doi: 10.1007/s00265-009-0807-5
- Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M. W., Dedhia, N., McCombie, W. R., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430, 471–473. doi: 10.1038/nature02651
- Liu, N., Zhang, L., Zhou, H., Zhang, M., Yan, X., Wang, Q., et al. (2013). Metagenomic insights into metabolic capacities of the gut microbiota in a fungus-cultivating termite (*Odontotermes yunnanensis*). *PLoS ONE* 8:e69184. doi: 10.1371/journal.pone.0069184
- Luchetti, A. (2005). Identification of a short interspersed repeat in the *Reticulitermes lucifugus* (Isoptera, Rhinotermitidae) genome. *DNA Seq.* 15, 304–307. doi: 10.1080/10425170500061467
- Luchetti, A., and Mantovani, B. (2009). Talua SINE biology in the genome of the *Reticulitermes* subterranean termites (Isoptera, Rhinotermitidae). *J. Mol. Evol.* 69, 589–600. doi: 10.1007/s00239-009-9285-7
- Miura, T., and Scharf, M. E. (2011). “Molecular basis underlying caste differentiation in termites,” in *Biology of Termites: A Modern Synthesis*, eds D. E. Bignell, Y. Roisin, and N. Lo (Dordrecht; Heidelberg; London; New York: Springer), 211–254.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321
- Nobre, T., Rouland-Lefevre, C., and Aanen, D. K. (2011). “Comparative biology of fungus cultivation in termites and ants,” in *Biology of Termites: A Modern Synthesis*, eds D. E. Bignell, Y. Roisin, and N. Lo (Dordrecht; Heidelberg; London; New York: Springer), 193–210.
- Noirot, C., and Darlington, J. P. E. C. (2000). “Termite nests: architecture, regulation and defence,” in *Termites: Evolution, Sociality, Symbiosis and Ecology*, eds T. Abe, D. E. Bignell, and M. Higashi (Dordrecht: Kluwer Academic Publishers), 121–139. doi: 10.1007/978-94-017-3223-9_6
- Noirot, C., and Pasteels, J. M. (1987). Ontogenic development and evolution of the worker caste in termites. *Experientia* 43, 851–860. doi: 10.1007/BF01951642
- Noirot, C., and Pasteels, J. M. (1988). The worker caste is polyphyletic in termites. *Sociobiology* 14, 15–20.
- Nygaard, S., Zhang, G., Schiott, M., Li, C., Wurm, Y., Hu, H., et al. (2011). The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Res.* 21, 1339–1348. doi: 10.1101/gr.121392.111
- Okot-Kotber, B. M. (1985). “Mechanisms of caste determination in a higher termite, *Macrotermes michaelseni* (Isoptera, Macrotermitidae),” in *Caste Differentiation in Social Insects*, eds J. A. L. Watson, B. M. Okot-Kotber, and C. Noirot (Oxford: Pergamon Press), 267–306. doi: 10.1016/B978-0-08-030783-1.50026-4
- Otani, S., Mikaelyan, A., Nobre, T., Hansen, L. H., Koné, N. A., Sørensen, S. J., et al. (2014). Identifying the core microbial community in the gut of fungus-growing termites. *Mol. Ecol.* 23, 4631–4644. doi: 10.1111/mec.12874
- Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S., et al. (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14500–14505. doi: 10.1073/pnas.1319718111
- Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S. J., Denell, R., Beeman, R. W., et al. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452, 949–955. doi: 10.1038/nature06784
- Roisin, Y. (2000). “Diversity and evolution of caste patterns,” in *Diversity and Evolution of Caste Patterns*, eds T. Abe, D. E. Bignell, and M. Higashi (Dordrecht: Kluwer Academic Publishers), 95–119.
- Rosengaus, R. B., Traniello, J. F. A., and Bulmer, M. S. (2011). “Ecology, behavior and evolution of disease resistance in termites,” in *Biology of Termites: A Modern Synthesis*, eds D. E. Bignell, Y. Roisin, and L. Lo (Dordrecht; Heidelberg; London; New York: Springer), 165–192.
- Ruelle, J. E. (1970). A revision of the termites of the genus *Macrotermes* from the Ethiopian Region (Isoptera: Termitidae). *Bull. Br. Mus. (Nat. Hist.) Entomol.* 24, 365–444.
- Rytz, R., Croset, V., and Benton, R. (2013). Ionotropic Receptors (IRs): chemosensory ionotropic glutamate receptors in *Drosophila* and beyond. *Insect Biochem. Mol. Biol.* 43, 888–897. doi: 10.1016/j.ibmb.2013.02.007
- Sackton, T. B., Lazzaro, B. P., Schlenke, T. A., Evans, J. D., Hultmark, D., and Clark, A. G. (2007). Dynamic evolution in the innate immune system in *Drosophila*. *Nat. Genet.* 39, 1461–1468. doi: 10.1038/ng.2007.60
- Schlenke, T. A., and Begun, D. J. (2003). Natural selection drives *Drosophila* immune system evolution. *Genetics* 164, 1471–1480.
- Schmid-Hempel, P. (1998). *Parasites in Social Insects*. Princeton, NJ: Princeton University Press.

- Schmid-Hempel, P. (2005). Evolutionary ecology of insect immune defenses. *Annu. Rev. Entomol.* 50, 529–551. doi: 10.1146/annurev.ento.50.071803.130420
- Shellman-Reeve, J. S. (1997). “The spectrum of eusociality in termites,” in *The Evolution of Social Behavior in Insects and Arachnids*, eds J. C. Choe and B. J. Crespi (Cambridge: Cambridge University Press), 52–93.
- Simard, F., Licht, M., Besansky, N. J., and Lehmann, T. (2007). Polymorphism at the *defensin* gene in the *Anopheles gambiae* complex: testing different selection hypotheses. *Infect. Genet. Evol.* 7, 285–292. doi: 10.1016/j.meegid.2006.11.004
- Simola, D. F., Wissler, L., Donahue, G., Waterhouse, R. M., Helmkamp, M., Roux, J., et al. (2013). Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res.* 23, 1235. doi: 10.1101/gr.155408.113
- Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Genet.* 8, 272–285. doi: 10.1038/nrg2072
- Smit, A. F. A., Hubley, R., and Green, P. (1996–2010). *RepeatMasker Open-3.0*. Available online at: <http://www.repeatmasker.org>
- Suen, G., Teiling, C., Li, L., Holt, C., Abouheif, E., Bornberg-Bauer, E., et al. (2011). The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. *PLoS Genet.* 7:e1002007. doi: 10.1371/journal.pgen.1002007
- Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., et al. (2014). Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* 5, e3636. doi: 10.1038/ncomms4636
- The Honeybee Genome Sequencing Consortium. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443, 931–949. doi: 10.1038/nature05260
- The International Pea Aphid Genomics Consortium. (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8:e1000313. doi: 10.1371/journal.pbio.1000313
- Thorne, B., and Traniello, J. (2003). Comparative social biology of basal taxa of ants and termites. *Annu. Rev. Entomol.* 48, 283–306. doi: 10.1146/annurev.ento.48.091801.112611
- van Oystaeyen, A., Oliveira, R. C., Holman, L., van Zweden, J. S., Romero, C., Oi, C. A., et al. (2014). Conserved class of queen pheromones stops social insect workers from reproducing. *Science* 343, 287–290. doi: 10.1126/science.1244899
- Van Zweden, J. S., and D’Ettorre, P. (2010). “Nestmate recognition in social insects and the role of hydrocarbons,” in *Insect Hydrocarbons: Biology, Biochemistry, and Chemical Ecology*, eds G. J. Blomquist and A.G. Bagnères (Cambridge UK, Cambridge University Press), 222–243.
- Viljakainen, L., and Pamilo, P. (2008). Selection on an antimicrobial peptide defensin in ants. *J. Mol. Evol.* 67, 643–652. doi: 10.1007/s00239-008-9173-6
- Walsh, A. M., Kortschak, R. D., Gardner, M. G., Bertozzi, T., and Adelson, D. L. (2013). Widespread horizontal transfer of retrotransposons. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1012–1016. doi: 10.1073/pnas.1205856110
- Weil, T., Hoffmann, K., Kroiss, J., Strohm, E., and Korb, J. (2009). Scent of a queen-cuticular hydrocarbons specific for female reproductives in lower termites. *Naturwissenschaften* 96, 315–319. doi: 10.1007/s00114-008-0475-8
- Werren, J. H., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., Colbourne, J. K., et al. (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327, 343–348. doi: 10.1126/science.1178028
- Wilson, E. O. (1971). *The Insect Societies*. Cambridge: Harvard University Press.
- Wood, T. G., and Thomas, R. J. (1989). “The mutualistic association between Macrotermitinae and Termitomyces,” in *Insect-Fungus Interactions*, eds N. Wilding, N. M. Collins, P. M. Hammond, and J. F. Webber (New York, NY: Academic Press), 69–92. doi: 10.1016/B978-0-12-751800-8.50009-4
- Yek, S. H., Boomsma, J. J., and Schiøtt, M. (2013). Differential gene expression in *Acromyrmex* leaf-cutting ants after challenges with two fungal pathogens. *Mol. Ecol.* 22, 2173–2187. doi: 10.1111/mec.12255
- Zdobnov, E. M., and Apweiler, R. (2001). InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–848. doi: 10.1093/bioinformatics/17.9.847
- Zhao, X., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi: 10.1093/nar/gkm286
- Zhou, X., Slone, J. D., Rokas, A., Berger, S. L., Liebig, J., Ray, A., et al. (2012). Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals caste-specific signatures of odor coding. *PLoS Genet.* 8:e1002930. doi: 10.1371/journal.pgen.1002930

Conflict of Interest Statement: The Associate Editor Jürgen Rudolf Gadau declares that, despite being affiliated with the same institute and having collaborated with the author Jürgen Liebig, the review process was handled objectively. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 September 2014; paper pending published: 25 November 2014; accepted: 09 January 2015; published online: 04 March 2015.

Citation: Korb J, Poulsen M, Hu H, Li C, Boomsma JJ, Zhang G and Liebig J (2015) A genomic comparison of two termites with different social complexity. *Front. Genet.* 6:9. doi: 10.3389/fgene.2015.00009

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2015 Korb, Poulsen, Hu, Li, Boomsma, Zhang and Liebig. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Omic research in termites: an overview and a roadmap

Michael E. Scharf*

Department of Entomology, Purdue University, West Lafayette, IN, USA

OPEN ACCESS

Edited by:

Juergen Rudolf Gadau, Arizona State University, USA

Reviewed by:

Judith Korb, University of Freiburg, Germany

Edward L. Vargo, North Carolina State University, USA

*Correspondence:

Michael E. Scharf, Department of Entomology, Purdue University, 901 West State Street, West Lafayette, IN 47907-2089, USA
mscharf@purdue.edu

Specialty section:

This article was submitted to Evolutionary and Population Genetics, a section of the journal *Frontiers in Genetics*

Received: 02 September 2014

Accepted: 13 February 2015

Published: 13 March 2015

Citation:

Scharf ME (2015) Omic research in termites: an overview and a roadmap. *Front. Genet.* 6:76.
doi: 10.3389/fgene.2015.00076

Many recent breakthroughs in our understanding of termite biology have been facilitated by “omics” research. Omic science seeks to collectively catalog, quantify, and characterize pools of biological molecules that translate into structure, function, and life processes of an organism. Biological molecules in this context include genomic DNA, messenger RNA, proteins, and other biochemicals. Other permutations of omics that apply to termites include sociogenomics, which seeks to define social life in molecular terms (e.g., behavior, sociality, physiology, symbiosis, etc.) and digestomics, which seeks to define the collective pool of host and symbiont genes that collaborate to achieve high-efficiency lignocellulose digestion in the termite gut. This review covers a wide spectrum of termite omic studies from the past 15 years. Topics covered include a summary of terminology, the various kinds of omic efforts that have been undertaken, what has been revealed, and to a degree, what the results mean. Although recent omic efforts have contributed to a better understanding of many facets of termite and symbiont biology, and have created important new resources for many species, significant knowledge gaps still remain. Crossing these gaps can best be done by applying new omic resources within multi-dimensional (i.e., functional, translational, and applied) research programs.

Keywords: holobiome, digestome, sociogenomics, symbiosis, metabolomics, DNA methylation, sociobiology, socioevolution

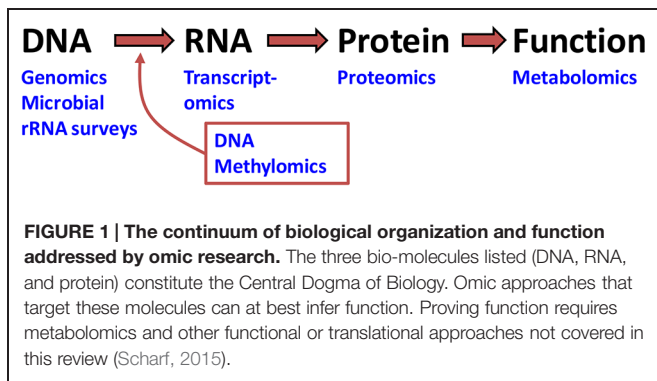
Introduction

Overview and Terminology

In a broad sense, the underlying goals of omic¹ science are to catalog, quantify, and characterize pools of biological molecules that translate into structure, function, and life processes of an organism or environment. The types of biological molecules receiving focus in omics² include genomic DNA, messenger RNA (mRNA), protein, and metabolites (**Figure 1**). DNA, mRNA, and protein are respectively the foci of genomics, transcriptomics, methylomics, and proteomics. Genomics, methylomics, and transcriptomics rely on nucleic acid sequencing, whereas proteomics utilizes peptide sequencing procedures. By contrast, metabolomics is rooted more in analytical chemistry and focuses on biochemicals, metabolites, or pathways. Another relevant omic approach is the cataloging of bacterial and protist symbionts using high-throughput 16S and 18S rRNA sequencing.

¹The singular term “omic” is used as an adjective in this review.

²The plural term “omics” is used as a noun.



Termite omic research has focused on the host termite, individual gut microbial symbionts or entire populations of gut microbes. In the latter case, these “meta” analyses focusing broadly on collective microbiota occurring in the gut microenvironment have been popular, particularly with microbiologists specializing in termite intestinal microbiology. Although it presents significant bioinformatic challenges, a more inclusive approach that considers host and symbionts together as a single functional unit is the best approach for appreciating the full functional capacity of termites. A fundamental advantage of omic research over more traditional organismal research is that it enables direct mechanistic insights into termite and symbiont physiology and biochemistry. The use of omic technologies has led to new insights into behavior, social structure, digestion, and host-symbiont/symbiont-symbiont interactions, and many other aspects of termite biology. However, also as addressed throughout this review, omic science has limits for being able to define biological function.

Termite Symbiosis and the Holobiont Concept

Termites are perhaps best known for their symbiotic associations with gut microbes (König et al., 2013; Brune, 2014) that are often linked to digestive processes, although lignocellulose digestion is not mediated entirely by gut microbes (Watanabe and Tokuda, 2010; **Figure 2A**). The more ancestral lower termites have tri-partite symbioses that include host, bacteria and protozoa; whereas in higher termites, symbiosis has been reduced to a two-way association between host and bacteria (but some higher termites also maintain ecto-symbiotic associations with fungi; Brune, 2014). The host component of termite symbiotic systems adds substantially to the digestive process both in terms of contributing enzymes and maintaining a favorable gut microenvironment for symbiosis and digestion to occur (Watanabe et al., 1998; Tartar et al., 2009; Scharf et al., 2011; Sethi et al., 2013a; Tokuda et al., 2014). Because of the high degree of interplay that occurs between the termite host and gut symbionts, a key idea moving forward will be to consider termites from the perspective of the “holobiont” (a single functional unit in which host and symbionts are physiologically tightly connected). Omic research has enabled a multifaceted systemic understanding of gut digestomes that is central to understanding

the termite holobiome from an applied perspective (Scharf, 2015).

Sociogenomics and Digestomics

The term *sociogenomics* was coined to describe the use of omic approaches for defining social life in molecular terms, which began with studies on the honey bee, *Apis mellifera* (Robinson et al., 2005). A parallel idea cited as rationale for many omic studies in social insects, including termites, is that solitary genes and traits were likely co-opted for new functions as solitary ancestors transitioned to social lifestyles (West-Eberhard, 2003; Nelson et al., 2007). Understanding such traits is essential for understanding termite social evolution (Miura and Scharf, 2011; **Figure 2B**). Another term used specifically in relation to digestive research is *digestomics*, which was coined to describe the collective pool of host and symbiont genes that collaborate to achieve high-efficiency lignocellulose digestion in the termite gut (Scharf and Tartar, 2008; Tartar et al., 2009; **Figure 2A**). Such terminology is useful because of the large number of symbionts that occupy termite guts and collaborate with the host in lignocellulose digestion. A related term is *termitosphere*, which is the full complement of gut and ectosymbiotic (nest) microbes present in termites, termite colonies, and their surrounding nest structures (Roose-Amsaleg et al., 2004; Bastien et al., 2013). Whether in relation to social, solitary or symbiont genes, proteins or other biomolecules, sociogenomic and digestomic research in termites has created an explosion of new sequence data.

Omic Studies in Termites: What has been Done?

Based on a recent literature survey (**Table 1**), at the time of writing this article around 70 papers had been published describing omic efforts in termite systems. These studies include all the themes introduced above, as well as microbial 16S and 18S surveys.

Taxonomic Distribution

In total, 82 termite species have been investigated using various omic approaches, with greater representation by lower than higher termites (72 vs. 28%). Among lower termites the top genera studied are important pest groups (*Reticulitermes* and *Coptotermes*), followed by non-pests from *Hodotermopsis*, *Mastotermes*, and *Cryptotermes*. Among higher termite genera, *Nasutitermes* dominate, followed by *Odontotermes*, *Trinervitermes*, and several other minor groups. Two termite genome sequences have now been published from the lower termite *Zootermopsis angusticollis* and the higher termite *Macrotermes natalensis* (see below).

Host vs. Symbiont Investigation

Of the various omic studies to date considering symbiosis and symbiotic partnerships in termite systems, the majority have taken an exclusive symbiont-oriented approach (>60%), whereas a minority have considered the host termite separately

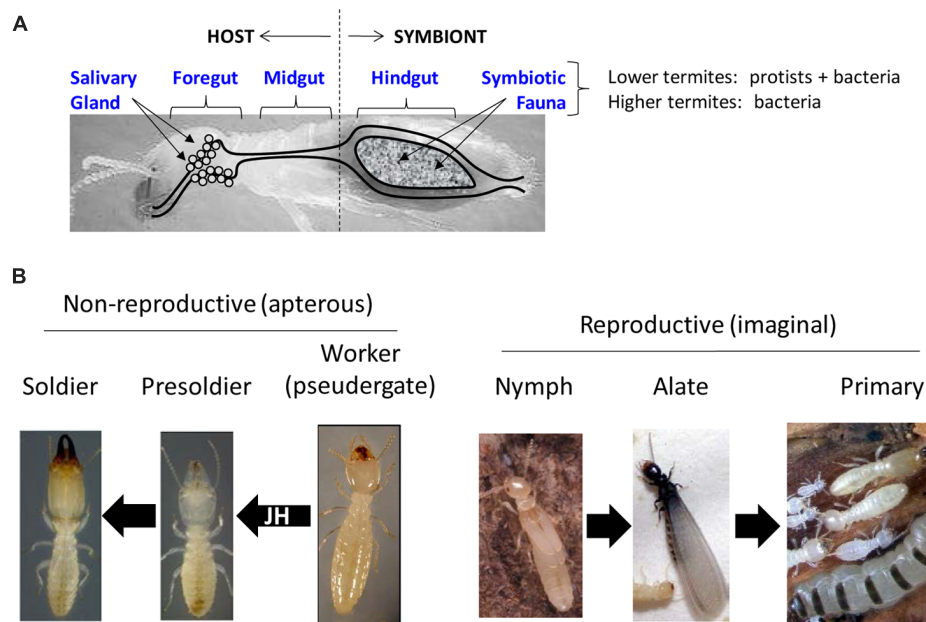


FIGURE 2 | Fundamental ideas behind digestomic and sociogenomic research in termites. (A) Key components associated with termite digestomes and digestomic research. Different gut regions have been studied in an attempt to dissect host and symbiont contributions to digestion. An important distinction between lower and higher termites is the presence of protist and bacterial symbiota in lower termites, and only bacteria in higher

termites. **(B)** Caste and phenotype-associated transitions addressed through sociogenomic research. Left: non-reproductive or “apterous” (wingless) phenotypes of lower termites. Presoldiers and soldiers differentiate from workers in response to elevated juvenile hormone (JH) titers. Right: nymphs give rise to alates that become primary reproductives; a process akin to typical hemimetabolous insect development.

(<20%). The remainder have considered host and symbiont together (~20%). In the latter category of host and symbiont combined, some studies have been a case of “accidental metatranscriptomics” (because protist symbionts have polyadenylated transcripts that are represented in cDNA libraries along with host transcripts; e.g., Scharf et al., 2003, 2005; Steller et al., 2010), but others have been deliberate metatranscriptomic studies (e.g., Tartar et al., 2009; Raychoudhury et al., 2013; Sen et al., 2013). The greater emphasis on gut symbiota compared to the host termite is likely because of the stereotypically well-recognized presence of gut microbes in termites.

Experimental Approaches and Types of Sequencing

In terms of experimental approaches taken, there has been an approximately equal split between descriptive and hypothesis-driven studies. Regarding the types of sequencing performed, transcriptomics and metatranscriptomics have been the dominant approaches (25 and 21% of studies), followed by microbial surveys for cataloging purposes (23%). The transcriptomic approaches used can be further divided into different methodologies such as cDNA library sequencing (Sanger, pyrosequencing or Illumina RNA-seq) and microarrays. Other efforts have targeted symbiont metagenomes (15%), symbiont or termite genomes (9%), proteomes (3%), and DNA methylomes (3%).

Omic Studies in Termites: What has been Revealed?

Genomics Host Termite Genomes

At present only two termite genome sequences are available (Table 1); one from the lower termite *Zootermopsis nevadensis* (Terrapon et al., 2014) and one from the higher termite *M. natalensis* (Poulsen et al., 2014). *Z. nevadensis* was selected for sequencing based on its small genome size of 562 Mb relative to other termites, most of which are over 1000 Mb (Koshikawa et al., 2008). The *Z. nevadensis* sequencing approach involved shotgun genome sequencing of genomic DNA from symbiont-free soldier heads ($n = 50$ and 150 heads for 2 and 20 kb libraries, respectively). The transcriptomes of castes and various phenotypes were also sequenced for both gene prediction and comparative transcriptomic purposes. Transcriptome data were also used to search for DNA methylation machinery and methylation/epigenetic differences among castes and developmental stages.

The *Z. nevadensis* genome provided the first hints into how termites differ at the genome level from their eusocial counterparts in the order Hymenoptera, which evolved sociality independently. For making socio-evolutionary comparisons, emphasis was placed on gene family expansions, male fertility, chemoreception, immunity, polyphenism/division of labor, and potential epigenetic caste regulation. An expansion of genes

TABLE 1 | A comprehensive literature summary of termite omic research, organized by approaches taken.

Omic approach taken	Termite group	Termite species	Host or symbiont	Tissue or fraction	Approach	Method	Major finding	Reference
Transcriptome	Lower	<i>Coptotermes formosanus</i>	Host	Whole-body polyphenic library	Hypothesis-driven (virgin vs. egg-laying queens)	Sanger sequencing + microarray	7663 ESTs sequenced that aligned into 4726 contigs; microarray analysis revealed 94 differentially expressed genes between virgin and reproductive queens	Husseneder et al. (2012)
			Combined	Whole workers	Hypothesis-driven	Sanger sequencing	1511 total unigenes (362 contigs + 1149 singletons)	Hussain et al. (2013)
		<i>Coptotermes gestral</i>	Host	Head tissue	Descriptive	Sanger sequencing	3003 high quality ESTs were obtained that aligned into 695 unigenes (245 contigs and 450 singletons)	Do et al. (2014)
		<i>Cryptotermes secundus</i>	Combined	Whole workers and neotenic reproductives	Hypothesis-driven	Representational difference analysis, Sanger sequencing	187 differentially expressed library clones were identified that aligned into 35 unigene contigs	Weil et al. (2007)
		<i>Hodotermopsis sjostedti</i>	Host	Head tissue of differentiated workers and soldiers	Hypothesis-driven	Differential display, Sanger sequencing	11 candidate bands were identified, including the <i>SOL1</i> gene, which was detected mostly in soldier heads	Miura et al. (1999)
			Host	Mandibular tissue of workers, presoldiers, and soldiers	Hypothesis-driven (JHA induced gene expression)	Fluorescent differential display (FDD), Sanger sequencing	81 candidate bands identified by FDD that aligned into 12 unigenes upregulated in mandibular tissue during soldier differentiation	Koshikawa et al. (2005)
			Host	Whole worker termites without guts	Hypothesis-driven (JHA up and downregulated genes)	FDD, Sanger sequencing	28 candidate bands identified by FDD; 18 aligned into ca. 5 unigenes	Cornette et al. (2006)
			Host	Worker brain and subesophageal ganglion	Hypothesis-driven (JHA induced gene expression)	FDD, Sanger sequencing	No differences in expression patterns detected between pseudergates and soldiers; five genes up-regulated in brain and/or SOG during differentiation	Ishikawa et al. (2010)
			Host	Whole worker termites without guts	Hypothesis-driven (JHA up and downregulated genes)	Subtractive libraries, filter arrays, and Sanger sequencing	87 and 64 JHA up- and downregulated clones identified	Cornette et al. (2013)
		<i>Mastotermes darwiniensis</i>	Host	Worker head, thorax and front legs	Hypothesis-driven (genes upregulated by <i>Metarhizium anisopliae</i> infection)	Subtractive cDNA library, Sanger sequencing	The number of differentially expressed clones was not specified	Thompson et al. (2003)
		<i>Reticulitermes flavipes</i>	Combined	Whole worker termites	Hypothesis-driven	Sanger sequencing	19 total unigenes (13 contigs + 6 singletons)	Gao et al. (2012)

(Continued)

TABLE 1 | Continued

Omic approach taken	Termite group	Termite species	Host or symbiont	Tissue or fraction	Approach	Method	Major finding	Reference
		Multi-species: <i>C. secundus</i> and <i>Cryptotermes cynocephalus</i>	Combined	Whole workers and neotenic reproductives	Hypothesis-driven	Representational difference analysis, Sanger sequencing	16 differentially expressed genes were identified in <i>C. cynocephalus</i> with significant homology to genes identified previously in <i>C. secundus</i>	Weil et al. (2009)
	Lower and higher	Multi-species: <i>H. sjostedti</i> , <i>Reticulitermes speratus</i> , and <i>Nasutitermes takasagoensis</i>	Host	Whole-body minus gut	Hypothesis-driven	454 pyrosequencing	> 1.2 million quality-filtered reads yielding >400 million bases for each of the three species. Caste transcriptomes compared by GO and orthology searches. Putative JH and caste differentiation genes annotated	Hayashi et al. (2013)
	Higher	<i>Macrotermes gilvus</i>	Symbiont	Fungal ectosymbiont (<i>Termitomyces</i>)	Descriptive	Subtractive cDNA library, Sanger sequencing	1,382 and 325 EST contigs were obtained for non-subtracted (lignocellulose fed) and subtracted (lignocellulose minus lab diet) cDNA libraries	Johjima et al. (2006)
		<i>N. takasagoensis</i>	Host	Soldier frontal gland	Descriptive	454 pyrosequencing	50,290 sequence reads were assembled into 1111 contigs (774 unigenes)	Hojo et al. (2012)
		<i>Odontotermes formosanus</i>	Host	Head tissue	Descriptive	Illumina sequencing	116,885 unigene sequences; 30,646 with significant identity	Huang et al. (2012)
		Species unknown	Symbiont	Fungal ectosymbiont of higher termite (<i>Termitomyces albuminosus</i>)	Descriptive	454 pyrosequencing	6494 candidate genes (3301 contigs + 3193 singletons)	Yang et al. (2012)
Proteome	Lower	<i>R. flavipes (santonensis)</i>	Symbiont	Gut symbiota (bacteria, protist)	Descriptive	LC-MS/MS (ion trap) and 2-D PAGE	Tubulins proved to be the most suitable protein family with which to identify flagellate populations from hindgut samples	Bauwens et al. (2013)

(Continued)

TABLE 1 | Continued

Omic approach taken	Termite group	Termite species	Host or symbiont	Tissue or fraction	Approach	Method	Major finding	Reference
	Higher	<i>Nasutitermes corniger</i>	Symbiont	Bacterial gut symbiota	Descriptive	LC-MS	886 proteins identified, 197 with known enzymatic function; very few cellulases identified	Burnum et al. (2011)
	Higher and lower	Multi-species: 12 species (10 lower, 2 higher)	Host	Labial glands of workers and soldiers	Descriptive	N-terminal peptide sequencing (Edman degradation)	Endogenous (host) endoglucanase cellulases were identified in worker labial glands of all species	Sillam-Dussès et al. (2012)
Metatranscriptome, metagenome, and 16S pyrosequencing	Higher	Multi-species: <i>Amitermes wheeleri</i> , <i>N. corniger</i>	Symbiont	Bacterial gut symbiota	Hypothesis-driven (differences between wood and dung feeders)	454 pyrosequencing	Firmicutes and Spirochaetes dominated in <i>A. wheeleri</i> , while Spirochaetes and Fibrobacteres dominated in <i>N. corniger</i>	He et al. (2013)
Metatranscriptome and proteome	Lower	<i>R. flavipes</i>	Combined	Worker termite gut and protist microbiota	Hypothesis-driven (comparison of cellulose vs. wood vs. lignin feeding)	454 pyrosequencing + LC-MS proteomics	347,798 sequence reads aligned into 97,254 singlets + 9553 differentially expressed contigs; proteome and transcriptome results showed congruence	Sethi et al. (2013a)
Metatranscriptome and proteome	Lower	<i>R. speratus</i>	Symbiont	Protist gut symbiota (hindgut lumen)	Descriptive	Sanger sequencing + proteomics	910 total clones sequenced; 580 candidate genes identified	Todaka et al. (2007)
Metatranscriptome	Lower	<i>C. formosanus</i>	Combined	Whole workers, nymphs, soldiers, and alates	Descriptive	Sanger sequencing (normalized polyphenic library)	25,939 candidate genes (16 691 contigs and 9248 singletons)	Zhang et al. (2012)
			Symbiont	Protist gut symbiota	Descriptive	454 pyrosequencing	75,122 candidate genes (2891 contigs + 72,231 singletons)	Xie et al. (2012)
		<i>H. sjostedti</i>	Combined	Worker termite gut (salivary gland, foregut, midgut, and hindgut) and protist microbiota	Descriptive	Sanger sequencing	Different compositions of expressed genes were identified across gut regions	Yuki et al. (2008)
		Multi-species: <i>R. speratus</i> , <i>H. sjostedti</i> , <i>Neotermes koshunensis</i> , <i>M. darwiniensis</i> , <i>Cryptocercus punctulatus</i>	Symbiont	Protist gut symbiota (hindgut lumen)	Hypothesis-driven	Sanger sequencing	910, 920, 1056, 1021, and 868 clones sequenced from each taxon ($n = 4775$) that aligned into 3780 unigene contigs; 77 full-length GHF7 cellulases were identified	Todaka et al. (2010)

(Continued)

TABLE 1 | Continued

Omic approach taken	Termite group	Termite species	Host or symbiont	Tissue or fraction	Approach	Method	Major finding	Reference
		<i>R. flavipes</i>	Combined	Whole-body polyphenic library (non-normalized)	Descriptive	Sanger sequencing (random clones)	88 random clones were sequenced that aligned into 49 unigene contigs	Wu-Scharf et al. (2003)
			Combined	Whole-body polyphenic library (non-normalized)	Hypothesis-driven (worker vs. soldier)	Filter arrays, Sanger sequencing	105 differentially expressed clones were identified that aligned into 34 unigene contigs	Scharf et al. (2003)
			Combined	Whole-body polyphenic library (non-normalized)	Hypothesis-driven (worker vs. immature reproductive)	Filter arrays, Sanger sequencing	68 differentially expressed clones were identified that aligned into 25 unigene contigs	Scharf et al. (2005)
			Combined	Worker termite gut and protist microbiota	Hypothesis-driven	Sanger sequencing	6555 total transcripts (3044 host, 3511 protist symbiont)	Tartar et al. (2009)
			Combined	Whole-body polyphenic library (soldier, worker, alate, early and late larvae)	Hypothesis-driven (comparisons among castes)	Sanger sequencing (random clones)	15,259 random clones sequenced representing 6991 total genes	Steller et al. (2010)
			Combined	Worker termite gut and protist microbiota	Hypothesis-driven (comparison of JH, soldier head extract, live soldiers, and reproductives)	Microarray	543 total gut genes differentially expressed after 24-h exposures (151 host + 392 protist symbiont)	Sen et al. (2013)
			Combined	Worker termite gut and protist microbiota	Hypothesis-driven (comparison of wood and cellulose/paper feeding)	Microarray	544 total gut genes differentially expressed after 7-days feeding periods (236 host + 301 protist symbiont)	Raychoudhury et al. (2013)
		<i>Zootermopsis angusticollis</i>	Symbiont	Bacterial gut symbiota (<i>Treponema spirachetes</i>)	Hypothesis-driven (comparison of two <i>Treponema</i> in co-culture)	Illumina sequencing	Total database size = 3,855,671 reads; 45% of reads were 16S and 23S rRNAs; >97% of all non-rRNA genes were unique	Rosenthal et al. (2011)
Metagenome and proteome	Higher	<i>N. corniger</i>	Symbiont	Bacterial gut symbiota (P3 luminal contents)	Descriptive	Sanger sequencing + 454 pyrosequencing + LC-MS proteomics	12 bacterial phyla and 216 phylotypes identified; >71 Mb of DNA sequenced; ~700 glycoside hydrolase domains corresponding to 45 different carbohydrate active enzymes were identified (including putative cellulases and hemicellulases)	Warnecke et al. (2007)
Metagenome and 16S survey	Higher	<i>Odontotermes yunnanensis</i>	Symbiont	Bacterial gut symbiota	Descriptive	454 pyrosequencing	548,807 total sequence reads; no evidence of lignases; 205 total cellulase and hemicellulase genes annotated	Liu et al. (2013)
Metagenome	LOWER	<i>C. gestroi</i>	Symbiont	Bacterial gut symbiota	Descriptive	Illumina de novo genome sequencing	316 candidate cellulase ORFs, 259 candidate hemicellulase ORFs, and 12 candidate pectinase ORFs	Do et al. (2014)

(Continued)

TABLE 1 | Continued

Omic approach taken	Termite group	Termite species	Host or symbiont	Tissue or fraction	Approach	Method	Major finding	Reference
		<i>R. flavipes</i> (<i>santonensis</i>)	Symbiont	Bacterial gut symbiota	Descriptive	Functional screening (beta glucosidase) + Sanger sequencing	9 beta glucosidase positive clones were identified from GH1, GH3, and GH4	Mattéotti et al. (2011b)
			Symbiont	Bacterial gut symbiota	Descriptive	Functional screening (beta glucosidase) + Sanger sequencing	1 beta glucosidase positive clone was identified (GH1)	Mattéotti et al. (2011a)
			Symbiont	Bacterial gut symbiota	Descriptive	Functional screening (xylosidase) + Sanger sequencing	1 putative endo-1,4-beta-xylanase was identified from GH11	Mattéotti et al. (2012)
	Higher	<i>Globitermes sulphureus</i>	Symbiont	Bacterial gut symbiota	Descriptive	Functional screening (beta glucosidase) + Sanger sequencing	1 beta glucosidase positive clones was identified and functionally expressed	Wang et al. (2012)
		<i>Macrotermes annandalei</i>	Symbiont	Bacterial gut symbiota	Descriptive	454 pyrosequencing (bacterial fosmid libraries grown under selective conditions)	13 positive clones identified encoding 1 xylanase and 12 beta-glucosidases	Liu et al. (2011)
		<i>Microcerotermes</i> sp.	Symbiont	Bacterial gut symbiota	Descriptive	Functional screening (cellulase and xylanase) + Sanger sequencing	Fourteen independent active clones (2 cellulases and 12 xylanases) were obtained by functional screening (GHF 5.8,10,11)	Nimchua et al. (2012)
		<i>Pseudacanthotermes militaris</i>	Symbiont	Gut and fungal comb bacteria	Descriptive	454 pyrosequencing	1.46 Mbp of metagenome sequence	Bastien et al. (2013)
		<i>Trinervitermes trinervoides</i>	Symbiont	Bacterial gut symbiota	Descriptive	Functional screening (esterase)+ Sanger sequencing	68 fosmid clones were identified with esterase activity, of which the 14 most active were sub cloned and sequenced	Rashmuse et al. (2012)
			Symbiont	Bacterial gut symbiota	Descriptive	Functional screening (feruloyl "FAE" esterase) + Sanger sequencing	Seven FAE-positive fosmid clones were identified	Rashmuse et al. (2014)
Metabolome	Lower	<i>H. sjostedti</i>	Combined	Worker termite gut	Descriptive	Isotope-ratio mass spectrometry (IR-MS)	Localized the majority of glucose release from 13C-cellulose to the foregut region	Tokuda et al. (2014)
		<i>Z. angusticollis</i>	Combined	Worker termite gut	Descriptive	TMAH thermocemical lysis coupled with GC-MS	Results transformed the view of lignin degradation in the termite gut	Gelb et al. (2008)

(Continued)

TABLE 1 | Continued

Omic approach taken	Termite group	Termite species	Host or symbiont	Tissue or fraction	Approach	Method	Major finding	Reference
		<i>C. formosanus</i>	Combined	Worker termite gut	Descriptive	TMAH thermoechemical lysis coupled with CP-MAS-NMR spectroscopy, and Py-GC/MS	During gut passage the native lignin macromolecular assembly undergoes structural modification but with conservation of the abundant β -O-4' interunit lignin linkage and retention of the original aromatic properties	Ke et al. (2011)
		<i>C. formosanus</i>	Combined	Worker termite gut	Descriptive	TMAH thermoechemical lysis coupled with GC-MS	Results suggest that the plant cell wall deconstruction process in <i>C. formosanus</i> consists of stepwise unlocking reactions that affect the lignin matrix and lignin-carbohydrate associations	Ke et al. (2013)
	Higher And lower	Multi-species: eight species (seven lower, one higher)	Host	Labial glands of workers and soldiers	Descriptive	HPLC MALDI-TOF and GC-TOF-MS	Hydroquinone and other glucose and benzene-linked compounds identified in labial gland secretions of workers and soldiers	Sillam-Dussès et al. (2012)
Genome	Lower	<i>C. formosanus</i>	Symbiont	Bacteroidales endosymbiont (phylotype CfP11-2) of the cellulolytic protist <i>Pseudotrichonympha grassii</i>	Descriptive	Combination of Sanger and 454 pyrosequencing	1,114,206 bp chromosome containing 758 putative protein-coding sequences, 38 transfer RNA genes, and 4 rRNA genes	Hongoh et al. (2008a)
		<i>M. darwiniensis</i>	Symbiont	Blattabacterium bacterial endosymbiont	Descriptive	Illumina sequencing	594 candidate genes identified (544 protein-coding + 40 RNA-coding)	Sabree et al. (2012)
		<i>R. flavipes</i>	Symbiont	Bacterial gut symbiont (Opitutaceae bacterium strain TAV1; <i>Verrucomicrobia</i>)	Descriptive	Combination of Illumina + 454 pyrosequencing	Genome contains 6,051 genes with 5,987 CDS; 64 structural RNAs were identified with the presence of one rRNA operon	Isanapong et al. (2012)
		<i>Reticulitermes lucifugus</i>	Symbiont	Bacterial gut symbiont (<i>Sebadella termitidis</i> strain NCTC 11300 ^T ; <i>Phylum Fusobacteria</i>)	Descriptive	Combined Sanger, Illumina, and 454 pyrosequencing	4,486,650 bp long genome containing 4,264 predicted genes (4,210 protein-coding, 54 RNAs)	Harmon-Smith et al. (2010)

(Continued)

TABLE 1 | Continued

Omic approach taken	Termite group	Termite species	Host or symbiont	Tissue or fraction	Approach	Method	Major finding	Reference
		<i>R. speratus</i>	Symbiont	Endomicrobia "TG-1" endosymbiont (phylogroup Rs-D17) of the cellulolytic protist <i>Trichonympha agilis</i>	Descriptive	Combination of Sanger and 454 pyrosequencing	1,125,857 bp chromosome encoding 761 putative protein-coding genes	Hongoh et al. (2008b)
Genome, transcriptome, and DNA methylome	Lower	<i>Zootermopsis nevadensis</i>	Host	Worker, soldier, reproductive, larvae	Descriptive	Illumina + 454 pyrosequencing	562 Mb genome sequenced with 98x coverage; 96 miRNA, and 17,737 protein coding genes were identified	Terrapon et al. (2014)
Genome, fungal symbiont genome, and gut microbial metagenome	Higher	<i>Macrotermes natalensis</i> (and <i>Termitomyces</i> sp. symbiont)	Combined	Genomic DNA of <i>M. natalensis</i> queen and <i>Termitomyces</i> homokaryon, metagenomic DNA of major worker, minor soldier, and queen gut	Descriptive	Illumina sequencing	First sequencing of a tripartite symbiotic system; 1.3 Gb host genome, 84 Mb fungal symbiont genome; 816 Mb gut prokaryotic metagenomes; major emphasis on cellulose digestion; greatly reduced gut microbiome in queens relative to major workers and minor soldiers	Poulsen et al. (2014)
DNA methylome	Lower	<i>Coptotermes lacteus</i>	Host	Workers, soldiers and nymphs	Descriptive	Methylation-targeted amplification fragment length polymorphism (AFLP)	Found evidence for DNA methylation, but no differences in methylation levels among castes	Lo et al. (2012)
	Lower and higher	Multi-species: <i>H. sjostedti</i> , <i>R. speratus</i> and <i>N. takasagoensis</i>	Host	Whole-body minus gut	Hypothesis-driven	454 Pyrosequencing	> 1.2 million filtered reads yielding >400 million bases for each of the three species. DNA methyltransferases putatively responsible for DNA methylation were represented in all three species	Hayashi et al. (2013)
	Lower	Multi-species: <i>R. flavipes</i> , <i>C. formosanus</i>	Host	Whole-body polyphenic libraries	Descriptive	Sanger sequencing	Signatures of high DNA methylation levels exist in <i>R. flavipes</i> and <i>C. formosanus</i> . Results suggest the presence of DNA methylation in <i>R. flavipes</i> and <i>C. formosanus</i> potentially at high levels or widely targeted across the lengths of genes, relative to other insect taxa	Glastad et al. (2013)
18S sequencing	Lower	Multi-species: <i>Reticulitermes</i> , <i>Zootermopsis</i> , <i>Cryptocercus</i>	Symbiont	Protist gut symbiota	Descriptive	454 pyrosequencing	Protist diversity estimated by 18S SSU sequencing is much higher than when estimated by protist morphology	Tai and Keeling (2013)

(Continued)

TABLE 1 | Continued

Omic approach taken	Termite group	Termite species	Host or symbiont	Tissue or fraction	Approach	Method	Major finding	Reference
	Lower	<i>Z. angusticollis</i>	Symbiont	Protist gut symbiota (single cell)	Descriptive	Sanger sequencing	Seven protists identified by rRNA sequence	Tai et al. (2013)
18S and bacterial 16S sequencing	Lower	Multi-species: 24 lower termites and three <i>Cryptocercus cockroaches</i>	Symbiont	Protist and bacterial gut symbiota	Hypothesis-driven	454 pyrosequencing	Although microbial communities are vertically inherited and codiversification with the host termite has had a prominent role in structuring symbiont communities, dispersal appears to have a larger role in community composition	Tai et al. (2015)
16S sequencing	Lower	<i>C. formosanus</i>	Symbiont (positions 27-1492)	Cuticular bacteria	Hypothesis-driven	Sanger sequencing	25 total ribotypes detected (20 and 14 from simple and extended families)	Husseneder et al. (2010b)
			Symbiont (positions 27-1492)	Bacterial gut symbiota (whole gut)	Hypothesis-driven	Sanger sequencing	1,876 total 16S reads that sorted into 213 bacteria ribotypes and 13 phyla	Husseneder et al. (2010a)
		Multi-species: <i>R. flavipes</i> (<i>santonensis</i>), <i>H. sjostedti</i> , <i>Z. nevadensis</i> , <i>M. darwiniensis</i> , <i>Kalotermes flavicollis</i> , <i>Neotermes castaneus</i> , <i>C. secundus</i>	Symbiont (positions 27-1492)	Bacterial endosymbionts of protist gut symbionts	Descriptive	Sanger sequencing	Each protist morphotype harbored "Endomicrobia" from unique phylogenetic lineages	Stingl et al. (2005)
		Multi-species: <i>R. flavipes</i> , <i>C. formosanus</i> , <i>Z. angusticollis</i>	Symbiont (positions 63-1492)	Spirochaete gut symbiota (whole gut)	Descriptive	Sanger sequencing	>21 new species of <i>Treponema</i> identified in each of the three species studied	Lilburn et al. (1999)
		<i>R. flavipes</i>	Symbiont (entire SSU region)	Bacterial gut symbiota (hindgut lumen)	Descriptive	Sanger sequencing + APDRA analysis	Six phyla and 261 species-level phylotypes estimated	Fisher et al. (2007)
			Symbiont (V5-V6 region)	Bacterial gut symbiota (hindgut lumen)	Hypothesis-driven	454 pyrosequencing	475,980 total 16S reads that sort into eight major bacterial phyla and 4761 species-level phylotypes (5% divergence level)	Boucias et al. (2013)
		<i>R. flavipes</i> (<i>santonensis</i>)	Symbiont positions (27-1492)	Bacterial gut symbiota (midgut, protozoa, hindgut fluid and wall)	Descriptive	Sanger sequencing + T-RFLP analysis	392 clones sequenced; seven major phyla and >200 species-level bacterial ribotypes identified	Yang et al. (2005)

(Continued)

TABLE 1 | Continued

Omic approach taken	Termite group	Termite species	Host or symbiont	Tissue or fraction	Approach	Method	Major finding	Reference
		<i>R. speratus</i>	Symbiont (positions 563–1114)	Bacterial gut symbiota (whole gut)	Descriptive	Sanger sequencing	1344 clones sequenced; 11 phyla and 268 species-level phylotypes identified	Hongoh et al. (2003)
		<i>Z. angusticollis</i>	Symbiont (positions 27–1492)	Bacterial gut symbiota	Hypothesis-driven (effects of antibiotic rifampicin)	Sanger sequencing	Six and 17 species-level OTUs were identified for rifampin and control treatments ($n = 85–87$ clones)	Rosengaus et al. (2011)
Higher		<i>Corritermes cumulans</i>	Symbiont (positions 21 or 27–907)	Whole gut	Descriptive	Sanger sequencing	>8 Phyla identified (species-level estimates not provided)	Grieco et al. (2013)
		<i>Cubitermes niokoloensis</i>	Symbiont (positions 338–518)	Bacterial symbionts from gut regions, soil and mound	Hypothesis-driven	Sanger sequencing + DGGE	212 total clones sequenced; 101 different species-level phylotypes identified	Fall et al. (2007)
		Multi-species: <i>Odontotermes somaliensis</i> , <i>Odontotermes</i> sp., <i>Microtermes</i> sp.	Symbiont (positions 27–1492)	Bacterial gut symbiota	Hypothesis-driven	Sanger sequencing	100, 100 and 96 clones sequenced from each taxon; 151 different phylotypes identified	Makonde et al. (2013)
		Multi-species:	Symbiont (positions 341–806)	Bacterial gut symbiota	Descriptive	454 pyrosequencing	Performed 16S sequencing on nine fungus-growing termite species from one geographic region of Ivory Coast; identified 16 phyla and 42 genera total, with 11 genera occurring in all nine species	Otani et al. (2014)
		<i>N. corniger</i>	Symbiont (V1–V2 and V8 regions)	Bacterial gut symbiota (P3 lumen)	Hypothesis-driven	454 pyrosequencing	2269 species-level OTUs of which 1617 and 652 were from the V1–V2 and V8 regions, respectively	Engelbrektson et al. (2010)
			Symbiont (V3–V4 region)	Bacterial gut symbiota (six whole gut regions)	Descriptive	454 pyrosequencing	3,200–26,000 16S reads per gut region (crop, midgut and paunch P1–P4) that sort into seven major bacterial phyla	Köhler et al. (2012)
		<i>N. takasagoensis</i>	Symbiont (positions 27–1390)	Bacterial gut symbiota (whole gut)	Hypothesis-driven	Sanger sequencing + T-RFLP analysis	388 total clones sequenced; 10 major phyla identified; 31–43 species-level phylotypes	Miyata et al. (2007)
		<i>T. trinervoides</i>	Symbiont (positions 1170–1492)	Bacterial gut symbiota	Hypothesis-driven (difference between grass and sugarcane feeding field colonies)	454 pyrosequencing	2274 and 2943 species-level OTUs sampled from sugarcane and grass feeding colonies (1% divergence level); nine major phyla sampled; Firmicutes and Bacteroidetes most common	Sanyika et al. (2012)

related to male fertility and upregulated gene expression in male reproductives are consistent with differences in mating biology between termites and Hymenoptera. Regarding chemoreception, divergent numbers of genes and gene families relative to Hymenoptera were identified, as were variations in chemoreception gene expression among castes. Regarding caste polyphenism and division of labor, caste-associated gene expression profiles were readily identifiable. Key caste-regulatory and reproduction-associated genes identified through preceding work (e.g., hexamerins, vitellogenins, and CYP genes) were further defined and verified as gene families at the genomic level. Interestingly, there are 76 cytochrome P450 genes in the *Z. nevadensis* genome; which is nearly 2x as many as encoded by the honey bee genome (Honey Bee Genome Sequencing Consortium, 2006). Lastly, DNA methylation signatures and patterns of alternative splicing provided some evidence to suggest epigenetic caste regulation (see later).

The *M. natalensis* sequencing considered not only the host genome, but also the entire tri-partite system of this higher fungus-growing termite. This included the 1.3 Gb host genome, the 84 Mb genome of the *Termitomyces* sp. fungal symbiont and 816 Mb of prokaryotic gut metagenome from major workers, minor soldiers, and queens. Emphasis was placed mostly on cellulose digestion, which revealed a rich complement of glycosyl hydrolases from host, fungi, and gut microbes that likely collaborate in lignocellulose digestion. Another major finding was that gut microbiota composition is reduced by over 50% in queens relative to workers and soldiers, suggesting that queen gut microbiota undergo substantial compositional changes during colony founding, which points toward the local environment or other external factors as sources of microbiota as incipient colonies grow and age. Moving forward, the *Z. nevadensis* and *M. natalensis* genomes will be important resources for termitologists, and will also provide important scaffolds for assembly of additional termite genomes that will facilitate study of genes related to many evolutionary and biological processes.

Individual Symbiont Genomes

Five individual symbiont genomes have been sequenced (Table 1), with several others published or in progress since the writing of this article. No protist genomes have yet been sequenced. Two bacterial endosymbionts of hindgut protists from *Coptotermes formosanus* and *Reticulitermes speratus* (phylum Elusimicrobia or “TG1”) were the first symbiont genomes sequenced; they were obtained from isolated individual cells after whole-genome amplification (Hongoh et al., 2008a,b). No lignocellulase genes were identified; however, both genomes encoded capabilities to fix nitrogen, recycle host nitrogen wastes for amino acid and cofactor biosynthesis, and import glucose and xylose as energy and carbon sources. The next symbiont genomes were from gut bacteria in the phyla Verrucomicrobia and Fusobacteria, from the termites *Reticulitermes flavipes* and *R. lucifugus* (Harmon-Smith et al., 2010; Isanapong et al., 2012). These genomes were from culturable isolates and were found to encode genes related to cellulose degradation and nitrogen fixation. Another example is the genome of an obligate fat body endosymbiont *Blattabacterium* from the basal termite

Mastotermes darwiniensis (Sabree et al., 2012). This bacterium displays a reduction in genome size and loss of genes required for amino acid production relative to free-living gut bacteria, which is consistent with its ability to recycle nitrogenous wastes and its role as a co-evolved endosymbiotic partner of the host termite.

Symbiont Metagenomes

At the time of writing this article, at least 12 prokaryotic metagenomes had been partially sequenced (Table 1). Most metagenome publications have reported on lignocellulase identification from genome sequences of gut bacteria that selectively grew on lignocellulose media (Liu et al., 2011; Mattéotti et al., 2011a,b, 2012; Nimchua et al., 2012; Rashamuse et al., 2012, 2014; Wang et al., 2012). Another study used targeted xylanase screening from gut and ectosymbiotic fungi-associated bacteria of the higher termite *Pseudacanthotermes militaris* (Bastien et al., 2013). Other studies took broader approaches to sequence from gut bacterial communities of higher termites. By combining metagenome sequencing with 16S surveys and metatranscriptomics, these studies revealed new information on bacterial cellulase diversity from termites with different symbiosis strategies (i.e., with and without fungal ectosymbionts; Warnecke et al., 2007; Liu et al., 2013) and from different feeding guilds (dung vs. wood; He et al., 2013). While these studies provided a wealth of new high-impact information on bacterial symbionts, they did not consider how symbionts from the gut and/or nest termitosphere collaborate with or complement the host termite.

Transcriptomics

Host Transcriptome

Around 15 transcriptomic studies to date have focused on physiological processes or tissues in the host termite (Table 1). Early studies looked for caste-biased gene expression, but the approaches employed had low resolving power and typically revealed only small numbers of differentially expressed genes. These studies mainly used subtractive hybridizations or cDNA “macro” arrays (reviewed by Miura and Scharf, 2011). Also, these early studies in lower termites often fell into the category of “accidental metatranscriptomics” as described earlier. The majority of focus in termite transcriptomic work has been on differences among castes or during caste differentiation (reviewed by Miura and Scharf, 2011). Mainly, newer studies are considered here.

Because of the importance of juvenile hormone (JH) to soldier caste differentiation and the reliability of JH treatment for inducing soldier caste differentiation, continuing focus has been placed on this transition in hypothesis-driven studies that combine JH assays with transcriptomics (e.g., Cornette et al., 2013; Sen et al., 2013). Caste-regulatory primer pheromones and the social environment have also been studied in the same context (Tarver et al., 2010; Sen et al., 2013). Other studies have included tissue-directed subtractive hybridizations, random/*de novo* cDNA library sequencing and/or cDNA oligonucleotide microarrays to reveal caste-biased gene expression (Weil et al., 2009; Ishikawa et al., 2010; Leonardo et al., 2011; Hojo et al.,

2012; Huang et al., 2012; Husseneder et al., 2012; Terrapon et al., 2014). The over-arching themes emerging from this work include caste and morphogenesis-associated gene expression, endocrine signaling, vitellogenesis, reproduction-related processes, and regulatory mechanisms that maintain juvenile worker states in lower termites.

The immune response is another aspect of host termite physiology investigated through transcriptomics. Four studies have revealed responses to immune challenges by both stereotypical and unprecedented immune-responsive genes (Thompson et al., 2003; Yuki et al., 2008; Gao et al., 2012; Hussain et al., 2013). Finally, an emerging theme has been to investigate pathogen-xenobiotic interactions at the transcriptome level (Husseneder and Simms, 2014; Sen et al., 2015).

Symbiont-Host Metatranscriptomes

In addition to host-targeted studies noted above, other studies have considered symbiont or host-symbiont metatranscriptome composition (Table 1). Early examples in this category showed worker-biased expression of protist cellulases (Scharf et al., 2003) and differential expression of symbiont cellulases between dispersing and non-dispersing adult reproductives (Scharf et al., 2005). Subsequent studies focused on metatranscriptome composition of bacteria, protist and/or fungal symbionts, mostly for the purpose of identifying digestive cellulases (reviewed by Scharf and Tartar, 2008). Recent work has probed deeper into gut metatranscriptomes by taking advantage of both traditional and next-generation sequencing technology (Todaka et al., 2010; Rosenthal et al., 2011; Xie et al., 2012; Zhang et al., 2012; He et al., 2013). Other work has sought to partition host and symbiont digestive contributions and identify candidate enzymes expressed specifically in response to wood (i.e., complex lignocellulose), cellulose and lignin feeding (Tartar et al., 2009; Raychoudhury et al., 2013; Sethi et al., 2013a).

One microarray study investigated gut metatranscriptome changes in responses to JH, primer pheromones and socio-environmental conditions, suggesting interesting linkages between gut symbiota and caste differentiation (Sen et al., 2013). Another microarray study investigated host and symbiont gene expression in response to pathogen and nicotinoid-insecticide challenges, providing new insights into immunological roles played by bacterial and protist gut symbionts in defending against invading fungal and bacterial pathogens (Sen et al., 2015), building on the ideas of extended disease resistance as conferred by fecal nest bacteria (Chouvenc et al., 2013) and gut microbiota (Rosengaus et al., 2014).

Proteomics

Proteomics (Table 1) is important to validate transcriptome studies, particularly for determining if a gene's presence and/or its transcription and translation are proportional. For example, proteomic studies in a higher termite were unable to identify most of the bacterial cellulase proteins predicted by metagenome sequencing (Warnecke et al., 2007; Burnum et al., 2011). Alternatively, proteomic studies in lower termites were able to identify both protist cellulases and other host lignocellulases initially identified via metatranscriptome sequencing

(Todaka et al., 2007; Sethi et al., 2013a). Another study investigated proteins present in labial gland secretions of 12 lower and higher termite species, identifying endogenous GHF9 cellulases as dominant components of worker labial gland secretions in most species investigated (Sillam-Dussès et al., 2012). Another study used proteomics to catalog gut microbial communities, but with limited resolution (Bauwens et al., 2013). Clearly, more proteomic efforts are needed to resolve issues related to: (1) congruency between nucleic acid and protein sequencing approaches, and (2) to verify open reading frames predicted by metagenome and transcriptome sequencing.

DNA Methylomes

Four studies to date have looked at methylation signatures across termite castes with somewhat differing results. A seminal study used a methylation-targeted amplification fragment length polymorphism (AFLP) approach in *Coptotermes lacteus* to look for methylation signature differences among castes (Lo et al., 2012). Evidence of methylation was found, but no significant caste-associated methylation patterns were identified.

A subsequent study was done *in silico* using database sequences from *R. flavipes* and *C. formosanus* (Glastad et al., 2013). In this study and the two described below, transcriptome data were mined to determine the specific distribution of CpG dinucleotides (i.e., 5'-3' cytosine followed by guanine), in order to predict DNA methylation levels *in silico*. Evidence of DNA methylation machinery and methylation signatures was found at high levels among expressed genes. Results also suggested that DNA methylation in *R. flavipes* is targeted to genes with ubiquitous (rather than differential) expression among castes and morphs. A third study examined host transcriptomes of three termite species that included two lower (*Hodotermopsis sjostedti*, *R. speratus*) and one higher termite (*Nasutitermes takasagoensis*; Hayashi et al., 2013). Pyrosequencing was done in combination with 69 caste and phenotypic libraries from the three termite species. Sequence analysis revealed that DNA methyltransferases potentially responsible for DNA methylation were present in each species, and verified the presence of methylation signatures. However, only limited evidence of caste-associated methylation profiles was detectable across the three species.

Finally, DNA methylation was assessed in *Z. nevadensis* as part of genome and transcriptome sequencing efforts (Terrapon et al., 2014). Transcriptome data were used to determine the specific distribution of CpG dinucleotides, in order to make *in silico* predictions of DNA methylation levels and explore for epigenetic differences among castes. In addition to verifying the presence of genes that encode for DNA methylation machinery (i.e., DNA methyltransferases 1 and 3), results showed greater methylation of genes rather than intergenic DNA, and a greater presence in introns than exons. This evidence, along with findings that alternatively spliced genes have greater degrees of methylation, suggests intronic methylation may impact alternative splicing.

While it is clear that DNA methylation exists in termites, so far inconclusive results have been obtained to suggest epigenetic caste regulation. As concluded previously in relation to genetic caste determination (Vargo and Husseneder, 2009), the field of

epigenetic caste regulation is in its infancy and epigenetic phenomena may or may not be relevant in natural colonies. More importantly, *in silico* methylation studies can only suggest that methylation may exist and which genes might be differentially methylated. Functional/translational research will be required to verify whether or not such genes truly are methylated, as well as the functions of those genes.

Metabolomics

Metabolomic studies are useful for assessing *in situ* processes, both as an exploratory approach and for functional/translational studies to verify nucleotide sequences. Soldier defensive secretions previously received much attention in this respect (Prestwich, 1984; Nelson et al., 2001). A more recent study investigated chemical components of labial gland secretions in soldier and worker termites from 7 lower and 1 higher termite (Sillam-Dussès et al., 2012). This study confirmed hydroquinone and other glucose and benzene-linked compounds as common labial gland secretions among most species.

Other metabolomic studies have focused on lignocellulose digestion. One main question addressed has been: *does lignin digestion or modification occur during passage through the termite gut?* Several studies over the past 25 years have addressed this question (reviewed by Ni and Tokuda, 2013) but recent metabolomic studies have been particularly informative (Geib et al., 2008; Ke et al., 2011, 2013). In general, findings are consistent regarding modification of lignin during passage through the gut, but evidence of actual lignin depolymerization has been more elusive. One possible reason for this could relate to insufficient detection procedures. Another possibility is that lignin-ether bonds, broken during depolymerization, only remain in this state for a short time and thus appear as intact lignin in frass. The induction of numerous antioxidant and detoxification enzymes by lignin feeding, as well as increased saccharification in the presence of lignin-associated phenoloxidases, supports the latter possibility (Sethi et al., 2013a). Despite convincing evidence of lignin modification during passage through the termite gut, and related omic studies revealing lignin-associated changes in host oxidative enzymatic machinery, the topic of lignin digestion/modification in termite guts remains contentious (Brune, 2014).

Another aspect of termite metabolomic research considers cellulose digestion and relative contributions of host and symbiont to this process. A recent metabolomic study investigated *in situ* digestion of ^{13}C -labeled crystalline cellulose by *H. sjostedti* (Tokuda et al., 2014). Novel insights obtained related to both cellulose digestion and nitrogen metabolism. The results not only confirmed preceding work showing that endogenous cellulose digestion by the host is substantial, but also suggested other novel possibilities; for example (i) a significant digestive contribution by hindgut bacteria is phosphorolysis of cello-oligosaccharides to glucose-1-phosphate, and (ii) essential amino acid acquisition occurs via lysis of hindgut microbes obtained through proctodeal trophallaxis. The rapid buildup of glucose observed in the foregut agrees well with prior studies showing that host foregut cellulases can produce high

levels of glucose directly from wood lignocellulose (Scharf et al., 2011; Sethi et al., 2013a,b). Additionally, higher glucose levels observed in the hindgut than other regions agrees with estimates that glucose release from lignocellulose is about 1/3 host and 2/3 symbiont (Scharf et al., 2011). However, since this study only focused on metabolite identification in gut tissue, it could not account for nutrients/metabolites transported out of the foregut and catabolized in other areas of the body.

Symbiont 16S and 18S Surveys

Bacterial 16S rRNA sequence surveys have been used extensively for cataloging bacteria and archaea (Wang and Qian, 2009), whereas 18S small subunit (SSU) rRNA surveys are just beginning to gain attention for cataloging protist symbionts (Tai and Keeling, 2013). Over 20 bacterial 16S surveys have been published to date using both cloning-dependent and -independent, high- and low-throughput approaches (Table 1). Highly variable species-level compositions have been obtained across the different termite species investigated, but, in general, six major bacterial phyla are represented across higher and lower termites: Bacteroidetes, Firmicutes, Spirochaetes, Proteobacteria, Fibrobacteres, and Elusimicrobia (Brune, 2014). Surveys conducted in parallel with higher-termite metagenome studies have been very informative for matching functional and taxonomic diversity (Warnecke et al., 2007; He et al., 2013); however, a study comparing multiple colonies through pyrosequencing of 16S amplicons found that bacterial compositions were different among colonies and likely influenced by local environment (Boucias et al., 2013). Additionally, 16S surveys revealed that lignocellulosic diet shifts have no short-term impacts on termite and cockroach microbiota composition (Sanyika et al., 2012; Boucias et al., 2013; Schauer et al., 2014). Another 16S survey of fungus-growing termites suggested a core microbiota of 42 genera that was shared among all nine termite species tested (Otani et al., 2014). This core microbiota was very different from other higher and lower termites, leading the authors to conclude the 42 common genera represent a core microbiota of fungus-growing termites. Conversely, since the termites were sampled from a limited geographic area it is possible that the core genera represent common microbes acquired from the local environment.

In comparison to prokaryotic 16S surveys, comparatively few protist 18S SSU surveys have been conducted (Table 1). These studies, conducted using a combination of cloning-dependent and independent approaches, have been transformative. Two studies provided new evidence to suggest greater protist symbiont diversity than originally indicated by traditional morphological identification (James et al., 2013; Tai et al., 2013). Two other studies used high-throughput 16S and 18S SSU sequencing to compare 24 lower termites with three wood-feeding cockroaches (Tai and Keeling, 2013; Tai et al., 2015). Like their predecessors, these studies found protist diversity to be higher than when estimated by morphology, and also that protist symbiont taxa tend to be highly endemic to

a host genus, which is different than relationships between termite hosts and bacterial symbiota. These findings illustrate the significant opportunities that exist for development of high-throughput techniques for assessing protist symbiont communities and studying protist-bacterial symbiont relationships.

Needs and Opportunities

Termite omic research in the last 10–15 years has led to a new era of understanding for termite and symbiont biology. Omics has also enabled the development of new unparalleled resources (i.e., transcriptome, genome, proteome, metabolome, symbiont meta-omic, and symbiont rDNA) useful for moving ahead with targeted functional work. The stage is now set for making significant headway in many aspects of termite research, including, but not limited to digestion, symbiosis, caste differentiation, and social evolution. However, key needs and opportunities remain in specific areas that seem particularly relevant for filling in knowledge gaps and potentially leading to transformative, paradigm-shifting outcomes.

Having the *Z. nevadensis* and *M. natalensis* genomes available not only facilitates further study of genes related to a range of evolutionary and biological processes, but these resources also provide important scaffolds for assembly of additional lower and higher termite genomes. Once multiple termite genomes are available, this would certainly better inform our view of termite social evolution. On the topic of host-symbiont “hologenomes,” sequencing more host genomes and symbiont metagenomes from the same termites concurrently (as recently done for *M. natalensis*), would provide unprecedented insights into the scope of interactions and synergies occurring in termite holobiomes. Such efforts could further reveal important differences between clades of higher and lower termites, leading to new evolutionary insights. Such datasets would also provide unmatched resources for advancing integrative sociogenomic, digestomic, termitosphere, and other research topics.

On the topic of proteomics, more studies are needed in species that have had genomes, transcriptomes, metagenomes, or metatranscriptomes sequenced. Combining proteomics with nucleic acid sequencing will better resolve gene prediction models and better test for congruency between transcription and translation profiles. On the topic of metabolomics, termite digestion remains an area much in need of metabolomic research focusing on how complex lignocellulose is broken down in termite guts and converted to energy. Also, tracking metabolites as they leave the gut and are utilized in the termite body would be very informative for testing hypotheses on the relative importance of nutrient flow into symbiont metabolic pathways.

On the topic of DNA methylomics, while it is now clear that DNA methylation happens in termites, so-far inconclusive results have been obtained regarding the role of DNA methylation in caste regulation. *In silico* methylation studies as performed

can only suggest that methylation may exist and which genes are potentially differentially methylated. Functional and translational research is needed to understand the roles of such genes.

Substantial opportunities and needs still remain for 16S and 18S rRNA-based symbiont cataloging. Protist 18S SSU cataloging capabilities in particular have recently been developed, and can continue to improve provided that several conditions are met, such as: (1) appropriate primers can be developed, (2) statistically sound sampling regimes can be developed at biologically relevant scales, (3) single-cell microbiology and other data sources can be integrated, and (4) appropriate analytical tools developed (Tai and Keeling, 2013). This line of research has already begun to transform the view of protist diversity and co-evolution with host termites but more studies are needed in different termite species with established omic resources.

Finally, regarding prokaryotic 16S surveys, much has already been done, but an important gap in knowledge is the extent to which environment influences bacterial microbiota composition. This is important information for understanding differences in behavior and physiology across the geographic range for a termite species, as well as potentially for limiting the extent to which generalizations can be made about the relative importance of individual microbes or core microbiota in gut communities.

Conclusion

This review has covered many aspects related to outcomes, findings and trends resulting from termite omic research. To date, omic research in diverse termite species has provided key insights into caste differentiation, digestion, pathogen defense and microbiomes, and most recently has provided two termite genome sequences. Termite omics has also created important tools and resources for conducting targeted, functional, translational, and applied research. However, these resources have only received limited attention to date for asking hypothesis-driven questions to elucidate the functional and evolutionary significance for pools of identified genes, proteins, and microbes. In recent years sequencing has rapidly moved into the realm of super high-throughput, with accompanying assembly and analyses requiring proportional super-computing power and bioinformatics expertise, but only limited resolution of biology or function. Transitioning from research that produces lists of genes, proteins and microbes, to research that determines their functional significance, is where the most important challenges lie for the next phases of termite science.

Funding

Work conducted in the author's laboratory was supported by the following funding sources: USDA-CSREES-NRI grant no. 2007-35607-17777, USDA-NIFA-AFRI grant nos. 2009-05245

and 2010-65106-30727, Consortium for Plant Biotechnology Research-DOE grant no. DE-FG36-02GO12026, DOE-SBIR grant nos. DE-FG02-08ER85063 and DE-85538 S08-II, NSF grant no. 1233484CBET, and the O.W. Rollins/Orkin Endowment at Purdue University. M.E.S. is an inventor on the following patents: US Patent No. 7,968,525, US Patent No. 8,445,240, US Provisional Patent No. 61/602,149, and US Provisional Patent No. 61/902,472.

References

- Bastien, G., Arnal, G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., et al. (2013). Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. *Biotechnol. Biofuels*. 6:78. doi: 10.1186/1754-6834-6-78
- Bauwens, J., Millet, C., Tarayre, C., Brasseur, C., Destain, J., Vandenbol, M., et al. (2013). Symbiont diversity in *Reticulitermes santonensis*: investigation strategy through proteomics. *Environ. Entomol.* 42, 882–887. doi: 10.1603/EN13112
- Boucias, D. G., Cai, Y., Sun, Y., Lietze, V. U., Sen, R., Raychoudhury, R., et al. (2013). The hindgut lumen prokaryotic microbiota of the termite *Reticulitermes flavipes* and its responses to dietary lignocellulose composition. *Mol. Ecol.* 22, 1836–1853. doi: 10.1111/mec.12230
- Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12, 168–180. doi: 10.1038/nrmicro3182
- Burnum, K. E., Callister, S. J., Nicora, C. D., Purvine, S. O., Hugenholtz, P., Warnecke, F., et al. (2011). Proteome insights into the symbiotic relationship between a captive colony of *Nasutitermes corniger* and its hindgut microbiome. *ISME J.* 5, 161–164. doi: 10.1038/ismej.2010.97
- Chouvenc, T., Efstathiou, C. A., Elliott, M. L., and Su, N. Y. (2013). Extended disease resistance emerging from the faecal nest of a subterranean termite. *Proc. Biol. Sci.* 280:20131885. doi: 10.1098/rspb.2013.1885
- Cornette, R., Koshikawa, S., Hojo, M., Matsumoto, T., and Miura, T. (2006). Caste-specific cytochrome P450 in the damp-wood termite *Hodotermopsis sjostedti*. *Insect Mol. Biol.* 15, 235–244. doi: 10.1111/j.1365-2583.2006.00632.x
- Cornette, R., Hayashi, Y., Koshikawa, S., and Miura, T. (2013). Differential gene expression in response to juvenile hormone analog treatment in the damp-wood termite *Hodotermopsis sjostedti*. *J. Insect Physiol.* 59, 509–518. doi: 10.1016/j.jinsphys.2013.02.002
- Do, T. H., Nguyen, T. T., Nguyen, T. N., Le, Q. G., Nguyen, C., Kimura, K., et al. (2014). Mining biomass-degrading genes through Illumina-based de novo sequencing and metagenomic analysis of free-living bacteria in the gut of the lower termite *Coptotermes gestroi* harvested in Vietnam. *J. Biosci. Bioeng.* 6, 665–671. doi: 10.1016/j.jbiosc.2014.05.010
- Engelbrektson, A., Kunin, V., Wrighton, K. C., Zvenigorodsky, N., Chen, F., Ochman, H., et al. (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J.* 4, 642–647. doi: 10.1038/ismej.2009.153
- Fall, S., Hamelin, J., Ndiaye, F., Assigbetse, K., Aragno, M., Chotte, J. L., et al. (2007). Differences between bacterial communities in the gut of a soil-feeding termite (*Cubitermes niokoloensis*) and its mounds. *Appl. Environ. Microbiol.* 73, 5199–5208. doi: 10.1128/AEM.02616-06
- Fisher, M., Miller, D., Brewster, C., Husseneder, C., and Dickerman, A. (2007). Diversity of gut bacteria of *Reticulitermes flavipes* as examined by 16S rRNA gene sequencing and amplified rDNA restriction analysis. *Curr. Microbiol.* 55, 254–259. doi: 10.1007/s00284-007-0136-8
- Gao, Q., Tancredi, S. E., and Thompson, G. J. (2012). Identification of mycosis-related genes in the eastern subterranean termite by suppression subtractive hybridization. *Arch. Insect Biochem. Physiol.* 80, 63–76. doi: 10.1002/arch.21026
- Geib, S. M., Filley, T. R., Hatcher, P. G., Hoover, K., Carlson, J. E., Jimenez-Gasco Mdel, M., et al. (2008). Lignin degradation in wood-feeding insects. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12932–12937. doi: 10.1073/pnas.0805257105
- Glastad, K. M., Hunt, B. G., and Goodisman, M. A. D. (2013). Evidence of a conserved functional role for DNA methylation in termites. *Insect Mol. Biol.* 22, 143–154. doi: 10.1111/imb.12010
- Grieco, M. A., Cavalcante, J. J., Cardoso, A. M., Vieira, R. P., Machado, E. A., Clementino, M. M., et al. (2013). Microbial community diversity in the gut of the South American termite *Cornitermes cumulans*. *Microb. Ecol.* 65, 197–204. doi: 10.1007/s00248-012-0119-6
- Harmon-Smith, M., Celia, L., Chertkov, O., Lapidus, A., Copeland, A., Glavina Del Rio, T., et al. (2010). Complete genome sequence of *Seballidella termitidis* type strain (NCTC 11300). *Stand. Genomic Sci.* 2, 220–227. doi: 10.4056/sigs.811799
- Hayashi, Y., Shigenobu, S., Watanabe, D., Toga, K., Saiki, R., Shimada, K., et al. (2013). Construction and characterization of normalized cDNA libraries by 454 pyrosequencing and estimation of DNA methylation levels in three distantly related termite species. *PLoS ONE* 8:e76678. doi: 10.1371/journal.pone.0076678
- He, S., Ivanova, N., Kirtan, E., Allgaier, M., Bergin, C., Scheffrahn, R. H., et al. (2013). Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS ONE* 8:e61126. doi: 10.1371/journal.pone.0061126
- Hojo, M., Maekawa, K., Saitoh, S., Shigenobu, S., Miura, T., Hayashi, Y., et al. (2012). Exploration and characterization of genes involved in the synthesis of diterpene defence secretion in nasute termite soldiers. *Insect Mol. Biol.* 21, 545–557. doi: 10.1111/j.1365-2583.2012.01162.x
- Honey Bee Genome Sequencing Consortium. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443, 931–949. doi: 10.1038/nature05260
- Hongoh, Y., Ohkuma, M., and Kudo, T. (2003). Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus*. *FEMS Microbiol. Ecol.* 44, 231–242. doi: 10.1016/S0168-6496(03)00026-6
- Hongoh, Y., Sharma, V. K., Prakash, T., Noda, S., Toh, H., Taylor, T. D., et al. (2008a). Genome of an endosymbiotic coupling N₂ fixation to cellulolysis within protist cells in termite gut. *Science* 322, 1108–1109. doi: 10.1126/science.1165578
- Hongoh, Y., Sharma, V. K., Prakash, T., Noda, S., Taylor, T. D., Kudo, T., et al. (2008b). Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5555–5560. doi: 10.1073/pnas.0801389105
- Huang, Q., Sun, P., Zhou, X., and Lei, C. (2012). Characterization of head transcriptome and analysis of gene expression involved in caste differentiation and aggression in *Odontotermes formosanus*. *PLoS ONE* 7:e50383. doi: 10.1371/journal.pone.0050383
- Hussain, A., Li, Y. F., Cheng, Y., Liu, Y., Chen, C. C., and Wen, S. Y. (2013). Immune-related transcriptome of *Coptotermes formosanus* Shiraki workers: the defense mechanism. *PLoS ONE* 8:e69543. doi: 10.1371/journal.pone.0069543
- Husseneder, C., Ho, H. Y., and Blackwell, M. (2010a). Comparison of the bacterial symbiont composition of the Formosan subterranean termite from its native and introduced range. *Open Microbiol. J.* 4, 53–66. doi: 10.2174/1874285801004010053
- Husseneder, C., Simms, D. M., Aluko, G. K., and Delatte, J. (2010b). Colony breeding system influences cuticular bacterial load of Formosan subterranean termite workers. *Environ. Entomol.* 39, 1715–1723. doi: 10.1603/EN09238
- Husseneder, C., McGregor, C., Lang, R. P., Collier, R., and Delatte, J. (2012). Transcriptome profiling of female alates and egg-laying queens of the Formosan subterranean termite. *Comp. Biochem. Physiol. D* 7, 14–27.
- Husseneder, C., and Simms, D. M. (2014). Effects of caste on the expression of genes associated with septic injury and xenobiotic exposure in the Formosan subterranean termite. *PLoS ONE* 9:e105582. doi: 10.1371/journal.pone.0105582
- Isanapong, J., Goodwin, L., Bruce, D., Chen, A., Detter, C., Han, J., et al. (2012). High-quality draft genome sequence of the *Opitutaceae* bacterium strain TAV1, a symbiont of the wood-feeding termite *Reticulitermes flavipes*. *J. Bacteriol.* 194, 2744–2745. doi: 10.1128/JB.00264-12
- Ishikawa, Y., Okada, Y., Ishikawa, A., Miyakawa, H., Koshikawa, S., and Miura, T. (2010). Gene expression changes during caste-specific neuronal development

- in the damp-wood termite *Hodotermopsis sjostedti*. *BMC Genomics* 11:314. doi: 10.1186/1471-2164-11-314
- James, E. R., Tai, V., Scheffrahn, R. H., and Keeling, P. J. (2013). *Trichonympha burlesquei* n. sp. from *Reticulitermes virginicus* and evidence against a cosmopolitan distribution of *Trichonympha agilis* in many termite hosts. *Int. J. Syst. Evol. Microbiol.* 63, 3873–3876. doi: 10.1099/ijs.0.054874-0
- Johjima, T., Taprab, Y., Noparatnaraporn, N., Kudo, T., and Ohkuma, M. (2006). Large-scale identification of transcripts expressed in a symbiotic fungus (*Termitomyces*) during plant biomass degradation. *Appl. Microbiol. Biotechnol.* 73, 195–203. doi: 10.1007/s00253-006-0570-8
- Ke, J., Laskar, D. D., and Chen, S. (2013). Tetramethylammonium hydroxide (TMAH) thermochemolysis for probing in situ softwood lignin modification in each gut segment of the termite. *J. Agric. Food Chem.* 61, 1299–1308. doi: 10.1021/jf3048548
- Ke, J., Laskar, D. D., Singh, D., and Chen, S. (2011). In situ lignocellulosic unlocking mechanism for carbohydrate hydrolysis in termites: crucial lignin modification. *Biotechnol. Biofuels* 4:17. doi: 10.1186/1754-6834-4-17
- Köhler, T., Dietrich, C., Scheffrahn, R. H., and Brune, A. (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78, 4691–4701. doi: 10.1128/AEM.00683-12
- König, H., Li, L., and Fröhlich, J. (2013). The cellulolytic system of the termite gut. *Appl. Microbiol. Biotechnol.* 97, 7943–7962. doi: 10.1007/s00253-013-5119-z
- Koshikawa, S., Cornette, R., Hojo, M., Maekawa, K., Matsumoto, T., and Miura, T. (2005). Screening of genes expressed in developing mandibles during soldier differentiation in the termite *Hodotermopsis sjostedti*. *FEBS Lett.* 579, 1365–1370. doi: 10.1016/j.febslet.2005.01.031
- Koshikawa, S., Miyazaki, S., Cornette, R., Matsumoto, T., and Miura, T. (2008). Genome size of termites (Insecta, Dictyoptera, Isoptera) and wood roaches (Insecta, Dictyoptera, Cryptocercidae). *Naturwissenschaften* 95, 859–867. doi: 10.1007/s00114-008-0395-7
- Leonardo, F. C., da Cunha, A. F., da Silva, M. J., Carazzolle, M. F., Costa-Leonardo, A. M., Costa, F. F., et al. (2011). Analysis of the workers head transcriptome of the Asian subterranean termite, *Coptotermes gestroi*. *Bull. Entomol. Res.* 101, 383–391. doi: 10.1017/S0007485310000556
- Lilburn, T. G., Schmidt, T. M., and Breznak, J. A. (1999). Phylogenetic diversity of termite gut spirochaetes. *Environ. Microbiol.* 1, 331–345. doi: 10.1046/j.1462-2920.1999.00043.x
- Liu, N., Yan, X., Zhang, M., Xie, L., Wang, Q., Huang, Y., et al. (2011). Microbiome of fungus-growing termites: a new reservoir for lignocellulase genes. *Appl. Environ. Microbiol.* 77, 48–56. doi: 10.1128/AEM.01521-10
- Liu, N., Zhang, L., Zhou, H., Zhang, M., Yan, X., Wang, Q., et al. (2013). Metagenomic insights into metabolic capacities of the gut microbiota in a fungus-cultivating termite (*Odontotermes yunnanensis*). *PLoS ONE* 8:e69184. doi: 10.1371/journal.pone.0069184
- Lo, N., Li, B., and Ujvari, B. (2012). DNA methylation in the termite *Coptotermes lacteus*. *Insect Soc.* 59, 257–261. doi: 10.1007/s00040-011-0213-7
- Makonde, H. M., Boga, H. I., Osiemo, Z., Mwirichia, R., Mackenzie, L. M., Göker, M., et al. (2013). 16S-rRNA-based analysis of bacterial diversity in the gut of fungus-cultivating termites (*Microtermes* and *Odontotermes* species). *Antonie Van Leeuwenhoek* 104, 869–883. doi: 10.1007/s10482-013-0001-7
- Mattéotti, C., Bauwens, J., Brasseur, C., Tarayre, C., Thonart, P., Destain, J., et al. (2012). Identification and characterization of a new xylanase from Gram-positive bacteria isolated from termite gut (*Reticulitermes santonensis*). *Protein Expr. Purif.* 83, 117–127. doi: 10.1016/j.pep.2012.03.009
- Mattéotti, C., Haubruge, E., Thonart, P., Francis, F., De Pauw, E., Portetelle, D., et al. (2011a). Characterization of a new β -glucosidase/ β -xylosidase from the gut microbiota of the termite (*Reticulitermes santonensis*). *FEMS Microbiol. Lett.* 314, 147–157. doi: 10.1111/j.1574-6968.2010.02161.x
- Mattéotti, C., Thonart, P., Francis, F., Haubruge, E., Destain, J., Brasseur, C., et al. (2011b). New glucosidase activities identified by functional screening of a genomic DNA library from the gut microbiota of the termite *Reticulitermes santonensis*. *Microbiol. Res.* 166, 629–642. doi: 10.1016/j.micres.2011.01.001
- Miura, T., Kamikouchi, A., Sawata, M., Takeuchi, H., Natori, S., Kubo, T., et al. (1999). Soldier caste-specific gene expression in the mandibular glands of *Hodotermopsis japonica*. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13874–13879. doi: 10.1073/pnas.96.24.13874
- Miura, T., and Scharf, M. E. (2011). “Molecular mechanisms underlying caste differentiation in termites,” in *Biology of Termites: A Modern Synthesis*, eds D. E. Bigness, Y. Roisin, and N. Lo (Dordrecht: Springer), 211–253.
- Miyata, R., Noda, N., Tamaki, H., Kinjo, K., Aoyagi, H., Uchiyama, H., et al. (2007). Influence of feed components on symbiotic bacterial community structure in the gut of the wood-feeding higher termite *Nasutitermes takasagoensis*. *Biosci. Biotechnol. Biochem.* 71, 1244–1251. doi: 10.1271/bbb.60672
- Nelson, C. M., Ihle, K. E., Fondrk, M. K., Page, R. E., and Amdam, G. V. (2007). The gene vitellogenin has multiple coordinating effects on social organization. *PLoS Biol.* 5:e62. doi: 10.1371/journal.pbio.0050062
- Nelson, L. J., Cool, L. G., Forschler, B. T., and Haverty, M. I. (2001). Correspondence of soldier defense secretion mixtures with cuticular hydrocarbon phenotypes for chemotaxonomy of the termite genus *Reticulitermes* in North America. *J. Chem. Ecol.* 27, 1449–1479. doi: 10.1023/A:1010325511844
- Ni, J., and Tokuda, G. (2013). Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnol. Adv.* 31, 838–850. doi: 10.1016/j.biotechadv.2013.04.005
- Nimchua, T., Thongaram, T., Uengwetwanit, T., Pongpattanakitsote, S., and Eurwilaichitr, L. (2012). Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. *J. Microbiol. Biotechnol.* 22, 462–469. doi: 10.4014/jmb.1108.08037
- Otani, S., Mikaelian, A., Nobre, T., Hansen, L. H., Koné, N. A., Sørensen, S. J., et al. (2014). Identifying the core microbial community in the gut of fungus-growing termites. *Mol. Ecol.* 23, 4631–4644. doi: 10.1111/mec.12874
- Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S., et al. (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14500–14505. doi: 10.1073/pnas.1319718111
- Prestwich, G. D. (1984). Defense mechanisms of termites. *Ann. Rev. Entomol.* 29, 201–232. doi: 10.1146/annurev.en.29.010184.001221
- Rashamuse, K., Mabizela-Mokoena, N., Sanyika, T. W., Mabvakure, B., and Brady, D. (2012). Accessing carboxylesterase diversity from termite hindgut symbionts through metagenomics. *J. Mol. Microbiol. Biotechnol.* 22, 277–286. doi: 10.1159/000342447
- Rashamuse, K., Ronneburg, T., Sanyika, W., Mathiba, K., Mmutlane, E., and Brady, D. (2014). Metagenomic mining of feruloyl esterases from termite enteric flora. *Appl. Microbiol. Biotechnol.* 98, 727–737. doi: 10.1007/s00253-013-4909-7
- Raychoudhury, R., Sen, R., Cai, Y., Sun, Y., Lietze, V. U., Boucias, D. G., et al. (2013). Comparative metatranscriptomic signatures of wood and paper feeding in the gut of the termite *Reticulitermes flavipes*. *Insect Mol. Biol.* 22, 155–171. doi: 10.1111/imb.12011
- Robinson, G. E., Grozinger, C. M., and Whitfield, C. W. (2005). Sociogenomics: social life in molecular terms. *Nat. Rev. Genet.* 6, 257–270. doi: 10.1038/nrg1575
- Roose-Amsaleg, C., Brygoo, Y., and Harry, M. (2004). Ascomycete diversity in soil-feeding termite nests and soils from a tropical rainforest. *Environ. Microbiol.* 6, 462–469. doi: 10.1111/j.1462-2920.2004.00579.x
- Rosengaus, R. B., Schultheis, K. F., Yalonetskaya, A., Bulmer, M. S., DuComb, W. S., Benson, R. W., et al. (2014). Symbiont-derived β -1,3-glucanases in a social insect: mutualism beyond nutrition. *Front. Microbiol.* 5:607. doi: 10.3389/fmicb.2014.00607
- Rosengaus, R. B., Zecher, C. N., Schultheis, K. F., Brucker, R. M., and Bordenstein, S. R. (2011). Disruption of the termite gut microbiota and its prolonged consequences for fitness. *Appl. Environ. Microbiol.* 77, 4303–4312. doi: 10.1128/AEM.01886-10
- Rosenthal, A. Z., Matson, E. G., Eldar, A., and Leadbetter, J. R. (2011). RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. *ISME J.* 5, 1133–1142. doi: 10.1038/ismej.2011.3
- Sabree, Z. L., Huang, C. Y., Arakawa, G., Tokuda, G., Lo, N., Watanabe, H., et al. (2012). Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*. *Appl. Environ. Microbiol.* 78, 204–210. doi: 10.1128/AEM.06540-11
- Sanyika, T. W., Rashamuse, K. J., Hennesy, F., and Brady, D. (2012). Luminal hindgut bacterial diversities of the grass and sugarcane feeding termite *Trinervitermes trinervoides*. *African J. Microbiol. Res.* 6, 2639–2648.
- Scharf, M. E. (2015). Termites as targets and models for biotechnology. *Ann. Rev. Entomol.* 60, 77–102. doi: 10.1146/annurev-ento-010814-020902
- Scharf, M. E., Karl, Z. J., Sethi, A., and Boucias, D. G. (2011). Multiple levels of synergistic collaboration in termite lignocellulose digestion. *PLoS ONE* 6:e21709. doi: 10.1371/journal.pone.0021709

- Scharf, M. E., Scharf, W. D., Pittendrigh, B. R., and Bennett, G. W. (2003). Caste- and development-associated gene expression in a lower termite. *Genome Biol.* 4:R62. doi: 10.1186/gb-2003-4-10-r62
- Scharf, M. E., and Tartar, A. (2008). Termite digestomes as sources for novel lignocellulases. *Biofuels Bioprod. Bioref.* 2, 540–552. doi: 10.1002/bbb.107
- Scharf, M. E., Wu-Scharf, D., Zhou, X., Pittendrigh, B. R., and Bennett, G. W. (2005). Gene expression profiles among immature and adult reproductive castes of the termite *Reticulitermes flavipes*. *Insect Mol. Biol.* 14, 31–44. doi: 10.1111/j.1365-2583.2004.00527.x
- Schauer, C., Thompson, C., and Brune, A. (2014). Pyrotag sequencing of the gut microbiota of the cockroach *Shelfordella lateralis* reveals a highly dynamic core but only limited effects of diet on community structure. *PLoS ONE* 9:e85861. doi: 10.1371/journal.pone.0085861
- Sen, R., Raychoudhury, R., Cai, Y., Sun, Y., Lietze, V. U., Boucias, D. G., et al. (2013). Differential impacts of juvenile hormone, soldier head extract and alternate caste phenotypes on host and symbiont transcriptome composition in the gut of the termite *Reticulitermes flavipes*. *BMC Genomics* 14:491. doi: 10.1186/1471-2164-14-491
- Sen, R., Raychoudhury, R., Cai, Y., Sun, Y., Lietze, V. U., Boucias, D. G., et al. (2015). Metatranscriptomic signatures of nicotinoid-pathogen synergy in the termite gut. *PLoS ONE* (in press).
- Sethi, A., Slack, J. M., Kovaleva, E. S., Buchman, G. W., and Scharf, M. E. (2013a). Lignin-associated metagene expression in a lignocellulose-digesting termite. *Insect Biochem. Mol. Biol.* 43, 91–101. doi: 10.1016/j.ibmb.2012.10.001
- Sethi, A., Kovaleva, E. S., Slack, J. M., Brown, S., Buchman, G. W., and Scharf, M. E. (2013b). A GHF7 cellulase from the protist symbiont community of *Reticulitermes flavipes* enables more efficient lignocellulose processing by host enzymes. *Arch. Insect Biochem. Physiol.* 84, 175–193. doi: 10.1002/arch.21135
- Sillam-Dussès, D., Krasulová, J., Vrkošlav, V., Pytelková, J., Cvačka, J., Kotalová, K., et al. (2012). Comparative study of the labial gland secretion in termites. *PLoS ONE* 7:e46431. doi: 10.1371/journal.pone.0046431
- Steller, M. M., Kambhampati, S., and Caragea, D. (2010). Comparative analysis of expressed sequence tags from three castes and two life stages of the termite *Reticulitermes flavipes*. *BMC Genomics* 11:463. doi: 10.1186/1471-2164-11-463
- Stingl, U., Radek, R., Yang, H., and Brune, A. (2005). “Endomicrobia”: cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* 71, 1473–1479. doi: 10.1128/AEM.71.3.1473-1479.2005
- Tai, V., James, E. R., Nalepa, C. A., Scheffrahn, R. H., Perlman, S. J., and Keeling, P. J. (2015). The role of host phylogeny varies in shaping microbial diversity in the hindguts of lower termites. *Appl. Environ. Microbiol.* 81, 1059–1070. doi: 10.1128/AEM.02945-14
- Tai, V., James, E. R., Perlman, S. J., and Keeling, P. J. (2013). Single-Cell DNA bar-coding using sequences from the small subunit rRNA and internal transcribed spacer region identifies new species of *Trichonympha* and *Trichomitopsis* from the hindgut of the termite *Zootermopsis angusticollis*. *PLoS ONE* 8:e58728. doi: 10.1371/journal.pone.0058728
- Tai, V., and Keeling, P. J. (2013). Termite hindguts and the ecology of microbial communities in the sequencing age. *J. Eukaryot. Microbiol.* 60, 421–428. doi: 10.1111/jeu.12048
- Tartar, A., Wheeler, M. M., Zhou, X., Coy, M. R., Boucias, D. G., and Scharf, M. E. (2009). Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *R. flavipes*. *Biotechnol. Biofuels* 2:25. doi: 10.1186/1754-6834-2-25
- Tarver, M. R., Zhou, X., and Scharf, M. E. (2010). Socio-environmental and endocrine influences on developmental and caste-regulatory gene expression in the eusocial termite *Reticulitermes flavipes*. *BMC Mol. Biol.* 11:28. doi: 10.1186/1471-2199-11-28
- Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., et al. (2014). Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* 5:3636. doi: 10.1038/ncomms4636
- Thompson, G. J., Crozier, Y. C., and Crozier, R. H. (2003). Isolation and characterization of a termite transferrin gene up-regulated on infection. *Insect Mol. Biol.* 12, 1–7. doi: 10.1046/j.1365-2583.2003.00381.x
- Todaka, N., Inoue, T., Saita, K., Ohkuma, M., Nalepa, C. A., Lenz, M., et al. (2010). Phylogenetic analysis of cellulolytic enzyme genes from representative lineages of termites and a related cockroach. *PLoS ONE* 5:e8636. doi: 10.1371/journal.pone.0008636
- Todaka, N., Moriya, S., Saita, K., Hondo, T., Kiuchi, I., Takasu, H., et al. (2007). Environmental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protist community of *Reticulitermes speratus*. *FEMS Microbiol. Ecol.* 59, 592–599. doi: 10.1111/j.1574-6941.2006.00237.x
- Tokuda, G., Tsuboi, Y., Kihara, K., Saitou, S., Moriya, S., Lo, N., et al. (2014). Metabolomic profiling of ¹³C-labelled cellulose digestion in a lower termite: insights into gut symbiont function. *Proc. Biol. Sci.* 281, 1789. doi: 10.1098/rspb.2014.0990
- Vargo, E. L., and Husseneder, C. (2009). Biology of subterranean termites: insights from molecular studies of *Reticulitermes*, and *Coptotermes*. *Annu. Rev. Entomol.* 54, 379–403. doi: 10.1146/annurev.ento.54.110807.090443
- Wang, Q., Qian, C., Zhang, X. Z., Liu, N., Yan, X., and Zhou, Z. (2012). Characterization of a novel thermostable β -glucosidase from a metagenomic library of termite gut. *Enzyme Microb. Technol.* 51, 319–324. doi: 10.1016/j.enzmictec.2012.07.015
- Wang, Y., and Qian, P. Y. (2009). Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS ONE* 4:e7401. doi: 10.1371/journal.pone.0007401
- Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T. H., Stege, J. T., et al. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450, 560–565. doi: 10.1038/nature06269
- Watanabe, H., Noda, H., Tokuda, G., and Lo, N. (1998). A cellulase gene of termite origin. *Nature* 394, 330–331. doi: 10.1038/28527
- Watanabe, H., and Tokuda, G. (2010). Cellulolytic systems in insects. *Annu. Rev. Entomol.* 55, 609–632. doi: 10.1146/annurev-ento-112408-085319
- Weil, T., Korb, J., and Rehli, M. (2009). Comparison of queen-specific gene expression in related lower termite species. *Mol. Biol. Evol.* 26, 1841–1850. doi: 10.1093/molbev/msp095
- Weil, T., Rehli, M., and Korb, J. (2007). Molecular basis for the reproductive division of labour in a lower termite. *BMC Genomics* 8:198. doi: 10.1186/1471-2164-8-198
- West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford: Oxford University Press.
- Wu-Scharf, D., Scharf, M. E., Pittendrigh, B. R., and Bennett, G. W. (2003). Expressed sequence tags from a polyphenic *Reticulitermes flavipes* cDNA library. *Sociobiology* 41, 479–490.
- Xie, L., Zhang, L., Zhong, Y., Liu, N., Long, Y., Wang, S., et al. (2012). Profiling the metatranscriptome of the protistan community in *Coptotermes formosanus* with emphasis on the lignocellulolytic system. *Genomics* 99, 246–255. doi: 10.1016/j.ygeno.2012.01.009
- Yang, H., Schmitt-Wagner, D., Stingl, U., and Brune, A. (2005). Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ. Microbiol.* 7, 916–932. doi: 10.1111/j.1462-2920.2005.00760.x
- Yang, F., Xu, B., Li, J., and Huang, Z. (2012). Transcriptome analysis of *Termitomyces albuminosus* reveals the biodegradation of lignocellulose. *Wei Sheng Wu Xue Bao* 52, 466–477.
- Yuki, M., Moriya, S., Inoue, T., and Kudo, T. (2008). Transcriptome analysis of the digestive organs of *Hodotermopsis sjostedti*, a lower termite that hosts mutualistic microorganisms in its hindgut. *Zoolog. Sci.* 25, 401–406. doi: 10.2108/zsj.25.401
- Zhang, D., Lax, A. R., Henrissat, B., Coutinho, P., Katiya, N., Nierman, W. C., et al. (2012). Carbohydrate-active enzymes revealed in *Coptotermes formosanus* transcriptome. *Insect Mol. Biol.* 21, 235–245. doi: 10.1111/j.1365-2583.2011.01130.x

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Scharf. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Epigenetics as an answer to Darwin's "special difficulty"

Brian R. Herb*

Center for Epigenetics and Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Edited by:

Greg J. Hunt, Purdue University, USA

Reviewed by:

Xianyun Mao, Guardian Analytics, USA

Keijo Viiri, University of Tampere, Finland

*Correspondence:

Brian R. Herb, Center for Epigenetics and Department of Medicine, Johns Hopkins University School of Medicine, 855 North Wolfe Street, Baltimore, MD, USA
e-mail: brianherb@jhmi.edu

Epigenetic modifications produce distinct phenotypes from the same genome through genome-wide transcriptional control. Recently, DNA methylation in honeybees and histone modifications in ants were found to assist the formation of caste phenotypes during development and adulthood. This insight allows us to revisit one of Darwin's greatest challenges to his natural selection theory; the derivation of multiple forms of sterile workers within eusocial species. Differential feeding of larvae creates two distinct developmental paths between queens and workers, with workers further refined by pheromone cues. Flexible epigenetic control provides a mechanism to interpret the milieu of social cues that create distinct worker sub-caste phenotypes. Recent findings suggest a distinct use for DNA methylation before and after adult emergence. Further, a comparison of genes that are differentially methylated and transcriptionally altered upon pheromone signaling suggests that epigenetics can play a key role in mediating pheromone signals to derive sub-caste phenotypes. Epigenetic modifications may provide a molecular mechanism to Darwin's "special difficulty" and explain the emergence of multiple sub-phenotypes among sterile individuals.

Keywords: epigenetics, evolution, genomics, developmental plasticity, eusociality, behavior, pheromones

A major defining feature of eusocial species is the division of labor among phenotypically distinct castes (Winston, 1987). The evolution of such a system required the eventual partitioning of all reproductive tasks to a single individual, leaving the remaining tasks to sterile relatives. This arrangement, however, posed a great challenge to Darwin's theory of natural selection (Darwin, 1859). The bedrock of his theory was that successful individuals passed on traits to the next generation. How then, could a sterile individual possess traits distinct from reproductive individuals and not have the means to pass them on? Darwin wrote:

... one special difficulty, which at first appeared to me insuperable, and actually fatal to the whole theory. I allude to the neuters or sterile females in insect-communities; for these neuters often differ widely in instinct and in structure from both the males and fertile females, and yet, from being sterile, they cannot propagate their kind.

Lacking a thorough understanding of the underlying genetics, Darwin nonetheless had the great insight that these sterile workers, being related to the reproductive member of the colony, can ensure the survival of their species by helping the colony as a whole. This insight did not fully answer the challenge, because he goes on to marvel, not at the existence of sterile workers *per se*, which he equates to the trait divergence between males and females, but rather how can multiple sub-phenotypes of sterile workers arise:

The great difficulty lies in the working ants differing widely from both the males and the fertile females in structure, as in the shape of the thorax, and in being destitute of wings and sometimes of eyes, and in instinct. As far as instinct alone is concerned, the wonderful difference in this respect between the workers and the perfect females would have been better exemplified by the hive-bee.

Darwin's curiosity might have been further heightened if he knew that for most social insects the reproductive and sterile females are genetically identical. Phenotypic difference in the absence of genetic difference falls in the realm of epigenetics, which is the study of heritable information other than the DNA sequence itself. Epigenetic information can be stored in the molecular form as methylation on the cytosine base of DNA or a variety of modifications to histone tails (Kouzarides, 2007; Jones, 2012). These epigenetic modifications play a key role in tissue development where drastically different organs are derived from the same genome (Irizarry et al., 2009). Here we explore the role of epigenetic modifications in caste determination and propose that epigenetic machinery is important to derive the multiple forms of sterile workers that vexed Darwin so long ago.

Honeybees (*Apis mellifera*) have unique control over the developmental fate of the females in the colony through differential feeding of the larvae and pupae. A nutrient rich diet of royal jelly produces a reproductive queen, and the absence of such diet produces facultatively sterile workers. This royal jelly contains royalactin, a potent activator of p70 s6 kinase that increases ovary development and shortens development time (Kamakura, 2011). Queen development is marked by an increase in Tor activity during the third to fifth instars, stimulating growth and increased metabolism. Increased Tor activity occurs at the developmental time point when queens and workers diverge into two irreversible paths, permanently locking in caste differences. RNAi knockdown of Tor causes larvae to prolong development, reduce growth and ultimately emerge as workers, even on a diet of royal jelly (Patel et al., 2007; Mutti et al., 2011).

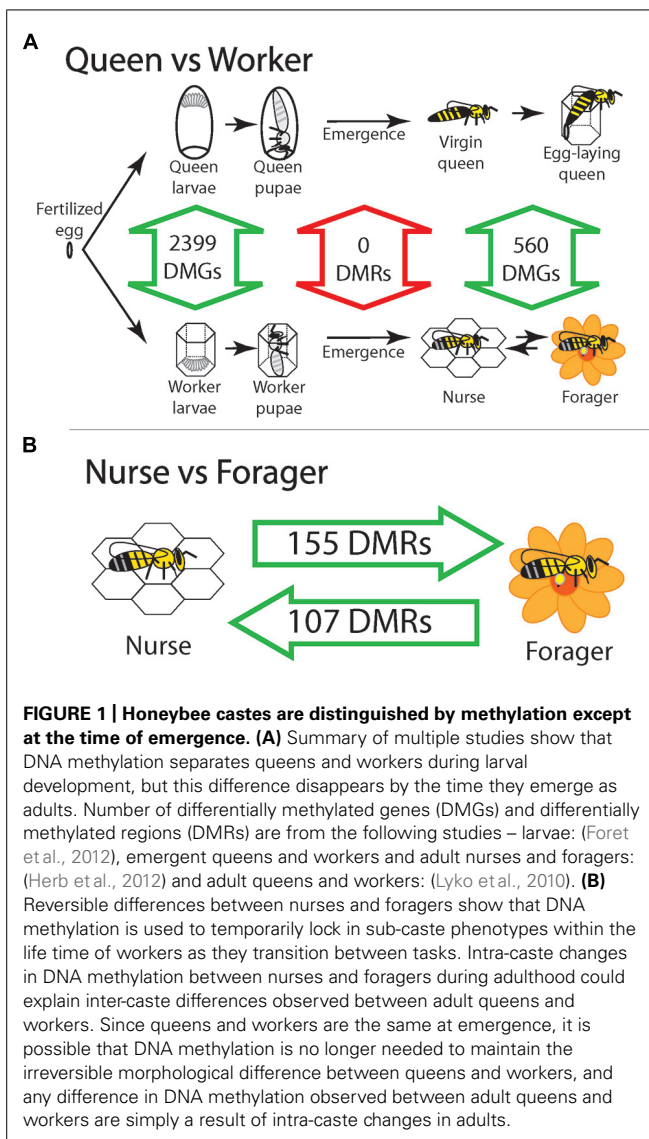
While honeybee hives have a single queen that lays millions of eggs over her 2–3 year lifespan, thousands of workers perform

the remaining tasks. The typical adult worker will first act as a nurse to raise the young, attend to the queen, and clean combs for the next generation. About 8 days post-emergence, the worker will transition into foraging tasks, which are metabolically taxing and accelerate physiological decline (Winston, 1987). However, the division of labor is dependent on the needs of the colony, and individuals within the hive are able to communicate these needs either through direct contact or by pheromones. Queen mandibular pheromone (QMP) is emitted by the queen in order to recruit nurse bees to care for her, suppress ovary growth in workers and discourage workers from raising a new queen. The brood translates its own needs by emitting brood pheromone (BP) to stimulate nurse bees to feed and care for the brood. BP also influences nurse bees to delay the transition into foraging, and existing foragers to skew their collecting toward the protein source pollen (Slessor et al., 2005). Workers further refine their tasks by physically interacting with fellow workers and recruiting them to specific tasks based on the needs of the hive (Seeley et al., 1998). Foragers themselves can also suppress nurse bees from foraging by emitting Ethyl oleate (Leoncini et al., 2004). So it is in this environment of constant signals that the worker bee refines her role throughout life. These signals form a basis for unlocking multiple phenotypes, the marvel of Darwin 150 years ago.

Honeybees use diet and social cues to separate genetically similar females into distinct roles, but what is the underlying molecular mechanism that integrates environmental stimuli and solidifies phenotype? Lacking a strong genetic candidate, epigenetic modifications can drive differentiation of multiple phenotypes as seen with cellular lineages in blood (Ji et al., 2010). The beauty of epigenetic mechanisms is that they can assist in maintaining a particular transcriptional state by storing information in the form of temporary chemical tags at the level of DNA itself. DNA methylation and histone modifications have been thoroughly studied in mammals and are known to play a major role in development and disease (Ho and Crabtree, 2010; Hansen et al., 2011). Genome-wide epigenetic modifications, like DNA methylation, can be context specific depending on their placement relative to genes and enhancers. A unique combination of DNA methylation and histone modifications in the promoter of a given gene can have a persistent repressive effect when these marks are bound by proteins that in turn establish larger protein complexes that as a whole suppress transcription. A good example of this process occurs during mammalian differentiation where pluripotency genes such as *OCT4* and *NANOG* are silenced by methylation of H3K9 by G9a, which in turn leads to condensing of chromatin by HP1 binding and eventual DNA methylation (Feldman et al., 2006; Smith and Meissner, 2013). This step-wise change in epigenetic modifications indicates different degrees of repression that become increasingly resistant to activation. These epigenetic modifications can be reversed, but require persistent signals, such as the expression of the reprogramming factors *OCT4*, *SOX2*, *MYC*, and *KLF4* to derive induced pluripotent stem cells (iPS; Doi et al., 2009). Another example of dynamic epigenetic change is during the activation of the *pS2* gene upon estrogen signaling. Time-course experiments showed active demethylation of DNA and recruitment of chromatin

remodeling proteins to the site of the *pS2* gene after estrogen signaling (Metivier et al., 2008). In addition to chemical modifications to DNA and histone tails, RNA itself can provide temporal control of gene expression through the binding of non-coding RNAs to DNA, proteins or other RNAs. Non-coding RNAs can organize chromatin structure on a large scale as evidenced by X chromosome inactivation by the ncRNA *Xist*, or control local expression in the case of the ncRNA *Air* interacting with the histone methyltransferase G9a to silence the *Slc22a3* gene during development (Nagano et al., 2008; Mercer and Mattick, 2013). Studies investigating DNA methylation differences between worker subcastes in honeybees (Herb et al., 2012) and between queens and workers in ants (Bonasio et al., 2012) have found that many differentially methylated genes are involved in non-coding RNA processing, suggesting a role for non-coding RNA in caste determination. While the fundamental role and scope of non-coding RNA in mammalian development is established (Mattick, 2011), the impact of non-coding RNA in social insects is just starting to be understood (Bonasio, 2012; Humann et al., 2013), therefore the focus of this perspective will only include epigenetic modifications that have been mapped genome-wide, namely DNA methylation and histone modifications. Overall, the temporal control of epigenetic modifications enforces a specific transcriptional state by storing information at the level of the DNA itself, which remembers that state until a new stimuli is encountered.

Only recently has the importance of epigenetics in social insects been appreciated through the discovery of DNA methylation in many species (Bonasio et al., 2010; Beeler et al., 2014). Social insects are an ideal test ground for studying the role of epigenetic mechanisms because they can derive multiple behavioral phenotypes from the same genome. The first major clue that epigenetics played a role in queen/worker differentiation came soon after the complete sequencing of the honeybee genome in 2006 (Consortium, 2006) when the presence of DNA methyltransferase enzymes confirmed a functional DNA methylation system (Wang et al., 2006). Kucharski et al. (2008) knocked down Dnmt3 in larvae and found that regardless of diet, most knockdowns developed queen features. This initial result inspired a genome-wide search for functional DNA methylation differences between queens and workers that resulted in three major studies that interrogated three developmental time points; larvae (Foret et al., 2012), adult emergence (Herb et al., 2012), and advanced age adults (Lyko et al., 2010). While differences between queens and workers were found across the genome in larvae and advanced age adults, there were no statistically significant differences at the time of adult emergence (Figure 1A). While these studies take different approaches to find regional changes in DNA methylation, the large number of differences found in larvae compared to the complete absence of differences between queens and workers strongly suggest that DNA methylation is required to maintain queen/worker differences during the larval stage, but are not required to separate newly emerged queens and workers when morphological differences are irreversible. DNA methylation appears to target many genes of the Tor pathway (Mutti et al., 2011; Foret et al., 2012), which has been implicated in queen worker developmental differentiation (Patel et al., 2007). It is possible that DNA



methylation assists in maintaining the activation of Tor pathway genes caused by royal jelly. Returning to Darwin's difficulty, we see that DNA methylation can assist in maintaining separate transcriptional programs for queens and workers in honeybees, providing a mechanism for maintaining caste differences. Further proof that epigenetic mechanisms help produce alternative phenotypes is illustrated by the finding that histone modifications, in particular H3K27ac, differentiate major from minor workers in the carpenter ant *Camponotus floridanus* (Simola et al., 2013). The remarkable size difference between ant worker sub-castes was particularly striking to Darwin and this result illustrates that epigenetic modifications, including histone modifications, can help produce multiple sterile worker phenotypes (Darwin, 1859). This example from ants bolsters the idea of epigenetic modifications solidifying differences initiated by diet during development, but how do epigenetic modifications help individuals navigate transitions throughout adult life as seen above with nurses and foragers?

While worker bees generally transition from nursing to foraging tasks over their lifetime, the timing of this transition and exact task they perform at any given point along this continuum is refined by social cues within the hive (Slessor et al., 2005). Specific pheromones can elicit a change in expression of hundreds of genes and recruit workers to a task or delay their transition into a new task (Grozing et al., 2003; Alaux et al., 2009). Although powerful, these signals must be regarded in the context of the hive, where the organization of the brood in the center and the storage of pollen and honey on the periphery create "task zones." Workers are born into a region where the queen is actively laying eggs and pheromones from the queen and brood are strongest and influence the newborn worker to assume nursing tasks. As workers age, they encounter returning foragers that present recruitment signals to elicit the nurses to transition into new roles (Winston, 1987; Whitfield, 2003). However, if upon the first interaction with a returning forager, a nurse flew out of the nest and began collecting nectar, or inversely upon a whiff of queen pheromone foragers began caring for the brood, the hive would be in chaos. Instead, tasks are performed for continuous periods and a transition requires repeated cues to initiate. The flexible control that epigenetics offers is an ideal mechanism for interpreting social cues within the hive and provide temporal control over gene expression.

Workers generally start their adult lives performing nursing tasks and transition into foraging tasks, but it is possible to revert foragers back to nursing tasks if the need arises (Amdam et al., 2005). This reversion schema includes two types of nurses, one set that has always performed nursing tasks, and one set that has had foraging experience that reverts back to nursing. When age-matched nurses, foragers and reverted nurses were compared, hundreds of differentially methylated regions (DMRs) distinguished these phenotypes. Incredibly 57 DMRs followed the behavioral reversion where DNA methylation levels changed during the nurse to forager transition, and changed back to original nurse levels during the reversion (Figure 1B). Genes associated with these 57 reversion DMRs had far reaching developmental and gene regulatory functions, including multiple genes containing DEAD-box helicase domains that act through chromatin remodeling to affect global gene expression. Importantly, distinct nurse and forager specific epigenetic signatures were identified, demonstrating that specific levels of DNA methylation in the brain are required to form sub-caste phenotypes. This study was also the first evidence of reversible methylation underlying a behavioral trait and demonstrates the flexible control epigenetic modifications can bestow on phenotype (Herb et al., 2012).

If pheromones and other social interactions impact the epigenome, then the socially defined task is remembered at the level of the genome itself. This also ensures an intrinsic "buffering" against instantly changing task upon the random encounter with workers performing different tasks. This buffering would ensure that only persistent social cues, reflective of the true needs of the hive, would cause the worker to switch tasks. This is not meant to detract from or oversimplify the complex network of social cues in the hive that organize labor in such an efficient way. Rather, coupling the flexible control of epigenetics with the spatial and social cues in the hive establishes a framework for understanding how

workers can achieve stabilized sub-caste phenotypes. If true, then genes under the influence of pheromones should also be regulated by epigenetic modifications. The influences of two major pheromones, QMP and BP, have been studied at the genome level using gene expression microarrays. Key genes *POE*, active in neuro-development, and the chromatin remodeling genes *HCF* and *ISWI* are all chronically regulated by QMP (Grozinger et al., 2003) and are also differentially methylated between nurses and foragers (Herb et al., 2012). In addition, genes regulated by BP and those differentially methylated between nurses and foragers (Alaux et al., 2009) share functional enrichment of helicase and nucleoside binding genes. BP influences the expression of the developmental genes *WIT*, *UNR*, and *BICD* (Alaux et al., 2009) and these genes are associated with reversible methylation between nurses and foragers (Herb et al., 2012). Epigenetic control of a core group of master regulatory genes may be sufficient to help maintain pheromone-induced phenotypes and help stabilize the division of labor in the hive. For example, the chromatin-remodeling gene *ISWI*, which is regulated by QMP and differentially methylated between nurses and foragers, remodels nucleosomes around gene promoters (Sala et al., 2011). The action of *Iswi* may facilitate the large-scale gene expression differences observed during the nurse to forager transition (Whitfield, 2003).

Social insects are masters at controlling the division of labor within their colonies to maximize efficiency and react to changing environmental conditions. However, the evolution of the sterile worker proved troublesome to Darwin, who struggled to incorporate the existence of multiple phenotypically distinct sterile workers in his natural selection theory that emphasized passing traits directly to the next generation. Through differential feeding and social cues, the colony as a whole has evolved numerous mechanisms to fine-tune the phenotypes of sterile workers to obtain an efficient division of labor. We can now integrate the action of epigenetic machinery to the evolution of social insects. Epigenetic information stabilizes phenotype and provides a mechanism for deriving multiple castes from the same genome. This extra layer of information works with established signaling pathways and regulatory programs to lock in gene expression patterns and interpret external stimuli. DNA methylation appears to play two major roles, distinguishing queens and workers during development and defining sub-castes within the lifetime of worker bees. Further, based on the limited presence of methylation across the honeybee genome, it seems that methylation has been reserved to act on select genes that have far reaching effects. These key genes are regulated by epigenetics but are initiated by social cues within the hive, directing the division of labor. As seen with histone modifications in ants and DNA methylation in honeybees, epigenetics plays an important role in social insects. Perhaps it will bear out that utilizing epigenetic machinery to derive additional worker phenotypes was critical to the evolution of eusociality in insects.

REFERENCES

- Alaux, C., Le Conte, Y., Adams, H. A., Rodriguez-Zas, S., Grozinger, C. M., Sinha, S., et al. (2009). Regulation of brain gene expression in honey bees by brood pheromone. *Genes Brain Behav.* 8, 309–319. doi: 10.1111/j.1601-183X.2009.00480.x
- Amdam, G. V., Aase, A. L., Seehuus, S. C., Kim Fondrk, M., Norberg, K., and Hartfelder, K. (2005). Social reversal of immunosenescence in honey bee workers. *Exp. Gerontol.* 40, 939–947. doi: 10.1016/j.exger.2005.08.004
- Beeler, S. M., Wong, G. T., Zheng, J. M., Bush, E. C., Remnant, E. J., Oldroyd, B. P., et al. (2014). Whole-genome DNA methylation profile of the jewel wasp (*Nasonia vitripennis*). *G3 (Bethesda)* 4, 383–388. doi: 10.1534/g3.113.008953
- Bonasio, R. (2012). Emerging topics in epigenetics: ants, brains, and noncoding RNAs. *Ann. N. Y. Acad. Sci.* 1260, 14–23. doi: 10.1111/j.1749-6632.2011.06363.x
- Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., et al. (2012). Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* 22, 1755–1764. doi: 10.1016/j.cub.2012.07.042
- Bonasio, R., Zhang, G., Ye, C., Mutti, N. S., Fang, X., Qin, N., et al. (2010). Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* 329, 1068–1071. doi: 10.1126/science.1192428
- Consortium, T. H. G. S. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443, 931–949. doi: 10.1038/nature05260
- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. London: John Murray.
- Doi, A., Park, I.-H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., et al. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* 41, 1350–1353. doi: 10.1038/ng.471
- Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., et al. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Mol. Cell Biol.* 8, 188–194. doi: 10.1038/ncb1353
- Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S. E., Robinson, G. E., et al. (2012). DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honeybees. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4968–4973. doi: 10.1073/pnas.1202392109
- Grozinger, C. M., Sharabash, N. M., Whitfield, C. W., and Robinson, G. E. (2003). Pheromone-mediated gene expression in the honey bee brain. *Proc. Natl. Acad. Sci. U.S.A.* 100(Suppl. 2), 14519–14525. doi: 10.1073/pnas.2335.884100
- Hansen, K. D., Timp, W., Bravo, H. C., Sabuncuyan, S., Langmead, B., McDonald, O. G., et al. (2011). Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* 43, 768–775. doi: 10.1038/ng.865
- Herb, B. R., Wolschin, F., Hansen, K. D., Aryee, M. J., Langmead, B., Irizarry, R., et al. (2012). Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nat. Neurosci.* 15, 1371–1373. doi: 10.1038/nn.3218
- Ho, L., and Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature* 463, 474–484. doi: 10.1038/nature08911
- Humann, F. C., Tiberio, G. J., and Hartfelder, K. (2013). Sequence and expression characteristics of long noncoding RNAs in honey bee caste development—potential novel regulators for transgressive ovary size. *PLoS ONE* 8:e78915. doi: 10.1371/journal.pone.0078915
- Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41, 178–186. doi: 10.1038/ng.298
- Ji, H., Ehrlich, L. I. R., Seita, J., Murakami, P., Doi, A., Lindau, P., et al. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* 467, 338–342. doi: 10.1038/nature09367
- Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13, 484–492. doi: 10.1038/nrg3230
- Kamakura, M. (2011). Royalactin induces queen differentiation in honeybees. *Nature* 473, 478–483. doi: 10.1038/nature10093
- Kouzariadis, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705. doi: 10.1016/j.cell.2007.02.005
- Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319, 1827–1830. doi: 10.1126/science.1153069
- Leoncini, I., Le Conte, Y., Costagliola, G., Plettner, E., Toth, A. L., Wang, M., et al. (2004). Regulation of behavioral maturation by a primer pheromone produced by adult worker honey bees. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17559–17564. doi: 10.1073/pnas.0407652101
- Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* 8:e1000506. doi: 10.1371/journal.pbio.1000506

- Mattick, J. S. (2011). The central role of RNA in human development and cognition. *FEBS Lett.* 585, 1600–1616. doi: 10.1016/j.febslet.2011.05.001
- Mercer, T. R., and Mattick, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* 20, 300–307. doi: 10.1038/nsmb.2480
- Metivier, R., Gallais, R., Tiffocche, C., Le Peron, C., Jurkowska, R. Z., Carmouche, R. P., et al. (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 452, 45–50. doi: 10.1038/nature06544
- Mutti, N. S., Dolezal, A. G., Wolschin, F., Mutti, J. S., Gill, K. S., and Amdam, G. V. (2011). IRS and TOR nutrient-signaling pathways act via juvenile hormone to influence honey bee caste fate. *J. Exp. Biol.* 214, 3977–3984. doi: 10.1242/jeb.061499
- Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R., et al. (2008). The air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322, 1717–1720. doi: 10.1126/science.1163802
- Patel, A., Fondrk, M. K., Kaftanoglu, O., Emore, C., Hunt, G., Frederick, K., et al. (2007). The making of a queen: TOR pathway is a key player in diphenic caste development. *PLoS ONE* 2:e509. doi: 10.1371/journal.pone.0000509
- Sala, A., Toto, M., Pinello, L., Gabriele, A., Di Benedetto, V., Ingrassia, A. M., et al. (2011). Genome-wide characterization of chromatin binding and nucleosome spacing activity of the nucleosome remodelling ATPase ISWI. *EMBO J.* 30, 1766–1777. doi: 10.1038/emboj.2011.98
- Seeley, T. D., Weidenmuller, A., and Kuhnholz, S. (1998). The shaking signal of the honey bee informs workers to prepare for greater activity. *Ethology* 104, 10–26. doi: 10.1111/j.1439-0310.1998.tb00026.x
- Simola, D. F., Ye, C., Mutti, N. S., Dolezal, K., Bonasio, R., Liebig, J., et al. (2013). A chromatin link to caste identity in the carpenter ant *Camponotus floridanus*. *Genome Res.* 23, 486–496. doi: 10.1101/gr.148361.112
- Slessor, K. N., Winston, M. L., and Le Conte, Y. (2005). Pheromone communication in the honeybee (*Apis mellifera* L.). *J. Chem. Ecol.* 31, 2731–2745. doi: 10.1007/s10886-005-7623-9
- Smith, Z. D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* 14, 204–220. doi: 10.1038/nrg3354
- Wang, Y., Jorda, M., Jones, P. L., Maleszka, R., Ling, X., Robertson, H. M., et al. (2006). Functional CpG methylation system in a social insect. *Science* 314, 645–647. doi: 10.1126/science.1135213
- Whitfield, C. W. (2003). Gene expression profiles in the brain predict behavior in individual honey bees. *Science* 302, 296–299. doi: 10.1126/science.1086807
- Winston, M. L. (1987). *The Biology of the Honey Bee*. Cambridge, MA: Harvard University
- Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 July 2014; paper pending published: 08 August 2014; accepted: 27 August 2014; published online: 12 September 2014.

Citation: Herb BR (2014) Epigenetics as an answer to Darwin's "special difficulty." *Front. Genet.* 5:321. doi: 10.3389/fgene.2014.00321

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Herb. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Epigenetics as an answer to Darwin's "special difficulty," Part 2: natural selection of metastable epialleles in honeybee castes

Douglas M. Ruden^{1*}, Pablo E. Cingolani², Arko Sen³, Wen Qu³, Luan Wang⁴, Marie-Claude Senut⁴, Mark D. Garfinkel⁵, Vincent E. Sollars⁶ and Xiangyi Lu⁴

¹ Department of Obstetrics and Gynecology, C. S. Mott Center for Human Growth and Development and Center for Urban Responses to Environmental Stressors, Institute of Environmental Health Sciences, Wayne State University, Detroit, MI, USA

² School of Computer Science and Genome Quebec Innovation Centre, McGill University, Montreal, QC, Canada

³ Department of Pharmacology, Wayne State University, Detroit, MI, USA

⁴ Institute of Environmental Health Sciences, Wayne State University, Detroit, MI, USA

⁵ Department of Biological Sciences, University of Alabama in Huntsville, Huntsville, AL, USA

⁶ Department of Biochemistry and Microbiology, Joan C. Edwards School of Medicine, Marshall University, Huntington, WV, USA

Edited by:

Greg J. Hunt, Purdue University, USA

Reviewed by:

Aaron Arthur Comeault, University of Sheffield, UK

Feng-Chi Chen, National Health Research Institutes, Taiwan

*Correspondence:

Douglas M. Ruden, Department of Obstetrics and Gynecology, C. S. Mott Center for Human Growth and Development and Center for Urban Responses to Environmental Stressors, Institute of Environmental Health Sciences, Wayne State University, 275 East Hancock, Room 002, Detroit, MI 48201, USA
e-mail: douglasr@wayne.edu

In a recent perspective in this journal, Herb (2014) discussed how epigenetics is a possible mechanism to circumvent Charles Darwin's "special difficulty" in using natural selection to explain the existence of the sterile-fertile dimorphism in eusocial insects. Darwin's classic book "On the Origin of Species by Means of Natural Selection" explains how natural selection of the fittest individuals in a population can allow a species to adapt to a novel or changing environment. However, in bees and other eusocial insects, such as ants and termites, there exist two or more castes of genetically similar females, from fertile queens to multiple sub-castes of sterile workers, with vastly different phenotypes, lifespans, and behaviors. This necessitates the selection of groups (or kin) rather than individuals in the evolution of honeybee hives, but group and kin selection theories of evolution are controversial and mechanistically uncertain. Also, group selection would seem to be prohibitively inefficient because the effective population size of a colony is reduced from thousands to a single breeding queen. In this follow-up perspective, we elaborate on possible mechanisms for how a combination of both epigenetics, specifically, the selection of metastable epialleles, and genetics, the selection of mutations generated by the selected metastable epialleles, allows for a combined means for selection amongst the fertile members of a species to increase colony fitness. This "intra-caste evolution" hypothesis is a variation of the epigenetic directed genetic error hypothesis, which proposes that selected metastable epialleles increase genetic variability by directing mutations specifically to the epialleles. Natural selection of random metastable epialleles followed by a second round of natural selection of random mutations generated by the metastable epialleles would allow a way around the small effective population size of eusocial insects.

Keywords: epigenetics, evolution, genomics, developmental plasticity, eusociality, group selection

DARWIN'S "SPECIAL PROBLEM"

Eusocial (Greek *eu*: "good/real" + "social") insects include the Hymenoptera (ants, bees, and wasps) and the Isoptera (termites). In honeybees, which is the focus of this perspective, a hive has caste differences; the diploid queen and the haploid drones are the sole reproducers, while the nurses, soldiers, guards, and foragers are "sub-castes" or "task groups" of the worker caste of sterile females that work together to benefit the group as a whole (Free, 1987). However, as pointed out by Herb (2014) in a previous perspective in this journal, having sterile females in a colony is a potentially fatal flaw in Darwin's theory of natural selection, which states that the fittest individuals pass their traits (i.e., genes) to the next generation. Darwin (1859) referred to sterile workers in insect communities as, "... one special difficulty, which at first appeared

insuperable, and actually fatal to the whole theory." Darwin (1871) later proposed a way around this "special difficulty" by proposing a "group selection" model for evolution of altruistic behaviors in eusocial insects. Darwin argued that "group selection" can occur when the benefits of altruism between castes are greater than the individual benefits of selfishness (egotism) within a subpopulation. Hamilton (1964), the great population geneticist Hamilton formalized the idea of group selection in a mathematical model, $rb > c$, where b represents the benefit to the recipient of altruism, c the cost to the altruist, and r their degree of relatedness. Kin selection takes into account the genetic relatedness of individuals in a group, was a further refinement of the group selection theory. In kin selection, $rb_k + b_e > c$, in which b_k is the altruistic benefit to kin and b_e is the altruistic benefit accruing to the group

as a whole (Wilson and Wilson, 2007). However, group- and kin-selection models are mathematically complex and remain controversial amongst many evolutionary theorists, such as Dawkins (1976) and Nowak et al. (2010), who argue that group selection is unlikely because, among many reasons, selfishness (i.e., “selfish genes” – a phrase Dawkins coined) would always predominate over altruism.

Here, we further elaborate on Herb's (2014) thesis that epigenetics might be a way around Darwin's “special difficulty.” We argue that epigenetic inheritance systems (EISs) can allow rapid evolution of traits specific for sterile workers and fertile queens. Epigenetics does not involve changes in the DNA sequence, but rather covalent, yet reversible, changes to the DNA in the form of 5-methylcytosine (5mC). EISs should work fine for short-term evolutionary changes. However, natural selection of DNA sequence variants would still be needed for long-term evolutionary changes. Histone modifications, long non-coding RNAs, prions, and other types of EISs will not be discussed in detail, but rather we will focus on 5mC, since 5mC represents a reversible change to the genome that can be modified by the environment (Chia et al., 2011). Heritable changes in 5mC, such as occurs in imprinted genes in mammals, are also called metastable epialleles (Rakyan et al., 2002; Dolinoy et al., 2007). The most important aspect of DNA methylation in the hypothesis presented in this paper is that, unlike histone modifications, 5mC is mutagenic and can lead to permanent changes to the DNA. Specifically, 5mC can undergo spontaneous deamination, which converts 5mC to T (Coulondre et al., 1978; Duncan and Miller, 1980). A hypothesis for how natural selection of metastable epialleles can lead to DNA mutations that permanently stabilize the epialleles into real alleles, the epigenetic directed genetic error (EDGE) hypothesis, is presented in the last section of this perspective.

The inspiration for many of the ideas in this perspective is a chapter in Jablonka and Lamb's (2005) excellent book “Evolution in Four Dimensions” on EISs. They created an imaginary planet named Jaynus where the variety of organisms all had exactly the same genome sequences, yet had many different phenotypes. They wrote:

Jaynus organisms have a genetic system that is based on DNA, and replication transcription, and translation are much the same as on Earth. However, there is one very extraordinary thing about the DNA of Jaynus creatures – every organism has exactly the same DNA sequences. From the simplest organism, a tiny unicellular creature, to the enormous fan-like colonial worms, the DNA is identical. Their genomes are large and complex, but no organism deviates from the universal standard sequences because there are cellular systems that check DNA and destroy any cell suspected of carrying a mutation.

In this perspective, we describe how intra-caste evolution is, in many respects, similar to how evolution proceeds on the mythical planet of Jaynus.

MECHANISMS OF CASTE DETERMINATION AND EPIGENETIC MODIFICATION IN HONEYBEES

At the extreme superorganismic phase, the level of selection becomes the genome of the queen and the sperm she stores, and the workers can be viewed as robotic extensions of her phenotype (Wilson and Nowak, 2014).

As beautifully described in the above quote, eusocial insects are even more extreme in some respects than the mythical organisms on Jaynus because they have evolved to a “superorganismic” stage in which the queen is the reproductive organ(ism) and the workers are the “robotic extensions” or the somatic cells of the superorganism. Kennedy et al. (2014) called the formation of eusocial insect colonies and super colonies “a new major transition in evolution.” We propose that the intra-caste evolutionary process in honeybees might share epigenetic mechanisms with those proposed on Jaynus. Darwin (1859) evolution utilizes the concepts of “survival of the fittest” in a population and “natural selection” of genetic variation to eventually form new species. Darwin, of course, did not know about either genetic or epigenetic variation, for his work pre-dated Gregor Mendel's discoveries (or, more accurately, rediscovery in the 20th century (Meneses Hoyos, 1960; Fairbanks and Rytting, 2001), but the modern interpretation of “natural selection” is selection of genetic variation. We propose that “intra-caste evolution” is a type of micro-evolution, which is small-scale evolution within a population, and refers to survival of the fittest members of a caste. We propose here that “intra-caste evolution” is initially based on natural selection of metastable epialleles of the most-fit caste (i.e., queen and worker) and sub-caste members (i.e., nurse, soldier, guard, and forager). As in the mythical Jaynus example, genetic selection probably cannot be the primary mechanism for selecting the fittest worker bee, since most worker bees cannot breed (however, see below). For example, the most efficient forager sub-caste of workers cannot be selected for by direct genetic selection, since workers, in most situations, are sterile females. However, a hive with more efficient foragers can be produced by group selection of metastable epialleles that produce an increased foraging efficiency. One phenotype that worker bees have evolved to increase foraging efficiency are the pollen baskets on the hind legs, which are present on workers but not on queens. A possible mechanism for the evolution of pollen baskets is presented later in this perspective.

If the “intra-caste evolution” hypothesis of honeybee castes is not mediated primarily by genetic means, then how are the desirable phenotypes, such as efficiency in foragers, transmitted to the next generation? We propose that, first, queens undergo a great deal of stress (i.e., malnutrition) when there is not an adequate amount of foraging being performed by the workers. The stress, in a mechanism that we present in a later section, leads to an activation of random stress-induced metastable epialleles, some of which increase the food-carrying capacity of pollen baskets. Second, the metastable epialleles which improve the fitness of the colony, such as those that serendipitously alter pollen baskets in workers in a manner that increases storage capacity, are selected over several generations by group selection. Third, random mutations can potentially be directed to the selected metastable epialleles by the EDGE mechanism, described in the final section of this perspective. The main reason for the need for the EDGE hypothesis is, we believe, because the “normal” background mutation and group selection processes are not adequate when the effective population size of a species is too low, as it arguably is in eusocial insects (i.e., only the queen breeds). The EDGE hypothesis provides an additional mechanism to increase

the mutation rate of specific genes required to ensure the survival of the colony.

CHEMICAL MEANS OF EPIGENETIC MODIFICATIONS IN HONEYBEE CASTES

In addition to the selection of the most-fit caste and sub-caste members in each generation by group selection, honeybees have evolved to produce royal jelly to alter the epigenetic and developmental machinery of their offspring. The active ingredients of royal jelly include a fatty acid, (E)-10-hydroxy-2-decenoic acid (10HDA), which accounts for up to 5% of royal jelly. The fatty acid 10HDA, interestingly, is an epigenetic modifier molecule with a histone deacetylase inhibitor (HDACi) activity (Spannhoff et al., 2011). HDACs remove acetyl groups from histones, which are present in actively transcribed genes to open up the chromatin, presumably by repellent ionic charges pushing the nucleosomes apart (Jenuwein and Allis, 2001). HDACi's inhibit the deacetylation of histones, which would lead to the acetyl groups remaining on histones, and therefore transcriptional activity would be high in the "queen-specific genes" of larvae fed royal jelly. Another component of royal jelly is the protein royalactin, which increases body size and ovary development in queens (Kamakura, 2011). The mechanism of action of royalactin is thought to be multifold: activation of mitogen-activated protein kinase (MAPK), which decreases developmental time, activation of p70 S6 kinase, which increases body size, and increasing juvenile hormone production, which is an essential hormone for ovary development (Kamakura, 2011). Interestingly, the same paper also showed that royalactin dramatically increases body size and ovary development when fed to the fruit fly, *Drosophila melanogaster* (Kamakura, 2011).

Based on the fact that royalactin has similar effects on the solitary fruit fly as on eusocial bees, we propose a theoretical epigenetic mechanism for how the queen's dependence on royal jelly for ovary development evolved. In our model, bees originally were solitary, like fruit flies, and every female fended for herself in terms of feeding and reproduction. However, when the food is in short supply, the absence of nutrients would lead to a reduction of reproductive fitness and a diminution in ovary development. Consequently, the population reaches a bottleneck when the food runs low, and only those few individuals that have sufficient nutrition survive. If the few survivors evolved the capacity to feed some of their offspring, which would be one of the first steps in eusocial evolution, then when the food runs too low, they can feed adequately only some of their offspring and leave the other offspring malnourished. The female offspring that are fed would develop ovaries, whereas the female offspring that were not sufficiently fed would develop atrophied ovaries and would be sterile. A decrease in reproductive fitness is a universal character of most animals during starvation (Carey et al., 2008). However, and this is key, both fertile and sterile offspring are produced by the same mother, in a manner that is dependent on how much food or the quality of food they were fed. If the sterile offspring provided a selective-advantage to the group as a whole, then those mothers that produced both fertile and sterile offspring would have a selective advantage over those mothers that produced only fertile offspring. After millions of years of fine-tuning this process,

the honeybee sterile-fertile dimorphism could have theoretically evolved by group selection.

As pointed out by an anonymous reviewer, there are at least three potential problems with our hypothesis on how honeybees evolved to produce royal jelly. First, the feeding behavior would need to be developed when food was scarce. Second, the sterile-fertile dimorphism would have to be maintained even though food became abundant again. Third, altruism would have to be developed when the sterile-fertile dimorphism emerged. It is hard to argue around these criticisms for a solitary insect such as *Drosophila*, and that might be why *Drosophila* and other solitary insects never evolved a sterile-fertile dimorphism. However, as suggested by Wilson and Nowak (2014), perhaps a way to circumvent all of these problems is the fact that the first step in eusocial evolution is probably the ability to form nests or colonies. This would allow the development of the dichotomy in bees, wasps, and ants of being a forager or staying in the nest to lay eggs. Since foraging is dangerous and taxing, if the workers are bringing the proto-queen pollen and nectar, then she is less inclined to forage for it. Once the proto-queen evolved the ability to produce royal jelly, then she would become the only fertile member of the colony – all of the workers could be chemically sterilized by withholding royal jelly. The development of altruism, in this case, could be an emergent property of the sterile-fertile dimorphism. As discussed further in a later section, there are many examples of emergent behavior in eusocial insects (Johnson, 2001), and we argue that altruism of sterile workers could be one of them.

In addition to royal jelly, honeybees have evolved an arsenal of other chemical weapons that subvert developmental and behavioral processes in the workers. For example, after about 8 days post-emergence, the nurse bees who take care of the eggs will transition into foragers, and foraging is more metabolically taxing because it requires the filling of baskets on the hind legs to transport the pollen (Free, 1987). However, the behavioral transition from nurse to forager depends on the needs of the hive, and individuals in the hive transmit these needs by both direct contact and by pheromones. The transition among sub-caste members in honeybees, but not some ant species that have physically different worker castes (Wilson and Nowak, 2014), is purely behavioral because all honeybee workers have pollen baskets despite the fact that only foragers use them – pollen baskets do not develop *de novo* in the nurse when she transforms into a worker.

Queen mandibular pheromone (QMP) is emitted by the queen to recruit nurses to her and to suppress ovary growth (Free, 1987). The larvae emit brood pheromone (BP) to stimulate nurse bees to feed and care for them (i.e., the brood). BP affects the nurses and foragers in different manners: it stimulates nurses to care for the brood and to delay their transition into foragers, while it stimulates foragers to collect nutrient-rich pollen to feed the brood (Slessor et al., 2005). Foragers, in turn, emit ethyl oleate (EO) to suppress nurse honeybees from foraging (Leoncini et al., 2004). Isoamyl acetate, which has a similar odor to the banana and pear, was found in Boch et al. (1962) to be an active component in the sting pheromone of the honeybee which is presumably released by honeybee guards when a hive is disturbed. It is through such chemical

(i.e., environmental) signals that the honeybees are able to epigenetically maintain the caste structure in a manner that circumvents, in most aspects, the need for the selection of genetic variation.

THE POTENTIAL ROLE OF NATURAL SELECTION OF GENETIC AND EPIGENETIC VARIATION IN DRIVING THE EVOLUTION OF CASTE SYSTEMS

As mentioned earlier, an important consideration regarding genetic-stabilization of the sterile-fertile dimorphism, is that the sterile workers can, in rare cases, develop ovaries. Removal of a queen can cause some workers to develop ovaries, in part because they are no longer exposed regularly to QMP (Herb et al., 2012, 2013). Also, nurses and foragers can revert back-and-forth rapidly in either direction in a manner that is dependent on the needs of the hive (Amdam et al., 2005). We believe that the occasional reversion of a sterile worker to a reproductive female is a critical mechanism for transmitting both metastable epialleles and genetic variation that is required for the worker caste. According to the EDGE hypothesis, genetic variation that is induced by the metastable epialleles, can be selected to increase worker specialization in the next generation. The purpose of the metastable epialleles, in the EDGE hypothesis, are not to circumvent the need for genetic variation, but rather to increase genetic variation in precisely the genes that need to be adapted for the organism, or superorganism in the case of eusocial insects, to survive the novel environment.

The selection of metastable epialleles in *Drosophila* is well-established in our laboratory (Ruden et al., 2003, 2008, 2009; Sollars et al., 2003; Ruden and Lu, 2008) and in other laboratories (Carrera et al., 1998; Ruden and Lu, 2008; Tariq et al., 2009; Gangaraju et al., 2010; Valtonen et al., 2012; Branco and Lemos, 2014; Le Thomas et al., 2014; Nystrand and Dowling, 2014; Somer and Thummel, 2014; Stern et al., 2014; Wei et al., 2015). We showed, for instance, that stress, or the inactivation of the chaperone protein Hsp90, can activate a metastable epiallele of the *Krueppel*^{Incomplete facets-1} (Kr^{If-1}) allele, which causes ectopic large bristle outgrowths (ELBOs) to protrude from the eyes (Sollars et al., 2003). We indicate the metastable epiallele with the nomenclature $[Kr^{If-1}]^*$ and showed that the metastable epiallele can be transmitted through both the male and female germlines for tens or even hundreds of generations (Ruden et al., 2003, 2008). What makes a metastable epiallele an example of an epigenetic variant rather than a genetic variant is the fact that a metastable epiallele, such as $[Kr^{If-1}]^*$, can be reverted back to the original allele, in this case Kr^{If-1} , in just one or two generations by negative selection (Sollars et al., 2003). Since *Drosophila* has very little DNA methylation, the metastable epialleles in *Drosophila* are probably not the result of differential DNA methylation. However, Gangaraju et al. (2010) presented evidence that the $[Kr^{If-1}]^*$ metastable epiallele requires Piwi and Pi RNAs, which are small non-coding RNAs in the germline and function similarly to siRNAs and miRNA (Ruden, 2011; Grentzinger et al., 2012). We are still actively trying to determine the exact nature of the $[Kr^{If-1}]^*$ metastable epiallele and how it is transmitted through both the male and female germlines. As discussed later, we believe that *Drosophila*,

and more generally most or all Dipterans (flies) and Coleopterans (beetles), lost DNA methylation because the presence of 5mC would slow down the syncytial blastoderm mitotic cycles, which at ~8 min are the fastest in the animal kingdom (Ruden and Jackle, 1995).

There is no direct laboratory evidence that selection of metastable epialleles occurs in eusocial insects, such as honeybees. However, there are at least three indirect indications that metastable epialleles that utilize differential DNA methylation occur in eusocial insects. First, Herb et al. (2012, 2013) showed that reverting foragers back to nurses reestablished the nurse-pattern of DNA methylation. This was the first evidence of reversible epigenetic changes associated with behavior. Second, Hunt et al. (2010) found that worker-biased proteins exhibited slower evolutionary rates than queen biased proteins or non-biased proteins. This is consistent with the idea that metastable epialleles must be transmitted through the germline, and the queen and fertile workers are the only females that produce eggs. Finally, as described in the next section, the bimodal distributions of CG content and/or DNA methylation in most insect genes suggests a role for differential DNA methylation and the existence of metastable epialleles in most insects.

MECHANISMS OF EPIGENETIC MODIFICATION IN HONEYBEES

How might an EIS in honeybees and other organisms evolve? In order to understand this, it is necessary to describe the patterns of DNA methylation in mammals and honeybees (Figures 1A,B). In mammals, ~60% of genes have so-called CpG islands in the promoter regions and 5' regions, which are defined as regions of higher than average CG content. DNA methylation of CpG islands in mammals occurs primarily at CpG sites in somatic cells but often at CHH (where H = C, A or T) sites in stem cells (reviewed in Patil et al., 2014). The degree of CpG island methylation is inversely proportional to gene expression for most genes; i.e., highly expressed genes have little CpG island DNA methylation, whereas, low-expressed genes have large amounts of CpG island DNA methylation (Figure 1A). Two mechanisms that CpG island DNA methylation in mammals are thought to function to reduce gene expression are by inhibiting binding of some transcriptional activation factors, such as AP1, which binds to GC-rich consensus sequences, and by increasing the binding of transcriptional inhibitory factors, such as MeCP2, which recruits HDACs to inhibit transcription (reviewed in Jones, 2012).

The CpG island DNA methylation story is the most-well-known aspect of epigenetic regulation of transcription in mammals. However, several studies have shown that gene-body DNA methylation also occurs in a manner that is mostly proportional to gene expression in both mammals and insects (Konu and Li, 2002; reviewed in Jones, 2012). In other words, highly expressed genes have the most gene-body DNA methylation, and this DNA methylation is mostly restricted to exon sequences (Figure 1B), but this is partly because exons, since they encode proteins, are CG rich compared to intronic and intergenic regions, which do not encode proteins. DNA methylation in mammals also occurs at repeat sequences, such as ALUs, SINES, LINES, and retroviruses,

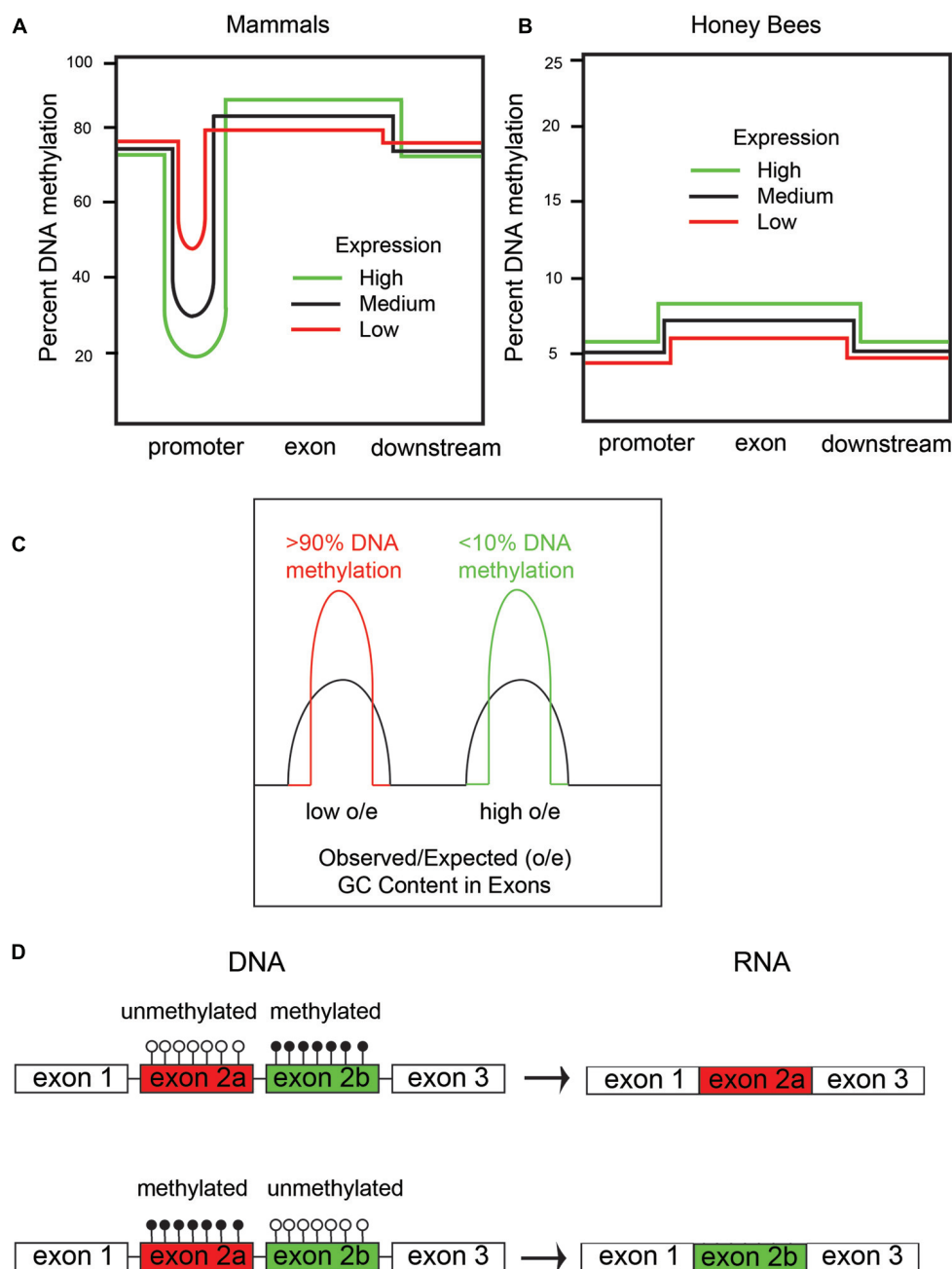


FIGURE 1 | DNA methylation in honeybees correlates with gene expression and alternative splicing. (A) There are two types of DNA methylation in mammals: (1) promoter DNA methylation, which inversely correlates with mRNA expression; and (2) exon DNA methylation, which positively correlates with mRNA expression. **(B)** Honeybees predominantly have DNA methylation in exons, which, like in mammals, positively correlates with gene expression. **(C)** There are two types of genes in honeybees: (1)

housekeeping genes with low observed/expected (o/e) CG content and high amounts of DNA methylation, and (2) caste-specific and developmental regulatory genes with a high o/e CG content and low amounts of DNA methylation. We have shown that the DNA methylation is at both CpG and CHH sites – CpG methylation primarily in exons and CHH methylation primarily in introns. **(D)** DNA methylation of cassette exons leads to their exclusion by alternative splicing in honeybees.

and this has been shown to prevent expression and, thereby, retro-transposition of the retroviruses to new genomic regions (Jones, 2012).

Interestingly, DNA methylation in honeybees occurs primarily in gene bodies, particularly in exons (Figure 1B). However,

in contrast to mammals, CpG islands are not apparent in the promoters of honeybee genes (i.e., there are very few genes with enriched CG-content in the promoter regions). Additionally, in the honeybee, little or no DNA methylation occurs in repeat or intergenic sequences (Lyko et al., 2010; Zemach et al., 2010; Chen

et al., 2011). Therefore, in honeybees, DNA methylation is not thought to epigenetically regulate expression of genes by controlling transcription factor binding to promoter regions, but rather is a consequence of gene expression. Gene body methylation in honeybees likely improves the fidelity of gene expression by allowing transcription to initiate only at the promoter and not at intergenic regions. Gene body DNA methylation in plants, for instance, has been shown to suppress intragenic transcriptional start sites and anti-sense transcription, presumably by preventing transcriptional activation proteins from binding to the gene body and inappropriately activating transcription from cryptic promoters (Zhang et al., 2006).

Originally, it was reported that most DNA methylation occurs primarily in CpG sequences in honeybees (Lyko et al., 2010; Zemach et al., 2010; Chen et al., 2011). However, we have shown, by analyzing our own data, and by reanalyzing the data from Lyko et al. (2010), that there is actually more CHH DNA methylation in honeybees than CpG DNA methylation (Cingolani et al., 2013). The other laboratories that analyzed DNA methylation in the honeybee used software that removed most of the CHH DNA methylation, presumably because this type of DNA methylation occurs in less complex regions of the genome (i.e., CG poor) and are therefore harder to align to the reference genome. Also, multiple CHH methylation events in a single next-generation DNA sequencing (NGS) read are often, sometimes improperly, interpreted as poorly converted by bisulfite and thrown out. However, we validated that most of the CHH methylation events are real by alternative methods, such as sequencing honeybee genomic DNA after immunoprecipitation with anti-5mC antibodies, and enzymatic digestion of DNA at 5-hydroxymethylcytosine (5hmC) sites (Cingolani et al., 2013). We did confirm, however, like the other groups, that CpG DNA methylation is primarily in exons. Interestingly, we also found that CHH DNA methylation is primarily in introns, partly because introns are larger and have a lower CG content (Cingolani et al., 2013).

We were also the first group to find significant amounts of 5hmC in bees (Cingolani et al., 2013). 5hmC is an oxidized form of cytosine, and is presumably produced by the honeybee ortholog to the ten-eleven-translocation (TET) protein, a dioxygenase that converts 5mC to 5hmC, and is involved in epigenetic reprogramming in mammals (reviewed in Chia et al., 2011). Wojciechowski et al. (2014) recently confirmed the presence of 5hmC in honeybees and characterized the enzymatic function of the TET enzyme. Because of the uncertainty of whether 5hmC is a stable epigenetic mark, as some investigators believe (including us), or a transient DNA modification in the de-methylation pathway, as most investigators believe, we will not discuss 5hmC further in this review but will await future clarification on this topic.

Genome sequencing the honeybee showed that it has an unusual genome structure that we believe facilitates the generation of metastable epialleles (Elango et al., 2009). In honeybees, there are two types of genes based on CG content in exons (Figure 1C). Highly expressed, so-called housekeeping genes, which are expressed in all cells, have a lower CG content than low-expressed genes. This bimodal distribution of CG content in genes, which are called isobars, was first observed by a bioinformatics analysis of the newly sequenced honeybee genome

(Jorgensen et al., 2007). The discovery of isobars in the honeybee genome was made prior to the mapping of the 5mC sites by whole-genome shotgun bisulfite sequencing by our laboratory and several other laboratories (Lyko et al., 2010; Zemach et al., 2010; Chen et al., 2011; Cingolani et al., 2013). Sodium bisulfite converts C to uracil (U) unless it is methylated (5mC), and whole genome shotgun bisulfite sequencing is used to map all of the 5mC sites in the genome (Xi and Li, 2009). Interestingly, all of the groups that performed whole-genome shotgun bisulfite sequencing to map the 5mC distribution in honeybees found, at first impression paradoxically, that the low-CG content genes have much more DNA methylation than the high-CG content genes (Lyko et al., 2010; Zemach et al., 2010; Chen et al., 2011; Figure 1C). We additionally found that CHH DNA methylation is also greater in the low-CG content genes than in the high-CG content genes. We proposed, since there is not a bimodal distribution of CHH sequences, that the same DNA methyltransferases (i.e., DNMT1 and DNMT3) methylate both CG and CHH sequences in a manner that is directly proportional to the level of gene expression (Cingolani et al., 2013). Figure 2 shows an example of a high-CG content, low-5mC gene (Ubx, Figure 2A) and a low-CG content, high-5mC gene (Actin, Figure 2B). Both Ubx and Actin will be discussed as examples throughout this perspective.

To reiterate, low-CG content genes have more 5mC than high-CG content genes. This is counter-intuitive because it indicates that the greater the CG content, the less the DNA methylation, despite there being more cytosines (specifically, CpG sites) to methylate. However, high-CG content genes having low DNA methylation makes biological sense for the same reason that CpG islands (by definition, with high CG content) have low DNA methylation. The biological sense is based on the fact that 5mC has a much higher (up to 10-fold) mutation rate to thymidine (T) than non-methylated cytosine (Rakyan et al., 2001). Therefore, the more highly expressed genes would have more 5mC (Figure 1B), and, consequently, more of the cytosines would become thymidine. Consequently, in highly expressed genes, the CG-content would be expected to become lower-and-lower as more-and-more CGs are converted to TGs. The reason for the higher mutation rate of 5mC-to-T compared with C-to-T is that 5mC spontaneously deaminates at the 6-position to form T, which is a natural DNA base. However, unmethylated C deaminates to U, which is normally not present in DNA, and there are enzymes [specifically uracil N-glycosylase (UNG)] to remove the U bases in DNA (Rakyan et al., 2001). The diagrams in Figures 1A,B are a simplification for clarity purposes because the most highly expressed genes, which we will call “ultra-high,” usually have less DNA methylation than the medium and highly expressed genes in the gene bodies in both insects and mammals. This might be because the ultra-high expressed genes may have lost so many of their CpGs that there are not enough remaining to allow them to enter the most highly methylated class – in other words, the amount of DNA methylation that can occur in genes is saturated and peaks before it reaches equilibrium. The genetic code for certain amino acids and intra-exon RNA-splicing enhancers requiring CGs in their consensus sequencings are likely two additional reasons for retaining a few CGs in housekeeping genes.

A. *A. mellifera* Ubx (993 bp; 97 CpGs)

atgaact**cg**t atttttgagca gactg**cg**ggt ggcttcta**cg** gaagccacca ccatcagaca
 ggagc**cg**cca gtcagcatca tgatccagcc a**cg**gcagc**cg** cctatcgaag tttccccc**c**
ggcct**cg**gta tgtcac**cg**ta **cg**cggtccacc caacaccatc atcacacctc ct**cg**t**cg**ttg
 ggcatacacc **cg**gg**cg**gtgg ga**cg**aaca**cg** aggc**cg**cccc aggatt**cg**cc gta**cg**atg**cg**
 ag**cg**t**cg**cgga **cg**gcttgcaa gctttatt**cg** a**cg**a**cg**cc**cg** aggcaactgg ccaca**cg**aca
 tcct**cg**tatt **cg**accacagc **gg**ccaaggac tgtaagcaac aggatcaagc at**cg**g**cg**cat
 cagaa**cg**gtt a**cg**c**cg**cagt gatggcagct gc**cg**c**cg**tca agga**cg**tgtg gcaat**cg**gct
 acct**cg**gggg **cg**aacagcca gagcaatt**cg** gtggtt**cg**cc cat**cg**g**cg**tg ccccc**cg**gaa
 ggg**cg**aggg ttggtagcta **cg**gtggtct**c** gtagg**cg**g**cg** atc**cg**gcat**c** g**ag**tcc**cg**gc
 aacaacagtt cct**cg**aggtc cctca**cg**t**cg** tcctggaaca cctgcagttt gaact**cg**tc**c**
gcgagccaac **cg**gttgcca**c** gcaactacat cagcaacca gcaaccatac gttctacccc
 tggatggcta tagcaggag**c** gaa**cg**gaatg **cg**cagg**cg**g**cg** gc**cg**ccagac ctata**cg**g**cg**c
 taccaga**cg**c t**cg**aactgga gaaggaattc caca**cg**aacc actacctcac tagg**cg**gag
cggat**cg**aga tggcacact**c** gctctgcctg a**cg**gaa**cg**gc agatcaagat ctggttccag
 aat**cg**g**cg**ga tgaagctgaa gaaggagata cagg**cg**atca aggagctgaa **cg**aacaggag
 aagcagg**cg**c agg**cg**cagaa ggcag**cg**gca gcag**cg**g**cg** **cg**gctg**cg**ca tcagcagcaa
 g**cg**gc**cg**gtg ggggac**cg**ga gggg**cg**caac tag

B. *A. mellifera* Actin (1131 bp; 40 CpGs)

atgtgtgatg a**cg**a**cg**tagc agcttttagta gtagataatg gatcaggcat gtgtaaggca
 ggatttgcag gagatgatgc tcca**cg**tgct gtattccctt ctattgtttg aagacctaga
 tatcaggga ttatggtttg tatgggacaa aaggatagtt atgtgggaga tgaagcacia
 agtaaaagag gtgtttttaac agttaaatat ccaattgaac atggtattgt aacaaattgg
 gatgatatgg aaaaaatatg gcatcataca ttctataatg agttaaggat tgctc**cg**gaa
 gaacacccc**cg** ttcttctgac tgaagctcct ttgaatccta aagctaatcg agagaagatg
 acacaaatta tgtttgaaac tttaattct ccagcaatgt atgt**cg**ccat acaagctgtt
 ttat**cg**cttt a**cg**cttcagg a**cg**cactaca ggcattgttc tagatag**cg**g agatggagtt
 tcacacacag taccaattta **cg**aggggatat gcattg**cg**c atgcaatttt a**cg**tatggat
 ttggctggt**c** gtgatttta**c** ggattatctt atgaaaattc taa**cg**gaaag aggatattct
 tttaccacta **cg**cgagag**cg** tgaaat**cg**tt **cg**agatatta aagaaaaatt gtgttatgtg
 gctttggact ttcagcaaga aatggcta**cg** gctgctggtt caagtgcttt ggaaaaaagc
 tatgaattac cagatggtca ggtgatcact atcggaaa**cg** aa**cg**tttc**cg** ttgcccagaa
 gcaatgtttc agccaagctt cct**cg**gcatg gaatcatg**cg** ggatacatga aacaacttat
 aattcaatta tgaaatg**cg**a tgttgatatt **cg**taaggatt tgtatg**cg**aa ttcagttcta
 tcggg**cg**gta ccaccatgta tc**cg**ggtatt g**cg**gacagga tgcaaaagga aataa**cg**gca
 ctggctccat ccacaatgaa aatcaagata at**cg**g**cg**cctc **cg**gaaaggaa atact**cg**gtc
 tggat**cg**gtg gtt**cg**atctt ggccagctta agta**cg**ttcc agcaaatgtg gatctcgaa
 caagagta**cg** acgagt**cg**gg accatccatt gtacacagaa agtgttttta a

FIGURE 2 | Examples of high CG and low CG genes. (A) *Apis mellifera* Ubx has 97 CG s in the coding region of a 993 base pair cDNA. The 5mC level of high CG -content genes, such as Ubx, is low. **(B)** *A. mellifera* actin has 40 CG s in the coding region of a 1131 base pair cDNA. The 5mC level of low CG -content genes, such as actin, is high.

As discussed in the next section, differential gene-body methylation might be a contributing factor to the emergence of eusociality. However, the bimodal distribution of CG content seems to be less of a contributor to eusociality than the bimodal distribution of DNA methylation. Bees, wasps, and ants all have bimodal distributions in DNA methylation in genes, but only bees and wasps have a bimodal distribution in CG content. In all three eusocial insects – bees, wasps and ants – the highly expressed genes are generally more methylated than the low-expressed genes (Sarda et al., 2012). Sarda et al. (2012) studied the evolution of gene-body DNA methylation in invertebrates and showed that silkworm (*Bombyx mori*), which has DNA methylation at appreciable levels in the genome, nevertheless does not have a bimodal peak of CG content in genes. This is similar to our finding of a unimodal peak of CHH sites but a bimodal peak of DNA methylation based on CG content, discussed earlier (Cingolani et al., 2013).

Interestingly, the silkworm has a bimodal peak in DNA methylation levels similar to the honeybee, in which highly expressed genes have higher levels of DNA methylation in the gene body. The unimodal peak in CG content but bimodal peak in DNA methylation levels seen in the silkworm genome also occurs in all ant species studied so far (Glastad et al., 2011; Bonasio et al., 2012; Bonasio, 2015). The bimodal peak in CG content in genes is not unique for the honeybee, however, because it is also seen in other invertebrates such as the sea anemone (*Nematostella vectensis*) and the sea squirt (*Ciona intestinalis*; Sarda et al., 2012). We conclude that while bimodal peaks in CG content and DNA methylation might facilitate the formation of metastable epialleles, they are not essential for the generation of metastable epialleles. In the next section, we explore the possibility that metastable epiallele hyper-mutability, a key component of the EDGE hypothesis, is an emergent property of bimodal levels of DNA methylation in eusocial insects.

METASTABLE EPIALLELE HYPER-MUTABILITY MIGHT BE AN EMERGENT PROPERTY OF BIMODAL LEVELS OF DNA METHYLATION

The movement from low-level rules to higher level sophistication is what we call emergence (Johnson, 2001).

The above quote is from Johnson's (2001) best-selling 2001 book, "Emergence: the connected lives of ants, brains, cities, and software." In the book, Johnson (2001) describes how a simple behavior, such as an increasing number of ants following a weak-and-winding scent trail laid down by one ant to a food supply, can lead to a complex behavior, such as all of the ants following a direct path to the food. Eusocial insects show many other examples of bottom-up behavior where workers follow simple rules that emerge into complex hive behaviors (Johnson, 2001). However, in contrast to human societies, there is little if any top-down behaviors in eusocial insects. For example, as mentioned above, the queen is best characterized as the "reproductive organ" in the hive and does little to influence the behaviors of the worker sub-castes (Johnson, 2001), who themselves follow simple rules that are programmed into their genomes and epigenomes. We believe that the differential methylation of genes based on the level of gene expression is just

such a simple rule that can lead to complex emergent phenomena, such as metastable epiallele hypermutability and, ultimately, eusociality.

We hypothesize that an emergent property of low-expressed genes having low levels of DNA methylation is that they become more susceptible to epigenetic control, for the simple fact that they have more unmethylated cytosines. Highly expressed genes with high levels of DNA methylation can also potentially become metastable epialleles, but this would require differential demethylation, such as by TET enzymes, in the germline cells after a stress response. In another review we presented a model for how oxidative stress can alter the function of the TET enzyme (Chia et al., 2011). However, what is the normal function(s) of gene body DNA methylation? In addition to preventing intragenic and antisense transcription within genes, mentioned above, one process that we and others have shown evidence to be regulated by gene body DNA methylation is alternative mRNA processing. For example, DNA methylation of cassette exons, at both CpG and CHH sites, correlates with their preferential exclusion in the mature mRNA (Lyko et al., 2010; Cingolani et al., 2013; **Figure 1D**). Furthermore, Li-Byarlay et al. (2013) have shown that RNA interference (RNAi) knockdown of DNMT3a, the *de novo* DNA methyltransferase, alters RNA splicing and causes intron retention in hundreds of genes in the honeybee fat bodies. How DNA methylation affects alternative mRNA splicing is not known in bees, but in mammals, DNA methylation inhibits the binding of the transcription factor CCCTC binding factor (CTCF), which affects alternative splicing (Shukla et al., 2011). We speculate that there might be some biophysical processes involved too, since methylated DNA has a higher melting temperature (T_m) than unmethylated DNA (Severin et al., 2011). Therefore, the increased T_m of methylated DNA might alter RNA polymerase translocation rates, cause pausing, and thereby affect the alternative mRNA splicing pattern.

One interesting observation is that most insects, such as honeybees, have relatively large amounts of DNA methylation (but much less than mammals), but *Drosophila* has very little DNA methylation (Lyko et al., 2000; Lyko, 2001). The reason for the scarcity in DNA methylation in *Drosophila* is that *Drosophila* appears to have lost Dnmt1, the maintenance DNA methyltransferase, which methylates hemizygous DNA after replication, and Dnmt3, the *de novo* DNA methyltransferase, which methylates unmethylated DNA. The existence of DNA methylation in *Drosophila* is controversial because the only cytosine methyltransferase orthologs in *Drosophila* is a homolog to DNA methyltransferase 2 (MT2), but this enzyme was shown to methylate transfer-RNA-Asp (tRNA_{Asp}) and presumably not DNA (Goll et al., 2006). However, the controversy appears to be resolved (at least to some in the field) by a recent paper that shows CHH methylation, albeit at very low levels, in *Drosophila* in a manner that is independent of MT2 (Capuano et al., 2014). The authors were able to detect low levels of 5mC in *Drosophila* embryos in a two-step protocol of first immunoprecipitation of DNA with anti-5mC antibodies, followed by bisulfite sequencing of the immunoprecipitated DNA fragments (Capuano et al., 2014). Our laboratory has similar evidence for low levels of 5mC in *Drosophila* and we speculate that it is generated non-enzymatically

by spontaneous methylation of cytosines by intrinsic alkylation of DNA.

We speculate that Dipterans (flies) and Coleopterans (beetles) lost DNA methyltransferases 1 and 3 because DNA methylation is redundant with histone modifications, such as H3K9me3 and H3K27me3, in repressing gene expression. Furthermore, we speculate that methylated DNA slows down DNA replication because of the higher melting temperature (T_m) of methylated DNA compared with unmethylated DNA (Severin et al., 2011), which we mentioned earlier in the discussion of mRNA splicing. The predicted slowing down of DNA replication by DNA methylation is important in *Drosophila* because the first 10 syncytial nuclear divisions in the blastoderm embryo are in “hyper-drive” and are less than 8–10 min in duration (a world record, to our knowledge). Therefore, any process that slows down these rapid divisions would presumably be selected against because the faster-developing siblings would breed sooner (Ruden and Jackle, 1995).

EPIGENETIC DIRECTED GENETIC ERRORS AND THE EVOLUTION OF CASTS IN HONEYBEES

Macroevolution requires selection of existing genetic variation to generate new species with greater fitness, but how does the sub-caste worker specialization increase when the effective population size of eusocial insects is so low (i.e., only one reproductive female per hive)? We mentioned group selection and kin-selection models at the beginning of this review, but they remain controversial in light of Dawkins's (1976) “selfish gene” hypothesis. Dawkins (1976) argued that “selfish genes” that benefit the immediate survival and propagation of the “vessel” (the organism) would have much greater (and more immediate) selective advantage than altruistic genes that benefited the group. We speculate again, as we did in several other reviews, that one possible mechanism to facilitate genetic variation in the evolution of species is what we call the EDGE hypothesis (Ruden, 2005; Ruden et al., 2005a, 2008; Ruden and Lu, 2008).

In the simplest version of the EDGE hypothesis, the first step is the intra-caste selection of metastable epialleles that increase the specialization of a worker. The metastable epialleles could initially be generated by a stressful (i.e., non-optimal) environment, which would lead to a functional inactivation of Hsp90 (Rutherford et al., 2007a), which is a chaperone for many chromatin remodeling proteins (Ruden and Lu, 2008), including the Trithorax (Trx) protein (Tariq et al., 2009; **Figures 3A,B**). Hsp90 has been called a “capacitor for morphological evolution” because many previously cryptic phenotypes are revealed when stress inactivates Hsp90 protein and this alters multiple signaling pathways (Rutherford and Lindquist, 1998; McLaren, 1999; Rutherford and Henikoff, 2003; Rutherford et al., 2007a,b). The Trx protein, since it is a client for Hsp90, is an environmentally sensitive component of the Trx Group (TrxG) complex of proteins that is involved in maintaining transcriptional memory (i.e., activation) of the Hox genes, such as the Ultrabithorax (Ubx) gene during early embryogenesis in insects (Orlando et al., 1998). One of the enzymatic functions of the TrxG complex is trimethylation of histone 3 at lysine 4 (H3K4me3), which is an activating mark for transcription (Jenuwein and Allis, 2001).

When Hsp90 is inactivated in a stressful environment, Hox genes such as Ubx would have lower expression, presumably because there would be less H4K4me3 histone marks at the promoters. Since stress inactivates the Trx protein (**Figure 3B**), then stress would be expected to cause an increase in the DNA methylation status of the Ubx gene. The reason for this is that, in the absence of the Trx protein, the gene would no longer be in an activated state but switch to a repressed state by the Polycomb Group (PcG) repressor proteins (Paro et al., 1998). It is not known whether this occurs in bees, but in mammals genes that are initially repressed by PcG proteins are often further repressed by intragenic DNA methylation during cellular differentiation (Deaton et al., 2011). We speculate that Ubx would have originally become a metastable epiallele in the proto-queen, who still has pollen baskets, because full pollen-baskets could immobilize her, and hence stress her, in the confines of the hive. In the EDGE hypothesis, the repair of base substitutions caused by methylated cytosines increases the mutation frequency of not only of the methylated cytosine, as mentioned above (Rakyan et al., 2001), but also neighboring bases because of error-prone DNA repair mechanisms (Ruden, 2005; Ruden et al., 2005a,b). This error-prone DNA repair could lead to an increase in the mutation frequency of genes with metastable epialleles, such as in the Ubx gene (**Figure 4D**). Through this “mutation-spreading” effect, the metastable epialleles could cause not only an increase in the mutation frequency of the exons, but also regulatory sequences in the adjacent promoters and introns. In other words, simply by becoming a metastable epiallele, the EDGE hypothesis predicts that the mutation frequency of a gene would increase. Fortuitously, genes with increased mutation frequencies are precisely those that need to be mutated to stabilize the metastable epialleles in a genetic manner.

How EDGE mutations generated in sterile workers are transmitted to the next generation in honeybees is a major issue that warrants discussion. One possible mechanism for transmitting the EDGE mutations to the next generation could be through honeybee workers who develop ovaries and become fertile after queen removal, as mentioned above (Feldmeyer et al., 2014). They could then directly transfer the mutations (as well as the metastable epialleles) to their offspring. Those mutations that are beneficial to the hive by stabilizing the metastable epialleles would have a selective advantage for the whole hive and would thereby be selected by group selection. An important consideration is that fertile-workers only have drone progeny (i.e., haploid males) and queens have both drone and worker progeny. This would necessitate that the metastable epialleles be transmitted through the male germline in the offspring of fertile workers. However, it is not clear whether worker-to-fertile-female conversions are frequent enough to explain the evolution of sterile-worker specializations.

Another possible mechanism for the transfer of EDGE mutations to the next generation is that the queen can transmit EDGE mutations to her offspring directly, without having to go through a worker-to-fertile-female conversion process. Metastable epialleles have to be in either the queen or the worker (by definition), but affect them in different manners. Therefore, the genes that become metastable epialleles would be predicted to have a

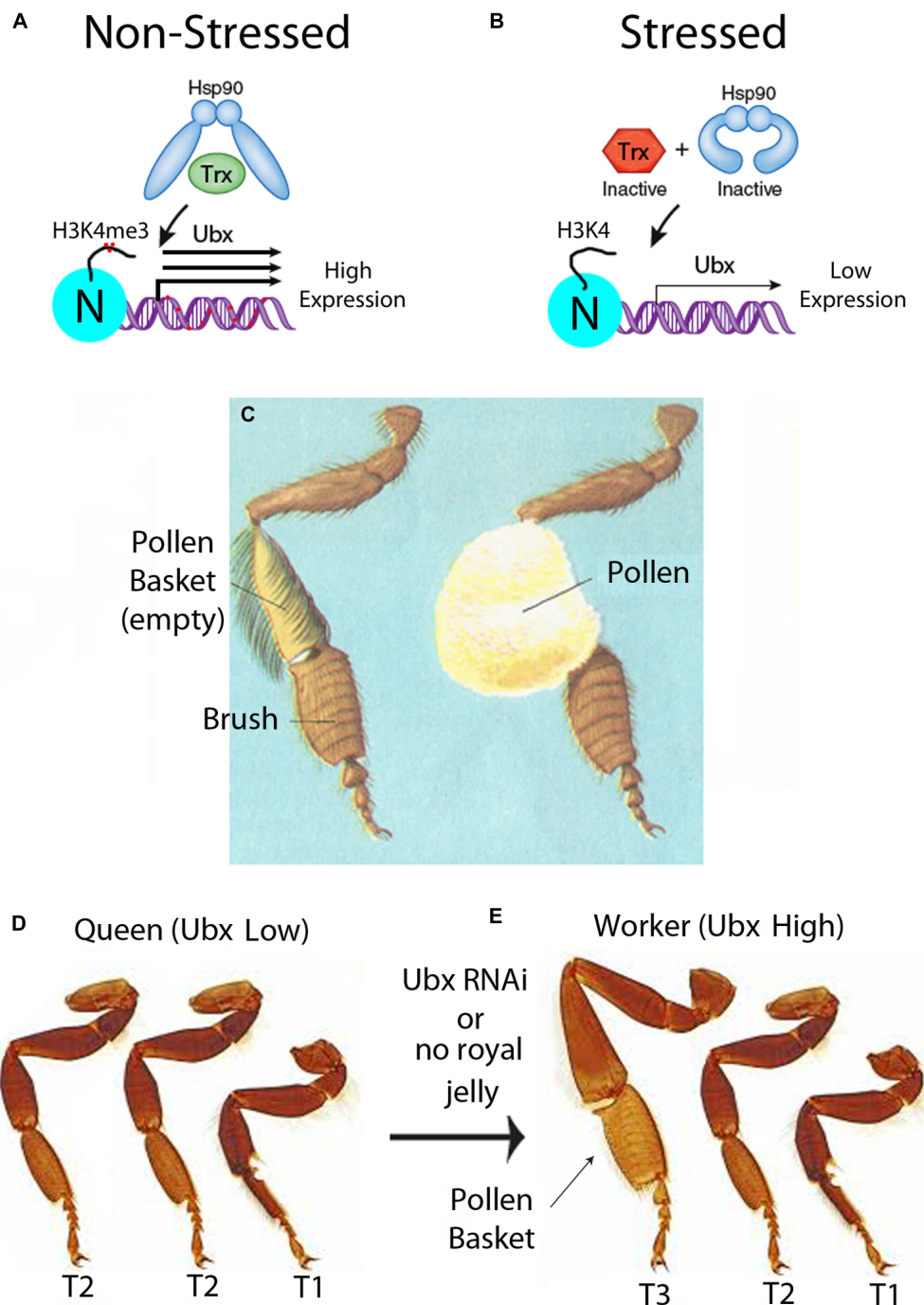
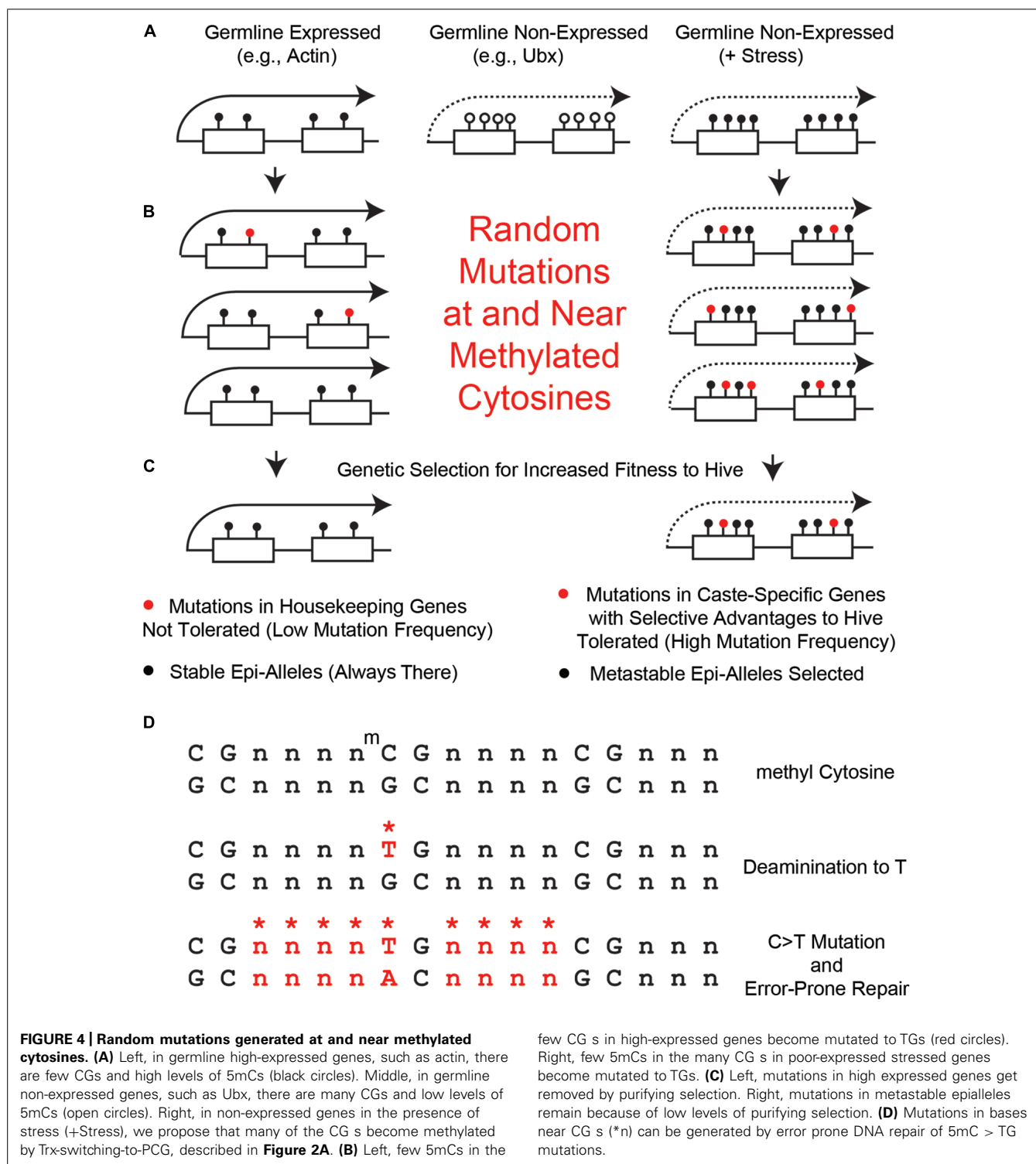


FIGURE 3 | Epigenetic control of development of the pollen basket in worker bees. (A) In unstressed conditions, Hsp90 is functional and activates Triethorax (Trx), through the chaperone activity of Hsp90. The Trx group (TrxG) proteins tri-methylate histone 3 lysine 4 (H3K4me3) on the promoter nucleosomes (N) (red dots), and increase expression of Hox genes such as Ubx. Transcriptional activation of Ubx in bees increases the DNA methylation of the gene body (red dots), as shown in **Figure 1B**. In the epigenetic directed error hypothesis (EDGE), the repair of base substitutions caused by methylated cytosines increases the mutation frequency of not only of the methylated cytosine but also neighboring bases. This could lead to an increase in the mutation frequency of genes with metastable epialleles, such as in the Ubx gene. This figure is modified from a previous review from our

laboratory and we retain the copyright (Ruden, 2011). (B) In stressed conditions, Hsp90 is inactive and cannot activate Trx, and transcription of Ubx is low. (C) Diagram of an empty and full pollen basket in forager bees. This diagram is used with permission from the Encyclopedia of Science, Copyright © The Worlds of David Darling (<http://www.daviddarling.info/>). (D) In queens, the pollen basket does not form because Ubx is low in T3. This causes an anterior transformation of the third thorax (T3) leg to look like the T2 leg. (E) In workers, the pollen basket forms because Ubx expression is high in T3. This figure represents a simplified representation of the homotic transformation that occurs when Ubx levels are reduced and are not meant to be accurate illustrations. This photograph is used with permission from Spike Walker, Wellcome Images, London (<http://wellcomelibrary.org/>).



higher mutation rate by the EDGE process. Some of these mutations would not affect the queen (and therefore reproduction), but might stabilize the metastable epialleles that affect the workers. Support for the EDGE hypothesis is the fact that queen-specific genes mutate faster than worker-specific genes (Hunt et al., 2010; Helantera and Uller, 2014). This makes sense since queens breed

much more frequently than fertile workers, which, as mentioned above, only occur when the queen is removed from the colony (Feldmeyer et al., 2014). Queens would therefore have a greater opportunity to transmit both metastable epialleles, and mutations in the metastable epialleles, to the offspring, than the fertile workers.

EPIGENETIC DIRECTED GENETIC ERRORS AND THE EVOLUTION OF POLLEN BASKETS IN HONEYBEE WORKERS

A recent paper that we believe supports the EDGE hypothesis for intra-caste evolution in honeybees discusses the dimorphism in pollen basket formation in genetically similar queens and workers. This fascinating paper shows that the Hox gene *Ubx*, mentioned throughout this perspective, promotes pollen basket formation on the tibia of the hind legs in bees in the third thoracic segment (T3) (Medved et al., 2014). The pollen basket is a hollow indentation on the large and mostly bristle-free tibia segment that the forager bees use to store and transport impressive amounts of pollen (Figure 3C). In the queen, who does not collect pollen, the tibia is covered with hairs that would otherwise inhibit pollen collection. The investigators showed that reduction of *Ubx* levels in the workers by injecting inhibitory RNA (RNAi) into the worker embryos caused the hind legs to resemble that of the queens and become reduced in bristles (Medved et al., 2014). In *Drosophila*, mutations in *Ubx*, combined with other mutations in the bithorax complex (BXC), produced the famous four-winged fly that won Edward Lewis the 1995 Nobel Prize in Physiology and Medicine (Crow and Bender, 2004). Normally, *Drosophila* one pair of wings on the second thoracic segment (T2) and one pair of halteres (balancer organs that counteract the wing movement) on the third thoracic segment (T3). Lewis (1978) explained the *Ubx* phenotype as causing an anterior homeotic transformation of T3 to T2, hence, the famous four-winged fly. *Ubx* has a conserved 60 amino acid homeobox (Hox) domain, which is nearly identical from *Drosophila* to humans, and a highly variable transcriptional regulatory domain. Hox genes, such as *Ubx*, not only regulate segmentation during embryogenesis, but they also affect subtle changes in limb, brain, and other organ development.

Medved et al. (2014) found that mutations in the Hox gene, *Ubx*, causes complex fate decisions in each segment of the honeybee T3 legs. A simplification of the results of the *Ubx*-RNAi experiments in honeybees is that the third leg has a partial homeotic transformation to the second leg by a similar T3-to-T2 homeotic transformation as seen in *Ubx*-mutant flies (Figures 3D,E). As mentioned earlier, the way the nurse honeybee controls pollen basket development in workers is by withholding royal jelly. In honeybees, the targets of *Ubx* are not known, but the authors speculated on what might be occurring in honeybees, based on what is known in the much better characterized *D. melanogaster* genetic system. In honeybee queens, when they are fed royal jelly as larvae, the HDACi activity in the royal jelly could possibly help in the activation of expression of the likely *Ubx*-target genes, such as grunge (*gug*) and Ataxin-2 (*Atx2*), which play a role in the formation of bristles in *Drosophila* (Erkner et al., 2002; Al-Ramahi et al., 2007). Consequently, the authors speculate, this might be one reason why the T3 tibia segments in queens have bristles in the area of the pollen basket, while workers do not (Medved et al., 2014).

EPIGENETIC DIRECTED GENETIC ERRORS IN NON-CG DINUCLEOTIDES IN METASTABLE EPIALLELES

In the EDGE hypothesis, we propose that methylated cytosines are mutagenic not only in the 5mC sites but also in the surrounding

bases. The reason we propose this broader-range of mutagenicity is because error-prone DNA repair mechanisms can increase the mutation frequency of surrounding bases while repairing 5mC > T base substitution mutations. Metastable epialleles, which have variable levels of 5mC, can occur in both somatic cells and germline cells, but they are generally referred to as simply “differentiated cells” when they occur in somatic cells. When metastable epialleles occur in somatic cells, they cannot be transmitted to the progeny. However, when a metastable epiallele occurs in a germline cell, then it can be transmitted to the progeny, as we and others have demonstrated in *Drosophila* (Sollars et al., 2003; Tariq et al., 2009).

It is not yet known whether there is a bimodal distribution of 5mC in bee germline cells, but for the sake of argument, let's assume for this perspective that it is similar to what occurs in somatic cells – i.e., housekeeping genes have low CG-content and high levels of 5mC and low-expressed genes have high CG-content and low levels of 5mC (Figure 4A). Therefore, housekeeping genes, such as Actin (Figure 4A, left) would never be metastable epialleles because their few CG sites are always heavily methylated – i.e., there can be no differential 5mC if it is always high. In contrast, low-expressed genes, such as *Ubx*, which is presumably not expressed at all in germline cells, would have high CG content but very little 5mC (Figure 4A, middle). We hypothesize that maternal stress can increase the DNA methylation in low-expressed genes, such as *Ubx*, and turn them into metastable epialleles (Figure 4A, right). This has not yet been demonstrated in any organism, but it should be possible to test this hypotheses in the laboratory once single-cell epigenomics techniques are further optimized (Teles et al., 2014).

In housekeeping genes, such as Actin, there would still be expected to be an increase in mutations near and surrounding the 5mC sites. However, since there is a great deal of purifying selection in housekeeping genes, that would make any deleterious mutations in such important structural genes selected against (Figures 4B,C, left). Also, the 5mC rate in housekeeping genes is so high that there has probably been a maximum change in CG-to-TG sequences so that no further such mutations can occur without having deleterious structural or regulatory changes to the gene. In contrast, in low-expressed genes, such as *Ubx*, there would be mutations in CG-sites that do not undergo as much purifying selection (Figures 4B,C, right). As mentioned above, while the *Ubx* Hox domain is a 60 amino acid sequence that is almost absolutely conserved from *Drosophila* to humans (Scott, 1986), the remaining amino acids, such as in the transcriptional regulatory domains, are amongst the most variable sequences in proteins (Ruden et al., 1991; Ruden, 1992).

In the *Ubx*-mutagenesis hypothesis for basket formation in honeybees, several questions arose during review of this manuscript. First, “How did *Ubx* changed its biological role without affecting fitness?” Second, “Is it possible that *Ubx* regulates both body plan and caste differentiation in honeybee but not in solitary insects?” Third, “Could some other gene(s) compensate for the supposed “functional loss” of *Ubx* in honeybee?” To answer these questions, we do not believe that the mutations in *Ubx* would necessarily affect fitness by causing a “functional loss.” Rather, we believe that the mutations in

Ubx were most likely regulatory mutations in the promoter and introns and they represent a functional gain rather than a functional loss. Developmental genes have large and complex regulatory regions, such as individual enhancers for each of the eight stripes in segmentation genes such as fushi tarazu (Ohtsuki et al., 1998). The Hox genes in *Drosophila*, such as Ubx and Antennapedia (Antp), have enhancer regions 10s or even 100s of kilobases from the promoter regions (Calhoun et al., 2002; Calhoun and Levine, 2003). The Ubx gene in *Drosophila* has a complex array of alternative spliced products and an unusual mechanism for splicing the 74 kb intron that involves multiple steps of re-splicing the intron (Hatton et al., 1998). This re-splicing mechanism avoids competition between distant splice sites and allows removal of the 74 kb intron as a series of smaller RNA fragments (Hatton et al., 1998). The diverse array of transcriptional and RNA splicing regulatory sequences should allow Ubx to evolve multiple additional roles in caste formation without the need for other genes to compensate for it proposed “functional loss.”

CONCLUSION

We propose an intra-caste model of evolution that is based on selection of metastable epialleles in worker bees that runs parallel to the macro-evolution and group selection of DNA mutations. Like the mythical world of Jaynus, the evolution of the most-fit sub-caste members occurs through the selection of metastable epialleles by group selection. However, our EDGE hypothesis expands upon the limited world of Jaynus, in which all of the organisms have exactly the same sequence, by proposing a mechanism to direct mutations to the metastable epialleles that were selected. These directed mutations can, in turn, stabilize and increase the penetrance of the metastable epialleles in future generations of superorganism colonies. Waddington, who is often considered the father of epigenetics, proposed a mechanism similar to the EDGE hypothesis in Waddington (1942) for the inheritance of acquired characteristics that were induced by stress. In follow-up experiments, in response to Waddington, we provide a possible epigenetic mechanism for how stress can reveal previously cryptic phenotypic information by the inactivation of Hsp90 (Ruden et al., 2003; Sollars et al., 2003). Finally, our EDGE hypothesis presented here provides a possible mechanism for the stabilization of metastable epialleles, thereby allowing the evolution of castes and sub-castes in eusocial insects.

ACKNOWLEDGMENTS

This research was supported by R01 ES012933 and R21 ES021893 and the WSU-NIEHS Center (P30 ES020957). We thank Greg Hunt for excellent editorial assistance and much needed expertise on eusocial insects. DMR worked as an undergraduate in Dr. Edward B. Lewis' laboratory in 1980–1981 dedicates this review to his memory.

REFERENCES

Al-Ramahi, I., Perez, A. M., Lim, J., Zhang, M. H., Sorensen, R., de Haro, M., et al. (2007). DAtaxin-2 mediates expanded Ataxin-1-induced neurodegeneration in a *Drosophila* model of SCA1. *PLoS Genet.* 3:e234. doi: 10.1371/journal.pgen.0030234

Amdam, G. V., Aase, A. L. T. O., Seehuus, S. C., Fondrk, M. K., Norberg, K., and Hartfelder, K. (2005). Social reversal of immunosenescence in honey bee workers. *Exp. Gerontol.* 40, 939–947. doi: 10.1016/j.exger.2005.08.004

Boch, R., Shearer, D. A., and Stone, B. C. (1962). Identification of isoamyl acetate as an active component in the sting pheromone of the honey bee. *Nature* 195, 1018–1020. doi: 10.1038/1951018b0

Bonasio, R. (2015). The expanding epigenetic landscape of non-model organisms. *J. Exp. Biol.* 218, 114–122. doi: 10.1242/jeb.110809

Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., et al. (2012). Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* 22, 1755–1764. doi: 10.1016/j.cub.2012.07.042

Branco, A. T., and Lemos, B. (2014). High intake of dietary sugar enhances bisphenol A (BPA) disruption and reveals ribosome-mediated pathways of toxicity. *Genetics* 197, 147–157. doi: 10.1534/genetics.114.163170

Calhoun, V. C., and Levine, M. (2003). Long-range enhancer-promoter interactions in the Scr-Antp interval of the *Drosophila* Antennapedia complex. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9878–9883. doi: 10.1073/pnas.1233791100

Calhoun, V. C., Stathopoulos, A., and Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* Antennapedia complex. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9243–9247. doi: 10.1073/pnas.142291299

Capuano, F., Mulleder, M., Kok, R., Blom, H. J., and Ralser, M. (2014). Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. *Anal. Chem.* 86, 3697–3702. doi: 10.1021/ac500447w

Carey, J. R., Harshman, L. G., Liedo, P., Muller, H. G., Wang, J. L., and Zhang, Z. (2008). Longevity-fertility trade-offs in the tephritid fruit fly, *Anastrepha ludens*, across dietary-restriction gradients. *Aging Cell* 7, 470–477. doi: 10.1111/j.1474-9726.2008.00389.x

Carrera, P., Abrell, S., Kerber, B., Walldorf, U., Preiss, A., Hoch, M., et al. (1998). A modifier screen in the eye reveals control genes for Kruppel activity in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U.S.A.* 95, 10779–10784. doi: 10.1073/pnas.95.18.10779

Chen, P. Y., Feng, S., Joo, J. W., Jacobsen, S. E., and Pellegrini, M. (2011). A comparative analysis of DNA methylation across human embryonic stem cell lines. *Genome Biol.* 12, R62. doi: 10.1186/gb-2011-12-7-r62

Chia, N., Wang, L., Lu, X., Senut, M. C., Brenner, C., and Ruden, D. M. (2011). Hypothesis: environmental regulation of 5-hydroxymethylcytosine by oxidative stress. *Epigenetics* 6, 853–856. doi: 10.4161/epi.6.7.16461

Cingolani, P., Cao, X., Khetani, R., Chen, C. C., Coon, M., Sammak, A., et al. (2013). Intronic non-CG DNA hydroxymethylation and alternative mRNA splicing in honey bees. *BMC Genomics* 14:666. doi: 10.1186/1471-2164-14-666

Coulondre, C., Miller, J. H., Farabaugh, P. J., and Gilbert, W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* 274, 775–780. doi: 10.1038/274775a0

Crow, J. F., and Bender, W. (2004). Edward B. Lewis 1918–2004. *Genetics* 168, 1773–1783.

Darwin, C. (1859). *The Origin of the Species* (Random House).

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. New York, NY: D. Appleton and company. doi: 10.1037/12293-000

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

Deaton, A. M., Webb, S., Kerr, A. R., Illingworth, R. S., Guy, J., Andrews, R., et al. (2011). Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res.* 21, 1074–1086. doi: 10.1101/gr.118703.110

Dolinoy, D. C., Das, R., Weidman, J. R., and Jirtle, R. L. (2007). Metastable epialleles, imprinting, and the fetal origins of adult diseases. *Pediatr. Res.* 61, 30R–37R. doi: 10.1203/pdr.0b013e31804575f7

Duncan, B. K., and Miller, J. H. (1980). Mutagenic deamination of cytosine residues in DNA. *Nature* 287, 560–561. doi: 10.1038/287560a0

Elango, N., Hunt, B. G., Goodisman, M. A., and Yi, S. V. (2009). DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11206–11211. doi: 10.1073/pnas.0900301106

Erkner, A., Roue, A., Charroux, B., Delaage, M., Holway, N., Core, N., et al. (2002). Grunge, related to human Atrophin-like proteins, has multiple functions in *Drosophila* development. *Development* 129, 1119–1129.

- Fairbanks, D. J., and Rytting, B. (2001). Mendelian controversies: a botanical and historical review. *Am. J. Bot.* 88, 737–752. doi: 10.2307/2657027
- Feldmeyer, B., Elsner, D., and Foitzik, S. (2014). Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Mol. Ecol.* 23, 151–161. doi: 10.1111/mec.12490
- Free, J. B. (1987). The biology of the honey-bee - winston, Ml. *Science* 238, 1591–1592. doi: 10.1126/science.238.4833.1591
- Gangaraju, V. K., Yin, H., Weiner, M. M., Wang, J., Huang, X. A., and Lin, H. (2010). *Drosophila* piwi functions in Hsp90-mediated suppression of phenotypic variation. *Nat. Genet.* 43, 153–158. doi: 10.1038/ng.743
- Glastad, K. M., Hunt, B. G., Yi, S. V., and Goodisman, M. A. (2011). DNA methylation in insects: on the brink of the epigenomic era. *Insect. Mol. Biol.* 20, 553–565. doi: 10.1111/j.1365-2583.2011.01092.x
- Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., et al. (2006). Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science* 311, 395–398. doi: 10.1126/science.1120976
- Grentzinger, T., Armenise, C., Brun, C., Mugat, B., Serrano, V., Pelisson, A., et al. (2012). piRNA-mediated transgenerational inheritance of an acquired trait. *Genome Res.* 22, 1877–1888. doi: 10.1101/gr.136614.111
- Hamilton, W. D. (1964). The evolution of social behavior. *J. Theor. Biol.* 7, 1–16. doi: 10.1016/0022-5193(64)90038-4
- Hatton, A. R., Subramaniam, V., and Lopez, A. J. (1998). Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. *Mol. Cell.* 2, 787–796. doi: 10.1016/S1097-2765(00)80293-2
- Helanterä, H., and Uller, T. (2014). Neutral and adaptive explanations for an association between caste-biased gene expression and rate of sequence evolution. *Front. Genet.* 5:297. doi: 10.3389/fgene.2014.00297
- Herb, B. R. (2014). Epigenetics as an answer to Darwins “special difficulty.” *Front. Genet.* 5:321. doi: 10.3389/fgene.2014.00321
- Herb, B. R., Wolschin, F., Hansen, K. D., Aryee, M. J., Langmead, B., Irizarry, R., et al. (2012). Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nat. Neurosci.* 15, 1371–1373. doi: 10.1038/nn.3218
- Herb, B. R., Wolschin, F., Hansen, K. D., Aryee, M. J., Langmead, B., Irizarry, R., et al. (2013). Reversible switching between epigenetic states in honeybee behavioral subcastes. *Integr. Comp. Biol.* 53, E91–E91.
- Hunt, B. G., Wyder, S., Elango, N., Werren, J. H., Zdobnov, E. M., Yi, S. V., et al. (2010). Sociality is linked to rates of protein evolution in a highly social insect. *Mol. Biol. Evol.* 27, 497–500. doi: 10.1093/molbev/msp225
- Jablonska, E., and Lamb, M. J. (2005). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. (Cambridge, MA: MIT Press).
- Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. *Science* 293, 1074–1080. doi: 10.1126/science.1063127
- Johnson, S. (2001). *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. (New York: Scribner).
- Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13, 484–492. doi: 10.1038/nrg3230
- Jorgensen, F. G., Schierup, M. H., and Clark, A. G. (2007). Heterogeneity in regional GC content and differential usage of codons and amino acids in GC-poor and GC-rich regions of the genome of *Apis mellifera*. *Mol. Biol. Evol.* 24, 611–619. doi: 10.1093/molbev/msl190
- Kamakura, M. (2011). Royalactin induces queen differentiation in honeybees. *Nature* 473, 478–483. doi: 10.1038/nature10093
- Kennedy, P., Uller, T., and Helanterä, H. (2014). Are ant supercolonies crucibles of a new major transition in evolution? *J. Evol. Biol.* 27, 1784–1796. doi: 10.1111/jeb.12434
- Konu, O. O., and Li, M. D. (2002). Correlations between mRNA expression levels and GC contents of coding and untranslated regions of genes in rodents. *J. Mol. Evol.* 54, 35–41. doi: 10.1007/s00239-001-0015-z
- Leoncini, I., Crauser, D., Robinson, G. E., and Le Conte, Y. (2004). Worker-worker inhibition of honey bee behavioural development independent of queen and brood. *Insect. Soc.* 51, 392–394. doi: 10.1007/s00040-004-0757-x
- Le Thomas, A., Marinov, G. K., and Aravin, A. A. (2014). A transgenerational process defines piRNA biogenesis in *Drosophila* virilis. *Cell Rep.* 8, 1617–1623. doi: 10.1016/j.celrep.2014.08.013
- Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565–570. doi: 10.1038/276565a0
- Li-Byarlay, H., Li, Y., Stroud, H., Feng, S. H., Newman, T. C., Kaneda, M., et al. (2013). RNA interference knockdown of DNA methyltransferase 3 affects gene alternative splicing in the honey bee. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12750–12755. doi: 10.1073/pnas.1310735110
- Lyko, F. (2001). DNA methylation learns to fly. *Trends Genet.* 17, 169–172. doi: 10.1016/S0168-9525(01)02234-X
- Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* 8:e1000506. doi: 10.1371/journal.pbio.1000506
- Lyko, F., Ramsahoye, B. H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. *Nature* 408, 538–540. doi: 10.1038/35046205
- McLaren, A. (1999). Too late for the midwife toad: stress, variability and Hsp90. *Trends Genet.* 15, 169–171. doi: 10.1016/S0168-9525(99)01732-1
- Medved, V., Huang, Z. Y., and Popadic, A. (2014). Ubx promotes corbicular development in *Apis mellifera*. *Biol. Lett.* 10, 20131021. doi: 10.1098/rsbl.2013.1021
- Meneses Hoyos, J. (1960). [The discovery and rediscovery of the laws of heredity. (The work of Johann Gregor Mendel)]. *Rev. Asoc. Médica Mex.* 40, 401–410.
- Nowak, M. A., Tarnita, C. E., and Wilson, E. O. (2010). The evolution of eusociality. *Nature* 466, 1057–1062. doi: 10.1038/nature09205
- Nystrand, M., and Dowling, D. K. (2014). Transgenerational interactions involving parental age and immune status affect female reproductive success in *Drosophila melanogaster*. *Proc. Biol. Sci.* 281, 20141242. doi: 10.1098/rspb.2014.1242
- Ohtsuki, S., Levine, M., and Cai, H. N. (1998). Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes Dev.* 12, 547–556. doi: 10.1101/gad.12.4.547
- Orlando, V., Jane, E. P., Chinwalla, V., Harte, P. J., and Paro, R. (1998). Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J.* 17, 5141–5150. doi: 10.1093/emboj/17.17.5141
- Paro, R., Strutt, H., and Cavalli, G. (1998). Heritable chromatin states induced by the Polycomb and trithorax group genes. *Novartis Found. Symp.* 214, 51–61.
- Patil, V., Ward, R. L., and Hesson, L. B. (2014). The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics* 9, 823–828. doi: 10.4161/epi.28741
- Rakyan, V. K., Blewitt, M. E., Druker, R., Preis, J. I., and Whitelaw, E. (2002). Metastable epialleles in mammals. *Trends Genet.* 18, 348–351. doi: 10.1016/S0168-9525(02)02709-9
- Rakyan, V. K., Preis, J., Morgan, H. D., and Whitelaw, E. (2001). The marks, mechanisms and memory of epigenetic states in mammals. *Biochem. J.* 356, 1–10. doi: 10.1042/0264-6021:3560001
- Ruden, D. M. (1992). Activating regions of yeast transcription factors must have both acidic and hydrophobic amino-acids. *Chromosoma* 101, 342–348. doi: 10.1007/BF00346013
- Ruden, D. M. (2005). Epigenetic regulatory mechanisms in cancer, development, and evolution. *Curr. Genomics* 6, 127–127. doi: 10.2174/1389202053971992
- Ruden, D. M. (2011). The (new) new synthesis and epigenetic capacitors of morphological evolution. *Nat. Genet.* 43, 88–89. doi: 10.1038/ng0211-88
- Ruden, D. M., Chen, L., Possidente, D., Possidente, B., Rasouli, P., Wang, L., et al. (2009). Genetical toxicogenomics in *Drosophila* identifies master-modulatory loci that are regulated by developmental exposure to lead. *Neurotoxicology* 30, 898–914. doi: 10.1016/j.neuro.2009.08.011
- Ruden, D. M., Garfinkel, M. D., Sollars, V. E., and Lu, X. (2003). Waddington's widget: Hsp90 and the inheritance of acquired characters. *Semin. Cell Dev. Biol.* 14, 301–310. doi: 10.1016/j.semcdb.2003.09.024
- Ruden, D. M., Garfinkel, M. D., Xiao, L., and Lu, X. Y. (2005a). Epigenetic regulation of trinucleotide repeat expansions and contractions and the “biased embryos” hypothesis for rapid morphological evolution. *Curr. Genomics* 6, 145–155. doi: 10.2174/1389202053971947
- Ruden, D. M., Xiao, L., Garfinkel, M. D., and Lu, X. (2005b). Hsp90 and environmental impacts on epigenetic states: a model for the trans-generational effects of diethylstilbestrol on uterine development and cancer. *Hum. Mol. Genet.* 14, R149–R155. doi: 10.1093/hmg/ddi103
- Ruden, D. M., and Jackle, H. (1995). Mitotic delay dependent survival identifies components of cell cycle control in the *Drosophila* blastoderm. *Development* 121, 63–73.
- Ruden, D. M., Jamison, D. C., Zeeberg, B. R., Garfinkel, M. D., Weinstein, J. N., Rasouli, P., et al. (2008). The EDGE hypothesis: epigenetically directed

- genetic errors in repeat-containing proteins (RCPs) involved in evolution, neuroendocrine signaling, and cancer. *Front. Neuroendocrinol.* 29:428–444. doi: 10.1016/j.yfrne.2007.12.004
- Ruden, D. M., and Lu, X. (2008). Hsp90 affecting chromatin remodeling might explain transgenerational epigenetic inheritance in *Drosophila*. *Curr. Genomics* 9, 500–508. doi: 10.2174/138920208786241207
- Ruden, D. M., Ma, J., Li, Y., Wood, K., and Ptashne, M. (1991). Generating yeast transcriptional activators containing no yeast protein sequences. *Nature* 350, 250–252. doi: 10.1038/350250a0
- Rutherford, S. L., and Henikoff, S. (2003). Quantitative epigenetics. *Nat. Genet.* 33, 6–8. doi: 10.1038/ng103-6
- Rutherford, S., Hirate, Y., and Swalla, B. J. (2007a). The Hsp90 capacitor, developmental remodeling, and evolution: the robustness of gene networks and the curious evolvability of metamorphosis. *Crit. Rev. Biochem. Mol. Biol.* 42, 355–372. doi: 10.1080/10409230701597782
- Rutherford, S., Knapp, J. R., and Csermely, P. (2007b). Hsp90 and developmental networks. *Adv. Exp. Med. Biol.* 594, 190–197. doi: 10.1007/978-0-387-39975-1_16
- Rutherford, S. L., and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature* 396, 336–342. doi: 10.1038/24550
- Sarda, S., Zeng, J., Hunt, B. G., and Yi, S. V. (2012). The evolution of invertebrate gene body methylation. *Mol. Biol. Evol.* 29, 1907–1916. doi: 10.1093/molbev/mss062
- Scott, M. P. (1986). More on the homeobox. *Bioessays* 5, 88–89. doi: 10.1002/bies.950050210
- Severin, P. M., Zou, X., Gaub, H. E., and Schulten, K. (2011). Cytosine methylation alters DNA mechanical properties. *Nucleic Acids Res.* 39, 8740–8751. doi: 10.1093/nar/gkr578
- Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., et al. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74–79. doi: 10.1038/nature10442
- Slessor, K. N., Winston, M. L., and Le Conte, Y. (2005). Pheromone communication in the honeybee (*Apis mellifera* L.). *J. Chem. Ecol.* 31, 2731–2745. doi: 10.1007/s10886-005-7623-9
- Sollars, V., Lu, X., Xiao, L., Wang, X., Garfinkel, M. D., and Ruden, D. M. (2003). Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nat. Genet.* 33, 70–74. doi: 10.1038/ng1067
- Somer, R. A., and Thummel, C. S. (2014). Epigenetic inheritance of metabolic state. *Curr. Opin. Genet. Dev.* 27, 43–47. doi: 10.1016/j.gde.2014.03.008
- Spannhoff, A., Kim, Y. K., Raynal, N. J. M., Gharibyan, V., Su, M. B., Zhou, Y. Y., et al. (2011). Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. *EMBO Rep.* 12, 238–243. doi: 10.1038/embor.2011.9
- Stern, S., Snir, O., Mizrahi, E., Galili, M., Zaltsman, I., and Soen, Y. (2014). Reduction in maternal Polycomb levels contributes to transgenerational inheritance of a response to toxic stress in flies. *J. Physiol.* 592, 2343–2355. doi: 10.1113/jphysiol.2014.271445
- Tariq, M., Nussbaumer, U., Chen, Y., Beisel, C., and Paro, R. (2009). Trithorax requires Hsp90 for maintenance of active chromatin at sites of gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1157–1162. doi: 10.1073/pnas.0809669106
- Teles, J., Enver, T., and Pina, C. (2014). Single-cell PCR profiling of gene expression in hematopoiesis. *Methods Mol. Biol.* 1185, 21–42. doi: 10.1007/978-1-4939-1133-2_3
- Valtonen, T. M., Kangassalo, K., Polkki, M., and Rantala, M. J. (2012). Transgenerational effects of parental larval diet on offspring development time, adult body size and pathogen resistance in *Drosophila melanogaster*. *PLoS ONE* 7:e31611. doi: 10.1371/journal.pone.0031611
- Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565. doi: 10.1038/150563a0
- Wei, Y., Schatten, H., and Sun, Q.-Y. (2015). Environmental epigenetic inheritance through gametes and implications for human reproduction. *Hum. Reprod. Update* 21, 194–208. doi: 10.1093/humupd/dmu061
- Wilson, D. S., and Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *Q. Rev. Biol.* 82, 327–348. doi: 10.1086/522809
- Wilson, E. O., and Nowak, M. A. (2014). Natural selection drives the evolution of ant life cycles. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12585–12590. doi: 10.1073/pnas.1405550111
- Wojciechowski, M., Rafalski, D., Kucharski, R., Misztal, K., Maleszka, J., Bochtler, M., et al. (2014). Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. *Open Biol.* 4:140110. doi: 10.1098/rsob.140110
- Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinform.* 10:232. doi: 10.1186/1471-2105-10-232
- Zemach, A., McDaniel, I. E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919. doi: 10.1126/science.1186366
- Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W., Chen, H., et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 126, 1189–1201. doi: 10.1016/j.cell.2006.08.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 November 2014; accepted: 08 February 2015; published online: 24 February 2015.

Citation: Ruden DM, Cingolani PE, Sen A, Qu W, Wang L, Senut M-C, Garfinkel MD, Sollars VE and Lu X (2015) Epigenetics as an answer to Darwin's "special difficulty," Part 2: natural selection of metastable epialleles in honeybee castes. *Front. Genet.* 6:60. doi: 10.3389/fgene.2015.00060

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2015 Ruden, Cingolani, Sen, Qu, Wang, Senut, Garfinkel, Sollars and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Cytosine modifications in the honey bee (*Apis mellifera*) worker genome

Erik M. K. Rasmussen^{1*} and Gro V. Amdam^{1,2}

¹ Department of Chemistry, Biotechnology and Food Science, Faculty of Veterinary Medicine and Biosciences, Norwegian University of Life Sciences, Aas, Norway

² School of Life Sciences, Arizona State University, Tempe, AZ, USA

Edited by:

Greg J. Hunt, Purdue University, USA

Reviewed by:

Bernardo Lemos, Harvard University, USA

Amy Lynn Toth, Iowa State University, USA

*Correspondence:

Erik M. K. Rasmussen, Department of Chemistry, Biotechnology and Food Science, Faculty of Veterinary Medicine and Biosciences, Norwegian University of Life Sciences, Christian Magnus Falsens vei 1, N-1432 Aas, Norway
e-mail: erik.rasmussen@nmbu.no

Epigenetic changes enable genomes to respond to changes in the environment, such as altered nutrition, activity, or social setting. Epigenetic modifications, thereby, provide a source of phenotypic plasticity in many species. The honey bee (*Apis mellifera*) uses nutritionally sensitive epigenetic control mechanisms in the development of the royal caste (queens) and the workers. The workers are functionally sterile females that can take on a range of distinct physiological and/or behavioral phenotypes in response to environmental changes. Honey bees have a wide repertoire of epigenetic mechanisms which, as in mammals, include cytosine methylation, hydroxymethylated cytosines, together with the enzymatic machinery responsible for these cytosine modifications. Current data suggests that honey bees provide an excellent system for studying the “social repertoire” of the epigenome. In this review, we elucidate what is known so far about the honey bee epigenome and its mechanisms. Our discussion includes what may distinguish honey bees from other model animals, how the epigenome can influence worker behavioral task separation, and how future studies can answer central questions about the role of the epigenome in social behavior.

Keywords: honey bee, methylation, demethylation, 5-hydroxymethylcytosine, social behavior

INTRODUCTION

Since the first honey bee methylome was sequenced in 2010, our understanding of the functional implications of DNA methylation in the honey bee has begun to unfold (Lyko et al., 2010). 5-methylcytosine (5mC) is believed to be involved in alternative splicing, caste differentiation and worker behavioral task separation (Lyko et al., 2010; Flores et al., 2012; Herb et al., 2012). Recently, several other cytosine modifications were discovered in mammalian genomes (Kriaucionis and Heintz, 2009; He et al., 2011; Ito et al., 2011). These modifications are believed to have separate functions from 5mC as they are distributed differently in the genome, and specific reader proteins for one of these modifications exist (Spruijt et al., 2013). Although studies to investigate cytosine modifications other than 5mC in bees have been performed, little is known about their functions and distributions (Cingolani et al., 2013; Wojciechowski et al., 2014). Here we review cytosine modifications and the enzymatic machinery responsible for their generation in different model organisms.

HONEY BEES

Nutritional cues lead female honey bee larvae into one of two developmental trajectories. The larvae either develop into a queen or into a worker (Winston, 1991). Queens are larger, highly fecund and long-lived (years), while the smaller workers are functionally sterile and shorter lived (weeks, months). Workers show a flexible physiological and behavioral progression that typically starts with care behavior toward siblings (nursing) and culminates in food collection (foraging) weeks later. Nursing is associated with enhanced somatic maintenance and slower aging than foraging

(Münch and Amdam, 2010). Yet, foragers can return to nursing tasks, and this behavioral reversion can put age-associated cognitive decline in reverse as well (Baker et al., 2012).

Honey bees, in other words, display a wide range of phenotypes that include complex social caste development and behavior, behavioral shifts, and plasticity of aging. Epigenetic mechanisms are already found to likely play major roles in queen-worker development as well as in worker behavioral progression and reversion (Kucharski et al., 2008; Spannhoff et al., 2011; Herb et al., 2012). These findings put the honey bee forward as a very interesting study organism to investigate the interplay between the social milieu and the epigenome. The use of the honey bee for complex epigenetic research is, furthermore, not diminished by the mainstream models fruit fly (*Drosophila melanogaster*) and nematode (*Caenorhabditis elegans*), since they do not have the full complement of the mammalian epigenetic machinery (Table 1).

EPIGENETIC MACHINERY

DNA methyltransferases (DNMTs) are enzymes that add a methyl group to the 5' carbon of the DNA base cytosine from the donor S-Adenosyl methionine (Law and Jacobsen, 2010). DNMT1 is the “maintenance” DNMT that copies the methylation pattern to the newly synthesized strand during DNA replication. DNMT3 is the *de novo* methyltransferase that can methylate specific loci independently of replication. DNMT2 is primarily an RNA methyltransferase that methylates t-RNA^{Asp} (Goll et al., 2006), however, DNA activity has been shown *in vivo* in the fruit fly (Phalke et al., 2009). The *de novo* and the maintenance DNMTs

Table 1 | Genomic copies of enzymes implicated in DNA methylation and demethylation and presence of epigenetically modified cytosines in select metazoan groups.

Organism	DNMT1	DNMT3	DNMT2	TET	TDG	5mC in CpG	5hmC	5fC	5caC
Nematode							?	?	?
Fly			•	•	•		?	?	?
Aphid	••	•	•	•		•	?	?	?
Jewel Wasp	•••	•	•	•	•	•	?	?	?
Bee	••	•	•	•	•	•	•	?	?
Mammals	•	•••	•	•••	•	•	•	•	•

Sources: (Kriaucionis and Heintz, 2009; Law and Jacobsen, 2010; Lyko et al., 2010; Walsh et al., 2010; Ito et al., 2011; Cingolani et al., 2013; Beeler et al., 2014) and assembled genomes available at <http://blast.ncbi.nlm.nih.gov>.

are found in a range of species including honey bees, mammals, aphids, and jewel wasps (Table 1). They are catalytically active in the honey bee (Wang et al., 2006), while fruit fly and nematode only contain a single copy of DNMT2. Nevertheless, 5mC originating from DNA has been reported in the fruit fly in both embryos and adult flies (Lyko et al., 2000), suggesting that DNMT2 has some DNA methylation activity *in vivo*. The impact of 5mC in the fruit fly genome is still debated, however (Phalke et al., 2010; Schaefer and Lyko, 2010).

In mammals, the ten eleven translocation (TET) enzyme is responsible for further oxidizing 5mC to 5-hydroxymethylcytosine (5hmC) that again can be oxidized to 5-formylcytosine (5fC), and ultimately 5-carboxylcytosine (5caC) (Tahiliani et al., 2009; He et al., 2011; Ito et al., 2011). 5fC and 5caC are recognized by the thymine DNA glycosylase (TDG), which is a part of the base excision repair pathway of the mammalian cell (Maiti and Drohat, 2011). The TET enzyme together with TDG are probably central to the mammalian active demethylation pathway (Pastor et al., 2013). Mammalian genomes harbor multiple TET enzyme genes, while bees, fruit flies, aphids, and jewel wasps only have one (Table 1). The RNA expression levels of the different mammalian TET enzymes vary greatly between developmental stages and cell types. The honey bee TET catalytic domain is catalytically active *in vitro*, and active transcription of the honey bee TET gene has been shown to vary in different stages of development as well as in different adult tissues (Wojciechowski et al., 2014). Interestingly, some species (including fruit fly) that contain only DNMT2 have well conserved TET orthologs, but their activity and function have not been deciphered (Dunwell et al., 2013).

The honey bee genome encodes several core histone modifying enzymes, which are also part of the epigenetic machinery of the honey bees (The Honeybee Genome Sequencing, 2006). However, the impact of and the mechanisms behind histone modifications are beyond the scope of this review.

5-METHYLCYTOSINE

The distribution and relative abundance of 5mC vary significantly between mammals, honey bee and fruit fly (Figure 1). 5mC is primarily located in a CpG dinucleotide context within repeat sequences and in proximity of promoter areas in mammals (Law and Jacobsen, 2010), whereas in bees methylated CpGs are

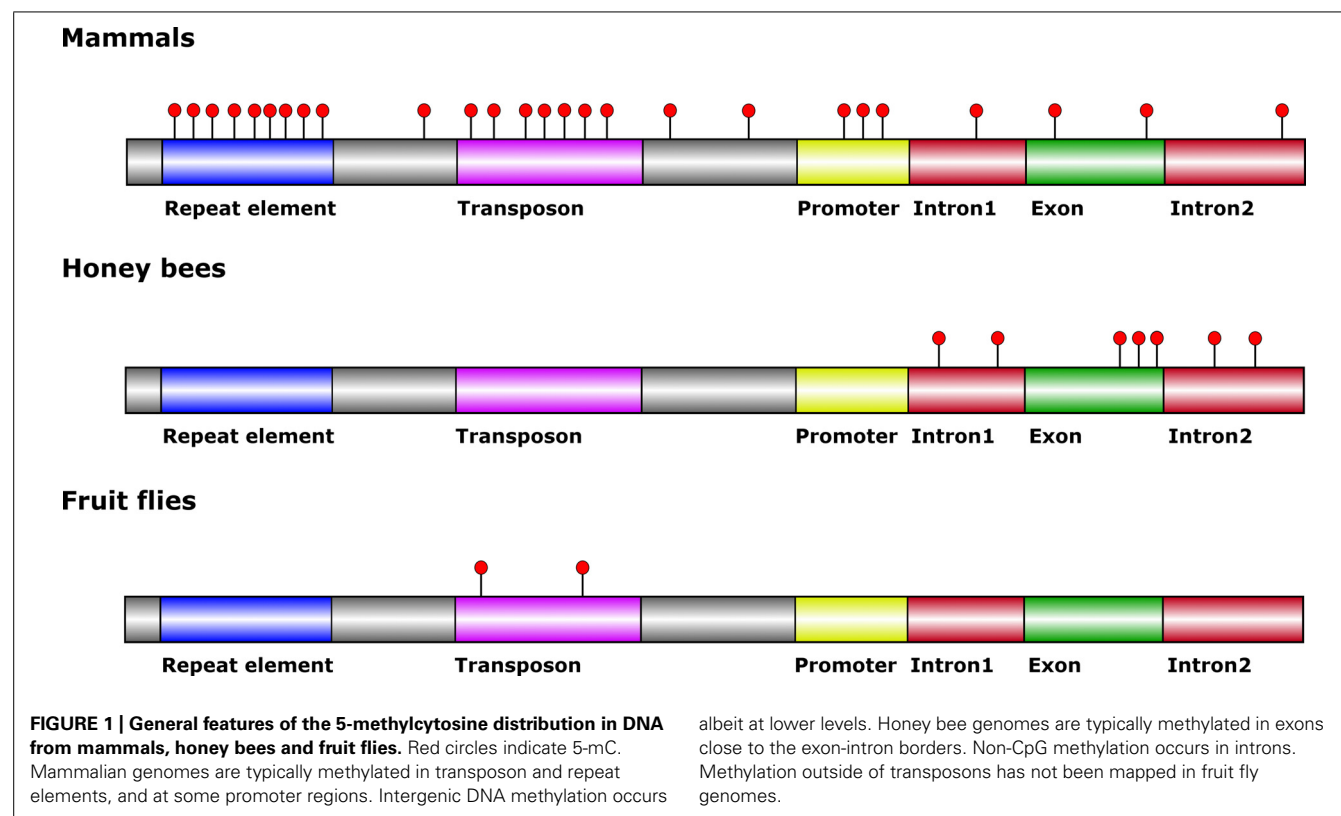
primarily located within genes (Lyko et al., 2010). However, 5mC can exist in a non-CpG dinucleotide context in both mammals and honey bees (Lister et al., 2009; Cingolani et al., 2013). In addition, the honey bee genome is much more sparsely methylated than mammalian genomes, thus reducing overall complexity and simplifying data analyses for studies conducted in bees. In the fruit fly genome, 5-mC is located within a non-CpG dinucleotide context and seems to be distributed randomly within the genome at an abundance 3- to 100-fold less when compared to honey bees and mammals (Mandrioli and Borsatti, 2006; Phalke et al., 2009). *C. elegans*, on the other hand, does not contain 5mC in its genome (Simpson et al., 1986).

The effect of 5mC on transcription varies between metazoans and genomic context. In mammalian promoters, 5mC is principally a repressive mark, silencing transcription (Bird, 2002). On the other hand, 5mC within gene bodies in mammals, honey bees, and the fruit fly, does not influence transcription levels to the same extent (Mandrioli and Borsatti, 2006; Flores et al., 2012). In honey bees, 5mC within gene bodies rather plays a role in the generation of alternative splice variants on the genome-wide level (Flores et al., 2012; Foret et al., 2012; Li-Byarlay et al., 2013). This role is not clearly defined in mammalian cells, as the role of 5mC in gene bodies differs between cell types and depends on whether 5mC is in a CpG context or not (Lister et al., 2009). These findings make honey bees an attractive system for studies on how 5mC influences the generation of alternative transcripts.

5-methylcytosine is found in multiple cell types, tissues, and life stages in both honey bees and mammals (Ikeda et al., 2011; Ziller et al., 2013). In *D. melanogaster*, 5mC is mostly found during early embryonic stages (Lyko et al., 2000). Although adult 5mC has been reported in fruit fly, the content is too low to be robustly detected by bisulfite sequencing, the gold standard in base resolution 5mC interrogation techniques, making further studies difficult with many established methods depending on bisulfite conversion (Capuano et al., 2014).

5-HYDROXYMETHYLCYTOSINE

The TET oxidative products of 5mC recently became a center of attention in mammalian epigenetic research. Many questions about TET and 5hmC dynamics have been answered in embryonic stem cells (Pastor et al., 2013), although 5hmC has been detected in different tissues at different life stages (Kriaucionis and Heintz,



2009; Ivanov et al., 2013). The abundance of 5hmC compared to 5mC is much lower ranging from 2- to 100-fold times less depending on tissue (Kriaucionis and Heintz, 2009; Song et al., 2012). The distribution of 5hmC does not seem to be directly linked to 5mC, as 5hmC is found more often in promoter areas and enhancers, and much less in repetitive elements (Pastor et al., 2011; Stroud et al., 2011; Yu et al., 2012). In addition, proteins capable of specifically binding 5hmC have been discovered, fueling the theory that 5hmC exists as separate epigenetic mark and not simply just as an intermediate in an active demethylation pathway (Frauer et al., 2011; Mellen et al., 2012; Spruijt et al., 2013). In honey bees, 5hmC has been characterized in multiple tissues, and its abundance seems to be highest in germ cells and the brain (7–10% of 5mC and about 4% of 5mC, respectively), following the trend in mammalian cell types (Wojciechowski et al., 2014). Only one study has attempted to map 5hmC in honey bees at a single nucleotide resolution (Cingolani et al., 2013). This same study, surprisingly, mapped the majority of 5hmC in head tissue to non-CpG intronic sequences. Further studies seems warranted to precisely quantify and map 5hmC in bees, especially in non-brain tissue, which has received less interest so far. To date, 5hmC together with 5fC and 5caC have not been identified in the fruit fly, aphid, jewel wasp, and *C. elegans* genomes. However, since *C. elegans* has no 5mC precursor or TET homolog, the existence of 5hmC, 5fC, and 5caC seems highly unlikely.

5-FORMYLCYTOSINE AND 5-CARBOXYLCYTOSINE

The recently identified nucleotides 5fC and 5caC have, so far, not accumulated the same level of information as their precursors

5mC and 5hmC. This situation is in part due to extremely low abundance, especially for 5caC, making robust detection difficult (in mammals 5caC is 10- to 1000-fold less abundant than 5hmC). Moreover, the molecular toolbox for investigating 5fC and 5caC is not as developed as it is for 5hmC (Song and He, 2013). Bisulfite sequencing for example, only discriminates between “methylated” and “unmethylated” cytosines, so that 5mC and 5hmC are identified as “methylated” and 5fC and 5caC as “unmethylated” (Pastor et al., 2013). Such data are therefore difficult to use as guidelines in narrowing down possible locations of 5fC and 5caC.

The extremely low abundance of 5caC suggest that this nucleotide is merely an intermediate step in complete demethylation (Song and He, 2013). Although 5fC is a more prominent epigenetic mark than 5caC, its function is still not fully understood. It is possible that 5fC might regulate transcription through stalling of RNA pol II (Kellinger et al., 2012), but further research is needed to elucidate the role of 5fC and 5caC in both vertebrates and invertebrates. In honey bees, 5fC and 5caC have not been investigated yet, though their precursors and catalytic enzyme have been reported (Lyko et al., 2010; Wojciechowski et al., 2014).

FUTURE WORK: EPIGENETICS AND WORKER BEHAVIOR

Epigenetic mechanisms have been linked to the queen-worker differentiation of honey bees (Kucharski et al., 2008), as well as to worker behavioral progression and reversion (Herb et al., 2012). Herb et al. (2012) bisulfite sequenced brains of age-matched nurses, foragers, and reverted workers (previous foragers now involved in care behavior). Their data revealed differentially

methylated regions (DMRs) between the behavioral groups indicating that DNA methylation can play a role in regulation of social behavior. These DMRs are associated with genes involved in development, nuclear pore formation, and ATP binding. RNA sequencing revealed that these same DMRs were connected to alternative splicing events. It is also very likely that the “behaviorally sensitive” DMRs of honey bees are hydroxymethylated at some point during either transition from nurse to forager, or reversion from forager to nurse. Since the study was conducted in adult brain tissue, which has no neurogenesis (Fahrback et al., 1995), dilution by replication would be unlikely or would only display minor effects. This situation makes these DMRs excellent candidates for investigating if 5hmC is associated with worker behavioral transitions, and if these hydroxymethylated regions are differentially hydroxymethylated between nurses, foragers, and reverted worker bees. Such a study could be the first to establish a putative link between hydroxymethylation and behavior.

Future studies should also dissect the role of TET in worker transitions from nurse to foragers, and back. Other candidate tissues than brain should include the fat body. This tissue is functionally homologous to liver and white adipose tissue and undergoes major remodeling during honey bee behavioral change (Chan et al., 2011). Functional implications of an RNA interference-mediated TET knockdown should provide insight into TET function. Studies can be conducted in honey bee larvae to investigate if TET knockdowns are capable of both queen and worker development. Similarly, consequences for behavioral plasticity can be studied in adult honey bee workers and perhaps link TET and its products with behavior for the first time.

Finally, a possible link between 5hmC and alternative splicing can be investigated by combining 5hmC sequencing at single nucleotide resolution with RNA sequencing of honey bee tissue samples. 5mC is reportedly implicated in the generation of alternative transcripts in the bee, but using methods not able to distinguish 5mC from 5hmC (Flores et al., 2012; Herb et al., 2012). Therefore, further studies that can map 5hmC alongside RNA sequencing data seems warranted, and could potentially give 5hmC a novel function in gene regulation.

CONCLUSION

The honey bee offers a system where the interplay between DNA methylation and social behavior can be studied in great detail. Published studies of the honey bee epigenome are dominated by questions surrounding queen and worker development, while the epigenetic dynamics of worker behavioral castes have only more recently gained attention. The readily identifiable social behaviors of worker honey bees make setting up precise, large scale experiments feasible (Münch et al., 2013). Better knowledge about honey bee epigenetics also has a dual purpose; increasing the understanding of epigenetic machineries in general, and gaining specific information about gene regulatory mechanisms in an economically important beneficial insect.

ACKNOWLEDGMENTS

We thank the two reviewers for their helpful comments. This study was supported by funds from the Research Council of Norway, grant no. 191699.

REFERENCES

- Baker, N., Wolschin, F., and Amdam, G. V. (2012). Age-related learning deficits can be reversible in honeybees *Apis mellifera*. *Exp. Gerontol.* 47, 764–772. doi: 10.1016/j.exger.2012.05.011
- Beeler, S. M., Wong, G. T., Zheng, J. M., Bush, E. C., Remnant, E. J., Oldroyd, B. P., et al. (2014). Whole-genome DNA methylation profile of the jewel wasp (*Nasonia vitripennis*). *G3 (Bethesda)* 4, 383–388. doi: 10.1534/g3.113.008953
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21. doi: 10.1101/gad.947102
- Capuano, F., Mulleder, M., Kok, R., Blom, H. J., and Ralser, M. (2014). Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. *Anal. Chem.* 86, 3697–3702. doi: 10.1021/ac500447w
- Chan, Q. W., Mutti, N. S., Foster, L. J., Kocher, S. D., Amdam, G. V., and Wolschin, F. (2011). The worker honeybee fat body proteome is extensively remodeled preceding a major life-history transition. *PLoS ONE* 6:e24794. doi: 10.1371/journal.pone.0024794
- Cingolani, P., Cao, X., Khetani, R. S., Chen, C. C., Coon, M., Sammak, A. A., et al. (2013). Intronic Non-CG DNA hydroxymethylation and alternative mRNA splicing in honey bees. *BMC Genomics* 14:666. doi: 10.1186/1471-2164-14-666
- Dunwell, T. L., McGuffin, L. J., Dunwell, J. M., and Pfeifer, G. P. (2013). The mysterious presence of a 5-methylcytosine oxidase in the *Drosophila* genome: possible explanations. *Cell Cycle* 12, 3357–3365. doi: 10.4161/cc.26540
- Fahrback, S. E., Strande, J. L., and Robinson, G. E. (1995). Neurogenesis is absent in the brains of adult honey bees and does not explain behavioral neuroplasticity. *Neurosci. Lett.* 197, 145–148. doi: 10.1016/0304-3940(95)11913-H
- Flores, K. B., Wolschin, F., Allen, A. N., Corneveaux, J. J., Huentelman, M., and Amdam, G. V. (2012). Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics* 13:480. doi: 10.1186/1471-2164-13-480
- Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S. E., Robinson, G. E., et al. (2012). DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honeybees. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4968–4973. doi: 10.1073/pnas.1202392109
- Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M. C., Antes, I., et al. (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* 6:e21306. doi: 10.1371/journal.pone.0021306
- Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C.-L., Zhang, X., et al. (2006). Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science* 311, 395–398. doi: 10.1126/science.1120976
- He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307. doi: 10.1126/science.1210944
- Herb, B. R., Wolschin, F., Hansen, K. D., Aryee, M. J., Langmead, B., Irizarry, R., et al. (2012). Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nat. Neurosci.* 15, 1371–1373. doi: 10.1038/nn.3218
- Ikeda, T., Furukawa, S., Nakamura, J., Sasaki, M., and Sasaki, T. (2011). CpG methylation in the hexamerin 110 gene in the European honeybee, *Apis mellifera*. *J. Insect. Sci.* 11:74. doi: 10.1673/031.011.7401
- Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., et al. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303. doi: 10.1126/science.1210597
- Ivanov, M., Kals, M., Kacevska, M., Barragan, I., Kasuga, K., Rane, A., et al. (2013). Ontogeny, distribution and potential roles of 5-hydroxymethylcytosine in human liver function. *Genome Biol.* 14:R83. doi: 10.1186/gb-2013-14-8-r83
- Kellinger, M. W., Song, C. X., Chong, J., Lu, X. Y., He, C., and Wang, D. (2012). 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 19, 831–833. doi: 10.1038/nsmb.2346
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930. doi: 10.1126/science.1169786
- Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319, 1827–1830. doi: 10.1126/science.1153069
- Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220. doi: 10.1038/nrg2719

- Li-Byarlay, H., Li, Y., Stroud, H., Feng, S., Newman, T. C., Kaneda, M., et al. (2013). RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12750–12755. doi: 10.1073/pnas.1310735110
- Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322. doi: 10.1038/nature08514
- Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* 8:e1000506. doi: 10.1371/journal.pbio.1000506
- Lyko, F., Ramsahoye, B. H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. *Nature* 408, 538–540. doi: 10.1038/35046205
- Maiti, A., and Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* 286, 35334–35338. doi: 10.1074/jbc.C111.284620
- Mandrioli, M., and Borsatti, F. (2006). DNA methylation of fly genes and transposons. *Cell. Mol. Life Sci.* 63, 1933–1936. doi: 10.1007/s00018-006-6039-1
- Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151, 1417–1430. doi: 10.1016/j.cell.2012.11.022
- Münch, D., and Amdam, G. V. (2010). The curious case of aging plasticity in honey bees. *FEBS Lett.* 584, 2496–2503. doi: 10.1016/j.febslet.2010.04.007
- Münch, D., Baker, N., Rasmussen, E. M. K., Shah, A. K., Kreibich, C. D., Heidem, L. E., et al. (2013). Obtaining specimens with slowed, accelerated and reversed aging in the honey bee model. *J. Vis. Exp.* 8, e50550. doi: 10.3791/50550
- Pastor, W. A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* 14, 341–356. doi: 10.1038/nrm3589
- Pastor, W. A., Pape, U. J., Huang, Y., Henderson, H. R., Lister, R., Ko, M., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473, 394–397. doi: 10.1038/nature10102
- Phalke, S., Nickel, O., and Reuter, G. (2010). Reply to “Lack of evidence for DNA methylation of Invader4 retroelements in *Drosophila* and implications for Dnmt2-mediated epigenetic regulation.” *Nat. Genet.* 42, 921–921. doi: 10.1038/ng1110-921
- Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M. C., and Reuter, G. (2009). Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nat. Genet.* 41, 696–702. doi: 10.1038/ng.360
- Schaefer, M., and Lyko, F. (2010). Lack of evidence for DNA methylation of Invader4 retroelements in *Drosophila* and implications for Dnmt2-mediated epigenetic regulation. *Nat. Genet.* 42, 920–921; author reply 921. doi: 10.1038/ng1110-920
- Simpson, V. J., Johnson, T. E., and Hammen, R. F. (1986). *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res.* 14, 6711–6719. doi: 10.1093/nar/14.16.6711
- Song, C. X., and He, C. (2013). Potential functional roles of DNA demethylation intermediates. *Trends Biochem. Sci.* 38, 480–484. doi: 10.1016/j.tibs.2013.07.003
- Song, C. X., Yi, C., and He, C. (2012). Mapping recently identified nucleotide variants in the genome and transcriptome. *Nat. Biotechnol.* 30, 1107–1116. doi: 10.1038/nbt.2398
- Spannhoff, A., Kim, Y. K., Raynal, N. J., Gharibyan, V., Su, M. B., Zhou, Y. Y., et al. (2011). Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. *EMBO Rep.* 12, 238–243. doi: 10.1038/embor.2011.9
- Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146–1159. doi: 10.1016/j.cell.2013.02.004
- Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., and Jacobsen, S. E. (2011). 5-hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* 12:R54. doi: 10.1186/gb-2011-12-6-r54
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935. doi: 10.1126/science.1170116
- The Honeybee Genome Sequencing, C. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443, 931–949. doi: 10.1038/nature05260
- Walsh, T. K., Brisson, J. A., Robertson, H. M., Gordon, K., Jaubert-Possamai, S., Tagu, D., et al. (2010). A functional DNA methylation system in the pea aphid, *Acyrtosiphon pisum*. *Insect Mol. Biol.* 19(Suppl. 2), 215–228. doi: 10.1111/j.1365-2583.2009.00974.x
- Wang, Y., Jorda, M., Jones, P. L., Maleszka, R., Ling, X., Robertson, H. M., et al. (2006). Functional CpG methylation system in a social insect. *Science* 314, 645–647. doi: 10.1126/science.1135213
- Winston, M. L. (1991). *The Biology of the Honey Bee*. Cambridge, MA: Harvard University Press.
- Wojciechowski, M., Rafalski, D., Kucharski, R., Misztal, K., Maleszka, J., Bochtler, M., et al. (2014). Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. *Open Biol.* 4:140110. doi: 10.1098/rsob.140110
- Yu, M., Hon, G. C., Szulwach, K. E., Song, C. X., Zhang, L., Kim, A., et al. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149, 1368–1380. doi: 10.1016/j.cell.2012.04.027
- Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T., Kohlbacher, O., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481. doi: 10.1038/nature12433

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 September 2014; accepted: 09 January 2015; published online: 06 February 2015.

Citation: Rasmussen EMK and Amdam GV (2015) Cytosine modifications in the honey bee (*Apis mellifera*) worker genome. *Front. Genet.* 6:8. doi: 10.3389/fgene.2015.00008

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2015 Rasmussen and Amdam. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Social parasitism and the molecular basis of phenotypic evolution

Alessandro Cini^{1†‡}, Solenn Patalano^{2,3‡}, Anne Segonds-Pichon³, George B. J. Busby^{2,4}, Rita Cervo¹ and Seirian Sumner^{2,5*}

¹ Dipartimento di Biologia, Università di Firenze, Firenze, Italy

² Institute of Zoology, Zoological Society of London, London, UK

³ The Babraham Institute, Babraham Research Campus – Cambridge, Cambridge, UK

⁴ Wellcome Trust Centre for Human Genetics, Oxford, UK

⁵ School of Biological Sciences, University of Bristol, Bristol, UK

Edited by:

Juergen Rudolf Gadau, Arizona State University, USA

Reviewed by:

Zachary Cheviron, University of Illinois, Urbana-Champaign, USA

René Massimiliano Marsano, University of Bari, Italy

*Correspondence:

Seirian Sumner, School of Biological Sciences, Life Sciences Building, 24 Tyndall Avenue, Bristol BS8 1TQ, UK
e-mail: seirian.sumner@bristol.ac.uk

† Present address:

Alessandro Cini, CRA – ABP Consiglio per la Ricerca in Agricoltura e l'Analisi Dell'Economia Agraria, Centro di Ricerca per l'Agrobiologia e la Pedologia, Cascine del Riccio, Firenze, Italy and Corpo Forestale dello Stato, Centro Nazionale Biodiversità Forestale "Bosco Fontana", Verona, Italy

‡ These authors have contributed equally to this work.

Contrasting phenotypes arise from similar genomes through a combination of losses, gains, co-option and modifications of inherited genomic material. Understanding the molecular basis of this phenotypic diversity is a fundamental challenge in modern evolutionary biology. Comparisons of the genes and their expression patterns underlying traits in closely related species offer an unrivaled opportunity to evaluate the extent to which genomic material is reorganized to produce novel traits. Advances in molecular methods now allow us to dissect the molecular machinery underlying phenotypic diversity in almost any organism, from single-celled entities to the most complex vertebrates. Here we discuss how comparisons of social parasites and their free-living hosts may provide unique insights into the molecular basis of phenotypic evolution. Social parasites evolve from a eusocial ancestor and are specialized to exploit the socially acquired resources of their closely-related eusocial host. Molecular comparisons of such species pairs can reveal how genomic material is re-organized in the loss of ancestral traits (i.e., of free-living traits in the parasites) and the gain of new ones (i.e., specialist traits required for a parasitic lifestyle). We define hypotheses on the molecular basis of phenotypes in the evolution of social parasitism and discuss their wider application in our understanding of the molecular basis of phenotypic diversity within the theoretical framework of phenotypic plasticity and shifting reaction norms. Currently there are no data available to test these hypotheses, and so we also provide some proof of concept data using the paper wasp social parasite/host system (*Polistes sulcifer*—*Polistes dominula*). This conceptual framework and first empirical data provide a spring-board for directing future genomic analyses on exploiting social parasites as a route to understanding the evolution of phenotypic specialization.

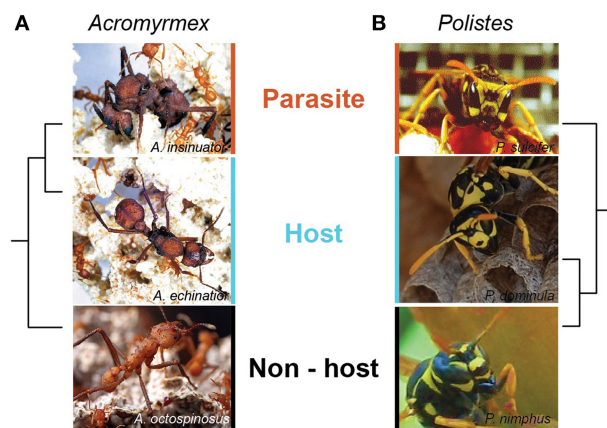
Keywords: phenotypic plasticity, social insects, *Polistes*, social parasites, genomics, gene expression

INTRODUCTION

THE MOLECULAR BASIS OF PHENOTYPIC DIVERSITY

Evolution plays with inherited traits to produce altered phenotypes which may be better adapted to fill a niche different to that of their ancestors. Ultimately, phenotypic traits arise at the level of the genes. A major outstanding question in evolutionary biology is what roles do losses, gains, co-options and modifications of genomic material play in the evolution of phenotypic diversity within and between species? (West-Eberhard, 2003; Kaessmann, 2010; Van Dyken and Wade, 2010; Wissler et al., 2013) Many species show phenotypic plasticity in the expression of alternative phenotypes from the same genotype, through variance in reaction norm responses to changes in the environment (Aubin-Horth and Renn, 2009). Such plasticity affects both short-term (ecological) and long-term (evolutionary) adaptation, and thus influences survival and fitness (Pfennig et al., 2010; Beldade et al., 2011; Hughes, 2012). Conditional expression of the genes underlying

polyphenisms facilitate gene, and consequently phenotypic, evolution (Van Dyken and Wade, 2010). Canalized developmental pathways shaped by evolution can result in heritable shifts in phenotype (Waddington, 1942). Genomic methods in modern evolutionary biology now allow us to dissect the molecular basis of such phenotypic diversity across a range of organisms, from genes to phenotypes (Tautz et al., 2010). But selection acts directly on phenotypes and only indirectly on the molecular machinery, and so an integrated study of key phenotypic traits in ecologically relevant settings and the genes associated with them is essential (West-Eberhard, 2005; Schwander and Leimar, 2011; Valcu and Kempenaers, 2014). Insects provide excellent models for studying these facets of phenotypic evolution within and across species (Nijhout, 2003; Moczek, 2010; Simpson et al., 2011), e.g., eusocial insect castes (Evans and Wheeler, 2001; Smith et al., 2008), male morphologies beetles (Moczek, 2009), asexual and sexual reproductive phases in aphids (Brisson and Stern, 2006).

Box 1 | Obligate social parasites and their hosts as models.

There are several features unique to hymenopteran obligate social parasite/host systems that make them ideal models for studying the molecular basis of phenotypic diversity.

1. Close phylogenetic relationships. Social parasites of hymenopterans are usually close relatives of their host, in a strict **(A)** or loose **(B)** sense, and thus share recent genomic ancestry (Lowe et al., 2002; Savolainen and Vepsäläinen, 2003; Sumner et al., 2004a; Huang and Dornhaus, 2008; Smith et al., 2013). Hosts are likely to represent the eusocial ancestor of the parasite at the molecular and phenotypic levels, providing the opportunity to compare how the parasite has diverged from its ancestral state.
2. Non-host sister species. Intriguingly, sister species of hosts are often resistant to their relative's social parasite, despite occurring sympatrically and sharing similar ecologies, phenotypes, life histories and environments. It is not known how non-hosts confer resistance, but comparisons of host and non-hosts at the molecular level may shed light on this.
3. Cryptic morphology. Although hymenopteran social parasites differ significantly to their hosts in life strategy and behavior, they are usually near indistinguishable from their hosts morphologically e.g., **(A)**: *Acromyrmex insinuator* (social parasite), *Acromyrmex echinator* (the host), and *Acromyrmex octospinosus* (non-host sister species); **(B)**: *Polistes sulcifer* (social parasite), *Polistes dominula* (the host) and *Polistes nimphus* (non-host sister species). This is important for genomic analyses of phenotypic plasticity where we are interested in understanding the molecular basis of traits other than morphology (e.g., behavior). Shared morphology between parasite and host therefore helps to controls to some extent for the machinery underlying morphological differences. Molecular analyses also help with social parasite species discovery, as the parasites may be cryptic at the morphological level, but not at the molecular level.
4. Trait losses and gains. Because both social parasite and host can be easily observed within and out of the nest, phenotypic traits can be easily identified, quantified and compared. Social parasites lack a wealth of free-living traits (e.g., maternal care, provisioning, nest-founding), but also exhibit novel traits (e.g., fighting ability, usurpation behaviors, chemical mimicry,

(Continued)

Box 1 | Continued

cryptic manipulation). Whilst these are well studied at the phenotypic level, we know nothing about how such losses and gains occur at the molecular level. Comparisons of the molecular bases of closely related host and social parasite traits will provide new insights into phenotypic evolution.

Photo credits: Alessandro Cini, Rita Cervo, Stefano Turillazzi and David Nash.

An ideal model system for determining the molecular basis of phenotypic evolution allows comparisons of related species which have evolved mutually exclusive traits and/or life histories (e.g., Arendt and Reznick, 2008; Schlichting and Wund, 2014). Parasites are good examples of species that have lost ancestral, free-living traits and gained new ones to evolve a specialized life-history that depends on exploiting the resources of other species. For example, endoparasitic worms have lost ancestral gut, head and light sensing organs, but have gained traits such as a specialized tegument, which protects them from host-stomach acids (Burton et al., 2012). Hosts co-evolve to combat parasitism, through enhanced immune responses and mechanisms for detecting infection; parasites manipulate their host to benefit the parasite's life cycle, often through an extended phenotype (Dawkins, 1982). Comparisons between parasites and their free-living relatives therefore present intriguing models for studying the molecular basis of phenotypic evolution (Dybdahl et al., 2014). However, these comparisons are complicated by co-evolution where frequency distributions of host and parasite genotypes (and traits) shift reciprocally and responsively over time, and moreover hosts and their parasites are rarely closely related species (Hamilton, 1980).

Insect social parasites and avian brood parasites differ from other parasites in that they exploit the parental behavior of the hosts rather than the physical resources of individuals. Such parasites have evolved several times in the animal kingdom. For example, cuckoldry occurs in more than 100 bird species, where the host pays the cost of raising unrelated chicks (Davies, 2000). Social parasites of eusocial insects (e.g., the Hymenoptera—bees, wasps and ants) are especially interesting as they are usually close relatives of their hosts, and have often entirely lost their worker caste (Savolainen and Vepsäläinen, 2003). The potential for using social parasites, especially of eusocial insects, as models for understanding the molecular basis of phenotypic plasticity has been recognized (West-Eberhard, 1989, 2003). However, we lack a defined theoretical framework and clear hypotheses to properly exploit this untapped niche using molecular studies. Advances in molecular technologies now make gene-level studies accessible in any organism. It is therefore timely to lay out a framework for exploiting social parasites and their hosts as models for understanding the genomic basis of phenotypic losses and gains in evolution. Here we identify the key traits of hymenopteran social parasites of eusocial insects that make them useful models for understanding phenotypic evolution at the molecular level. We define specific, testable hypotheses on the

Box 2 | An example test system: the paper wasp social parasite *Polistes sulcifer* and its free-living host, the eusocial *Polistes dominula*.



The molecular basis of phenotypes in *Polistes* has received some attention over the last few years (Sumner et al., 2006; Toth et al., 2007, 2009, 2010; Daugherty et al., 2011; Ferreira et al., 2013). *P. sulcifer* is the obligate social parasite of its close relative, the primitively eusocial wasp *P. dominula* (Choudhary et al., 1994). The life-history and behaviors of the social parasite-host system *Polistes sulcifer*-*Polistes dominula*, are well known (reviewed in Cervo, 2006), but we lack molecular analyses on the social parasites.

Both species have an annual lifecycle (Pardi, 1996; Cervo, 2006). Host colonies (blue line) are founded in spring (March–April) by one or more foundresses, among which a reproductive hierarchy is soon established through the mean of dominance interactions (Pardi, 1946). The first brood emerges around the end of May or early June and develops into workers. At the end of the summer, reproductives (males and females) emerge on the nest, leave the colony and mate. Males die soon after mating. Mated females cluster together in sheltered places to overwinter. Those who survive overwinter found new colonies the following spring (Pardi, 1996). Parasite females (orange line) emerge later than their hosts (late May) from overwintering (Cervo and Turillazzi, 1996) and migrate from their overwintering sites to pre-emergence host nests (Cervo and Dani, 1996; Cervo, 2006). Parasites find host colonies using visual and chemical stimuli (Cervo et al., 1996; Cini et al., 2011a). Nest usurpation takes place during a small window of time (late May–early June) (Cervo and Turillazzi, 1996; Ortolani et al., 2008) and it involves violent fights between hosts and parasites (Turillazzi et al., 1990; Cini et al., 2011b). Parasites display a novel behavior during this time (restlessness) (Ortolani et al., 2008). If the parasite is successful she becomes the sole egg-layer of the nest, adopting both the behaviors and chemical signatures of the host queen (Turillazzi et al., 2000; Sledge et al., 2001; Dapporto et al., 2004). After colony usurpation, the social parasite and un-parasitised host queens share the same environmental and social conditions (temperature, microclimate, diet etc.). Photo credits: Alessandro Cini, Rita Cervo and Stefano Turillazzi.

molecular basis of shared and contrasting traits in the evolution of social parasitism within the conceptual framework of shifting reaction norms and phenotypic plasticity (e.g., Aubin-Horth and Renn, 2009; Fusco and Minelli, 2010). We also provide a first test of some of these hypotheses, as proof of concept for our conceptual model and a spring-board for future genomic analyses on the evolution of phenotypic adaptation (see Supplementary Materials).

SOCIAL PARASITISM IN EUSOCIAL INSECTS

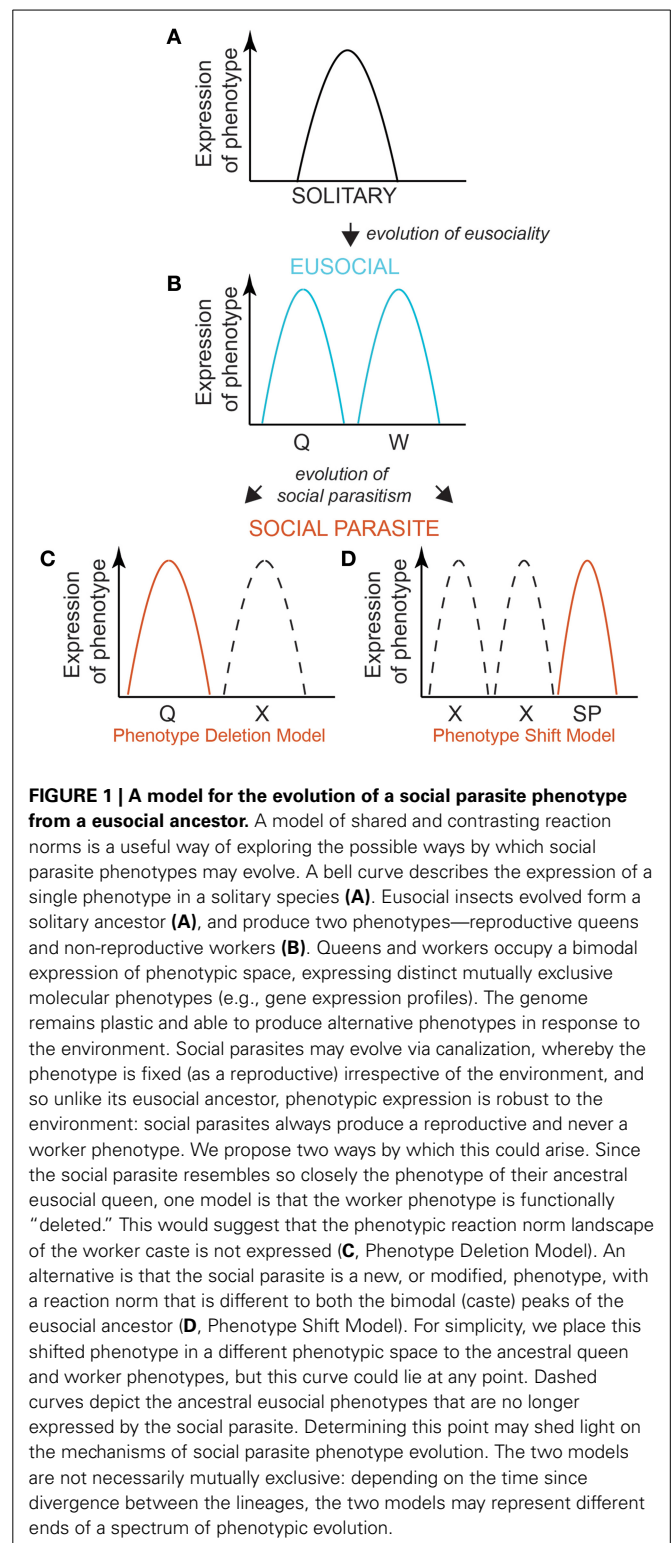
There are over 14,000 eusocial species in the Hymenoptera (bees, wasps and ants) representing over 11 independent origins of eusociality. Their societies are defined by a division of reproductive labor in the form of queen and worker castes, overlapping of generations, and cooperative brood care. Social parasitism has evolved multiple times independently in the eusocial insects: three times in wasps (once in Polistinae *Polistes*—Choudhary

et al., 1994; Cervo, 2006; twice in Vespinae—genus *Vespa* and *Dolichovespula*, Carpenter and Perera, 2006); at least 12 times in bees [three times in bumblebees - *Bombus* (subgenus *Psithyrus*, *Thoracobombus*) and *Alpinobombus*, (Alford, 1975; Cameron et al., 2007; Hines and Cameron, 2010)]; seven times in Allodapinae (Tierney et al., 2008; Smith et al., 2013); twice in Halictidae (*Dialictus* genus, Gibbs et al., 2012); and multiple times in the ants (Huang and Dornhaus, 2008; Buschinger, 2009).

There are several features of social parasite/host systems that make them ideal models for studying the molecular basis of phenotypic diversity. Their easily observable behaviors (e.g., paper wasps *Polistes*, Cervo, 2006, and leafcutting ants *Acromyrmex*; Sumner et al., 2004a) facilitate an integrated study of the behavioral phenotype with the molecular one. Social parasites are usually closely related to their hosts and thus share recent genomic (and phenotypic) ancestry (**Box 1**) (Choudhary et al., 1994; Lowe et al., 2002; Savolainen and Vepsäläinen, 2003; Sumner et al., 2004a; Huang and Dornhaus, 2008; Smith et al., 2013). Obligate social parasites depend on their host for their entire life cycle, and so have lost many of the essential free-living traits such as the ability to found a nest, produce an effective worker caste and raise offspring (Sumner et al., 2004b; Cervo, 2006; Buschinger, 2009). They have also evolved new traits, e.g., the ability to manipulate the host worker force so that parasitic offspring are raised as if they were host offspring. Full release from free-living traits means there are few restrictions on phenotypic evolution. This may facilitate phenotypic diversity at the molecular level. Obligate social parasites of eusocial insects therefore allow a direct comparison of the molecular basis of traits with recent, shared evolutionary history and contrasting traits that have evolved (and persist) within the same environmental context (see **Box 1**).

A MODEL

Eusocial species evolve from solitary ancestors. Solitary phenotypes occupy a normal distribution of variation, determined by their individual threshold level of response to environmental cues (**Figure 1A**). Queen and worker castes are alternative phenotypes that arise from the same genome, via bi-modal developmental pathways of individuals with evolved differences in their response thresholds to an environmental cue (Wheeler, 1986; Nijhout, 2003; Page and Amdam, 2007; **Figure 1B**). These alternative phenotypes arise through differential expression of shared genes, possibly via epigenetic regulation (Sumner, 2006; Smith et al., 2008; Patalano et al., 2012; Yan et al., 2014). This bi-modal landscape of phenotypic fitness is the ancestral basis from which social parasites must evolve. There are two likely routes by which specialized social parasites evolve from their eusocial ancestor. They may lose the worker phenotype and thus share a phenotypic fitness landscape with just the queens of their social ancestor (De Visser and Krug, 2014). Their phenotype response therefore becomes genetically fixed (canalized) by genetic assimilation, with selection favoring the loss of plasticity such that the genotype no longer responds to the caste-relevant environmental cue (“Phenotype Deletion Model” **Figure 1C**). Alternatively, they may evolve an entirely new phenotype with a novel/contrasting phenotype-response curve (“Phenotype Shift Model” **Figure 1D**), by genetic accommodation whereby there is



selection for altered patterns of gene expression and associated phenotypic effects (West-Eberhard, 2003; Schlichting and Wund, 2014). Under either scenario, the pre-existing polyphenism of the eusocial ancestor facilitates the evolution of the specialist social parasite. Determining which route evolution takes is

important: in the Deletion Model (**Figure 1C**) co-option of conserved genomic processes would be paramount, but with silencing of the worker response threshold (e.g., using the existing machinery used by queens to silence worker expression); in the Phenotype Shift Model (**Figure 1D**) novel genomic processes (e.g., brought about via mutation) would be important in generating a new range of response thresholds to the environment. The timing since speciation between the eusocial ancestor and social parasite is also important to consider as this may mean the two models are not mutually exclusive: the longer the time since the two lineages split, the more differences each lineage may accumulate. There may be a transition from the phenotype shift model to the phenotype deletion model for traits, gene expression or genes, depending on the time since the two lineages split.

HYPOTHESES AND PREDICTIONS

Here we present some testable hypotheses for these models. These hypotheses and predictions are specific to obligate social parasites and their eusocial insect hosts, but they may also be of general relevance to furthering our understanding of the molecular basis of phenotypic diversity. The empirical approach we suggest requires a combined analysis of individual-level behavioral monitoring with subsequent quantitative analyses of the many components of the molecular phenotype (Pavey et al., 2010), e.g., transcription (RNAseq/transcriptomics; Ferreira et al., 2013), protein synthesis (proteomics; Begna et al., 2012), regulatory elements (e.g., microRNAs; Greenberg et al., 2012) and epigenetic modifications (Kucharski et al., 2008; Lyko et al., 2010; Bonasio et al., 2012; Simola et al., 2013). In **Figure 2**, we illustrate schematic regions of shared and contrasting trait-associated molecular phenotypes, which we refer to in our hypotheses, and suggest this as a useful way of making sense of complex genomics datasets.

HYPOTHESIS 1: CONSERVED MOLECULAR PROCESSES UNDERLIE CONVERGENT PHENOTYPES

Conserved genes, like the *Hox* gene family (Lee et al., 2003; Fernald, 2004), underlie convergent phenotypes, suggesting that phenotypic variation can evolve using shared genes and regulatory mechanisms differently (Shubin et al., 2009; Stern, 2013). By this mechanism, evolution re-uses the same ingredients (or “toolkit”) in different organisms, but tinkers with the recipe to produce different outcomes. By expressing genes at different times in development and/or in different parts of the body, the same genes can be used in different combinations, generating phenotypic diversity and innovation. Animals look different not because the molecular machinery is different, but because different parts of the machinery are activated to differing degrees, at different times, in different places and in different combinations. The number of combinations is huge, and so this is a compelling and simple explanation for the development of complex and diverse phenotypes from even a small number of genes. For example, the human genome has a mere 19,000 protein-coding genes (Ezkurdia et al., 2014), and yet humans are arguably one of the most complex products of evolution, and differ in significant ways from close relatives with similar gene sets. “Toolkit” genes are old, present in all animals and often share functions across species. Conserved toolkit genes associated

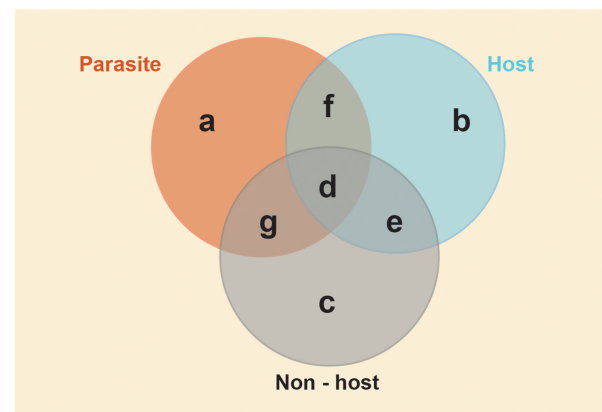


FIGURE 2 | Conceptual framework for predictions on shared and contrasting genomic/phenotypic diversity in social parasite/host relationships. Venn diagram depicting predicted shared and contrasting molecular phenotypes of non-hosts, hosts and social parasites. We define the molecular phenotype to include contrasting patterns of gene expression (significant up or down regulation), gene regulatory elements (e.g., non-coding RNAs, microRNAs, DNA methylation, histone modifications), gene interaction networks (e.g., correlated co-expression) and protein synthesis. Each area represents the molecular phenotype of the specific suite of traits. Overlapping areas indicate putatively shared molecular phenotypes. The yellow shaded area indicates the shared environment of the three species, which we predict will cause similar responses in molecular phenotypes of all three species. Conserved generic traits (area d): Molecular processes underlying traits shared by all species, and thus putatively inherited from their common ancestor. These will include fundamental machinery for growth, cell repair, metabolism, as well as more specific traits of interest that are shared among queens and social parasites such as aggression and reproduction. Identifying the molecular phenotype of this area allows tests of the genetic toolkit hypothesis. Parasite-specific (area a): Molecular processes underlying traits that have evolved in the parasite to facilitate its specialized parasitic life style, for example enhanced fighting ability, usurpation behaviors, cryptic mimicry. Identifying the molecular phenotype of this area addresses the question of whether newly evolved phenotypic traits require new genes/pathways or simply re-use existing ancestral genes/pathways. Free-living specific (area e): Molecular processes underlying free-living traits that no longer provide a fitness advantage to social parasites, e.g., parental care traits and nest founding. Identifying the molecular phenotype of this area allows us to determine what happens at the molecular level when phenotypic traits are lost, e.g., are there changes in regulation/expression, loss of processes/genes? Restricting this to traits/genes shared by free-living host and non-host species is likely to represent the traits present in the eusocial ancestor of the social parasite, and exclude processes that may have evolved subsequently. These latter processes may be associated with social parasite resistance (areas c and g) in sympatric non-hosts, host response to parasitism (area b) and co-evolved traits (area f) in host and parasite that are absent from the non-host.

with convergent social behaviors have been detected in a range of eusocial insects (Toth and Robinson, 2007; Fischman et al., 2011; Woodard et al., 2011; Toth et al., 2014), but recent work has also revealed that eusocial lineages also harbor novel (taxonomically restricted) genes that are associated with eusocial behaviors (Ferreira et al., 2013; Simola et al., 2013; Feldmeyer et al., 2014; Sumner, 2014).

Closely related social parasites and their hosts are especially powerful models for asking to what extent conserved molecular

processes underlie similar phenotypes in species with shared, recent genomic inheritance. The toolkit hypothesis predicts that host queens and social parasites will share the same molecular phenotype (i.e., express the same genes and proteins), because they are both reproductive specialists. Support for this hypothesis would suggest that social parasites are simply a reduced form of the social phenotype, expressing the reproductive component, but suppressing the worker component of their ancestors (i.e., the Phenotype Deletion Model; **Figure 1C**). Alternatively, if gene conservation is not supported, this may suggest that social parasitism evolves via Phenotype Shift (**Figure 1D**), or a combination of the two processes. This can be tested by looking at shared transcriptional patterns between social parasites and their host queens (See **Figure 2**; molecular processes underlying traits in areas d & f).

Preliminary data suggest that expression of toolkit genes is not conserved in the evolution of a social parasite, supporting the Phenotype Shift Model (**Figure 1D**). Analyses of gene expression profiles for putative toolkit genes thought to be important in castes of *Polistes* paper wasps reveal that social parasites and their host queens have distinct expression patterns (**Figure 3A**, see Supplementary Materials). This is unlikely to be a species-level effect since host workers are equally as distinct from their conspecific queens (**Figures 3A,B**). Importantly, gene expression differences between social parasites and queens were greater than among social parasites, suggesting that social parasite gene expression is not strongly overlapping with the queens among these putative toolkit genes (**Figure 3B**). Quantitative transcriptome sequencing (e.g., RNAseq) would allow a comprehensive test of this. However, these preliminary data suggest that social parasites evolve via a Phenotype Shift Model (**Figure 1D**), and that they may be a more complex phenotype than simply a partial genomic expression of the ancestral social state (as suggested by the Phenotype Deletion Model, **Figure 1C**). We predict that the shared molecular components between host and parasite will be few and limited to fundamental processes, e.g., egg production and protein storage, as characteristics of any reproductively active insect.

HYPOTHESIS 2: CONSERVED MOLECULAR PROCESSES UNDERLIE RESPONSE TO A SHARED ENVIRONMENT

Molecular phenotypes (e.g., gene expression, regulation and protein synthesis) are highly labile and can change responsively to environmental variation. A key question is whether different organisms use the same genes to respond to the same environmental cues. There will be strong selection for the social parasites to be able to accurately detect and respond to their host's environmental cues since they share the same intimate environment on the nest. Moreover, the social parasite must synchronize its life cycle and behavior perfectly with the host's life cycle (Cervo, 2006; Ortolani et al., 2008). The molecular processes underlying responsiveness to their shared environment may therefore be conserved. The Phenotype Deletion Model (**Figure 1C**) makes the implicit assumption that the phenotypes of host and parasite arise via different responses to the same environmental cue. Conversely, the Phenotype Shift Model (**Figure 1D**) is compatible with either a response to the same

cue (but with a novel threshold), or a response to a *new* cue (i.e., one that evokes no caste-related response in the eusocial host).

One important phenotype-environment response trait in both hosts and social parasites is the ability to respond to the switch from a solitary to social environment. Many eusocial insects have a solitary phase, when a single queen founds a new colony and raises her first brood alone, and then switches to a eusocial phase when her workers emerge (see **Box 2**). Likewise, social parasites have a solitary phase, during which they need to locate and successfully infiltrate a host colony, followed by a social phase where the parasite takes over the role of the queen in a society of host workers (see **Box 2**). The Phenotype Deletion Model predicts that the social parasite co-opts the molecular plasticity of its eusocial ancestor. Thus, we would expect the same genes to change in both the social parasite, its eusocial host and any co-occurring related eusocial non-hosts (see **Box 1**) when each shifts from a solitary to a eusocial phase. In **Figure 2** the social environment is depicted by the yellow shaded area surrounding the three species spheres. Since all three species (social parasite, host and non-host) occupy similar societies, we predict that each will respond to a shift between solitary (nest founding/nest searching) and eusocial (established queens on host/non-hosts, and established parasite queens on host colonies) environments using similar changes in their molecular phenotypes. Conversely, if the social parasites evolve via Phenotype Shifting, we would not necessarily expect host and social parasite to respond to the same cue, using the same molecular processes. A test of this requires comparisons of transcription, protein synthesis and regulatory elements in the solitary and eusocial forms of the reproductive phenotypes in each species (**Figure 2**, area d).

Among the toolkit genes we analyzed, insulin growth factor (*IGF*) is a putative candidate gene for response to changes in the social environment. We observed up-regulation of *IGF* in social parasites brains when they shift from solitary to social living, whilst *IGF* shows no change in expression in the constant eusocial environments of the host (**Figure 3C**). In our *Polistes* test system (see **Box 2**), both host and parasite over-winter as newly mated queens, but the parasite overwinters alone whilst the host overwinters in socially active aggregations (Dapporto and Palagi, 2006; Cini and Dapporto, 2009). If social context influences gene expression, hosts should show no significant change in the expression of genes responsive to social environment since they remain in a social phase during the winter and summer. Conversely, social parasites shift between solitary (overwintering) and social phases, and expression of genes responsive to social environment should reflect this dynamic, as seen with *IGF* in our system (**Figure 3C**). Recent work in a free-living species of *Polistes* has highlighted the importance of social environment in gene expression (Toth et al., 2014). Further analyses will reveal whether host/non-host species in the solitary founding phase also show similar patterns of response to environment as found in the social parasite (**Figure 2**, area d). Other likely candidate genes for this response include juvenile hormone-binding proteins and hexamerins, which are up-regulated in gregarious/social forms relative to solitary phases in the migratory locust (Kang et al., 2004).

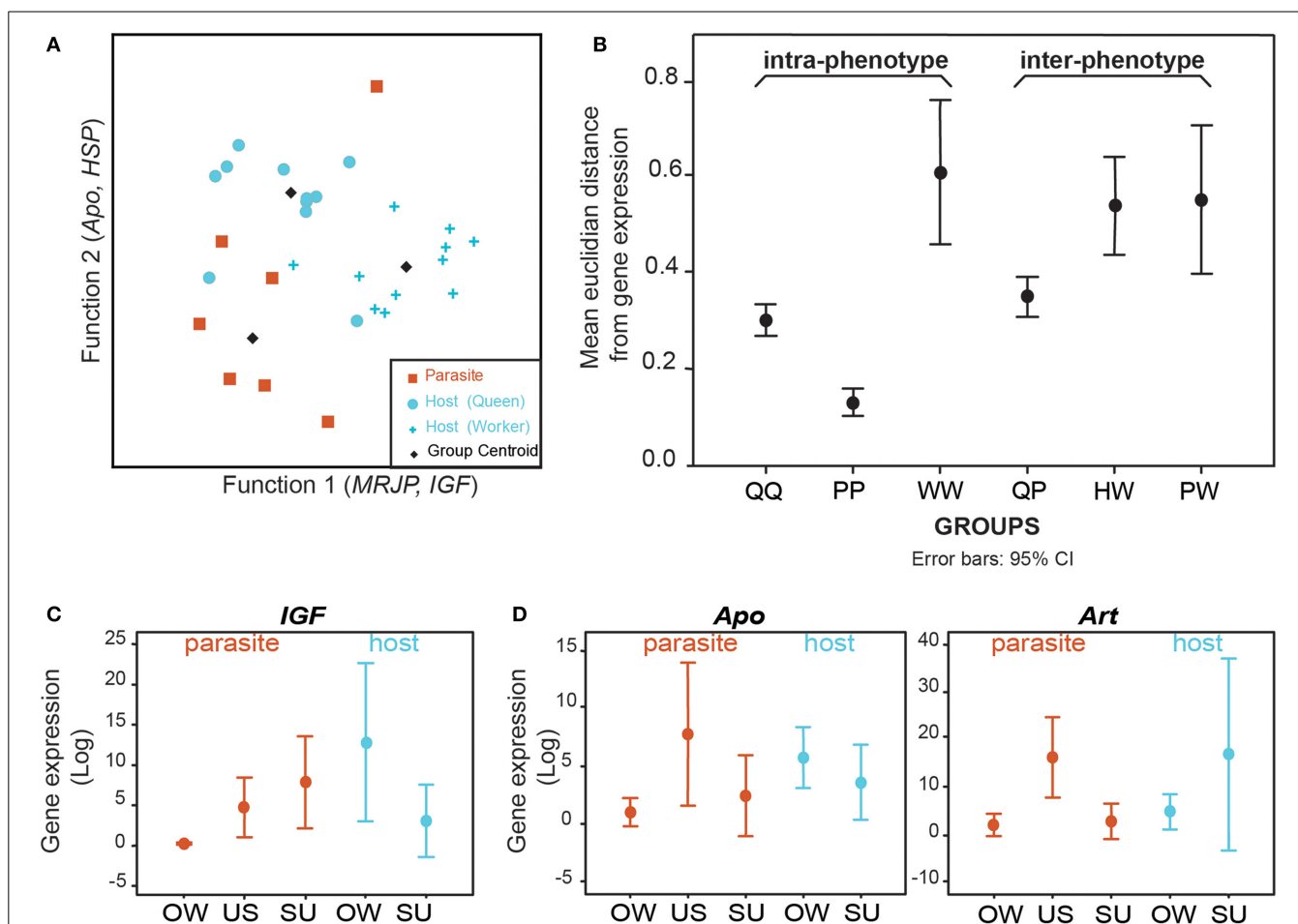


FIGURE 3 | Brain gene expression data from the social parasite *Polistes sulcifer* and its social host *Polistes dominula*. Comparison of expression levels for five “toolkit” genes that are differentially expressed among queens and workers in *Polistes* (chosen from: Sumner et al., 2006; Toth et al., 2007; Ferreira et al., 2013). *Arrestin* (*Art*) is expressed in response to light; *Apolipoprotein* (*Apo*) is involved in general metabolic processes and lipid transport; *Heat Shock Protein 70kDa* (*HSP*) is involved in response to heat stimulus; insulin growth factor (*IGF*) responds to nutrition; *Major Royal Jelly Protein* (*MRJP*) is a yellow protein associated with reproductive behaviors. We compared individual-level gene expression across three phenotypes: social parasites (P), host queens (Q) and host workers (W). **(A)** Discriminant analyses revealed three distinct clusters, corresponding to the 3 phenotypes. Function 1 closely correlates with gene expression of *MRJP* and *IGF* and discriminates between social parasites and workers while function 2 closely correlates with *Apo* and *HSP* and discriminates social parasites from queens. 79.3% of individuals grouped into non-overlapping clusters. Cross validation analyses correctly classified 69% of samples. **(B)** Euclidean distances in gene expression among phenotypes showing greater inter-phenotype differences than intra-phenotypes (*t*-test, $t = -2.114$, $df = 376$, $p = 0.035$, $n = 126$ vs. 252). Gene expression differences between social parasites and queens were greater than among social parasites (Mann Whitney test, $U = 233$, $p = 0.0005$, $n = 72$ vs. 15). **(C,D)** Gene

expression dynamics across the seasons (OW, overwinter; US, usurpation; SU, summer). **(C)** Changes in social environment experienced by the social parasites are accompanied by changes in *IGF* gene expression (within social parasites: Mann Whitney test, $U = 4.0$, $p = 0.0183$, $n = 8$ vs. 5; between species: Mann Whitney test, $U = 8.0$, $p = 0.1498$, $n = 7$ vs. 5). **(D)** *Apo* and *Art* are upregulated during usurpation compared to the pre and post usurpation periods (Kruskal Wallis test, *Apo*: $H = 8.525$, $p = 0.0141$; *Art*: $H = 8.842$, $p = 0.0120$). Expression levels of *Apo* and *Art* are significantly higher in usurping social parasites than in overwintering social parasites but no differences occur between overwintering and summer period [*Apo*: Mann Whitney post hoc pair wise comparisons US vs. OW $p = 0.0112$, US vs. SU, $p = 0.0230$; OW vs. SU $p = 0.341$, $n = 9$ (OW) vs. 5 (US) vs. 7 (SU), *Art*: Mann Whitney post hoc pair wise comparisons US vs. OW, $p = 0.00848$, US vs. SU, $p = 0.01421$; OW vs. SU $p = 0.9485$, $n = 8$ (OW) vs. 4 (US) vs. 6 (SU)]. No changes were observed in the expression levels for *Art* and *Apo* in the host species (Mann Whitney test, *Apo*: OW vs. SU Hosts $U = 12.0$, $p = 0.2343$, $n = 7$ vs. 6; *Art*: $U = 14.0$, $p = 0.366$, $n = 7$ vs. 6). No significant changes in *MRJP* and *HSP* gene expression dynamic across season were observed in parasites (Mann Whitney test, *MRJP*: $U = 4$, $p = 0.176$; *HSP*: $U = 13.0$, $p = 0.236$), or in the hosts who remain in a social environment throughout (Mann Whitney test, *MRJP*: $U = 8.0$, $p = 0.246$; *HSP*: $U = 6.0$, $p = 0.226$) (data not shown).

HYPOTHESIS 3: TRAIT LOSSES AND GAINS WILL BE REFLECTED AT THE MOLECULAR LEVEL

Phenotypically, social parasites exhibit a functional deletion of parental care traits (West-Eberhard, 2003). It is this

observation that forms the basis of the Phenotype Deletion Model (Figure 1C). At the molecular level, selection for the genes/gene functions associated with parental care will be relaxed as their expression no longer has any fitness consequence. Such genes

may be subject to rapid evolution, loss or other modifications (Hunt and Carrano, 2010; Hunt et al., 2011). This means genomic changes can be fixed rather than conditionally expressed (Van Dyken and Wade, 2010). Genes identified as important in parental care in host species therefore, are predicted to be lost (or not expressed) in social parasites. These traits can be easily recognized in the host (**Figure 2**, area b), thus providing a base-line of “absent” traits to compare with in the parasite (**Figure 2**, area a). Comparisons of molecular phenotypes of social parasites and their host (and non-host) workers are promising routes to defining the genes, regulatory processes and pathways involved in parental care in free-living species. Such analyses would provide a test of the Phenotype Deletion Model, and it also raises intriguing questions regarding the fate of the molecular processes involved in ancestral maternal care: does the parasite lose these genes/functions? In what sense are they “lost”; via their coding potential? What are the molecular processes that prevent these ancestral molecular processes from being expressed?

The evolution of social parasitism is accompanied by release from the evolutionary constraints experienced by a free-living species (Sumner et al., 2004b). This may allow the evolution of new/modified traits, not found in their free-living ancestor (West-Eberhard, 2003). For example, exaggerated morphological traits that enhance a social parasite’s fitness e.g., enlarged Dufours glands in Vespine social parasites (Jeanne, 1977); enlarged mandibles (Cervo, 1994; Cervo and Dani, 1996); specific usurpation behaviors in Polistine social parasites (Ortolani et al., 2008); reduced scopae and mouthparts in Allodapinae social parasites (Michener, 1970; Smith et al., 2013); specialized piercing mandibles in slave making ants (Buschinger, 2009). Other traits include mechanisms of effective manipulation and deception of the host, such as chemical insignificance to elude host recognition and chemical mimicry to integrate into the host colony (Lenoir et al., 2001; Bagnères and Lorenzi, 2010; Bruschini et al., 2010) or suppression of host queens/workers reproduction (e.g., Cervo and Lorenzi, 1996; Vergara et al., 2003). A key question is whether these novel traits arise through co-opted conserved molecular processes, or via *de novo* birth of novel genes and/or re-organization of existing genomic material.

Novel traits that have evolved in a range of different taxa have recently been associated with taxonomically restricted genes (Khalturin et al., 2008; Johnson and Tsutsui, 2011; Ferreira et al., 2013; Looso et al., 2013; Harpur et al., 2014), and this includes the eusocial Hymenoptera (Simola et al., 2013; Wissler et al., 2013; Sumner, 2014). We predict that social parasites will harbor a higher proportion of new genes, gene functions, or novel gene networks relative to their free-living eusocial hosts. Additionally, ancestral genes may be modified substantially in function through modulation of their expression patterns, regulatory roles or protein production (**Figure 2**, area a).

Analyses of gene expression dynamics in *Polistes* social parasite brains at the pre-usurpation (OW), usurping (U) and post-usurpation (SU) phases of their life cycle (see **Box 2**), revealed significant changes in the expression of Arrestin (*Art*) and Apolipoprotein (*Apo*) (**Figure 3D**). These genes are significantly

up-regulated during usurpation—a critical period in a social parasite’s life which, if not executed correctly during a narrow temporal window, could result in zero fitness (Turillazzi et al., 1990; Cervo and Turillazzi, 1996). During this phase, a novel behavior is exhibited—restlessness—(Ortolani et al., 2008), which is not found in the host (or non-host). No such variation of *Art* and *Apo* expression was detected in the host queens suggesting that these expression patterns are specific to the parasite’s novel behavior, potentially due to the acquisition of regulatory mechanisms that enhance gene expression variability. Unbiased genome-wide RNAseq analyses are required to determine whether putative novel genes are also involved in usurpation behaviors. New genes may be important drivers of phenotypic evolution (Chen et al., 2013). Studies on social parasites and their hosts will therefore help identify some such novel genes, and facilitate further exploration of the role of novel genes in phenotypic evolution. Such phenotype-led gene discovery is likely to be a rich, untapped resource.

HYPOTHESIS 4: RESISTANCE TO SOCIAL PARASITISM IN NON-HOSTS WILL BE REFLECTED AT THE MOLECULAR LEVEL

Comparison of social parasites, hosts and non-hosts has the potential to reveal the molecular processes associated with host response to parasitism (**Figure 2**, area b), for example in host worker rebellions to the presence of social parasites in *Protomognathus americanus* ants (Achenbach and Foitzik, 2009), and resistance to social parasitism as found in sympatric non-host sister species (**Figure 2** area c). In *Polistes dominula*, workers respond to parasite queens as if they were the host (mother) queen (Cervo et al., 1990; Cervo, 2006) suggesting that the parasite manipulates host workers successfully. However, recent work suggests that after several weeks of parasitism, workers are able to detect and respond to the parasite as they show some level of ovarian development, perhaps priming themselves for direct reproduction (Cini et al., 2014). Examining the molecular changes that take place in workers over the social parasite’s life cycle may reveal important insights into the dynamic interactions of host and social parasite genomes, in a similar way to pathogens and their hosts (Riddell et al., 2011; Dybdahl et al., 2014).

Non-host sister species that occur sympatrically to the host in parasitized populations are powerful models for studying the molecular basis of social parasite resistance. For example, the free-living leafcutter ant *Acromyrmex octospinosus* co-occurs with its sister species *Acromyrmex echinator*, and yet is resistant to parasitism by *Acromyrmex insinuator* (Sumner et al., 2004a; **Box 1A**); *Polistes nimphus* occurs alongside *P. dominula* and is resistant to invasion by *P. sulcifer* (Cervo, 2006; **Box 1B**). Phenotypically, there is no explanation for why co-occurring close relatives of hosts and social parasite are not also vulnerable to social-parasitism. We hypothesize that there will be key differences in the transcriptional and/or regulatory processes of hosts and non-hosts, which may confer resistance to non-hosts (**Figure 2**, area c). These may include novel processes (or novel usage of conserved genes) that have evolved in the non-host since speciation. Functional genomics (e.g., RNAi, cross-species expression experiments) provide powerful tools to test candidate genes or regulatory elements involved resistance.

CONCLUSIONS AND FUTURE PERSPECTIVES

Comparative genomic analyses of obligate social parasites with their eusocial hosts and non-hosts are powerful approaches to studying losses and gains in phenotypic evolution. These analyses promise important insights into how genomes give rise to phenotypic diversity. We outline two scenarios for the evolution of social parasites from their eusocial ancestors. The scant data available to date suggest that the social parasite phenotype is distinct from their eusocial ancestor counter-part (i.e., eusocial queens). Social parasites therefore may not evolve through simple “deletion” (silencing) of the worker phenotype and its associated molecular functions (West-Eberhard, 2005). Based on recent empirical findings on the molecular basis of phenotypic evolution in other organisms, we predict that the evolution of new genes as well as the re-use of old ones will be important in the generation of the novel traits that characterize this new phenotype. We also predict that the full social parasite phenotype (defined as a combined consideration of the behavioral and molecular phenotype, Nachtomy et al., 2007) will be more complex than perceived from classical behavioral studies. Crucially, social parasites may retain the machinery for detecting and responding to the environment, just like their social ancestor and their free-living social hosts. The molecular processes associated with response to the environment, rather than behavior, are likely to be conserved (e.g., toolkit genes).

Our model and predictions are preliminary, but are relevant more widely to non-hymenopteran social parasites, as social parasitism of parental care has evolved multiple times in different taxa of the animal kingdom, e.g., birds (Davies, 2000); lycaenid butterflies (Fiedler, 2006); freshwater fishes (Baba et al., 1990). In each case, the social parasite is a highly specialized species that has lost the traits associated with caring for its own young, and evolved new traits that enable it to successfully insinuate its young into the home of its chosen host. More generally, our framework may also be relevant to phenotypic evolution in non-social parasites that are closely related to their hosts, such as in fungi, red algae and mistletoe, cynipids wasps, gall inducing aphids (West-Eberhard, 2003) and parasitoids (e.g., *Nasonia*, Werren et al., 2010).

ACKNOWLEDGMENTS

This work was funded by Fondation Fyssen and Accademia dei Lincei-Royal Society grants (AC), Università di Firenze (RC), Wellcome Trust (SP and ASP), RCUK & NERC (NE/G000638/1) (SS, SP), L'Oreal for Women in Science Fellowship (SS, GB).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2015.00032/abstract>

REFERENCES

- Achenbach, A., and Foitzik, S. (2009). First evidence for slave rebellion: enslaved ant workers systematically kill the brood of their social parasite *Protomognathus americanus*. *Evolution* 63, 1068–1075. doi: 10.1111/j.1558-5646.2009.00591.x
- Alford, D. V. (1975). *Bumblebees*. London: Davis Poynter.
- Arendt, J., and Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* 23, 26–32. doi: 10.1016/j.tree.2007.09.011
- Aubin-Horth, N., and Renn, S. C. P. (2009). Genomic reaction norms: using integrative biology to understand molecular mechanisms of phenotypic plasticity. *Mol. Ecol.* 18, 3763–3780. doi: 10.1111/j.1365-294X.2009.04313.x
- Baba, R., Nagata, Y., and Yamagishi, S. (1990). Brood parasitism and egg robbing among three freshwater fish. *Anim. Behav.* 40, 776–778. doi: 10.1016/S0003-3472(05)80707-9
- Bagnères, A. G., and Lorenzi, M. C. (2010). “Chemical deception/mimicry using cuticular hydrocarbons,” in *Insect Hydrocarbons: Biology, Biochemistry and Chemical Ecology*, eds G. J. Blomquist and A.-G. Bagnères (Cambridge, MA: Cambridge University Press), 282–324.
- Begna, D., Han, B., Feng, M., Fang, Y., and Li, J. (2012). Differential expressions of nuclear proteomes between honeybee (*Apis mellifera* L.) Queen and worker larvae: a deep insight into caste pathway decisions. *J. Proteome Res.* 11, 1317–1329. doi: 10.1021/pr200974a
- Beldade, P., Mateus, A. R. A., and Keller, R. A. (2011). Evolution and molecular mechanisms of adaptive developmental plasticity. *Mol. Ecol.* 20, 1347–1363. doi: 10.1111/j.1365-294X.2011.05016.x
- Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., et al. (2012). Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* 22, 1–10. doi: 10.1016/j.cub.2012.07.042
- Brisson, J. A., and Stern, D. L. (2006). The pea aphid, *Acyrtosiphon pisum*: an emerging genomic model system for ecological, developmental and evolutionary studies. *Bioessays* 28, 747–755. doi: 10.1002/bies.20436
- Bruschini, C., Cervo, R., and Turillazzi, S. (2010). Pheromones in social wasps. *Vit. Horm. Phero* 83, 447–492. doi: 10.1016/S0083-6729(10)83019-5
- Burton, J., Bogitsh, C. E. C., and Oeltmann, T. N. (2012). *Human Parasitology*. Academic Press.
- Buschinger, A. (2009). Social parasitism among ants: a review (Hymenoptera: Formicidae). *Myrmec. News* 12, 219–235.
- Cameron, S. A., Hines, H. M., and Williams, P. H. (2007). A comprehensive phylogeny of the bumble bees (*Bombus*). *Biol. J. Linnean Soc.* 91, 161–188. doi: 10.1111/j.1095-8312.2007.00784.x
- Carpenter, J. M., and Perera, E. P. (2006). Phylogenetic relationships among yellowjackets and the evolution of social parasitism (Hymenoptera: Vespidae, Vespinae). *Am. Mus. Novit.* 3507, 1–19. doi: 10.1206/0003-0082(2006)3507[1:PRAYAT]2.0.CO;2
- Cervo, R. (1994). Morphological adaptations to the parasitic life in *Polistes sulcifer* and *P. atrimandibularis* (Hymenoptera Vespidae). *Ethol. Ecol. Evol.* 6, 61–66. doi: 10.1080/03949370.1994.10721975
- Cervo, R. (2006). *Polistes* wasps and their social parasites: an overview. *Ann. Zool. Fenn* 43, 531–549.
- Cervo, R., Bertocci, F., and Turillazzi, S. (1996). Olfactory cues in host nest detection by the obligate social parasite *Polistes sulcifer* (Hymenoptera, Vespidae). *Behav. Processes* 36, 213–218. doi: 10.1016/0376-6357(95)00030-5
- Cervo, R., and Dani, F. R. (1996). “Social parasitism and its evolution in *Polistes*,” in *Natural History and Evolution of Paper-Wasps*, eds S. Turillazzi and M. J. West-Eberhard (New York, NY: Oxford University Press), 98–112.
- Cervo, R., and Lorenzi, M. C. (1996). Inhibition of host queen reproductive capacity by the obligate social parasite *Polistes atrimandibularis* (Hymenoptera, Vespidae). *Ethology* 102, 1042–1047. doi: 10.1111/j.1439-0310.1996.tb01180.x
- Cervo, R., Lorenzi, M. C., and Turillazzi, S. (1990). Different strategies of host nest invasion in two species of *Sulcopolistes* (Hymenoptera, Vespidae). *Ethol. Ecol. Evol.* 3, 302–303. doi: 10.1080/08927014.1990.9525432
- Cervo, R., and Turillazzi, S. (1996). Host nest preference and nest choice in the cuckoo paper wasp *Polistes sulcifer* (Hymenoptera: Vespidae). *J. Insect Behav.* 9, 297–306. doi: 10.1007/BF02213872
- Chen, S., Krinsky, B. H., and Long, M. (2013). New genes as drivers of phenotypic evolution. *Nat. Rev. Genet.* 14, 645–660. doi: 10.1038/nrg3521
- Choudhary, M., Strassmann, J. E., Queller, D. C., Turillazzi, S., and Cervo, R. (1994). Social parasites in Polistine wasps are monophyletic: implications for sympatric speciation. *Proc. R. Soc. Lond. B* 257, 31–35. doi: 10.1098/rspb.1994.0090
- Cini, A., Bruschini, C., Poggi, L., and Cervo, R. (2011b). Fight or fool? physical strength, instead of sensory deception, matters in host nest invasion by a wasp social parasite. *Anim. Behav.* 81, 1139–1145. doi: 10.1016/j.anbehav.2011.02.017
- Cini, A., Bruschini, C., Signorotti, L., Pontieri, L., Turillazzi, S., and Cervo, R. (2011a). The chemical basis of host nest detection and chemical integration in a cuckoo paper wasp. *J. Exp. Biol.* 214, 3698–3703. doi: 10.1242/jeb.059519

- Cini, A., and Dapporto, L. (2009). Autumnal helpers of *Polistes dominulus* represent a distinct behavioural phenotype. *Ann. Zool. Fenn.* 46, 423–430. doi: 10.5735/086.046.0603
- Cini, A., Nieri, R., Dapporto, L., Monnin, T., and Cervo, R. (2014). Almost royal: incomplete suppression of host worker ovarian development by a social parasite wasp. *Behav. Ecol. Sociobiol.* 68, 467–475. doi: 10.1007/s00265-013-1661-z
- Dapporto, L., Cervo, R., Sledge, M. F., and Turillazzi, S. (2004). Rank integration in dominance hierarchies of host colonies by the paper wasp social parasite *Polistes sulcifer* (Hymenoptera, Vespidae). *J. Insect Physiol.* 50, 217–223. doi: 10.1016/j.jinsphys.2003.11.012
- Dapporto, L., and Palagi, E. (2006). Wasps in the shadow: looking at the pre-hibernating clusters of *Polistes dominulus*. *Ann. Zool. Fenn.* 43, 583–594.
- Daugherty, T. H. F., Toth, A. L., and Robinson, G. E. (2011). Nutrition and division of labor: effects on foraging and brain gene expression in the paper wasp *Polistes metricus*. *Mol. Ecol.* 20, 5337–5347. doi: 10.1111/j.1365-294X.2011.05344.x
- Davies, N. B. (2000). *Cuckoos, Cowbirds and Other Cheats*. London: T. & A. D. Poyser.
- Dawkins, R. (1982). *The Extended Phenotype*. Oxford: Oxford University Press.
- De Visser, J. A. G. M., and Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genetics* 15, 480–490. doi: 10.1038/nrg3744
- Dybdahl, M. F., Jenkins, C. E., and Nuismer, S. L. (2014). Identifying the molecular basis of host-parasite coevolution: merging models and mechanisms. *Am. Nat.* 184, 1–13. doi: 10.1086/676591
- Evans, J. D., and Wheeler, D. E. (2001). Gene expression and the evolution of insect polyphenisms. *Bioessays* 23, 62–68. doi: 10.1002/1521-1878(200101)23:1<62::AID-BIES1008>3.3.CO;2-Z
- Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., et al. (2014). Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Hum. Mol. Genet.* 23, 5866–5878. doi: 10.1093/hmg/ddu309
- Feldmeyer, B., Elsnér, D., and Foitzik, S. (2014). Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Mol. Ecol.* 23, 151–161. doi: 10.1111/mec.12490
- Fernald, R. D. (2004). Evolving eyes. *Intern. J. Dev. Biol.* 48, 701–705. doi: 10.1387/ijdb.041888rf
- Ferreira, P. G., Patalano, S., Chauhan, R., Ffrench-Constant, R., Gabaldón, T., Guigó, R., et al. (2013). Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol.* 14:R20. doi: 10.1186/gb-2013-14-2-r20
- Fiedler, K. (2006). Ant-associates of Palaearctic lycaenid butterfly larvae (Hymenoptera: Formicidae; Lepidoptera: Lycaenidae)—a review. *Myrmec. Nachrichten* 9, 77–87.
- Fischman, B. J., Woodard, S. H., and Robinson, G. E. (2011). Molecular evolutionary analyses of insect societies. *Proc. Natl. Acad. Sci.* 108 10847–10854. doi: 10.1073/pnas.1100301108
- Fusco, G., and Minelli, A. (2010). Phenotypic plasticity in development and evolution: facts and concepts. Introduction. *Phil. Trans. R. Soc. London B* 365, 547–556. doi: 10.1098/rstb.2009.0267
- Gibbs, J., Albert, J., and Packer, L. (2012). Dual origins of social parasitism in North American *Dialictus* (Hymenoptera: Halictidae) confirmed using a phylogenetic approach. *Cladistics* 28, 195–207. doi: 10.1111/j.1096-0031.2011.00373.x
- Greenberg, J. K., Xia, J., Zhou, X., Thatcher, S. R., Gu, X., Ament, S. A., et al. (2012). Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes Brain Behav.* 11, 660–670. doi: 10.1111/j.1601-183X.2012.00782.x
- Hamilton, W. D. (1980). Sex versus non-sex versus parasite. *Oikos* 35, 282–290. doi: 10.2307/3544435
- Harpur, B. A., Kent, C. F., Molodtsova, D., Lebon, J. M., Alqarni, A. S., Owayss, A. A., et al. (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. doi: 10.1073/pnas.1315506111
- Hines, H. M., and Cameron, S. A. (2010). The phylogenetic position of the bumble bee inquiline *Bombus inexpectatus* and implications for the evolution of social parasitism. *Insectes Soc.* 57, 379–383. doi: 10.1007/s00040-010-0094-1
- Huang, M. H., and Dornhaus, A. (2008). A meta-analysis of ant social parasitism: host characteristics of different parasitism types and a test of Emery's rule. *Ecol. Entomol.* 33, 589–596. doi: 10.1111/j.1365-2311.2008.01005.x
- Hughes, A. L. (2012). Evolution of adaptive phenotypic traits without positive Darwinian selection. *Heredity* 108, 347–353. doi: 10.1038/hdy.2011.97
- Hunt, B. G., Ometto, L., Wurm, Y., Shoemaker, D., Soojin, V. Y., Keller, L., et al. (2011). Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15936–15941. doi: 10.1073/pnas.1104825108
- Hunt, G., and Carrano, M. T. (2010). “Models and methods for analyzing phenotypic evolution in lineages and clades,” in *Quantitative Methods in Paleobiology*, eds J. Alroy and G. Hunt (Boulder, CO: The Paleontological Society), 245–269.
- Jeanne, R. L. (1977). Behavior of the obligate social parasite *Vespa arctica* (Hymenoptera: Vespidae). *J. Kansas Ent. Soc.* 50, 541–557.
- Johnson, B. R., and Tsutsui, N. D. (2011). Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* 12:164. doi: 10.1186/1471-2164-12-164
- Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20, 1313–1326. doi: 10.1101/gr.101386.109
- Kang, L., Chen, X., Zhou, Y., Liu, B., Zheng, W., Li, R., et al. (2004). The analysis of large-scale gene expression correlated to the phase changes of the migratory locust. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17611–17615. doi: 10.1073/pnas.0407753101
- Khalturin, K., Anton-Erxleben, F., Sassmann, S., Wittlieb, J., Hemmrich, G., and Bosch, T. C. (2008). A novel gene family controls species-specific morphological traits in *Hydra*. *PLoS Biol.* 6:e278. doi: 10.1371/journal.pbio.0060278
- Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319, 1827–1830. doi: 10.1126/science.1153069
- Lee, P. N., Callaerts, P., and Couet, H. G., De (2003). Cephalopod Hox genes and the origin of morphological novelties. *Nature* 424, 1061–1065. doi: 10.1038/nature01872
- Lenoir, A., d'Ettorre, P., Errard, C., and Hefetz, A. (2001). Chemical ecology and social parasitism in ants. *Ann. Rev. Entomol.* 46, 573–599. doi: 10.1146/annurev.ento.46.1.573
- Looso, M., Preussner, J., Sousounis, K., Bruckskotten, M., Michel, C. S., Lignelli, E., et al. (2013). A de novo assembly of the newt transcriptome combined with proteomic validation identifies new protein families expressed during tissue regeneration. *Genome Biol.* 14:R16. doi: 10.1186/gb-2013-14-2-r16
- Lowe, R. M., Ward, S. A., and Crozier, R. H. (2002). The evolution of parasites from their hosts: intra- and interspecific parasitism and Emery's rule. *Proc. R. Soc. Lond. B* 269, 1301–1305. doi: 10.1098/rspb.2002.2008
- Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* 8:e1000506. doi: 10.1371/journal.pbio.1000506
- Michener, C. D. (1970). Social parasites among African allopapine bees. *Zool. J. Linn. Soc.* 49, 199–215. doi: 10.1111/j.1096-3642.1970.tb00736.x
- Moczek, A. P. (2009). “Developmental plasticity and the origins of diversity: a case study on horned beetles,” in *Phenotypic Plasticity in Insects: Mechanisms and Consequences*, eds T. N. Ananthakrishnan and D. Whitman (Plymouth: Science Publishers), 81–134.
- Moczek, A. P. (2010). Phenotypic plasticity and diversity in insects. *Phil. Trans. R. Soc. London B* 365, 593–603. doi: 10.1098/rstb.2009.0263
- Nachtomy, O., Shavit, A., and Yakhini, Z. (2007). Gene expression and the concept of the phenotype. *Stud. Hist. Phil. Biol. Biomed. Sci.* 38, 238–254. doi: 10.1016/j.shpsc.2006.12.014
- Nijhout, H. F. (2003). Development and evolution of adaptive polyphenisms. *Evol. Dev.* 5, 9–18. doi: 10.1046/j.1525-142X.2003.03003.x
- Ortolani, I., Turillazzi, S., and Cervo, R. (2008). Spring usurpation restlessness: a wasp social parasite adapts its seasonal activity to the host cycle. *Ethology* 114, 782–788. doi: 10.1111/j.1439-0310.2008.01525.x
- Page, R. E., and Amdam, G. V. (2007). The making of a social insect: developmental architectures of social design. *Bioessays* 29, 334–343. doi: 10.1002/bies.20549
- Pardi, L. (1946). La ‘Dominazione’ e il ciclo ovario annuale in polistes gallicus (L.). Ricerche sui Polistini VII. *Boll. Ist. Entom.* 25–84.
- Pardi, L. (1996). “Polistes: analysis of a society,” in *Natural History and Evolution of Paper-Wasps*, eds S. Turillazzi and M. J. West-Eberhard (New York, NY: Oxford University Press), 98–112.
- Patalano, S., Hore, T. A., Reik, W., and Sumner, S. (2012). Shifting behaviour: epigenetic reprogramming in eusocial insects. *Curr. Opin. Cell Biol.* 24, 367–373. doi: 10.1016/j.ceb.2012.02.005

- Pavey, S. A., Collin, H., Nosil, P., and Rogers, S. M. (2010). The role of gene expression in ecological speciation. *Ann. NY Acad. Sci.* 1206, 110–129. doi: 10.1111/j.1749-6632.2010.05765.x
- Pfennig, D. W., Wund, M. A., Snell-Rood, E. C., Cruickshank, T., Schlichting, C. D., and Moczek, A. P. (2010). Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol. Evol.* 25, 459–467. doi: 10.1016/j.tree.2010.05.006
- Riddell, C. E., Sumner, S., Adams, S., and Mallon, E. B. (2011). Pathways to immunity: temporal dynamics of the bumblebee (*Bombus terrestris*) immune response against a trypanosomal gut parasite. *Insect Molec. Biol.* 20, 529–540. doi: 10.1111/j.1365-2583.2011.01084.x
- Savolainen, R., and Vepsäläinen, K. (2003). Sympatric speciation through intraspecific social parasitism. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7169–7174. doi: 10.1073/pnas.1036825100
- Schlichting, C. D., and Wund, M. A. (2014). Phenotypic plasticity and epigenetic marking: an assessment of evidence for genetic accommodation. *Evolution* 68, 656–672. doi: 10.1111/evo.12348
- Schwander, T., and Leimar, O. (2011). Genes as leaders and followers in evolution. *Trends Ecol. Evol.* 26, 143–151. doi: 10.1016/j.tree.2010.12.010
- Shubin, N., Tabin, C., and Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature* 457, 818–823. doi: 10.1038/nature07891
- Simola, D. F., Wissler, L., Donahue, G., Waterhouse, R. M., Helmkamp, M., Roux, J., et al. (2013). Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res.* 23, 1235–1247. doi: 10.1101/gr.155408.113
- Simpson, S. J., Sword, G. A., and Lo, N. (2011). Polyphenism in insects. *Curr. Biol.* 21, R738–R749. doi: 10.1016/j.cub.2011.06.006
- Sledge, M. F., Dani, F. R., Cervo, R., Dapporto, L., and Turillazzi, S. (2001). Recognition of social parasites as nest-mates: adoption of colony-specific host cuticular odours by the paper wasp parasite *Polistes sulcifer*. *Proc. R. Soc. Lond. B* 268, 2253–2260. doi: 10.1098/rspb.2001.1799
- Smith, C. R., Toth, A. L., Suarez, A. V., and Robinson, G. E. (2008). Genetic and genomic analyses of the division of labour in insect societies. *Nat. Rev. Genet.* 9, 735–748. doi: 10.1038/nrg2429
- Smith, J. A., Chenoweth, L. B., Tierney, S. M., and Schwarz, M. P. (2013). Repeated origins of social parasitism in allopapine bees indicate that the weak form of Emery's rule is widespread, yet sympatric speciation remains highly problematic. *Biol. J. Linn. Soc.* 109, 320–331. doi: 10.1111/bij.12043
- Stern, D. L. (2013). The genetic causes of convergent evolution. *Nat. Rev. Genet.* 14, 751–764. doi: 10.1038/nrg3483
- Sumner, S. (2006). Determining the molecular basis of sociality in insects: progress, prospects and potential in sociogenomics. *Ann. Zool. Fenn.* 43, 423–442.
- Sumner, S. (2014). The importance of genomic novelty in social evolution. *Mol. Ecol.* 23, 26–28. doi: 10.1111/mec.12580
- Sumner, S., Aanen, D. K., Delabie, J., and Boomsma, J. J. (2004a). The evolution of social parasitism in *Acromyrmex* leaf-cutting ants: a test of Emery's rule. *Insectes Soc.* 51, 37–42. doi: 10.1007/s00040-003-0723-z
- Sumner, S., Hughes, W. O., Pedersen, J. S., and Boomsma, J. J. (2004b). Ant parasite queens revert to mating singly. *Nature* 428, 35–36. doi: 10.1038/428035a
- Sumner, S., Pereboom, J. J., and Jordan, W. C. (2006). Differential gene expression and phenotypic plasticity in behavioural castes of the primitively eusocial wasp, *Polistes canadensis*. *Proc. R. Soc. Lond. B* 273, 19–26. doi: 10.1098/rspb.2005.3291
- Tautz, D., Ellegren, H., and Weigel, D. (2010). Next generation molecular ecology. *Mol. Ecol.* 19, 1–3. doi: 10.1111/j.1365-294X.2009.04489.x
- Tierney, S. M., Smith, J. A., Chenoweth, L., and Schwarz, M. P. (2008). Phylogenetics of allopapine bees: a review of social evolution, parasitism and biogeography. *Apidologie* 39, 3–15. doi: 10.1051/apido:2007045
- Toth, A. L., Bilof, K. B. J., Henshaw, M. T., Hunt, J. H., and Robinson, G. E. (2009). Lipid stores, ovary development, and brain gene expression in *Polistes metricus* females. *Insectes Soc.* 56, 77–84. doi: 10.1007/s00040-008-1041-2
- Toth, A. L., and Robinson, G. E. (2007). Evo-devo and the evolution of social behavior. *Trends Genet.* 23, 334–341. doi: 10.1016/j.tig.2007.05.001
- Toth, A. L., Tooker, J. F., and Radhakrishnan, S. (2014). Shared genes related to aggression, rather than chemical communication, are associated with reproductive dominance in paper wasps (*Polistes metricus*). *BMC Genomics* 15:75. doi: 10.1186/1471-2164-15-75
- Toth, A. L., Varala, K., Henshaw, M. T., Rodriguez-Zas, S. L., Hudson, M. E., and Robinson, G. E. (2010). Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc. R. Soc. Lond. B* 277, 2139–2148. doi: 10.1098/rspb.2010.0090
- Toth, A. L., Varala, K., Newman, T. C., Miguez, F. E., Hutchison, S. K., Willoughby, D. A., et al. (2007). Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science* 318, 441–444. doi: 10.1126/science.1146647
- Turillazzi, S., Cervo, R., and Cavallari, I. (1990). Invasion of the nest of *Polistes dominulus* by the social parasite *Sulcopolistes sulcifer* (Hymenoptera, Vespidae). *Ethology* 84, 47–59. doi: 10.1111/j.1439-0310.1990.tb00784.x
- Turillazzi, S., Sledge, M. F., Dani, F. R., Cervo, R., Massolo, A., and Fondelli, L. (2000). Social hackers: integration in the host chemical recognition system by a paper wasp social parasite. *Naturwissen* 87, 172–176. doi: 10.1007/s001140050697
- Valcu, C. M., and Kempenaers, B. (2014). Proteomics in behavioral ecology. *Behav. Ecol.* 26, 1–15. doi: 10.1093/beheco/aru096
- Van Dyken, J. D., and Wade, M. J. (2010). The genetic signature of conditional expression. *Genetics* 184, 557–570. doi: 10.1534/genetics.109.110163
- Vergara, C. H., Schroder, S., Almanza, M. T., and Wittmann, D. (2003). Suppression of ovarian development of *Bombus terrestris* workers by *B. terrestris* queens, *Psithyrus vestalis* and *Psithyrus bohemicus* females. *Apidol* 34, 563–568. doi: 10.1051/apido:2003056
- Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565. doi: 10.1038/150563a0
- Werren, J. H., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., and Colbourne, J. K. (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327, 343–348. doi: 10.1126/science.1178028
- West-Eberhard, M. J. (1989). Phenotypic plasticity and the origins of diversity. *Annu. Rev. Ecol. Syst.* 249–278. doi: 10.1146/annurev.es.20.110189.001341
- West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. New York, NY: Oxford University Press.
- West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proc. Natl. Acad. Sci. U.S.A.* 102 (Suppl.), 6543–6549. doi: 10.1073/pnas.0501844102
- Wheeler, D. E. (1986). Developmental and physiological determinants of caste in social Hymenoptera: evolutionary implications. *Am. Nat.* 128, 13–34. doi: 10.1086/284536
- Wissler, L., Gadau, J., Simola, D. F., Helmkamp, M., and Bornberg-Bauer, E. (2013). Mechanisms and dynamics of orphan gene emergence in insect genomes. *Genome Biol. Evol.* 5, 439–455. doi: 10.1093/gbe/evt009
- Woodard, S. H., Fischman, B. J., Venkat, A., Hudson, M. E., Varala, K., Cameron, S. A., et al. (2011). Genes involved in convergent evolution of eusociality in bees. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7472–7477. doi: 10.1073/pnas.1103457108
- Yan, H., Simola, D. F., Bonasio, R., Liebig, J., Berger, S. L., and Reinberg, D. (2014). Eusocial insects as emerging models for behavioural epigenetics. *Nat. Rev. Genet.* 15, 677–688. doi: 10.1038/nrg3787

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 September 2014; accepted: 23 January 2015; published online: 18 February 2015.

Citation: Cini A, Patalano S, Segonds-Pichon A, Busby GBJ, Cervo R and Sumner S (2015) Social parasitism and the molecular basis of phenotypic evolution. *Front. Genet.* 6:32. doi: 10.3389/fgene.2015.00032

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2015 Cini, Patalano, Segonds-Pichon, Busby, Cervo and Sumner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Function and evolution of microRNAs in eusocial Hymenoptera

Eirik Søvik¹, Guy Bloch² and Yehuda Ben-Shahar^{1*}

¹ Department of Biology, Washington University in St. Louis, St. Louis, MO, USA, ² Department of Ecology, Evolution, and Behavior, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

OPEN ACCESS

Edited by:

Greg J. Hunt,
Purdue University, USA

Reviewed by:

Darren E. Hagen,
University of Missouri, USA
Roberto Bonasio,
University of Pennsylvania, USA

*Correspondence:

Yehuda Ben-Shahar,
Department of Biology, Washington
University in St. Louis, Campus
Box 1137, One Brookings Drive,
St. Louis, MO 63130, USA
benshahary@wustl.edu

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 23 March 2015

Accepted: 14 May 2015

Published: 27 May 2015

Citation:

Søvik E, Bloch G and Ben-Shahar Y
(2015) Function and evolution
of microRNAs in eusocial
Hymenoptera.
Front. Genet. 6:193.
doi: 10.3389/fgene.2015.00193

The emergence of eusociality (“true sociality”) in several insect lineages represents one of the most successful evolutionary adaptations in the animal kingdom in terms of species richness and global biomass. In contrast to solitary insects, eusocial insects evolved a set of unique behavioral and physiological traits such as reproductive division of labor and cooperative brood care, which likely played a major role in their ecological success. The molecular mechanisms that support the social regulation of behavior in eusocial insects, and their evolution, are mostly unknown. The recent whole-genome sequencing of several eusocial insect species set the stage for deciphering the molecular and genetic bases of eusociality, and the possible evolutionary modifications that led to it. Studies of mRNA expression patterns in the brains of diverse eusocial insect species have indicated that specific social behavioral states of individual workers and queens are often associated with particular tissue-specific transcriptional profiles. Here, we discuss recent findings that highlight the role of non-coding microRNAs (miRNAs) in modulating traits associated with reproductive and behavioral divisions of labor in eusocial insects. We provide bioinformatic and phylogenetic data, which suggest that some Hymenoptera-specific miRNA may have contributed to the evolution of traits important for the evolution of eusociality in this group.

Keywords: miRNA, Aculeata, Hymenoptera, eusociality, non-coding RNAs

Introduction

Most insect species are solitary, and behavioral interactions with conspecifics are primarily restricted to reproductive behaviors such as male–female courtship and male–male competition. This is in sharp contrast to social insects, where groups of genetically related individuals often live together in a colonial lifestyle. The size and stability of these colonies vary from a few individuals sharing a nest for a short period of time, to large perennial colonies composed of thousands of individuals (Hölldobler and Wilson, 2009). The most advanced form of animal social organization is termed “eusociality” (Crespi and Yanega, 1995), marked by the presence of sterile workers that often forgo own reproduction in order to support the reproduction of other colony members. Although eusociality is relatively rare in most taxonomic animal lineages, eusocial species have been immensely successful. Current projections estimate eusocial insects to represent the largest proportion of the global animal biomass (Hölldobler and Wilson, 2009). Although the reasons for this remarkable success are not well-understood, it is commonly assumed that the social lifestyle of these animals must have played a major role in their current ecological dominance (Wilson, 1990). For example, it is thought that specialization in task performance (division of labor) amongst eusocial workers enables colonies to maximize the exploitation of their environment. In contrast,

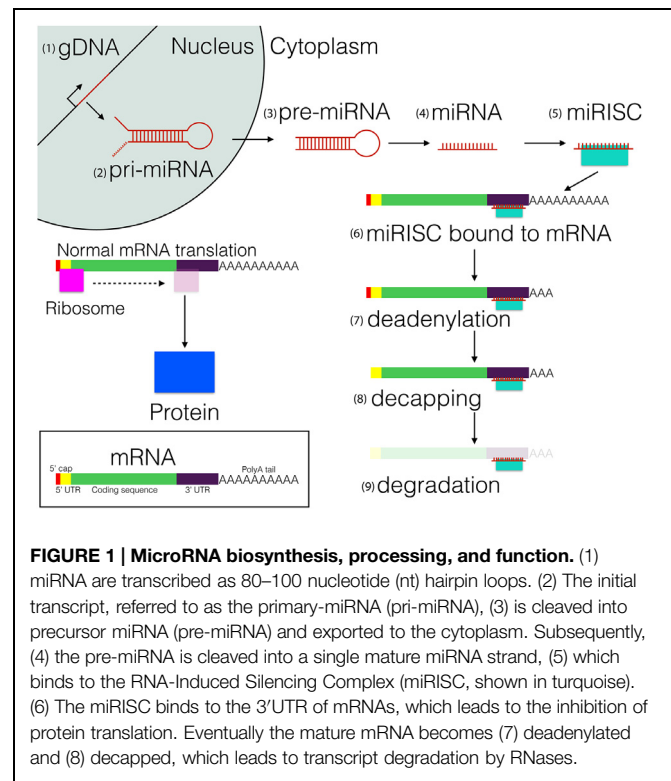
solitary insects have to multitask independent activities, including foraging and brood rearing (Wilson, 1985).

The recent sequencing of genomes from diverse social and solitary Hymenoptera clades offers a unique opportunity for identifying genome-level molecular events that may have supported the emergence of specific traits associated with the evolution of eusociality (“eusocial traits”). The ability to compare whole-genome sequences, gene expression patterns, and other molecular properties of species with diverse forms of social lifestyles, has generated novel mechanistic and evolutionary insights into these complex behaviors. This approach has been used most successfully in studies of the division of labors in worker tasks (Smith et al., 2008) and reproduction (Schwander et al., 2010), both of which are hallmarks of eusociality (Wilson, 1985). To date, the efforts to decipher the evolution of eusocial traits, and the mechanisms that support them, have focused on protein-coding genes (Keller and Ross, 1998; Page and Amdam, 2007; Fischman et al., 2011; Woodard et al., 2011). In contrast, how non-coding regulatory RNAs may have played a role in the evolution of eusociality is understudied. Here, we examine the emerging role of an important class of small, non-coding RNAs, which are collectively referred to as “microRNAs” (miRNAs), in regulating social behaviors. We discuss their possible role in regulating eusocial traits in social Hymenoptera at the developmental, physiological, and evolutionary time scales.

miRNAs: History and Background

During the early days of the molecular biology revolution, the majority of research on gene regulation was limited to transcriptional mechanisms of protein coding genes as originally defined by the “Central dogma of molecular biology” (Crick, 1970). However, the discovery of the regulatory function of non-coding RNAs indicated that the early views on gene regulation and their associated phenotypic outcomes, were oversimplified and required major revisions to the dogma. We now know that in addition to transcriptional regulation (Lee and Young, 2000; Yan et al., 2015), gene functions are also regulated by factors such as post-transcriptional RNA editing (Gott and Emeson, 2000), mRNA splicing (Breitbart et al., 1987), RNA degradation (Bushati and Cohen, 2007), and diverse post-translational protein modifications (Braakman and Bulleid, 2011). More recently, regulatory non-coding RNAs have also emerged as important factors that regulate phenotypic variation via diverse molecular mechanisms (Qureshi and Mehler, 2012; Bonasio and Shiekhattar, 2014).

miRNAs are short (18–24 nucleotides) non-coding RNAs, which in animals seem to act primarily by repressing protein translation via interaction with the 3'UTR of mRNAs (Figure 1). miRNAs were first discovered in the nematode *Caenorhabditis elegans*, where the miRNA *cel-lin-4* was shown to be necessary for the temporal timing of key developmental events (Lee et al., 1993). Because of their short length and the nature of their molecular interaction with mRNA targets, it has been hypothesized that a single miRNA can potentially regulate the function of multiple protein-coding genes (Bartel, 2009), and



thus act as a pleiotropic genetic factor (Bartel, 2004). It is estimated that between one and two thirds of mRNAs encoded by animal genomes are regulated by miRNAs (Berezikov, 2011). As a result, it is likely that miRNAs play some roles in the regulation of most biological processes in animal cells (Bushati and Cohen, 2007).

miRNAs in Development and Function of Nervous Systems

Various miRNAs have been implicated in neuronal development (Alvarez-Garcia and Miska, 2005; Wienholds and Plasterk, 2005). There is evidence that miRNAs play important roles in fine tuning the temporal and spatial regulation of protein translation during development (Aboobaker et al., 2005; Wienholds et al., 2005). For example, miRNAs have been shown to affect canonical signaling pathways that are important for nervous system development, such as the MAPK and Notch signaling pathways (Lai et al., 2005; Chiba, 2006; Louvi and Artavanis-Tsakonas, 2006; Paroo et al., 2009; Zhu et al., 2010). It has been hypothesized that in these essential developmental pathways, miRNAs reduce the impact of stochastic variability in mRNA transcript levels on actual protein levels, which subsequently buffers the effects of environmental perturbations on cellular functions (Wu et al., 2009). Thus, some miRNAs evolved to maintain the robust association between gene expression patterns and fixed developmental traits (Peterson et al., 2009).

In contrast to their role on constraining plasticity during development, miRNAs seem to play a role in enhancing plasticity

in the context of behavior and neuronal functions. This has been demonstrated in several recent studies, which implicated multiple miRNA genes in the regulation of neuronal plasticity (Fiore et al., 2011; Siegel et al., 2011; McNeill and Van Vactor, 2012). For example, *miR-132* and *miR-134* have been implicated in the growth and pruning of mammalian dendritic spines (Schratt et al., 2006; Impey et al., 2010), *miR-133b* in neurotransmitter vesicle size (Kim et al., 2007), and others in different aspects of neuronal plasticity (Schratt, 2009).

Given the emerging importance of miRNAs for neuronal plasticity, it is perhaps not surprising that distinct miRNA genes have been implicated in the regulation of behavioral plasticity as well, including entrainment of the circadian clock in mammals (Na et al., 2009; Bartok et al., 2013), positive and negative responses to specific odors in *Drosophila* (Li et al., 2013), and the social response to unfamiliar conspecifics in mice (Gascon et al., 2014). Specific miRNA genes have also been implicated in processes associated with learning and memory, including in social insects. For example, expression levels of several miRNAs are associated with spatial learning (Qin et al., 2013), and long-term olfactory memory (Cristino et al., 2014) in the honey bee. In addition, studies in *Drosophila melanogaster* showed that blocking the action of *dme-miR-276a* in the mushroom bodies, a key neuroanatomical structure necessary for many cognitive functions (Heisenberg, 2003), leads to inhibition of long-term olfactory memory formation via direct interaction with the dopamine receptor *DopR* (Li et al., 2013).

In addition to neuronal functions of miRNAs, some miRNAs can also affect behavior via their actions in non-neuronal tissues. For example, manipulations of the expression of *miR-184* is implicated in the synthesis and release of insulin (Morita et al., 2013), a conserved and ubiquitously important neuroendocrine factor that is secreted from non-neuronal cells in all animal lineages (Ament et al., 2008; Wolschin et al., 2011).

The Possible Role of miRNAs in the Regulation of Traits Associated with Eusociality

Developmental Plasticity: Caste Differentiation

The completion of the honey bee genome revealed many conserved candidate miRNAs (Weinstock et al., 2006). Because of the known functions of miRNAs in the regulation of various developmental processes, it has been suggested that miRNAs are likely to contribute to the developmental processes of reproductive caste (queen-worker) differentiation (Weaver et al., 2007; Bonasio et al., 2010). In this context, it was recently reported that the expression level of the miRNA *ame-miR-71* is higher in workers relative to queens during the pupal stage (Weaver et al., 2007). A subsequent study revealed that many additional miRNAs are differentially expressed between larvae that are destined to develop as either queens or workers (Shi et al., 2014). These differences in miRNA expression levels are consistent with the hypothesis that miRNAs are involved in the

regulation of caste determination and differentiation. However, functional analyses of these miRNAs is needed to establish genetic causation between changes in the expression of specific miRNAs and the development of reproductive traits.

In contrast to species such as the honey bee, in which caste differentiation occurs early during larval development, in some eusocial species such as the ant *Harpegnathos saltator*, females retain the potential to become reproductive individuals (gamergates) throughout life. Although gamergates are morphologically worker-like, they reproduce and behave like a queen following the loss of the primary queen (Peeters et al., 2000). In this species, the transition of workers into gamergates is associated with a significant reduction in the global expression levels of several miRNA genes (Bonasio et al., 2010). How global miRNA down-regulation occurs, and why it might be important for the regulation of reproductive division of labor in this species, are not yet known.

Surprisingly, recent reports suggest that exogenous miRNAs can also affect reproductive caste-determination in honey bees. Guo et al. (2013) reported that miRNAs are present in the honey bee larval food. A comparison of short RNAs found in worker food versus “royal jelly” (food that induces queen development) indicated that the overall amount of miRNAs that are fed to worker-destined larvae is significantly higher than in food given to queen-destined larvae. Furthermore, queen-destined larvae that were fed with royal jelly supplemented with the worker-enriched miRNA *ame-miR-184* developed some worker-like morphologies (e.g., smaller body and shorter wings). This remarkable finding suggests that in honey bees, the consumption of exogenous miRNAs could play an important role in the differentiation of totipotent larvae into either sterile workers or reproductive queens. In this context, the conserved role of *miR-184* in the regulation of neuroendocrine functions across different animal taxa (Morita et al., 2013) is particularly alluring. In agreement with this hypothesis, genetic pathways that are targeted by *miR-184* in mammals are also important for queen versus worker differentiation in bees (Wolschin et al., 2011; Foret et al., 2012), suggesting that perhaps these observed effects of *miR-184* are conserved to the same pathways across mammals and insects.

Behavioral Plasticity: Division of Labor

One of the best-studied aspects of eusociality is the division of labor between workers. In some eusocial insects, such as the honey bee, division of labor relates to age (Robinson, 1992; Naug and Gadagkar, 1998; Kim et al., 2012). Young worker bees (typically <14 days of age) typically perform in-hive tasks, such as brood care (“nursing”) or food handling, and later in life (typically at around 3 weeks of age) they transition to foraging outside the hive. This well-characterized form of behavioral development has emerged as an excellent model for the molecular mechanisms involved in social behavioral plasticity (Robinson et al., 1997, 2005; Denison and Raymond-Delpech, 2008; Bloch and Grozinger, 2011). Gene expression studies, mostly using brain tissue, have demonstrated that division of labor in honey bees, and several other eusocial species, is associated with task-specific mRNA transcriptional profiles (Whitfield et al., 2003;

Adams et al., 2008; Daugherty et al., 2011; Liu et al., 2011; Oxley et al., 2014).

Three recent studies also examined the possible association between changes in brain miRNAs transcript levels and division of labor in honey bees (Behura and Whitfield, 2010; Greenberg et al., 2012; Liu et al., 2012). All three studies found that the expression levels of several miRNAs are upregulated in the brains of foragers relative to bees that perform in-hive duties (Table 1).

The association of miRNA transcript levels with specific behavioral states in colonies of eusocial insects is not limited to reproductive and worker divisions of labor. For example, reproductive queens in diverse eusocial species mate only once in their lifetime (Woyke, 1955). In honey bees, newly eclosed virgin queens (gynes) leave the hive for their sole “nuptial flight” during which they copulate with 10–20 males. After mating, they spend the rest of their lives laying eggs inside the hive. Thus, virgin and mated queens represent two distinct behavioral and physiological states (Winston, 1987). A recent study of the miRNA transcriptome in virgin and mated honey bee queens identified two different genes (*ame-miR-124* and *ame-miR-275*), which are differentially expressed in virgin and mated queens (Wu et al., 2014). While the precise function of these miRNAs

in honey bees is not known, previous reports indicate that *miR-124* is an evolutionary conserved, brain-enriched miRNA that plays a role in neural development and plasticity in invertebrates, birds, and mammals (Cao et al., 2007; Makeyev et al., 2007; Rajasethupathy et al., 2009), and more specifically in the development and function of the peripheral sensory system in *C. elegans* (Clark et al., 2010). *miR-275* is also conserved across insects, and has been implicated in the regulation of egg laying behavior in *Aedes aegypti* (Bryant et al., 2010). Wu et al. (2014) speculated that the upregulation of *ame-miR-124* miRNA in virgin queens might be related to the modulation of sensory and/or other neuronal functions associated with mating behaviors, while the increased expression of *ame-miR-275* in mated queens might be important for the newly mated queens to initiate egg-laying behavior.

A Case for the Possible Role of miRNAs in the Evolution of Eusociality

Why eusociality evolved multiple times within Hymenoptera but is rare in other insect orders is still a mystery. Several evolutionary models have attempted to explain this phenomenon by proposing various ultimate selective forces that may have driven the repeated rise of eusocial traits in this insect order (Hamilton, 1964; Andersson, 1984; Nowak et al., 2010). Although the regulation of phenotypes associated with eusociality has been independently linked to key regulatory pathways such as insulin and juvenile hormone signaling (Page and Amdam,

TABLE 1 | microRNAs that are differentially expressed in honey bee foragers and nurses.

miRNAs	Behura and Whitfield (2010)	Greenberg et al. (2012)	Liu et al. (2012)
<i>ame-let-7</i>	Nurses		Nurses
<i>ame-Bantam</i>		Foragers	
<i>ame-miR-9</i>	Foragers		
<i>ame-miR-12</i>	Foragers		
<i>ame-miR-13a</i>	Nurses		
<i>ame-miR-13b</i>	Nurses		Foragers*
<i>ame-miR-14</i>	Nurses		
<i>ame-miR-31a</i>			Nurses
<i>ame-miR-92a</i>	Foragers		Foragers
<i>ame-miR-124</i>	Nurses		
<i>ame-miR-133</i>			Foragers
<i>ame-miR-184</i>		Foragers	
<i>ame-miR-210</i>	Foragers		Foragers
<i>ame-miR-219</i>	Foragers		
<i>ame-miR-263</i>	Foragers		
<i>ame-miR-275</i>			Nurses
<i>ame-miR-276</i>	Nurses		
<i>ame-miR-278</i>			Foragers
<i>ame-miR-279</i>			Nurses
<i>ame-miR-283</i>	Foragers		
<i>ame-miR-2796</i>		Foragers	

miRNAs that were differentially expressed in at least two of the studies are highlighted in red. Denoted worker group (foragers/nurses) expressed significantly higher levels relative to the other group. Behura and Whitfield (2010) measured expression of pri-miRNA using qRT-PCR, Liu et al. (2012) relied on RNA sequencing of mature miRNA, while Greenberg et al. (2012) measured mature miRNA using northern blots. *qRT-PCR analysis showed a trend that was opposite to the RNA-seq data.

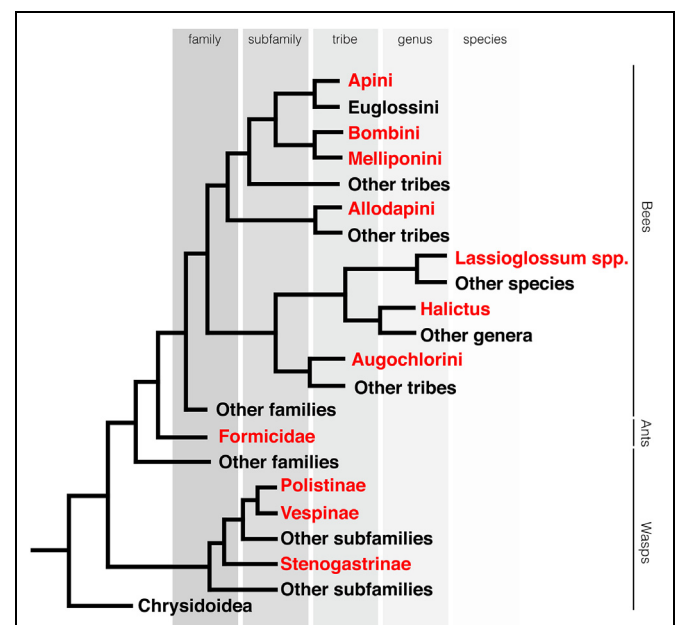


FIGURE 2 | Eusociality evolved multiple times in hymenoptera.

Phylogeny of the Aculeata. Clades containing eusocial species highlighted in red. Phylogeny is based on Danforth et al. (2013) and Johnson et al. (2013). Each branch represents the lowest taxonomic classification level that is solely comprised of eusocial species.

TABLE 2 | Genomes analyzed.

Order	Species	Common name	Eusocial	NCBI BioProject ID
Ixodida	<i>Ixodes scapularis</i>	Deer tick	No	34667
Hemiptera	<i>Acyrtosiphon pisum</i>	Pea aphid	No	29489
Coleoptera	<i>Tribolium castaneum</i>	Red flour beetle	No	15718
Lepidoptera	<i>Bombyx mori</i>	Silkworm	No	205630
Diptera	<i>Drosophila melanogaster</i> *	Fruit fly	No	164
Diptera	<i>Aedes aegypti</i>	Mosquito	No	19731
Hymenoptera	<i>Athaliae rosae</i>	Turnip sawfly	No	167403
Hymenoptera	<i>Microplitis demolitor</i> *	Parasitoid wasp	No	251518
Hymenoptera	<i>Nasonia vitripennis</i> *	Parasitoid wasp	No	20073
Hymenoptera	<i>Nasonia longicornis</i>	Parasitoid wasp	No	20225
Hymenoptera	<i>Nasonia girulta</i>	Parasitoid wasp	No	20223
Hymenoptera	<i>Apis mellifera</i>*	Honey bee	Yes	13343
Hymenoptera	<i>Apis dorsata</i>	Honey bee	Yes	174631
Hymenoptera	<i>Apis florea</i>	Honey bee	Yes	86991
Hymenoptera	<i>Bombus impatiens</i>*	Bumble bee	Yes	70395
Hymenoptera	<i>Bombus terrestris</i>	Bumble bee	Yes	68545
Hymenoptera	<i>Lasioglossum albipes</i>	Sweat bee	Facultative	174755
Hymenoptera	<i>Megachile rotundata</i>*	Leafcutter bee	No	87021
Hymenoptera	<i>Harpegnathos saltator</i>	Jumping ant	Yes	50203
Hymenoptera	<i>Camponotus floridanus</i>	Carpenter ant	Yes	50201
Hymenoptera	<i>Atta cephalotes</i>*	Leafcutter ant	Yes	48091
Hymenoptera	<i>Solenopsis invicta</i>	Fire ant	Yes	49629
Hymenoptera	<i>Pogonomyrmex barbatus</i>	Harvester ant	Yes	45797
Hymenoptera	<i>Polistes dominula</i>	Paper wasp	Yes	Unpublished

The following genomes were analyzed for the presence or absence of miRNAs. We performed an initial BLAST search of annotated miRNAs from *D. melanogaster*, *A. mellifera*, and *N. vitripennis* in the species denoted by *. Candidate miRNAs identified as either present only in the genomes of eusocial species (red) or only in Aculeate species (bold), were subsequently analyzed in all genomes listed.

2007; Toth and Robinson, 2007; Bloch and Grozinger, 2011), the actual molecular events that supported traits contributing to eusociality remain elusive. Here, we propose that the molecular evolution of specific miRNAs could have contributed to the phenotypic evolution of eusociality. We propose that these miRNAs may have contributed to the emergence of eusociality by either introducing new regulatory nodes to ancestral behavioral genetic networks, and/or by supporting novel behavioral genetic networks.

The primary sequence of mature miRNAs is often completely conserved across long phylogenetic distances. Consequently, conserved miRNAs are likely to regulate similar target protein-coding genes in distant taxa, and thus support analogous phenotypes across phylogeny (Lee et al., 2007). Given their broad pleiotropic function, novel miRNAs can modify complex developmental or physiological genetic programs. Because of this, it has been suggested by several investigators that, similarly, to the evolution of protein regulatory networks (e.g., evolution of novel transcription factors), novel miRNAs could lead to evolutionary innovations (Sempere et al., 2006; Lee et al., 2007; Niwa and Slack, 2007; Tarver et al., 2012) such as the establishment of new body plans, or novel behavioral traits (Peterson et al., 2009).

Consistent with this premise, the evolution of bilateral animals from eumetazoans was associated with a great expansion in the number of miRNA genes (Niwa and Slack, 2007).

Other examples include the many novel miRNA genes found within placental mammals, and their clade-specific expansion in primates (Sempere et al., 2006). Although the evolution of eusociality is considered a major evolutionary transition event (Maynard Smith and Szathmari, 1995), the hypothesis that it was also associated with the evolution of novel miRNAs has not been previously suggested. We reasoned that the monophyletic Aculeata clade is ideal for testing this hypothesis since, based on current phylogenetic models (Danforth et al., 2013; Johnson et al., 2013), eusociality has independently emerged in this group multiple times (Figure 2). Below we discuss two independent, non-mutually exclusive hypotheses for the possible involvement of miRNAs in the evolution of eusociality.

Hypothesis 1: Specific miRNAs have been Repeatedly Associated with Eusocial Evolution in Hymenoptera

Here, we hypothesize that, similarly to the evolution of novel transcription factors, the repeated evolution of specific new miRNAs, either *de novo* or via duplication events, facilitated the evolution of some eusocial traits in multiple independent clades that currently display eusociality. Under this hypothesis, novel miRNAs in current eusocial species act as essential nodes in genetic networks that support eusocial traits. If true, we expect that specific miRNAs would be more likely to be present in the

genomes of eusocial species in comparison to related solitary species.

As an initial test of this hypothesis, we searched for miRNA genes in the sequenced genomes of species in the Aculeata clade, which includes all living eusocial species in Hymenoptera. We first generated a list of all known annotated miRNA genes available in miRBase for the eusocial honey bee *Apis mellifera*, the solitary wasp *Nasonia vitripennis*, and the fruit fly *D. melanogaster* (Griffiths-Jones et al., 2006). Next, we searched for the presence or absence of each annotated miRNA in several representative hymenopteran genomes (Table 2) using BLASTN (Altschul et al., 1997). We only scored a miRNA as “present” if an exact match to the mature miRNA sequence was found in the genome (Figure 3). Consistent with data from other animal clades (Tarver et al., 2012), we found that most annotated miRNA genes aligned with phylogeny rather than with the presence or absence of eusociality. Nevertheless, five miRNA genes (*ame-miR-281*, *ame-miR-306*, *ame-miR-279c*, *ame-miR-279d*, and *ame-miR-6065*) seem to be associated with the expression of eusocial traits independent of phylogeny (Figure 3).

To further refine our results, we subsequently extended the bioinformatic analyses to all available sequenced hymenopteran genomes, as well as several non-hymenopteran insect species, which served as outgroups (Figure 4A). Although the low sequence coverage for some of the analyzed ant genomes could lead to higher false-negative discovery rate, we reasoned that the likelihood that certain miRNAs will be falsely missing from all analyzed genomes is very low. Future miRNA sequencing data from many of the species studied here should further help reducing the possibility of false-negatives.

This analysis revealed that three out of the five putative eusociality-associated miRNAs were unique to Hymenoptera (*ame-miR-281*, *ame-miR-306*, and *ame-miR-279c*), and one possibly unique to Aculeata (*ame-miR-6065*). The phylogenetic distribution of these five miRNAs indicated that multiple eusociality-associated miRNAs might have been gained and lost during the Hymenoptera radiation. In addition, we found that two eusociality-associated miRNA genes (*miR-306* and *miR-6065*) were lost in the eusocial wasp *Polistes dominula*. Markedly, two of the eusocial-related miRNAs (*miR-281* and *miR-6065*) were also present in the genome of the facultative eusocial bee *Lasioglossum albipes*. One possible explanation for this finding is that these specific miRNAs are important for traits associated with basal levels of sociality such as communal living, overlapping generations, and reproductive division of labor (Kocher et al., 2013).

Our analysis also revealed that two of the candidate eusociality-related miRNAs (*mir-279c* and *mir-279d*) belong to a single conserved miRNA-family (Cayirlioglu et al., 2008; Hartl et al., 2011; Luo and Sehgal, 2012; Mohammed et al., 2014). The most parsimonious interpretation of these observed phylogenetic patterns is that *miR-279d* is conserved across Arthropoda, but was lost in Diptera and Hymenoptera, and then reappeared via duplications in eusocial Aculeates. In contrast, *miR-279c* seems to have specifically evolved in Hymenoptera prior to the divergence of Aculeata, and was subsequently lost from non-social Aculeate species. The identification of members of the



FIGURE 3 | Phylogenetic distributions of miRNAs in Hymenoptera genomes. miRNAs present in each genome are shown in yellow, while those absent are shown in blue. miRNAs present in all or only one species are not shown. Data are clustered based on the phylogenetic relationships between the species analyzed, with eusocial species shown in red. Genes framed in black are present only in eusocial species. Genes framed in red are present only in Aculeata. The fruit fly *Drosophila melanogaster* served as the outgroup. Phylogeny based on Danforth et al. (2013) and Johnson et al. (2013).

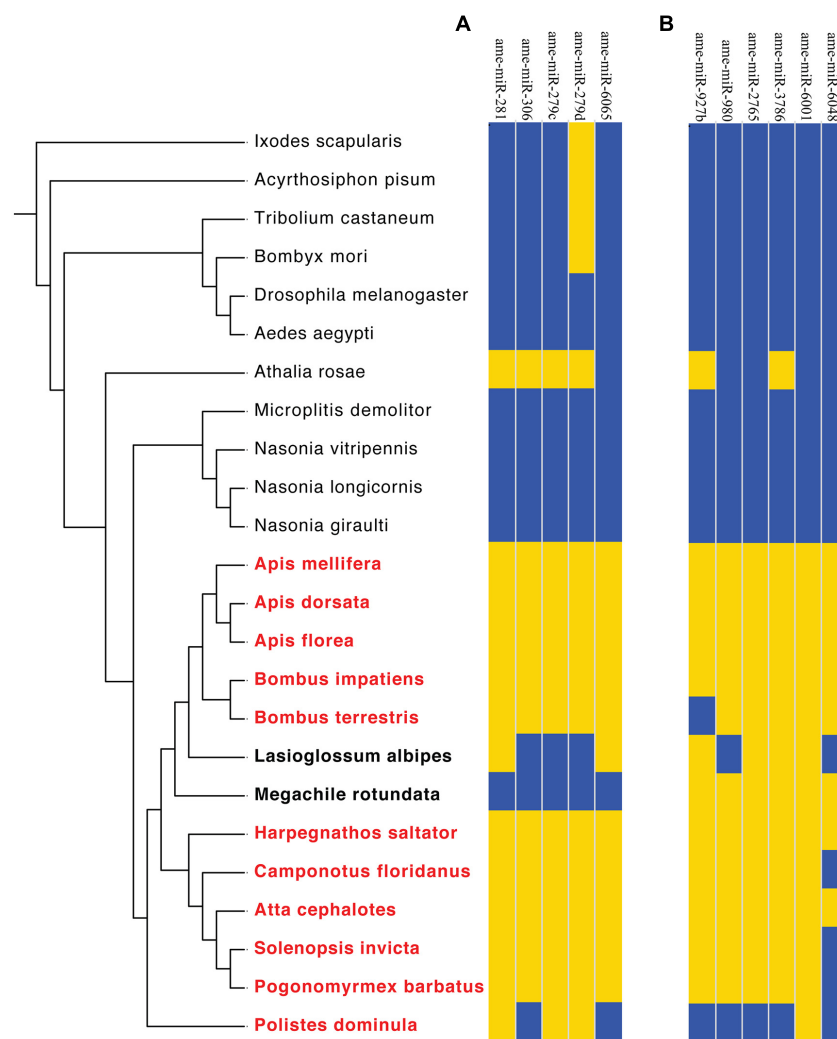


FIGURE 4 | Phylogenetic distributions of Aculeata-specific miRNA.

miRNAs present in each genome are shown in yellow, while those absent are shown in blue. **(A)** High resolution analysis of eusociality-associated candidate miRNAs from **Figure 3**. **(B)** High resolution analysis of

Aculeata-specific candidate miRNAs. Data are clustered based on the phylogenetic relationships between the analyzed genomes, with Aculeate species in bold and eusocial species in red. *L. albipes* is facultative eusocial. Species phylogeny as in **Figure 3**.

mir-279 family as possible candidate genes for the evolution of eusociality is in agreement with findings about their differential regulation between nurses and foragers (**Table 1**), and possible functions in *Drosophila*. For example, members of the *mir-279* family have been implicated in regulating neuronal development (Hartl et al., 2011), olfactory receptivity (Cayirlioglu et al., 2008; Hartl et al., 2011), and circadian rhythms (Luo and Sehgal, 2012). It is interesting to note that plasticity in both circadian rhythms (Bloch, 2010) and olfactory neurons has been shown to be associated with worker and reproductive divisions of labor in eusocial Hymenoptera (López-Riquelme et al., 2006; Zube and Rössler, 2008; Mysore et al., 2009). Although preliminary, these findings suggest that members of the *mir-279* gene family are prime candidates for studies on the possible roles of specific miRNA in the evolution of eusociality-related traits.

To further test this hypothesis it will be necessary to increase the phylogenetic resolution of our analyses by studying the miRNA repertoire encoded by the genomes of additional social and solitary insects. It will also require the development of tools that will allow the manipulation of focal miRNA expression to causally determine their effect on behavioral and physiological traits related to eusociality. The recent progress in genome-editing techniques for honey bees and other social insects (Wang et al., 2013; Schulte et al., 2014) suggest that this will be feasible in the near future. Another complementary approach will be to study the protein-coding genetic networks that eusociality-associated miRNAs are interacting with. By identifying the genes involved, their spatial and temporal expression patterns, and the possible physiological and behavioral processes they modulate, a higher resolution picture of the genetics that support eusociality could emerge.

Hypothesis 2: Aculeate-Specific miRNAs were Required for Eusocial Evolution

The second hypothesis we consider is that the presence of specific miRNAs in the pre-eusocial Aculeate genome might have “primed” certain species to evolve eusociality. In other words, specific miRNAs, already present in the genome of the solitary Aculeate ancestor were required, but not sufficient, for the emergence of eusocial traits. Under this hypothesis, specific miRNAs already present in the ancestral solitary aculeate increased the probability of emergence of specific behavioral and physiological traits in response to selective pressures that favored eusociality.

If true, we expect that specific miRNAs should be present in all Aculeate genomes, but absent from all other hymenopteran genomes, as eusociality has never been observed in hymenopteran species outside of the Aculeata. Our initial analysis revealed six Hymenoptera specific miRNA genes (*ame-miR-927b*, *ame-miR-980*, *ame-miR-2765*, *ame-miR-3786*, *ame-miR-6001*, and *ame-miR-6048*; **Figure 3**). However, two of these genes were also present in the sawfly *Athalia rosae* (*ame-miR-927b* and *ame-miR-3786*), and therefore are not specific to Aculeata. Three additional genes (*ame-miR-980*, *ame-miR-2765*, and *ame-miR-6048*) appear to have originated after the divergence of Vespidae and therefore did not fulfill the above criteria (**Figure 4B**). Thus, our analysis revealed *ame-miR-6001* as the single Aculeate-specific miRNA candidate gene that should be tested in the context of the above hypothesis. Similarly to Hypothesis 1 (see Hypothesis 1: Specific miRNAs have been Repeatedly Associated with Eusocial Evolution in Hymenoptera), the possible role of *miR-6001* in the repeated evolution of eusocial traits in Aculeata is hypothetical. Directly testing the hypothesis we put forward here will require extensive molecular, biochemical, and phenotypic studies of its possible physiological and behavioral roles in eusocial traits.

References

- Aboobaker, A., Tomancak, P., Patel, N., Rubin, G. M., and Lai, E. C. (2005). *Drosophila* microRNAs exhibit diverse spatial expression patterns during embryonic development. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18017–18022. doi: 10.1073/pnas.0508823102
- Adams, H. A., Southey, B. R., Robinson, G. E., and Rodriguez-Zas, S. L. (2008). Meta-analysis of genome-wide expression patterns associated with behavioral maturation in honey bees. *BMC Genomics* 9:503. doi: 10.1186/1471-2164-9-503
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Alvarez-García, I., and Miska, E. A. (2005). MicroRNA functions in animal development and human disease. *Development* 132, 4653–4662. doi: 10.1242/dev.02073
- Ament, S. A., Corona, M., Pollock, H. S., and Robinson, G. E. (2008). Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4226–4231. doi: 10.1073/pnas.0800630105
- Andersson, M. (1984). The evolution of eusociality. *Annu. Rev. Ecol. Syst.* 15, 165–189. doi: 10.1146/annurev.es.15.110184.001121
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5
- Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002

A Look to the Future

To date, the majority of data about the function of miRNAs in social insects come from studies of the European honey bee, *A. mellifera*. Thus, additional molecular and evolutionary analyses of non-hymenopteran eusocial insects as well as other eusocial and solitary clades in the Hymenoptera are required in order to better understand miRNA functions in the context of eusociality. Furthermore, to establish causation between the action of specific miRNAs and eusocial traits, new *in vivo* genetic and molecular techniques to manipulate social insects are required. Recent advances in molecular genetics of social insects (Yan et al., 2014), and the successful generation of transgenic honey bees (Ben-Shahar, 2014; Schulte et al., 2014), suggest that such studies might be possible in the near future. Furthermore, the development of pharmacological reagents that can block or mimic the action of specific miRNAs (e.g., antagomirs), would represent another important step in that direction (Cristino et al., 2014).

Author Contributions

ES collected and analyzed genomic data. ES, GB, and YB-S gathered and synthesized relevant literature. ES, GB, and YB-S wrote the manuscript.

Acknowledgments

We would like to thank Amy Toth for granting us access to unpublished genomic data of *P. dominula*. This work was supported by grants from the US National Science Foundation (1322783) to YB-S and the US-Israel Binational Science Foundation (BSF2012807) to GB.

- Bartok, O., Kyriacou, C. P., Levine, J., Sehgal, A., Kadener, S., and B, P. R. S. (2013). Adaptation of molecular circadian clockwork to environmental changes: a role for alternative splicing and miRNAs. *Proc. Biol. Sci.* 280:20130011. doi: 10.1098/rspb.2013.0011
- Behura, S. K., and Whitfield, C. W. (2010). Correlated expression patterns of microRNA genes with age-dependent behavioural changes in honeybee. *Insect Mol. Biol.* 19, 431–439. doi: 10.1111/j.1365-2583.2010.01010.x
- Ben-Shahar, Y. (2014). A piggyBac route to transgenic honeybees. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8708–8709. doi: 10.1073/pnas.1407876111
- Berezikov, E. (2011). Evolution of microRNA diversity and regulation in animals. *Nat. Rev. Genet.* 12, 846–860. doi: 10.1038/nrg3079
- Bloch, G. (2010). The social clock of the honeybee. *J. Biol. Rhythms* 25, 307–317. doi: 10.1177/0748730410380149
- Bloch, G., and Grozinger, C. M. (2011). Social molecular pathways and the evolution of bee societies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 2155–2170. doi: 10.1098/rstb.2010.0346
- Bonasio, R., and Shiekhattar, R. (2014). Regulation of transcription by long noncoding RNAs. *Annu. Rev. Genet.* 48, 433–455. doi: 10.1146/annurev-genet-120213-092323
- Bonasio, R., Zhang, G., Ye, C., Mutti, N. S., Fang, X., Qin, N., et al. (2010). Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* 329, 1068–1071. doi: 10.1126/science.1192428

- Braakman, I., and Bulleid, N. J. (2011). Protein folding and modification in the mammalian endoplasmic reticulum. *Annu. Rev. Biochem.* 80, 71–99. doi: 10.1146/annurev-biochem-062209-093836
- Breitbart, R. E., Andreadis, A., and Nadal-Ginard, B. (1987). Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu. Rev. Biochem.* 56, 467–495. doi: 10.1146/annurev.bi.56.070187.002343
- Bryant, B., Macdonald, W., and Raikhel, A. S. (2010). microRNA miR-275 is indispensable for blood digestion and egg development in the mosquito *Aedes aegypti*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22391–22398. doi: 10.1073/pnas.1016230107
- Bushati, N., and Cohen, S. M. (2007). microRNA functions. *Annu. Rev. Cell Dev. Biol.* 23, 175–205. doi: 10.1146/annurev.cellbio.23.090506.123406
- Cao, X., Pfaff, S. L., and Gage, F. H. (2007). A functional study of miR-124 in the developing neural tube A functional study of miR-124 in the developing neural tube. *Genes Dev.* 21, 531–536. doi: 10.1101/gad.1519207
- Cayirlioglu, P., Kadow, I. G., Zhan, X., Okamura, K., Suh, G. S. B., Gunning, D., et al. (2008). Hybrid neurons in a microRNA mutant are putative evolutionary intermediates in insect CO₂ sensory systems. *Science* 319, 1256–1260. doi: 10.1126/science.1149483
- Chiba, S. (2006). Notch signaling in stem cell systems. *Stem Cells* 24, 2437–2447. doi: 10.1634/stemcells.2005-0661
- Clark, A. M., Goldstein, L. D., Tevlin, M., Tavaré, S., Shaham, S., and Miska, E. A. (2010). The microRNA miR-124 controls gene expression in the sensory nervous system of *Caenorhabditis elegans*. *Nucleic Acids Res.* 38, 3780–3793. doi: 10.1093/nar/gkq083
- Crespi, B. J., and Yanega, D. (1995). The definition of eusociality. *Behav. Ecol.* 6, 109–115. doi: 10.1093/beheco/6.1.109
- Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561–563. doi: 10.1038/227561a0
- Cristino, A. S., Barchuk, A. R., Freitas, F. C. P., Narayanan, R. K., Biergans, S. D., Zhao, Z., et al. (2014). Neuroligin-associated microRNA-932 targets actin and regulates memory in the honeybee. *Nat. Commun.* 5, 5529. doi: 10.1038/ncomms6529
- Danforth, B. N., Cardinal, S., Praz, C., Almeida, E. A., and Michez, D. (2013). The impact of molecular data on our understanding of bee phylogeny and evolution. *Annu. Rev. Entomol.* 58, 57–78. doi: 10.1146/annurev-ento-120811-153633
- Daugherty, T. H. F., Toth, A. L., and Robinson, G. E. (2011). Nutrition and division of labor: effects on foraging and brain gene expression in the paper wasp *Polistes metricus*. *Mol. Ecol.* 20, 5337–5347. doi: 10.1111/j.1365-294X.2011.05344.x
- Denison, R., and Raymond-Delpech, V. (2008). Insights into the molecular basis of social behaviour from studies on the honeybee, *Apis mellifera*. *Invertebr. Neurosci.* 8, 1–9. doi: 10.1007/s10158-008-0066-6
- Fiore, R., Khudayberdiev, S., Saba, R., and Schratt, G. (2011). MicroRNA function in the nervous system. *Prog. Mol. Biol. Transl. Sci.* 102, 47–100. doi: 10.1016/B978-0-12-415795-8.00004-0
- Fischman, B. J., Woodard, S. H., and Robinson, G. E. (2011). Molecular evolutionary analyses of insect societies. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10847–10854. doi: 10.1073/pnas.1100301108
- Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S. E., Robinson, G. E., et al. (2012). DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4968–4973. doi: 10.1073/pnas.120392109
- Gascon, E., Lynch, K., Ruan, H., Almeida, S., Verheyden, J. M., Seeley, W. W., et al. (2014). Alterations in microRNA-124 and AMPA receptors contribute to social behavioral deficits in frontotemporal dementia. *Nat. Med.* 20, 1444–1451. doi: 10.1038/nm.3717
- Gott, J. M., and Emeson, R. B. (2000). Functions and mechanisms of RNA editing. *Annu. Rev. Genet.* 34, 499–531. doi: 10.1146/annurev.genet.34.1.499
- Greenberg, J. K., Xia, J., Zhou, X., Thatcher, S. R., Gu, X., Ament, S. A., et al. (2012). Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes Brain Behav.* 11, 660–670. doi: 10.1111/j.1601-183X.2012.00782.x
- Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34, D140–D144. doi: 10.1093/nar/gkj112
- Guo, X., Su, S., Skogerboe, G., Dai, S., Li, W., Li, Z., et al. (2013). Recipe for a busy bee: microRNAs in honey bee caste determination. *PLoS ONE* 8:e81661. doi: 10.1371/journal.pone.0081661
- Hamilton, W. D. (1964). The genetical evolution of social behaviour: I, II. *J. Theor. Biol.* 7, 1–52. doi: 10.1016/0022-5193(64)90038-4
- Hartl, M., Loschek, L. F., Stephan, D., Siju, K. P., Knappmeyer, C., and Kadow, I. C. G. (2011). A new Prospero and microRNA-279 pathway restricts CO₂ receptor neuron formation. *J. Neurosci.* 31, 15660–15673. doi: 10.1523/JNEUROSCI.2592-11.2011
- Heisenberg, M. (2003). Mushroom body memoir: from maps to models. *Nat. Rev. Neurosci.* 4, 266–275. doi: 10.1038/nrn1074
- Hölldobler, B., and Wilson, E. O. (2009). *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies*. New York, NY: W.W. Norton.
- Impey, S., Davare, M., Lesiak, A., Lasiek, A., Fortin, D., Ando, H., et al. (2010). An activity-induced microRNA controls dendritic spine formation by regulating Rac1-PAK signaling. *Mol. Cell. Neurosci.* 43, 146–156. doi: 10.1016/j.mcn.2009.10.005
- Johnson, B. R., Borowiec, M. L., Chiu, J. C., Lee, E. K., Atallah, J., and Ward, P. S. (2013). Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. *Curr. Biol.* 23, 2058–2062. doi: 10.1016/j.cub.2013.08.050
- Keller, L., and Ross, K. G. (1998). Selfish genes: a green beard in the red fire ant. *Nature* 394, 573–575. doi: 10.1038/29064
- Kim, B., Kim, K. W., and Choe, J. C. (2012). Temporal polyethism in Korean yellowjacket foragers, *Vespula koreensis* (Hymenoptera, Vespidae). *Insectes Soc.* 59, 263–268. doi: 10.1007/s00040-011-0212-8
- Kim, J., Inoue, K., Ishii, J., Vanti, W. B., Voronov, S. V., Murchison, E., et al. (2007). A microRNA feedback circuit in midbrain dopamine neurons. *Science* 317, 1220–1224. doi: 10.1126/science.1140481
- Kocher, S. D., Li, C., Yang, W., Tan, H., Yi, S. V., Yang, X., et al. (2013). The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. *Genome Biol.* 14:R142. doi: 10.1186/gb-2013-14-12-r142
- Lai, E. C., Tam, B., and Rubin, G. M. (2005). Pervasive regulation of *Drosophila* Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. *Genes Dev.* 19, 1067–1080. doi: 10.1101/gad.1291905
- Lee, C.-T., Risom, T., and Strauss, W. M. (2007). Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. *DNA Cell Biol.* 26, 209–218. doi: 10.1089/dna.2006.0545
- Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854. doi: 10.1016/0092-8674(93)90529-Y
- Lee, T., and Young, R. (2000). Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* 34, 77–137. doi: 10.1146/annurev.genet.34.1.77
- Li, W., Cressy, M., Qin, H., Fulga, T., Van Vactor, D., and Dubnau, J. (2013). MicroRNA-276a functions in ellipsoid body and mushroom body neurons for naive and conditioned olfactory avoidance in *Drosophila*. *J. Neurosci.* 33, 5821–5833. doi: 10.1523/JNEUROSCI.4004-12.2013
- Liu, F., Li, W., Li, Z., Zhang, S., Chen, S., and Su, S. (2011). High-abundance mRNAs in *Apis mellifera*: comparison between nurses and foragers. *J. Insect Physiol.* 57, 274–279. doi: 10.1016/j.jinsphys.2010.11.015
- Liu, F., Peng, W., Li, Z., Li, W., Li, L., Pan, J., et al. (2012). Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: comparison between nurses and foragers. *Insect Mol. Biol.* 21, 297–303. doi: 10.1111/j.1365-2583.2012.01135.x
- López-Riquelme, G. O., Malo, E. A., Cruz-López, L., and Fanjul-Moles, M. L. (2006). Antennal olfactory sensitivity in response to task-related odours of three castes of the ant *Atta mexicana* (hymenoptera: formicidae). *Physiol. Entomol.* 31, 353–360. doi: 10.1111/j.1365-3032.2006.00526.x
- Louvi, A., and Artavanis-Tsakonas, S. (2006). Notch signalling in vertebrate neural development. *Nat. Rev. Neurosci.* 7, 93–102. doi: 10.1038/nrn1847

- Luo, W., and Sehgal, A. (2012). Regulation of circadian behavioral output via a microRNA-JAK/STAT circuit. *Cell* 148, 765–779. doi: 10.1016/j.cell.2011.12.024
- Makeyev, E. V., Zhang, J., Carrasco, M. A., and Maniatis, T. (2007). The microRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* 27, 435–448. doi: 10.1016/j.molcel.2007.07.015
- Maynard Smith, J., and Szathmari, E. (1995). *The Major Transitions in Evolution*. Oxford: Oxford University Press.
- McNeill, E., and Van Vactor, D. (2012). MicroRNAs shape the neuronal landscape. *Neuron* 75, 363–379. doi: 10.1016/j.neuron.2012.07.005
- Mohammed, J., Siepel, A., and Lai, E. C. (2014). Diverse modes of evolutionary emergence and flux of conserved microRNA clusters. *RNA* 20, 1850–1863. doi: 10.1261/rna.046805.114
- Morita, S., Horii, T., Kimura, M., and Hatada, I. (2013). MiR-184 regulates insulin secretion through repression of Slc25a22. *PeerJ* 1:e162. doi: 10.7717/peerj.162
- Mysore, K., Subramanian, K. A., Sarasij, R. C., Suresh, A., Shyamala, B. V., VijayRaghavan, K., et al. (2009). Caste and sex specific olfactory glomerular organization and brain architecture in two sympatric ant species *Camponotus sericeus* and *Camponotus compressus* (Fabricius, 1798). *Arthropod. Struct. Dev.* 38, 485–497. doi: 10.1016/j.asd.2009.06.001
- Na, Y. J., Sung, J. H., Lee, S. C., Lee, Y. J., Choi, Y. J., Park, W. Y., et al. (2009). Comprehensive analysis of microRNA-mRNA co-expression in circadian rhythm. *Exp. Mol. Med.* 41, 638–647. doi: 10.3858/emmm.2009.41.9.070
- Naug, D., and Gadagkar, R. (1998). The role of age in temporal polyethism in a primitively eusocial wasp. *Behav. Ecol. Sociobiol.* 42, 37–47. doi: 10.1007/s002650050409
- Niwa, R., and Slack, F. J. (2007). The evolution of animal microRNA function. *Curr. Opin. Genet. Dev.* 17, 145–150. doi: 10.1016/j.gde.2007.02.004
- Nowak, M. A., Tarnita, C. E., and Wilson, E. O. (2010). The evolution of eusociality. *Nature* 466, 1057–1062. doi: 10.1038/nature09205
- Oxley, P. R., Ji, L., Fetter-Pruneda, I., McKenzie, S. K., Li, C., Hu, H., et al. (2014). The genome of the *Clonal raider* ant *Cerapachys biroi*. *Curr. Biol.* 24, 451–458. doi: 10.1016/j.cub.2014.01.018
- Page, R. E., and Amdam, G. V. (2007). The making of a social insect: developmental architectures of social design. *Bioessays* 29, 334–343. doi: 10.1002/bies.20549
- Paroo, Z., Ye, X., Chen, S., and Liu, Q. (2009). Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. *Cell* 139, 112–122. doi: 10.1016/j.cell.2009.06.044
- Peeters, C., Liebig, J., and Hölldobler, B. (2000). Sexual reproduction by both queens and workers in the ponerine ant *Harpegnathos saltator*. *Insectes Soc.* 47, 325–332. doi: 10.1007/PL00001724
- Peterson, K. J., Dietrich, M. R., and McPeck, M. A. (2009). MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *Bioessays* 31, 736–747. doi: 10.1002/bies.200900033
- Qin, Q.-H., Wang, Z.-L., Tian, L.-Q., Gan, H.-Y., Zhang, S.-W., and Zeng, Z.-J. (2013). The integrative analysis of microRNA and mRNA expression in *Apis mellifera* following maze-based visual pattern learning. *Insect Sci.* 21, 619–636. doi: 10.1111/1744-7917.12065
- Qureshi, I. A., and Mehler, M. F. (2012). Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat. Rev. Neurosci.* 13, 528–541. doi: 10.1038/nrn3234
- Rajasethupathy, P., Fiumara, F., Sheridan, R., Betel, D., Puthanveetil, S. V., Russo, J. J., et al. (2009). Characterization of small RNAs in *Aplysia* reveals a role for miR-124 in constraining synaptic plasticity through CREB. *Neuron* 63, 803–817. doi: 10.1016/j.neuron.2009.05.029
- Robinson, G. E. (1992). Regulation of division of labor in insect societies. *Annu. Rev. Entomol.* 37, 637–665. doi: 10.1146/annurev.en.37.010192.003225
- Robinson, G. E., Fahrbach, S. E., and Mark, L. (1997). Insect societies and the molecular biology of social behavior. *Bioessays* 19, 1099–1108. doi: 10.1002/bies.950191209
- Robinson, G. E., Grozinger, C. M., and Whitfield, C. W. (2005). Sociogenomics: social life in molecular terms. *Nat. Rev. Genet.* 6, 257–270. doi: 10.1038/nrg1575
- Schratt, G. (2009). microRNAs at the synapse. *Nat. Rev. Neurosci.* 10, 842–849. doi: 10.1038/nrn2763
- Schratt, G. M., Tuebing, F., Nigh, E. A., Kane, C. G., Sabatini, M. E., Kiebler, M., et al. (2006). A brain-specific microRNA regulates dendritic spine development. *Nature* 439, 283–289. doi: 10.1038/nature04367
- Schulte, C., Theilenberg, E., Müller-Borg, M., Gempe, T., and Beye, M. (2014). Highly efficient integration and expression of piggyBac-derived cassettes in the honeybee (*Apis mellifera*). *Proc. Natl. Acad. Sci. U.S.A.* 111, 9003–9008. doi: 10.1073/pnas.1402341111
- Schwander, T., Lo, N., Beekman, M., Oldroyd, B. P., and Keller, L. (2010). Nature versus nurture in social insect caste differentiation. *Trends Ecol. Evol.* 25, 275–282. doi: 10.1016/j.tree.2009.12.001
- Sempere, L. F., Cole, C. N., McPeck, M. A., and Peterson, K. J. (2006). The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J. Exp. Zool. B Mol. Dev. Evol.* 306, 575–588. doi: 10.1002/jez.b.21118
- Shi, Y.-Y., Zheng, H.-J., Pan, Q.-Z., Wang, Z.-L., and Zeng, Z.-J. (2014). Differentially expressed microRNAs between queen and worker larvae of the honey bee (*Apis mellifera*). *Apidologie* 46, 35–45. doi: 10.1007/s13592-014-0299-9
- Siegel, G., Saba, R., and Schratt, G. (2011). microRNAs in neurons: manifold regulatory roles at the synapse. *Curr. Opin. Genet. Dev.* 21, 491–497. doi: 10.1016/j.gde.2011.04.008
- Smith, C. R., Toth, A. L., Robinson, G. E., and Suarez, A. V. (2008). Genetic and genomic analyses of the division of labour in insect societies. *Nat. Rev. Genet.* 9, 735–748. doi: 10.1038/nrg2429
- Tarver, J. E., Donoghue, P. C. J., and Peterson, K. J. (2012). Do miRNAs have a deep evolutionary history? *Bioessays* 34, 857–866. doi: 10.1002/bies.201200055
- Toth, A. L., and Robinson, G. E. (2007). Evo-devo and the evolution of social behavior. *Trends Genet.* 23, 334–341. doi: 10.1016/j.tig.2007.05.001
- Wang, Y., Li, Z., Xu, J., Zeng, B., Ling, L., You, L., et al. (2013). The CRISPR/Cas system mediates efficient genome engineering in *Bombyx mori*. *Cell Res.* 23, 1414–1416. doi: 10.1038/cr.2013.146
- Weaver, D. B., Anzola, J. M., Evans, J. D., Reid, J. G., Reese, J. T., Childs, K. L., et al. (2007). Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genome Biol.* 8:R97. doi: 10.1186/gb-2007-8-6-r97
- Weinstock, G. M., Robinson, G. E., Gibbs, R. A., Worley, K. C., Evans, J. D., Maleszka, R., et al. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443, 931–949. doi: 10.1038/nature05260
- Whitfield, C. W., Cziko, A.-M., and Robinson, G. E. (2003). Gene expression profiles in the brain predict behavior in individual honey bees. *Science* 302, 296–299. doi: 10.1126/science.1086807
- Wienholds, E., Kloosterman, W. P., Miska, E., Alvarez-Saavedra, E., Berezikov, E., de Bruijn, E., et al. (2005). MicroRNA expression in zebrafish embryonic development. *Science* 309, 310–311. doi: 10.1126/science.1114519
- Wienholds, E., and Plasterk, R. H. A. (2005). MicroRNA function in animal development. *FEBS Lett.* 579, 5911–5922. doi: 10.1016/j.febslet.2005.07.070
- Wilson, E. (1990). *Success and Dominance in Ecosystems: The Case of the Social Insects*. Oldendorf/Luhe: Ecology Institute.
- Wilson, E. O. (1985). The sociogenesis of insect colonies. *Science* 228, 1489–1495. doi: 10.1126/science.228.4707.1489
- Winston, M. L. (1987). *The Biology of the Honey Bee*. Cambridge, MA: Harvard University Press.
- Wolschin, F., Mutti, N. S., and Amdam, G. V. (2011). Insulin receptor substrate influences female caste development in honeybees. *Biol. Lett.* 7, 112–115. doi: 10.1098/rsbl.2010.0463
- Woodard, S. H., Fischman, B. J., Venkat, A., Hudson, M. E., Varala, K., Cameron, S. A., et al. (2011). Genes involved in convergent evolution of eusociality in bees. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7472–7477. doi: 10.1073/pnas.1103457108
- Woyke, J. (1955). Multiple mating of the honeybee queen (*Apis mellifica* L.) in one nuptial flight. *Bull. Acad. Polon. Sci.* 3, 175–180.
- Wu, C.-I., Shen, Y., and Tang, T. (2009). Evolution under canalization and the dual roles of microRNAs: a hypothesis. *Genome Res.* 19, 734–743. doi: 10.1101/gr.084640.108

- Wu, X., Wang, Z., Zhang, F., Shi, Y., and Zeng, Z. (2014). Mating flight causes genome-wide transcriptional changes in sexually mature honeybee queens. *J. Asia Pac. Entomol.* 17, 37–43. doi: 10.1016/j.aspen.2013.10.001
- Yan, H., Bonasio, R., Simola, D. F., Liebig, J., Berger, S. L., and Reinberg, D. (2015). DNA methylation in social insects: how epigenetics can control behavior and longevity. *Annu. Rev. Entomol.* 60, 435–452. doi: 10.1146/annurev-ento-010814-020803
- Yan, H., Simola, D. F., Bonasio, R., Liebig, J., Berger, S. L., and Reinberg, D. (2014). Eusocial insects as emerging models for behavioural epigenetics. *Nat. Rev. Genet.* 15, 677–688. doi: 10.1038/nrg3787
- Zhu, Q.-Y., Liu, Q., Chen, J.-X., Lan, K., and Ge, B.-X. (2010). MicroRNA-101 targets MAPK phosphatase-1 to regulate the activation of MAPKs in macrophages. *J. Immunol.* 185, 7435–7442. doi: 10.4049/jimmunol.1000798
- Zube, C., and Rössler, W. (2008). Caste- and sex-specific adaptations within the olfactory pathway in the brain of the ant *Camponotus floridanus*. *Arthropod. Struct. Dev.* 37, 469–479. doi: 10.1016/j.asd.2008.05.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Søvik, Bloch and Ben-Shahar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviors

Daria Molodtsova¹, Brock A. Harpur¹, Clement F. Kent^{1†}, Kajendra Seevananthan² and Amro Zayed^{1*}

¹ Department of Biology, York University, Toronto, ON, Canada

² Department of Computer Science and Engineering, York University, Toronto, ON, Canada

Edited by:

Greg J. Hunt, Purdue University, USA

Reviewed by:

Brian R. Johnson, University of California, Davis, USA

Sandra M. Rehan, University of New Hampshire, USA

*Correspondence:

Amro Zayed, Department of Biology, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada
e-mail: zayed@yorku.ca

† Present address:

Clement F. Kent, HHMI Janelia Farm Research Campus, Ashburn, VA, USA

It is increasingly apparent that genes and networks that influence complex behavior are evolutionary conserved, which is paradoxical considering that behavior is labile over evolutionary timescales. How does adaptive change in behavior arise if behavior is controlled by conserved, pleiotropic, and likely evolutionary constrained genes? Pleiotropy and connectedness are known to constrain the general rate of protein evolution, prompting some to suggest that the evolution of complex traits, including behavior, is fuelled by regulatory sequence evolution. However, we seldom have data on the strength of selection on mutations in coding and regulatory sequences, and this hinders our ability to study how pleiotropy influences coding and regulatory sequence evolution. Here we use population genomics to estimate the strength of selection on coding and regulatory mutations for a transcriptional regulatory network that influences complex behavior of honey bees. We found that replacement mutations in highly connected transcription factors and target genes experience significantly stronger negative selection relative to weakly connected transcription factors and targets. Adaptively evolving proteins were significantly more likely to reside at the periphery of the regulatory network, while proteins with signs of negative selection were near the core of the network. Interestingly, connectedness and network structure had minimal influence on the strength of selection on putative regulatory sequences for both transcription factors and their targets. Our study indicates that adaptive evolution of complex behavior can arise because of positive selection on protein-coding mutations in peripheral genes, and on regulatory sequence mutations in both transcription factors and their targets throughout the network.

Keywords: *Apis mellifera*, network hubs, natural selection, evo devo, social evolution

INTRODUCTION

Understanding the genetics and evolution of complex traits is a central goal in biology. Behavior is a complex phenotype that exhibits a high degree of variation within an individual's lifetime, within and between populations of the same species, and between species. Behavioral genetics research conducted over the past decade has emphasized the role of conserved genes in behavioral evolution. There is good evidence that behavior, like most complex phenotypes, is controlled by gene regulatory networks that exhibit modularity and pleiotropy, and that genes and gene networks that influence behavior in one organism also influence similar behaviors in evolutionary distant species (Anholt and Mackay, 2004; Reaume and Sokolowski, 2011; Zayed and Robinson, 2012). This conservation of gene action on behavior has allowed researchers to study behavioral evolution within the framework of Evolutionary Developmental Biology (i.e., evo devo) (Carroll, 2008). The synthesis of behavioral genetics and evo devo has led to many insights (Linksvayer and Wade, 2005; Toth and Robinson, 2007, 2009), including the existence of a

genetic tool kit for behavior (i.e., conserved gene modules that influence basic forms of behavior across species), and that complex behaviors can evolve through the co-option of genetic modules that control simple forms of behavior. In contrast to the evo devo paradigm, there is a burgeoning body of literature suggesting that novel taxonomically-restricted genes are important, and perhaps most prominent, in behavioral evolution (Johnson and Tsutsui, 2011; Chen et al., 2013; Ferreira et al., 2013; Simola et al., 2013; Harpur et al., 2014; Jasper et al., 2014; Sumner, 2014). Fortunately, genomics-enabled research on a variety of model and non-model organisms is providing a wealth of information on the contribution of novel and conserved genes to the genetic architecture of complex traits. Along with population genomic data on levels of selection acting on genes and regulatory sequences, evolutionary biologists are at the verge of ultimately testing the different theories of phenotypic evolution.

The different paradigms of phenotypic evolution make distinct predictions about the relative contribution of regulatory and protein-coding sequence changes. On one end of the spectrum,

the evo devo paradigm emphasizes the role of adaptive regulatory sequence evolution (Wray, 2007; Carroll, 2008) because of the assumption that genes with multiple functions, or genes that interact with other genes, are expected to experience a great deal of constraint at their amino acid sequence (Fisher, 1930). Others have challenged this central assumption of the evo devo paradigm by arguing that seemingly “conserved” proteins, including transcription factors, have several features that allow them to “escape” the constraining influence of pleiotropy thereby allowing adaptive evolution via amino-acid changing mutations (Lynch and Wagner, 2008; Wagner and Lynch, 2008); such features include alternative splicing, modularity at the level of protein domain and structure, and the presence of mutable short or simple sequence motifs. At the other end of the spectrum, there is a growing interest in novel taxonomically restricted genes that are free to evolve new functions without suffering from the constraining effect of pleiotropy (Chen et al., 2013). Empirical evidence do not fully support any one of these three paradigms over the others—there is population genetic evidence for both adaptive protein sequence evolution and adaptive coding sequence evolution in many organisms (Andolfatto, 2005; Hoekstra and Coyne, 2007; Halligan et al., 2010, 2013; Harpur et al., 2014; Wallberg et al., 2014). However, most previous tests of these paradigms involved correlating general rates of protein evolution with molecular features of genes and their position in regulatory networks (e.g., Hahn and Kern, 2005; Kim et al., 2007; Davila-Velderrain et al., 2014); data on the actual levels of positive or negative selection on coding sequences (Assis and Kondrashov, 2014) are seldom used. Moreover, we know very virtually nothing about how pleiotropy and the structure of gene regulatory networks affect patterns of regulatory sequence evolution.

The honey bee *Apis mellifera* has emerged as a model organism for studying the genetics and evolution of complex behaviors (Hunt et al., 2007; Page et al., 2012; Zayed and Robinson, 2012). Here we use several powerful genomic resources developed for the honey bee to examine if regulatory networks that influence behavior follow the predictions of the evo devo paradigm for phenotypic evolution. Chandrasekaran et al. (2011) recently constructed a brain transcriptional regulatory network (TRN) influencing several aspects of worker behavior, including behavioral maturation, foraging, and colony defense. The honey bee brain TRN is highly amenable to studies of how connectedness and network topology constrain behavioral and molecular evolution, especially given the recent availability of a large population genomic dataset for the honey bee (Harpur et al., 2014), which consists of genome wide polymorphism data for 39 *A. mellifera* diploid genomes and genome wide divergence data between *A. mellifera* and its sister species *A. cerana*.

We used the honey bee population genomic dataset to study the strength of selection on protein and putative *cis*-regulatory sequences of genes in the bee brain TRN. We tested the following hypotheses from the evo devo paradigm: (1) Highly connected TFs and target genes are predicted to experience stronger negative selection on nonsynonymous mutations relative to weakly connected TFs and target genes and (2) Genes with signs of adaptive amino acid sequence evolution are expected to be less central within the regulatory network. The evo devo paradigm does

not explicitly make predictions about the relationship between pleiotropy and regulatory sequence evolution, but rather predicts that the evolution of regulatory sequences should be less constrained relative to protein sequence evolution, and that regulatory mutations are more likely to fuel adaptive evolution. We compared the average selection coefficient on mutations in putative *cis*-regulatory regions of strongly and weakly connected genes within the TRN to explore how network properties influence regulatory sequence evolution. Our study provides an important glimpse into the evolution of regulatory networks that influence complex behaviors.

MATERIALS AND METHODS

SEQUENCING, ALIGNMENT, SNP CALLING AND MODIFIED McDONALD-KREITMAN (MK) TESTS

We recently sequenced 40 honey bee genomes, each at approximately 40X coverage, using Illumina Hi-Seq technology (Harpur et al., 2014). Alignment and polymorphism identification were described in detail by Harpur et al. (2014). We used a Bayesian implementation of the McDonald-Kreitman (MK) test, using SnIPRE (Eilertson et al., 2012), to determine the population size scaled selection coefficient γ for 12,303 genes in the honey bee genome. Here, we used the population genomics dataset to study selection acting on putative *cis*-regulatory regions of the honey bee genome. We first estimated the number of polymorphic mutations in *A. mellifera*, and the number of fixed mutations between *A. mellifera* and its sister species *A. cerana*, in putative *cis*-regulatory regions of honey bee genes. Because the regulatory sequences of the honey bee genome have not been characterized, we considered the 1000 bp sequence upstream of each gene's start codon as a putative *cis*-regulatory region (Davidson, 2006; Li et al., 2006; Myers, 2014). We excluded upstream sequences that overlapped genes encoded by the complementary DNA strand, resulting in putative *cis*-regulatory regions with an average size of 905 bp. These regions are expected to contain most of the sequences important for transcriptional and translational control, including the 5'UTR and important transcription factor binding sites (Davidson, 2006; Li et al., 2006; Myers, 2014). Our cut-off would have certainly excluded some regulatory sequences that reside far upstream of genes (Negre et al., 2011)—sequences that are currently very difficult to annotate in the honey bee. Despite this important caveat, our population genomic analyses (see results) show an overall signature of negative purifying selection within 1 Kb upstream of genes, which is consistent with such regions having a functional role related to gene regulation (Dunham et al., 2012; Wittkopp and Kalay, 2012). Following, Torgerson et al. (2009), we studied the evolution of *cis*-regulatory regions using a modified MK test by comparing the ratio of fixed:polymorphic mutations in a *cis*-regulatory sequence of a gene to same ratio for silent sites in the same gene. The modified MK test was implemented using SnIPRE (Eilertson et al., 2012), which allowed us to estimate the average population size scaled selection coefficients on regulatory sequence mutations. Similar to Harpur et al. (2014), we only used polymorphism data from African honey bee genomes, which represent a large population that is minimally impacted by human management (Harpur et al., 2012; Kent et al., 2012).

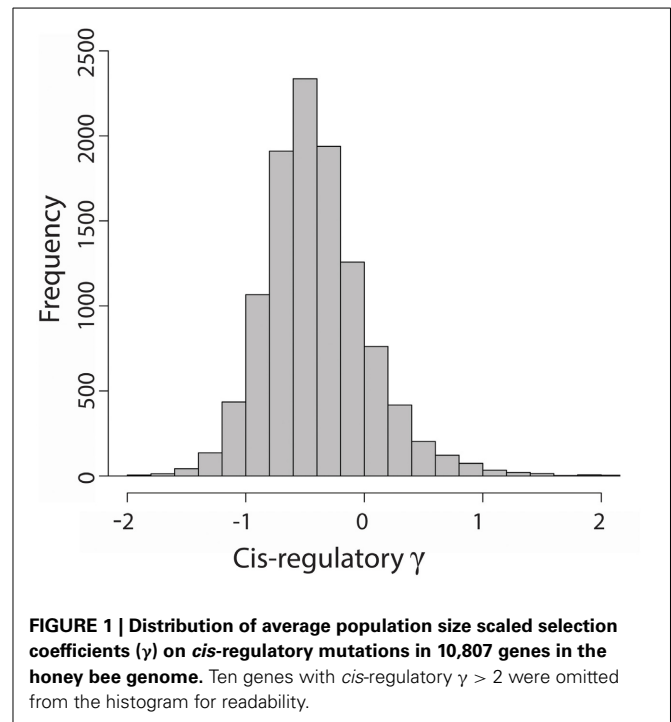
TRN CONSTRUCTION AND ANALYSIS

The honey bee brain TRN (Chandrasekaran et al., 2011) is freely available online (Web: <http://price.systemsbiology.net/honeybee-transcriptional-regulatory-network>). The dataset consisted of microarray probes for TFs and their targets in the bee brain TRN. We remapped the array probes to the honey bee's official gene set OGS v3.2 (Elsik et al., 2014) using Blastn v. 2.2.28+. We only retained probes that had perfect matches to OGS v3.2 gene predictions. We were able to blast match microarray probes to 191 transcription factors and 1597 target genes. We restricted our analyses to 184 TFs and 1521 target genes that had γ estimates for coding and putative regulatory sequences. We estimated the number of target genes for every transcription factor (k ranged from 1 to 161), and the number of transcription factors regulating every target (k ranges from 1 to 15). We plotted the regulatory network using Gephi (Bastian et al., 2009) and produced a directed graph with 1504 nodes and 5149 edges representing transcription factor—target interactions. Gephi was used to estimate *betweenness* centrality of the genes in the network. We used the R package *powerLaw* (Gillespie, 2014) to fit a power law distribution to TRN connectedness using established methods (Clauset et al., 2009). Statistical tests were carried out using R. We used a one-tailed test to compare the γ of hub and non-hub TFs and targets, given *a priori* theoretical expectations and empirical findings regarding the relationship between pleiotropy/connectedness and molecular evolution. All other p -values are two-tailed. It is important to note that the honey bee brain TRN was developed by first selecting honey bee TFs that had robust orthologs to *Drosophila* TFs (Chandrasekaran et al., 2011); the bee brain TRN is thereby enriched for old taxonomically-conserved TFs and target genes. Our study of the bee brain TRN can therefore illuminate how ancestral gene networks influencing behaviors evolve, but tell us little about the role of taxonomically-restricted genes in behavioral evolution—a topic that we recently discussed elsewhere (Harpur et al., 2014).

RESULTS

SELECTION ON REGULATORY AND CODING SEQUENCES IN THE HONEY BEE GENOME

We had previously estimated the average population size scaled selection coefficient γ on nonsynonymous mutations in 12,303 genes in the honey bee genome since divergence between *A. mellifera* and *A. cerana* (ca. 5 MYA) (Harpur et al., 2014). Here we used a variant of the MK test (Torgerson et al., 2009; implemented using Eilertson et al., 2012) to estimate the average γ on mutations in putative *cis*-regulatory sequences by comparing the ratio of polymorphic:fixed mutations within 1 kb upstream of a gene's start codon to the ratio of polymorphic:fixed synonymous mutations at the same gene. We were able to estimate γ on the putative *cis*-regulatory sequences of 10,807 genes in the honey bee genome (Figure 1). We found most (93%) *cis*-regulatory sequences to have estimates of γ consistent with neutral or nearly neutral evolution ($-1 < \gamma < 1$). About 6% of *cis*-regulatory sequences have $\gamma < -1$, indicative of negative purifying selection, while 1% of sequences have signs of positive selection ($\gamma > 1$). In contrast to evolution of protein coding sequences (average $\gamma \sim 0$), the average mutation in *cis*-regulatory regions appear to be



weakly deleterious (average $\gamma = -0.4$). This pattern was previously observed in humans (Torgerson et al., 2009) and most likely results from an observational bias: sequences from rapidly evolving regulatory regions will have many mismatches between *A. mellifera* and *A. cerana*, which results in lower alignment scores and coverage, and would have been removed from the dataset based on our quality control filters. As such, direct comparisons of the selection coefficient on coding and regulatory mutations are not appropriate. Instead, we examined the influence of a gene's connectedness and position within the TRN on regulatory and protein sequence evolution in separate analyses.

NETWORK TOPOLOGY AND EVOLUTION OF TFs AND THEIR TARGET GENES

We studied patterns of selection on coding and regulatory mutations in 170 transcription factors (TFs) and 1334 of their target genes in the honey bee brain TRN. Similar to other regulatory networks (Babu et al., 2004; Nicolau and Schoenauer, 2009), the honey bee brain TRN is approximately scale-free, whereby the distribution of connectedness (k) between the network nodes (i.e., genes) has a very long tail (Supplementary Information Figure S1). The bee brain TRN contained a large number of genes with a small number of connections, and a small number of genes with a large number of connections—often called “hub” genes. The number of connections, k , between nodes in a scale-free network follows a power law, at least above a certain value of k (Nicolau and Schoenauer, 2009). Connectedness varied between 1 and 161 in the honey bee brain TRN, and we found the tail of the connectedness distribution to follow a power law ($x_{\min} = 42$, $\alpha = 3.00$; H_0 = power law: Goodness of fit: 0.088, $p = 0.32$). We elected to analyse the dataset by categorizing genes

as hub or non-hub, following Wang et al. (2010a), because analyses based on linear models or correlations do not adequately deal with the properties of regulatory networks (i.e., the distribution of connections within the TRN is not normal). Following Wang et al. (2010a), we considered the top 20% of most connected TFs as hubs ($k > 44$ connections). Hub TFs were more central in the network as evidenced by a significantly higher estimate of eigenvector centrality relative to non-hub TFs (Wilcoxon test, $p < 2.2e-16$). We found that hub TFs had a significantly lower mean coding γ than non-hub transcription factors (Figure 2A, Wilcoxon 1-tailed $p = 0.0025$), and that hub TFs were significantly enriched for genes with negative coding γ (Chi square enrichment $p = 0.015$) relative to non-hub TFs. In contrast to coding γ , hub TFs and non-hub TFs did not significantly differ with respect to *cis*-regulatory γ (Figure 2C, Wilcoxon 1-tailed $p = 0.27$). Hub and non-hub TFs did not significantly differ in terms of sequence coverage and length at regulatory and coding sites (Supplementary Information Table 1).

Similar to TFs, we used connectedness to classify target genes in the TRN into hubs (top 20%) and non-hubs based on k . Hub target genes within the TRN were regulated by four or more TFs, and were significantly more central within the network relative to non-hub target genes (Wilcoxon $p = 2.2e-16$). Similar to the

differences between hub TFs and non-hub TFs, hub target genes had significantly lower coding γ (Figure 2B, Wilcoxon 1-tailed $p = 0.0425$), but not *cis*-regulatory γ (Figure 2D, Wilcoxon 1-tailed $p = 0.12$) relative to non-hub target genes. Hub and non-hub target genes did not significantly differ in terms of sequence coverage and length at regulatory and coding sites (Supplementary Information Table 1).

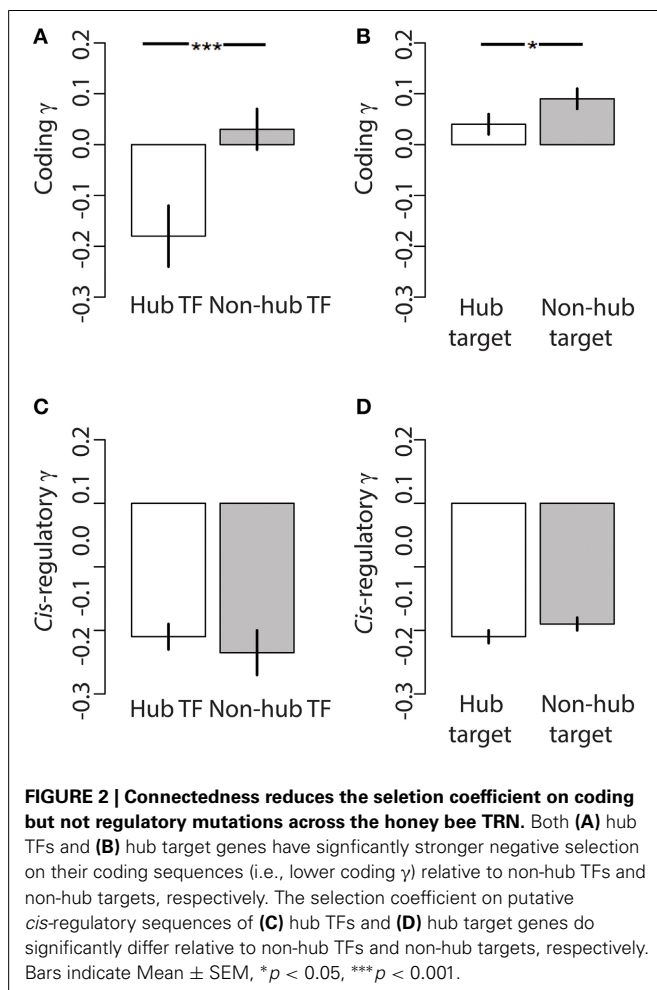
WHERE IS POSITIVE SELECTION ACTING WITHIN THE TRN?

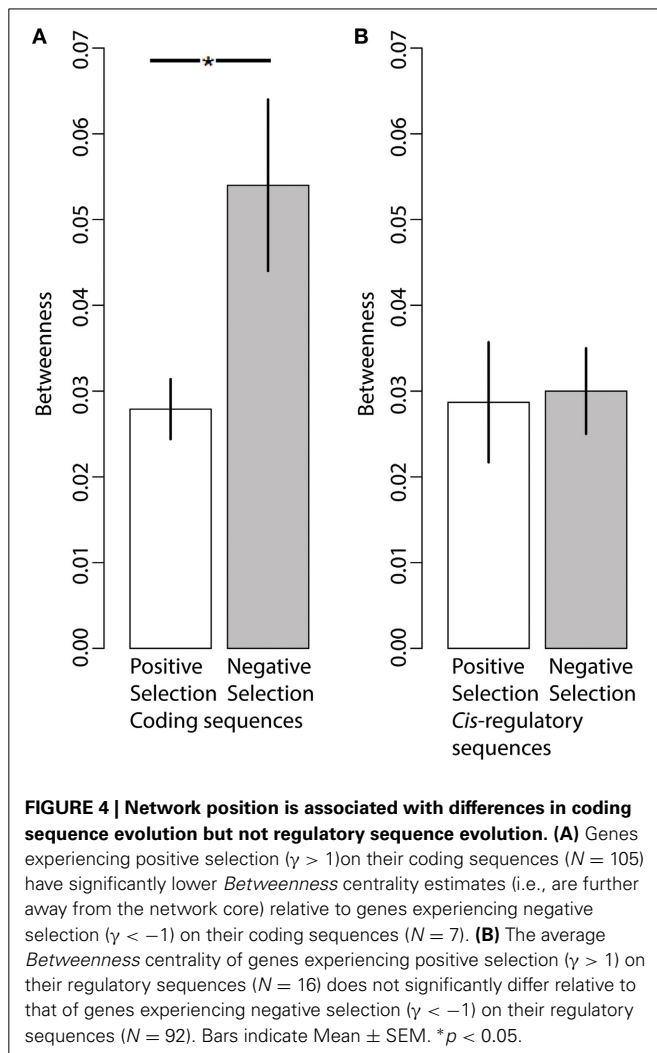
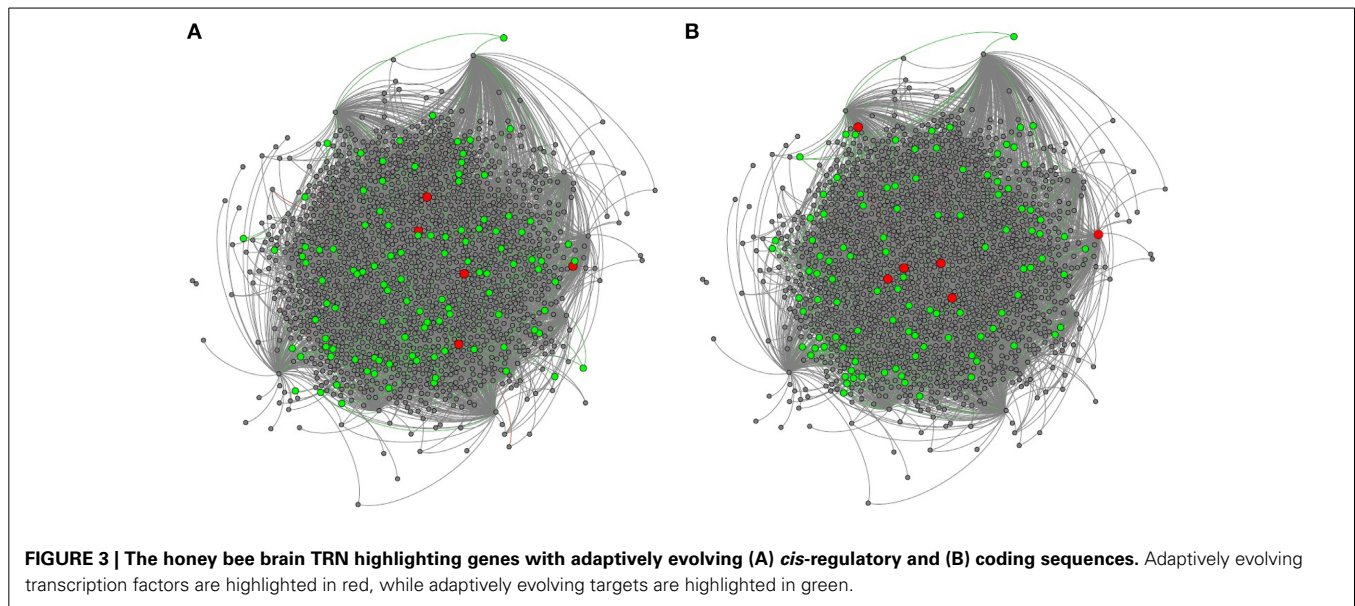
We mapped all genes with signatures of positive selection on coding and *cis*-regulatory sequences in the TRN (Figure 3). We also estimated *betweenness* for each gene in the TRN; *betweenness* is a global measure of centrality (Borgatti and Everett, 2006) which ranges from 1, indicating most central or at the core of the network, to 0, indicating the outside perimeter or the periphery of the network. We compared the average *betweenness* of genes with substantial signs of positive ($\gamma > 1$) and negative ($\gamma < -1$) selection. We found that proteins with signatures of positive selection on their coding sequences had significantly lower *betweenness* relative to proteins with signatures of negative selection, indicating that adaptively evolving proteins are often more distant from the network core relative to proteins with signs of negative selection (Figure 4A, Wilcoxon, two tailed $p = 0.04$). In contrast, we did not find a significant difference in the *betweenness* of genes with positive selection on their *cis*-regulatory sequences relative to those with negative selection on their *cis*-regulatory sequences (Figure 4B, Wilcoxon two-tailed $p = 0.4$). This indicates that genes with regulatory sequences experiencing positive selection reside in approximately the same locations within the TRN as genes with regulatory sequences experiencing negative selection.

DISCUSSION

We examined how gene position within a network influenced the average selection coefficient γ on putative *cis*-regulatory and replacement mutations in 1504 genes in the honey bee brain TRN. Our results support a “mosaic” view of phenotypic evolution by illuminating how the scale-free properties of regulatory networks (Wang et al., 2010b; Le Nagard et al., 2011; Wagner and Zhang, 2011) facilitate adaptive evolution involving both coding and regulatory mutations.

Several lines of evidence suggest that the most connected, and likely most pleiotropic, proteins within the bee brain TRN experience the greatest levels of purifying selection, as predicted by Fisher (1930) and the Evo Devo paradigm (Carroll, 2008). Despite the large number of factors that influence the rate of molecular evolution of genes (Xia et al., 2009) we consistently found that the most connected genes in the TRN had the strongest signatures of negative selection on their coding sequence. In brief, transcription factors that regulate hundreds of target genes experience, on average, stronger negative selection on their coding sequence relative to transcription factors that regulate a few target genes (Figure 2A). Hub transcription factors likely have to interact with many other co-factors, in addition to binding target promoter sites, which may be responsible for the stronger levels of purifying selection on their amino acid sequence. Similar to hub transcription factors, hub target genes that are regulated by





many transcription factors experience stronger negative selection on their coding sequence relative to target genes that are regulated by a few transcription factors. Target genes that are regulated by multiple TFs may be expressed in multiple tissues or during multiple contexts relative to target genes regulated by a few TFs, resulting in greater pleiotropy and stronger purifying selection, as evident from our analysis (**Figure 2B**). It is important to note that several genes within the TRN had signs of adaptive protein evolution; most of these genes were transcription factor targets, and most resided near the periphery of the TRN. Lynch and Wagner (2008) and Wagner and Lynch (2008) previously argued that proteins, including conserved TFs, have features that allow them to escape from the negative effects of pleiotropy. Our population genomic data are not fully consistent with the Lynch and Wagner hypotheses because the most central and most connected TFs or targets do experience stronger levels of negative selection vs. peripheral and weakly connected TFs or targets; a relationship that is more inline with the classic evo devo paradigm. We strongly believe that the structure of TRNs hold the key for reconciling the predictions of the evo devo paradigm with the empirical data showing that amino-acid changes do contribute to adaptive evolution. The classic evo devo paradigm assumes that *most* genes are constrained by pleiotropy, while studies of TRN structure clearly show that only a *few* genes are highly connected and central, while *most* genes are weakly connected and peripheral. Although pleiotropy does appear to curtail adaptive protein sequence evolution of the *few* most connected and most central genes within a TRN, adaptive protein evolution is still a powerful evolutionary force for *most* TRN genes that reside *at* the network periphery.

In stark contrast to the influence of TRN topology on protein coding evolution, we found that connectedness matters little with respect to levels of selection on putative *cis*-regulatory regions. The average selection coefficient on regulatory sequence mutations of hub transcription factors was similar to

that of non-hub transcription factors (Figure 2C). Similarly, the selection coefficient on regulatory sequences of hub target genes was similar to those of non-hub target genes. Genes with signs of adaptive regulatory sequence evolution were found in similar locations within the TRN as genes with negative selection on their regulatory sequences. Our analysis indicates that network properties do not significantly shape the selection pressures acting on regulatory sequences within the TRN. It is not clear how this evidence support the evo devo paradigm because the evo devo paradigm does not make explicit predictions about the relationship between pleiotropy, connectedness and regulatory sequence evolution. On one hand, our finding that putative *cis*-regulatory sequences evolve independently of TRN connectedness and topology appears to support an important assumption of the evo devo paradigm: pleiotropy or connectedness of a protein only influences the protein's amino acid sequence, not its *cis*-regulatory sequence. On the other hand, another interpretation of the evo devo paradigm suggests that the most connected and pleiotropic genes should have the greatest levels of adaptive regulatory evolution, while the least connected genes should have the least levels of adaptive regulatory evolution (i.e., regulatory sequence evolution compensates for constrained amino acid sequences); our findings do not support this idea. It would appear that adaptive regulatory sequence evolution can occur throughout any compartment of the regulatory network.

Our analyses shed light on the evolution of regulatory networks influencing complex behavior. Highly connected genes within the honey bee brain TRN exhibit stronger patterns of purifying selection on amino acid replacement mutations similar to highly connected genes in other types of networks studied so far. Also, genes with signs of adaptive protein evolution tend to be concentrated at the network periphery, as previously documented for proteins in the Human Interactome (Kim et al., 2007). We found that connectedness does not influence the strength of selection on regulatory sequences of genes in the bee brain TRN. Our study suggests that the properties of regulatory networks, with a few large modules and many small modules, allows for both coding and regulatory sequence mutations to contribute to adaptive evolution. Based on our findings, we expect adaptive evolution of regulatory networks influencing complex traits to proceed through positive selection on coding mutations in peripheral genes and on regulatory mutations in TFs and their targets across the regulatory network. We had previously presented strong evidence that novel taxonomically-restricted genes have the highest rates of adaptive protein evolution in the honey bee genome (Harpur et al., 2014). A recent analysis also pointed to an increased expansion of regulatory sequences in social genomes (Simola et al., 2013). Going forward, it will be important to study how novel taxonomically restricted genes interact with conserved TRN modules with expanded regulatory features to influence the evolution of complex behaviors in social insects.

ACKNOWLEDGMENTS

We thank K. Eilertson for help with SnIPRE, the N. Price and G. Robinson groups for constructing the honey bee brain TRN and for making this valuable dataset publically available, and

two anonymous referees for providing insightful comments on our manuscript. This study was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada and an Early Researcher Award from the Ontario Ministry of Research and Innovation (to Amro Zayed). Brock A. Harpur was supported by an Elia Doctoral Scholarship from York University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2014.00431/abstract>

Figure S1 | Degree distribution of connectedness in the TRN.

REFERENCES

- Andolfatto, P. (2005). Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437, 1149–1152. doi: 10.1038/nature04107
- Anholt, R. R. H., and Mackay, T. F. C. (2004). Quantitative genetic analyses of complex behaviours in *Drosophila*. *Nat. Rev. Genet.* 5, 838–849. doi: 10.1038/nrg1472
- Assis, R., and Kondrashov, A. S. (2014). Conserved proteins are fragile. *Mol. Biol. Evol.* 31, 419–424. doi: 10.1093/molbev/mst217
- Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M., and Teichmann, S. A. (2004). Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* 14, 283–291. doi: 10.1016/j.sbi.2004.05.004
- Bastian, M., Heymann, J., and Jacomy, M. (2009). “Gephi: an open source software for exploring and manipulating networks,” in *International AAAI Conference on Weblogs and Social Media* (Paris).
- Borgatti, S. P., and Everett, M. G. (2006). A graph-theoretic perspective on centrality. *Soc. Netw.* 28, 466–484. doi: 10.1016/j.socnet.2005.11.005
- Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36. doi: 10.1016/j.cell.2008.06.030
- Chandrasekaran, S., Ament, S. A., Eddy, J. A., Rodriguez-Zas, S. L., Schatz, B. R., Price, N. D., et al. (2011). Behavior-specific changes in transcriptional modules lead to distinct and predictable neurogenomic states. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18020–18025. doi: 10.1073/pnas.1114093108
- Chen, S. D., Krinsky, B. H., and Long, M. Y. (2013). New genes as drivers of phenotypic evolution. *Nat. Rev. Genet.* 14, 645–660. doi: 10.1038/nrg3521
- Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *Siam Rev.* 51, 661–703. doi: 10.1137/070710111
- Davidson, E. H. (2006). *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. Burlington, VT; San Diego, CA: Academic.
- Davila-Velderrain, J., Servin-Marquez, A., and Alvarez-Buylla, E. R. (2014). Molecular evolution constraints in the floral organ specification gene regulatory network module across 18 angiosperm genomes. *Mol. Biol. Evol.* 31, 560–573. doi: 10.1093/molbev/mst223
- Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C., Doyle, F., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247
- Eilertson, K. E., Booth, J. G., and Bustamante, C. D. (2012). SnIPRE: selection inference using a poisson random effects model. *PLoS Comput. Biol.* 8:E1002806. doi: 10.1371/journal.pcbi.1002806
- Elsik, C. G., Worley, K. C., Bennett, A. K., Beyre, M., Camara, F., Childers, C. P., et al. (2014). Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* 15:86. doi: 10.1186/1471-2164-15-86
- Ferreira, P. G., Patalano, S., Chauhan, R., Ffrench-Constant, R., Gabaldon, T., Guigo, R., et al. (2013). Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol.* 14:R20. doi: 10.1186/gb-2013-14-2-r20
- Fisher, R. A. (1930). *The Genetic Theory of Natural Selection*. New York, NY: Dover.
- Gillespie, C. S. (2014). Fitting heavy tailed distributions: the powerlaw package. *R package version 0.20.25* (Newcastle, UK).

- Hahn, M. W., and Kern, A. D. (2005). Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.* 22, 803–806. doi: 10.1093/molbev/msi072
- Halligan, D. L., Kousathanas, A., Ness, R. W., Harr, B., Eory, L., Keane, T. M., et al. (2013). Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet.* 9:e1003995. doi: 10.1371/journal.pgen.1003995
- Halligan, D. L., Oliver, F., Eyre-Walker, A., Harr, B., and Keightley, P. D. (2010). Evidence for pervasive adaptive protein evolution in wild mice. *PLoS Genet.* 6:e1000825. doi: 10.1371/journal.pgen.1000825
- Harpur, B. A., Kent, C. F., Molodtsova, D., Lebon, J. M., Alqarni, A. S., Owayss, A. A., et al. (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. doi: 10.1073/pnas.1315506111
- Harpur, B. A., Minaei, S., Kent, C. F., and Zayed, A. (2012). Management increases genetic diversity of honey bees via admixture. *Mol. Ecol.* 21, 4414–4421. doi: 10.1111/j.1365-294X.2012.05614.x
- Hoekstra, H. E., and Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61, 995–1016. doi: 10.1111/j.1558-5646.2007.00105.x
- Hunt, G. J., Amdam, G. V., Schlipalius, D., Emore, C., Sardesai, N., Williams, C. E., et al. (2007). Behavioral genomics of honeybee foraging and nest defense. *Naturwissenschaften* 94, 247–267. doi: 10.1007/s00114-006-0183-1
- Jasper, W. C., Linksvayer, T. A., Atallah, J., Friedman, D., Chiu, J. C., and Johnson, B. R. (2014). Large scale coding sequence change underlies the evolution of post-developmental novelty in honey bees. *Mol. Biol. Evol.* doi: 10.1093/molbev/msu292
- Johnson, B. R., and Tsutsui, N. D. (2011). Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* 12:164. doi: 10.1186/1471-2164-12-164
- Kent, C. F., Minaei, S., Harpur, B. A., and Zayed, A. (2012). Recombination is associated with the evolution of genome structure and worker behavior in honey bees. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18012–18017. doi: 10.1073/pnas.1208094109
- Kim, P. M., Korb, J. O., and Gerstein, M. B. (2007). Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20274–20279. doi: 10.1073/pnas.0710183104
- Le Nagard, H., Chao, L., and Tenaillon, O. (2011). The emergence of complexity and restricted pleiotropy in adapting networks. *BMC Evol. Biol.* 11:326. doi: 10.1186/1471-2148-11-326
- Li, Y., Lee, K. K., Walsh, S., Smith, C., Hadingham, S., Sorefan, K., et al. (2006). Establishing glucose- and ABA-regulated transcription networks in Arabidopsis by microarray analysis and promoter classification using a Relevance Vector Machine. *Genome Res.* 16, 414–427. doi: 10.1101/gr.4237406
- Linksvayer, T. A., and Wade, M. J. (2005). The evolutionary origin and elaboration of sociality in the aculeate Hymenoptera: maternal effects, sib-social effects, and heterochrony. *Q. Rev. Biol.* 80, 317–336. doi: 10.1086/432266
- Lynch, V. J., and Wagner, G. P. (2008). Resurrecting the role of transcription factor change in developmental evolution. *Evolution* 62, 2131–2154. doi: 10.1111/j.1558-5646.2008.00440.x
- Myers, A. J. (2014). “The genetics of gene expression: multiple layers and multiple players,” in *The OMICS: Applications in Neuroscience*, ed G. Coppola (Oxford: Oxford University Press), 132–154.
- Negre, N., Brown, C. D., Ma, L. J., Bristow, C. A., Miller, S. W., Wagner, U., et al. (2011). A cis-regulatory map of the *Drosophila* genome. *Nature* 471, 527–531. doi: 10.1038/nature09990
- Nicolau, M., and Schoenauer, M. (2009). On the evolution of scale-free topologies with a gene regulatory network model. *Biosystems* 98, 137–148. doi: 10.1016/j.biosystems.2009.06.006
- Page, R. E., Rueppell, O., and Amdam, G. V. (2012). Genetics of reproduction and regulation of honeybee (*Apis mellifera* L.) social behavior. *Annu. Rev. Genet.* 46, 97–119. doi: 10.1146/annurev-genet-110711-155610
- Reaume, C. J., and Sokolowski, M. B. (2011). Conservation of gene function in behaviour. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 2100–2110. doi: 10.1098/rstb.2011.0028
- Simola, D. F., Wissler, L., Donahue, G., Waterhouse, R. M., Helmkamp, M., Roux, J., et al. (2013). Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res.* 23, 1235–1247. doi: 10.1101/gr.155408.113
- Sumner, S. (2014). The importance of genomic novelty in social evolution. *Mol. Ecol.* 23, 26–28. doi: 10.1111/mec.12580
- Torgerson, D. G., Boyko, A. R., Hernandez, R. D., Indap, A., Hu, X., White, T. J., et al. (2009). Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. *PLoS Genet.* 5:e1000592. doi: 10.1371/journal.pgen.1000592
- Toth, A. L., and Robinson, G. E. (2007). Evo-devo and the evolution of social behavior. *Trends Genet.* 23, 334–341. doi: 10.1016/j.tig.2007.05.001
- Toth, A. L., and Robinson, G. E. (2009). Evo-devo and the evolution of social behavior: brain gene expression analyses in social insects. *Cold Spring Harb. Symp. Quant. Biol.* 74, 419–426. doi: 10.1101/sqb.2009.74.026
- Wagner, G. P., and Lynch, V. J. (2008). The gene regulatory logic of transcription factor evolution. *Trends Ecol. Evol.* 23, 377–385. doi: 10.1016/j.tree.2008.03.006
- Wagner, G. P., and Zhang, J. (2011). The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* 12, 204–213. doi: 10.1038/nrg2949
- Wallberg, A., Han, F., Wellhagen, G., Dahle, B., Kawata, M., Haddad, N., et al. (2014). A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nat. Genet.* Available Online at: <http://www.nature.com/ng/journal/vaop/ncurrent/abs/ng.3077.html> - supplementary information.
- Wang, Y., Franzosa, E. A., Zhang, X. S., and Xia, Y. (2010a). Protein evolution in yeast transcription factor subnetworks. *Nucleic Acids Res.* 38, 5959–5969. doi: 10.1093/nar/gkq353
- Wang, Z., Liao, B. Y., and Zhang, J. (2010b). Genomic patterns of pleiotropy and the evolution of complexity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18034–18039. doi: 10.1073/pnas.1004666107
- Wittkopp, P. J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13, 59–69. doi: 10.1038/nrg3095
- Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8, 206–216. doi: 10.1038/nrg2063
- Xia, Y., Franzosa, E. A., and Gerstein, M. B. (2009). Integrated assessment of genomic correlates of protein evolutionary rate. *PLoS Comput. Biol.* 5:e1000413. doi: 10.1371/journal.pcbi.1000413
- Zayed, A., and Robinson, G. E. (2012). Understanding the relationship between brain gene expression and social behavior: lessons from the honey bee. *Annu. Rev. Genet.* 46, 591–615. doi: 10.1146/annurev-genet-110711-155517

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 October 2014; accepted: 21 November 2014; published online: 23 December 2014.

Citation: Molodtsova D, Harpur BA, Kent CF, Seevananthan K and Zayed A (2014) Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviors. *Front. Genet.* 5:431. doi: 10.3389/fgene.2014.00431

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Molodtsova, Harpur, Kent, Seevananthan and Zayed. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Neutral and adaptive explanations for an association between caste-biased gene expression and rate of sequence evolution

Heikki Helanterä^{1*} and Tobias Uller^{2,3}

¹ Department of Biosciences, Centre of Excellence in Biological Interactions, University of Helsinki, Helsinki, Finland

² Department of Zoology, Edward Grey Institute, University of Oxford, Oxford, UK

³ Department of Biology, University of Lund, Sölvegatan, Lund, Sweden

Edited by:

Greg J. Hunt, Purdue University, USA

Reviewed by:

Jacob A. Tennesen, Oregon State University, USA

Alex Wong, Carleton University, Canada

*Correspondence:

Heikki Helanterä, Department of Biosciences, Centre of Excellence in Biological Interactions, University of Helsinki, PO Box 65, FI-00014 Helsinki, Finland
e-mail: heikki.helantera@helsinki.fi

The castes of social insects provide outstanding opportunities to address the causes and consequences of evolution of discrete phenotypes, i.e., polymorphisms. Here we focus on recently described patterns of a positive association between the degree of caste-specific gene expression and the rate of sequence evolution. We outline how neutral and adaptive evolution can cause genes that are morph-biased in their expression profiles to exhibit historical signatures of faster or slower sequence evolution compared to unbiased genes. We conclude that evaluation of different hypotheses will benefit from (i) reconstruction of the phylogenetic origin of biased expression and changes in rates of sequence evolution, and (ii) replicated data on gene expression variation within versus between morphs. Although the data are limited at present, we suggest that the observed phylogenetic and intra-population variation in gene expression lends support to the hypothesis that the association between caste-biased expression and rate of sequence evolution largely is a result of neutral processes.

Keywords: polymorphism, social insects, phenotypic plasticity, neutral evolution, antagonistic pleiotropy

INTRODUCTION

Polymorphic populations are comprised of distinct interbreeding phenotypes. These include males and females, alternative male mating morphs and different forms of resource, dispersal, or defense polymorphisms (West-Eberhard, 2003). Notably, polymorphisms also include the social insect castes, such as queens and workers. Different forms of polymorphisms vary in several respects (Box 1), but they have in common that determination and maintenance of morph-specific phenotypes involve differential expression of genes. Polymorphisms are outstanding model systems for studying the relative importance of changes in regulatory versus coding sequences (Carroll, 2005; Hoekstra and Coyne, 2007), and the interchangeability of genes and environments in phenotypic evolution (West-Eberhard, 2003; Schwander and Leimar, 2011; Uller and Helanterä, 2011).

The recent increase in availability of large scale gene expression data through microarray and whole transcriptome sequencing has facilitated quantitative and qualitative description of the developmental genetic basis of polymorphism. In social insects, caste biased expression patterns have been investigated transcriptome wide in, for example, honeybees *Apis mellifera* (Grozinger et al., 2007), bumblebees *Bombus terrestris* (Colgan et al., 2011), *Polistes* wasps (Sumner et al., 2006; Ferreira et al., 2013) and ants such as *Solenopsis invicta* (Hunt et al., 2011, 2013) and *Temnothorax longispinosus* (Feldmeyer et al., 2014). Comparisons can be made with data from other polymorphic systems, including males versus females of laboratory model species

with genotypic sex determination (e.g., mice and *Drosophila* spp., (Ranz et al., 2003; Zhang et al., 2007; Mank et al., 2008; Meisel, 2011), and a variety of environmentally induced polymorphisms, such as horn polymorphism in beetles (Snell-Rood et al., 2011), feeding type polymorphism in toad tadpoles (Leichty et al., 2012), and dispersal polymorphisms in pea aphids (Purandare et al., 2014). Sequencing methods, sampling design, pooling of samples, statistical power and definitions of what qualifies as a morph biased expression pattern vary extensively among studies, and the proportions of genes or transcripts that are classified as morph-biased range from a few to several tens of percents. For example, between 7.5% (Hunt et al., 2011) and 40% (Grozinger et al., 2007) of studied genes have been classified as caste biased in social insects. In males and females, Naurin et al. (2011) describe only 1.6–2.4% of genes as sex biased in two bird species, whereas as many as 90% of genes were reported to exhibit sex biased expression in a study of *Drosophila* (Innocenti and Morrow, 2010). Approximately half of the genes in the pea aphid show a biased expression pattern according to either morph or sex (Purandare et al., 2014). However, across all these examples, very few genes, if any, are exclusively expressed in one morph. That is, gene expression patterns vary between morphs in degree, and not in an on-or-off manner.

Morph-biased gene expression has a wide range of causes and consequences that are of interest to developmental and evolutionary biologists. In the rest of this commentary we focus on the intriguing observation that worker or queen biased genes

Box 1 | Animal polymorphism.

Different forms of polymorphisms share many features, but there are also important differences that may affect how the developmental genetic regulation of their determination and function will evolve. This is an area that would benefit strongly from theory and comparative analyses, and here we can only provide a brief summary of some morph features that may be important for the relationship between patterns of gene expression and sequence evolution.

All morphs are by definition discrete phenotypes, but the degree to which morphs differ from each other varies dramatically between systems. Queens and workers are among their more extreme polymorphisms, but even within social insects the extent to which they differ morphologically (e.g., size and shape) and physiologically (e.g., reproductive activity, lifespan) is quite variable (Bourke, 1999). In some systems where morphs represent, for example, alternative reproductive strategies, it is not necessarily the case that the average difference in gene expression between morphs is greater than differences within morphs (e.g., throughout the season). Other morphs are not even functionally different and hence might only differ consistently in gene expression during morph determination (e.g., some color or pattern morphs).

Morphs also sometimes differ genetically. The most familiar example is the sexes in mammals. It is unclear if and how the extent to which morph determination relies on the presence of specific genes versus specific environments should affect the evolution of morph-biased gene expression. Comparisons of closely related species with environment- versus genotype-dependent sex or caste determination would be informative. Related to this is the extent to which genome evolution in species with genotypic morph determination parallels that of sex chromosome evolution, which not only affects morph-specific gene content but also the extent of antagonistic selection across the genome (Connallon and Clark, 2010).

Polymorphisms can be maintained by different forms of selection. In some cases the average fitness of morphs may be equal when averaged across contexts, for example due to frequency-dependent selection. In other cases one morph is adopted under poor conditions and is therefore maintained also when it exhibits consistently lower fitness on average. The selective dynamics affect the frequency of morphs within populations and hence the strength of selection on genes with morph-specific expression or function (Van Dyken and Wade, 2010). But it is also possible that different forms of selection create particular signatures in terms of sequence evolution, for example their tendency to maintain nucleotide polymorphism within populations (Nielsen, 2005). The outcome of these processes will also be affected by the age of polymorphisms. For example, we suggest that antagonistic selection may be more common in the early stages of morph evolution when resolution of antagonism through expression patterns has not had time to evolve yet, and neutral evolution of both sequence and expression pattern more common in highly canalized morphs.

Morphs are defined at the level of the individual organism. However, in social insects where the colony, or even the supercolony, can function as an individual organism (Queller and Strassmann, 2009), it is possible that morph-biased gene expression is not analogous to, for example, sex-specific gene expression, but more similar to the differential expression of genes among tissues. Some predictions for the rate of sequence evolution of tissue-specific and morph-specific genes compared to constitutively expressed genes are shared, but others differ (in particular when not all morphs are capable of reproduction). Analogously to genes that are over-expressed in one tissue, to understand the actual strength of selection on worker-biased genes it may be necessary to understand how this expression pattern contributes to the performance of the colony (i.e., the reproductive unit).

in ants and social bees appear to evolve faster at the sequence level than do genes with no expression bias (Hunt et al., 2010, 2011; Feldmeyer et al., 2014). This is not just a social insect phenomenon. For example, it has repeatedly been shown in fruit flies and mice that both male and female biased genes evolve faster than unbiased genes (for recent studies see Meisel, 2011; Assis et al., 2012; Grath and Parsch, 2012, reviewed in Parsch and Ellegren, 2013). The same pattern has also been found with respect to sex-specific reproductive functions in *Arabidopsis* (Gossmann et al., 2014), tadpole feeding morphs in spadefoot toads (Leichty et al., 2012), horn polyphenisms in beetles (Snell-Rood et al., 2011), and dispersal morphs in pea aphids (Purandare et al., 2014).

There are a number of potential explanations for these patterns. To the extent that the morphs reflect different reproductive roles, as is the case for males and females, queens and workers, and dispersing sexuals and sedentary asexuals, faster sequence evolution of biased genes can partly be explained by faster evolution of reproductive genes (Meisel, 2011; Wright and Mank, 2013). Fast evolution of reproductive genes has been attributed to sexual selection, including sexual conflict and sperm competition, which is expected to increase the rate of sequence evolution (Swanson and Vacquier, 2002). Similar arguments should apply to the various reproductive conflicts in insect societies (Rice and

Holland, 1997). Rapid evolution of reproduction-related genes may also be partly due to faster evolution of tissue specific genes when the observed sex biases arise from genes that are expressed in sex-specific reproductive tissues (Meisel, 2011). Furthermore, when interpreting results of transcriptomes of whole individuals, it needs to be kept in mind that they may reflect differences in the size or composition of tissues between morphs, rather than differences in gene expression at a cell level. More generally, studies in model organisms have established that a large number of additional factors could underlie a correlation between gene expression patterns and evolutionary rates, including expression breadth, overall expression levels, DNA methylation patterns, architecture of regulatory sequences, and potential for pleiotropy (Lemos et al., 2005a; Larracuente et al., 2008; Meisel, 2011; Park et al., 2012; Warnefors and Kaessmann, 2013). Teasing these explanations apart is major challenge in emerging model organisms such as social insects. However, the correlation between morph biased expression pattern and fast sequence evolution rate remains significant even if many of these are controlled for statistically (e.g., Snell-Rood et al., 2011; Grath and Parsch, 2012; Warnefors and Kaessmann, 2013), suggesting that this relationship may be a fundamental feature of the evolution of genomes.

A recent study in fire ants suggested that caste-biased genes evolved faster at sequence level even before they became

morph-biased or, indeed, before the evolution of the castes (as shown by comparisons with the solitary, monomorphic wasp *Nasonia vitripennis*; Hunt et al., 2011). Interestingly, a similar pattern was found in toads where repeated evolution of polymorphism has taken place within a single genus (Leichty et al., 2012). A second interesting finding is that genes that vary in expression levels among morphs also seem to vary extensively in their expression levels within morphs. This is particularly well documented in fire ants (Hunt et al., 2013), and appears to occur also in polymorphic toads (Leichty et al., 2012), and sex biased genes in birds and fruit flies (Mank et al., 2007; Mank and Ellegren, 2009). As described below, both of these observations provide important pieces of evidence for addressing the role of neutral and selective explanations for associations between biased gene expression and rate of sequence evolution across the genome.

In this paper, we provide an overview of five different scenarios that predict a relationship between caste-biased gene expression and accelerated sequence evolution. We draw from both studies of sex biased gene expression in model organisms and from the diverse but less studied polymorphic organisms. We conclude that the association between caste-biased gene expression and rate of sequence evolution will be better understood if we address the contribution of selective and neutral processes for both inter-individual and phylogenetic divergence in gene expression. This requires both more detailed analyses of individual and context-dependent variation in gene expression, and establishing whether the strength of selection on gene sequences is a cause or a consequence of changing patterns of gene expression.

ROUTES TO COUPLING OF MORPH-BIASED EXPRESSION AND RATE OF SEQUENCE EVOLUTION

Contemporary patterns of variation in DNA sequence and gene expression partly reflect a mix of selective and stochastic events accumulating over evolutionary time (as well as current conditions experienced by the focal individuals). The relative importance of neutral and adaptive evolution for genome evolution is a contentious issue. DNA sequences diverge as a result of accumulation of changes that are neutral with respect to fitness or too weakly selected to be purged, but they also diverge because of repeated fixation of mutations due to selection. Similarly, divergence in gene regulation can represent both selection and drift. Here we ask how these processes can cause genes with caste-biased expression to exhibit evidence of accelerated sequence evolution.

NEUTRALITY

Assuming morphs are adaptive, at least some morph-bias in gene expression is a result of selection. However, in many species the number of genes that contribute to morph determination or maintenance of morph-specific phenotypes may be quite small. Thus, it is possible that a large proportion of variation in gene expression between morphs is a result of neutral evolution. Genes whose expression level is under weak selection are expected to be less precisely regulated due to accumulation of near-neutral regulatory mutations (Khaitovich et al., 2005, 2006). By chance, some of the accumulating regulatory mutations may result in morph-biased expression. Thus, morph specific expression patterns can

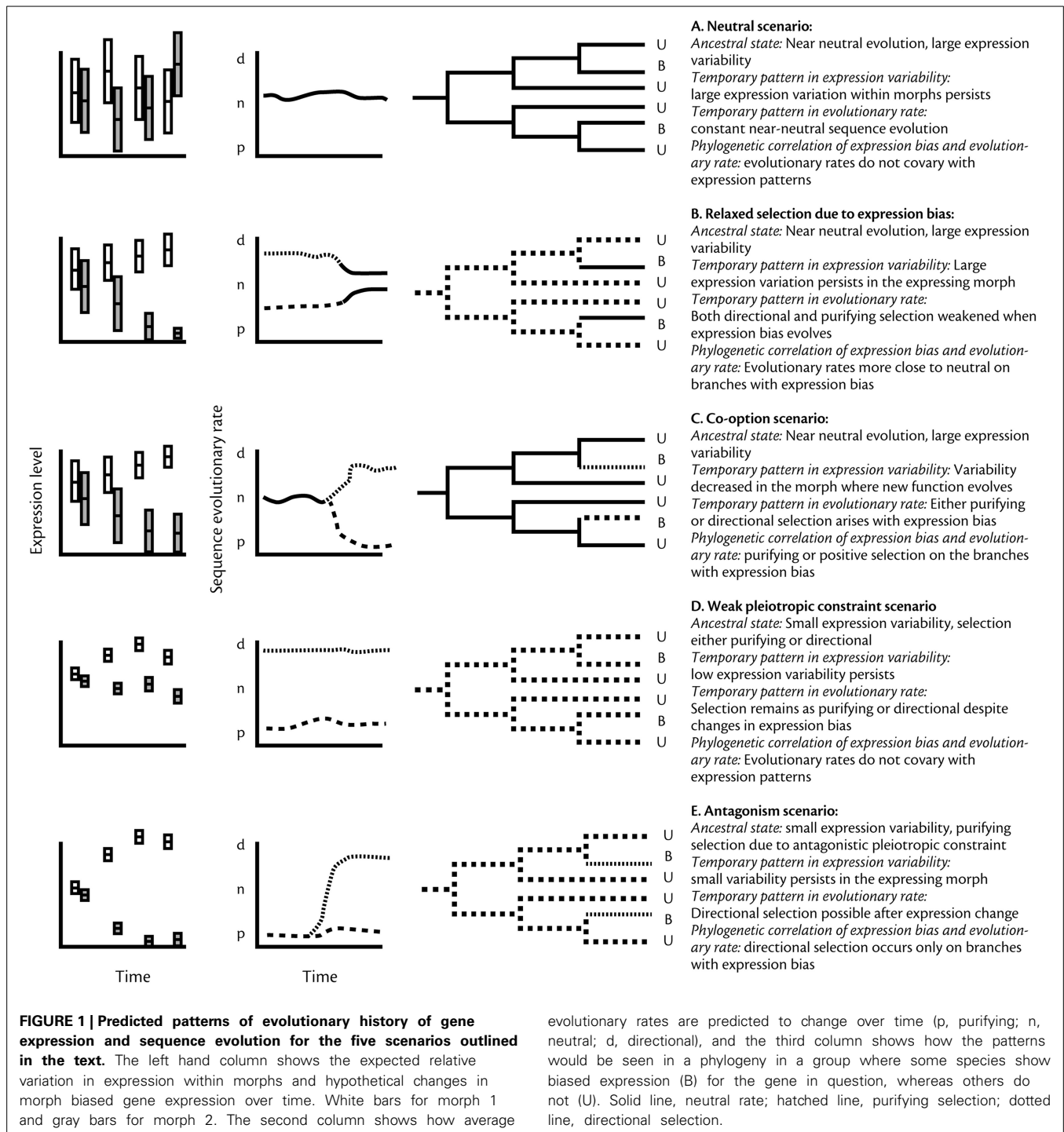
arise through drift. This suggests that, in any given data set, morph-biased expression partly reflects a history of weak purifying selection on gene regulation. This can create a link between rate of sequence evolution and biased gene expression if genes that are under weak selection with respect to sequence are also under weak selection in terms of expression, and consequently more likely to have a drifted toward biased expression pattern than constrained genes. This is likely to often be the case given the observed correlations between gene essentiality and expression noise (Fraser et al., 2004), and expression divergence and sequence divergence among species (Lemos et al., 2005b; Zhang et al., 2007; Mcmanus et al., 2010).

This process should result in the pattern observed in *S. invicta* (Hunt et al., 2011), where genes that presently exhibit morph-biased expression evolved faster even before the evolution of morphs or before morph biased expression pattern arose (Figure 1). Also, since it implies that regulation of expression is not under strong selection or constraint, we expect such genes to show substantial variation in their expression both between and within morphs—a pattern also shown in *S. invicta*. The temporal and phylogenetic patterns of sequence and expression evolution that would correspond to this scenario are shown in Figure 1A. Assessing neutrality of gene expression is compromised by the lack of a widely accepted neutral baseline (comparable to comparison of synonymous and non-synonymous amino acid changes in sequence data), but theory is advancing fast in this area and several alternatives have recently been proposed (Gout et al., 2010; Warnefors and Eyre-Walker, 2012; Rohlf et al., 2014).

Relaxed selection due to expression bias

Interpreting a correlation between expression bias and evolutionary rate in social insects is complicated by the fact that genes expressed in workers only have indirect effects mediated through the queen genotype (Linksvayer and Wade, 2009; Hall and Goodisman, 2012). The strength of selection on worker biased genes thus depends on the genetic similarity or kinship between the worker expressing a gene, and the queen that is reproducing in the nest. Relaxed selection may also occur in polymorphic species where all morphs reproduce. A very general model of polymorphic expression suggests that, all else being equal, expression bias itself may directly contribute to evolutionary rate (Snell-Rood et al., 2010; Van Dyken and Wade, 2010). This is because once expression of a gene falls below functional levels in some individuals (e.g., one of several morphs), those genes are under direct selection in a subset of the population and hence under weaker selection than constitutively expressed genes. Thus, both directional and purifying selection become relaxed following morph-biased gene expression, which allows mildly deleterious alleles to accumulate at a higher rate than is the case for constitutively expressed genes. In one such scenario, a gene under weak selection may drift to non-detectable expression levels in one morph (assuming the strength of selection on sequence and expression are correlated as in Figure 1A), leading to further relaxation of selection and hence neutral sequence evolution (Figure 1B).

Although this is an attractive hypothesis, the extremely low number of genes that are morph-specific, rather than simply



morph-biased, may suggest that few genes in fact are under relaxed selection. However, if we assume that there is a general correlation between expression level and strength of selection (Pál et al., 2001; Lemos et al., 2005a; Meisel, 2011), or that genes with a low expression level may be below a threshold value for being functional, the logic holds for genes with non-zero expression. Consistent with these predictions, genes expressed in rarer morphs of pea aphids appear to evolve faster due to relaxed

purifying selection than genes biased toward the more common morphs (Purandare et al., 2014). Furthermore, the findings that both queen biased and worker biased genes evolve faster than unbiased genes (Hunt et al., 2010, 2011) suggests that fast evolution is not only due to reproductive genes (expressed in queens) evolving fast, which is broadly consistent with the general theory of relaxed selection. Nevertheless, widespread positive selection on worker biased genes in honeybees (Harpur et al.,

2014) suggests that relaxed selection is not necessarily a major force limiting adaptive evolution of genes with worker biased expression.

This relaxation of selection due to gene copies in non-expressing or non-reproducing individuals being invisible to selection should apply to all genes with extreme expression bias and not only those that drift to this situation. For genes that are under positive selection before morph bias evolves, evolution could slow down and approach neutrality due to morph bias, whereas genes historically under purifying selection could start accumulating mutations and shift toward neutrality due to weakened purifying selection with morph bias. As a result, many different patterns of historical signatures of sequence evolution are possible.

SELECTION

Consistent stabilizing (purifying) or directional (positive) selection can generate both slower and faster rates of sequence evolution compared to the neutral expectation. Harmful mutations in genes whose sequence is essential for organismal function are rapidly purged, resulting in slow rates of evolutionary change. On the other hand, mutations in genes that cause functional changes to phenotypes can be consistently and repeatedly selected if conditions change, such as the case in evolutionary “arms races.” Thus, if these patterns of selection covary with expression patterns, it could contribute to the observed relationship between caste-biased gene expression and the rate of gene sequence divergence.

Co-option of neutral genes to morph specific function

Genes with high rate of sequence evolution due to weak purifying selection may not only drift toward morph-biased expression, as described above, but also be more likely to become co-opted for morph specific functions. This is because weak selection enables the accumulation of genetic variation that can become functional in novel contexts (True and Carroll, 2002). Co-option can potentially occur during morph evolution. Alternatively, genes may become morph biased after the evolution of morphs even if they did not play a role in their original divergence. This has potential implications for the evolution of both the sequence and regulation of those genes. Both positive and purifying selection following co-option are possible and can make the rate of sequence evolution change from near-neutral toward faster or slower or, comparing site by site, increase both the proportions of sites under positive and purifying selection, respectively (Figure 1C).

In this scenario, following evolutionary rates of gene sequences over time should reveal a change from expectations of neutrality toward signatures of selection. Consequently co-opted genes should contribute to the observed correlation of expression bias and fast sequence evolution only through those genes that became positively selected following co-option, as only these genes continue to evolve fast. At the level of expression, co-option of historically “near-neutral” gene sequences should result in further selection for precise gene regulation and hence a reduction in expression noise over evolutionary time in lineages with morphs (Figure 1C). The role of co-option in the evolution

of morph biased gene expression has not been directly studied in social insects, and doing so requires more information on the extent to which morph-biased genes also have morph-biased fitness effects. For example, studies showing that weak selection on sequence precedes morph biased expression (Hunt et al., 2011; Leichty et al., 2012) have not demonstrated that the subsequent expression bias reflects morph specific function rather than continued weak selection. In contrast, positive selection of worker biased genes in honeybees is also consistent with a co-option scenario, but the historical data on evolutionary rates before caste biased expression evolved is lacking (Harpur et al., 2014). Outside social insects it has been observed that up-regulated expression in one morph is linked to higher fitness effects in that morph (Connallon and Clark, 2011; Hall and Goodisman, 2012), but it is unknown whether the fitness effects caused selection for biased expression, or if biased expression arose first followed by compensatory changes to maintain morph fitness.

Evolution under weak pleiotropic constraint

Genes typically have multiple functional targets, which may constrain their evolution. Functional constraint contributes to the overall strength of selection on sequence and expression and is therefore implicit in much of what has already been discussed. However, the literature also emphasizes a more constructive role of weak pleiotropy where it directly causes particular fast evolving genes to become morph biased. Genes that are expressed in a context or tissue specific manner (Duret and Mouchiroud, 2000; Zhang and Li, 2004), are likely to have low number of interactions with other gene products (e.g., Assis et al., 2012) and therefore be free to evolve under directional selection. Furthermore, it has been suggested that genes that have a regulatory architecture that allows precise regulation, which also decreases pleiotropic constraint, are more likely to exhibit context sensitive expression patterns (Grishkevich and Yanai, 2013), such as morph biased expression. While the correlation of pleiotropic constraint and expression pattern has not yet been tested in social insects, it is supported by sex-specific gene expression patterns in mice, chicken and fruit flies, where sex biased genes appear to exhibit weak pleiotropic constraints (Mank et al., 2007; Meisel, 2011; see also below). If precisely regulated genes, with potentially low pleiotropic constraint, are more likely to evolve morph biased expression patterns, this should create a consistent positive correlation between morph biased expression and high rate of sequence evolution, caused by positive selection, over evolutionary time (Figure 1D).

Fast sequence evolution due to positive selection has been shown to occur in worker biased genes in the honeybee (Harpur et al., 2014), in caste biased genes in seven ant genomes (Roux et al., 2014) and male biased genes in *Drosophila* (reviewed in Wright and Mank, 2013). However, additional data is necessary before this can be taken as support for rapid evolution due to relaxation from pleiotropic constraints. If fast evolving genes with expression bias are indeed only weakly constrained by pleiotropy, we expect their evolutionary rate to be high already before the biased expression pattern evolved. Furthermore, under highly precise regulation we expect relatively low expression variation

among individuals within morphs. Consistent directional selection is thought to be rare, but may be particularly likely for genes involved in reproduction, immunity and social and reproductive conflicts (Swanson and Vacquier, 2002; Summers and Crespi, 2005; Obbard et al., 2009). Because many morphs, notably sex and caste, have different reproductive functions the morph biased genes that evolve fast under directional selection may largely be composed of reproductive and conflict related genes. This could also explain why such genes evolve fast before they become morph biased (or before morphs evolve). In social insects such genes may be found among genes involved in recognition and responses to hormones, as suggested by Roux et al. (2014). Outside social insects, genes with a conserved male biased expression, likely to be involved in reproductive function and often expressed in sex specific tissues, have been shown to evolve faster than other sex biased genes in *Drosophila* (Grath and Parsch, 2012). The often narrow tissue wide expression profiles of sex biased genes may also support that genes without pleiotropic effects are more likely to become morph biased, but without temporal data it is difficult to tease apart what is cause and consequence for this association.

Morph antagonistic selection

One source of antagonistic selection is when an allele has beneficial effects on one morph but negative effects in another morph. This form of antagonistic selection has been discussed frequently with respect to sex biased gene expression (Rice and Chippindale, 2001; Morrow et al., 2008; Innocenti and Morrow, 2010), but here we emphasize that the same logic applies to any polymorphism, including social insect castes (see Hall et al., 2013; Holman, 2014, for specific models on caste antagonistic selection). If an allele has opposite fitness effects in two or more morphs, selection should favor suppression of expression in the morph(s) where it has negative consequences (Rice and Chippindale, 2001). Following the evolution of suppression of gene expression, antagonistic pleiotropy is relaxed, which enables genes to respond to directional selection and exhibit fast sequence evolution (Gadagkar, 1997). This follows the general logic described above, but under this scenario the changes in expression pattern and evolutionary rates are predicted to occur concurrently, i.e., sequence evolution accelerates when biased expression evolves (Figure 1E). Also, genes under positive selection should be highly regulated (Fraser et al., 2004; Wang and Zhang, 2011), and thus vary relatively little in their expression pattern within each morph. Many sex biased genes do evolve under positive selection (reviewed in Wright and Mank, 2013), and positive selection in worker biased genes has recently been demonstrated in the honeybee (Harpur et al., 2014). The expression history of these genes, and the timing of possible changes in evolutionary rates, is largely uncharted. Support for the theory would require data showing that these genes began to evolve under positive selection following the evolution of morph-biased expression.

Actual mapping of loci with antagonistic fitness effects is currently out of reach for any social insect system, but studies have recently been conducted in *Drosophila* (Innocenti and Morrow, 2010; Parsch and Ellegren, 2013). Only a minor proportion of genes with sex biased expression showed sexually antagonistic fitness effects in a hemiclinal analysis (Innocenti and Morrow,

2010), which suggests that ongoing selection for suppression is not a major explanation for biased gene expression. However, there are several reasons why some of the antagonistic fitness effects may go undetected in such coarse scale analyses (outlined in Parsch and Ellegren, 2013), and the contemporary lack of antagonistic fitness variation in sex biased genes could be a signal of a resolved ancestral conflict (Innocenti and Morrow, 2010). Because a considerable proportion of unbiased genes appear to have antagonistic fitness effects (Innocenti and Morrow, 2010), it is also possible that constraints such as intersexual genetic correlation may limit an evolutionary response to sexual antagonism in terms of biased expression. Alternatively, alleles at unbiased loci that show antagonistic fitness effects may have arisen so recently that the resulting conflict has not yet been resolved through morph-biased suppression of expression.

Summary of scenarios

It is important to keep in mind that even when the scenarios make mutually exclusive predictions (summarized in Figure 1), they still represent processes that can co-occur and overlap. For example, if antagonistic fitness drives the system to morph-specific expression, this leads to relaxation of selection as well. Similarly, co-option and relaxed selection can be seen as alternative interpretations of a similar process that on the one hand allows exploration of the phenotypic space and on the other hand may lead to accumulation of slightly harmful mutations and “polymorphism load.” Finally, although weak pleiotropic constraint can be viewed as the principal reason for why some genes become morph biased in their expression, it is also an important determinant of evolutionary rates in other scenarios simply by dictating the overall selection on both gene sequence and expression.

DISCUSSION

Recent research provides partial support for several of the scenarios described above, but the data on social insects are still very limited and it is unknown how much can be generalized from other polymorphic systems (Box 1). It is likely that several processes contribute to some extent. Thus, the empirical task is assessing the relative contribution of the different processes rather than forcing a single explanation to any given pattern. We suggest that doing so relies on two critical types of data—gene expression variation between and within species—both of which are limited in published studies to date.

First, phylogenetic mapping of the rates of sequence evolution and patterns of gene expression (Figure 1) is necessary for revealing the temporal order of changes in gene expression and sequence divergence. To date, comparisons are typically weak in terms of phylogenetic rigor (e.g., making use of two-species comparisons), and availability of a number of relevant contrasts is clearly a major challenge for future work. With more species, mapping rates of sequence divergence on a phylogenetic tree can determine whether fast-evolving genes are more likely to show morph-biased expression than slow-evolving genes. In cases where ancestral monomorphic populations are extant, such comparisons can be carried out by comparing evolutionary rates in lineages with and without morphs (e.g., Leichy et al., 2012), although in many cases replication is limited by

the number of independent origins of morphs. For the advanced eusocial Hymenoptera, the number of independent origins of morphs is a serious limiting factor—but social taxa that comprise multiple origins of sociality, such as halictid bees or snapping shrimps, could be fruitful model systems for replicated studies. For cases where number of independent replicates is limiting, or where monomorphic outgroups are unavailable, the phylogenetic reconstruction of expression patterns has to be carried out gene-by-gene (see e.g., Grath and Parsch, 2012). Making the comparisons at a relevant phylogenetic scale can reveal both the effects of idiosyncratic features of specific polymorphic taxa, and possible convergent features shared across independent evolutionary origins of polymorphisms.

Furthermore, separating weak purifying selection from positive selection as causes of fast sequence change places demands on sequence data. Simple summary statistics such as the average dn/ds ratios per gene can reveal interesting patterns of average evolutionary rates but are unlikely to capture the complexity of the process. McDonald-Kreitman type tests are a more powerful and suitable method for detecting genes under positive selection (see e.g., Harpur et al., 2014 for a recent example) when a small number of taxa are analyzed. Furthermore, given that some of the scenarios predict concurrent changes in both the strength of positive and purifying selection, investigating site specific signatures of selection e.g., using maximum likelihood methods (Yang, 2007) in larger phylogenetic data sets might be necessary for thoroughly teasing apart the contributions of all the different processes (Nielsen, 2005).

Second, we suggest that it will be necessary to establish the patterns of variation among individuals within morphs for teasing apart adaptive and non-adaptive scenarios. This is because the processes that drive the rate of sequence divergence phylogenetically are also expected to generate different patterns of variation in gene expression among individuals within populations and species. Large expression variation among individuals in morph-biased genes would support the idea that genes become morph-biased because they are under relatively weak selection. In contrast, if genes evolve fast as a result of directional selection, this should be associated with precise gene regulation and hence biased genes should exhibit low expression variation within and between individuals of a given morph. Individual-level data on gene expression therefore provide one potential source of information that can help to evaluate the reasons for biased expression, which also sets demands for replication and careful study design for future studies. Given the large size of many social insects, replication at an individual (see e.g., Morandin et al., 2014) and tissue level (Johnson et al., 2013) should be feasible.

Unfortunately, interpretation of gene expression variation is difficult. On the one hand, variation may represent lack of precise regulation, which causes noisy expression (Fraser et al., 2004). On the other hand, gene expression data may show substantial variation simply because of variable external or internal states not controlled during data collection (Figure 2). While it has been shown that expression is inherently noisier in non-essential genes in model organisms such as yeast, interpreting patterns of expression variability that underlie complex phenotypes is far from straightforward given the large numbers of genes that

exhibit context-dependence that is unrelated to morph-specific function. For example, the proportions of genes that are differently expressed across life stages (Ometto et al., 2011; Perry et al., 2014), social environments (Manfredini et al., 2013), and genotypes (Nipitwattanaphon et al., 2013) are sometimes comparable in magnitude to morph-biased proportions. It has also been shown in studies focusing on single genes, such as vitellogenin, that caste bias is sensitive to seasonal and contextual variation (Azevedo et al., 2011; Libbrecht et al., 2013; Morandin et al., 2014). Furthermore, factors such as individual condition in *Drosophila* (Wyman et al., 2010), behavior in zebrafish (Rey et al., 2013), presence of social and sexual stimuli in swordtails (Cummings et al., 2008), and abiotic environmental conditions (Yampolsky et al., 2012) have been demonstrated to co-vary with expression patterns. These results suggest that without proper replication it cannot be assumed that all observed variation within morphs is stochastic and a sign of weak regulation. Also the observation that morph bias varies extensively between life stages (Ometto et al., 2011) and tissues (Mank et al., 2008) suggests that the more we understand the causes of variation in expression patterns, the fewer genes will be consistently classified as morph biased (Meisel, 2011).

Importantly, assessing any adaptive scenario for gene expression variation is only possible when compared against a suitable neutral expectation. While the neutral evolution of morph biased expression patterns has been directly assessed in only a few cases, studies of selection acting on gene expression patterns in general may shed some light on this issue. There are several recently suggested neutral scenarios in the literature (Gout et al., 2010; Warnefors and Eyre-Walker, 2012; Smith et al., 2013; Rohlf et al., 2014) but empirical studies that address neutral expectations have focused on species divergence in gene expression and not morph-biased expression. It has been suggested that the factors that cause gene expression to diverge among species (e.g., non-essentiality) also expose genes to evolve context specific expression patterns (Grishkevich and Yanai, 2013). Whether general conclusions about selection on gene expression also apply to caste specific patterns remains an open question, but we suggest that they may very well do. This is supported by the finding of enriched signatures of adaptive regulatory evolution in genes underlying worker behavioral plasticity in honeybees (Harpur et al., 2014), and the extensive diversification of regulatory elements in social insects in general (Simola et al., 2013). Overall, the evidence for selection on gene expression is mixed, but a prevalence of stabilizing selection has been suggested (Gilad et al., 2006; Khaitovich et al., 2006; Warnefors and Eyre-Walker, 2012). In contrast, the relatively large turnover in the set of morph biased genes [caste biased genes between two species of *Polistes* paper wasps (Ferreira et al., 2013), between two species of *Cryptotermes* termites (Weil et al., 2009), sex biased genes among species of *Drosophila* (Metta et al., 2006; Zhang et al., 2007; Jiang and Machado, 2009; Assis et al., 2012) and between zebra finch *Taeniopygia guttata* and common whitethroat *Sylvia communalis* (Naurin et al., 2011)] supports that neutral processes play a large role, implying that genes acquire or lose morph biased expression largely due to drift. This is consistent with studies comparing a small numbers of genes in closely related species that have shown

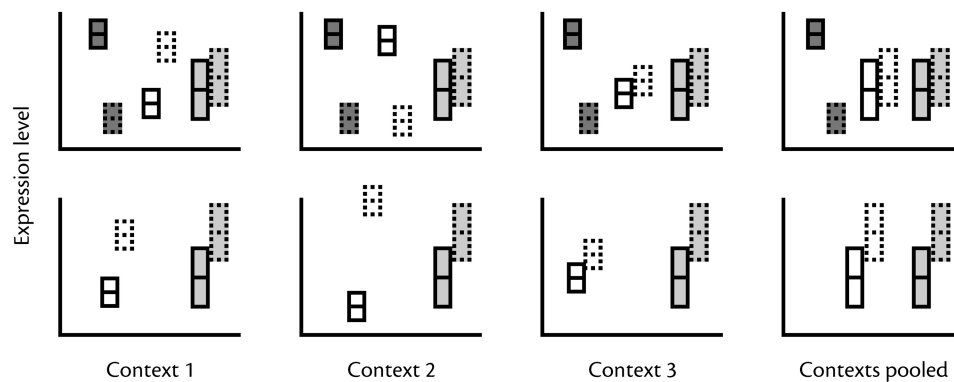


FIGURE 2 | Why variability in expression pattern may be difficult to interpret. Boxes with solid and dotted outlines refer to expression in two different morphs. Top row: Within contexts gene expression that are consistently morph-biased (dark gray boxes) or morph-biased in a context-dependent manner (e.g., body size, age, past or current social experience) (white boxes) will show a morph

biased pattern. In contrast, pooling samples across contexts makes context-dependent expression indistinguishable from expression variability *per se* (light gray boxes). Bottom row: In pooled samples genes regulated according to context (white boxes) are indistinguishable from genes with an expression pattern that is un-biased but highly variable (gray boxes).

that caste biases may be evolutionarily labile (Weil et al., 2009; Morandin et al., 2014).

CONCLUSIONS

Recent data suggests a relationship between the rate of sequence evolution and morph-biased gene expression in social insects and other polymorphic taxa, but its causes remain poorly understood. Morph-biased genes can evolve faster for several reasons. We suggest that the majority of morph-biased genes are under relatively weak selection, which can also explain why those genes evolve faster before the evolution of morphs. This suggests that adaptive scenarios should be treated with caution unless further supporting evidence can be provided. However, we also suggest that genes that ancestrally have been under weak selection, and therefore show high accumulation of mutations, may be co-opted in morph evolution and hence continue to evolve fast because of directional selection. Alternatively, co-option can lead to a reduction in the rate of evolution because of purifying selection following the onset of morph-biased expression. There are therefore several different possible genomic signatures of the evolution of morphs. Distinguishing between adaptive and (near-)neutral scenarios for the coupling of the rate of sequence evolution and morph-biased expression will require data to be replicated in several dimensions (individuals, contexts, morphs, species) at a level that is only now beginning to be possible in any taxa, including social insects. Many of the reported correlations to date are weak, and the patterns are likely to be refined by carefully assessing different functional classes of genes, more detailed studies of tissue specific expression, and studies that directly assess the evolution of gene regulation.

ACKNOWLEDGMENTS

The authors thank Tim Linksvayer, Jonna Kulmuni, and the two reviewers for comments. Heikki Helanterä is funded by the Academy of Finland (135990) and the Centre of Excellence in Biological Interactions. Tobias Uller is funded by the

Royal Society of London and the Knut and Alice Wallenberg Foundations.

REFERENCES

- Assis, R., Zhou, Q., and Bachtrog, D. (2012). Sex-biased transcriptome evolution in drosophila. *Genome Biol. Evol.* 4, 1189–1200. doi: 10.1093/gbe/evs093
- Azevedo, D. O., Zanoncio, J. C., Delabie, J. H. C., and Serrão, J. E. (2011). Temporal variation of vitellogenin synthesis in *Ectatomma tuberculatum* (Formicidae: Ectatomminae) workers. *J. Insect Physiol.* 57, 972–977. doi: 10.1016/j.jinsphys.2011.04.015
- Bourke, (1999). Colony size, social complexity and reproductive conflict in social insects. *J. Evol. Biol.* 12, 245–257. doi: 10.1046/j.1420-9101.1999.00028.x
- Carroll, S. B. (2005). Evolution at two levels: on genes and form. *PLoS Biol.* 3:e245. doi: 10.1371/journal.pbio.0030245
- Colgan, T., Carolan, J., Bridgett, S., Sumner, S., Blaxter, M., and Brown, M. (2011). Polyphenism in social insects: insights from a transcriptome-wide analysis of gene expression in the life stages of the key pollinator, *Bombus terrestris*. *BMC Genomics* 12:623. doi: 10.1186/1471-2164-12-623
- Connallon, T., and Clark, A. G. (2010). Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution* 64, 3417–3442. doi: 10.1111/j.1558-5646.2010.01136.x
- Connallon, T., and Clark, A. G. (2011). Association between Sex-Biased gene expression and mutations with Sex-Specific phenotypic consequences in drosophila. *Genome Biol. Evol.* 3, 151–155. doi: 10.1093/gbe/evr004
- Cummings, M. E., Larkins-Ford, J., Reilly, C. R. L., Wong, R. Y., Ramsey, M., and Hofmann, H. A. (2008). Sexual and social stimuli elicit rapid and contrasting genomic responses. *Proc. Biol. Sci.* 275, 393–402. doi: 10.1098/rspb.2007.1454
- Duret, L., and Mouchiroud, D. (2000). Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol. Biol. Evol.* 17, 68–70. doi: 10.1093/oxfordjournals.molbev.a026239
- Feldmeyer, B., Elsner, D., and Foitzik, S. (2014). Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Mol. Ecol.* 23, 151–161. doi: 10.1111/mec.12490
- Ferreira, P. G., Patalano, S., Chauhan, R., Ffrench-Constant, R., Gabaldon, T., Guigo, R., et al. (2013). Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol.* 14:R20. doi: 10.1186/gb-2013-14-2-r20
- Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J., and Eisen, M. B. (2004). Noise minimization in eukaryotic gene expression. *PLoS Biol.* 2:834–838. doi: 10.1371/journal.pbio.0020137

- Gadagkar, R. (1997). The evolution of caste polymorphism in social insects: genetic release followed by diversifying evolution. *J. Genet.* 76, 167–179. doi: 10.1007/BF02932215
- Gilad, Y., Oshlack, A., and Rifkin, S. A. (2006). Natural selection on gene expression. *Trends Genet.* 22, 456–461. doi: 10.1016/j.tig.2006.06.002
- Gossmann, T. I., Schmid, M. W., Grossniklaus, U., and Schmid, K. J. (2014). Selection-Driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. *Mol. Biol. Evol.* 31, 574–583. doi: 10.1093/molbev/mst226
- Gout, J. F., Kahn, D., Duret, L., and Parametrium Post-Genomics, Consortium (2010). The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet.* 6:e1000944. doi: 10.1371/annotation/c55d5089-ba2f-449d-8696-2bc8395978db
- Grath, S., and Parsch, J. (2012). Rate of amino acid substitution is influenced by the degree and conservation of Male-Biased transcription over 50 Myr of drosophila evolution. *Genome Biol. Evol.* 4, 346–359. doi: 10.1093/gbe/evs012
- Grishkevich, V., and Yanai, I. (2013). The genomic determinants of genotype x environment interactions in gene expression. *Trends Genet.* 29, 479–487. doi: 10.1016/j.tig.2013.05.006
- Grozinger, C. M., Fan, Y. L., Hoover, S. E. R., and Winston, M. L. (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. doi: 10.1111/j.1365-294X.2007.03545.x
- Hall, D. W., and Goodisman, M. A. D. (2012). The effects of kin selection on rates of molecular evolution in social insects. *Evolution* 66, 2080–2093. doi: 10.1111/j.1558-5646.2012.01602.x
- Hall, D. W., Yi, S. V., and Goodisman, M. A. D. (2013). Kin selection, genomics and caste-antagonistic pleiotropy. *Biol. Lett.* 9:20130309. doi: 10.1098/rsbl.2013.0309
- Harpur, B. A., Kent, C. F., Molodtsova, D., Lebon, J. M. D., Alqarni, A. S., Owayss, A. A., et al. (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. doi: 10.1073/pnas.1315506111
- Hoekstra, H. E., and Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61, 995–1016. doi: 10.1111/j.1558-5646.2007.00105.x
- Holman, L. (2014). Caste load and the evolution of reproductive skew. *Am. Nat.* 183, 84–95. doi: 10.1086/674052
- Hunt, B. G., Ometto, L., Keller, L., and Goodisman, M. A. D. (2013). Evolution at two levels in fire ants: the relationship between patterns of gene expression and protein sequence evolution. *Mol. Biol. Evol.* 30, 263–271. doi: 10.1093/molbev/mss234
- Hunt, B. G., Ometto, L., Wurm, Y., Shoemaker, D., Yi, S. V., Keller, L., et al. (2011). Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15936–15941. doi: 10.1073/pnas.1104825108
- Hunt, B. G., Wyder, S., Elango, N., Werren, J. H., Zdobnov, E. M., Yi, S. V., et al. (2010). Sociality is linked to rates of protein evolution in a highly social insect. *Mol. Biol. Evol.* 27, 497–500. doi: 10.1093/molbev/msp225
- Innocenti, P., and Morrow, E. H. (2010). The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biol.* 8:e1000335. doi: 10.1371/journal.pbio.1000335
- Jiang, Z. F., and Machado, C. A. (2009). Evolution of sex-dependent gene expression in three recently diverged species of *Drosophila*. *Genetics* 183, 1175–1185. doi: 10.1534/genetics.109.105775
- Johnson, B., Atallah, J., and Plachetzki, D. (2013). The importance of tissue specificity for RNA-seq: highlighting the errors of composite structure extractions. *BMC Genomics* 14:586. doi: 10.1186/1471-2164-14-586
- Khaitovich, P., Enard, W., Lachmann, M., and Paabo, S. (2006). Evolution of primate gene expression. *Nat. Rev. Genet.* 7, 693–702. doi: 10.1038/nrg1940
- Khaitovich, P., Paabo, S., and Weiss, G. (2005). Toward a neutral evolutionary model of gene expression. *Genetics* 170, 929–939. doi: 10.1534/genetics.104.037135
- Larracuente, A. M., Sackton, T. B., Greenberg, A. J., Wong, A., Singh, N. D., Sturgill, D., et al. (2008). Evolution of protein-coding genes in *Drosophila*. *Trends Genet.* 24, 114–123. doi: 10.1016/j.tig.2007.12.001
- Leichty, A. R., Pfennig, D. W., Jones, C. D., and Pfennig, K. S. (2012). Relaxed genetic constraint is ancestral to the evolution of phenotypic plasticity. *Integr. Comp. Biol.* 52, 16–30. doi: 10.1093/icb/ics049
- Lemos, B., Bettencourt, B. R., Meiklejohn, C. D., and Hartl, D. L. (2005a). 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. doi: 10.1093/molbev/msi122
- Lemos, B., Meiklejohn, C. D., Caceres, M., and Hartl, D. L. (2005b). Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* 59, 126–137. doi: 10.1111/j.0014-3820.2005.tb00900.x
- Libbrecht, R., Corona, M., Wende, F., Azevedo, D. O., Serrão, J. E., and Keller, L. (2013). Interplay between insulin signaling, juvenile hormone, and vitellogenin regulates maternal effects on polyphenism in ants. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11050–11055. doi: 10.1073/pnas.1221781110
- Linksvayer, T. A., and Wade, M. J. (2009). Genes with social effects are expected to harbor more sequence variation within and between species. *Evolution* 63, 1685–1696. doi: 10.1111/j.1558-5646.2009.00670.x
- Manfredini, F., Riba-Grognuz, O., Wurm, Y., Keller, L., Shoemaker, D., and Grozinger, C. M. (2013). Sociogenomics of cooperation and conflict during colony founding in the fire ant *Solenopsis invicta*. *PLoS Genet.* 9:e1003633. doi: 10.1371/journal.pgen.1003633
- Mank, J. E., and Ellegren, H. (2009). Are sex-biased genes more dispensable? *Biol. Lett.* 5, 409–412. doi: 10.1098/rsbl.2008.0732
- Mank, J. E., Hultin-Rosenberg, L., Axelsson, E., and Ellegren, H. (2007). Rapid evolution of female-biased, but not male-biased, genes expressed in the avian brain. *Mol. Biol. Evol.* 24, 2698–2706. doi: 10.1093/molbev/msm208
- Mank, J. E., Hultin-Rosenberg, L., Zwahlen, M., and Ellegren, H. (2008). Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. *Am. Nat.* 171, 35–43. doi: 10.1086/523954
- Mcmanus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., and Wittkopp, P. J. (2010). Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* 20, 816–825. doi: 10.1101/gr.102491.109
- Meisel, R. P. (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. doi: 10.1093/molbev/msr010
- Metta, M., Gudavalli, R., Gibert, J. M., and Schlotterer, C. (2006). No accelerated rate of protein evolution in male-biased *Drosophila pseudoobscura* genes. *Genetics* 174, 411–420. doi: 10.1534/genetics.106.057414
- Morandini, C., Havukainen, H., Kulmuni, J., Dhaygude, K., Trontti, K., and Helanterä, H. (2014). Not only for egg yolk-functional and evolutionary insights from expression, selection and structural analyses of Formica ant vitellogenins. *Mol. Biol. Evol.* 31, 2181–2193. doi: 10.1093/molbev/msu171
- Morrow, E. H., Stewart, A. D., and Rice, W. R. (2008). Assessing the extent of genome-wide intralocus sexual conflict via experimentally enforced gender-limited selection. *J. Evol. Biol.* 21, 1046–1054. doi: 10.1111/j.1420-9101.2008.01542.x
- Naurin, S., Hansson, B., Hasselquist, D., Kim, Y.-H., and Bensch, S. (2011). The sex-biased brain: sexual dimorphism in gene expression in two species of songbirds. *BMC Genomics* 12:37. doi: 10.1186/1471-2164-12-37
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.* 39, 197–218. doi: 10.1146/annurev.genet.39.073003.112420
- Nipitwattanaphon, M., Wang, J., Dijkstra, M. B., and Keller, L. (2013). A simple genetic basis for complex social behaviour mediates widespread gene expression differences. *Mol. Ecol.* 22, 3797–3813. doi: 10.1111/mec.12346
- Obbard, D. J., Welch, J. J., Kim, K.-W., and Jiggins, F. M. (2009). Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet.* 5:e1000698. doi: 10.1371/journal.pgen.1000698
- Ometto, L., Shoemaker, D., Ross, K. G., and Keller, L. (2011). Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. *Mol. Biol. Evol.* 28, 1381–1392. doi: 10.1093/molbev/msq322
- Pál, C., Papp, B., and Hurst, L. D. (2001). Highly expressed genes in yeast evolve slowly. *Genetics* 158, 927–931.
- Park, J., Xu, K., Park, T., and Yi, S. V. (2012). What are the determinants of gene expression levels and breadths in the human genome? *Hum. Mol. Genet.* 21, 46–56. doi: 10.1093/hmg/ddr436
- Parsch, J., and Ellegren, H. (2013). The evolutionary causes and consequences of sex-biased gene expression. *Nat. Rev. Genet.* 14, 83–87. doi: 10.1038/nrg3376

- Perry, J. C., Harrison, P. W., and Mank, J. E. (2014). The ontogeny and evolution of Sex-Biased gene expression in *Drosophila melanogaster*. *Mol. Biol. Evol.* 31, 1206–1219. doi: 10.1093/molbev/msu072
- Purandare, S. R., Bickel, R. D., Jaquiere, J., Rispe, C., and Brisson, J. A. (2014). Accelerated evolution of morph-biased genes in pea aphids. *Mol. Biol. Evol.* 31, 2073–2083. doi: 10.1093/molbev/msu149
- Queller, D. C., and Strassmann, J. E. (2009). Beyond society: the evolution of organismality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3143–3155. doi: 10.1098/rstb.2009.0095
- Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D., and Hartl, D. L. (2003). Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* 300, 1742–1745. doi: 10.1126/science.1085881
- Rey, S., Boltana, S., Vargas, R., Roher, N., and Mackenzie, S. (2013). Combining animal personalities with transcriptomics resolves individual variation within a wild-type zebrafish population and identifies underpinning molecular differences in brain function. *Mol. Ecol.* 22, 6100–6115. doi: 10.1111/mec.12556
- Rice, W. R., and Chippindale, A. K. (2001). Intersexual ontogenetic conflict. *J. Evol. Biol.* 14, 685–693. doi: 10.1046/j.1420-9101.2001.00319.x
- Rice, W. R., and Holland, B. (1997). The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. *Behav. Ecol. Sociobiol. (Print)*. 41, 1–10. doi: 10.1007/s002650050357
- Rohlf, R. V., Harrigan, P., and Nielsen, R. (2014). Modeling gene expression evolution with an extended Ornstein-Uhlenbeck process accounting for within-species variation. *Mol. Biol. Evol.* 31, 201–211. doi: 10.1093/molbev/mst190
- Roux, J., Privman, E., Moretti, S., Daub, J. T., Robinson-Rechavi, M., and Keller, L. (2014). Patterns of positive selection in seven ant genomes. *Mol. Biol. Evol.* 31, 1661–1685. doi: 10.1093/molbev/msu141
- Schwander, T., and Leimar, O. (2011). The evolution of novel cues for ancestral phenotypes. *Trends Ecol. Evol. (Amst)*. 26, 436–437. doi: 10.1016/j.tree.2011.06.001
- Simola, D. F., Wissler, L., Donahue, G., Waterhouse, R. M., Helmkampf, M., Roux, J., et al. (2013). Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res.* 23, 1235–1247. doi: 10.1101/gr.155408.113
- Smith, J. D., McManus, K. F., and Fraser, H. B. (2013). A novel test for selection on cis-regulatory elements reveals positive and negative selection acting on mammalian transcriptional enhancers. *Mol. Biol. Evol.* 30, 2509–2518. doi: 10.1093/molbev/mst134
- Snell-Rood, E. C., Cash, A., Han, M. V., Kijimoto, T., Andrews, J., and Moczek, A. P. (2011). Developmental decoupling of alternative phenotypes: insights from the transcriptomes of horn-polyphenic beetles. *Evolution* 65, 231–245. doi: 10.1111/j.1558-5646.2010.01106.x
- Snell-Rood, E. C., Van Dyken, J. D., Cruickshank, T., Wade, M. J., and Moczek, A. P. (2010). Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity. *Bioessays* 32, 71–81. doi: 10.1002/bies.200900132
- Summers, K., and Crespi, B. (2005). Cadherins in maternal-foetal interactions: red queen with a green beard? *Proc. Biol. Sci.* 272, 643–649. doi: 10.1098/rspb.2004.2890
- Sumner, S., Pereboom, J. J. M., and Jordan, W. C. (2006). Differential gene expression and phenotypic plasticity in behavioural castes of the primitively eusocial wasp, *Polistes canadensis*. *Proc. Biol. Sci.* 273, 19–26. doi: 10.1098/rspb.2005.3291
- Swanson, W. J., and Vacquier, V. D. (2002). The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* 3, 137–144. doi: 10.1038/nrg733
- True, J. R., and Carroll, S. B. (2002). GENE Co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* 18, 53–80. doi: 10.1146/annurev.cellbio.18.020402.140619
- Uller, T., and Helanterä, H. (2011). From the origin of sex-determining factors to the evolution of sex-determining systems. *Q. Rev. Biol.* 86, 163–180. doi: 10.1086/661118
- Van Dyken, J. D., and Wade, M. J. (2010). The genetic signature of conditional expression. *Genetics* 184, 557–570. doi: 10.1534/genetics.109.110163
- Wang, Z., and Zhang, J. (2011). Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc. Natl. Acad. Sci. U.S.A.* 108, 67–76. doi: 10.1073/pnas.1100059108
- Warnefors, M., and Eyre-Walker, A. (2012). A selection index for gene expression evolution and its application to the divergence between humans and chimpanzees. *PLoS ONE* 7:e34935. doi: 10.1371/journal.pone.0034935
- Warnefors, M., and Kaessmann, H. (2013). Evolution of the correlation between expression divergence and protein divergence in mammals. *Genome Biol. Evol.* 5, 1324–1335. doi: 10.1093/gbe/evt093
- Weil, T., Korb, J., and Rehli, M. (2009). Comparison of Queen-Specific gene expression in related lower termite species. *Mol. Biol. Evol.* 26, 1841–1850. doi: 10.1093/molbev/msp095
- West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford: Oxford University Press.
- Wright, A. E., and Mank, J. E. (2013). The scope and strength of sex-specific selection in genome evolution. *J. Evol. Biol.* 26, 1841–1853. doi: 10.1111/jeb.12201
- Wyman, M. J., Agrawal, A. F., and Rowe, L. (2010). Condition-dependence of the sexually dimorphic transcriptome in *Drosophila melanogaster*. *Evolution* 64, 1836–1848. doi: 10.1111/j.1558-5646.2009.00938.x
- Yampolsky, L. Y., Glazko, G. V., and Fry, J. D. (2012). Evolution of gene expression and expression plasticity in long-term experimental populations of *Drosophila melanogaster* maintained under constant and variable ethanol stress. *Mol. Ecol.* 21, 4287–4299. doi: 10.1111/j.1365-294X.2012.05697.x
- Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Zhang, L., and Li, W.-H. (2004). Mammalian housekeeping genes evolve more slowly than Tissue-Specific genes. *Mol. Biol. Evol.* 21, 236–239. doi: 10.1093/molbev/msh010
- Zhang, Y., Sturgill, D., Parisi, M., Kumar, S., and Oliver, B. (2007). Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature* 450, 233–237. doi: 10.1038/nature06323

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 May 2014; accepted: 08 August 2014; published online: 29 August 2014.

Citation: Helanterä H and Uller T (2014) Neutral and adaptive explanations for an association between caste-biased gene expression and rate of sequence evolution. *Front. Genet.* 5:297. doi: 10.3389/fgene.2014.00297

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Helanterä and Uller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Developmental regulation of ecdysone receptor (*EcR*) and *EcR*-controlled gene expression during pharate-adult development of honeybees (*Apis mellifera*)

Tathyana R. P. Mello¹, Aline C. Aleixo¹, Daniel G. Pinheiro², Francis M. F. Nunes³, Márcia M. G. Bitondi⁴, Klaus Hartfelder⁵, Angel R. Barchuk^{6*} and Zilá L. P. Simões⁴

¹ Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil

² Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, São Paulo, Brazil

³ Departamento de Genética e Evolução, Centro de Ciências Biológicas e da Saúde, Universidade Federal de São Carlos, São Carlos, Brazil

⁴ Departamento de Biologia, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil

⁵ Departamento de Biologia Celular, Molecular e de Bioagentes Patogênicos, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, São Paulo, Brazil

⁶ Laboratório de Biologia Animal Integrativa, Departamento de Biologia Celular, Tecidual e do Desenvolvimento, Instituto de Ciências Biomédicas, Universidade Federal de Alfenas, Alfenas, Brazil

Edited by:

Greg J. Hunt, Purdue University, USA

Reviewed by:

Xavier Belles, Institute of Evolutionary Biology (CSIC-UPF), Spain

Joanna Kelley, Stanford University, USA

*Correspondence:

Angel R. Barchuk, Departamento de Biologia Celular, Tecidual e do Desenvolvimento, Instituto de Ciências Biomédicas, Universidade Federal de Alfenas, Rua Gabriel Monteiro Silva, 700, CEP 37130-000 Campus Sede, Alfenas, Brazil
e-mail: barchuk@unifal-mg.edu.br; arbarchuk@yahoo.com

Major developmental transitions in multicellular organisms are driven by steroid hormones. In insects, these, together with juvenile hormone (JH), control development, metamorphosis, reproduction and aging, and are also suggested to play an important role in caste differentiation of social insects. Here, we aimed to determine how *EcR* transcription and ecdysteroid titers are related during honeybee postembryonic development and what may actually be the role of *EcR* in caste development of this social insect. In addition, we expected that knocking-down *EcR* gene expression would give us information on the participation of the respective protein in regulating downstream targets of *EcR*. We found that in *Apis mellifera* females, *EcR-A* is the predominantly expressed variant in postembryonic development, while *EcR-B* transcript levels are higher in embryos, indicating an early developmental switch in *EcR* function. During larval and pupal stages, *EcR-B* expression levels are very low, while *EcR-A* transcripts are more variable and abundant in workers compared to queens. Strikingly, these transcript levels are opposite to the ecdysteroid titer profile. 20-hydroxyecdysone (20E) application experiments revealed that low 20E levels induce *EcR* expression during development, whereas high ecdysteroid titers seem to be repressive. By means of RNAi-mediated knockdown (KD) of both *EcR* transcript variants we detected the differential expression of 234 poly-A⁺ transcripts encoding genes such as CYPs, MRJPs and certain hormone response genes (*Kr-h1* and *ftz-f1*). *EcR-KD* also promoted the differential expression of 70 miRNAs, including highly conserved ones (e.g., miR-133 and miR-375), as well honeybee-specific ones (e.g., miR-3745 and miR-3761). Our results put in evidence a broad spectrum of *EcR*-controlled gene expression during postembryonic development of honeybees, revealing new facets of *EcR* biology in this social insect.

Keywords: honey bee, adult development, 20E, ecdysteroid, juvenile hormone, JH, RNAi, miRNA

INTRODUCTION

Most multicellular organisms go through developmental transitions that enable them to cope with environmental changes and/or broaden their niche possibilities. Such transitions are generally timed and synchronized by morphogenetic hormones in a broad range of species, including insects, amphibians, metamorphic fish, tunicates, echinoderms, and plants. In insects, developmental transitions, such as larval and metamorphic molts, are driven by steroid hormones (ecdysteroids) acting in conjunction with juvenile hormone (JH). These hormones also control reproduction and aging (Flatt et al., 2005; Gáliková et al., 2011), and, in social insects, play

important roles in caste polyphenism (Hartfelder and Emlen, 2012).

The steroid hormone ecdysone is produced by the prothoracic glands. After secretion, it is transported via the hemolymph to its target organs. Due to its lipophilic nature it passes directly into the cytoplasm of target and/or modification center cells (Iga and Kataoka, 2012; Ono, 2014), where it can be modified to 20-hydroxyecdysone (20E) by a 20-monooxygenase encoded by the *shade* gene, a member of the cytochrome P450 family (CYP314a1) known as Halloween (Petryk et al., 2003). The mode of action of JH, which is a sesquiterpenoid morphogenetic molecule, has only recently become clear (for review see

Bellés and Santos, 2014), both in terms of its receptor and downstream cascade, as well as its molecular interaction with ecdysteroids. Produced by the *corpora allata* in the retrocerebral complex, JH relies on binding proteins for its transport in the hemolymph to target cells. There, it first binds to its intracellular receptor, the Methoprene-tolerant (Met) protein, which then forms a complex with Taiman (Charles et al., 2011). This dimeric hormone-receptor complex then regulates the expression of target genes.

Knowledge on the mechanism of action of insect ecdysteroids initiated with the early work of Clever and Carlson (for a historical review see Bellés and Santos, 2014), which eventually resulted in the so-called Ashburner model (Ashburner et al., 1974), which proposed a general model for the action of ecdysone, based on its participation in the regulation of gene expression (puffing) in the polytenic salivary gland chromosomes during *Drosophila melanogaster* molting and metamorphosis. Briefly, the model states that ecdysone associates with an intracellular receptor protein to activate early genes encoding transcription factors, which then activate late genes and, on the other, inhibit the transcription of previously activated early genes. The receptor protein and certain other members of this cascade belong to a large family of proteins, the nuclear hormone receptors (NR, see Fahrback et al., 2012). NR proteins are generally comprised of four independent but functionally interacting domains. A/B is a highly variable domain that may contain a motif (AF-1) driving ligand-independent transcription. The second, the C domain, is a DNA-binding domain (DBD), the most conserved region of NRs. The D or hinge domain provides a link between DBD and the next domain, LBD, a multifunctional domain that mediates ligand binding, dimerization, and interaction with heat shock proteins, nuclear localization, and transactivation functions. Functional NRs form homodimers and/or heterodimers that recognize specific DNA sequences. In the absence of a ligand molecule they act as repressors maintaining target genes inhibited by co-repressor complexes. In the presence of hormone they are activators of target genes by recruiting co-activator proteins and displacing co-repressors (Hill et al., 2013; Yamanaka et al., 2013; Evans and Mangelsdorf, 2014).

The functional ecdysone receptor is a heterodimeric NR formed by the *Ecdysone Receptor* (*EcR*) and the *ultraspiracle* (*usp*) gene products (for a comprehensive review, see Hill et al., 2013). USP is an ortholog of the vertebrate retinoid-X receptor (RXR) (Yao et al., 1992) and is most commonly considered a kind of orphan NR. Though its ligand is not known, its participation as a mediator of JH action has been postulated, probably through a direct binding of JH (Barchuk et al., 2004). Furthermore, the EcR/USP complex can also bind to the *let-7-C* gene after a 20E pulse has triggered the larval-to-pupal metamorphic molt, thus inducing the transcription of a cluster of three microRNAs (miR-100, let-7 and miR-125). These then post-transcriptionally regulate the expression of genes involved in neuromuscular morphogenesis, leading to adult body characteristics (Chawla and Sokol, 2012; see also Rubio and Bellés, 2013). The EcR protein can also *per se* regulate the expression of target genes (Davis and Li, 2013), thus adding extra levels of complexity to the mechanisms

and gene regulatory networks involving hormone/transcription factor activities.

In insects showing caste polyphenism, there is evidence that ecdysteroids are important players in caste differentiation, not only during post embryonic, but possibly even during embryonic development (Schwander et al., 2008). The role of ecdysteroids in caste development and regulation of adult reproduction is currently best understood in bees, especially so in the bumblebee *Bombus terrestris* (Geva et al., 2005) and in the honeybee *Apis mellifera* (Hartfelder and Engels, 1998), where they participate in the regulation of the differential morphogenesis programs by interacting with JH and possibly other mediating environmental modulators.

Receptor proteins mediating ecdysteroids action in social insects have been studied mainly in the honeybee (The Honey Bee Genome Sequencing Consortium, 2006; Velarde et al., 2006), where USP and EcR cDNAs have been cloned (Barchuk et al., 2004; Takeuchi et al., 2007), and the expression profiles of the respective genes were determined in several organs, tissues, and conditions (Barchuk et al., 2004, 2008; Takeuchi et al., 2007; Velarde et al., 2009). However, and despite all these works, several responses to differential hormone signaling in honeybee caste development are still poorly understood (Barchuk et al., 2007). For instance, ecdysteroid titers in developing females are higher in queens during the second half of the last larval instar (Rachinsky et al., 1990) and differ in their profiles during pupal and pharate-adult development of queens and workers (Pinto et al., 2002). These hormone titer differences are associated with the differential development of specific structures (e.g., brain and ovary, Barchuk et al., 2007) and also the onset of vitellogenin synthesis (Barchuk et al., 2002), but this is essentially correlative information lacking functional support. Herein we aimed at determining the extent to which *EcR* transcription follows ecdysteroids titers during honeybee postembryonic development and can actually mediate the action of molecular determinants of caste development in honeybees. Moreover, we expected that knocking-down *EcR* gene expression during pharate-adult development would bring to light new downstream targets of EcR.

MATERIALS AND METHODS

BEES

Embryos and the successive developmental phases in the larval and pupal stages, as well as newly-emerged adults were obtained from *A. mellifera* colonies (Africanized hybrids) maintained at the Experimental Apiary of the University of São Paulo at Ribeirão Preto, Brazil. The developmental phases of workers and queens (Table 1) were identified according to Rembold (1987) and Michelette and Soares (1993). Immediately after sampling, the bees were immersed in TRIzol reagent (Life Technologies) and frozen at -80°C until RNA extraction.

NORTHERN BLOT ANALYSIS

Approximately 15 μg of total RNA extracted from queens and workers at the PP1 and Pb developmental stages were subjected to electrophoresis in a denaturing 1.5% agarose/formaldehyde gel, and the RNA was then transferred to a PVF (Polyvinylidene Fluoride, GE) membrane using a VacuGene XL Vacuum Blotting

Table 1 | Developmental stages (embryonic, larval, pupal, adult) of *A. mellifera* considered in this work.

Abbreviation	Developmental stage
E	Embryo
L1	First instar larva
L2	Second instar larva
L3	Third instar larva
L4	Fourth instar larva
L5F1	Fifth instar larva, feeding phase 1
L5F2	Fifth instar larva, feeding phase 2
L5F3	Fifth instar larva, feeding phase 3
L5S1	Fifth instar larva, cocoon-spinning phase 1
L5S2	Fifth instar larva, cocoon-spinning phase 2
L5S3	Fifth instar larva, cocoon-spinning phase 3
PP1	Fifth instar larva, prepupa 1
PP2	Fifth instar larva, prepupa 2
PP3	Fifth instar larva, prepupa 3
Pw	White-eyed pupa, unpigmented cuticle
Pp	Pink-eyed/pharate-adult transition, unpigmented cuticle
Pdp	Dark pink-eyed pharate-adult, unpigmented cuticle
Pb	Brown-eyed pharate-adult, unpigmented cuticle
Pbl	Brown-eyed pharate-adult, light pigmented cuticle
Pbm	Brown-eyed pharate-adult, intermediary pigmented cuticle
Pbd	Brown-eyed pharate-adult, dark pigmented cuticle
NE	Newly emerged adult

system (GE Healthcare). An *EcR* cDNA fragment of 160 bp encoding the 3' part of the DNA-binding domain was used for probe synthesis by means of the Random Primers DNA Labeling System (Life Technologies) and Redivue ³²P-nucleotides (Amersham). After 3 h of hybridization at 42°C, the membranes were washed during 20 min with 0.1 × SSC solution containing 0.1% SDS and then exposed to a Super Sensitive ST film and the bands revealed with Cyclone™ Storage Phosphor System (PerkinElmer).

HORMONE TREATMENTS

For the analysis of the *EcR* expression response to artificially augmented levels of hormones, workers at the brown-eyed pupal phase (Pb) were removed from the brood frames and maintained in an incubator at 34°C and 80% relative humidity. For the ecdysone response, three groups of 3–7 workers were injected with 5 µg of 20-hydroxyecdysone (20E; Sigma) dissolved in 2 µL Ringer saline containing 12.5% ethanol. For the JH response, a similar number of Pb-phase workers received a topical application of 10 µg JH-III (Fluka) dissolved in 2 µL acetone. Controls received 2 µL of the respective solvents. The amounts of applied hormone were based on previous experiments in which we had examined their effects on inducing gene expression during pupal stage (Barchuk et al., 2002, 2004). RNA was isolated from fat bodies after 1, 12, and 24 h (independent experiments). Fat bodies were obtained via a longitudinal incision in isolated abdomens, which were then kept under gentle agitation in Petri dishes containing 0.9% NaCl. The resultant suspension of dispersed fat body cells was centrifuged during 1 min at 2500 × g and the pellet was

transferred into TRIzol reagent and frozen at –80°C until RNA extraction. We used fat bodies because this allowed us to specifically assay this metabolically important organ, especially with regard to vitellogenin (*vg*) gene expression in honeybees.

RNA EXTRACTION, REVERSE TRANSCRIPTION AND QUANTITATIVE PCR ASSAYS

Total RNA was isolated using TRIzol (Life Technologies), following the manufacturer's protocol, and purified by column purification (RNeasy Mini Kit, QIAGEN), as described previously (Barchuk et al., 2004, 2007). For the quantification of mRNA levels (except those validating the RNA-Seq data), first strand cDNA was synthesized by reverse transcription from 2 µg of RNA with SuperScript II Reverse Transcriptase (Life Technologies) and an oligo(dT)_{12–18} primer (Life Technologies). For the validation of the RNA-Seq libraries, cDNA was synthesized using NCode™ miRNA First-Strand cDNA Synthesis and qRT-PCR (Invitrogen) kits and their instructions, adding a DNase (Promega) treatment step.

Comparative analyses of transcript levels were performed by Real Time quantitative PCR (qPCR) using a 7500 Real-Time PCR System (Applied Biosystems) or a StepOne Plus system (Applied Biosystems). Amplifications were carried out in 20 µL reaction mixtures, each containing 10 µL of SYBR® Green Master Mix 2× (Applied Biosystems), 0.8 µL of a 10 mM stock solution of each of the gene-specific forward and reverse primers (Table S1), and 1 µL of first-strand cDNA diluted 1:4 (or 1:10, for cDNA samples used to validate RNA-Seq data) in ultrapure water. The sequences of forward primers were identical to the mature miRNA sequences available at miRBase, but replacing U by T, while the reverse Universal qPCR primer is supplied by NCode kit. Reaction conditions were 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min (or 33 s for miRNA amplification). Three biological replicates were run in three technical replicates each. *Actin related protein 1* (*Arp1*, GenBank accession number NM_001185145.1), *rpl32* (accession number NM_001011587.1), or a *U5 snRNA* gene were used as reference genes (for confirmation, cDNAs for all three reference genes were partially sub-cloned and sequenced in our laboratory). Relative quantities of transcripts were calculated using the comparative Ct method (Applied Biosystems, User bulletin#2). Statistical analyses were carried out with Statistica version 7.0 (<http://statistica.software.informer.com/>).

ASSESSING GENE TRANSCRIPTION PATTERNS ASSOCIATED TO *EcR* FUNCTION DURING HONEYBEE DEVELOPMENT USING RNAi

dsRNA synthesis and treatment

We employed a general protocol for dsRNA synthesis and injection in honeybees (Pbl phase) (Amdam et al., 2003). For *EcR* dsRNA synthesis, a 391 bp clone of *EcR* cDNA was amplified to serve as template, this comprising a fragment shared by the two transcript variants (A and B). The primers with the respective recognition site for T7 RNA polymerase (underlined) were: *EcR*-forward 5'-TAATACGACTCACTATAGGGCGAGAAT GGCGAGGAAGTACGAC and *EcR*-reverse 5'-TAATACGACTCACTATAGGGCGATTCTTGAACTTGAGGCTGAAG. A green fluorescent protein (*GFP*) gene clone was used as template to

synthesize the respective dsRNA used as a non-target control (*GFP-forward* 5'-TAATACGACTCACTATAGGCGGAAGTGGAGAGGGTGAAGGTGA-3' and *GFP-reverse* 5'-TAATACGACTCACTATAGGCGGAGGTAAAAGGACAGGGCCATC-3'; see Nunes et al., 2013a). The amplification products were visualized and retrieved after agarose gel electrophoresis and purified using QIAquick™ (QIAGEN) columns. *In vitro* transcription reactions were performed by using the RiboMax™ T7 system (Promega) and the obtained dsRNA was isolated using TRIzol LS reagent (Invitrogen), subjected to a denaturation step at 98°C for 5 min, followed by 30 min at room temperature, and diluted with nuclease free water to a final concentration of 2.5 µg/µL. The dsRNA quality was assessed by agarose gel electrophoresis.

Pbl-phase workers ($n = 30$ for each experimental group) received an intra-abdominal injection of 2 µL of *EcR* dsRNA solution (2.5 µg/µL). Controls of the same developmental phase received the same volume of *GFP* dsRNA solution. dsRNA-injected bees were kept in an incubator at 34°C and 80% relative humidity until adult eclosion (~2 days), when they were transferred to TRIzol reagent (Invitrogen) and frozen at -80°C until RNA extraction. Total RNA extraction and first strand cDNA synthesis were carried out as described above. *EcR* knockdown efficiencies were assessed by RT-qPCR using variant-specific primers (*EcRA-F*, *EcRB-F*, and *EcRA/B-R*; see Table S1). Bees not used for gene expression analysis were used for evaluation of the adult phenotype.

Analysis of gene expression patterns by RNA-Seq

RNA pools of equal concentration from each group of *EcR*- and *GFP*-dsRNA treated bees were used for RNA-sequencing. Libraries were prepared using the TruSeq RNA™ Sample Preparation kit (Illumina) for poly-A⁺ RNA, and the TruSeq™ Small RNA Sample preparation kit (Illumina) for small RNAs (shorter than 200 nt). These libraries were shipped to the University of North Carolina (Chapel Hill, USA) facility where they were sequenced on an Illumina platform (Genome Analyzer II, Life Sciences).

RNA-Seq reads for the poly-A⁺ RNA library were first submitted to adapter clipping using Scythe (Buffalo, 2011) (v.0.981—default parameters) for the 3'-end adapter and CutAdapt (Martin, 2011) (v.1.1—minimum overlap of 5 bp) for the 5'-end adapter. The next step was read trimming based on quality scores (mean $Q \geq 25$), Ns (number of N bases lower than 10%) and poly-A tail prediction (minimum of 5 bp of A/T at both ends). This step was performed using PRINSEQ (v.0.19.5) (Schmieder and Edwards, 2011), which also filtered very small reads (length < 15 bp). An alignment against the *A. mellifera* genome (assembly version 4.5) was run using TopHat (Trapnell et al., 2009) (v.2.0.7), guided by the respective RefSeq (Release 55) transcript coordinates. The genomic alignments were then submitted to Cufflinks (Trapnell et al., 2010) (v.2.0.2) for transcript assembly, estimation of their abundances and testing for differential expression between *EcR*-KD and control samples. The Cufflinks procedures were also guided by the RefSeq transcript coordinates. The expression estimates were properly normalized considering ambiguous alignments, and corrected for fragment bias (Roberts et al., 2011). The Poisson fragment dispersion model was used in the comparison

analysis. Cufflinks calculates the FPKM (Fragments Per Kilobase of exon per Million fragments mapped), log₂-fold-change and q -value (p -value adjusted by False Discovery Rate, FDR). However, the log₂-fold-change was recalculated after adding an offset of 1 to FPKM values in order to enable comparison involving samples without expression (zero) and to reduce the variability of the log ratios for low expression values (less than one). The functional annotation was done using Blast2GO (Conesa et al., 2005) (v.2.5), InterProScan (Mulder and Apweiler, 2007) (v.5-RC6), RefSeq transcript annotation and finding the Reciprocal-Best-Hit of *A. mellifera* RefSeq proteins against *D. melanogaster* proteins database (FlyBase r5.49) using blastp. The Blast2GO annotation pipeline was run based on blastp results of RefSeq proteins against nr database.

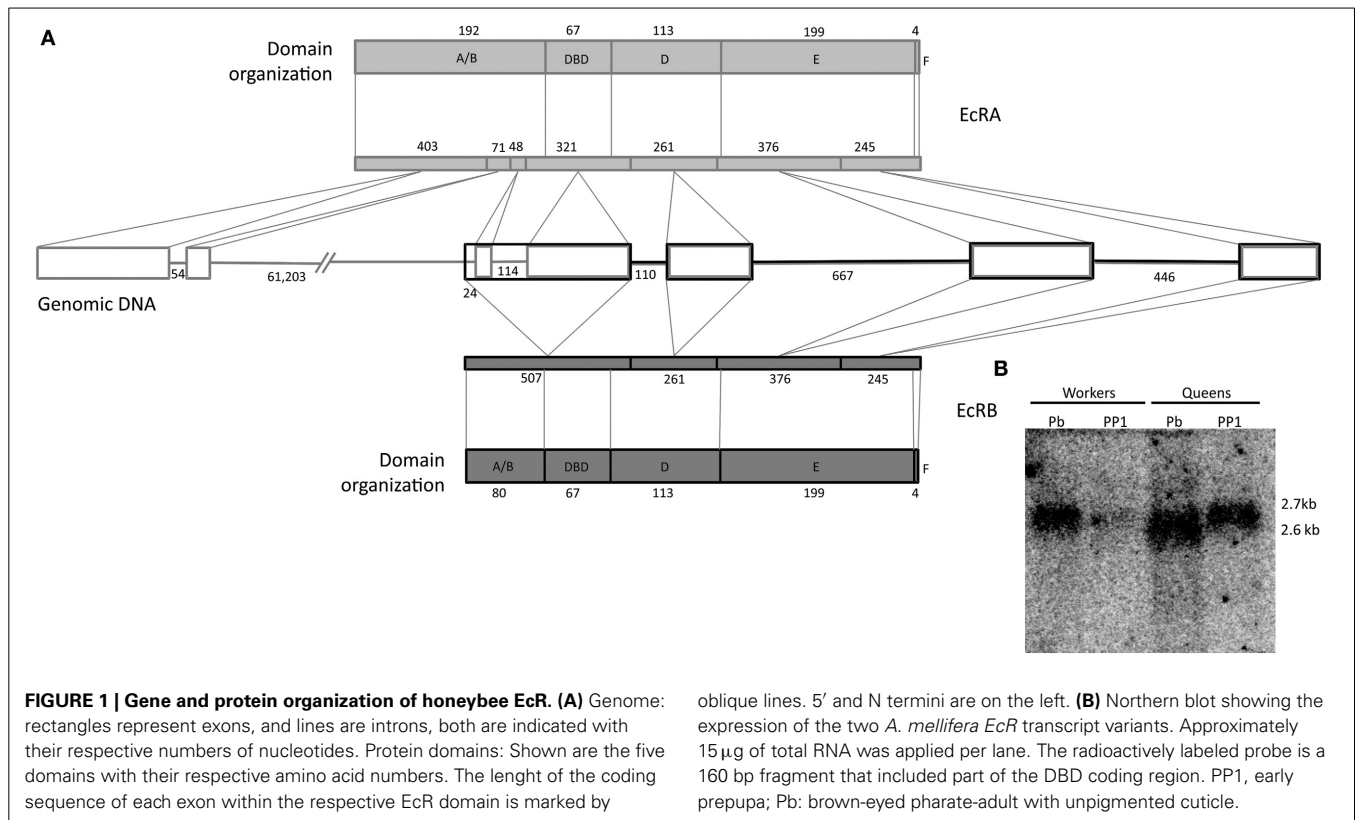
Computational processing of the Small RNA-Seq reads comprised the following steps: (i) initial sequence quality filtering based on unidentified bases; (ii) rRNA read filtering based on matches against SILVA database (Release 115); (iii) sequence adapter clipping using CutAdapt and Scythe; (iv) read trimming based on quality scores, Ns and poly-A⁺ tail prediction. All of these procedures were performed using PRINSEQ in the same way as described above. After each one of these preprocessing steps, an alignment against the *A. mellifera* genome (assembly version 4.5) was performed using the reads that had not already been aligned at each previous alignment step. Finally, all the alignment results were concatenated and transformed into a proper format to identify miRNAs. For this purpose, any splitted alignments were excluded.

Genomic alignments were performed using TopHat and the other alignments were performed using Bowtie2 (Langmead and Salzberg, 2012) (v.2.0.6). miRNA digital expression (MDE) levels were obtained by analysis with miRDeep2 (Friedländer et al., 2012) (v.2.0.0.5*), which provides the counts of reads mapped to the *A. mellifera* miRNA dataset in miRBase (Release 19). The original miRDeep2 code was modified to provide read counts for mature miRNAs instead of each precursor, and then the log₂-fold-change was calculated and statistical significance was assessed using the method proposed by Audic and Claverie (1997) with adjustment by FDR.

RESULTS

EcR-A AND EcR-B TRANSCRIPT VARIANT IDENTIFICATION IN HONEYBEES

Two transcript variants, *EcR-A* (Accession numbers NM_001098215.2) and *EcR-B* (NM_001159355.1) of 2635 and 2782 nucleotides, respectively, have been identified for the *A. mellifera EcR* gene (Takeuchi et al., 2007 and Watanabe et al., 2010). The difference in nucleotide length was shown to reside within the 5' end, resulting in amino acid sequence variation in the N- modulator A/B domain. Conceptual translation of the nucleotide sequences resulted in a putative *EcR-A* protein consisting of 629 amino acid residues and an *EcR-B* protein of 557 amino acids, both sharing a 452 amino acid sequence in the carboxy terminal (Figure 1A). Northern blot analysis using a C-terminal *EcR* probe showed hybridization bands of approximately 2.7 kb and 2.6 kb (Figure 1B), mainly in queen samples, but since we did not aim at quantifying, the respective band



density does not necessarily represent difference in transcript levels between the two castes. Nonetheless, this result reveals that the two transcripts indeed have small differences in length, this supporting the *in silico* evidence.

DEVELOPMENTAL PROFILES OF THE EcR TRANSCRIPT VARIANTS A AND B

Using variant-specific primers we quantified the transcript levels of *EcR-A* and *EcR-B* covering the entire postembryonic development for honeybee queens and workers (Figure 2). Three major findings are worthy of note: (i) transcripts representing the *EcR-B* variant are predominant in embryos (Mann–Whitney Test, $P \leq 0.05$), but these transcript levels decline at the transition to the first larval instar, and it is the *EcR-A* variant which is then predominantly expressed during the end of the larval stage (fifth instar) and pupal stage; (ii) at several time-points, *EcR* expression is higher in workers than in queens (Mann–Whitney Test, $P \leq 0.05$); and (iii) there is a clear discrepancy between circulating ecdysteroid levels and the developmental expression of *EcR-A*.

Major caste differences in *EcR-A* expression were seen to accompany the larval/pupal metamorphic molt. As soon as the larvae were no longer fed by nurse bees and the brood cells were closed, the *EcR-A* levels were increased by two orders of magnitude in cocoon-spinning worker larvae (S1–S3 phases). A rise was also seen in *EcR-A* levels in cocoon spinning queen larvae, but this was significantly lower than in workers (Mann–Whitney Test, $P \leq 0.05$). A similar pattern was also seen for

the *EcR-B* variant, but at much lower modulation. Interestingly, the *EcR* expression levels were then decreased for both variants and in both castes at the onset of the prepupal development (PP1), marked by the appearance of apolysis fluid separating the fifth instar larval cuticle from the newly synthesized pupal cuticle in the head region. A new rise in the transcript levels of both variants was then seen at the end of the prepupal development (PP3), but this was primarily evident in workers (Mann–Whitney Test, $P \leq 0.05$). *EcR-A* and *EcR-B* transcript variants remained at low levels during the pupal and early pharate-adult stages (Pw to Pb1 phases) before they showed another steady increase, but again mainly so in workers (Mann–Whitney Test, $P \leq 0.05$).

TRANSCRIPTIONAL RESPONSE OF EcR TO ARTIFICIALLY AUGMENTED ECDYSTEROID AND JH TITERS

So as to better understand the relationship between hemolymph hormone titers and hormone receptor expression, especially the remarkable divergence in the pupal stage, we treated Pb-phase workers and queens, as these are at the transition from pupal development *per se* to the pharate adult stage, with JH and 20E. At the Pb-phase the ecdysteroid titer is rapidly declining in both castes after having gone through the maximum peak at the preceding Pp phase (Pinto et al., 2002), while JH levels are still basal (Rembold, 1987). The transcriptional responses for the two *EcR* variants assayed by RT-qPCR revealed a general repressive effect of both hormones at 24 h after application (Figure 3). In queens, 20E injection elicited a repressive effect on both *EcR* variants.

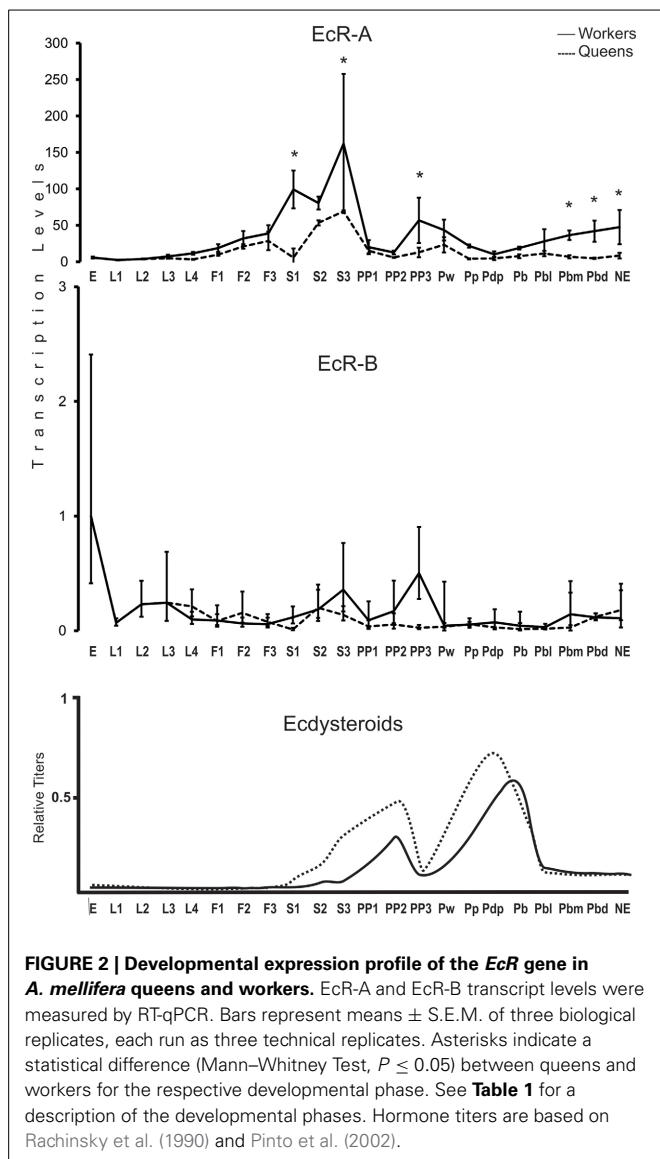


FIGURE 2 | Developmental expression profile of the *EcR* gene in *A. mellifera* queens and workers. *EcR-A* and *EcR-B* transcript levels were measured by RT-qPCR. Bars represent means \pm S.E.M. of three biological replicates, each run as three technical replicates. Asterisks indicate a statistical difference (Mann–Whitney Test, $P \leq 0.05$) between queens and workers for the respective developmental phase. See Table 1 for a description of the developmental phases. Hormone titers are based on Rachinsky et al. (1990) and Pinto et al. (2002).

Mean transcript levels were diminished at 12 h after 20E injection and were significantly lower at 24 h (Mann–Whitney Test, $P \leq 0.05$). In workers this was the case only for the *EcR-B* transcript and only at 24 h (Figure 3A).

The effect of exogenous JH on *EcR* expression was not as clear-cut as that elicited by 20E treatment. While there was no apparent effect on *EcR-A* transcripts in workers, the *EcR-B* levels showed slightly elevated means at all time points (Figure 3B), and these were significantly higher at 12 h following hormone treatment (Mann–Whitney Test, $P \leq 0.05$). Interestingly, in the queen caste the response to JH treatment appeared to be opposite to that seen in workers, with mean *EcR-A* and *EcR-B* transcript levels diminished already at 1 h after treatment and significant differences apparent at 1 h in the case of *EcR-B* and at 24 h for *EcR-A* (Mann–Whitney Test, $P \leq 0.05$). These results indicate a repressor effect of high circulating ecdysteroid levels on *EcR* expression in both castes and a differential response to JH, with workers responding positively and queens negatively to elevated JH levels.

EcR KNOCKDOWN IN PHARATE-ADULT HONEYBEE WORKERS SIGNIFICANTLY DOWNREGULATES THE EXPRESSION OF CANDIDATE TARGET GENES

So as to understand the role of the *EcR* gene in honeybee development, beyond the correlation analysis between transcript levels of the two *EcR* variants and hormone levels, we experimentally decreased the *EcR* gene functionality by an RNA interference approach. We herein focused on the *EcR* response in workers during the pharate-adult to adult transition because only one of the two transcript variants, viz. *EcR-A*, undergoes a gradual increase at this developmental interval, and only so in the worker caste (Figure 2). We expected this to give not only more clear-cut results and insights into the role of the predominant *EcR* variant, but also into still very little understood aspects of morphogenetic processes taking place in developing adult honeybees.

The dsRNA fragment used in this experiment represented an *EcR* region shared by the two transcript variants and its injection resulted in a reduction of 79.8 and 74.9% for *EcR-A* and *EcR-B* mRNA levels, respectively ($P < 0.001$, Student's *t*-test; see Figure 4). A mortality of 10% was observed in both *EcR*- (KD) and GFP-dsRNA treated (control) bees. A proportion of dsRNA-injected bees showed alterations in cuticle pigmentation and wing development, similar to previously reported observations by Barchuk et al. (2008) when studying *ultraspiracle* function. Based on the strong knockdown response we next assayed the transcriptional response of four candidate target genes, these being a homolog of the *D. melanogaster ftz-f1* gene, the *vg* gene, and two genes involved in adult cuticle formation (*AmelCPR14* and *BursA*). The *ftz-f1* gene was included in this analysis because in *D. melanogaster* it acts as a competence factor for the response to 20E; furthermore, *EcR* also inhibits *ftz-f1* expression in *D. melanogaster* mid-prepupa, thus temporarily impairing the larval-to-pupal transition in response to the second 20E peak (King-Jones and Thummel, 2005). In pharate-adult honeybees, the levels of *ftz-f1* transcripts were seen to increase (data not shown) concomitantly with the levels of *EcR-A*, suggesting a synergistic action of the two genes. In addition, the increase in the levels of the two genes coincides with the increase in the expression of genes encoding enzymes and proteins needed for the complete differentiation of the adult cuticle (Soares et al., 2007, 2011, 2013; Elias-Neto et al., 2010). Similarly, in *D. melanogaster* the expression of *ftz-f1* has recently been related to adult cuticle formation and eclosion (Sultan et al., 2014). The analysis of *ftz-f1* transcript levels in newly emerged workers ($N = 12$), i.e., approximately 2 days after injecting dsRNAs, showed that *ftz-f1* expression was significantly decreased in *EcR*-KD bees ($P \leq 0.05$, Student's *t*-test) (Figure 4). A significant effect of the *EcR* knockdown was also seen for the cuticular protein gene *AmelCPR14*, but not for the *BursA* gene that encodes a subunit of the neurohormone Bursicon. The significant reduction in the expression of a cuticular protein gene following *EcR*-RNAi is consistent with the ecdysteroid-related expression of these genes in developing honeybees (Soares et al., 2007, 2011, 2013). An interesting though not easily explained finding was that *vg* gene expression was not significantly affected by reducing *EcR* function, although the mean *vg* transcript levels were slightly reduced

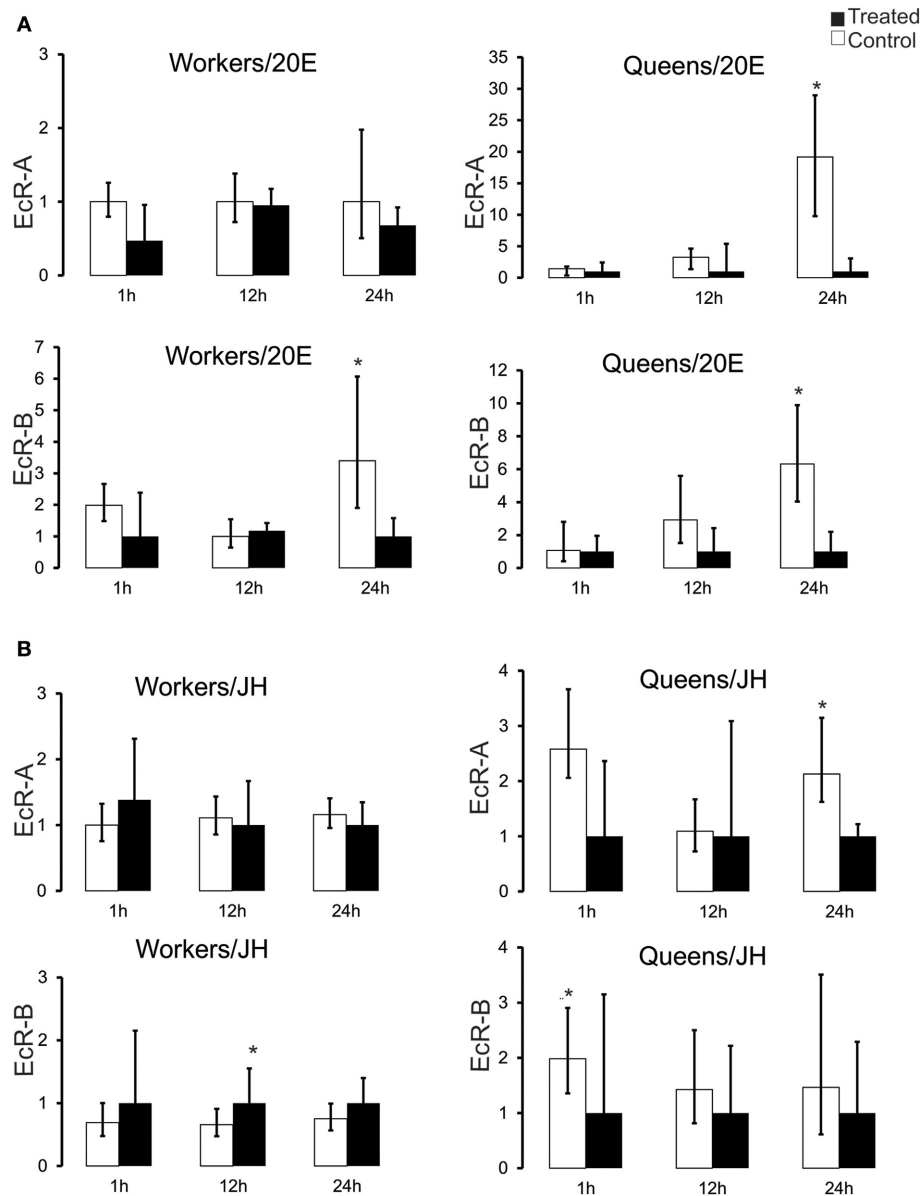


FIGURE 3 | EcR expression response to experimentally augmented levels of (A) 20-hydroxyecdysone (20E) and (B) juvenile hormone (JH) in honeybee castes. Aliquots of 5 μ g of 20E or 10 μ g of JH-III were applied to brown-eyed pharate-adults with unpigmented cuticle (Pb phase). RNA samples from fat bodies were obtained after 1, 12, and 24 h after hormone

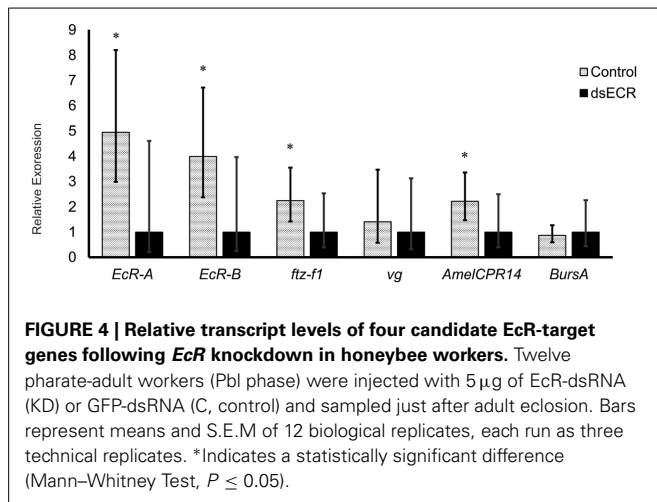
applications. Bars represent means \pm S.E.M. of three samples. Each biological replicate consisted of 3–7 Pb-phase workers and each was run as three technical replicates. Statistical differences (* $P = 0.05$) in gene expression between treated and control groups were assessed by Mann–Whitney Test.

compared to the non-target dsRNA control. This was surprising as *vg* gene expression has been shown to gradually increase in pharate-adult honeybee females, and this increase was thought to be related to ecdysone levels (Barchuk et al., 2002; Piulachs et al., 2003).

EcR KNOCKDOWN AFFECTS THE POLY-A⁺ PROFILE OF NEWLY EMERGED WORKERS

So as to understand EcR functions during the pharate-adult to adult transition of honeybee workers on a more global scale we compared the poly-A⁺ transcriptomes of EcR-KD and

GFP-injected (control) bees. After filtering of the raw data we obtained 112,659,148 reads for the KD and 71,050,536 reads for the control samples. Most of these reads were 50 nt long. This data has been submitted to the Sequence Read Archive (SRA, NCBI, <http://www.ncbi.nlm.nih.gov/sra>) under the Accession Number SRX700299. As we had only one RNA sample set per group (two libraries, no replicates), the estimate obtained by Cuffdiff analysis was that 234 loci were differentially expressed [absolute log₂ (fold change) > 1; q -value = 0.05; FPKM > 5 in at least one library] (Table S2). Among these, 121 code for known protein products, and 100 of these were upregulated in KD pharate-adults and 21



were downregulated (overexpressed in control bees; **Table 2**). The five times higher number of differentially expressed genes in the EcR-KD group indicates that during the pharate-adult to adult transition more genes may be repressed by ecdysone than are induced.

In terms of functional assignments the following conclusions can be drawn. Seven genes among the ones upregulated in the KD group code for cytochrome P450 proteins (**Table 2** and Figure S1A), six of these belonging to CYP clade 3 (CYP6AS2, CYP6AS3, CYP6AS4, CYP6AS5, CYP6AS12, and CYP6BD1) and one (CYP305D1) to CYP clade 2 [for clade assignments of honeybee cytochrome P450 genes see (Claudianos et al., 2006)]. A second protein family that was well-represented among the upregulated genes in the KD group is that encoding Major Royal Jelly Proteins (MRJP1 and MRJP9) and an MRJP-associated protein, apisimin. A third class is represented by hormone response-related genes: a gene encoding a JH-inducible protein, a gene encoding a honeybee eclosion hormone (EH) homolog, and *krüppel-homolog 1*, an immediate response gene regulated by the JH receptor (Bellés and Santos, 2014). Nonetheless, the genes with the highest differential expression index are three genes encoding transcripts of unknown function and without conserved domain evidence (LOC100576540, LOC727013, and LOC727546). The fourth highest upregulated gene in the KD group encodes an α -glucosidase, an enzyme that converts the disaccharide sucrose into glucose and fructose and is, thus, critically involved in carbohydrate metabolism. Another three genes in the top gene list are also related to metabolic functions, these being transcripts for a glycine N-methyltransferase-like, a glycine-methanol-choline (GMC) oxidoreductase 3 and a lipase 3-like protein. Furthermore, three genes upregulated in KD bees, the GMC oxidoreductase 3, a UDP-glycosyltransferase (LOC 413043) and a glucuronosyltransferase (LOC 725997), could be related to ecdysteroid metabolism and function.

The genes downregulated in EcR-KD bees are listed at the bottom of **Table 2**. They are represented with positive fold change values, as these were calculated as relative to the control group. In contrast to the upregulated genes, those that were downregulated are not as clearly associated to putative functions during

the pharate-adult to adult transition, except for the LOC724735 and *Grp* genes that encode structural cuticle proteins needed for the construction of the adult cuticle at this stage. The gene with the highest overexpression index in the control group codes for a Niemann-Pick type protein (NPC2), that is, genes involved in cholesterol metabolism-related syndromes and diseases (another *npc2*-type gene was found slightly overexpressed in the KD bees). Next are three transcripts possibly related to venom gland function, encoding a phospholipase, scapin and a putative mast cell degranulating peptide (**Table 2** and Figure S1A). Also down-regulated was the *brown* gene, which encodes an ABC-2 type transporter protein, and a gene coding for a Major Royal Jelly Protein (MRJP3).

A more global analysis on the entire set of differentially expressed genes was done based on Gene Ontology (Blast2GO and InterProScan) using Fisher and Kolmogorov–Smirnov statistics. This confirmed that the poly-A⁺ RNAs representing genes upregulated in the KD group are enriched in proteins participating in metabolic pathways, particularly ones with catalytic and oxidoreductase activities (Table S3).

So as to validate the poly-A⁺ RNA-sequencing results we then chose two genes revealed as upregulated in the KD group (*Cyp6as5*, a P450 protein coding gene, and *kr-h1*, a gene encoding the JH response factor Krüppel homolog-1) and two down-regulated genes (LOC406145, *secp* and LOC724386, *npc2*). For these we designed or selected from the literature gene-specific primers and ran RT-qPCR assays. The expression pattern was confirmed for all four genes (Figure S1A), thus providing further evidence that the 234 poly-A⁺ RNA coding genes found as differentially expressed between treated and control bees are under EcR control.

EcR KNOCKDOWN AFFECTS THE miRNA PROFILE OF NEWLY EMERGED WORKERS

We obtained a total of 31,171,886 and 33,683,147 reads of small RNAs from the KD and control sequence libraries, respectively. This data has been submitted to the Sequence Read Archive (SRA, NCBI, <http://www.ncbi.nlm.nih.gov/sra>) under the Accession Number SRX700299. After filtering the raw data, we focused on the discovery of miRNAs linked to the EcR network. A total of 4,436,511 reads of the KD samples (~13.2%) and 10,557,117 reads of the control sample (~33.9%) mapped to known honeybee mature miRNAs (available in miRBase version 19), suggesting that EcR disruption causes a general downregulation of miRNA families. We considered as “expressed” those miRNAs with more than 10 reads represented in at least one library. By doing so we retrieved a total of 132 known miRNAs expressed in newly emerged workers, most of them (124) in both conditions (Table S4). In order to find a set of miRNAs whose transcription is significantly affected by the EcR pathway, we filtered the Cuffdiff results by selecting miRNAs with expression differences higher than 1.2-fold and a q -value < 0.05 between KD and control bees. We found 60 downregulated and 10 upregulated miRNAs in KD samples compared to controls (**Table 3**). These data were then further validated by RT-qPCR assays for the following miRNAs: miR-14, miR-100, miR-125, miR-133, miR-375, miR-3728, and miR-3771 (Figure S1B).

Table 2 | Protein-coding genes (121) that were differentially expressed in EcR knockdown (EcR-KD) bees ($FC \geq 2$; $q\text{-value} \leq 0.0015$).

Gene ID	Symbol	Annotation	EcR-KD	Control	log2 FC	FC
100576540	LOC100576540	Uncharacterized protein	102,312	14,520	-2.817	7.935
727013	LOC727013	Uncharacterized protein	1184,780	181,732	-2.705	7.316
727546	LOC727546	Uncharacterized protein	836,105	128,550	-2.701	7.297
409889	AGLU2	Alpha glucosidase 2	447,420	74,553	-2.585	6.684
724275	LOC724275	FBgn0051324	9486	1630	-2.541	6.455
409677	Cyp6as5	Cytochrome P450 6AS5	159,673	33,817	-2.239	5.014
551405	LOC551405	4-nitrophenylphosphatase-like	12,995	2764	-2.233	4.987
406093	LOC406093	Apisimin	943,098	205,225	-2.200	4.841
727213	LOC727213	FBgn0036592	11,369	2517	-2.175	4.732
552832	LOC552832	Glycine N-methyltransferase-like	97,380	23,780	-2.034	4.137
726965	LOC726965	CHK kinase-like; Protein kinase-like domain; Protein of unknown function DUF227; juvenile hormone-inducible protein	13,029	3194	-2.028	4.115
677664	Obp15	Odorant binding protein 15	44,643	11,157	-2.001	4.002
410747	GMCOX3	GMC oxidoreductase 3	14,032	3601	-1.962	3.850
409628	CDase	Neutral ceramidase	90,558	23,540	-1.944	3.778
411353	LOC411353	Lipase 3-like	358,486	93,415	-1.940	3.764
406069	Kr-h1	Krüppel homolog 1	19,403	5207	-1.898	3.601
725165	LOC725165	Aquaporin-4-like	73,542	19,826	-1.891	3.577
412209	CYP6AS4	Cytochrome P450 6AS4, transcript variant 1	154,962	42,262	-1.874	3.514
552388	LOC552388	Major royal jelly protein 1-like	34,026	9661	-1.816	3.299
413908	CYP6AS12	Cytochrome P450 6AS12	71,569	20,875	-1.778	3.160
724312	LOC724312	Vanin-like protein 1-like	16,428	4810	-1.772	3.141
100578616	LOC100578616	Uncharacterized protein	65,980	19,436	-1.763	3.109
726377	LOC726377	Eclosion hormone-like	7230	2141	-1.756	3.083
100576979	LOC100576979	Apidaecins type 73-like	41,592	12,496	-1.735	3.010
411257	Hbg2	Alpha-glucosidase	213,375	66,083	-1.691	2.860
409709	LOC409709	Glucocerebrosidase	59,543	18,587	-1.680	2.821
724367	LOC724367	Protein lethal(2)essential for life-like	61,621	19,247	-1.679	2.818
726372	LOC726372	Trypsin-1-like	267,513	83,558	-1.679	2.818
411330	LOC411330	EF-hand domain; EF-hand-like domain; EF-Hand 1, calcium-binding site	16,958	5405	-1.649	2.721
726362	LOC726362	Luciferin 4-monooxygenase-like	18,728	6023	-1.637	2.679
408379	TpnCIIIa	Troponin C type IIIa	384,660	126,632	-1.603	2.569
411188	LOC411188	Lactate dehydrogenase	12,273	4046	-1.601	2.563
100578106	LOC100578106	Leucine-rich repeat, cysteine-containing subtype	198,220	65,924	-1.588	2.522
100576902	LOC100576902	Uncharacterized protein	7740	2611	-1.567	2.457
408788	LOC408788	Glucuronosyltransferase	13,777	4672	-1.560	2.434
724126	LOC724126	GNAT domain; Acyl-CoA N-acyltransferase	38,134	12,950	-1.558	2.428
551223	CYP305D1	Cytochrome P450 305D1	17,548	5976	-1.554	2.415
100576134	LOC100576134	FBgn0036665	37,976	13,190	-1.526	2.327
409873	Mrjp9	Major royal jelly protein 9	47,842	16,644	-1.523	2.320
552320	LOC552320	FBgn0263216	38,886	13,772	-1.497	2.242
409626	SP40	Serine protease 40	251,004	89,994	-1.480	2.190
408395	LOC408395	Venom carboxylesterase-6-like	39,507	14,192	-1.477	2.182
406142	LOC406142	Hymenoptaecin	1436,990	516,557	-1.476	2.179

(Continued)

Table 2 | Continued

Gene ID	Symbol	Annotation	EcR-KD	Control	log2 FC	FC
726412	LOC726412	Solute carrier family 22 member 8-like, transcript variant 1; solute carrier family 22 member 8-like, transcript variant 2	9953	3585	−1.473	2.170
726737	LOC726737	Venom acid phosphatase Acph-1-like	7268	2629	−1.467	2.153
551560	CYP6BD1	Cytochrome P450 6BD1	106,633	39,084	−1.448	2.097
100577477	LOC100577477	Uncharacterized protein	5897	2167	−1.444	2.085
726611	LOC726611	Uncharacterized protein	75,931	28,272	−1.425	2.031
100576555	LOC100576555	Cytochrome b561, eukaryote; Cytochrome b561/ferric reductase transmembrane	36,621	13,730	−1.415	2.003
725381	LOC725381	Uncharacterized protein	163,884	62,397	−1.393	1.941
409716	apd-3	Apidermin 3	70,996	27,036	−1.393	1.940
724239	LOC724239	Kynurenine/alpha-aminoadipate aminotransferase, mitochondrial-like	8440	3274	−1.366	1.867
100576895	LOC100576895	Fatty acyl-CoA reductase	10,988	4279	−1.361	1.851
100578995	LOC100578995	Vanin-like protein 1-like	5902	2301	−1.359	1.846
413043	LOC413043	UDP-glycosyltransferase	293,945	116,620	−1.334	1.779
413575	LOC413575	Facilitated trehalose transporter Tret1-like	128,316	51,327	−1.322	1.747
725114	LOC725114	Trypsin Inhibitor-like, cysteine rich domain	6914	2772	−1.319	1.739
725273	CTL1	C-type lectin 1	43,402	17,672	−1.296	1.680
725204	LOC725204	Tyrosine aminotransferase-like	8411	3454	−1.284	1.648
726934	SPH50	Serine protease homolog 50	13,552	5672	−1.257	1.579
724993	LOC724993	FBgn0036202	213,583	89,752	−1.251	1.564
725997	LOC725997	Glucuronosyltransferase	26,069	10,989	−1.246	1.553
551458	LOC551458	Leucine-rich repeat-containing protein 20-like, transcript variant 2; leucine-rich repeat-containing protein 20-like, transcript variant 1	107,500	46,024	−1.224	1.498
724756	LOC724756	Gadd45	92,117	39,653	−1.216	1.479
100578635	LOC100578635	Uncharacterized protein	822,735	356,743	−1.206	1.453
413117	LOC413117	Proton-coupled amino acid transporter 4-like	41,718	18,176	−1.199	1.437
552366	LOC552366	Hypothetical LOC552366	15,048	6559	−1.198	1.435
724721	LOC724721	Dehydrogenase/reductase SDR family member 11-like	356,294	158,024	−1.173	1.376
100578329	LOC100578329	Putative fatty acyl-CoA reductase CG5065-like	44,956	19,991	−1.169	1.367
412458	LOC412458	Dehydrogenase/reductase SDR family member 11-like	337,026	151,686	−1.152	1.327
726803	LOC726803	FBgn0037126	305,241	139,173	−1.133	1.284
406115	Apid73	Apidaecin	38,277	17,505	−1.129	1.274
552301	SP35	Serine protease 35	156,146	71,778	−1.121	1.257
724654	LOC724654	Cytochrome b5 type B-like	285,205	131,483	−1.117	1.248
409740	LOC409740	Clavesin-1-like	18,383	8491	−1.114	1.242
724899	Lys-2	Lysozyme 2	15,635	7245	−1.110	1.232
552193	LOC552193	Proton-coupled amino acid transporter 4-like	12,681	5892	−1.106	1.223
406065	Wat	Worker-enriched antennal transcript	187,038	87,451	−1.097	1.203

(Continued)

Table 2 | Continued

Gene ID	Symbol	Annotation	EcR-KD	Control	log2 FC	FC
408622	LOC408622	Phenylalanine hydroxylase, transcript variant 2	19,241	9009	−1.095	1.199
725038	LOC725038	Protein npc2 homolog, similar to that of GenelD724386	619,928	290,849	−1.092	1.192
551094	LOC551094	Fatty-acid amide hydrolase 2-A-like, transcript variant 2	6524	3061	−1.092	1.191
408807	LOC408807	Hypothetical LOC408807	22,855	10,746	−1.089	1.185
409638	LOC409638	Elongation of very long chain fatty acids protein AAEL008004-like	112,535	52,938	−1.088	1.184
406109	Obp6	Odorant binding protein 6	12,054	5672	−1.088	1.183
727522	LOC727522	Insect allergen-related	16,753	7972	−1.071	1.148
100577337	LOC100577337	Glucose dehydrogenase	6428	3072	−1.065	1.134
100577132	LOC100577132	Calcium calmodulin-dependent protein kinase kinase 2	7270	3491	−1.058	1.120
408299	LOC408299	Purine nucleoside phosphorylase-like	159,071	76,456	−1.057	1.117
408733	LOC408733	Pinocchio	210,857	101,474	−1.055	1.113
727367	LOC727367	Glucose dehydrogenase [acceptor]-like	29,151	14,038	−1.054	1.111
410087	LOC410087	Protein lethal(2)essential for life-like, transcript variant 1	101,566	48,915	−1.054	1.111
100576662	LOC100576662	Uncharacterized protein	60,105	29,083	−1.047	1.097
411615	CYP6AS2	Cytochrome P450 6AS2	76,760	37,313	−1.041	1.083
551044	Gld2	Glucose dehydrogenase 2, transcript variant 1	5627	2756	−1.030	1.060
726871	LOC726871	Synaptic vesicle glycoprotein 2C-like	77,575	38,301	−1.018	1.037
406144	LOC406144	Abaecin	979,123	484,043	−1.016	1.033
408868	LOC408868	Inositol-1-(or 4-)monophosphatase	87,544	43,471	−1.010	1.020
408421	LOC408421	CHK kinase-like; Protein kinase-like domain; Protein of unknown function DUF227	391,146	194,281	−1.010	1.019
725344	LOC725344	Histone H2B.3-like	13,796	6863	−1.007	1.015
726690	CYP6AS3	Cytochrome P450 6AS3	34,408	17,138	−1.006	1.011
725217	LOC725217	Armadillo-type fold	9786	19,784	1.016	1.031
552421	LOC552421	Glycogenin-1-like	37,050	75,154	1.020	1.041
100577901	LOC100577901	FBgn0030763	18,696	38,195	1.031	1.062
100578838	LOC100578838	Chymotrypsin inhibitor-like	48,355	99,598	1.042	1.087
100579045	LOC100579045	Uncharacterized protein	3038	6320	1.057	1.116
412399	LOC412399	Organic cation transporter protein-like	8913	19,674	1.142	1.305
412797	LOC412797	Facilitated trehalose transporter Tret1-like, transcript variant 1	7785	17,708	1.186	1.406
100578811	LOC100578811	Transmembrane protein 223-like	7839	18,013	1.200	1.441
724735	LOC724735	Endocuticle structural glycoprotein SgAbd-2-like	2941	7068	1.265	1.600
678674	LOC678674	Venom allergen Api m 6	1994	5014	1.330	1.769
552281	Grp	Glycine-rich cuticle protein	7302	18,853	1.369	1.873
100578552	LOC100578552	Chitin binding domain	4141	10,943	1.402	1.965
406145	LOC406145	Secapin	32,484	92,219	1.505	2.266
100578873	LOC100578873	Allergen Api m 6-like	1799	5256	1.547	2.393
100576769	LOC100576769	Mast cell degranulating peptide-like	31,784	95,466	1.587	2.518
406141	Pla2	Phospholipase A2	5068	15,586	1.621	2.627

(Continued)

Table 2 | Continued

Gene ID	Symbol	Annotation	EcR-KD	Control	log2 FC	FC
100578863	LOC100578863	Uncharacterized protein	1641	5082	1.631	2.660
725163	LOC725163	Trypsin Inhibitor-like, cysteine rich domain	5520	22,669	2.038	4.153
406121	Mrjp3	Major royal jelly protein 3	1102	5592	2.343	5.491
412203	bw	Brown	1347	7074	2.392	5.724
724386	LOC724386	Protein npc2 homolog	757,358	4202,210	2.472	6.111

Negative values of log2 FC indicate genes that are upregulated in EcR-KD bees (highlighted in red); positive values indicate downregulation in EcR-KD bees (highlighted in green). Genes shown in bold had their transcription patterns validated by qPCR. Expression values measured as FPKM (Fragments Per Kilobase of exon per Million fragments mapped). This list contains only the genes with FPKM values of 5 for any of the two samples.

DISCUSSION

THE HONEYBEE EcR TRANSCRIPT VARIANTS AND THEIR DEVELOPMENTAL REGULATION

The existence of more than one EcR isoform is commonplace in insects, including the honeybee, for which two transcript variants, *EcR-A* and *EcR-B* had been found (Takeuchi et al., 2007). First shown for *D. melanogaster* (Talbot et al., 1993) and then for the red flour beetle *Tenebrio molitor* (Mouillet et al., 1997), the extensive review of insect EcR isoforms by Watanabe et al. (2010) showed a high similarity in their nucleotide and amino acid sequences in most of their functional domains, except for the N-terminal region including the variable A/B modulator domain, which might allow for the recruitment of different co-activators/co-repressors (Tora et al., 1988; Kato et al., 1995; Watanabe et al., 2010).

First, we confirmed by northern blotting the expression of the two *EcR* variants in honeybee queens and workers. Then, we compared their temporal expression profiles to the hemolymph ecdysteroid titers of fifth instar queen and worker larvae (F1-PP3 phases) (Rachinsky et al., 1990). The results for the developmental expression profiles of the two ecdysteroid receptor variants are surprising in two aspects. First, contrasting with the hormone titers, which are higher in queens than in workers, the *EcR* transcript levels were found to be higher in workers, especially so for *EcR-A*. Second, there was a marked drop in *EcR* expression at the beginning of the prepupal phase (PP1), i.e., exactly when the hemolymph ecdysteroid levels increase to reach a developmental peak at the subsequent PP2 phase. Strikingly as well, the transcript levels for both EcR variants remained at low or basal levels during the pupal and early pharate-adult stages (Pw to Pbl phases), even though the ecdysteroid hemolymph titers are at a maximum during this period (Feldlaufer et al., 1985; Pinto et al., 2002).

The switch from EcR-B expression in the embryonic stage to EcR-A as the predominant isoform in the fifth larval instar and pupal stage reflects a change in the processing of an eventual long pre-mRNA, or a shift in transcription start site utilization (our RNA-Seq data are in support of the latter possibility and even suggest the existence of a third EcR transcript variant). Since the *EcR* gene is known to be induced after an ecdysteroid pulse (Karim and Thummel, 1992; Davis and Li, 2013), the production of *EcR-B* mRNA in honeybee embryos would require the presence of steroid hormones, which is indeed the case. Makisterone

A, the predominant ecdysteroid in *A. mellifera* (Feldlaufer et al., 1985), has been shown to be present in ovaries in quite large amounts (Feldlaufer et al., 1986a), and unpublished data from our laboratory also confirm the presence of ecdysteroids in developing embryos. High levels of ecdysteroids in ovaries have also been shown for bumblebee queens (Geva et al., 2005) and queens of a swarm-founding neotropical wasp, *Polybia micans* (Kelstrup et al., 2014). Embryonic ecdysteroids can be synthesized by enzymatic conversion from inactive conjugates stored during oogenesis (Dorn, 2000) or, as seen in mosquitoes, transferred by males during copulation (Baldini et al., 2013).

Since makisterone A is the predominant ecdysteroid compound in queen ovaries (Feldlaufer et al., 1986a) and also in pupal-stage hemolymph (Feldlaufer et al., 1986b) and 20E is not negligible in prepupal hemolymph (Rachinsky et al., 1990), the observed embryonic-to-larval EcR isoform switch may be linked to variation in the ecdysteroid composition circulating in the hemolymph, throughout a bee’s life cycle. This could not only be responsible for the observed differential *EcR* transcription, but also for the formation of different hormone/receptor complexes with potentially different target genes thus, governing separate physiological processes. 20E, for example, might have retained a role in reproductive physiology, as suggested by Takeuchi et al. (2007), whereas makisterone A may have been co-opted for governing postembryonic development, as suggested for *Dysdercus fasciatus* (Feldlaufer et al., 1991). Nonetheless, for honeybees such “division of labor” in ecdysteroid compounds is still highly speculative, especially since the ecdysteroid hemolymph levels in adult honeybee queens and workers are continuously low, this making it rather unlikely that these steroid hormones may play a major role in the reproductive female physiology (Hartfelder et al., 2002). Instead, they seem to be preferentially stored in the developing follicles.

The second and third major findings mentioned above are that the *EcR-A* transcript levels are higher in worker than in queen development, and that there is no positive, but rather an apparently negative correlation between hormone levels and hormone receptor transcript levels. This stands in stark contrast to the developmental pattern of the hemolymph ecdysteroid titers in the two castes, which are higher in queens than in workers, particularly so during larval-pupal metamorphosis (Rachinsky et al., 1990). The ecdysteroid titer in last instar queen larva rises as soon as the brood cells are closed and the larvae start to spin their

Table 3 | miRNAs that were differentially expressed in EcR-knockdown bees.

Effect	miRNA	Read counts		Normalized number of reads		Fold change
		Control	KD	Control	KD	
Downregulated	ame-miR-3771-3p	12	1	0.79	0.13	6.06
	ame-miR-6067-5p	19	2	1.25	0.25	4.99
	ame-miR-6046-3p	31	6	2.04	0.76	2.68
	ame-miR-6057-5p	81	17	5.34	2.16	2.48
	ame-miR-100-5p	1,117,633	257,905	73,616.85	32,712.83	2.25
	ame-miR-3770-5p	57	14	3.75	1.78	2.11
	ame-miR-133-3p	7726	2092	508.90	265.35	1.92
	ame-miR-375-3p	64,153	17,597	4225.66	2232.01	1.89
	ame-miR-1-3p	78,433	21,559	5166.27	2734.56	1.89
	ame-miR-6043-3p	352	98	23.19	12.43	1.87
	ame-miR-2765-5p	1903	548	125.35	69.51	1.80
	ame-miR-125-5p	118,218	35,904	7786.85	4554.09	1.71
	ame-miR-6047a-3p	298	92	19.63	11.67	1.68
	ame-miR-927a-5p	129,279	40,144	8515.42	5091.89	1.67
	ame-miR-971-3p	110	35	7.25	4.44	1.64
	ame-miR-87-3p	14,514	4783	956.02	606.68	1.58
	ame-miR-263a-5p	857,064	283,563	56,453.55	35,967.31	1.57
	ame-miR-137-3p	3351	1119	220.73	141.93	1.56
	ame-miR-1175-3p	2179	731	143.53	92.72	1.55
	ame-miR-210-3p	3330	1121	219.34	142.19	1.55
	ame-miR-316-5p	21,370	7287	1407.61	924.29	1.53
	ame-miR-6038-5p	445	158	29.31	20.04	1.46
	ame-miR-219-5p	70	25	4.61	3.17	1.45
	ame-miR-980-3p	235	86	15.48	10.91	1.41
	ame-miR-92a-3p	190	70	12.52	8.88	1.41
	ame-miR-92b-3p	42,256	15,578	2783.34	1975.92	1.40
	ame-miR-279a-3p	55,440	20,620	3651.75	2615.45	1.39
	ame-miR-279c-3p	6610	2482	435.39	314.82	1.39
	ame-let-7-5p	119,020	44,921	7839.67	5697.81	1.38
	ame-miR-989-3p	528	201	34.78	25.49	1.37
	ame-miR-252a-5p	19,053	7232	1254.99	917.31	1.37
	ame-miR-184-3p	1,460,229	568,579	96,183.15	72,118.91	1.34
	ame-miR-6041-3p	95	37	6.26	4.69	1.34
	ame-miR-252b-5p	420	164	27.66	20.80	1.33
	ame-miR-3747b-5p	793	310	52.23	39.32	1.33
	ame-miR-2788-3p	1137	445	74.89	56.44	1.33
	ame-miR-317-3p	6091	2391	401.21	303.28	1.32
	ame-miR-6012-3p	122	48	8.04	6.09	1.32
	ame-miR-3791-3p	697	277	45.91	35.13	1.31
	ame-miR-71-5p	1667	660	109.80	83.71	1.31
	ame-miR-34-5p	3245	1289	213.74	163.50	1.31
	ame-miR-3718a-3p	3817	1517	251.42	192.42	1.31
	ame-miR-927b-5p	3108	1250	204.72	158.55	1.29
	ame-miR-929-5p	3423	1379	225.47	174.91	1.29
	ame-miR-305-5p	62,480	25,699	4115.47	3259.68	1.27
	ame-miR-124-3p	2142	882	141.09	111.87	1.26
	ame-miR-996-3p	44,757	18,495	2948.08	2345.92	1.26
	ame-miR-10-5p	1,146,362	477,869	75,509.19	60,613.20	1.25
	ame-miR-276-3p	808,220	339,646	53,236.27	43,080.91	1.24

(Continued)

Table 3 | Continued

Effect	miRNA	Read counts		Normalized number of reads		Fold change
		Control	KD	Control	KD	
	ame-miR-3477-5p	46,537	19,658	3065.32	2493.43	1.23
	ame-miR-6001-5p	173	73	11.40	9.26	1.23
	ame-miR-283-5p	42,234	17,884	2781.89	2268.42	1.22
	ame-miR-263b-5p	41,279	17,642	2718.99	2237.72	1.21
	ame-miR-14-3p	94,785	40,489	6243.35	5135.65	1.21
	ame-miR-79-3p	3701	1589	243.78	201.55	1.21
	ame-miR-3719-3p	2799	1217	184.37	154.37	1.20
	ame-miR-12-5p	89,841	38,948	5917.70	4940.19	1.20
	ame-miR-279b-3p	650	281	42.81	35.64	1.20
	ame-miR-6039-5p	494	214	32.54	27.14	1.20
	ame-miR-6051-3p	343	149	22.59	18.90	1.20
Upregulated	ame-miR-3728-5p	105	124	6.92	15.73	2.27
	ame-miR-965-5p	21	20	1.38	2.54	1.84
	ame-miR-6005-3p	76	64	5.01	8.12	1.62
	ame-miR-3798-3p	342	259	22.53	32.85	1.45
	ame-miR-6052-5p	43	31	2.83	3.93	1.39
	ame-miR-3761-5p	342	232	22.53	29.43	1.31
	ame-miR-3720-5p	1586	1004	104.47	127.35	1.22
	ame-miR-6001-3p	27,917	17,745	1,838.85	2250.79	1.22
	ame-miR-306-5p	186,621	117,263	12,292.45	14,873.71	1.21
	ame-miR-9a-5p	88,013	55,087	5797.29	6987.27	1.21

Only miRNAs with Fold-change > 1.2 are listed.

cocoons (S1-stage), while in workers this was only seen in the late spinning (S3) to early prepupal (PP1) phases. Furthermore, the peak in edysteroid levels reached during the prepupal phase (PP2) is twice as high in queens compared to workers (Rachinsky et al., 1990). The negative correlation between edysteroid levels and *EcR* expression is particularly evident at two time points: in prepupae, when edysteroid levels are high in the PP1–PP2 phases, just as the *EcR*-A expression pattern undergoes a valley, and in the pupal stage, when the edysteroid levels are high in both castes at the Pp phase, before dropping in the Pb–Pbl phases (Pinto et al., 2002). It is only after this drop in circulating hormone levels that *EcR*-A transcription is resumed, particularly so in the worker caste, and strikingly, it is during the subsequent Pbm and Pbd phases that the edysteroid levels are again lower in workers than in queens (Pinto et al., 2002).

This apparent negative correlation between hormone and hormone receptor levels was suggestive of a repressive action of high concentrations of circulating edysteroids on the expression of their receptor gene. To test this we manipulated the endogenous hormone levels by treating Pb pharate-adults with either 20E or JH. The results of the 20E injection experiments showed that a prolonged and excessive presence of edysteroids had a repressive effect on *EcR*-A and *EcR*-B expression, especially so in queens (Figure 3A). Interestingly, workers seem to be more resilient to this repressor effect, as there was no significant reduction in *EcR*-A transcript levels, comparable to that seen in queens, or for

EcR-B in both castes. Such resilience was also denoted in the JH application experiment, where *EcR*-A transcript levels in workers remained little affected compared to those in queens and for *EcR*-B in both castes 24 h after the 20E injection. Strikingly, JH appeared to have opposite effects on *EcR*-B expression in the two female castes, showing a positive effect in workers and a negative one in queens. These differences of hormone effects on *EcR*-expression related to caste certainly deserve a closer look in future experiments.

Repressive effects of high concentrations of edysteroids on *EcR*-expression are, however, not new and are likely to be a general feature of hormone systems that underly cyclical events in morphogenesis and physiology. For instance, similar results were described for *Manduca sexta*, where low concentrations of 20E induced *EcR* expression while high concentrations repressed the expression of this gene (Jindra et al., 1996). Like ours, these results suggest that the *EcR* gene responds positively to a slight increase in edysteroids, whereas high hormone levels are repressive. In fact, as we could see, *EcR* expression actually appears to precede the rise in hormone levels, for instance in the S1–S3 phases, when the circulating edysteroid levels start increasing (Rachinsky et al., 1990), but *EcR*-A and also *EcR*-B transcript levels have already undergone a steep rise. A repetition of this pattern can be inferred for the pupal ecdysis event, occurring between PP3 and Pw, when the edysteroid titers undergo a sharp drop, but *EcR*-A and *EcR*-B are on the rise (mainly in workers), and drop once the edysteroid

titers build up to maximal values in the Pp phase. It is such cyclical events, the molts, that are synchronized by the ecdysone/ecdysones receptor complex action, and this is primarily seen in the epidermis, the main organ of cuticle synthesis. In the honeybee, several cuticle protein genes were shown to be regulated by ecdysteroids (Soares et al., 2007, 2011, 2013; Elias-Neto et al., 2010).

RNAi-MEDIATED KNOCKDOWN REVEALS EcR REGULATED GENES IN DEVELOPING ADULTS

Upon comparing the sequencing results of the poly-A⁺ libraries for EcR knockdown (EcR-KD) and control groups, the Cutdiff analysis classified 234 loci as differentially expressed. Among these, 121 were annotated as coding for known protein products or, from another point of view, 113, i.e., one half, represent loci for unknown, not annotated products, which could be either proteins or long non-coding RNAs. Especially the latter are still “dark matter” in the honeybee genome, represented by many ESTs in the databases, but only four long non-coding RNAs are so far characterized to some detail (Sawata et al., 2002; Humann et al., 2013).

Among the genes with known orthologs or sequence similarity in functional domains, 100 were overexpressed (fold change > 1) in the EcR-KD group and 21 in the control group, this indicating that apparently more genes are repressed by the ecdysone/EcR receptor complex than are activated. Furthermore, a Gene Ontology and KEGG pathway analysis showed that there is little overlap in gene functions between the two sets of differentially expressed genes (DEGs). As mentioned above, cytochrome P450 genes are strongly represented among the DEGs. While cytochrome P450 genes are a large gene family, strongly related to detoxification processes, this family has undergone considerable reduction in honeybee genome evolution (Claudianos et al., 2006). This reduction is, however, denoted only in certain clades of the P450 enzymes, but not in the clades comprising the genes found in our EcR-RNAi experiment. Unfortunately, there is no further functional or tissue/cell type information available for the five cytochrome P450 genes, especially whether or not they may be related to steroid synthesis or metabolism. Nonetheless, similar findings as the ones we report here were also denoted by Davis and Li (2013) in their genomic screen for ecdysone and EcR-dependent gene expression in *D. melanogaster*.

A second group of overrepresented genes that called attention was the hormone response-related genes, as these may provide a link between JH and ecdysteroid action during the pharate-adult to adult transition in honeybees. For this group we found three genes as overexpressed in the EcR-KD group, viz. a JH-induced protein, *kr-h1*, and an Eclosion hormone-like (*EH-like*) gene. *kr-h1* is certainly the most interesting gene in this set, as it represents a direct readout of the activity of the JH response in target tissues (Lozano and Belles, 2011; Bellés and Santos, 2014). As *kr-h1* has previously been identified in a screen for ecdysone-response genes in *D. melanogaster* (Beckstead et al., 2005), the current identification of this gene in the EcR-KD group provides experimental evidence toward a mechanistic explanation for the modulation of vitellogenin induction in honeybee pharate-adults, where *vg* expression is caste-specifically induced by JH and counteracted by ecdysteroids (Barchuk et al., 2002). Overexpression of an *EH-like*

gene in the EcR-KD group was not unexpected, as EH is synthesized in response to declining ecdysteroid titers and is part of the ecdysis triggering signaling cascade (Zitnan and Adams, 2012). Interestingly, other three upregulated genes in EcR-KD bees may have roles in ecdysteroid metabolism and function. Several GMC oxidoreductase genes in diverse insects, including *A. mellifera*, are clustered in an evolutionary conserved tandem array with potential to be co-regulated for a common function related to ecdysteroid metabolism (Iida et al., 2007). The products of LOC413043 and LOC725997 may regulate ecdysteroid titer and function since the enzymes encoded by these genes catalyze the transfer of glucose from UDP-glucose to ecdysteroids, and thus are possibly related to ecdysteroid inactivation (O'Reilly and Miller, 1989).

Overexpression of members of the Major Royal Jelly Protein (MRJP) family can be interpreted in the context of a repressive action of the ecdysone/EcR receptor complex on genes of the adult honeybee life cycle (the only exception being the gene coding for the MRJP3, which was overexpressed in control bees). The *mrjp* gene family with its nine members is a lineage-specific extension in the genus *Apis*, from a single *mrjp-like* gene within the *yellow* genes complex (Drapeau et al., 2006). Even though these proteins are highly expressed in the hypopharyngeal glands of nurse worker bees, constituting the major protein fraction of the glandular secretions fed to larvae (royal jelly and worker jelly), expression of the *mrjp* genes is neither exclusive to this tissue nor is it restricted to the worker caste. Especially *mrjp9* has been shown to be broadly expressed, in different tissues of adult workers and also in queens and even drones (Buttstedt et al., 2013). In contrast to *mrjp9*, *mrjp1* expression is more tissue-specific, being highest in heads (viz. hypopharyngeal glands) of nurse bees, with expression levels being considerably lower in other body parts, castes and sexes (Buttstedt et al., 2013). MRJP1 is the predominant MRJP moiety in royal jelly, present as oligomers of MRJP1 subunits, which are held together by apisimin, a small 5 kDa protein (Tamura et al., 2009). ESTs corresponding to *apisimin* were found as overrepresented in the EcR-KD group, indicating that its expression is co-regulated with that of *mrjp1*. But this co-regulation is not due to genomic proximity, as the *mrjp/yellow* gene cluster maps to chromosome 11, whereas *apisimin* is located in chromosome 6. Interestingly, an MRJP1 monomer, royalactin, was found to be an important factor in caste development, acting through the Egfr signaling pathway (Kamakura, 2011).

Among the EcR-KD group genes we also identified *obp15*, which encodes a putative odorant binding protein. Some *obp* genes were also found to be under negative EcR control in *D. melanogaster*, including the *obp15* and *obp6* genes (Davis and Li, 2013). Forêt and Maleszka (2006) had previously shown that *obp15* is expressed in the antennae of adult bees and also in young larvae but not in pupae. The high ecdysteroid levels in honeybee hemolymph during the pupal to pharate-adult transition, thus, appear to repress *obp15* expression, and possibly also other members of this complex gene family.

Among the genes overrepresented in the transcriptome of the control group (downregulated in EcR-KD bees), the first in the top ten list is annotated as *npc2*. Genes of this family are associated with Niemann-Pick syndromes and diseases affecting

cholesterol metabolism (Carstea et al., 1997). In *D. melanogaster*, NPC mutations cause intracellular enrichment of cholesterol, reduced ecdysteroidogenesis and death in the first larval instar. The fact that this condition could be fully rescued when an excess of dietary cholesterol was given to these mutants indicated that the ecdysone biosynthesis pathway is intact, but precursor processing is not (Huang et al., 2007). Interestingly, in honeybees, as in other insects, the major ecdysteroid is not ecdysone or its derivative 20E, but makisterone A, an ecdysteroid methylated at C24 (Feldlaufer et al., 1985, 1986a,b; Rachinsky et al., 1990), possibly due to a lack or restriction in C24-demethylation of a phytosterol precursor. The expression of two other genes overrepresented in the transcriptome of the control group has been shown to be dependent on the ecdysteroid titer. The protein encoded by LOC724735, an endocuticle structural protein (Márcia M. G. Bitondi, unpublished results) and also the *Grp* gene, renamed as *tweedle1* (*AmelTwdl1*) (Soares et al., 2011), were induced in the integument by the ecdysteroid pulse that promotes the pupal to pharate-adult transition. Thus, the functionality of the ecdysone/EcR complex is necessary for the activation of these genes. Interestingly, *mrjp3*, the third among the genes overrepresented in control bees (and thus induced by the EcR pathway), encodes one of the main MRJPs produced by nurse bees (Buttstedt et al., 2013). The *mrjp3* gene thus seems to be highly expressed by the time of adult emergence and the first days of adult life. Unlike the *mrjp1* and *mrjp2* genes, *mrjp3* reaches negligible expression levels in foragers, which, together with its distinctive amino acid sequence (Drapeau et al., 2006), supports the notion of its main function as food protein supplier to larvae by nurse bees. However, the fact that *mrjp1*, another MRJP gene highly expressed in nurse bees, was found to be repressed by the ecdysteroid pathway (see above), suggests the *mrjp3* gene is regulated in a distinct mode from the other *mrjp* genes.

miRNAs AS ACTORS IN THE EcR REGULATORY NETWORK

Here we demonstrate that the RNAi-mediated knockdown of *EcR* function perturbs the expression of 70 miRNAs (~1/3 of the honeybee miRNAs known to date). Most of these (60) were downregulated and 10 were upregulated and we assume that these down and upregulated miRNAs are “induced” or “repressed,” respectively, by the EcR pathway as bees undergo the pharate-adult to adult transition.

Among the miRNAs that showed significant changes in abundance following *EcR* knockdown, most had already been identified in a large-scale sequencing project (Chen et al., 2010), but had no function(s) associated. Our data now lead to infer that these miRNAs are, at least, closely associated with EcR action and, consequently, connected to pupal-adult metamorphosis. In addition to these miRNAs of yet unclear functions, we also found conserved and functionally well-defined miRNAs, such as let-7, miR-1, miR-133, miR-375, miR-184, and miR-34. For example, miR-133 and miR-1 are both clustered in the mouse and fly genomes, and they play important roles in muscle development and differentiation in vertebrates and invertebrates (Sokol and Ambros, 2005; Chen et al., 2006; Boutz et al., 2007). In the honeybee, however, we found these two miRNAs to be located far apart from one another on chromosome 16. Nonetheless, they still seem

to be linked in their cooperative functions, such as formation and physiology of flight muscle tissue. miR-133 has also been implicated in dopamine production (Yang et al., 2014), and high levels of dopamine were shown to coincide with rapid growth and compartmentalization of the antennal lobe neuropil, suggesting a role in the developing brain of the honeybee (Kirchhof et al., 1999). Furthermore, dopamine-derivatives are substrates for oxidation by laccases (Andersen, 2010) that are involved in tanning of the developing adult cuticle (Elias-Neto et al., 2010). Members of the *D. melanogaster* let-7-C locus (a cluster containing the let-7, miR-100, and miR-125 genes) are also found in the honeybee genome. In *D. melanogaster* they are expressed in neuromusculature development of pupae and adults, and knockout flies showed disturbances in flight, reproduction and locomotion (Sokol et al., 2008). Moreover, ecdysteroid signaling was shown to be linked to the expression levels of the let-7-C cluster genes, as well as of miR-14 and miR-34 during insect development (for review see Kucherenko and Shcherbata, 2013).

Many of the miRNAs affected by *EcR* knockdown in honeybees (let-7, miR-1, miR-9a, miR-12, miR-14, miR-34, miR-79, miR-92b, miR-124, miR-184, miR-210, miR-219, miR-263a, miR-276, miR-279, miR-283, miR-305, miR-306, miR-316, miR-317) have previously been reported as putatively involved in the regulation of *D. melanogaster* immune genes, particularly those belonging to the JNK, Imd and Toll signaling pathways (Fullaondo and Lee, 2012). Accordingly, ecdysone and the ecdysone receptor complex (EcR/USP) are considered critical for innate cellular immunity (Flatt et al., 2008; Regan et al., 2013). Among these miRNAs, miR-184 is highly and/or broadly expressed in a number of tissues and developmental stages of vertebrates (Wienholds and Plasterk, 2005) and invertebrates (Jagadeeswaran et al., 2010), including *A. mellifera* (Chen et al., 2010; Nunes et al., 2013b). Moreover, several studies reported a wide spectrum of roles for miR-184, such as germline differentiation, axis formation of the egg chamber, anteroposterior patterning and cellularization of the embryo, gastrulation and neuroectoderm formation, apoptosis, and processes involved in the development and differentiation of imaginal discs (head, wing, and eyes) (see Iovino et al., 2009; Li et al., 2011, and references therein). The ecdysone response of miR-184 seen here in pharate-adult honeybees is associated with a period of extensive tissue remodeling, suggesting that miR-184 may play a role in the differentiation of honeybee imaginal disc-derived structures and maintenance of their tissue identities. Interestingly, the EcR mRNA has predicted binding sites for miR-14 (data not shown), and our global gene expression assays revealed a down-regulation of this miRNA in bees silenced for *EcR* gene function. These results suggest that in *A. mellifera*, *EcR* gene expression is regulated in a loop-type mechanism involving miR-14, as already demonstrated for *D. melanogaster* (Varghese and Cohen, 2007; for a comprehensive review see Yamanaka et al., 2013).

CONCLUDING REMARKS

Our results suggest a differential use of EcR isoforms during the honeybee life-cycle stages. We could show that there is a generally positive *EcR* gene response to slight increases in ecdysteroids, whereas high levels of these hormones are repressive. The EcR knockdown experiments revealed that the expression of

several hormone response-related genes (e.g., *kr-h1*) is contingent on a functional ecdysone/EcR receptor complex, thus providing a possible link between JH and ecdysteroid action during preimaginal honeybee development. These knockdown experiments also highlighted the relevance of a set of miRNAs involved in the regulation of immune response genes and in the general morphogenesis processes during pharate-adult development (e.g., *miR-184* and *let-7* locus genes). Within this framework and on the background of current knowledge on honeybee biology, our results highlight the relevance of the drop in the ecdysteroid pathway function for the appropriate timing in the expression of adult-specific genes, such as the Major Royal Jelly Protein (MRJP) family members.

AUTHOR CONTRIBUTIONS

Tathyana R. P. Mello, Aline C. Aleixo, Angel R. Barchuk, and Zilá L. P. Simões conceived the project; Tathyana R. P. Mello, Aline C. Aleixo, Daniel G. Pinheiro, Francis M. F. Nunes, Klaus Hartfelder, and Angel R. Barchuk performed the experiments; Tathyana R. P. Mello, Aline C. Aleixo, Daniel G. Pinheiro, Francis M. F. Nunes, Márcia M. G. Bitondi, Klaus Hartfelder, Angel R. Barchuk, and Zilá L. P. Simões analyzed and interpreted the data and drafted the MS. All authors approved the final version of the MS.

FUNDING

This study was funded by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), grants 2011/03171-5; 2008/1446-4; 2008/10757-3.

ACKNOWLEDGMENTS

We thank Dr. Nilce M. M. Rossi, Roseli de Aquino P. Ferreira and Diana Gras for lab facility and support in Northern blotting assays, and Luiz Aguiar for technical assistance in the apiary. We also thank Felipe Martelli, Denyse C. Lago, Fabiano Abreu, and Camilla Valente Pires for help with the qPCR assays.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/journal/10.3389/fgene.2014.00445/abstract>

REFERENCES

- Amdam, G. V., Simões, Z. L. P., Guidugli, K. R., Norberg, K., and Omholt, S. W. (2003). Disruption of vitellogenin gene function in adult honeybees by intra-abdominal injection of double-stranded RNA. *BMC Biotechnol.* 3:e1. doi: 10.1186/1472-6750-3-1
- Andersen, S. O. (2010). Insect cuticular sclerotization: a review. *Insect Biochem. Mol. Biol.* 40, 166–178. doi: 10.1016/j.ibmb.2009.10.007
- Ashburner, M., Chihara, C., Meltzer, P., and Richards, G. (1974). Temporal control of puffing activity in polytene chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* 38, 655–662. doi: 10.1101/SQB.1974.038.01.070
- Audic, S., and Claverie, J. M. (1997). The significance of digital gene expression profiles. *Genome Res.* 7, 986–995.
- Baldini, F., Gabrieli, P., South, A., Valim, C., Mancini, F., and Catteruccia, F. (2013). The interaction between a sexually transferred steroid hormone and a female protein regulates oogenesis in the malaria mosquito *Anopheles gambiae*. *PLoS Biol.* 11:e1001695. doi: 10.1371/journal.pbio.1001695
- Barchuk, A. R., Bitondi, M. M. G., and Simões, Z. L. P. (2002). Effects of juvenile hormone and ecdysone on the timing of vitellogenin appearance in hemolymph of queen and worker pupae of *Apis mellifera*. *J. Insect Sci.* 2, 1–8. doi: 10.1673/031.002.0101
- Barchuk, A. R., Cristino, A. S., Kucharski, R., Costa, L. F., Simões, Z. L. P., and Maleszka, R. (2007). Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Dev. Biol.* 7:70. doi: 10.1186/1471-213X-7-70
- Barchuk, A. R., Figueiredo, V. L. C., and Simões, Z. L. P. (2008). Downregulation of *ultraspiracle* gene expression delays pupal development in honeybees. *J. Insect Physiol.* 54, 1035–1040. doi: 10.1016/j.jinsphys.2008.04.006
- Barchuk, A. R., Maleszka, R., and Simões, Z. L. P. (2004). *Apis mellifera* *ultraspiracle*: cDNA sequence and rapid up-regulation by juvenile hormone. *Insect Mol. Biol.* 13, 459–467. doi: 10.1111/j.0962-1075.2004.00506.x
- Beckstead, R. B., Lam, G., and Thummel, C. S. (2005). The genomic response to 20-hydroxyecdysone at the onset of *Drosophila* metamorphosis. *Genome Biol.* 6:R99. doi: 10.1186/gb-2005-6-12-r99
- Bellés, X., and Santos, C. G. (2014). The MEKRE93 (Methoprene tolerant-Krüppel homolog 1-E93) pathway in the regulation of insect metamorphosis, and the homology of the pupal stage. *Insect Biochem. Mol. Biol.* 52, 60–68. doi: 10.1016/j.ibmb.2014.06.009
- Boutz, P. L., Chawla, G., Stoilov, P., and Black, D. L. (2007). MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes Dev.* 21, 71–84. doi: 10.1101/gad.1500707
- Buffalo, V. (2011). *Scythe*. Available online at: <https://github.com/vsbuffalo/scythe>
- Buttstedt, A., Moritz, R. F. A., and Erler, S. (2013). More than royal food - Major royal jelly protein genes in sexuals and workers of the honeybee *Apis mellifera*. *Front. Zool.* 10:e72. doi: 10.1186/1742-9994-10-72
- Carstee, E. D., Morris, J. A., Coleman, K. G., Loftus, S. K., Zhang, D., Cummings, C., et al. (1997). Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. *Science* 277, 228–231. doi: 10.1126/science.277.5323.228
- Charles, J.-P., Iwema, T., Epa, V. C., Takaki, K., Rynes, J., and Jindra, M. (2011). Ligand-binding properties of a juvenile hormone receptor, Methoprene-tolerant. *Proc. Natl. Acad. Sci. U.S.A.* 108, 21128–21133. doi: 10.1073/pnas.1116123109
- Chawla, G., and Sokol, N. S. (2012). Hormonal activation of *let-7-C* microRNAs via EcR is required for adult *Drosophila melanogaster* morphology and function. *Development* 139, 1788–1797. doi: 10.1242/dev.077743
- Chen, J.-F., Mandel, E. M., Thomson, J. M., Wu, Q., Callis, T. E., Hammond, S. M., et al. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat. Genet.* 38, 228–233. doi: 10.1038/ng1725
- Chen, X., Yu, X., Cai, Y., Zheng, H., Yu, D., Liu, G., et al. (2010). Next-generation small RNA sequencing for microRNAs profiling in the honey bee *Apis mellifera*. *Insect Mol. Biol.* 19, 799–805. doi: 10.1111/j.1365-2583.2010.01039.x
- Claudianos, C., Ranson, H., Johnson, R. M., Biswas, S., Schuler, M. A., Berenbaum, M. R., et al. (2006). A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Mol. Biol.* 15, 615–636. doi: 10.1111/j.1365-2583.2006.00672.x
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Davis, M. B., and Li, T. (2013). Genomic analysis of the ecdysone steroid signal at metamorphosis onset using ecdysoneless and EcRnull *Drosophila melanogaster* mutants. *Genes Genomics* 35, 21–46. doi: 10.1007/s13258-013-0061-0
- Dorn, A. (2000). “Arthropoda—insecta: embryology,” in *Reproductive Biology of Invertebrates – Progress in Developmental Endocrinology*, Vol. 10, Part B, ed A. Dorn (Chichester: Wiley & Sons), 72–116.
- Drapeau, D. M., Albert, S., Kucharski, R., Prusko, C., and Maleszka, R. (2006). Evolution of the yellow/major royal jelly protein family and the emergence of social behavior in honey bees. *Genome Res.* 16, 1385–1394. doi: 10.1101/gr.5012006
- Elias-Neto, M., Soares, M. P. M., Simões, Z. L. P., Hartfelder, K., and Bitondi, M. M. G. (2010). Developmental characterization, function and regulation of a Laccase2 encoding gene in the honey bee, *Apis mellifera* (Hymenoptera, Apinae). *Insect Biochem. Mol. Biol.* 40, 241–251. doi: 10.1016/j.ibmb.2010.02.004
- Evans, R. M., and Mangelsdorf, D. J. (2014). Nuclear receptors, RXR, and the Big Bang. *Cell* 157, 255–266. doi: 10.1016/j.cell.2014.03.012
- Fährbach, S. E., Smagghe, G., and Velarde, R. A. (2012). Insect nuclear receptors. *Annu. Rev. Entomol.* 57, 83–106. doi: 10.1146/annurev-ento-120710-100607

- Feldlaufer, M. F., Herbert, E. W. J., and Svoboda, J. A. (1985). Makisterone A: the major ecdysteroid from the pupae of the honey bee, *Apis mellifera*. *Insect Biochem.* 15, 597–600. doi: 10.1016/0020-1790(85)90120-9
- Feldlaufer, M. F., Herbert, E. W. Jr., Svoboda, J. A., and Thompson, M. J. (1986b). Biosynthesis of makisterone A and 20-hydroxyecdysone from labeled sterols by the honey bee, *Apis mellifera*. *Arch. Insect Biochem. Physiol.* 3, 415–421. doi: 10.1002/arch.940030502
- Feldlaufer, M. F., Svoboda, J. A., and Herbert, E. W. Jr. (1986a). Makisterone A and 24-methylcholesterol from the ovaries of the honey bee, *Apis mellifera* L. *Experientia* 42, 200–201. doi: 10.1007/BF01952468
- Feldlaufer, M. F., Weirich, G. F., Lusby, W. R., and Svoboda, J. A. (1991). Makisterone C a 29-carbon ecdysteroid from developing embryos of the cotton stainer bug, *Dysdercus fasciatus*. *Arch. Insect Biochem. Physiol.* 18, 71–79. doi: 10.1002/arch.940180202
- Flatt, T., Heyland, A., Rus, F., Porpiglia, E., Sherlock, C., Yamamoto, R., et al. (2008). Hormonal regulation of the humoral innate immune response in *Drosophila melanogaster*. *J. Exp. Biol.* 211, 2712–2724. doi: 10.1242/jeb.014878
- Flatt, T., Tu, M. P., and Tatar, M. (2005). Hormonal pleiotropy and the juvenile hormone regulation of *Drosophila* development and life history. *Bioessays* 27, 999–1010. doi: 10.1002/bies.20290
- Forêt, S., and Maleszka, R. (2006). Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Res.* 16, 1404–1413. doi: 10.1101/gr.5075706
- Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52. doi: 10.1093/nar/gkr688
- Fullaondo, A., and Lee, S. Y. (2012). Identification of putative miRNA involved in *Drosophila melanogaster* immune response. *Dev. Comp. Immunol.* 36, 267–273. doi: 10.1016/j.dci.2011.03.034
- Gálíková, M., Klepsatel, P., Senti, G., and Flatt, T. (2011). Steroid hormone regulation of *C. elegans* and *Drosophila* aging and life history. *Exp. Gerontol.* 46, 141–147. doi: 10.1016/j.exger.2010.08.021
- Geva, S., Hartfelder, K., and Bloch, G. (2005). Reproductive division of labor, dominance, and ecdysteroid levels in hemolymph and ovary of the bumble bee *Bombus terrestris*. *J. Insect Physiol.* 51, 811–823. doi: 10.1016/j.jinsphys.2005.03.009
- Hartfelder, K., Bitondi, M. M. G., Santana, W. C., and Simões, Z. L. P. (2002). Ecdysteroid titers and reproduction in queens and workers of the honey bee and of a stingless bee: loss of ecdysteroid function at increasing levels of sociality? *Insect. Mol. Biol.* 32, 211–216. doi: 10.1016/S0965-1748(01)00100-X
- Hartfelder, K., and Emlen, D. (2012). “Endocrine control of insect polyphenism,” in *Insect Endocrinology*, ed L. I. Gilbert (San Diego, CA: Academic Press), 464–522.
- Hartfelder, K., and Engels, W. (1998). Social insect polymorphism: hormonal regulation of plasticity in development and reproduction in the honeybee. *Curr. Top. Dev. Biol.* 40, 45–77. doi: 10.1016/S0070-2153(08)60364-6
- Hill, R. J., Billas, I. M. L., Bonneton, F., Graham, L. D., and Lawrence, M. C. (2013). Ecdysone receptors: from the Ashburner model to structural biology. *Annu. Rev. Entomol.* 58, 251–271. doi: 10.1146/annurev-ento-120811-153610
- Huang, X., Suyama, K., Buchanan, J., Zhu, A. J., and Scott, M. P. (2007). A *Drosophila* model of the Niemann-Pick type C lysosome storage disease: dnpc1a is required for molting and sterol homeostasis. *Development* 132, 5115–5124. doi: 10.1242/dev.02079
- Humann, F. C., Tiberio, G. J., and Hartfelder, K. (2013). Sequence and expression characteristics of long noncoding RNAs in honey bee caste development – potential novel regulators for transgressive ovary size. *PLoS ONE* 8:e78915. doi: 10.1371/journal.pone.0078915
- Iga, M., and Kataoka, H. (2012). Recent studies on insect hormone metabolic pathways mediated by Cytochrome P450 enzymes. *Biol. Pharm. Bull.* 35, 838–843. doi: 10.1248/bpb.35.838
- Iida, K., Cox-Foster, D. L., Yang, X., Ko, W.-Y., and Cavener, D. R. (2007). Expansion and evolution of insect GMC oxidoreductases. *BMC Evol. Biol.* 7:75. doi: 10.1186/1471-2148-7-75
- Iovino, N., Pane, A., and Gaul, U. (2009). miR-184 has multiple roles in *Drosophila* female germline development. *Dev. Cell* 17, 123–133. doi: 10.1016/j.devcel.2009.06.008
- Jagadeeswaran, G., Zheng, Y., Sumathipala, N., Jiang, H., Arrese, E. L., Soulages, J. L., et al. (2010). Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development. *BMC Genomics* 11:52. doi: 10.1186/1471-2164-11-52
- Jindra, M., Malone, F., Hiruma, K., and Riddiford, L. M. (1996). Developmental profiles and ecdysteroid regulation of the mRNAs for two ecdysone receptor isoforms in the epidermis and wings of the tobacco hornworm, *Manduca sexta*. *Dev. Biol.* 180, 258–272. doi: 10.1006/dbio.1996.0299
- Kamakura, M. (2011). Royalactin induces queen differentiation in honeybees. *Nature* 473, 478–483. doi: 10.1038/nature10093
- Karim, F. D., and Thummel, C. S. (1992). Temporal coordination of regulatory gene expression by the steroid hormone ecdysone. *EMBO J.* 11, 4083–4093.
- Kato, S., Endoh, H., Masuhiro, Y., Kitamoto, T., and Uchiyama, S., Sasaki H., et al. (1995). Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science* 270, 1491–1494. doi: 10.1126/science.270.5241.1491
- Kelstrup, H., Hartfelder, K., Nascimento, F. S., and Riddiford, L. M. (2014). Reproductive status, endocrine physiology and chemical signaling in the Neotropical, swarm-founding eusocial wasp *Polybia micans* Ducke (Vespidae: Epiponini). *J. Exp. Biol.* 217, 2399–2410. doi: 10.1242/jeb.096750
- King-Jones, K., and Thummel, C. S. (2005). Nuclear receptors, a perspective e from *Drosophila*. *Nat. Rev. Genet.* 6, 311–323. doi: 10.1038/nrg1581
- Kirchhof, B. S., Homberg, U., and Mercer, A. R. (1999). Development of dopamine-immunoreactive neurons associated with the antennal lobes of the honey bee, *Apis mellifera*. *J. Comp. Neurol.* 411, 643–653.
- Kucherenko, M. M., and Shcherbata, H. R. (2013). Steroids as external temporal codes act via microRNAs and cooperate with cytokines in differential neurogenesis. *Fly* 7, 173–183. doi: 10.4161/fly.25241
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Li, P., Peng, J., Hu, J., Xu, Z., Xie, W., and Yuan, L. (2011). Localized expression pattern of miR-184 in *Drosophila*. *Mol. Biol. Rep.* 38, 355–358. doi: 10.1007/s11033-010-0115-1
- Lozano, J., and Belles, X. (2011). Conserved repressive function of Krüppel homolog 1 on insect metamorphosis in hemimetabolous and holometabolous species. *Sci. Rep.* 1:e163. doi: 10.1038/srep00163
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- Michelette, E. R. F., and Soares, A. E. E. (1993). Characterization of preimaginal developmental stages of Africanized honey bee workers (*Apis mellifera* L.). *Apidologie* 24, 431–440. doi: 10.1051/apido:19930410
- Mouillet, J. F., Delbecq, J. P., Quennedey, B., and Delachambre, J. (1997). Cloning of two putative ecdysteroid receptor isoforms from *Tenebrio molitor* and their developmental expression in the epidermis during metamorphosis. *Eur. J. Biochem.* 248, 856–863. doi: 10.1111/j.1432-1033.1997.00856.x
- Mulder, N., and Apweiler, R. (2007). InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol. Biol.* 396, 59–70. doi: 10.1007/978-1-59745-515-2_5
- Nunes, F. M. F., Aleixo, A., Barchuk, A. R., Bomtorin, A. D., Grozinger, C., and Simões, Z. L. P. (2013a). Non-target effects of Green Fluorescent Protein (GFP)-derived double-stranded RNA (dsRNA-GFP) used in honey bee RNA interference (RNAi) assays. *Insects* 4, 90–103. doi: 10.3390/insects4010090
- Nunes, F. M. F., Ihle, K. E., Mutti, N. S., Simões, Z. L. P., and Amdam, G. V. (2013b). The gene vitellogenin affects microRNA regulation in honey bee (*Apis mellifera*) fat body and brain. *J. Exp. Biol.* 216, 3724–3732. doi: 10.1242/jeb.089243
- O'Reilly, D. R., and Miller, L. R. (1989). A baculovirus blocks insect molting by producing ecdysteroid UDP-glucosyl transferase. *Science* 245, 1110–1112.
- Ono, H. (2014). Ecdysone differentially regulates metamorphic timing relative to 20-hydroxyecdysone by antagonizing juvenile hormone in *Drosophila melanogaster*. *Dev. Biol.* 391, 32–42. doi: 10.1016/j.ydbio.2014.04.004
- Petryk, A., Warren, J. T., Marques, G., Jarcho, M. P., Gilbert, L. I., Kahler, J., et al. (2003). Shade is the p450 enzyme that mediates the hydroxylation of ecdysone to the steroid insect molting hormone 20-hydroxyecdysone. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13773–13778. doi: 10.1073/pnas.2336088100
- Pinto, L. Z., Hartfelder, K., Bitondi, M. M. G., and Simões, Z. L. P. (2002). Ecdysteroid titers in pupae of highly social bees relate to distinct modes of caste development. *J. Insect Physiol.* 48, 783–790. doi: 10.1016/S0022-1910(02)00103-8
- Piulachs, M. D., Guidugli, K. R., Barchuk, A. R., Cruz, J., Simões, Z. L. P., and Bellés, X. (2003). The vitellogenin of the honey bee, *Apis mellifera*: structural analysis of the cDNA and expression studies. *Insect Biochem. Mol. Biol.* 33, 459–465. doi: 10.1016/S0965-1748(03)00021-3

- Rachinsky, A., Strambi, C., Strambi, A., and Hartfelder, K. (1990). Caste and metamorphosis - hemolymph titers of juvenile hormone and ecdysteroids in last instar honeybee larvae. *Gen. Comp. Endocrinol.* 79, 31–38. doi: 10.1016/0016-6480(90)90085-Z
- Regan, J. C., Brandão, A. S., Leitão, A. B., Mantas, D. A. R., Sucena, E., Jacinto, A., et al. (2013). Steroid hormone signaling is essential to regulate innate immune cells and fight bacterial infection in *Drosophila*. *PLoS Pathog.* 9:e1003720. doi: 10.1371/journal.ppat.1003720
- Rembold, H. (1987). Caste-specific modulation of juvenile hormone titers in *Apis mellifera*. *Insect Biochem.* 17, 1003–1006. doi: 10.1016/0020-1790(87)90110-7
- Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L., and Pachter, L. (2011). Improving RNA-seq expression estimates by correcting for fragment bias. *Genome Biol.* 12:R22. doi: 10.1186/gb-2011-12-3-r22
- Rubio, M., and Bellés, X. (2013). Subtle roles of microRNAs let-7, miR-100 and miR-125 on wing morphogenesis in hemimetabolous metamorphosis. *J. Insect Physiol.* 59, 1089–1094. doi: 10.1016/j.jinsphys.2013.09.003
- Sawata, M., Yoshino, D., Takeuchi, H., Kamikouchi, A., Ohashi, K., and Kubo, T. (2002). Identification and punctate nuclear localization of a novel noncoding RNA, Ks-1, from the honeybee brain. *RNA* 8, 772–785. doi: 10.1017/S1355838202028790
- Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026
- Schwander, T., Humbert, J.-Y., Brent, C. S., Cahan, S. H., Chapuis, L., Renai, E., et al. (2008). Maternal effect on female caste determination in a social insect. *Curr. Biol.* 18, 265–269. doi: 10.1016/j.cub.2008.01.024
- Soares, M. P. M., Barchuk, A. R., Quirino Simões, A. C., Cristino, A. S., Freitas, F. C. P., Canhos, L. L., et al. (2013). Genes involved in thoracic exoskeleton formation during the pupal-to-adult molt in a social insect model, *Apis mellifera*. *BMC Genomics* 14:576. doi: 10.1186/1471-2164-14-576
- Soares, M. P. M., Elias-Neto, M., Simões, Z. L. P., and Bitondi, M. M. G. (2007). A cuticle protein gene in the honeybee: Expression during development and in relation to the ecdysteroid titer. *Insect Biochem. Mol. Biol.* 37, 1272–1282. doi: 10.1016/j.ibmb.2007.07.014
- Soares, M. P. M., Silva-Torres, F. A., Elias-Neto, M., Nunes, F. M. F., Simões, Z. L. P., and Bitondi, M. M. G. (2011). Ecdysteroid-dependent expression of the *Tweddle* and *Peroxidase* genes during adult cuticle formation in the honey bee, *Apis mellifera*. *PLoS ONE* 6:e20513. doi: 10.1371/journal.pone.0020513
- Sokol, N. S., and Ambros, V. (2005). Mesodermally expressed *Drosophila* microRNA-1 is regulated by Twist and is required in muscles during larval growth. *Genes Dev.* 19, 2343–2354. doi: 10.1101/gad.1356105
- Sokol, N. S., Xu, P., Jan, Y. N., and Ambros, V. (2008). *Drosophila* let-7 microRNA is required for remodeling of the neuromusculature during metamorphosis. *Genes Dev.* 22, 1591–1596. doi: 10.1101/gad.1671708
- Sultan, A.-R. S., Oishi, Y., and Ueda, H. (2014). Function of the nuclear receptor FTZ-F1 during the pupal stage in *Drosophila melanogaster*. *Dev. Growth Diff.* 56, 245–253. doi: 10.1111/dgd.12125
- Takeuchi, H., Paul, R. K., Matsuzaka, E., and Kubo, T. (2007). EcR-A expression in the brain and ovary of the honeybee (*Apis mellifera* L.). *Zool. Sci.* 24, 596–603. doi: 10.2108/zsj.24.596
- Talbot, W. S., Swyryd, E. A., and Hogness, D. S. (1993). *Drosophila* tissues with different metamorphic responses to ecdysone express different ecdysone receptor isoforms. *Cell* 73, 1323–1337. doi: 10.1016/0092-8674(93)90359-X
- Tamura, S., Amano, S., Kono, T., Kondoh, J., Yamaguchi, K., Kobayashi, S., et al. (2009). Molecular characteristics and physiological functions of major royal jelly protein 1 oligomer. *Proteomics* 9, 5534–5543. doi: 10.1002/pmic.200900541
- The Honey Bee Genome Sequencing Consortium (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443, 931–949. doi: 10.1038/nature05260
- Tora, L., Gronemeyer, H., Turcotte, B., Gaub, M. P., and Chambon, P. (1988). The N-terminal region of the chicken progesterone receptor specifies target gene activation. *Nature* 333, 185–188. doi: 10.1038/333185a0
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Varghese, J., and Cohen, S. M. (2007). microRNA miR-14 acts to modulate a positive autoregulatory loop controlling steroid hormone signaling in *Drosophila*. *Genes Dev.* 21, 2277–2282. doi: 10.1101/gad.439807
- Velarde, R. A., Robinson, G. E., and Fahrback, S. E. (2006). Nuclear receptors of the honey bee: annotation and expression in the adult brain. *Insect Mol. Biol.* 15, 583–595. doi: 10.1111/j.1365-2583.2006.00679.x
- Velarde, R. A., Robinson, G. E., and Fahrback, S. E. (2009). Coordinated responses to developmental hormones in the Kenyon cells of the adult worker honey bee brain (*Apis mellifera* L.). *J. Insect Physiol.* 55, 59–69. doi: 10.1016/j.jinsphys.2008.10.006
- Watanabe, T., Takeuchi, H., and Kubo, T. (2010). Structural diversity and evolution of the N-terminal isoform-specific region of ecdysone receptor-A and -B1 isoforms in insects *BMC Evol. Biol.* 10:40. doi: 10.1186/1471-2148-10-40
- Wienholds, E., and Plasterk, R. H. (2005). MicroRNA function in animal development. *FEBS Lett.* 579, 5911–5922. doi: 10.1016/j.febslet.2005.07.070
- Yamanaka, N., Rewitz, K. F., and O'Connor, M. B. (2013). Ecdysone control of developmental transitions: lessons from *Drosophila* research. *Annu. Rev. Entomol.* 58, 497–516. doi: 10.1146/annurev-ento-120811-153608
- Yang, M., Wei, Y., Jiang, F., Wang, Y., Guo, X., He, J., et al. (2014). MicroRNA-133 inhibits behavioral aggregation by controlling dopamine synthesis in locusts. *PLoS Genet.* 10:e1004206. doi: 10.1371/journal.pgen.1004206
- Yao, T. P., Sagraves, W. A., McKeown, M., and Evans, R. M. (1992). *Drosophila ultraspiracle* modulates ecdysone receptor function via heterodimer formation. *Cell* 71, 63–72. doi: 10.1016/0092-8674(92)90266-F
- Zitnan, D., and Adams, M. E. (2012). “Neuroendocrine regulation of ecdysis,” in *Insect Endocrinology*, ed L. I. Gilbert (San Diego, CA: Academic Press), 253–309.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 September 2014; accepted: 04 December 2014; published online: 22 December 2014.

Citation: Mello TRP, Aleixo AC, Pinheiro DG, Nunes FMF, Bitondi MMG, Hartfelder K, Barchuk AR and Simões ZLP (2014) Developmental regulation of ecdysone receptor (EcR) and EcR-controlled gene expression during pharate-adult development of honeybees (*Apis mellifera*). *Front. Genet.* 5:445. doi: 10.3389/fgene.2014.00445

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2014 Mello, Aleixo, Pinheiro, Nunes, Bitondi, Hartfelder, Barchuk and Simões. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Biased Allele Expression and Aggression in Hybrid Honeybees may be Influenced by Inappropriate Nuclear-Cytoplasmic Signaling

Joshua D. Gibson^{1*}, Miguel E. Arechavaleta-Velasco², Jennifer M. Tsuruda³ and Greg J. Hunt¹

¹ Department of Entomology, Purdue University, West Lafayette, IN, USA, ² CENID-Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, México, Mexico, ³ Public Service and Agriculture, Clemson University, Clemson, SC, USA

OPEN ACCESS

Edited by:

Melanie April Murphy,
University of Wyoming, USA

Reviewed by:

Guy Bloch,
The Hebrew University of Jerusalem,
Israel

Bart Pannebakker,
Wageningen University, Netherlands

Douglas Mark Ruden,
Wayne State University, USA
Eamonn Mallon,
University of Leicester, UK

*Correspondence:

Joshua D. Gibson
gibson85@purdue.edu

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 18 August 2015

Accepted: 20 November 2015

Published: 01 December 2015

Citation:

Gibson JD, Arechavaleta-Velasco ME,
Tsuruda JM and Hunt GJ (2015)
Biased Allele Expression
and Aggression in Hybrid Honeybees
may be Influenced by Inappropriate
Nuclear-Cytoplasmic Signaling.
Front. Genet. 6:343.
doi: 10.3389/fgene.2015.00343

Hybrid effects are often exhibited asymmetrically between reciprocal families. One way this could happen is if silencing of one parent's allele occurs in one lineage but not the other, which could affect the phenotypes of the hybrids asymmetrically by silencing that allele in only one of the hybrid families. We have previously tested for allele-specific expression biases in hybrids of European and Africanized honeybees and we found that there was an asymmetric overabundance of genes showing a maternal bias in the family with a European mother. Here, we further analyze allelic bias in these hybrids to ascertain whether they may underlie previously described asymmetries in metabolism and aggression in similar hybrid families and we speculate on what mechanisms may produce this biased allele usage. We find that there are over 500 genes that have some form of biased allele usage and over 200 of these are biased toward the maternal allele but only in the family with European maternity, mirroring the pattern observed for aggression and metabolic rate. This asymmetrically biased set is enriched for genes in loci associated with aggressive behavior and also for mitochondrial-localizing proteins. It contains many genes that play important roles in metabolic regulation. Moreover we find genes relating to the piwi-interacting RNA (piRNA) pathway, which is involved in chromatin modifications and epigenetic regulation and may help explain the mechanism underlying this asymmetric allele use. Based on these findings and previous work investigating aggression and metabolism in bees, we propose a novel hypothesis; that the asymmetric pattern of biased allele usage in these hybrids is a result of inappropriate use of piRNA-mediated nuclear-cytoplasmic signaling that is normally used to modulate aggression in honeybees. This is the first report of widespread asymmetric effects on allelic expression in hybrids and may represent a novel mechanism for gene regulation.

Keywords: parental effects, cytoplasmic incompatibility, hybrid incompatibility, aggression, Africanized, *Apis mellifera*, PIWI, PIWI-interacting small RNAs

INTRODUCTION

The honeybee (*Apis mellifera*) is becoming a promising model for understanding epigenetic processes, which affect gene expression without modifying the DNA sequence. Honeybees possess all of the key genes in the methylation machinery (Wang et al., 2006) and DNA methylation plays a role in queen caste determination (Kucharski et al., 2008). Additionally, honeybee post-translational histone protein modifications have been characterized and may play a role in caste determination (Spannhoff et al., 2011; Dickman et al., 2013). However, we are just beginning to learn how these epigenetic processes regulate gene expression in honeybees. For example, methylation events in the bee genome have been shown to be plastic and related to behavioral castes (Herb et al., 2012), though they appear to be associated with alternative splicing rather than large alterations in transcriptional abundance (Flores et al., 2012; Foret et al., 2012).

One major hypothesis for epigenetic control of gene expression in honeybees, the kinship theory of genomic imprinting, predicts a bias in the expression of alleles in a parent-specific manner due to differences in relatedness between nestmates in order to enhance inclusive fitness (Haig, 2002; Queller, 2003). Another hypothesis predicts a bias in expression toward the maternal allele in order to maintain a match between co-adapted nuclear alleles and the maternally inherited mitochondria (Wolf, 2009). There is considerable empirical support for the former theory while there is little evidence in support of the latter (Haig, 2004). Only the kinship theory predicts a paternal expression bias, but both theories predict a maternal bias for some genes and in these cases support for one theory over the other can only be distinguished by the functions of the biased genes. In either case, this differential allelic expression must include an epigenetic mechanism because the expression bias of an allele is affected by the parent from which it is inherited and not solely by its genotype (i.e., European or Africanized alleles). Until recently, it was not known if there were gene expression effects in honeybees that were consistent with epigenetic regulation. However, there are phenotypic parental effects on aggression and metabolic rate that have been identified in studies utilizing hybrids from crosses between European and African subspecies that occur in patterns that suggest that epigenetic processes may be involved (Harrison and Hall, 1993; Guzman-Novoa et al., 2005; Oldroyd et al., 2014).

We previously tested for epigenetic effects by identifying parent-specific gene expression (PSGE) in honeybees by utilizing reciprocal F1 worker families derived from crosses of European (*A.m. carnica*) and Africanized bees (invasive hybrids between African *A.m. scutellata* and European honeybees; Kocher et al., 2015). Transcriptomes of workers in the two reciprocal families (differing in the lineage from which each parent is derived) were sequenced and read counts of heterozygous SNPs were used to test for PSGE. This experimental design allows us to assess epigenetic effects on transcription because the genotype (European or Africanized) and parent-of-origin of an allele will differ between the two families. Each allele will be inherited maternally in one family and paternally in the other. We found that PSGE is present in the honeybee (1–2% of tested loci) and

that a bias toward maternal expression was common. A set of 46 genes showed consistent, symmetric parental biases in both families (maternal or paternal in both families) and several of these genes have functions that are predicted by the kinship theory of genomic imprinting. Surprisingly, a strong maternal bias occurred primarily in the family with European maternity (EA hybrids hereafter) and 215 genes were maternally biased exclusively in this family compared to only 24 genes that were maternally biased exclusively in the Africanized maternity family (AE hybrids hereafter). This was the first evidence of PSGE in honeybees and, while PSGE has been studied in many organisms, to the best of our knowledge the observed asymmetric pattern of PSGE (PSGE in only one family) has not been documented in other species.

Asymmetries in phenotypic effects between reciprocal hybrid families are commonly observed, including honeybee hybrids (Harrison and Hall, 1993; Guzman-Novoa et al., 2005; Turelli and Moyle, 2007). These asymmetric phenotypic effects require that there be some asymmetry in the expression of genes inherited from the parents, which could include sex chromosomes, cytoplasmic factors, differentially imprinted genes, or maternal effects (Wolf et al., 2014). We propose that these asymmetric phenotypic effects in honeybees are due to asymmetric PSGE that is the result of inappropriate signaling in the hybrids. Specifically, we propose that these wide crosses disrupt, in the hybrids, nuclear-cytoplasmic signaling pathways, and epigenetic processes that are utilized differentially in the parental lineages. This disruption leads to inappropriate signaling in these pathways that influences epigenetic chromosomal modifications (in an allele-specific manner), ultimately resulting in asymmetric phenotypic effects. In support of this, some of the maternally biased genes in our previous study were located within quantitative trait loci (QTL) that influence honeybee stinging behavior, a trait which is asymmetrically expressed in these hybrids (Hunt et al., 1998, 2007; Guzman-Novoa et al., 2005). In addition, the maternally biased genes for one of the three tissue samples analyzed (first instar larvae) were enriched for nuclear-encoded proteins known to localize to the mitochondria, supporting a connection between the cytoplasm, asymmetric PSGE, and asymmetric hybrid effects. Here we undertake a more comprehensive analysis of the transcriptome data to identify additional genes that show bias in these hybrids, to characterize their function and chromosomal localization with respect to QTL, and to test for differential gene expression between the two families.

MATERIALS AND METHODS

Previous and New Analyses

The work presented here utilizes a dataset originally published in Kocher et al. (2015). The sequencing data are available in the NCBI Short Read Archive, project number PRJNA277772. This dataset consisted of the cross utilizing Africanized honeybees (AHB) and the European honeybee (EHB) subspecies *A.m. carnica* (described in “crosses” below) and the analyses leading to the expression levels of the alleles in the reciprocal hybrid families (described in “PSGE of alleles” below). We produced a

new gene set by altering the criteria for a gene to be considered to show PSGE and differences in these criteria are described below. New analyses were performed using this gene set, constituting the results presented here. The new work also includes new crosses utilizing the European subspecies *A.m. ligustica* and *A. m. carnica* that provided workers used in the stinging behavior assays.

Crosses

Crosses for evaluating PSGE were previously described in Kocher et al. (2015) and biased expression patterns were further analyzed from this dataset. Briefly, EHB and AHB colonies were maintained at INIFAP facilities near Villa Guerrero, Estado de México, Mexico. There are several closely related subspecies of EHB that are often used by beekeepers and two of these were among the colonies tested for use in our crosses, Carniolan bees, *A.m. carnica*, and Italian bees, *A.m. ligustica*. The two most aggressive AHB colonies and the most docile EHB colonies (one each of *A.m. carnica* and *A.m. ligustica*) were chosen based on differences in stinging behavior and response to queen mandibular pheromone. Daughter queens and drones were raised from the parental colonies. Pairs of reciprocal crosses were performed using single-drone queen instrumental insemination between one AHB and the *A.m. ligustica* parental colony and between the AHB and the *A.m. carnica* parental colony. For crosses with *A. m. carnica*, two EA families and four AE families were tested for individual stinging behavior (see below). For crosses with *A. m. ligustica*, three EA and three AE families were tested.

PSGE of Alleles

Expression levels of the alleles in the two reciprocal crosses are from Kocher et al. (2015). Briefly, all transcriptome data is derived from the two reciprocal crosses utilizing the *A.m. carnica* queen (designated EA and AE for those with EHB and AHB maternity, respectively). Transcriptomes were sequenced from cDNA libraries of pooled first instar larvae (two libraries per family), pooled adults (guard bees, two libraries per family), and individual adult brains (three libraries per family). Single nucleotide polymorphisms (SNPs) differentiating European and Africanized alleles were identified by sequencing genomic DNA of the queen and drone parents of these two crosses to ensure the European and Africanized alleles were homozygous and different, resulting in F1 offspring that are heterozygous at all tested SNPs. All reads were mapped to the honeybee reference genome (Amel4.0; The Honeybee Genome Sequencing Consortium, 2006). Using counts of reads at each SNP, a general linear interactive mixed model (SAS, Cary, NC, USA) was utilized to assess expression of each allele for all transcripts containing diagnostic SNPs. The analysis in Kocher et al. (2015) required the bias to be in the same direction (maternal or paternal) in both directions of the cross (EA and AE hybrids) based on significant parent FDR < 0.05, and a bias of at least 0.6 (maternal or paternal reads/total reads; Wang and Clark, 2014) in order to search for consistent parent-of-origin effects. For the current analysis, we relaxed this criteria and only required that the bias be present in one direction of the cross. These genes were then placed into bias categories based on the expression

levels of their alleles in each family relative to the parent-of-origin of that allele (e.g., Maternal bias, EA maternal/AE maternal; European bias, EA maternal/AE paternal, etc.; see Figure 1).

Individual Stinging Behavior

We tested the stinging behavior of individual bees from our crosses by measuring the time that each bee took to sting a black suede patch after being stimulated with electrical current (assay described in Shorter et al., 2012). In total 573 bees were tested from the four F1 reciprocal colonies. These colonies are designated AL (AHB queen × *A.m. ligustica* drone), LA (*A.m. ligustica* queen × AHB drone), AC (AHB queen × *A.m. carnica* drone), and CA (*A.m. carnica* queen × AHB drone) in Figure 2. Data was transformed using the natural log function to fit a normal distribution and was analyzed under a one way analysis of variance to test for differences in the stinging behavior of the four F1 reciprocal crosses. Least squares means *t*-tests were performed to compare the means of the four groups.

Differential Expression Analysis

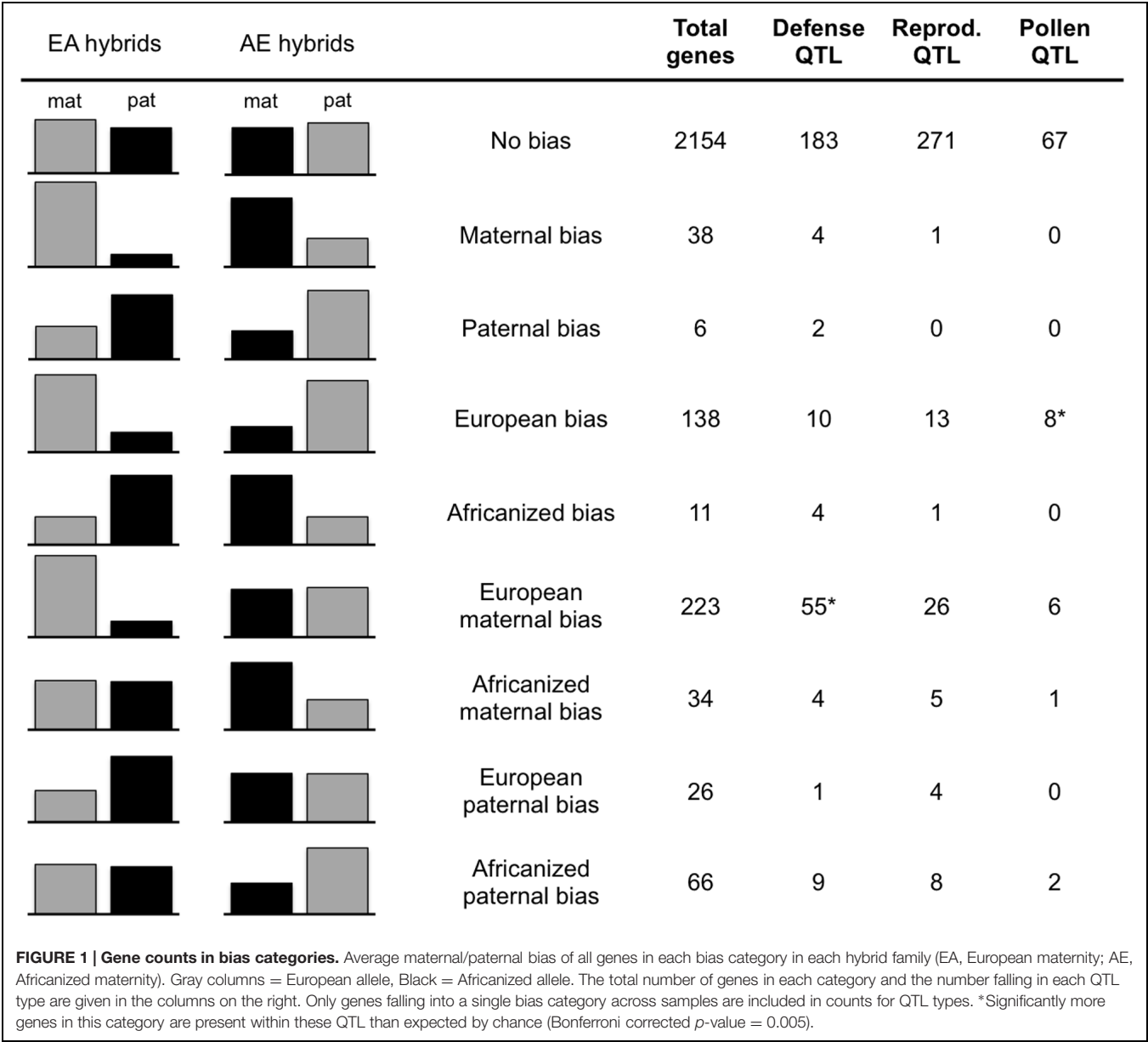
We assessed differential gene expression (DGE) between comparable stages/tissues in the EA and AE families using CLC Genomics Workbench version 7.5 (CLC Bio, Boston MA, USA) employing the Empirical Analysis of DGE option, which implements the “Exact Test” of Robinson et al. (2010). Genes were considered significantly differentially expressed if the False Discovery Rate corrected *p*-value was less than 0.05.

Overlap of Biased Genes with Known QTL

We assessed whether genes showing asymmetric maternal PSGE lie within QTL influencing traits related to colony defense (Hunt et al., 1998, 1999, 2007; Arechavaleta-Velasco et al., 2003; Shorter et al., 2012), reproduction (Oxley et al., 2008; Linksvayer et al., 2009; Rueppell et al., 2011), and pollen foraging behavior (Hunt et al., 1995, 2007; Page et al., 2000). We used the diagnostic SNPs within biased genes to determine their location in the Amel4.5 assembly. In cases where physical locations of markers were given in the QTL studies, we used these to identify the bounds of the QTL (typically the 1.5 LOD support interval). When information on physical locations of markers wasn't available, we used the sets of candidate genes from these projects to identify the range of the QTL. We then compared the positions of the biased genes with the ranges of these QTL to determine overlap.

Genomic Clustering of Biased Genes

In addition to identifying biased genes that are within previously identified QTL, we also assessed whether any biased genes were in physical clusters within the genome by visualizing their positions on SNP-based linkage maps (Arechavaleta-Velasco et al., 2012; Tsuruda et al., 2012). Once putative clusters of significantly biased genes were identified, we also looked at the allelic expression patterns of all tested genes (regardless of significance) within the putative cluster.



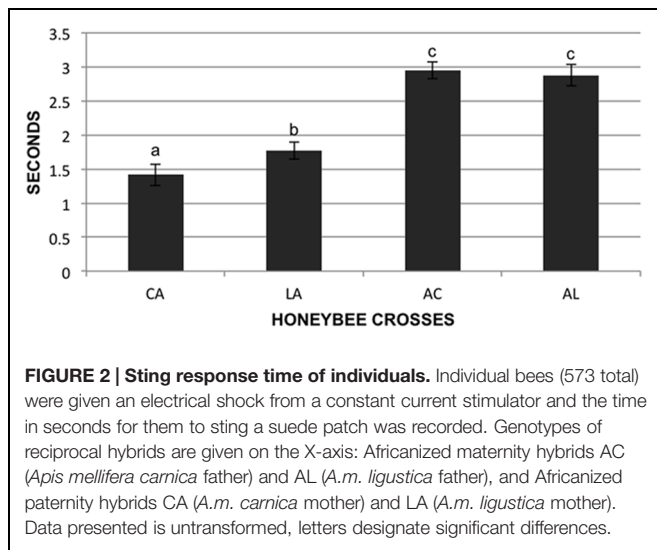
Statistical Tests of Enrichment/Overlap

We utilized goodness of fit tests to determine whether the genes in our bias categories (see **Figure 1**) were enriched for mitochondrial-localizing genes, significantly overlapped other gene sets, or were overrepresented in QTL. In all cases we only report results if they were significant after Bonferroni correction. We tested for enrichment of mitochondrial-localizing proteins by performing reciprocal BLASTs of the AmelOGS3.2 peptide sequences against a set of *Drosophila melanogaster* genes with proteins that are known to localize to mitochondria (Pagliarini et al., 2008; Smith et al., 2012). Expected values for our biased gene categories were calculated using the proportion of genes in the total honeybee gene set that match the *Drosophila* mitochondrial-localizing set. We also tested for significant overlap of our genes in each bias category with our own

differentially expressed gene set, genes differentially expressed between aggressive and non-aggressive bees (Alaux et al., 2009), and for overrepresentation within QTL. We used the proportion of the total official gene set represented in each of these groups to calculate the expected number of genes in each of our bias categories. We also tested for Gene Ontology (GO) term enrichment using the best reciprocal matching *D. melanogaster* genes by utilizing the Gene Ontology Consortium’s enrichment analysis pipeline (geneontology.org).

Use of Animals in Research

This research did not require IRB approval because we only used invertebrates in this study, which are exempt from IRB approval. Despite not requiring approval, we made every effort to minimize any potential suffering of the bees used in this research.



RESULTS

Genes Showing PSGE Bias

We found that out of the 2663 unique transcripts that we could test, 509 exhibited biased expression with one of the parental alleles used more than the other (≥ 0.6 bias) in at least one reciprocal hybrid family. In addition to the previously reported genes that show a parent-of-origin effect (either maternal or paternal in both families), we found evidence for biased PSGE in all other potential categories of bias (maternal or paternal only, allele-specific, or no bias in either family; **Figure 1**). Over 40% (223 genes) of the biased genes had a maternal bias only in the hybrids with European maternity (EA hybrids; **Figure 1**). Out of these 509 transcripts, 33 fell into more than one category of bias due to differences between samples (Supplementary Tables S1 and S2). In the majority of these cases, the category shift was due to a small level of bias in the AE family (near the 60% cutoff value) relative to the greater bias in the EA family. This is evident because the number of transcripts falling into more than one category decreases by 85% (to 5 total) simply by increasing the cutoff value to 70% bias. The EA maternal bias is much more robust, as indicated by the decrease of only <2% in the EA maternal-only category with the same change in criteria (Supplementary Table S2 and Figure S1). Nevertheless, to ensure unambiguous results we removed these genes for tests of enrichment of mitochondrial-localizing genes and presence in QTLs.

Differential Expression Analysis

A total of 160 unique genes were differentially expressed between the EA and AE families in at least one stage and six of these genes were differentially expressed in both guards and another stage while only one gene was shared between larvae and brains (Supplementary Table S3). One-hundred and one of these genes have a more than twofold change in expression.

Parental Effect on Stinging

There were significant differences in the honeybee individual stinging behavior between the four F1 reciprocal crosses ($F = 53.64$; $df = 3,567$; $p < 0.01$). The bees of the colonies with European maternity (CA and LA) stung significantly faster than the bees of the colonies with Africanized mothers (AC and AL; $p < 0.05$). There were differences between the two crosses with European maternity, the bees with *A.m. carnica* maternity (CA) stung faster than the bees of the cross with *A.m. ligustica* maternity (LA; $p < 0.05$), but there were no differences in the time to sting for the two crosses with Africanized maternity (AC and AL; **Figure 2**; $p > 0.05$).

Genomic Clustering of Biased Genes

We identified two regions containing clusters of genes that all appear to be biased, both of which overlap with defense-related QTL. One on chromosome 3 lies within the Sting-2 QTL, a region associated with increased colony-level stinging behavior (Hunt et al., 1998, 2007). There are 12 genes within a region of ~410 kb that show a significant maternal bias of greater than 90% in the European maternity family (**Figure 3A**). There is only one additional gene within this region that could be tested and this gene also shows >90% maternal bias in this family. The second cluster lies on chromosome 12 within the bounds of a QTL associated with production of the active component of alarm pheromone, isopentyl acetate (Hunt et al., 1999). This region is ~600 kb in length and there are 29 genes that could be tested within this region. Similar to the cluster within the Sting-2 QTL, 27 of these genes show a significant maternal bias (>90% maternal in 23 of these genes) in the European maternity family and the remaining two genes show the same pattern of extreme maternal bias (**Figure 3B**).

Overlap of Biased Genes with Other Data Sets

We found 164 of our biased genes overlapping with known QTL associated with traits for defense, reproduction and foraging behavior based on the position of these genes in OGS3.2 (Elsik et al., 2014). Within these QTL, just two of the gene bias categories were over-represented relative to the expected number based on OGS3.2 (**Figure 1**). The European maternal biased (only maternal bias in EA) gene set is overrepresented in defensive QTL with 55 genes (expect 36.3, $p < 0.002$), and the EA maternal AE paternal (European biased) gene set is overrepresented in pollen hoarding QTL with eight genes (expect 2.9, $p < 0.0031$).

We tested whether any of our biased categories are enriched for genes whose proteins are known to localize to mitochondria and found that the genes that are maternally biased only in the EA family are significantly enriched in each of the three samples (larvae [10/78], $p = 0.0005$; brains [16/140], $p = 0.0001$, adults [7/49] $p = 0.0016$). Moreover, 15 of the 17 genes that have the same bias in all three sample types are maternally biased only in the EA family and are highly enriched for mitochondrial-localizing genes (6/15, $p = 9.4 \times 10^{-11}$). Despite genomic clustering of some of our biased genes, there is no clustering of mitochondrial-localizing genes. We also

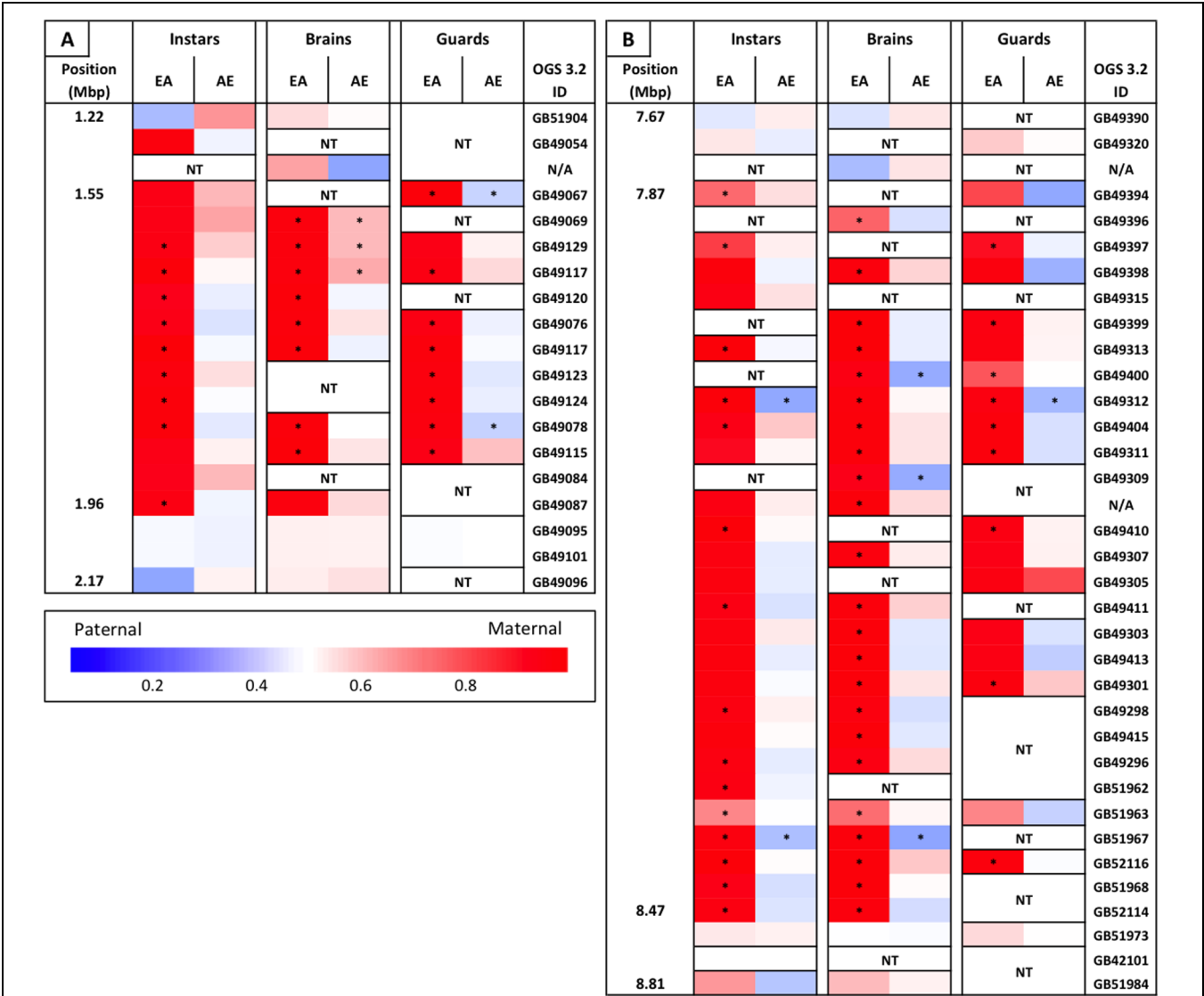


FIGURE 3 | Heat map of maternally biased gene clusters. (A) Gene cluster on chromosome 3, overlapping a QTL associated with stinging behavior. **(B)** Gene cluster on chromosome 12, overlapping a QTL associated with alarm pheromone production (isopentyl acetate). Position along the chromosome (in Mbp), relative allele usage of each reciprocal family within each sample, and OGS 3.2 gene ID is given for every tested gene within these clusters. Relative allele usage calculated as maternal read count/total read count. *Statistically significant allelic bias. NT, not tested. N/A in gene ID column are transcripts that had no clear match to a protein coding gene.

determined the extent to which our gene list overlapped with genes that were differentially regulated in aggressive vs. non-aggressive bees (Alaux et al., 2009). Significantly more genes overlapped between this study and our biased gene list than expected by chance (115 genes, $p = 0.017$), though this is not significant if we correct for testing each individual bias category. None of the individual categories are significantly enriched for overlapping genes, even without multiple test correction (Supplementary Table S1). Similarly, 43 of our differentially expressed genes overlap with those of Alaux et al. (2009), though there is no pattern to the overlap in regards to the up or down regulation of the genes in each study (Supplementary Table S3).

DISCUSSION

We previously found evidence of PSGE in honeybees (Kocher et al., 2015). However, we also found over 200 genes that showed highly biased expression toward the maternal allele, but only in the family with European maternity (EA hybrids). This asymmetric bias in expression is not predicted by theories of genomic imprinting. Similarly, if the pattern we observed were due to allelic effects (Africanized or European alleles preferentially expressed), we would expect to see a maternal bias in one family and a paternal bias in the other, but we did not. There have been several cases reported in which asymmetric hybrid phenotypic effects have resulted from disrupted genomic

imprinting in hybrids. One example is found in two species of deer mice, *Peromyscus maniculatis* and *P. polionotus*, in which the hybrid family with a *P. polionotus* mother exhibits offspring overgrowth while the reciprocal family exhibits undersized offspring (Loschiavo et al., 2007; Duselis and Vrana, 2010; reviewed in Wolf et al., 2014). These asymmetric hybrid effects are expected because of the imprinting that is predicted to occur in *P. maniculatis*, but not in *P. polionotus*, based on the conflict hypothesis of imprinting (Moore and Haig, 1991). *P. maniculatis* is polygamous while *P. polionotus* is monogamous, and therefore there is selection for *P. maniculatis* fathers to increase their own offspring's growth at the expense of other male's offspring from the same mother. This inclusive fitness incentive doesn't exist in a monogamous system. In *P. maniculatis* there is also a selective advantage for the mother to counteract this with mechanisms that allow all of her offspring to receive equal nutrition. The hybrid families end up with asymmetric offspring growth because the genomic conflict (and hence the balanced offspring growth) is disrupted in these crosses. *P. polionotus* parents don't have this conflict and so don't counteract the growth effects of their *P. maniculatis* partners, resulting in small offspring with *P. maniculatis* mothers and large offspring with *P. maniculatis* fathers. We extend this idea to hybrids between races of honeybees by proposing that the disruption of these parental effects (whether imprinting or other heritable factors) results in inappropriate signaling within established nuclear-cytoplasmic signaling pathways that leads to allele-specific changes in expression in only one of the hybrid families. To the best of our knowledge, there are no other examples of PSGE biases that show a widespread asymmetric pattern such as we find in these reciprocal hybrids.

Our biased gene set is dominated by a single category of bias, genes that are maternally biased only in the European maternity family (EA hybrids; **Figure 1**). The second most abundant category is bias toward European alleles, which may be influenced by similar processes. Previous studies investigating several trait differences between AHB and EHB found phenotypic patterns in hybrid crosses that are similar to the asymmetry we see in allelic bias. EA bees repeatedly exhibit high Africanized-like aggression while AE hybrids exhibit levels of aggression intermediate to the parents (Guzman-Novoa et al., 2005). We tested whether our biased genes may play a role in this asymmetric aggression by assessing the positions of the biased genes in the Amel4.5 assembly to see how they may fit with previously identified QTL associated with aggressive behavior, as well as QTL associated with reproduction and foraging (Elsik et al., 2014). We found that 89 out of the total of 509 biased genes lie within QTL for defensive traits, 58 genes are within QTL for reproductive traits and 17 are within QTL for pollen foraging behavior (**Figure 1**). There were significantly more genes than expected by chance between the EA maternal-only bias category and defensive QTL ($p = 0.0029$) and between the overall European bias category (EA maternal and AE paternal) and pollen foraging QTL ($p = 3.39 \times 10^{-5}$). The connection between the European bias group and pollen foraging is interesting given that the propensity for pollen collection has been shown to vary between European and Africanized bees (Pesante et al., 1987; Page et al., 2000). However, unlike the

genes with an EA maternal-only bias, when we increase our bias cutoff criteria from 60 to 70%, the number of genes in the pollen foraging QTL category is reduced by more than 60% and the enrichment within these QTL disappears, indicating that these genes are not highly biased in either family (Supplementary Table S2). We also tested whether our biased gene set is enriched for genes that are differentially expressed between aggressive and non-aggressive bees (Alaux et al., 2009). While the entire set of biased genes shows a slight enrichment for these genes (115 biased genes overlapping with 2254 differentially expressed genes; $p = 0.017$), no individual category of bias is significantly enriched.

Another asymmetric phenotype is that EA hybrids have asymmetrically low flight metabolic capacity (based on whole body CO₂ measurements) relative to both the parents and AE hybrids, which could indicate that there is an incompatibility between maternally derived European mitochondria and paternally derived nuclear genes in EA hybrids (Harrison and Hall, 1993). A separate study of aggression in honeybees found that aggression and brain metabolic rates are related, as the brains of highly aggressive bees showed significantly reduced oxidative metabolism relative to non-aggressive bees (Alaux et al., 2009). Reducing the rate of oxidative phosphorylation both in bees and in *Drosophila*, even in the whole body, increased aggression and therefore it appears that brain metabolic rate plays a causal role in aggressive behavior in insects (Li-Byarlay et al., 2014). Given this connection, differential gene expression associated with aggression in the parents (Alaux et al., 2009) would lead us to expect to find that metabolic genes show differential expression between our reciprocal families. Despite this expectation, genes that were differentially expressed between these families are not enriched for any functional GO category or for mitochondrial-localizing proteins (Supplementary Table S3; See Supplementary File 1 for discussion of differentially expressed genes). The lack of enrichment for genes showing expression differences between aggressive and non-aggressive bees or for any functional GO categories makes interpretation of our differentially expressed gene set difficult. Interestingly, only 13 of the 509 biased genes are also differentially expressed between the two families. This is significantly more overlap than expected given the small number of differentially expressed genes ($p = 0.0008$, Supplementary Table S1), however, there was no pattern to the overlap between up/down regulation and bias category. The fact that the vast majority of the biased genes are not differentially expressed means that in general there is a combination of allele-specific silencing and dosage compensation and that this process is occurring for many genes in only one of the reciprocal families. These results are reminiscent of the increased expression (e.g., *Drosophila* males) or silencing (e.g., mammalian females) that occurs on sex chromosomes to maintain comparable expression in both sexes (Disteche, 2012).

If nuclear-mitochondrial interactions are involved in the asymmetric phenotype of EA aggressive behavior through changes in metabolism, then we expect that this phenotype would have an inherent physiological basis and not necessarily be influenced by social interactions. Therefore we used an aversive

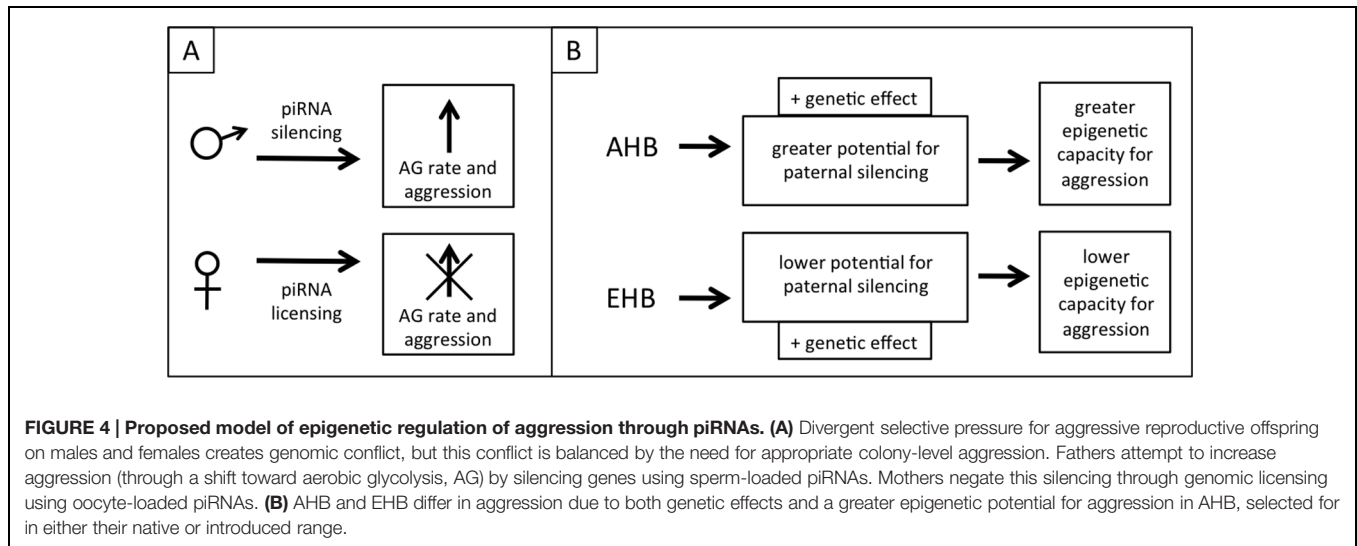
stimulus of individuals in a lab assay to test for an asymmetric phenotype outside the colony environment. We used hybrids with two European mitochondrial backgrounds (*A.m. ligustica* and *carnica*) for these tests. As seen at the colony level, bees with European mothers reacted more aggressively (faster to sting) than those with Africanized mothers but also that bees with *A.m. carnica* mothers were significantly more aggressive than those with *A.m. ligustica* mothers (**Figure 2**) even though the *A.m. carnica* parental source was less aggressive than the *A.m. ligustica* parent (data not shown). The stinging behavior QTL discussed above were also identified in a cross with *A.m. carnica* mothers (Hunt et al., 1998).

In addition to the significant overlap with QTL associated with aggressive behavior, another clue that the asymmetric PSGE in the EA family may be tied to both the asymmetric hybrid aggression and metabolic deficit is the fact that we found highly significant enrichment for mitochondrial-localizing proteins in every sample (Supplementary Table S1). This contrasts with results of our previous analyses that focused on 46 genes that showed consistent parental effects in both families. That study only found significant enrichment of mitochondrial proteins for biased genes in larvae of the EA family (Kocher et al., 2015). It is important to note that if the asymmetric bias in EA hybrids were due solely to incompatible interactions between nuclear genes and their proteins that directly interact with mitochondria, we would expect this enrichment to be very high (approaching 100%). However, enrichment only reaches ~8% in this biased gene set, which implies that this asymmetric bias may be due to dysfunctional signaling involving the mitochondrial and nuclear genomes rather than a direct result of nuclear-mitochondrial dysfunction. The reduced oxidative brain metabolism associated with aggression in bees isn't necessarily an overall reduction in energy metabolism, as studies show this reduction in oxidative metabolism is accompanied by a shift toward aerobic glycolysis (AG; glycolysis in the presence of oxygen; Chandrasekaran et al., 2015). The shift toward AG is mediated by mitochondrial retrograde signaling (signals from mitochondria that regulate nuclear transcription), a process that is normally used to maintain energy homeostasis (Liu and Butow, 2006). These connections may implicate retrograde signaling in the modulation of aggression in bees (Li-Byarlay et al., 2014).

In addition to this newfound connection with aggression in bees, the shift away from oxidative metabolism and toward AG is a well established phenomenon in cancer cells, known as the Warburg effect (Warburg, 1956). This metabolic shift is thought to play an important role in cell proliferation in both cancer cells and in normal, non-cancer cells (Lunt and Vander Heiden, 2011). AG seems to be especially important in brain tissue, as developing brain tissue shows this same metabolic transition and brain areas with increased synaptic activity exhibit major changes in lactate concentrations, a byproduct of AG (Barros, 2013; Gershon et al., 2013; Goyal et al., 2014). GO analysis of our EA maternal-only biased gene set revealed an enrichment for 80 GO terms but these fall into a few broad categories that include cellular morphogenesis (particularly neurogenesis), behavior, and regulation and cell signaling (Supplementary Table S4). These

categories are consistent with both the behavioral changes in the EA family as well as the cellular processes associated with a metabolic switch toward AG. Although genes that were maternally biased in EA were enriched for mitochondrial-localizing proteins, this set is not enriched for any GO categories directly involved in energy metabolism. However, the biased gene set does contain many genes that likely play a role in the retrograde response that elicits the switch toward AG. Moreover, this gene set has many genes that play a role in transcriptional regulation, including the piwi-interacting RNA (piRNA) pathway, which is a small RNA pathway involved in chromatin modifications (Huang et al., 2013). While we couldn't test for these small RNAs due to the size selection involved in our library preparation, these genes may provide a link between the metabolic shift to AG involving mitochondrial signaling and a potential mechanism for the asymmetric maternal-only bias that we see in gene expression (see Supplementary File 1 for a discussion of our interpretation of the connection these genes provide).

The piRNA pathway acts by modifying chromatin to inhibit transcription, as compared to posttranscriptional silencing initiated by other small RNA pathways (Huang et al., 2013). The piRNA pathway is primarily involved in the suppression of transposable elements, but recent studies have shown that piRNAs also play important roles in epigenetic modulation and genomic imprinting (Brennecke et al., 2008; Chang et al., 2009; Huang et al., 2013; Le Thomas et al., 2013). Chromatin modifications may help to explain our lack of differential expression of biased genes between the hybrid families: if the paternal allele is unable to be expressed due to these modifications, then any signal in the cell to increase expression (e.g., transcription factor binding) will only be able to act on the maternal allele, resulting in both allelic bias and a lack of differential expression. Chromatin modifications may also explain why two of the defensive QTL identified above contained large clusters of significantly biased genes in which every gene that could be tested showed the same pattern of >90% maternal bias only in the EA family (**Figure 3**). In both of these QTL we were only able to test a subset of all the genes in the region due to non-informative SNPs and/or insufficient read counts (Sting 2, 14/57; alarm pheromone, 27/59), however, the consistent level of bias in tested genes and their broad distribution across these clusters indicates that this bias likely occurs across all genes within these clusters. Within these clusters the same genes are biased across all sample types (though not always significant due to read counts), which indicates that this pattern is likely present throughout the lifespan of the individuals and across all tissues. It is possible that there is an inversion in these regions in the AHB lineage resulting in this pattern of expression bias, but we are unable to test for this due to low coverage of our genomic DNA from this lineage. This possibility seems unlikely, however, as we would expect to see a comparable reduction in Africanized (i.e., maternal) expression in the AE family in these regions and previous independent studies of recombination in EA hybrids haven't indicated the expected loss of recombination within these regions (Hunt et al., 1998; Shorter et al., 2012; Ross et al., 2015). Given the size of these clusters (~500 kb)



and the near complete silencing of the paternal alleles across tissues and life stages it seems likely that differential chromatin modifications in the homologous chromosomes contribute to this asymmetric pattern. While it is possible that these chromatin modifications occur independent of the piRNA pathway (e.g., chromosomal conformational changes that can be assessed using the Hi-C technique; Belton et al., 2012), these also seem less likely to be responsible for the overall asymmetric bias than the piRNA pathway due the genes involved in this bias (see Supplementary File 1) and the fact that >80% of these asymmetrically biased genes lie outside of these clusters (223 genes total with 39 in clusters). Given that the piRNA pathway acts in a sequence specific manner and may therefore be able to act on individual genes (Huang et al., 2013), we consider these other possibilities as alternative hypotheses to the model we propose below.

Piwi-interacting RNAs are loaded into oocytes, where they serve as a sequence-specific transgenerational epigenetic memory of both gene silencing and activation (maternal licensing), and in self/non-self recognition. piRNAs have also recently been found in mature sperm (Johnson and Spence, 2011; Shirayama et al., 2012; Pantano et al., 2015). The piRNA pathway has been shown to be involved in epigenetic regulation of phenotypic traits in both mice (white tail tip, WTT; Yuan et al., 2015) and fruit flies (ectopic long bristle outgrowths on the eyes, ELBOs; Sollars et al., 2003; Gangaraju et al., 2011). Both of these phenotypes occur at a low frequency in populations (naturally for WTT mice and artificially induced for ELBOs in *Drosophila*) and represent an epigenetic capacity within these populations that is normally suppressed through the piwi/piRNA pathway. The epigenetic capacity for these phenotypes can be released through selection for these traits (Sollars et al., 2003; Ruden et al., 2015; Yuan et al., 2015). These phenotypes are initially expressed in individuals with certain mutant alleles but the phenotype can occur in offspring that lack the causal mutant (Ruden et al., 2015; Yuan et al., 2015). Perhaps most intriguingly, the ELBOs in *Drosophila* can be maintained in the population (over 100 generations) as

long as selection for the trait is maintained (Ruden et al., 2008). This epigenetic selection may help to explain our asymmetric hybrid effects.

We speculate that the metabolic switch toward AG in honeybee brains and the associated aggression is a phenotypic trait that has a partially epigenetic basis, mediated through the piwi/piRNA pathway. An epigenetic switch to aggression is at least implied by theory regarding genomic imprinting in honeybees, as honeybees exhibit extreme polyandry and drones should have an evolutionary drive to produce daughters that are more selfish in regards to producing their own offspring, including more aggressive daughter queens that may be successful in queen duels, therefore inheriting the nest including the worker bees and other resources. In a population there should also be simultaneous selective pressure on the queens to suppress this selfish and aggressive behavior, resulting in intragenomic conflict similar to the *Peromyscus* mouse example given earlier (Queller, 2003). This level of intracolony aggression also needs to be balanced with the need for appropriate extra-colony aggression (i.e., colony defense). The genomic conflict could occur through paternal piRNA silencing to increase aggression and maternal piRNA licensing to mitigate the paternal silencing and reduce aggression. Similar to the WTT and ELBOs discussed above this is a phenotype that would normally be suppressed (or canalized) but which would vary in extent between populations due to differing selective pressures (e.g., within AHB and EHB lineages; Figure 4).

The implication of our admittedly speculative model is that wider crosses in honeybees can result in increased aggression in one of the hybrid families because the cross disrupts the balance of genomic conflict for/against aggression. This could occur in any case where the extent of genomic conflict differs between lineages and might explain why beekeepers who cross different races of bees sometimes report higher aggression in one of the reciprocal families, though individual crosses would need to be investigated to gain a full understanding of this phenomenon (Adam, 1983). The selection for highly aggressive AHB colonies

to use for our crosses might have resulted in the epigenetic release of this phenotype in that parent colony (i.e., simultaneously selected for increased paternal silencing of alleles that leads to aggression and less maternal opposition to this silencing), which explains the asymmetrically high aggression in the EA family (**Figure 4**). The importance of selecting for aggressive traits in the AHB parent may explain why another study that used EHB \times AHB crosses similar to ours, but that didn't select for differential aggression, did not have these same asymmetries in allele-specific expression (Galbraith and Grozinger, personal communication).

This piRNA mediated aggression model can be tested in multiple ways. Isolating and sequencing small RNAs from both sperm and eggs of the parent colonies would allow us to determine whether piRNAs are present and whether they target the genes that show biased allele usage. If they are present, the total RNA from sperm derived from drones of AHB colonies selected for high and low aggression could be injected into eggs from an EHB queen crossed with an AHB drone from a colony selected to be docile (the eggs must still be F1 hybrids to ensure both alleles are present for sequence specificity). Allele use can then be assessed in the resulting offspring, with the prediction that eggs injected with RNA derived from the "aggressive drones" will result in biased allele use in these offspring while those injected with RNA from the "docile drones" will not. Further experiments could then be performed to analyze brain metabolism and aggression in both sets of offspring. This model can also be tested through RNAi-mediated knockdown of the piRNA machinery in the parent EHB and AHB queens and drones (from colonies selected for high and low aggression) and subsequent testing of hybrid allele usage, metabolism, and aggression. Biased allele use and aggression would be expected to be lower when the machinery is knocked down paternally. Knockdowns could also be used in crosses between the highly aggressive AHB colonies. The cross of AHB^{piRNA-} queens \times AHB^{piRNA+} drones would be expected to have high aggression (perhaps even higher than the parents) due to a loss of maternal licensing and AHB^{piRNA+} queen \times AHB^{piRNA-} drones would be expected to have lower aggression (**Figure 4**).

REFERENCES

- Adam, B. (1983). *In Search of the Best Strains of Bees and the Results of the Evaluations of the Crosses and Races*. Hebdon Bridge: Northern Bee Books.
- Alaux, C., Sinha, S., Hasadsri, L., Hunt, G. J., Guzman-Novoa, E., Degrandi-Hoffman, G., et al. (2009). Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15400–15405. doi: 10.1073/pnas.0907043106
- Arechavaleta-Velasco, M. E., Alcalá-Escamilla, K., Robles-Rios, C., Tsuruda, J. M., and Hunt, G. J. (2012). Fine-scale linkage mapping reveals a small set of candidate genes influencing honey bee grooming behavior in response to varroa mites. *PLoS ONE* 7:e47269. doi: 10.1371/journal.pone.0047269
- Arechavaleta-Velasco, M. E., Hunt, G. J., and Emore, C. (2003). Quantitative trait loci that influence the expression of guarding and stinging behaviors of individual honey bees. *Behav. Genet.* 33, 357–364. doi: 10.1023/A:1023458827643
- Barros, L. F. (2013). Metabolic signaling by lactate in the brain. *Trends Neurosci.* 36, 396–404. doi: 10.1016/j.tins.2013.04.002
- Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* 58, 268–276. doi: 10.1016/j.jmeth.2012.05.001
- Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., and Hannon, G. J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322, 1387–1392. doi: 10.1126/Science.1165171
- Chandrasekaran, S., Rittschof, C. C., Djukovic, D., Gu, H., Raftery, D., Price, N. D., et al. (2015). Aggression is associated with aerobic glycolysis in the honey bee brain. *Genes Brain Behav.* 4, 158–166. doi: 10.1111/gbb.12201

The patterns that we observed in this family likely represent an inappropriate utilization of the retrograde signaling pathway because the maternal bias in gene expression in the EA family seems to occur in all tissues across life stages and not just in the brains as described for these metabolic changes in aggressive bees (Alaux et al., 2009; Li-Byarlay et al., 2014). Similarly, the metabolic deficiency in EA hybrids occurs at the whole body level even though aggressive bees are known to have increased oxidative metabolic rates at the whole body level in response to alarm pheromone (Moritz et al., 1985). Taken together, these results indicate that the whole organism-level asymmetric hybrid effects on allelic gene expression, metabolism, and aggression may be due to perturbations of established nuclear-mitochondrial signaling pathways that normally modulate brain metabolism and aggression in honeybees. While these results mark an important step in our understanding of aggression and describe a new pattern of hybrid gene regulation, additional work is necessary to better understand how differential allele expression acts on these traits, how this allelic expression is controlled and how these signaling pathways modulate aggression in the context of honeybee natural history.

AUTHOR CONTRIBUTIONS

GH and MA planned hybrid crosses, collected samples and performed stinging behavior assays. MA performed the crosses. JG, JT, and GH performed bioinformatics analyses. JG and GH wrote the manuscript with input from all authors.

ACKNOWLEDGMENTS

We would like to thank Sarah Kocher for helpful comments on the manuscript. This project was funded by NSF IOS – 1145509.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fgene.2015.00343>

- Chang, S., Wen, S. M., Chen, D. H., and Jin, P. (2009). Small regulatory RNAs in neurodevelopmental disorders. *Hum. Mol. Genet.* 18, R18–R26. doi: 10.1093/Hmg/Ddp072
- Dickman, M. J., Kucharski, R., Maleszka, R., and Hurd, P. J. (2013). Extensive histone post-translational modification in honey bees. *Insect. Biochem. Mol. Biol.* 43, 125–137. doi: 10.1016/J.Ibmb.2012.11.003
- Disteche, C. M. (2012). Dosage compensation of the sex chromosomes. *Annu. Rev. Genet.* 46, 537–560. doi: 10.1146/annurev-genet-110711-155454
- Duselis, A. R., and Vrana, P. B. (2010). Aberrant growth and pattern formation in *Peromyscus* hybrid placental development. *Biol. Reprod.* 83, 988–996. doi: 10.1095/biolreprod.110.085654
- Elsik, C. G., Worley, K. C., Bennett, A. K., Beye, M., Camara, F., Childers, C. P., et al. (2014). Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* 15:86. doi: 10.1186/1471-2164-15-86
- Flores, K., Wolschin, F., Corneveaux, J. J., Allen, A. N., Huentelman, M. J., and Amdam, G. V. (2012). Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics* 13:480. doi: 10.1186/1471-2164-13-480
- Foret, S., Kucharski, R., Pellegrini, M., Feng, S. H., Jacobsen, S. E., Robinson, G. E., et al. (2012). DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4968–4973. doi: 10.1073/Pnas.1202392109
- Gangaraju, V. K., Yin, H., Weiner, M. M., Wang, J., Huang, X. A., and Lin, H. (2011). *Drosophila* Piwi functions in Hsp90-mediated suppression of phenotypic variation. *Nat. Genet.* 43, 153–158. doi: 10.1038/ng.743
- Gershon, T. R., Crowther, A. J., Tikunov, A., Garcia, I., Annis, R., Yuan, H., et al. (2013). Hexokinase-2-mediated aerobic glycolysis is integral to cerebellar neurogenesis and pathogenesis of medulloblastoma. *Cancer Metabol.* 1:2. doi: 10.1186/2049-3002-1-2
- Goyal, M. S., Hawrylycz, M., Miller, J. A., Snyder, A. Z., and Raichle, M. E. (2014). Aerobic glycolysis in the human brain is associated with development and neonatal gene expression. *Cell Metab.* 19, 49–57. doi: 10.1016/J.Cmet.2013.11.020
- Guzman-Novoa, E., Hunt, G. J., Page, R. E., Uribe-Rubio, J. L., Prieto-Merlos, D., and Becerra-Guzman, F. (2005). Paternal effects on the defensive behavior of honeybees. *J. Hered.* 96, 376–380. doi: 10.1093/Jhered/Esi038
- Haig, D. (2002). *Genomic Imprinting and Kinship*. New Brunswick: Rutgers University Press.
- Haig, D. (2004). Genomic imprinting and kinship: how good is the evidence? *Annu. Rev. Genet.* 38, 553–585. doi: 10.1146/annurev.genet.37.110801.142741
- Harrison, J. F., and Hall, H. G. (1993). African european honeybee hybrids have low nonintermediate metabolic capacities. *Nature* 363, 258–260. doi: 10.1038/363258a0
- Herb, B. R., Wolschin, F., Hansen, K. D., Aryee, M. J., Langmead, B., Irizarry, R., et al. (2012). Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nat. Neurosci.* 15, 1371–1373. doi: 10.1038/Nn.3218
- Huang, X. A., Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2013). A major epigenetic programming mechanism guided by piRNAs. *Dev. Cell* 24, 502–516. doi: 10.1016/j.devcel.2013.01.023
- Hunt, G. J., Amdam, G. V., Schlipalius, D., Emore, C., Sardesai, N., Williams, C. E., et al. (2007). Behavioral genomics of honeybee foraging and nest defense. *Naturwissenschaften* 94, 247–267. doi: 10.1007/s00114-006-0183-1
- Hunt, G. J., Collins, A. M., Rivera, R., Page, R. E., and Guzman-Novoa, E. (1999). Quantitative trait loci influencing honeybee alarm pheromone levels. *J. Hered.* 90, 585–589. doi: 10.1093/Jhered/90.5.585
- Hunt, G. J., Guzman-Novoa, E., Fondrk, M. K., and Page, R. E. (1998). Quantitative trait loci for honey bee stinging behavior and body size. *Genetics* 148, 1203–1213.
- Hunt, G. J., Page, R. E., Fondrk, M. K., and Dullum, C. J. (1995). Major quantitative trait loci affecting honey-bee foraging behavior. *Genetics* 141, 1537–1545.
- Johnson, C. L., and Spence, A. M. (2011). Epigenetic licensing of germline gene expression by maternal RNA in *C. elegans*. *Science* 333, 1311–1314. doi: 10.1126/Science.1208178
- Kocher, S. D., Tsuruda, J. M., Gibson, J. D., Emore, C. M., Arechavala-Velasco, M. E., Queller, D. C., et al. (2015). A search for parent-of-origin effects on honey bee gene expression. *G3* 5, 1657–1662. doi: 10.1534/g3.115.017814
- Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319, 1827–1830. doi: 10.1126/Science.1153069
- Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., et al. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Gene Dev.* 27, 390–399. doi: 10.1101/Gad.209841.112
- Li-Byarlay, H., Rittschof, C. C., Massey, J. H., Pittendrigh, B. R., and Robinson, G. E. (2014). Socially responsive effects of brain oxidative metabolism on aggression. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12533–12537. doi: 10.1073/pnas.1412306111
- Linksvayer, T. A., Rueppell, O., Siegel, A., Kaftanoglu, O., Page, R. E., and Amdam, G. V. (2009). The genetic basis of transgressive ovary size in honeybee workers. *Genetics* 183, 693–707. doi: 10.1534/Genetics.109.105452
- Liu, Z., and Butow, R. A. (2006). Mitochondrial retrograde signaling. *Annu. Rev. Genet.* 40, 159–185. doi: 10.1146/annurev.genet.40.110405.090613
- Loschiavo, M., Nguyen, Q. K., Duselis, A. R., and Vrana, P. B. (2007). Mapping and identification of candidate loci responsible for *Peromyscus hybrid* overgrowth. *Mamm. Genome* 18, 75–85. doi: 10.1007/s00335-006-0083-x
- Lunt, S. Y., and Vander Heiden, M. G. (2011). Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. *Annu. Rev. Cell Dev. Biol.* 27, 441–464. doi: 10.1146/annurev-cellbio-092910-154237
- Moore, T., and Haig, D. (1991). Genomic imprinting in mammalian development - a parental tug-of-war. *Trends Genet.* 7, 45–49. doi: 10.1016/0168-9525(91)90040-W
- Moritz, R. F. A., Southwick, E. E., and Breh, M. (1985). A metabolic test for the quantitative analysis of alarm behavior of honeybees (*Apis mellifera* L.). *J. Exp. Zool.* 235, 1–5. doi: 10.1002/Jez.1402350102
- Oldroyd, B. P., Allsopp, M. H., Roth, K. M., Remnant, E. J., Drewell, R. A., and Beekman, M. (2014). A parent-of-origin effect on honeybee worker ovary size. *Proc. Roy. Soc. B.* 281:2013238. doi: 10.1098/Rspb.2013.2388
- Oxley, P. R., Thompson, G. J., and Oldroyd, B. P. (2008). Four quantitative trait loci that influence worker sterility in the honeybee (*Apis mellifera*). *Genetics* 179, 1337–1343. doi: 10.1534/Genetics.108.087270
- Page, R. E., Fondrk, M. K., Hunt, G. J., Guzman-Novoa, E., Humphries, M. A., Nguyen, K., et al. (2000). Genetic dissection of honeybee (*Apis mellifera* L.) foraging behavior. *J. Hered.* 91, 474–479. doi: 10.1093/Jhered/91.6.474
- Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S. E., et al. (2008). A mitochondrial protein compendium elucidates complex I disease biology. *Cell* 134, 112–123. doi: 10.1016/J.Cell.2008.06.016
- Pantano, L., Jodar, M., Bak, M., Balleza, J. L., Tommerup, N., Oliva, R., et al. (2015). The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. *RNA* 21, 1085–1095. doi: 10.1261/rna.046482.114
- Pesante, D. G., Rinderer, T. E., and Collins, A. M. (1987). Differential pollen collection by africanized and european honeybees in venezuela. *J. Apicult. Res.* 26, 24–29.
- Queller, D. C. (2003). Theory of genomic imprinting conflict in social insects. *BMC Evol. Biol.* 3:15. doi: 10.1186/1471-2148-3-15
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Ross, C. R., DeFelice, D. S., Hunt, G. J., Ihle, K. E., Amdam, G. V., and Rueppell, O. (2015). Genomic correlates of recombination rate and its variability across eight recombination maps in the western honey bee (*Apis mellifera* L.). *BMC Genomics* 16:107. doi: 10.1186/s12864-015-1281-2
- Ruden, D. M., Cingolani, P. E., Sen, A., Qu, W., Wang, L., Senut, M. C., et al. (2015). Epigenetics as an answer to Darwin's "special difficulty," Part 2: natural selection of metastable epialleles in honeybee castes. *Front. Genet.* 6:60. doi: 10.3389/fgene.2015.00060
- Ruden, D. M., Jamison, D. C., Zeeberg, B. R., Garfinkel, M. D., Weinstein, J. N., Rasouli, P., et al. (2008). The EDGE hypothesis: epigenetically directed genetic errors in repeat-containing proteins (RCPs) involved in evolution, neuroendocrine signaling, and cancer. *Front. Neuroendocrinol.* 29:428. doi: 10.1016/j.yfrne.2007.12.004
- Rueppell, O., Metheny, J. D., Linksvayer, T., Fondrk, M. K., Page, R. E., and Amdam, G. V. (2011). Genetic architecture of ovary size and

- asymmetry in European honeybee workers. *Heredity* 106, 894–903. doi: 10.1038/Hdy.2010.138
- Sequencing Consortium (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443, 931–949. doi: 10.1038/Nature05260
- Shirayama, M., Seth, M., Lee, H. C., Gu, W., Ishidate, T., Conte, D., et al. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell* 150, 65–77. doi: 10.1016/j.cell.2012.06.015
- Shorter, J. R., Arechavaleta-Velasco, M., Robles-Rios, C., and Hunt, G. J. (2012). A genetic analysis of the stinging and guarding behaviors of the honey bee. *Behav. Genet.* 42, 663–674. doi: 10.1007/s10519-012-9530-5
- Smith, A. C., Blackshaw, J. A., and Robinson, A. J. (2012). MitoMiner: a data warehouse for mitochondrial proteomics data. *Nucleic Acids Res.* 40, D1160–D1167. doi: 10.1093/Nar/Gkr1101
- Sollars, V., Lu, X. Y., Xiao, L., Wang, X. Y., Garfinkel, M. D., and Ruden, D. M. (2003). Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nat. Genet.* 33, 70–74. doi: 10.1038/Ng1067
- Spannhoff, A., Kim, Y. K., Raynal, N. J. M., Gharibyan, V., Su, M. B., Zhou, Y. Y., et al. (2011). Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. *EMBO Rep.* 12, 238–243. doi: 10.1038/Emborep.2011.9
- Tsuruda, J. M., Harris, J. W., Bourgeois, L., Danka, R. G., and Hunt, G. J. (2012). High-resolution linkage analyses to identify genes that influence varroa sensitive hygiene behavior in honey bees. *PLoS ONE* 7:e48276. doi: 10.1371/journal.pone.0048276
- Turelli, M., and Moyle, L. C. (2007). Asymmetric postmating isolation: darwin's corollary to Haldane's rule. *Genetics* 176, 1059–1088. doi: 10.1534/Genetics.106.065979
- Wang, X., and Clark, A. G. (2014). Using next-generation RNA sequencing to identify imprinted genes. *Heredity* 113, 156–166. doi: 10.1038/Hdy.2014.18
- Wang, Y., Jorda, M., Jones, P. L., Maleszka, R., Ling, X., Robertson, H. M., et al. (2006). Functional CpG methylation system in a social insect. *Science* 314, 645–647. doi: 10.1126/science.1135213
- Warburg, O. (1956). On the origin of cancer cells. *Science* 123, 309–314. doi: 10.1126/science.123.3191.309
- Wolf, J. B. (2009). Cytonuclear interactions can favor the evolution of genomic imprinting. *Evolution* 63, 1364–1371. doi: 10.1111/J.1558-5646.2009.00632.X
- Wolf, J. B., Oakey, R. J., and Feil, R. (2014). Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. *Heredity* 113, 167–175. doi: 10.1038/hdy.2014.11
- Yuan, S., Oliver, D., Schuster, A., Zheng, H., and Yan, W. (2015). Breeding scheme and maternal small RNAs affect the efficiency of transgenerational inheritance of a paramutation in mice. *Sci. Rep.* 5:9266. doi: 10.1038/srep09266

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Gibson, Arechavaleta-Velasco, Tsuruda and Hunt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The evolutionary dynamics of major regulators for sexual development among Hymenoptera species

Matthias Biewer^{1,2}, Francisca Schlesinger^{1,3} and Martin Hasselmann^{1,2*}

¹ Population Genetics of Social Insects, Institute of Genetics, University of Cologne, Cologne, Germany, ² Livestock Population Genomics Group, Institute of Animal Science, University of Hohenheim, Stuttgart, Germany, ³ Institute of Bee Research, Hohen Neuendorf, Germany

OPEN ACCESS

Edited by:

Juergen Rudolf Gadau,
Arizona State University, USA

Reviewed by:

Frederic Guy Brunet,
The Pennsylvania State University,
USA

Ehab Abouheif,
McGill University, Canada

*Correspondence:

Martin Hasselmann,
Livestock Population Genomics
Group, Institute of Animal Science,
University of Hohenheim,
Garbenstrasse 17, 70599 Stuttgart,
Germany
martin.hasselmann@
uni-hohenheim.de

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 29 September 2014

Accepted: 16 March 2015

Published: 10 April 2015

Citation:

Biewer M, Schlesinger F and
Hasselmann M (2015) The
evolutionary dynamics of major
regulators for sexual development
among Hymenoptera species.
Front. Genet. 6:124.
doi: 10.3389/fgene.2015.00124

All hymenopteran species, such as bees, wasps and ants, are characterized by the common principle of haplodiploid sex determination in which haploid males arise from unfertilized eggs and females from fertilized eggs. The underlying molecular mechanism has been studied in detail in the western honey bee *Apis mellifera*, in which the gene *complementary sex determiner* (*csd*) acts as primary signal of the sex determining pathway, initiating female development by *csd*-heterozygotes. *Csd* arose from gene duplication of the *feminizer* (*fem*) gene, a *transformer* (*tra*) ortholog, and mediates in conjunction with *transformer2* (*tra2*) sex-specific splicing of *fem*. Comparative molecular analyses identified *fem/tra* and its downstream target *doublesex* (*dsx*) as conserved unit within the sex determining pathway of holometabolous insects. In this study, we aim to examine evolutionary differences among these key regulators. Our main hypothesis is that sex determining key regulators in Hymenoptera species show signs of coevolution within single phylogenetic lineages. We take advantage of several newly sequenced genomes of bee species to test this hypothesis using bioinformatic approaches. We found evidences that duplications of *fem* are restricted to certain bee lineages and notable amino acid differences of *tra2* between *Apis* and non-*Apis* species propose structural changes in Tra2 protein affecting co-regulatory function on target genes. These findings may help to gain deeper insights into the ancestral mode of hymenopteran sex determination and support the common view of the remarkable evolutionary flexibility in this regulatory pathway.

Keywords: gene duplications, sex determination, adaptive evolution, regulatory changes, pathway evolution

Introduction

Understanding the evolution of biological pathways and the driving processes shaping them still belongs to the central questions in biology. Studying genetic networks and their underlying selectional and developmental processes can provide important insights into the divers evolutionary trajectories of molecules between species (Pires-da Silva and Sommer, 2003; Wilkins, 2007; Fani and Fondi, 2009; Davidson, 2010; Peter and Davidson, 2011). As one common process, gene duplication has been identified to play a key role in providing novel or modified gene functions resulting from various forms of selection acting on the paralogous copies (Innan and Kondrashov, 2010). Following the model of neofunctionalization, a paralogous copy may

acquire a novel function not present in the gene from which it arose. Besides positive selection promoting the fixation of advantageous mutations in this copy, exon or domain shuffling may further contribute to the evolution of a neofunctionalized gene. Among others, the well-established duplication-degeneration-complementation (DDC) model provides a basis for the evolution of an modified (subfunctionalized) function in the paralogous gene (Force et al., 2005).

The sex determination pathway in honey bees constitutes a well-studied example, in which gene duplication has been identified to play a major role in its evolutionary history (Hassellmann et al., 2008). Common for all hymenopteran species (ants, wasps, and bees) is the principle of haplodiploidy in which males are haploid and develop from unfertilized eggs, whereas females are diploid and arise from fertilized eggs (Bull, 1983). The underlying molecular signals and regulatory key genes involved in the sex determination pathway have been studied in greater detail for only two hymenopteran species, the parasitic wasp *Nasonia vitripennis* (Beukeboom et al., 2007; Verhulst et al., 2013; van de Zande and Verhulst, 2014) and the western European honey bee *Apis mellifera* (Beye et al., 2003; Hassellmann and Beye, 2004; Hassellmann et al., 2008; Gempe et al., 2009) with an estimated divergence time of about 170 million years ago (Werren et al., 2010).

With the now available new genomes of bee species covering a divergence time of about 100 million years, we are closing the so far existing gap between *Apis* and *Nasonia*. Consequently, we can now study the evolution of the sex determination pathway and the driving forces shaping key components on a refined scale. Thus, one of the obvious questions is whether lineage specific events such as gene duplications can be observed to affect key regulator coevolution. Within the sex determination pathway of insects, a conserved unit of genes has been identified, giving rise to a transductional core downstream of the primary signal (Bopp et al., 2014) that transmits the information of the primary signal and releases male/female specific developmental regulatory signals to a variety of target genes. We hypothesize that the core unit of sex determining genes is relative conserved in all Hymenopterans; however, evolutionary processes may have shaped these genes and the additional cofactors lineage specifically.

In the honeybee, the primary signal is the gene *complementary sex determiner* (*csd*), which arose from gene duplication of its copy *feminizer* (*fem*) (Beye et al., 2003; Hassellmann et al., 2008). The molecular decision of male or female development is mediated by a multiallelic system of protein-protein interaction, in which a heterozygous conformation leads to female development, while homo- and hemizygotes develop into males. The evolutionary history of the paralogous genes has been shaped by contrasting forms of selection in *Apis*: after the duplication, *csd* experienced strong positive selection, following the model of neofunctionalization, whereas *fem* evolved under strong purifying selection (Hassellmann et al., 2010). The formation of specific protein regions such as a hypervariable region (HVR) and a protein-interacting coiled-coil motif are important for the rise and function of *csd*-alleles. Molecular functional analysis provided evidence for sex-specific splicing of *fem*, initiated by the

allelic state of *csd*, in which heterozygote *csd* lead to female-specific *fem*-mRNA splicing. Acting as binary switch gene, *fem* transcripts are maintained and enhanced by an autoregulatory feedback loop of the Fem protein (Hassellmann et al., 2008; Gempe et al., 2009). This serine-arginine (SR) rich protein and its ortholog *transformer* (*tra*) are differentially spliced, either to a female functional or to a male nonfunctional isoform, as found for other insect species (Butler et al., 1986; Pane et al., 2002; Sarno et al., 2010; Verhulst et al., 2010). The processing of sex-specific information by the *fem/tra* gene is conserved in these insects and the sex determining pathway converged at this level (Gempe et al., 2009).

The absence of an RNA recognition motif (RRM) domain in *Apis fem* requires a cofactor protein for RNA binding to mediate the sex-specific splicing process. It has been shown by Nissen et al. (2012) that the Transformer2 (Tra2) protein in conjunction with the Csd protein transmit the sex-specific splicing of *fem*-mRNA. Tra2 is evolutionary conserved among insects and characterized by a single, 80–90 amino acid long RRM domain, flanked by two SR domains. Two sequence elements (RNP1 and RNP2) have been shown to be directly involved in RNA recognition. With this ability to recognize RNA motifs, Tra2 facilitates the *fem/tra* autoregulatory splicing loop, which can be found in other insect species, except *Drosophila* (Gempe et al., 2009; Salvemini et al., 2009; Hediger et al., 2010; Sarno et al., 2010).

The female-specific active Fem (Tra)/Tra2 complex regulates the differential splicing of the downstream target for sex-specific development, *doublesex* (*dsx*). The gene *doublesex* (*dsx*) represents the key gene in sex determination of insects as the most downstream component of the pathway regulating sex-specific phenotypes (Burtis and Baker, 1989; Cline and Meyer, 1996). Acting as transcription factor, *dsx* encodes a protein with a zinc-finger DNA-binding domain (DM domain). In all insect species studied so far, gene structure and pattern of sex-specific splicing is generally conserved (Cho et al., 2007). Female and male transcripts consists of two oligomerization domains (OD1 and OD2) harboring DNA and protein interaction functions. The use of different splice sites at the C-terminal region results in OD2 sequence variation that alters the female- and male-specific regulation of target genes, which regulates the sex-specific splicing of pre-mRNA into male or female isoforms for the particular development as an essential transductional core of the pathway.

Among hymenopteran non-*Apis* species, the molecular basis of sex determination is best understood for the phylogenetically most basal parasitic wasp *Nasonia*, in which an alternative mode of haplodiploid sex determination evolved (Verhulst et al., 2013). Similar to what is known for many other dipteran insects, *transformer* mRNA of *Nasonia vitripennis* (*Nvtra* mRNA) is maternally provided to all eggs, however only in fertilized eggs *Nvtra* transcription can initiate and maintain female *Nvtra* mRNA by an autoregulatory feedback loop. In unfertilized eggs, the maternally provided genome induces low level of *Nvtra* expression, leading to the hypothesis of genomic imprinting as sex determination mechanism (Verhulst et al., 2010). Recent findings indicate that alleles of an trans-acting factor (*womanizer*), likely to be maternally provided may have been recruited as novel

component in the sex determining pathway (Verhulst et al., 2013).

There is increasing evidence that the initial signals of sex determining pathways may evolve rapidly, contributing to the astonishing diversity of species. The underlying processes driving this rapid evolution may be gene duplications, accompanied by the gain of novel or modified function and changes in the selective regime under which the key genes evolve. In our study we provide evidence for the importance of instantaneously occurring events such as gene duplications and lineage specific mutations that affect key regulator coevolution within the sex determination pathway of hymenopteran species.

Materials and Methods

Sequence Data

Genome assemblies and annotations of recently sequenced bee species (Kapheim et al., in revision) were used to identify gene copies of interest (*feminizer*—*fem*, *transformer2*—*tra2*, *doublesex*—*dsx*), taking Amell vs. 4.5, OGS 3.2 as reference and using various blast parameters to avoid non-detection errors. Hidden Markov profile searches (Eddy, 1998) were performed to search specifically for *fem* paralogs in bee genomes using HMMer3 on protein (HMMsearch) and nucleotide (nHMMer) level (Eddy, 2009). Multiple sequence alignments were generated using MUSCLE (Edgar, 2004) and optimized manually. To reduce the loss of informative sites due to incomplete or misleading annotations, experimentally proven and publicly available data were used for some species and GenBank and OrthoDB entries were used for *fem* and paralogous copies, *tra2* and *dsx* sequences. The sequences used for our analyses for comparing functional and evolutionary relationships were retrieved from GenBank and OrthoDB. Accession numbers are given in the Supplementary Tables 1–3.

Evolutionary Analyses

Genealogies were reconstructed after applying Model Test (Posada and Crandall, 1998) on the dataset to determine the evolutionary substitution model that fitted the data best. The model with the lowest BIC (Bayesian Information Criterion) scores was considered best for describing the substitution pattern. Non-uniformity of evolutionary rates among sites was modeled by using a discrete Gamma distribution (+G) with 5 rate categories. Evolutionary trees were constructed using the maximum likelihood method (JTT model) implemented in MEGA6 (Tamura et al., 2013). Examination of exonic splicing regulatory elements (ESR) was performed on the ESR search website (<http://esrsearch.tau.ac.il/>) using the highest number of available parameters (Fairbrother et al., 2002). Further, analyses of conserved protein domains and protein function were performed with conserved domain search module (Marchler-Bauer et al., 2011) implemented on www.ncbi.nlm.nih.gov and InterPro (Hunter et al., 2012), <http://www.ebi.ac.uk/interpro/>). The program COIL (implemented online under: http://embnet.vital-it.ch/software/COILS_form.html) was used to search specifically for predicted coiled-coil regions. The COIL

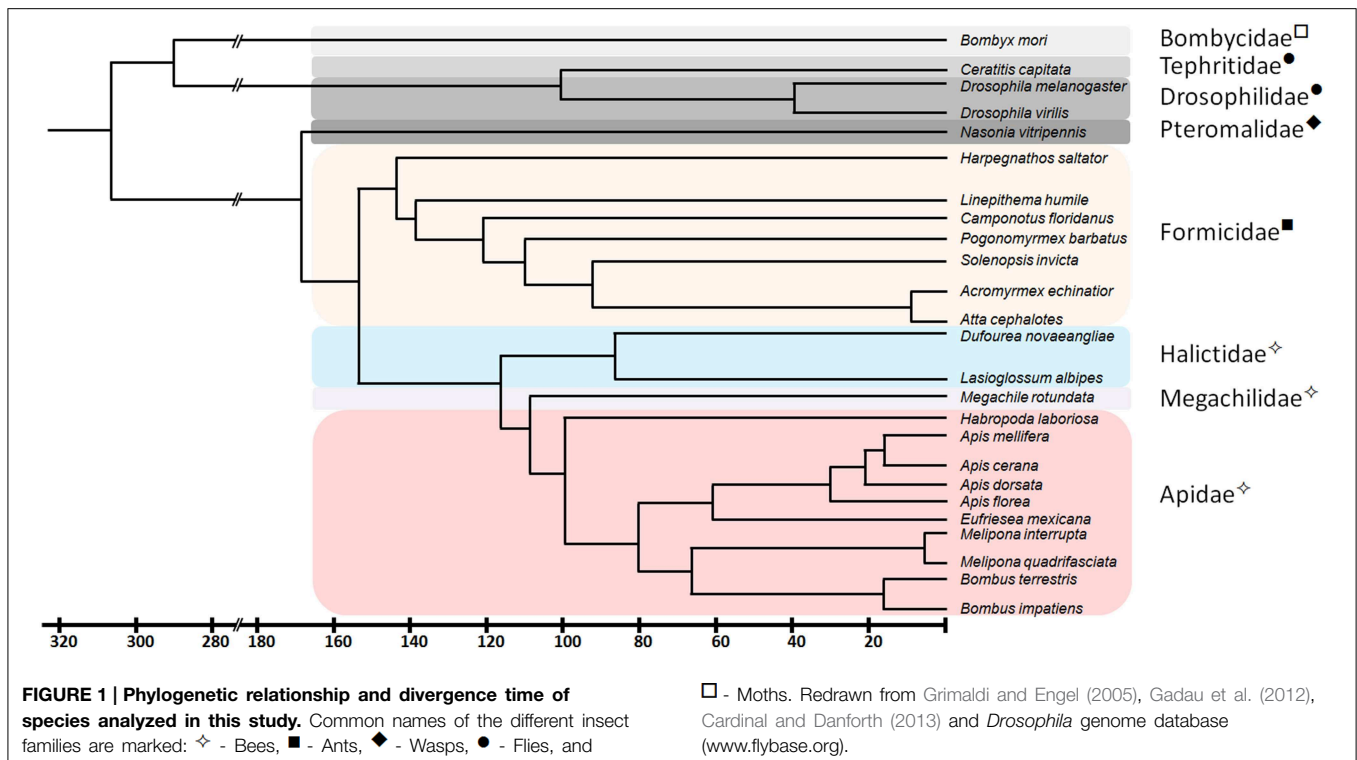
program compares the query sequence to a database of known coiled coils and derives a similarity score. The probability that the sequence will form a coiled-coil motif is obtained within the program by comparing the similarity score against the distribution of scores in globular and coiled-coil proteins. Sequence-based motifs were identified and analyzed using the MEME suite package (<http://meme.nbcr.net/meme/>, Bailey et al., 2009). The significance of the motif is determined by first finding the most statistically significant (low *E*-value) motifs. Motifs are shown as sequence logos, represented by position-specific probability matrices that specify the probability of each possible letter appearing at each possible position in an occurrence of the motif. Displayed as stacks of letters at each position in the motif, the total height of the stack is the “information content” of that position in the motif in bits.

Results

Diversification of *Feminizer* Gene Duplicates

In a previous study (Kapheim et al., in revision), we have identified *fem* paralogs and orthologs of recently obtained genomic resources of bees representing different levels of social organization (Figure 1). Representative species were analyzed from Apini (the western European honey bee *Apis mellifera* and the dwarf honey bee *Apis florea*), Bombini (the buff-tailed bumble bees *Bombus terrestris* and *Bombus impatiens*), Euglossini (the orchid bee *Eufriesea mexicana*), Meliponini (*Melipona quadrifasciata*), Megachilini (the leafcutter bee *Megachile rotundata*) and Halictini (*Lasioglossum albipes* and *Habropoda laboriosa*). We noticed that the occurrence of *fem* duplications varies among different lineages in conjunction with varying signs of diversifying and negative selection. When including *transformer* (*tra*) orthologous sequences of seven ant species and one parasitic wasp, the sequences fall into two major clades, separating *ant-ra* from the remaining sequences (Figure 2). All genes share an arginine-serine rich and a proline rich domain, establishing these copies as strong candidates to be involved in protein interaction and splicing processes. For the paralogous genes *fem* and *csd* within the *Apis* lineage, evidence for both processes has been given by numerous functional studies (Hasselman et al., 2008; Gempe et al., 2009; Nissen et al., 2012). In the bumble bee *Bombus terrestris*, for *fem* and its paralogous copy *fem1* several splice forms were identified (Biewer et al., in revision); in the stingless bee *Melipona interrupta* the single copy *fem* gene is characterized by two splice forms (Brito et al., unpublished).

Here, we focus on the evolutionary dynamic of Fem proteins among bees using amino acid sequence motifs. We follow the hypothesis that characteristic motifs should be found in all bees harboring changes in species-specific paralogs of *fem*. These could hint to lineage-specific modifications of protein interaction in the sex determination pathways. Our hypothesis is supported by the previous study of Koch et al. (2014) showing the independent origin of *fem* paralogs in *Apis* and *Bombus* (and Ants) and thus different evolutionary fates, for which the multiallelic evolution of *csd* stands as one remarkable example (Hasselman et al., 2008; Lechner et al., 2014).



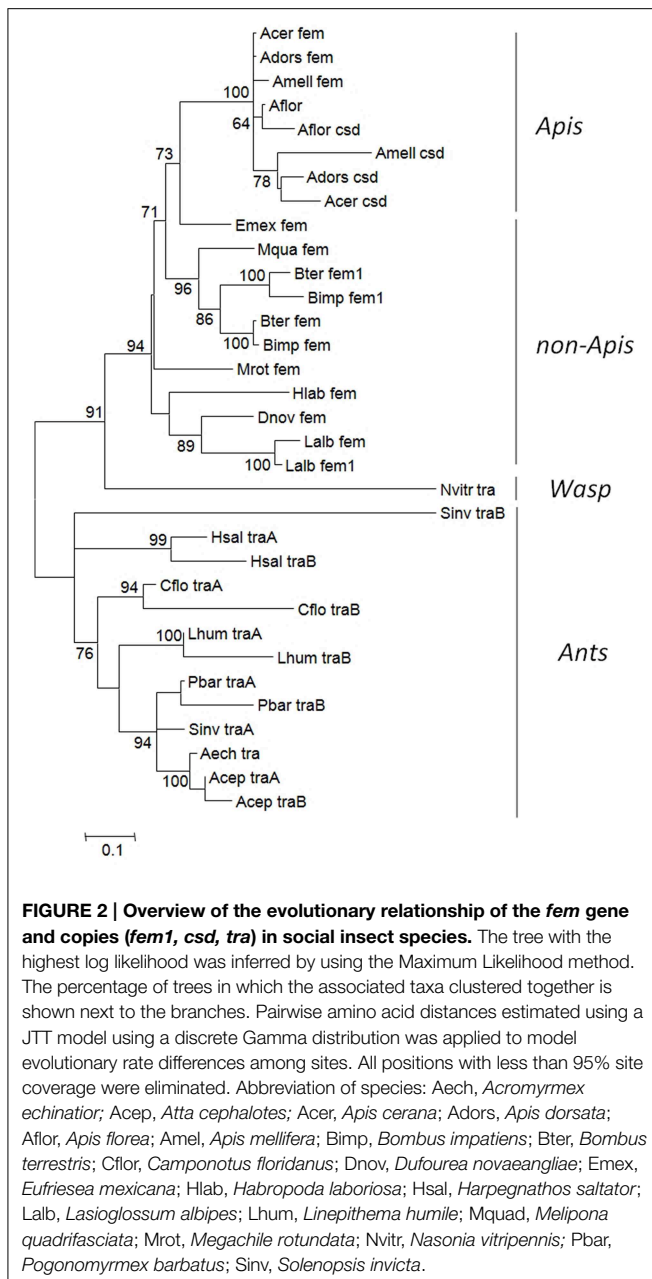
In order to test our hypothesis, we first evaluated the amino acid motifs in Fem protein sequences of bee species and the wasp *Nasonia vitripennis* using the MEME program package (see Materials and Methods). Six motifs with the best scoring *E*-values (*E*-values ranging from $1.0e^{-488}$ to $8.0e^{-143}$) were detected, represented by sequence logos (see Figure 3 and Supplementary Figure 1). The relative positions of these motifs in the protein are located in the N-terminal and C-terminal, as well as in-between regions of the protein (Supplementary Figure 2). Sequence logos illustrate the evolutionary conservation of several amino acids, the most prominent ones are Glutamic acid (E), Arginine (R), Lysine (K), Glutamine (G) and Proline (P), as well as variable positions, giving rise to species-specific divergence.

Next, we evaluated the phylogenetic signal for each motif by constructing genealogies based on the maximum likelihood algorithm. Amino acid divergence of *Apis* compared to non-*Apis* bees is most pronounced for motif 1, 5, and 6, resulting in three separated and clearly supported (78–98% bootstrap support) clusters. The sequence clustering is less obvious for motif 3 and an unresolved branching pattern results from motif 2 and 4. Interestingly, motif 5 locates in direct vicinity of the predicted coiled-coil (cc)-motif, identified to be specifically evolved in *csd* of *A. mellifera*, *A. cerana* and *A. dorsata* by positive selection of six non-synonymous changes (Hasselmann et al., 2008). No such cc-motif can be detected on the homolog positions in *A. florea csd* (Biewer et al., in revision) and in those of *fem* or its paralogs for other hymenopteran species (Supplementary Figure 3). However, we identified the presence of a cc-motif in the region of motif 3 for *A. florea csd* (Biewer et al., in revision) that coincides with an α and β sheet PLP-dependent transferase-like structure

predicted by the SMART tool (<http://smart.embl-heidelberg.de/>) in non-*Apis* bees. We conclude that at least regions of motif 3 and 5 are candidates for having resulted from lineage-specific evolution in protein interaction processes associated with the sex determination pathway in bees.

Lineage Specific Coevolution of *Transformer2*

Subsequently, we followed the hypothesis that coevolutionary signs should be detectable within the *tra2* gene as major co-regulator in the sex determining pathway, if the key regulator *fem* and (if present) its paralogous copies evolved with a modified function. Therefore we first aligned *Tra2* protein sequences from orthologs of 10 bee species and three other insect species (*N. vitripennis*, *B. mori*, *D. melanogaster*) and focused on the RNA recognition motif (RRM), which is flanked by two SR rich regions. The RRM contains about 80 aa and forms a $\beta\alpha\beta\beta\alpha\beta$ barrel motif, whereas on the third β sheet the two conserved elements RNP1 and RNP2 are located, known to be directly involved in the RNA recognition of *dsx* in *D. melanogaster* (Chandler et al., 1997), Figure 4. No amino acid changes between bee species exist in RNP2 whereas the remaining part of the RRM show pronounced differences among the species. Two observations are of particular interest: First, all non-bee species compared to the bee species show numerous amino acid changes, ranging from 9 aa (*Nvit*) to 28 aa (*Dmel*) which reflects their phylogenetic distance. Second, within the bee species, the *Apis* species *A. mellifera* and *A. florea* are consistently different for 9 amino acids that are otherwise conserved in bees, two of them locate in the RNP1 region. In addition, *Bombus* and *Melipona* species have one common amino acid replacement, as compared to the other species. When



compared over full length, *Apis*-Tra2 shows 21 of otherwise fixed amino acid differences compared to non-*Apis* species. In previous analyses (Kapheim et al., in revision), we noticed that the RRM domain is on average more divergent between *Apis* and non-*Apis* species than outside of the domain ($P < 0.1$), predominantly for the downstream region ($P < 0.01$). These unexpected findings could hint to an *Apis* specific functional association of *tra2* with *fem*, depending on the lineage specific *fem* copies and their function.

We compared the relative evolutionary rate of *tra2* bee sequences to further evaluate the differences between *Apis* and non-*Apis* species. Using Tajima's relative rate test, we tested the null hypothesis of equal molecular clock rate between *Mint*/*Amel*,

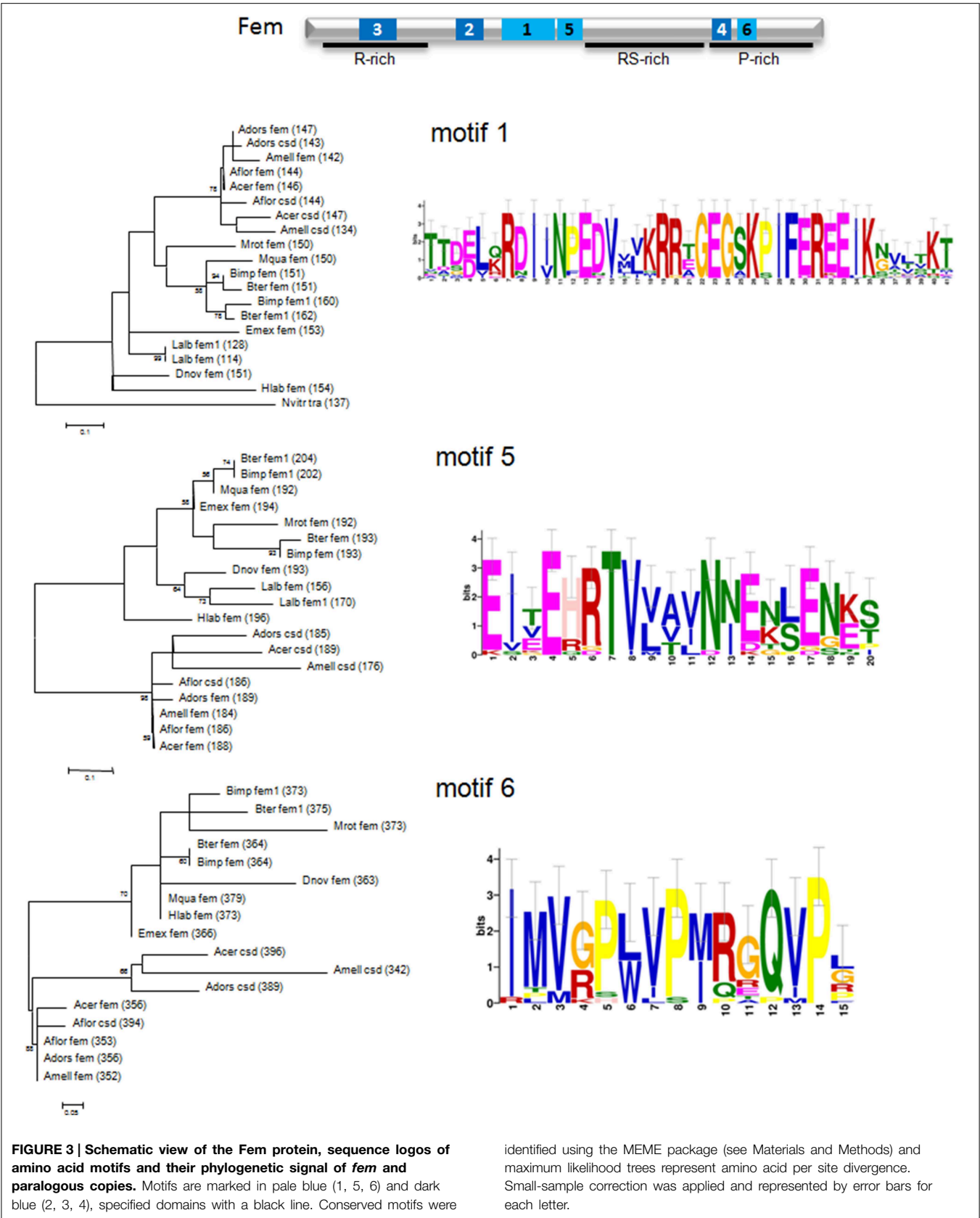
Mint/*Bter*, and *Mint*/*Bter* using four non-*Apis* species (*Mrot*, *Emex*, *Hlab*, *Nvit*) as outgroup. Tests on molecular evolutionary rates of *fem* reveal a higher rate in *Apis* compared to non-*Apis* species ($P < 0.01$ for all pairwise comparisons, using different outgroups). No difference in evolutionary rate was detected between non-*Apis* (*Mint*/*Bter*) *tra2* comparison ($P > 0.5$). To test, whether these evolutionary rate differences is specific to *tra2* or a general phenomenon among *Apis* and non-*Apis* species, two reference genes were analyzed (*elongation factor 1* and *GB11211*—a gene know to be located in close vicinity of the *fem* gene within the sex determination locus, Hasselmann et al., 2010). No rate differences were detected between *Apis* and non-*Apis* for both genes ($P > 0.05$).

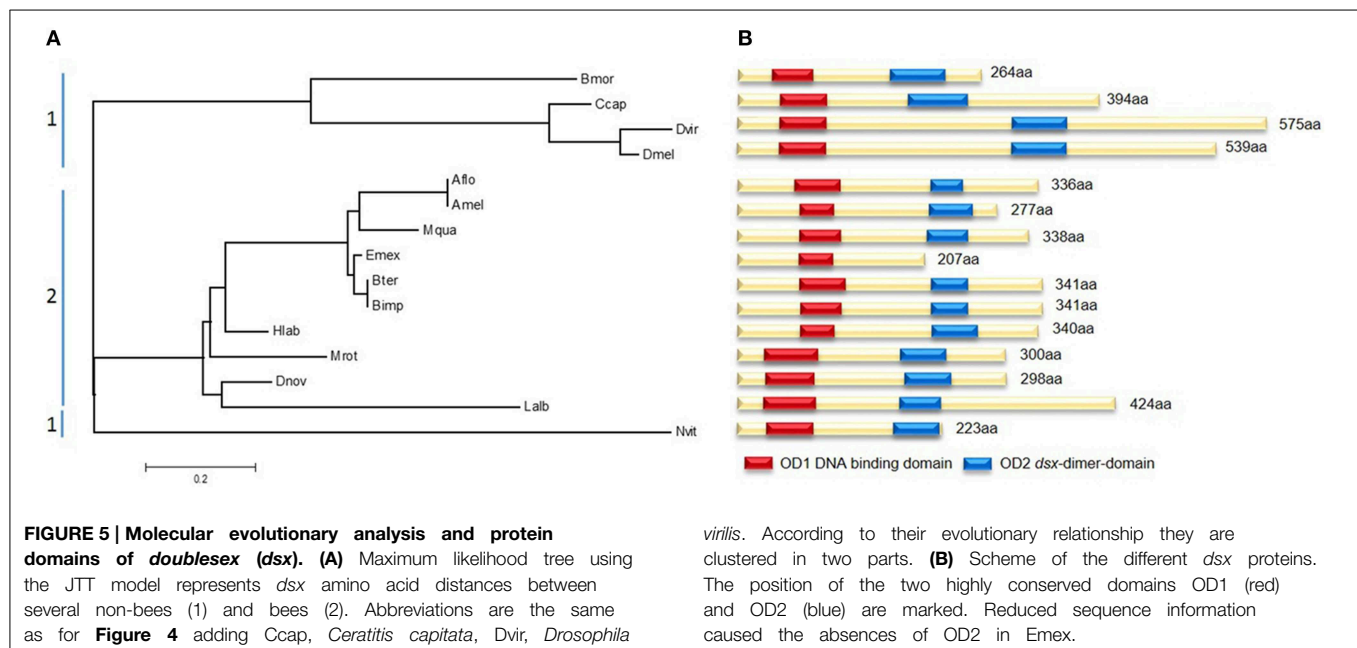
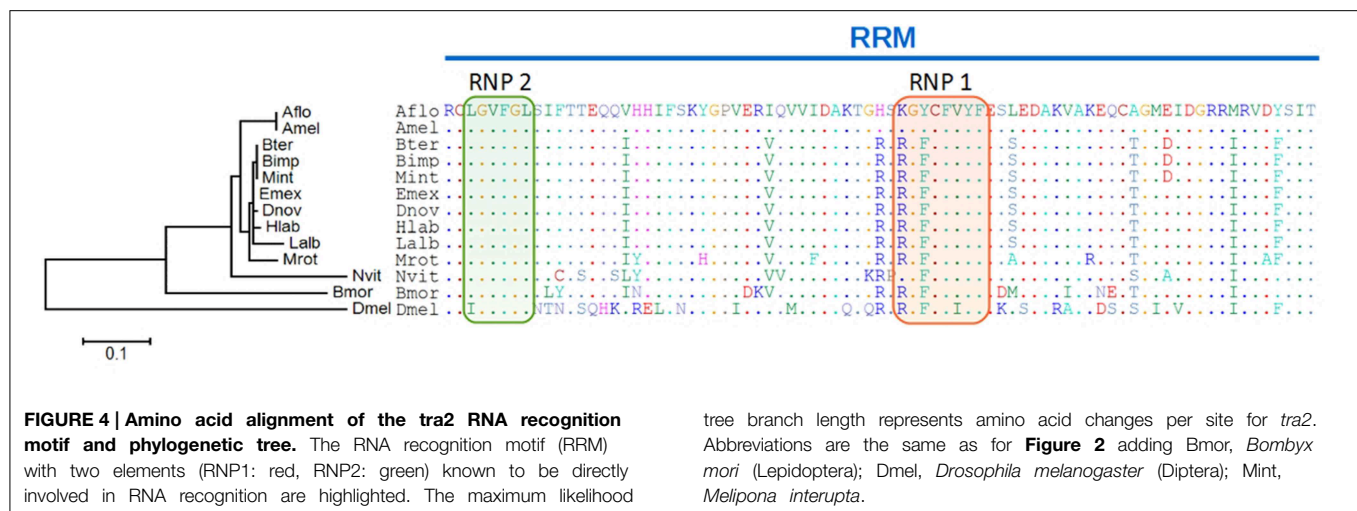
The Evolutionary Conserved Key Regulator *Doublesex*

Sequence analyses of different bee and non-bee species indicate fundamental changes in the initial regulatory elements of the sex determining pathways. Although the gene *doublesex* (*dsx*), which is located toward the bottom of the pathway, shows large amino acid sequence divergence between species, two major domains remain highly conserved (Figure 5). OD1 harbors a DNA-binding domain containing a zinc-finger, while OD2 includes a dimerization domain which was found in all analyzed species except *Eufriesea mexicana*, which could be due to poor sequence quality. The evolutionary tree of *dsx* shows a distinct segregation between bee and non-bee species (Figure 5A). This might be not only due to evolutionary distances by nucleotide changes, but also by structural changes. All non-bee species (except the wasp *Nvit*) showed a female-specific exon which was not present in the bee species. In *D. melanogaster* this exon contains six 13-nucleotide repeats, which are exonic splicing regulatory elements (ESR) and are essential for Tra2 binding to *dsx* (Baker, 1989). This repeats were not specifically found in the other non-bee species (e.g., *Bmor*, *Ccap*), whereas the presence of the female specific exon might suggests a similar mechanism of protein-binding to *dsx* as it was found in *Drosophila* and other dipteran species (data not shown; Ruiz et al., 2007). Since this female specific exon seems to be absent in the bee species, one hypothesis could be that these ESR are located in other positions of the gene. We tested for this and did not find any evidence of similar ESR in other positions of the gene (data not shown). Alternatively, bees might have evolved other regulatory elements transmitting the Tra2 binding to *dsx*. This hypothesis should be tested in future experimental studies.

Discussion

Studying the evolution of genetic components within regulatory pathways may shed light on the flexibility of how similar requirements are satisfied by different approaches in nature. This ubiquitous phenomenon, known as developmental system drift (DSD) has been identified to establish homologous conserved traits by developmental mechanism that are diverged among species (True and Haag, 2001; Abouheif and Wray, 2002; Nahmad et al., 2008). Here, we focused on major regulators of the sex determination pathway in social insect species, elucidating their





evolutionary dynamic. The transductional core of the sex determining pathway [*fem*(*tra*)/*dsx* complex] is evolutionarily conserved in insects over more than 280 million years of divergence (Diptera/Hymenoptera). Upstream initial signals regulating the sex-specific splicing of *fem/tra* may evolve within much shorter time, being consistent with the bottom-up theory (Wilkins, 1995) and the hour-glass model recently developed by Bopp et al. (2014). The different copy numbers of *fem* duplications found in bee genomes (this study and Kapheim et al., in revision) would allow either lineage specific gene loss (in *Mqua*, *Mrot*, *Dnov*, and *Hlab*) from a single ancestral duplication event or independent gene duplications (in *Apis*, *Bter*, *Bimp*, *Lalb*).

Our data from a variety of bee species now provide evidence for different evolutionary fates of the key regulator *fem* in bees. Gene duplications of *fem* in only some of the bee lineages in

conjunction with diversifying selection seem to be the major force driving the evolution of *fem* and its paralogous copies. We identified amino acid motifs in *fem* and its copies that coincides with the prediction of protein structures (e.g., coiled-coil) known to be involved in protein interaction processes. The amino acid divergence between *Apis* and non-*Apis* species on these motifs favors the hypothesis that functional constraints may have shaped these parts of the protein differently. Among them, motif 5 reveals highest divergence between *Apis* and non-*Apis* species (73% total aa divergence, Supplementary Table 4) whereas to the low overall divergence (0.4% aa) between both groups hints to an lineage specific accumulation of amino acid changes. Recent analyses of Koch et al. (2014) provide evidence for independent gene duplication of *fem* in *Apis* and *Bombus* and reject the hypothesis of concerted evolution between *fem/csd* and *fem/fem1* as

proposed by Privman et al. (2013) and Schmieder et al. (2012). By these processes, primary sex determining signals may evolve rapidly including modified function of known key regulators. This hypothesis could be supported by the greater divergence of the Tra2-RRM-domain, particularly between *Apis* and non-*Apis* bees, indicating a lineage specific functional interaction of *tra2/csd*, *tra2/fem* or *tra2/fem1*.

Evolutionary Changes in *tra2* but Not in *Dsx* between *Apis* and Non-*Apis* Bees

There are several indications that *tra2* in *Apis* has evolved differently compared to other bee species. The *tra2* genealogy (Figure 4) does not match to the species phylogeny (Figure 1) derived from seven genes. The *tra2* sequences of *Apis* cluster in a separate branch from phylogenetic closely related groups and evolve with higher evolutionary rate. Reflected by the high number of *Apis*-specific amino acid changes we suggest a modified function of *tra2* compared to non-*Apis*. Changes in the amino acid composition on 21 sites, 9 of them inside the RRM-domain, led us to conclude that target molecule specificities in binding sites may have been modified. These target molecules could be *fem* and/or *dsx*. Our evolutionary analysis of Dsx protein indicates a rather high degree of structural conservation (Figure 5B). Consequently, and in agreement with the widely accepted hypothesis of bottom-up evolution in sex determining pathways (Wilkins, 1995), we have reasons to assume that *dsx* has retained its conserved function and that the structural changes in Tra2 were driven by *fem* evolution.

Coevolutionary Model of *tra2* and *Fem*/Paralog Complex in *Apis* and Non-*Apis*

The evolution of novel or modified gene function may affect the function of associated genes (Innan and Kondrashov, 2010), a characteristic that we have noticed already in the evolution of *fem* in *Apis* in a previous study (Hasselman et al., 2010). In that study we found that *fem* in *Apis* evolves under stronger functional constraints than in non-*Apis*, likely due to the origin of the novel function raised by *csd*. Often known as coevolution, molecular changes among closely interacting genes may lead to lineage-specific modification of protein function. The concept of gene-for-gene evolution has been introduced and widely described in plant-pathogen interactions, with natural selection and genetic drift as the major evolutionary processes driving this form of coevolution (Thompson and Burdon, 1992; Dodds et al., 2006). Our present results led us to propose a model of coevolutionary changes in sex determining key regulator *tra2* and *fem* with its paralogs, depending on their presence or absence.

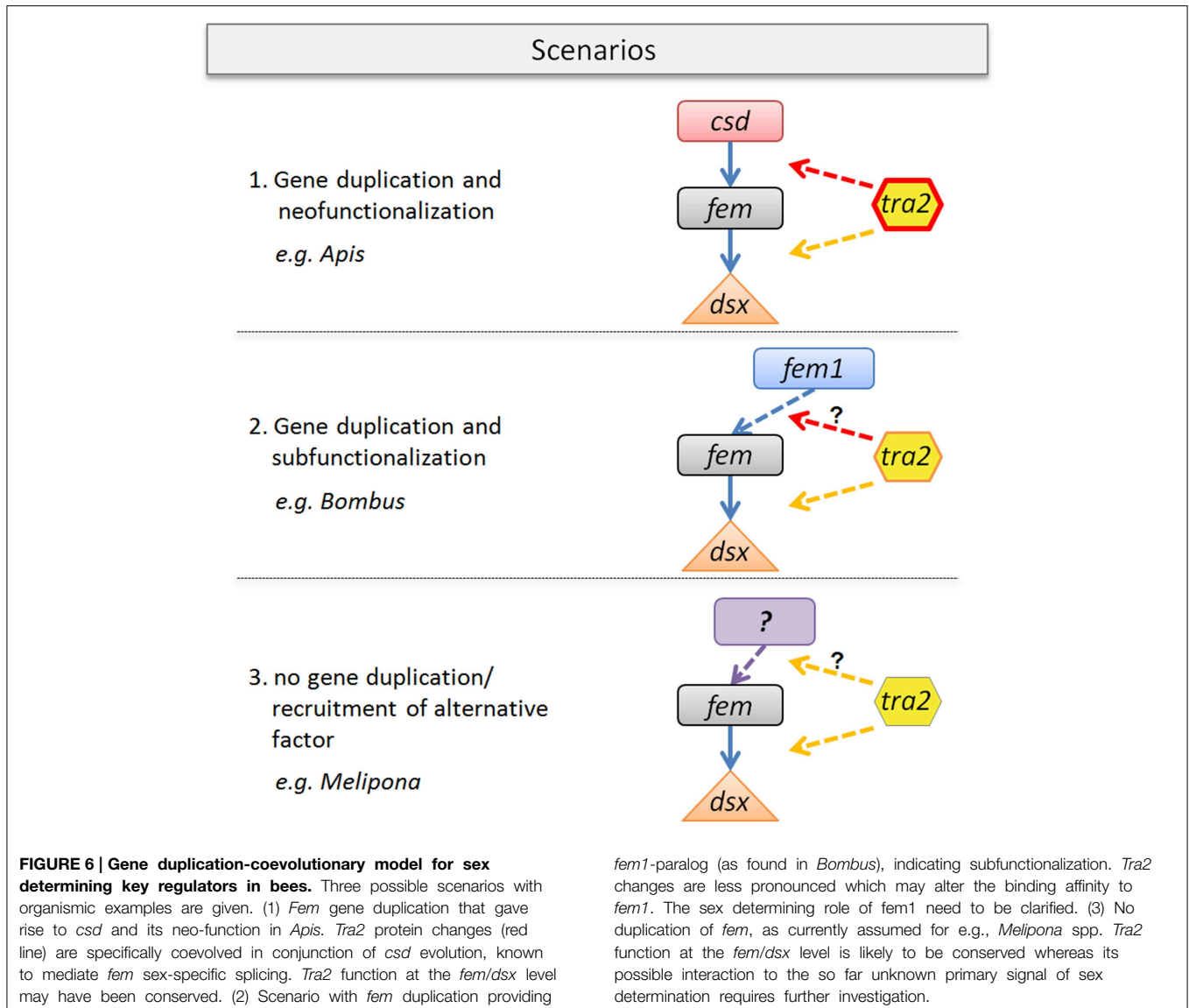
We propose three scenarios that may impact the evolution of the *tra2/fem*/paralog gene complex. Scenario one resembles the best studied case so far (Figure 6) found to be established in the *Apis* lineage. In this scenario, the evolution of the multiallelic *csd* operating as primary signal of sex determination following the model of neofunctionalization was accompanied by lineage specific changes of the Tra2 protein. Tra2 has been proven to mediate *fem* mRNA sex-specific splicing, transmitting the information of the allelic composition at *csd* to its downstream target (Nissen et al., 2012). Consequently, our data of numerous *Apis*-specific

amino acid substitutions (Figure 4) within and outside of the RRM domain indicates a coevolutionary, fast-evolving process forced by the strong directional evolution that has acted on *csd* (Hasselman et al., 2008). In addition, Tra2 has been proposed to interact with the genes *fem* and *dsx* to act on regulating sex-specific splicing of *dsx* (Nissen et al., 2012). To disentangle which of the amino acid changes are directly associated to this twofold functions of *tra2*, future *in vitro* studies are needed.

The second scenario illustrates duplication events of *fem*, as found in e.g., *Bombus*, giving rise to the paralogous copy *fem1* (Sadd et al., in press). The proposed model of subfunctionalization (Figure 6) is supported by the absence of allelic variation in *fem1* which is in contrast to the *fem* paralog in *Apis* (*csd*) (Biewer et al., in revision). Another difference between *csd* and *fem1* is the occurrence of various splice transcripts in the latter and their absence in *csd* (Gempe et al., 2009). We hypothesize that the numerous amino acid differences in Tra2 are associated with its modified binding specificity in *Bombus* (dotted arrow), driven by a different evolutionary fate of the *fem* paralog. Still, it remains up to further investigation to identify the primary signal of sex determination in *Bombus* and the position of *fem1* within the sex determination pathway.

Our last scenario three (Figure 6) is stimulated by the observation that in some bee species (e.g., *Melipona*) obviously no *fem* duplication exist. This result is not only supported by bioinformatic approaches on newly sequenced genome data (this study and Kapheim et al., in revision) but also by various experimental setups (Brito et al., unpublished). In this scenario *tra2* function is likely to be related to the sex determination pathway based on its evolutionary conservation (this study) and on its constant expression over early (egg) and late (larvae, pupae, adults) developmental stages in *Melipona interrupta* (Schlesinger and Hasselman, unpublished data). Gene expression studies can add another useful dimension to examine coevolution among genes as interacting proteins are often precisely coexpressed, (Fraser et al., 2004), ultimately leading to a better understanding of protein interaction processes within regulatory pathways. Further, analyses will likely elucidate the primary signal of sex determination in *Melipona*, a system on which various alternative models to explain the determination of different sexes have been developed in the past, including empirical evidence for a complementary mode of sex determination resulting from controlled crossing experiments (Kerr, 1987; Carvalho, 2001).

Our comparative analyses of major regulators of sex determination in hymenopteran species provide further support to the wide range of evolutionary possibilities for shaping the sex determination regulatory pathway, consistent with the concept of DSD. Driving forces affecting the evolutionary dynamic of sex determining key regulators are gene duplication, selection and coevolution. More instantaneously occurring events such as transposon mediated translocation of genes or fragments and recombination events may lead to gene copy number variations, including pseudogenization (Lonnig and Saedler, 2002). These processes are likely to be common in hymenopteran species, as high recombination frequencies in bees and ants (Beye et al., 2006; Sirvio et al., 2006; Meznar et al., 2010) and transposable elements near sex determining genes (Beye et al., 2003;



Koch et al., 2014) have been observed. For the hymenopteran wasp species *Nasonia vitripennis* a non-complementary sex determining system has been recently proposed, based on maternal effected genomic imprinting (van de Zande and Verhulst, 2014). To ensure male development in unfertilized eggs, a *womanizer* factor, which is maternally silenced during oogenesis and affects *tra* expression, has been described, opening the road to study the probably highly divergent alternative mechanism that has evolved in course of wasp and bee divergence. The challenge for future studies on species with newly sequenced genomes will be to test evolutionary predictions raised by bioinformatic analyses using functional experiments.

Acknowledgments

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) to MH. We thank two reviewers for their helpful comments and Paul D'Alvise for carefully reading a previous version of the manuscript.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2015.00124/abstract>

References

- Abouheif, E., and Wray, G. A. (2002). Evolution of the gene network underlying wing polyphenism in ants. *Science* 297, 249–252. doi: 10.1126/science.1071468
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, 202–208. doi: 10.1093/nar/gkp335
- Baker, B. S. (1989). Sex in flies: the splice of life. *Nature* 340, 521–524. doi: 10.1038/340521a0
- Beukeboom, L. W., Kamping, A., Louter, M., Pijnacker, L. P., Katju, V., Ferree, P. M., et al. (2007). Haploid females in the parasitic wasp *Nasonia vitripennis*. *Science* 315, 206. doi: 10.1126/science.1133388
- Beye, M., Gattermeier, I., Hasselmann, M., Gempe, T., Schioett, M., Baines, J. F., et al. (2006). Exceptionally high levels of recombination across the honey bee genome. *Genome Res.* 16, 1339–1344. doi: 10.1101/gr.5680406
- Beye, M., Hasselmann, M., Fondrk, M. K., Page, R. E., and Omholt, S. W. (2003). The Gene *csd* is the primary signal for sexual development in the honey bee and encodes a SR-type protein. *Cell* 114, 419–429. doi: 10.1016/S0092-8674(03)00606-8
- Bopp, D., Saccone, G., and Beye, M. (2014). Sex determination in insects: variations on a common theme. *Sex. Dev.* 8, 20–28. doi: 10.1159/000356458
- Bull, J. J. (1983). *Evolution of Sex Determining Mechanisms*. Menlo Park, CA: Cummings Publishing Company.
- Burtis, K. C., and Baker, B. S. (1989). *Drosophila* doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* 56, 997–1010. doi: 10.1016/0092-8674(89)90633-8
- Butler, B., Pirrotta, V., Irminger-Finger, I., and Nöthinger, R. (1986). The sex-determining gene *tra* of *Drosophila*: molecular cloning and transformation studies. *EMBO J.* 5, 3607–3613.
- Cardinal, S., and Danforth, B. N. (2013). Bees diversified in the age of eudicots. *Proc. Biol. Sci.* 280:20122686. doi: 10.1098/rspb.2012.2686
- Carvalho, G. A. (2001). The number of sex alleles (CSD) in a Bee population and its practical importance (Hymenoptera: Apidae). *J. Hym. Res.* 10, 10–15. Available online at: <http://biostor.org/reference/293>
- Chandler, D., McGuffin, M. E., Piskur, J., Yao, J., Baker, B. S., and Mattox, W. (1997). Evolutionary conservation of regulatory strategies for the sex determination factor transformer-2. *Mol. Cell. Biol.* 17, 2908–2919.
- Cho, S., Huang, Z. Y., and Zhang, J. (2007). Sex-specific splicing of the honeybee doublesex gene reveals 300 million years of evolution at the bottom of the insect sex-determination pathway. *Genetics* 177, 1733–1741. doi: 10.1534/genetics.107.078980
- Cline, T. W., and Meyer, B. J. (1996). Vive la difference: males vs females in flies vs worms. *Annu. Rev. Genet.* 30, 637–702. doi: 10.1146/annurev.genet.30.1.637
- Davidson, E. H. (2010). Emerging properties of animal gene regulatory networks. *Nature* 468, 911–920. doi: 10.1038/nature09645
- Dodds, P. N., Lawrence, G. J., Catanzariti, A. M., Catanzariti, A. M., Teh, T., Wang, C. A., et al. (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8888–8893. doi: 10.1073/pnas.0602577103
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* 14, 755–763. doi: 10.1093/bioinformatics/14.9.755
- Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 23, 205–211. doi: 10.1142/9781848165632_0019
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Fairbrother, W. G., Yeh, R. F., Sharp, P. A., and Burge, C. B. (2002). Predictive identification of exonic splicing enhancers in human genes. *Science* 297, 1007–1013. doi: 10.1126/science.1073774
- Fani, R., and Fondi, M. (2009). Origin and evolution of metabolic pathways. *Phys. Life Rev.* 6: 23–52. doi: 10.1016/j.plrev.2008.12.003
- Force, A., Cresko, W. A., Pickett, F. B., Proulx, S. R., Amemiya, C., and Lynch, M. (2005). The origin of subfunctions and modular gene regulation. *Genetics* 170, 433–446. doi: 10.1534/genetics.104.027607
- Fraser, H. B., Hirsh, A. E., Wall, D. P., and Eisen, M. B. (2004). Coevolution of gene expression among interacting proteins. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9033–9038. doi: 10.1073/pnas.0402591101
- Gadau, J., Helmkamp, M., Nygaard, S., Roux, J., Simola, D. F., Smith, C. R., et al. (2012). The genomic impact of 100 million years of social evolution in seven ant species. *Trends Genet.* 28, 14–21. doi: 10.1016/j.tig.2011.08.005
- Gempe, T., Hasselmann, M., Schioett, M., Hause, G., Otte, M., and Beye, M. (2009). Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway. *PLoS Biol.* 7:e1000222. doi: 10.1371/journal.pbio.1000222
- Grimaldi, D., and Engel, M. S. (2005). *Evolution of the Insects*. Cambridge: Cambridge University Press.
- Hasselmann, M., and Beye, M. (2004). Signatures of selection among sex-determining alleles of the honey bee. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4888–4893. doi: 10.1073/pnas.0307147101
- Hasselmann, M., Gempe, T., Schioett, M., Nunes-Silva, C. G., Otte, M., and Beye, M. (2008). Evidence for the evolutionary nascent of a novel sex determination pathway in honeybees. *Nature* 454, 519–522. doi: 10.1038/nature07052
- Hasselmann, M., Lechner, S., Schulte, C., and Beye, M. (2010). Origin of a function by tandem gene duplication limits the evolutionary capability of its sister copy. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13378–13383. doi: 10.1073/pnas.1005617107
- Hediger, M., Henggeler, C., Meier, N., Perez, R., Saccone, G., and Bopp, D. (2010). Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. *Genetics* 184, 155–170. doi: 10.1534/genetics.109.109249
- Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., et al. (2012). InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* 40, D306–D312. doi: 10.1093/nar/gks456
- Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108. doi: 10.1038/nrg2689
- Kerr, W. E. (1987). Sex determination in bees. XXI. Number of XO-heteroalleles in a natural population of *Melipona compressipes fasciculata* (Apidae). *Insectes Soc.* 34, 274–279.
- Koch, V., Nissen, I., Schmitt, B. D., and Beye, M. (2014). Independent evolutionary origin of fem paralogous genes and complementary sex determination in hymenopteran insects. *PLoS ONE* 9:e91883. doi: 10.1371/journal.pone.0091883
- Lechner, S., Ferretti, L., Schoning, C., Kinuthia, W., Willemsen, D., and Hasselmann, M. (2014). Nucleotide variability at its limit? insights into the number and evolutionary dynamics of the sex-determining specificities of the honey bee *apis mellifera*. *Mol. Biol. Evol.* 31, 272–287. doi: 10.1093/molbev/mst207
- Lonnig, W. E., and Saedler, H. (2002). Chromosome rearrangements and transposable elements. *Annu. Rev. Genet.* 36, 389–410. doi: 10.1146/annurev.genet.36.040202.092802
- Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39, D225–D229. doi: 10.1093/nar/gkq1189
- Meznar, E. R., Gadau, J., Koeniger, N., and Rueppell, O. (2010). Comparative linkage mapping suggests a high recombination rate in all honeybees. *J. Hered.* 101(Suppl. 1), S118–S126. doi: 10.1093/jhered/esq002
- Nahmad, M., Glass, L., and Abouheif, E. (2008). The dynamics of developmental system drift in the gene network underlying wing polyphenism in ants: a mathematical model. *Evol. Dev.* 10, 360–374. doi: 10.1111/j.1525-142X.2008.00244.x
- Nissen, I., Muller, M., and Beye, M. (2012). The Am-*tra2* gene is an essential regulator of female splice regulation at two levels of the sex determination hierarchy of the honeybee. *Genetics* 192, 1015–1026. doi: 10.1534/genetics.112.143925
- Pane, A., Salvemini, M., Delli, B. P., Polito, C., and Saccone, G. (2002). The transformer gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. *Development* 129, 3715–3725. Available online at: <http://dev.biologists.org/content/129/15/3715.long>
- Peter, I. S., and Davidson, E. H. (2011). Evolution of gene regulatory networks controlling body plan development. *Cell* 144, 970–985. doi: 10.1016/j.cell.2011.02.017
- Pires-da Silva, A., and Sommer, R. J. (2003). The evolution of signalling pathways in animal development. *Nat. Rev. Genet.* 4, 39–49. doi: 10.1038/nrg977
- Posada, D., and Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818. doi: 10.1093/bioinformatics/14.9.817
- Privman, E., Wurm, Y., and Keller, L. (2013). Duplication and concerted evolution in a master sex determiner under balancing selection. *Proc. Biol. Sci.* 280, 20122968. doi: 10.1098/rspb.2012.2968

- Ruiz, M. F., Eirin-Lopez, J. M., Stefani, R. N., Perondini, A. L., Selivon, D., and Sánchez, L. (2007). The gene doublesex of *Anastrepha* fruit flies (Diptera, Tephritidae) and its evolution in insects. *Dev. Genes Evol.* 217, 725–731. doi: 10.1007/s00427-007-0178-8
- Sadd, B., Barribeau, S. M., Bloch, G., Bourke, A. F., Colling, D., Dearden, P. K., et al. (in press). The genomes of two key bumblebee species with primitive eusocial organisation. *Genome Biol.*
- Salvemini, M., Robertson, M., Aronson, B., Atkinson, P., Polito, L. C., and Saccone, G. (2009). *Ceratitis capitata* transformer-2 gene is required to establish and maintain the autoregulation of Cctra, the master gene for female sex determination. *Int. J. Dev. Biol.* 53, 109–120. doi: 10.1387/ijdb.082681ms
- Sarno, F., Ruiz, M. F., Eirin-Lopez, J. M., Perondini, A. L., Selivon, D., and Sánchez, L. (2010). The gene transformer-2 of *Anastrepha* fruit flies (Diptera, Tephritidae) and its evolution in insects. *BMC Evol. Biol.* 10:140. doi: 10.1186/1471-2148-10-140
- Schmieder, S., Colinet, D., and Poirie, M. (2012). Tracing back the nascence of a new sex-determination pathway to the ancestor of bees and ants. *Nat. Commun.* 3, 895. doi: 10.1038/ncomms1898
- Sirvio, A., Gadau, J., Rueppell, O., Lamatsch, D., Boomsma, J. J., Pamilo, P., et al. (2006). High recombination frequency creates genotypic diversity in colonies of the leaf-cutting ant *Acromyrmex echinatior*. *J. Evol. Biol.* 19, 1475–1485. doi: 10.1111/j.1420-9101.2006.01131.x
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Thompson, J. N., and Burdon, J. J. (1992). Gene-for-gene coevolution between plants and parasites. *Nature* 360, 121–125. doi: 10.1038/360121a0
- True, J. R., and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* 3, 109–119. doi: 10.1046/j.1525-142x.2001.003002109.x
- van de Zande, L., and Verhulst, E. C. (2014). Genomic imprinting and maternal effect genes in haplodiploid sex determination. *Sex. Dev.* 8, 74–82. doi: 10.1159/000357146
- Verhulst, E. C., Beukeboom, L. W., and van de Zande, L. (2010). Maternal control of haplodiploid sex determination in the wasp *Nasonia*. *Science* 328, 620–623. doi: 10.1126/science.1185805
- Verhulst, E. C., Lynch, J. A., Bopp, D., Beukeboom, L. W., and van de Zande, L. (2013). A new component of the *Nasonia* sex determining cascade is maternally silenced and regulates transformer expression. *PLoS ONE* 8:e63618. doi: 10.1371/journal.pone.0063618
- Werren, J. H., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., Colbourne, J. K., et al. (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327, 343–348. doi: 10.1126/science.1178028
- Wilkins, A. S. (1995). Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays* 17, 71–77. doi: 10.1002/bies.950170113
- Wilkins, A. S. (2007). Between “design” and “bricolage”: genetic networks, levels of selection, and adaptive evolution. *Proc. Natl. Acad. Sci. U.S.A.* 104(Suppl. 1), 8590–8596. doi: 10.1073/pnas.0701044104

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Biewer, Schlesinger and Hasselmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Metatranscriptomic analyses of honey bee colonies

Cansu Ö. Tozkar^{1*}, Meral Kence¹, Aykut Kence^{1†}, Qiang Huang² and Jay D. Evans^{2*}

¹ Ecological Genetics Laboratory, Department of Biological Sciences, Middle East Technical University, Ankara, Turkey, ² Bee Research Laboratory, United States Department of Agriculture-Agricultural Research Service, Beltsville, MD, USA

OPEN ACCESS

Edited by:

Juergen Rudolf Gadau,
Arizona State University, USA

Reviewed by:

Marcial Escudero,
Doñana Biological Station-Consejo
Superior de Investigaciones
Científicas, Spain
Robert Brucker,
Harvard University, USA

*Correspondence:

Cansu Ö. Tozkar,
Ecological Genetics Laboratory,
Department of Biological Sciences,
Middle East Technical University,
Dumlupınar Bulvarı No: 1, Cankaya,
Ankara 06800, Turkey
tozkar@metu.edu.tr;
Jay D. Evans,
Bee Research Laboratory,
USDA-ARS, 10300 Baltimore Avenue
Bldg. 306 BARC-E, Beltsville,
MD 20705-0000, USA
jay.evans@ars.usda.gov
† Deceased.

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal *Frontiers in
Genetics*

Received: 16 November 2014

Accepted: 25 February 2015

Published: 19 March 2015

Citation:

Tozkar CÖ, Kence M, Kence A,
Huang Q and Evans JD (2015)
Metatranscriptomic analyses of honey
bee colonies. *Front. Genet.* 6:100.
doi: 10.3389/fgene.2015.00100

Honey bees face numerous biotic threats from viruses to bacteria, fungi, protists, and mites. Here we describe a thorough analysis of microbes harbored by worker honey bees collected from field colonies in geographically distinct regions of Turkey. Turkey is one of the World's most important centers of apiculture, harboring five subspecies of *Apis mellifera* L., approximately 20% of the honey bee subspecies in the world. We use deep ILLUMINA-based RNA sequencing to capture RNA species for the honey bee and a sampling of all non-endogenous species carried by bees. After trimming and mapping these reads to the honey bee genome, approximately 10% of the sequences (9–10 million reads per library) remained. These were then mapped to a curated set of public sequences containing ca. Sixty megabase-pairs of sequence representing known microbial species associated with honey bees. Levels of key honey bee pathogens were confirmed using quantitative PCR screens. We contrast microbial matches across different sites in Turkey, showing new country recordings of Lake Sinai virus, two *Spiroplasma* bacterium species, symbionts *Candidatus Schmidhempelia bombi*, *Frischella perrara*, *Snodgrassella alvi*, *Gilliamella apicola*, *Lactobacillus* spp.), neogregarines, and a trypanosome species. By using metagenomic analysis, this study also reveals deep molecular evidence for the presence of bacterial pathogens (*Melissococcus plutonius*, *Paenibacillus larvae*), *Varroa destructor*-1 virus, Sacbrood virus, and fungi. Despite this effort we did not detect KBV, SBPV, Tobacco ringspot virus, VdMLV (*Varroa* Macula like virus), *Acarapis* spp., *Tropilaelaps* spp. and *Apocephalus* (phorid fly). We discuss possible impacts of management practices and honey bee subspecies on microbial retinues. The described workflow and curated microbial database will be generally useful for microbial surveys of healthy and declining honey bees.

Keywords: *Apis mellifera*, pollination, colony collapse disorder, RNA sequencing, bioinformatics, honey bee viruses, trypanosomes

Introduction

The honey bee (*Apis mellifera* L.) has ecological importance as a natural pollinator of wild flora and crops. Moreover, managed honey bees have economical importance with hive products including honey, pollen, wax, propolis, and royal jelly (Maheshwari, 2003). Recently, declines of managed colonies have been noted on many continents. Several causes of these large-scale losses have been reported, including honey bee parasites (*Varroa destructor*, *Acarapis woodi*);

pathogens (*Nosema* spp. and bee viruses); pesticides, contaminated water, use of antibiotics, poor nutrition, and migratory beekeeping practices (Kevan et al., 2007; Higes et al., 2008; Naug, 2009; vanEngelsdorp et al., 2009; Bacandritsos et al., 2010; vanEngelsdorp and Meixner, 2010). The high density of individuals and the exchange of food among *A. mellifera* colony members create a favorable environment for bacterial, viral, fungal, and protist pathogens and several studies have noted an increase in diversity and infection rates of pathogens in failing bee colonies. Here we will focus on honey bee pathogens and parasites and the use of modern sequencing techniques to identify these agents in healthy and declining colonies.

Among the honey bee pathogens, viruses are of special concern. Viruses are widespread in honey bees although most often without noticeable symptoms (Ball and Bailey, 1997). Multiple viral infections have been diagnosed in many bee colonies (Chen et al., 2004). At least 18 different viruses exist in honey bees (Bailey and Ball, 1991) with six of them; Sacbrood virus (SBV), Deformed wing virus (DWV), Acute bee paralysis virus (ABPV), Black queen cell virus (BQCV), Chronic bee paralysis virus (CBPV), and Kashmir bee virus (KBV) most commonly linked to bee disease. These viruses have strong impacts on managed bee populations, pollination services, and honey production on several continents (Allen and Ball, 1996; Nordstrom et al., 1999; Ellis and Munn, 2005; De Miranda et al., 2010). DWV and ABPV have been linked to parasitic mite loads, while Chronic bee paralysis virus (CBPV) and the related Lake Sinai viruses are also widespread and tied to significant losses in honey bee colonies (Runckel et al., 2011; Ravoet et al., 2013). Sacbrood virus is the only common virus of developing bees, and this virus is not generally implicated in adult bee mortality or morbidity (Anderson and Gibbs, 1989).

Two species of Microsporidia, *Nosema apis* and *Nosema ceranae*, are widespread parasites of adult honey bees. *N. apis* is a long-standing infection agent of the European honey bee, *A. mellifera*, (Zander, 1909) that causes nosemosis, a disease with mild virulence. *N. ceranae* was first found as a parasite of the Asian honey bee, *Apis cerana* (Fries et al., 1996) although this species is now widespread throughout the range of *A. mellifera* (Fries et al., 1996, 2006; Paxton et al., 2007) arguably thanks to worldwide trade in bees (Klee et al., 2007) and perhaps pollen supplements. Honey bees can be co-infected with both *Nosema* species (Fries, 2010). Additionally, *N. ceranae* seems to be replacing *N. apis* worldwide (Klee et al., 2007). In association with *Nosema* infections, several viruses can significantly affect the apparent virulence of *Nosema* (Bailey and Ball, 1991).

Trypanosomes are a varied group of parasites that infect insects (Merzlyak et al., 2001). *Crithidia bombi*, a trypanosome that infects bumble bees, has effects on behavior (Gegear et al., 2005) and longevity during stressful conditions (Brown et al., 2003). Trypanosomatid parasites parasitize the mid- and hind-guts of their hosts (Lange and Lord, 2012) and can be widespread (Langridge and McGhee, 1967; Schmid-Hempel and Tognazzo, 2010). The current role of trypanosomes in honey bee health is not clear (Schwarz and Evans, 2013), although they have been recognized as possible correlates

with bee declines in two field surveys (Runckel et al., 2011; Ravoet et al., 2013). The most common honey bee trypanosomatid currently is distinct from the species *Crithidia mellificae* described by Langridge and McGhee (1967), and has recently been named as *Lotmaria passim* (Schwarz et al., 2015).

Spiroplasmas are particularly virulent pathogens that are found in various environments and implicated as pathogens of plants, vertebrates and insects. They have a seasonal occurrence associated with the nectarines and surfaces of flowers (Markham and Townsend, 1981; Williamson et al., 1989). Adult honey bees are parasitized by two species of bacteria, *Spiroplasma apis* (Mouches et al., 1983) and *Spiroplasma melliferum* (Clark et al., 1985). Upon invading the hemolymph, these bacteria can cause a fatal disease called spiroplasmosis or May disease. Gut microbiota of animals living in social communities may influence their health with their functions related to nutrition, immune responses and resistance against pathogens (Dillon and Dillon, 2004; Round and Mazmanian, 2009). Surveys of 16S rDNA sequences from the honey bee indicate the presence of eight predominant species (defined as strains sharing >97% 16S rRNA identity) which account for 95% of the resident bacteria (Moran et al., 2012). These species include the beta-proteobacterium *Snodgrassella alvi* (family *Neisseriaceae*) and the gamma-proteobacterium *Gilliamella apicola* (family *Orbaceae*), the dominant Gram-negative members of the gut community, with each comprising up to 30–39% of the microbiota (Moran et al., 2012).

The surveillance and discovery of novel pathogens via high-throughput sequencing can provide a relatively unbiased view of pathogens and microbes associated with insects and other arthropods (Bishop-Lilly et al., 2010; Ma et al., 2011; Vayssier-Taussat et al., 2013). Recent efforts based on these technologies have uncovered novel and unexpected taxa associated with honey bees (e.g., Cox-Foster et al., 2007; Runckel et al., 2011; Cornman et al., 2013) and have provided estimates of normal microbial levels vs. those of diseased colonies. Turkey is one of the World's most important centers of apiculture, with managed and wild populations of five subspecies of *Apis mellifera* L. including *A. m. caucasica*, *A. m. syriaca*, *A. m. anatoliaca*, *A. m. meda* and an ecotype in Thrace belonging to the carnica subspecies group which is distinctly different from the subspecies found in Anatolia (**Figure 1**) (Kandemir et al., 2000; Bodur et al., 2007). These subspecies cover approximately 20% of honey bee subspecies in the world. Our purpose here was to determine the regional prevalence of bacterial pathogens, viruses, fungi, parasites, protists, and symbionts and to compare their loads in areas where migratory and stationary beekeeping is practiced in Turkey. Monitoring viruses and other disease agents can help solve problems related to the health of stationary and migratory honey bee colonies and limit or avoid their spread. Although commercial migratory beekeeping practices are necessary for pollination and crop production, their effects on honey bee colony health and pathogen transmission should be addressed. We used transcriptome analyses to survey a wide range of pathogens and to detect unexpected or rare taxa. RNA-seq technology allows the precise detection of rare transcripts by mapping reads against inclusive sequence databases,



FIGURE 1 | The provinces and the distribution of honey bee subspecies in this study.

reducing the repetitive effort and possible biases of conventional molecular diagnostics (Minoche et al., 2011). Identification and confirmation of select virus and pathogen loads was confirmed by quantitative RT-PCR. Our survey was aligned with the European Honey bee colony loss network (www.COLOSS.org) survey, and was conducted to determine the colony losses in many regions representing honey bee diversity in Turkey. By using metagenomic analysis, our results provide new incidence records of the virome and microbiome in search of etiologically unexpected or previously unknown agents among 10 distinct provinces in Turkey and suggest a higher viral prevalence, and increased losses, in migratory beekeeping operations. *Apis mellifera* is an economically and ecologically important model organism and identification of pathogens and other microbes can have extensive implications for current practices in apiculture and agriculture.

Materials and Methods

Sampling

Colony loss surveys were carried out in eight regions and 158 beekeeping operations in 2010. Among those, 98 were migratory beekeepers and 60 were stationary. In 2011, surveys of 221 beekeepers from seven different regions were evaluated for this study. Samples analyzed for microbial loads comprised adult honey bees collected from field colonies in Turkey during summer and fall of 2010 and 2011. Efforts were made to collect from different regions of Turkey, and from a diverse set of beekeepers (**Figure 1**). Adult bees were collected from 134 colonies from different regions of Turkey from 38, 48, and 51 beekeepers in 2010-Fall, 2010-Spring, and 2011 respectfully and the samples were kept frozen until molecular diagnosis.

RNA Isolation

RNA was extracted from a pooled sample of 50 bees from each sampled colony, using an acid-phenol RNA extraction method (Evans et al., 2013).

cDNA Synthesis and Real Time qPCR

RNA extracts were used to generate first and second-strand cDNAs using random hexamer primers and the reverse transcriptase Superscript II® (Invitrogen™), as described in vanEngelsdorp et al. (2009). Pathogen loads were estimated using real-time quantitative-PCR (qPCR) and a Bio-Rad CFX-96™ thermocycler. Complementary DNA (cDNA) was generated from 1 µg RNA template and was amplified in a separate 20 µl final reaction volume of Sso-Fast™ SYBR® Green reaction mix (Bio-Rad™) for each diagnostic primer pair. We used published primers to survey for SBV, KBV, IAPV, DWV, ABPV, BQCV, trypanosomes (vanEngelsdorp et al., 2009), AFB (Evans et al., 2006), *Nosema ceranae* (Fries et al., 2013) and *Nosema apis* (Schwarz and Evans, 2013), *Spiroplasma apis* and *Spiroplasma melliferum* (Schwarz et al., 2014). Honey bee ribosomal protein S5 (RPS5) was used to normalize for cDNA content and to filter samples for degradation or experimental losses.

We used a thermal profile of 95°C for 30 s followed by 95°C for 5 s and 60°C for 30 s. Steps two and three were repeated for a total of 50 cycles and included plate reads for fluorescence during each 60°C step. Following the cycle program, products were denatured for 10 s at 95°C., reannealed and then a dissociation profile was measured between 69 and 95°C at an increment of 0.5°C to provide evidence for reaction fidelity (Evans et al., 2006). Annealing temperature of 55°C was used for the CytbSF (trypanosomatid *L. passim*) primers. Positive and negative control reactions were run on each 96-well plate. Pathogen loads ($\Delta\Delta CT$) were determined as the difference between the CT of RPS5 and the CT of each target (ΔCT), scaled up from the minimal ΔCT across all samples.

PCR Purification and Sequencing

RT-PCR products were selected for sequencing to confirm the identities of products indicating the trypanosome, *L. passim*, and the bacteria *S. apis* and *S. melliferum*. These PCR products were purified using QIAquick PCR purification kit according to protocol recommended by manufacturer, and then were sequenced commercially by Macrogen (Rockville, MD, USA) DNA sequence

similarity with trypanosomes, *S. apis* and *S. melliferum*, was confirmed using the BLAST search tool (Altschul et al., 1990) from the U.S. National Institutes of Health and searches against the National Center for Biotechnology Information (NCBI) *nr* database.

High Throughput Sequencing and Data Analysis

High throughput sequencing was performed with the pooled samples from stationary colonies of each region and all RNA samples were pooled before sequencing ($n = 6$ ILLUMINA RNA libraries). Libraries were run on paired-end 100-cycle reactions and flow cells using Illumina Hi-Seq 2000 machines at the University of Maryland Institute of Genome Sciences. Sequences for all six libraries have been deposited at the US National Institutes of Health NCBI “Honey Bee Disease Database” Bioproject (PRJNA52851).

After trimming and quality control of the generated sequences, transcriptomic analysis was done with the support of CLC Genomics Workbench 7.0.3 (CLC Bio, Aarhus, Denmark) by mapping sequencing reads and counting and distributing the reads across genes and transcripts based on annotated reference genes (Figure 2). During the alignment, two mismatches were allowed with deletion cost of three, insertion cost of three, length fraction of 0.8, similarity fraction of 0.8 and maximum 10 hits for a read. For statistical analyses, proportions-based tests were used for the comparison of the counts by considering the proportions that they comprise of the total sum of counts in each sample. Multi-group comparison was done by weighted *t*-type test statistic Baggerly's test (Baggerly et al., 2003). FDR (False Discovery Rate) corrected *p*-values were calculated according to the methods of Benjamini and Hochberg (1995) to determine the statistical significance of the pathogen load.

Real time q-PCR data statistics were performed by using JMPTM (SAS Institute, Cary, NC, USA, v.9). A One-Way analysis of variance was used to test differences between group means. The total variability in the response was divided into two parts; within-group variability and between-group variability. The differences between the group means were considered to be significant if the between-group variability was broader relative to the within-group variability. Multiple comparisons of group means were done by using pooled variance estimates for these means. Student's *t*-tests were computed for each pair of group levels and individual pairwise comparisons. Matrix of correlation coefficients that summarized the strength of the linear relationships between each pair of response variables was calculated and Pearson product-moment correlations for each pair of variables were listed. Correlations and the significance probabilities were calculated by the pairwise deletion method and the count values differed if any pair had a missing value for either variable.

Results

RT-qPCR Results

Kashmir bee virus (KBV), Israeli acute paralysis virus (IAPV), American foulbrood (AFB) and Sacbrood virus (SBV) were not detected by PCR in any of the samples, nor was the microsporidian pathogen *N. apis* (except for only three colonies in 2011)

by RT-PCR. Present in these samples were Acute bee paralysis virus (ABPV), Deformed wing virus (DWV), Black queen cell virus (BQCV), and *N. ceranae*. Mixed infections of ABPV, DWV, BQCV, and *N. ceranae* were detected (Tables 1, 2).

The distribution of bee pathogens was significantly different among provinces and with beekeeping practices in 2010 (Table 3 in Supplementary Material). Generally, DWV loads were higher in Bitlis, Hatay, Muğla, and Ardahan than in other regions. ABPV was the most common virus in Bitlis and was especially high in the samples of migratory beekeepers. Among provinces that were sampled, BQCV loads were lowest in Edirne and highest in colonies of migratory beekeepers sampled in Ardahan, Hatay, and Muğla (Table 2). In 2011, DWV, BQCV, and *N. ceranae* loads were significantly different among the regions (Table 3 in Supplementary Material). DWV loads were higher in Muğla, Hatay, and Yığılca. BQCV levels were high in Ardahan and Artvin and was not detected at all in Yığılca. *N. ceranae* was more frequent among Artvin, Ardahan, and Kırklareli but not detected among stationary colonies of Hatay (Table 2). ABPV occurrence was the highest in samples from stationary apiaries of Yığılca and Ardahan and samples from migratory ones of Hatay and Muğla. The lowest level was in the stationary colonies of Hatay. ABPV loads showed difference between Yığılca-Kırklareli and Yığılca-Muğla ($p = 0.0398$ and $p = 0.0478$).

ABPV loads in Muğla and Ankara, BQCV loads in Ardahan, Muğla, and *N. ceranae* loads in Hatay were significantly higher in the samples of migratory beekeepers than the samples of stationary beekeepers in 2010. Among 2011 samples, the pathogen loads of migratory colonies were also higher in Muğla and Hatay but the results were not significant (Table 3 in Supplementary Material).

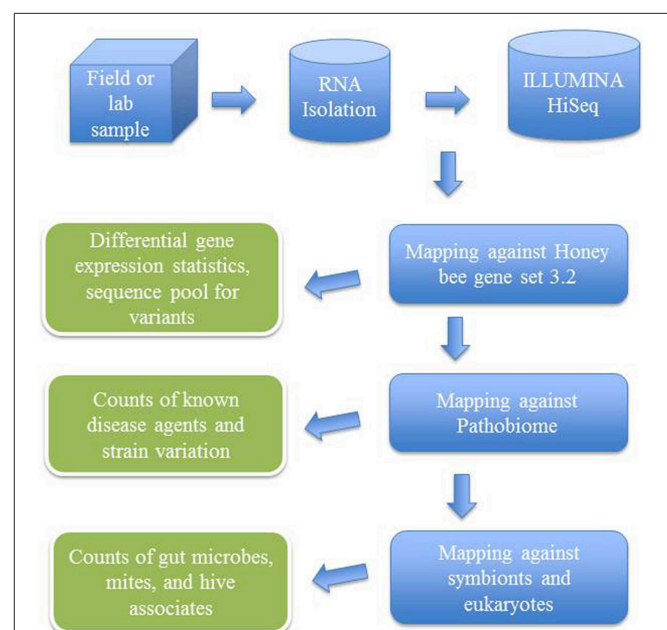


FIGURE 2 | Flow chart showing analytical steps for mapping ILLUMINA RNA-seq reads against an annotated bee microbial dataset.

TABLE 1 | The results of pathogen analysis of honey bee samples collected from different regions of Turkey.

Sampling date	Locations	Pathogens
2010–2011	Muğla	ABPV, BQCV, DWV, <i>Nosema ceranae</i> , Trypanosomes
2010–2011	Hatay	ABPV, BQCV, DWV, <i>Nosema ceranae</i> , Trypanosomes
2010	Ankara	ABPV, BQCV, DWV, <i>Nosema ceranae</i> , Trypanosomes
2011	Yığılca	ABPV, BQCV, DWV, <i>Nosema ceranae</i> , Trypanosomes
2010–2011	Kırklareli	ABPV*, BQCV, DWV, <i>Nosema ceranae</i> , Trypanosomes
2010–2011	Ardahan	ABPV, BQCV, DWV, <i>Nosema ceranae</i> *, Trypanosomes
2010–2011	Artvin	ABPV, BQCV, DWV, <i>Nosema ceranae</i> *, Trypanosomes
2010	Bitlis	ABPV, BQCV, DWV, <i>Nosema ceranae</i> , Trypanosomes
2010	Edirne	BQCV, DWV, Trypanosomes
2010	Elazığ	BQCV, DWV, Trypanosomes

*Detected among samples of 2011 only.

In 2010, *N. ceranae* was widespread in migratory colonies of Muğla, Hatay, and Bitlis and less frequent in samples from Ankara and Kırklareli. *N. ceranae* was not found in samples of beekeepers from Edirne, Artvin, Ardahan, and Elazığ in 2010 (Table 2). In contrast, *N. ceranae* showed high incidence in all sites except for the stationary colonies of Hatay in 2011 (Table 2). *N. ceranae* loads were higher among migratory colonies than stationary ones in both years. *N. apis* was not observed among the samples of 2010 and most of the samples of 2011 although the analysis was repeated twice. In 2011, *N. apis* was detected in three stationary colonies in total, from three beekeepers in Muğla-Bodrum, Düzce-Nas, and Ardahan-Posof.

Trypanosome loads were significantly different among regions in both years. Trypanosome abundance was higher within the samples of migratory beekeepers than those of stationary ones (Table 3 in Supplementary Material). Ten/ten positive samples were confirmed by DNA sequencing as reflecting the 28S rRNA locus of trypanosomes. In general trypanosome levels were higher among samples from Muğla, Ankara, Artvin and Ardahan compared to the samples from Kırklareli, Edirne and Elazığ, for 2010. In 2011, levels were especially high in samples from Artvin, Yığılca, and Muğla, and low in samples of stationary colonies in Hatay and Kırklareli (Table 2). Seasonal variation was observed among samples collected in 2010 (ANOVA $p = 0.0023$). Specifically, trypanosome loads were highest in spring as opposed to fall. In Ardahan and Hatay, trypanosomes were not detected among any of the fall samples. Among the samples with trypanosome infection, thirteen 2010 samples for which products were assayed by DNA sequencing were all confirmed as being part of the “SF” strain of *L. passim* (identified as *C. mellificae*) (Runckel et al., 2011) and 10 were positive among 51 samples in 2011, based on qPCR and melt-curve analyses (seven products were confirmed by DNA sequencing). four of the “SF” positive samples were from Muğla, six from Hatay, eight from Artvin-Ardahan, four from Yığılca, and one from Elazığ. GAPDH sequence analyses from Artvin and Yığılca all matched *L. passim*, haplotype A (Morimoto et al., 2013), indicating that this haplotype was more common among the samples when compared with haplotype B.

S. melliferum was detected for two samples in Artvin province in 2010 while 13 samples were positive for Artvin, Ardahan, Muğla, Yığılca, and Kırklareli during 2011. Six of the positive samples were confirmed by DNA sequencing. *S. apis* was detected in only two samples from Muğla and Elazığ in 2010 and in one sample from Yığılca in 2011.

Dual infections of DWV and BQCV were detected in both migratory and stationary beekeepers in 2010 (50–100% of honey bees in six provinces sampled). Dual infections with these two viruses were less commonly observed in 2011 (18–40% of honey bees in four of the provinces sampled). Triple infections of ABPV, BQCV, and DWV were seen in three regions (50–100% of honey-bee samples) in 2010, whereas occurred less (18–40% of honey bee samples) at four regions in 2011. Among 2010 samples, *N. ceranae* loads were correlated with DWV loads ($r = 0.25$, $p = 0.0186$). ABPV loads were correlated with both DWV ($r = 0.40$, $p = 0.0001$) and *N. ceranae* ($r = 0.35$, $p = 0.0009$). BQCV correlated with DWV within samples from Ankara, Bitlis, Edirne, and Muğla. In Bitlis, ABPV levels were positively correlated with DWV ($r = 0.89$, $p = 0.0061$). In Hatay, BQCV correlated with *N. ceranae* ($r = 0.47$, $p = 0.0443$). Among 2011 samples ABPV showed correlations with DWV ($r = 0.29$, $p = 0.0381$) and in some regions with *N. ceranae*. A positive correlation was observed between trypanosome load and *N. ceranae* ($r = 0.28$, $p = 0.0083$) and between trypanosomes and BQCV ($r = 0.59$, $p < 0.0001$).

According to the survey results, there was an increase in colony losses in 2011 when compared with 2010 losses (Figure 3). Kırklareli, Muğla, Hatay, and Ankara had relatively higher colony losses in both years while Ardahan and Artvin tended to have lower losses than the other provinces. High colony losses were also observed in Edirne and Bitlis in 2010 and especially in Yığılca in 2011. Average colony losses of migratory beekeepers were significantly higher than those of the stationary beekeepers.

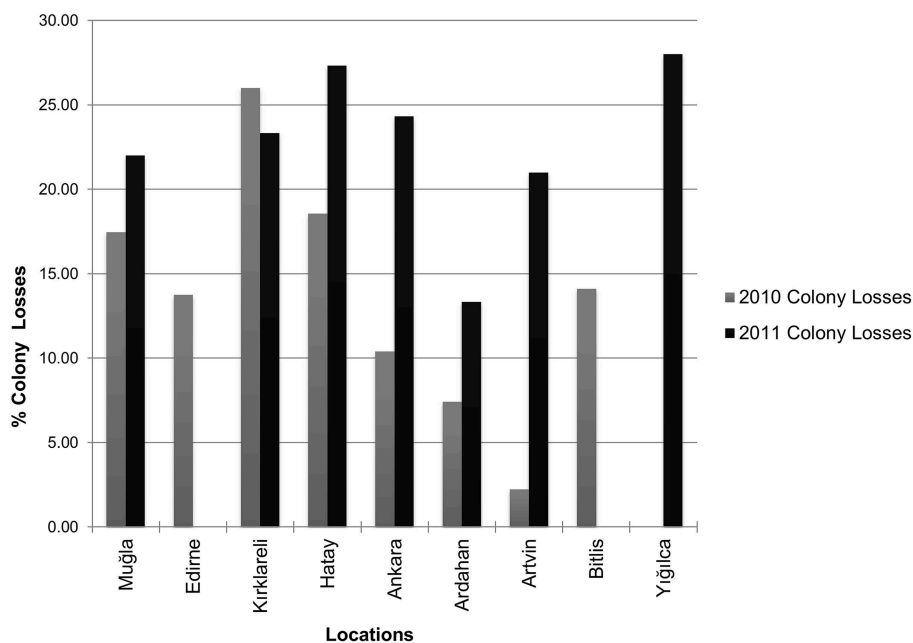
Read Mapping Results

DWV reads were detected in sequenced RNA libraries from all of the regions (Figure 4 and Table 4 in Supplementary Material), although they were most highly represented in Hatay, Yığılca, and Muğla, matching the results from real time q-PCR. We found strong evidence for *Varroa destructor* virus-1 as well, a first for these regions. Overall, the DWV/VDV group of iflaviruses was the predominant viral taxon (Figure 4). RNA sequences of BQCV were highest in the Ardahan region, although BQCV reads were also abundant in other provinces. Consistent with the fact that BQCV was not detected in Yığılca by RT-qPCR very few reads were found in this province during RNA sequencing. According to the sequence analysis, ABPV contigs were common in most regions. Consistent with RT-PCR results, they were most frequently observed in Yığılca and Ardahan. Very few CBPV reads were found in Kırklareli province and this virus was not detected or was very few in other regions. In contrast, the related Lake Sinai viruses were highly represented both in diversity across this large clade, and in abundance. IAPV and KBV were notably rare or were not present across all regions. Sacbrood virus was similarly rare, although 1885 SBV reads were detected in Ardahan

TABLE 2 | Normalized pathogen loads across sites and between stationary and migratory beekeepers for 2010 and 2011.

	DWV		ABPV		BQCV		<i>N. ceranae</i>		TRYP	
	2010	2011	2010	2011	2010	2011	2010	2011	2010	2011
STATIONARY BEEKEEPERS										
All sites	4.99	11.59	1.34	13.87	5.36	6.31	0.19	10.97	8.82	14.36
Muğla	7.02	12.57	0.37	11.6	5.83	3.94	0.42	3.81	ND	11.74
Hatay	10.99	15.16	3.42	8.69	8.86	5.99	ND	ND	12.36	1.91
Yığılca	/	17.55	/	19.40	/	ND	/	9.48	/	20.83
Kırklareli	3.88	9.15	ND	12.37	4.26	5.02	0.91	15.20	0.94	7.04
Ardahan	7.56	9.00	2.57	17.80	ND	13.00	ND	13.67	ND	16.28
Artvin	ND	6.13	2.26	13.37	5.78	9.92	ND	23.67	24.90	28.37
Ankara	4.31	/	0.78	/	11.64	/	ND	/	21.59	/
Edirne	1.14	/	ND	/	1.15	/	ND	/	1.92	/
MIGRATORY BEEKEEPERS										
All sites	4.81	17.08	6.32	18.09	8.13	5.99	1.54	4.69	13.92	15.89
Muğla	6.62	17.77	3.60	15.42	8.64	6.52	2.25	5.34	23.04	18.54
Hatay	8.21	16.38	5.05	20.75	10.32	5.45	2.69	4.04	14.77	13.24
Ardahan	ND	/	4.63	/	13.16	/	ND	/	26.24	/
Ankara	ND	/	4.91	/	7.28	/	0.71	/	2.15	/
Bitlis	11.56	/	19.72	/	6.33	/	3.57	/	12.92	/
Elazığ	2.44	/	ND	/	3.05	/	ND	/	4.38	/

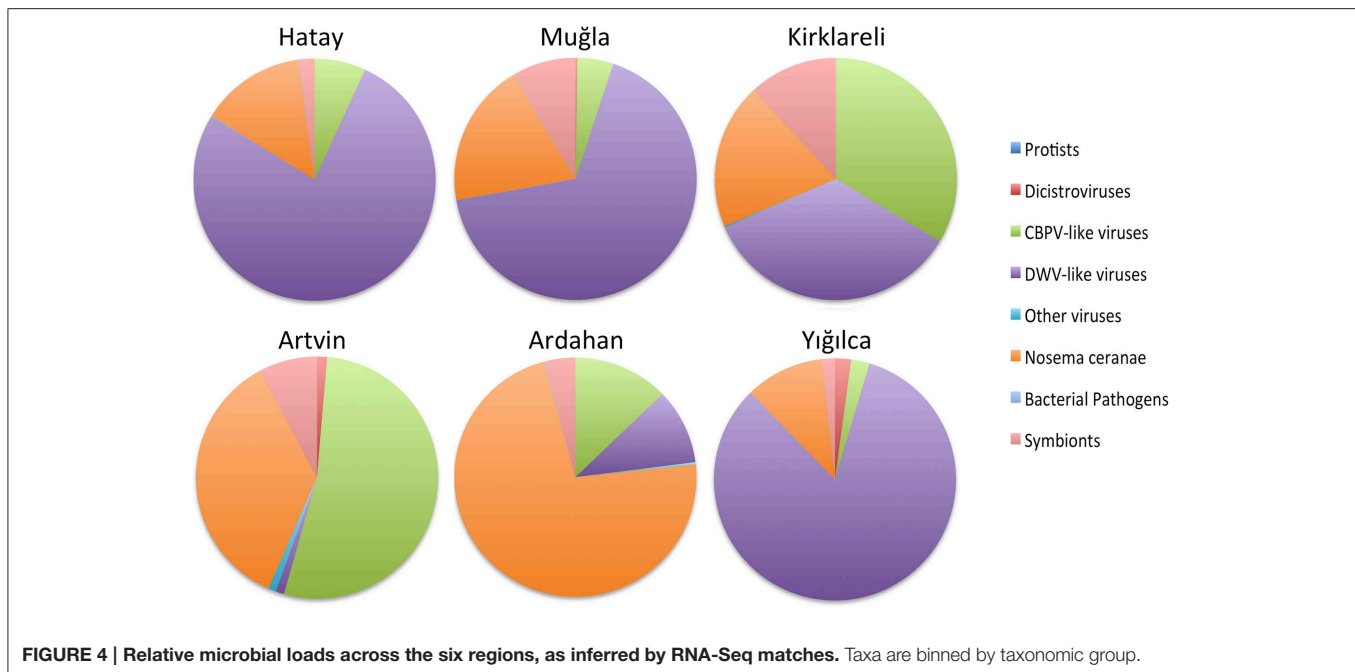
Abundances are log-based-two scale so every increase by one number reflects a doubling of pathogen RNA. ND, Not detected; /, Sampling was not done.

**FIGURE 3 | Percent colony losses of surveyed beekeepers by location in 2010 and 2011.**

region. SBPV reads (Slow Bee Paralysis Virus) were not observed in any of the regions (Table 4 in Supplementary Material).

Nosema ceranae was highly prevalent in our RNA-seq analysis. *N. ceranae* was especially common in Artvin province, again matching our RT-qPCR analyses. *N. ceranae* was not detected in

stationary colonies of Hatay province but it was evident there by RNA-seq analysis. We found minimal, if any reads matching *N. apis* in the RNASeq data, those present were largely the result of regions with high sequence similarity to *N. ceranae*. Reads for the betaproteobacterium, *Snodgrassella alvi* were common



among all of the provinces with highest frequency in Kırklareli, Muğla, Yığılca, and Ardahan. The gamma-proteobacterium, *Gilliamella apicola* was prevalent in all regions, while the related *Frischella perrara* was rare (Table 4 in Supplementary Material). Consistent with both RT-qPCR and DNA sequencing, the presence of *Spiroplasma melliferum* was confirmed in most of the regions. The abundance of congener *S. apis* was confirmed with the RNA-seq analysis despite being missed by PCR, yet *S. apis* remained the minor of the two species.

Discussion

Evaluation of Real Time q-PCR Results

The surveyed regions comprise important centers of beekeeping in Turkey. Kashmir Bee Virus (KBV) and Israel Acute Bee Paralysis Virus (IAPV) were not detected in sampled colonies by RT-qPCR, while the related virus Acute Bee Paralysis Virus (ABPV) was abundant in most of the samples and was present in the regions where colony declines were observed. Historically, KBV has been found most frequently in the United States and Australia (Allen and Ball, 1996) and this virus is thought to be an exotic bee virus in central Europe (Berenyi et al., 2006). Our discovery of just one of these three species could reflect climatic and environmental conditions (Anderson, 1991) or differential abilities of these viruses to infect different honey bee subspecies.

DWV was detected in both migratory and stationary operations in all of the regions, consistent with the high prevalence of this virus worldwide (Tentcheva et al., 2004; Berenyi et al., 2006; Kukielka et al., 2008; Nielsen et al., 2008; Welch et al., 2009). The presence of DWV has been reported in colonies of *Apis mellifera* L. in Ordu province of Turkey (Gülmez et al., 2009). BQCV was the second most widespread bee virus in our study. BQCV is also the second-most prevalent virus in honey bee colonies in

Asia and Europe (Tentcheva et al., 2004) and its presence was confirmed in many studies (Berenyi et al., 2006; Kukielka et al., 2008; Welch et al., 2009; Choe et al., 2012). BQCV was previously reported in 21.42% of 28 bee samples from six provinces of the Black Sea Region in Turkey (Gümüşova et al., 2010). BQCV was more prevalent in migratory bee colonies, consistent with the results found for migratory bees sampled in the U.S. (Welch et al., 2009).

In most of the regions we studied, BQCV and DWV were found together more often than expected by chance. This result is consistent with the findings in a previous study of virus infections (Chen et al., 2004) for which these distantly related viruses were found coinfecting honey bees at a high frequency. Triple infections of ABPV, BQCV, and DWV were detected in some colonies. Nielsen et al. (2008) also found a high incidence of dual and triple infections. According to Chen et al. (2004), 50% of colonies in the USA had dual infections while 7% had triple infections. Simultaneous multiple infections were also common in Austria (Berenyi et al., 2006), Southwest England (Baker and Schroeder, 2008), Brazil (Teixeira et al., 2008), Jordan (Haddad et al., 2008), France (Tentcheva et al., 2004), and Hungary (Forgach et al., 2008).

In Turkey, the presence of *N. apis* was confirmed earlier (Aydin et al., 2005; Muz et al., 2010; Whitaker et al., 2010). Using qPCR, we detected *N. apis* in only three colonies, indicating the replacement of *N. apis* by *N. ceranae* in many regions included in this study. Worldwide, *N. ceranae* is now far more common than *N. apis* (Chen et al., 2009; Valera et al., 2011; Yoshiyama and Kimura, 2011).

In 2006, *N. ceranae* was found in the provinces of Artvin, Hatay, and Muğla (Whitaker et al., 2010). Collapsed colonies from the Hatay overwintering region and Southeastern Marmara were found to show infections of *N. ceranae* (Muz et al., 2010). In our 2010 samples, *N. ceranae* was observed in Bitlis, Hatay,

and Muğla and not in Edirne, Artvin, Ardahan, and Elazığ in 2010. Interestingly, *N. ceranae* was widely common for all of the regions involved in the study in 2011. Detection of *N. ceranae* among 2011 samples of Ardahan-Artvin might be the reason for the increase in colony losses in these provinces in 2011. The incidence of *N. ceranae* in Kırklareli province in 2011 was much more than the previous year. *N. ceranae* loads of samples from migratory beekeepers were significantly higher than the stationary ones and *N. ceranae* loads were correlated the viruses ABPV, BQCV and DWV. We propose that the infectivity of *N. ceranae* expands with migratory beekeeping activities and in association with different viruses and trypanosomatids. In July, the Thrace region is a highly frequented location by migratory beekeepers seeking to harvest sunflower honey. It is thought that these beekeepers in this region had colony losses because of the application of pesticides rather than honey bee disease factors. Hatay and Ankara are also important locations for migratory beekeepers. Caucasian bees which are used by most of the migratory beekeepers, are raised in Ardahan and Artvin. Bitlis also suffered high colony losses in 2010, combined with high disease loads among samples of migratory beekeepers. In contrast, while honeybee samples in Yığılca were derived from stationary colonies, high colony losses and high pathogen loads were observed. Probably some other factors were contributing to colony losses in this province. All five viruses and pathogens were also detected in Muğla, the center of migratory beekeeping. These results could reflect the differences in pathogen exposure of local and migratory colonies, varying resistance levels, or perhaps a differential ability to handle stress. Genetic impacts of migratory beekeeping has become an important concern in Turkey and elsewhere. Population structure can be disturbed with hybridization, leading to a loss of regionally adaptive traits and perhaps decreasing colony fitness.

Trypanosomes were highly prevalent in our samples. Trypanosomes have also been reported in Australia (Langridge and McGhee, 1967), China (Yang et al., 2013), France (Dainat et al., 2012), Japan (Morimoto et al., 2013), Switzerland (Schmid-Hempel and Tognazzo, 2010), USA (vanEngelsdorp et al., 2009; Runckel et al., 2011), and Spain (Orantes-Bermejo, 1999). Trypanosome levels were higher in summer samples than the fall samples in our samples. In southern Spain, this parasite appears in July and August (Orantes-Bermejo, 1999) and *L. passim* (recently named as the primary trypanosomatid in lieu of *C. mellificae*, Schwarz et al., 2015) was common in summer colonies in Belgium (Ravoet et al., 2013). These findings are in contrast with the results of Runckel et al. (2011) which show a peak in *L. passim* levels in January in U.S. colonies.

Synergistic effects can make colonies more vulnerable to other pathogens (Cornman et al., 2012) and we noted a positive correlation between *L. passim* and *N. ceranae* prevalence. Similarly, this relationship was documented in field surveys from the U.S. (Runckel et al., 2011) and the co-occurrence of *L. passim* and *N. ceranae* in summer was tied to higher colony mortality in Belgium (Ravoet et al., 2013). Complex dynamic immune responses of honey bees to both *Nosema* and trypanosomatids were recently reported (Schwarz and Evans, 2013) and it will be interesting to test further for mechanistic explanations for any synergisms between these parasites.

Metagenomic Sequencing

As in prior studies, iflaviruses related to Deformed wing virus (DWV) were especially prevalent. While DWV amounts were higher, reads for the closely related *Varroa destructor*-1 virus (VDV-1) were fairly abundant in Hatay, Yığılca, and Muğla. CBPV reads were relatively rare in most of the regions and were not detected in Yığılca and Ardahan. CBPV was reported in seven of 28 (25%) samples from six provinces of Black Sea region in Turkey (Gümüsova et al., 2010). CBPV was detected in four of 96 apiaries in survey study of Denmark, in 73% of the samples from Greek and 9% of apiaries in China (Nielsen et al., 2008; Bacandritsos et al., 2010; Ai et al., 2012). A recently described relative of CBPV, Lake Sinai virus, was prevalent in all regions, second only to the DWV and VDV group. In fact, this group was the most prevalent virus in Kırklareli and Ardahan provinces. The pathogenic or epidemiological significance of Lake Sinai viruses are not well-known. The LSV species complex is diverse, with members sharing between 70 and 99% sequence identity, hence this group is often missed in screenings based on PCR. LSV4 appears to be especially abundant, along with LSV1 and LSV2. There was significant variation across the regions in the specific lineages seen for this group. LSV2 was the most abundant single component of the honey bee microbiome in the study of Runckel et al. (2011). The presence of LSV was also confirmed in honey bees from Spain by high-throughput sequencing (Granberg et al., 2013). A new fourth strain of Lake Sinai Virus (LSV) was identified in the study of Ravoet et al. (2013).

In this study, despite tens of millions of microbial gene reads, few reads matching IAPV, KBV, and SBV were seen. IAPV, KBV, and SBV were not detected using RT-qPCR, indicating that RNA-seq sensitivity was higher than that of qPCR, or that current qPCR primers for this group must be redesigned to capture all strains. Given the RNA-Seq data, we can consider that these pathogens are quite rare in our samples. KBV has been absent in some other European surveys (e.g., Berenyi et al., 2006; Forgach et al., 2008) and rare in surveys in France, Denmark, and United Kingdom (Tentcheva et al., 2004; Ward et al., 2007; Nielsen et al., 2008). IAPV is prevalent in the Middle East and Australia (Maori et al., 2007; Palacios et al., 2008) and this species was reported in 71 samples containing 10 bees; each from 20 provinces in Turkey (Ozkirim and Schiesser, 2013). In this study IAPV wasn't confirmed by RT-qPCR in our samples and RNA sequence analysis revealed that this virus was not present in Muğla and Hatay and was incredibly rare among samples from Artvin, Yığılca, and Ardahan. SBV was similarly scarce in this study. This conflicts with results of Tentcheva et al. (2004) but overlaps with the results of Baker and Schroeder (2008) and Forgach et al. (2008). As with most surveys, we examined only adult honey bees. SBV causes a fatal disease in honey bee larvae (Bailey, 1975), thus brood samples could provide more evidence for its prevalence.

Apis mellifera filamentous virus (AmFV) was not detected within our samples, nor was VdMLV (*Varroa* Macula-like virus). A plant-pathogenic RNA virus, tobacco ringspot virus (TRSV) was not detected in any of the RNA samples in our study.

Recent culture-independent studies reported eight bacterial phylotypes inhabiting the gut of the honey bee, *Apis mellifera* from several continents (Jeyaprakash et al., 2003; Mohr and Tebbe, 2006; Babendreier et al., 2007; Cox-Foster et al., 2007; Olofsson and Vasquez, 2008). Colonies founded by swarms, interactions within the colony, intercolony interactions like robbing food in neighboring hives and mixing of colonies by beekeepers all might affect the gut microbiota (Engel et al., 2012). In our study, unique gene reads of *Snodgrassella alvi* and *Gilliamella apicola* were highly represented and showed differences among provinces because of geographic, environmental and subspecies differences of hosts. The differences in social behaviors of the subspecies, the dietary sources and exposure to varying pathogens and pesticides might influence the abundance of these bacteria among regions. One gamma-proteobacterial member of the gut microbiota *Candidatus Schmidhempelia bombi*, was present in 90% of bumble bee individuals in the study of Martinson et al. (2014). This symbiont was prevalent in all of our surveyed locations and widely represented in Hatay, Yıgılca, and Muğla. The recently described gamma-proteobacterium *Frischella perrara* (Engel et al., 2013) was present in our study albeit at low levels.

Lactobacillales symbionts have been proposed as actors in both nutrition and parasite defenses of honey bees. *Lactobacillales* stimulate the innate immune system, arguably increasing honey bee defenses against disease agents (Evans and Lopez, 2004). Along with their impacts on immunity, the microbial symbionts have been proposed to nutritionally compete with pathogens by occupying the available niches (Crotti et al., 2012). In this study, very low number of 16S *Lactobacillus* reads were observed among regions. Overall bacterial loads were especially high in Hatay, Yıgılca, and Artvin. The prevalence of the bacterial pathogens *S. apis* and *S. melliferum* is low among Belgian honey bee colonies (Ravoet et al., 2013), and these bacteria are present only seasonally in North and South America (Runckel et al., 2011; Schwarz et al., 2014). We found low *Spiroplasma* levels, with highest incidence in Artvin province. *S. melliferum* was the more common species, matching results from the Americas (Schwarz et al., 2014). The bacterial brood diseases European Foulbrood (EFB) caused by the bacterium *Melissococcus plutonius* (Bailey et al., 1983) and American Foulbrood (AFB) caused by the bacterium *Paenibacillus larvae* (Genersch et al., 2006) are globally important diseases of honey bees. *P. larvae* reads were not common among our samples. Similarly, *M. plutonius* was not prevalent among the regions but was more ubiquitous in Hatay, Muğla, and Yıgılca. Environmental conditions at these sites can be conducive for the expression of the disease. Like AFB, EFB transmission is also linked to larval immune responses (Evans, 2004), hygienic behavior (Spivak and Reuter, 2001) as well as interaction between *M. plutonius* and the intestinal microbiota of the honey bee larvae (Gilliam, 1997; Olofsson and Vasquez, 2008), nutritional and stress conditions, weather and geography (Bailey, 1961).

Sequences for *Ascosphaera apis*, the causative agent for Chalkbrood disease, were generally rare, with the highest incidence from samples of in Hatay province. Similarly, neogregarines (nominally *Apicystis bombi*) persisted at extremely low levels

among regions in this study. Among arthropod parasites of honey bees, we found no genetic evidence for the presence of the tarsonemid tracheal mite, *Acarapis woodi* within our samples. The Asian parasitic mite *Tropilaelaps* is considered more dangerous to *A. mellifera* than the parasitic mite *Varroa destructor* (Rath et al., 1995), and this mite is worthy of screening. Our deep sequencing analysis showed no sign of *Tropilaelaps*. The presence of phorid flies (*Apocephalus borealis*) in the study of Ravoet et al. (2013) proves their existence in Europe, but our deep sequencing did not reveal signs of this parasite.

In conclusion, we screened bee-derived RNA against the most complete sequence set for honey bee associates used to date. We assessed levels of known and novel parasites, pathogens, and symbionts. We present quantitative data for bacterial pathogens (*Melissococcus plutonius*, *Paenibacillus larvae*, *S. apis*, *S. melliferum*), protists (*Apicystis*, trypanosomatids), viruses (Lake Sinai virus, Chronic bee paralysis virus, Deformed wing virus, *Varroa destructor* virus, Sacbrood, and Dicistroviruses), symbionts (*Candidatus Schmidhempelia bombi*, *Frischella perrara*, *Snodgrassella alvi*, *Gilliamella apicola*, *Lactobacillus* spp., *Acetobacteraceae*), microsporidia and fungi in *A. mellifera* colonies in distinct regions of Turkey. The presence of KBV, SBPV, Tobacco ringspot virus, VdMLV (Varroa Macula like), *Acarapis* spp., *Tropilaelaps clareae* and *Apocephalus* (phorid fly) were also examined. As in other countries, bee viruses were correlated with colony losses in Turkey. In comparison with 2010, the increase in pathogen loads in 2011 might be a factor for increased colony losses observed in this study. It seems likely that migratory beekeeping practices enable the spread of disease factors among honey bees in places where they visit and causing an important threat to the honey bee colonies. In addition, the impacts of parasites and pathogens varies between regions, perhaps reflecting different honey bee genetic traits. Migratory beekeeping was correlated with both higher disease loads and a potential risk of dispersing regional parasites and pathogens across the country. This practice also allows for greater gene flow between migratory honey bee populations and local populations. Current diversity and local genetic structure can be preserved with selection strategies and establishing broad areas of isolation to reduce the risks of migratory beekeeping practices. While experimental work and longitudinal analyses will be needed to confirm causes of bee declines, our analyses, reference sequences, and strategy will help reduce the set of likely causes.

Acknowledgments

It is with sadness that we note the passing of COLOSS Management Committee member Dr. AK on February 1, 2014. We dedicate this paper to the memory of Professor AK, for initiating the study and for his teaching, mentoring, collegiality and friendship. The work was part of the project “The effects of the genetic differences of Turkish honey bees on colony vitality” and the work was funded by “The Science and Technological Research Council of Turkey” (TUBITAK Project Number:109T547). The work was also supported by Middle East Technical University (BAP Project number:BAP-01-08-2012-004. The molecular diagnostics

part was technically and financially supported by the USDA-ARS Bee Research Laboratory. We would like to express our gratitude to Dr. Ryan Schwarz for technical advice, Dawn Lopez and Margaret Smith for their laboratory guidance, and the scientists of the BRL for their feedback. We appreciate the advice and help of Prof. Dr. Muhsin Dogaroglu, Prof. Dr. Banu Yucel and Dr. Hasan Huseyin Inal in the initiation of the project. We are also grateful to Assist. Prof. Dr. Mustafa Muz, Okan Can Arslan,

Mehmet Kayim, Mert Kukrer, and Mustafa Nail Cirik for their contributions to the field work.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fgene.2015.00100/abstract>

References

- Ai, H., Yan, X., and Han, R. (2012). Occurrence and prevalence of seven bee viruses in *Apis mellifera* and *Apis cerana* apiaries in China. *J. Invertebr. Pathol.* 109, 160–164. doi: 10.1016/j.jip.2011.10.006
- Allen, M., and Ball, B. (1996). The incidence and world distribution of honey bee viruses. *Bee World* 77, 141–162.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Anderson, D. L. (1991). Pests and pathogens of the honeybee (*Apis mellifera* L.) in Fiji. *J. Apic. Res.* 29, 53–59.
- Anderson, D. L., and Gibbs, A. J. (1989). Transuparial transmission of Kashmir bee virus and sacbrood virus in the honey bee (*Apis mellifera*). *Ann. Appl. Biol.* 114, 1–7. doi: 10.1111/j.1744-7348.1989.tb06781.x
- Aydin, L., Çakmak, Y., Güleğen, E., and Wells, H. (2005). Honey bee nosema disease in the Republic of Turkey. *J. Apic. Res.* 44, 196–197. doi: 10.3896/IBRA.1.44.4.13
- Babendreier, D., Joller, D., Romeis, J., Bigler, F., and Widmer, F. (2007). Bacterial community structures in honeybee intestines and their response to two insecticidal proteins. *FEMS. Microbiol. Ecol.* 59, 600–610. doi: 10.1111/j.1574-6941.2006.00249.x
- Bacandritsos, N., Granato, A., Budge, G., Papanastasiou, I., Roinioti, E., Caldon, M., et al. (2010). Sudden deaths and colony population decline in Greek honey bee colonies. *J. Invertebr. Pathol.* 105, 335–340. doi: 10.1016/j.jip.2010.08.004
- Baggerly, K., Deng, L., Morris, J., and Aldaz, C. (2003). Differential expression in SAGE: accounting for normal between-library variation. *Bioinformatics* 19, 1477–1483. doi: 10.1093/bioinformatics/btg173
- Bailey, L. (1961). European foulbrood. *Am. Bee J.* 101, 89–92.
- Bailey, L. (1975). Recent research on honey bee viruses. *Bee World* 56, 55–64.
- Bailey, L., and Ball, B. V. (1991). *Honey Bee Pathology, 2nd Edn.* London: Academic Press.
- Bailey, L., Ball, B. V., and Perry, J. N. (1983). Association of viruses with 2 protozoal pathogens of the honeybee. *Ann. Appl. Biol.* 103, 13–20. doi: 10.1111/j.1744-7348.1983.tb02735.x
- Baker, A., and Schroeder, D. (2008). Occurrence and genetic analysis of picorna-like viruses infecting worker bees of *Apis mellifera* L. populations in Devon, South West England. *J. Invertebr. Pathol.* 98, 239–242. doi: 10.1016/j.jip.2008.02.010
- Ball, B. V., and Bailey, L. (1997). “Viruses,” in *Honey Bee Pests Predators and Diseases*, eds R. A. Morse and K. Flottum (Medina: AI Root Co), 11–31.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B. Stat. Methodol.* 57, 289–299.
- Berenyi, O., Bakonyi, T., Derakhshifar, I., Koglbberger, H., and Nowotny, N. (2006). Occurrence of six honeybee viruses in diseased Austrian apiaries. *Appl. Environ. Microbiol.* 72, 2414–2420. doi: 10.1128/AEM.72.4.2414-2420.2006
- Bishop-Lilly, K. A., Turell, M. J., Willner, K. M., Butani, A., Nolan, N. M., Lentz, S. M., et al. (2010). Arbovirus detection in insect vectors by rapid, high-throughput pyrosequencing. *PLoS Negl. Trop. Dis.* 4:e878. doi: 10.1371/journal.pntd.0000878
- Bodur, C., Kence, M., and Kence, A. (2007). Genetic structure of honeybee, *Apis mellifera* L. (Hymenoptera: Apidae) populations of Turkey inferred from microsatellite analysis. *J. Apic. Res.* 46, 50–56. doi: 10.3896/IBRA.1.46.1.09
- Brown, M. J. F., Schmid-Hempel, R., and Schmid-Hempel, P. (2003). Strong context-dependent virulence in a host-parasite system: reconciling genetic evidence with theory. *J. Anim. Ecol.* 72, 994–1002. doi: 10.1046/j.1365-2656.2003.00770.x
- Chen, Y., Zhao, Y., Hammond, J., Hsu, H. T., Evans, J., and Feldlaufer, M. (2004). Multiple virus infections in the honey bee and genome divergence of honey bee viruses. *J. Invertebr. Pathol.* 87, 84–93. doi: 10.1016/j.jip.2004.07.005
- Chen, Y. P., Evans, J. D., Zhou, L., Boncristiani, H., Kimura, K., Xiao, T., et al. (2009). Asymmetrical coexistence of *Nosema ceranae* and *Nosema apis* in honey bees. *J. Invertebr. Pathol.* 101, 204–209. doi: 10.1016/j.jip.2009.05.012
- Choe, S. E., Nguyen, L. T. K., Noh, J. H., Koh, H. B., Jean, Y. H., Kweon, C. H., et al. (2012). Prevalence and distribution of six bee viruses in Korean *Apis cerana* populations. *J. Invertebr. Pathol.* 109, 330–333. doi: 10.1016/j.jip.2012.01.003
- Clark, T. B., Whitcomb, R. F., Tully, J. G., Mouches, C., Saillard, C., Bove, J. M., et al. (1985). *Spiroplasma melliferum*, a new species from the honey bee (*Apis mellifera*). *Int. J. Syst. Bacteriol.* 35, 296–308. doi: 10.1099/00207713-35-3-296
- Cornman, R. S., Boncristiani, H., Dainat, B., Chen, Y. P., vanEngelsdorp, D., Weaver, D., et al. (2013). Population genomic variation within RNA viruses of the Western honey bee, *Apis mellifera*, inferred from deep sequencing. *BMC Genomics* 14:154. doi: 10.1186/1471-2164-14-154
- Cornman, R. S., Tarry, D. R., Chen, Y., Jeffreys, L., Lopez, D., Pettis, J. S., et al. (2012). Pathogen webs in collapsing honey bee colonies. *PLoS ONE* 7:e43562. doi: 10.1371/journal.pone.0043562
- Cox-Foster, D. L., Conlan, S., Holmes, E. C., Palacios, G., Evans, J. D., Moran, N. A., et al. (2007). A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 318, 283–287. doi: 10.1126/science.1146498
- Crotti, E., Balloi, A., Hamdi, C., Sansonno, L., Marzorati, M., Gonella, E., et al. (2012). Microbial symbionts: a resource for the management of insect-related problems. *Microb. Biotechnol.* 5, 307–317. doi: 10.1111/j.1751-7915.2011.00312.x
- Dainat, B., Evans, J. D., Gauthier, L., and Neumann, P. (2012). “*Crithidia mellificae* is widespread in Europe and can be used as a predictive marker of honeybee colony losses,” in *Proceedings of the 5th European Conference of Apidology* (Halle an der Saale).
- De Miranda, J. R., Cordon, G., and Budge, G. (2010). The Acute bee paralysis virus-Kashmir bee virus-Israeli acute paralysis virus complex. *J. Invert. Pathol.* 103, 30–47. doi: 10.1016/j.jip.2009.06.014
- Dillon, R. J., and Dillon, V. M. (2004). The gut bacteria of insects: nonpathogenic interactions. *Annu. Rev. Entomol.* 49, 71–92. doi: 10.1146/annurev.ento.49.061802.123416
- Ellis, J. D., and Munn, P. A. (2005). The worldwide health status of honey bees. *Bee World* 86, 88–101. Available online at: http://www.ibra.org.uk/articles/20080611_27
- Engel, P., Kwong, W. K., and Moran, N. A. (2013). *Frischella perrara* gen. nov., sp. nov., a gammaproteobacterium isolated from the gut of the honeybee, *Apis mellifera*. *Int. J. Syst. Evol. Microbiol.* 63, 3646–3651. doi: 10.1099/ijs.0.049569-0
- Engel, P., Martinson, V. G., and Moran, N. A. (2012). Functional diversity within the simple gut microbiota of the honey bee. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11002–11007. doi: 10.1073/pnas.1202970109
- Evans, J., Aronstein, K. A., Chen, Y., Hetru, C., Imler, J.-L., Jiang, H., et al. (2006). Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect. Mol. Biol.* 15, 645–656. doi: 10.1111/j.1365-2583.2006.00682.x
- Evans, J. D. (2004). Transcriptional immune responses by honey bee larvae during invasion by the bacterial pathogen, *Paenibacillus larvae*. *J. Invertebr. Pathol.* 85, 105–111. doi: 10.1016/j.jip.2004.02.004

- Evans, J. D., Chen, Y. P., Cornman, R. S., De La Rua, P., Foret, S., Foster, L., et al. (2013). "Standard methodologies for molecular research in *Apis mellifera*," in *The COLOSS BEEBOOK, Volume I: Standard Methods for Apis Mellifera Research*, eds V. Dietemann, J. D. Ellis, and P. Neumann (Treforest, UK: International Bee Research Association), 52.
- Evans, J. D., and Lopez, D. L. (2004). Bacterial probiotics induce an immune response in the honey bee (*Hymenoptera: Apidae*). *J. Econ. Entomol.* 97, 752–756. doi: 10.1093/jee/97.3.752
- Forgach, P., Bakonyi, T., Tapaszi, Z., Nowotny, N., and Rusvai, M. (2008). Prevalence of pathogenic bee viruses in Hungarian apiaries: situation before joining the European Union. *J. Invertebr. Pathol.* 98, 235–238. doi: 10.1016/j.jip.2007.11.002
- Fries, I. (2010). *Nosema ceranae* in European honey bees (*Apis mellifera*). *J. Invertebr. Pathol.* 103, 73–79. doi: 10.1016/j.jip.2009.06.017
- Fries, I., Chauzat, M. P., Chen, Y. P., Doublet, V., Genersch, E., Gisder, S., et al. (2013). Standard methods for *Nosema* research. *J. Apic. Res.* 52, 1–28. doi: 10.3896/IBRA.1.52.1.14
- Fries, I., Feng, F., daSilva, A., Slemenda, S. B., and Pieniazek, N. J. (1996). *Nosema ceranae* n. sp. (*Microspora, Nosematidae*), morphological and molecular characterization of a microsporidian parasite of the Asian honey bee *Apis cerana* (*Hymenoptera, Apidae*). *Eur. J. Protistol.* 32, 356–365. doi: 10.1016/S0932-4739(96)80059-9
- Fries, I., Martin, R., Meana, A., Garcia-Palencia, P., and Higes, M. (2006). Natural infections of *Nosema ceranae* in European honey bees. *J. Apic. Res.* 45, 230–233. doi: 10.3896/IBRA.1.45.4.13
- Gegeer, R. J., Otterstatter, M. C., and Thomson, J. D. (2005). Does parasitic infection impair the ability of bumblebees to learn flower-handling techniques? *Anim. Behav.* 70, 209–215. doi: 10.1016/j.anbehav.2004.09.025
- Genersch, E., Forsgren, E., Pentikainen, J., Ashiralieva, A., Rauch, S., Kilwinski, J., et al. (2006). Reclassification of *Paenibacillus larvae* subsp. *pulvificiens* and *Paenibacillus larvae* subsp. *larvae* as *Paenibacillus larvae* without subspecies differentiation. *Int. J. Syst. Evol. Microbiol.* 56, 501–511. doi: 10.1099/ijs.0.63928-0
- Gilliam, M. (1997). Identification and roles of non-pathogenic microflora associated with honey bees. *FEMS Microbiol. Lett.* 155, 1–10. doi: 10.1016/S0378-1097(97)00337-6
- Granberg, F., Vicente-Rubiano, M., Rubio-Guerri, C., Karlsson, O. E., Kukiella, D., Belák, S., et al. (2013). Metagenomic detection of viral pathogens in spanish honeybees: coinfection by aphid lethal paralysis, Israel acute paralysis and lake sinai viruses. *PLoS ONE* 8:e57459. doi: 10.1371/journal.pone.0057459
- Gülmez, Y., Bursalı, A., and Tekin, S. (2009). First molecular detection and characterization of deformed wing virus (DWV) in honeybees (*Apis mellifera* L.) and mites (*Varroa destructor*) in Turkey. *Afr. J. Biotechnol.* 8, 3698–3702. Available online at: <http://www.ajol.info/index.php/ajb/article/viewFile/61919/49978>
- Gümüşova, S. O., Albayrak, H., Kurt, M., and Yazici, Z. (2010). Prevalence of three honey bee viruses in Turkey. *Veterinarski Arhiv.* 80, 779–785. Available online at: <http://www.vet.unizg.hr/vetarhiv/papers/2010-80-6-10.pdf>
- Haddad, N., Brake, M., Migdadi, H., and de Miranda, J. R. (2008). First detection of honey bee viruses in Jordan by RT-PCR. *Jordan J. Agric. Sci.* 4, 242–246. Available online at: <https://journals.ju.edu.jo/JJAS/article/view/1003/996>
- Higes, M., Martín-Hernández, R., Botías, C., Bailón, E., González-Porto, A., Barrios, L., et al. (2008). How natural infection by *Nosema ceranae* causes honey bee colony collapse. *Environ. Microbiol.* 10, 2659–2669. doi: 10.1111/j.1462-2920.2008.01687.x
- Jeyaprakash, A., Hoy, M. A., and Allsopp, M. H. (2003). Bacterial diversity in worker adults of *Apis mellifera capensis* and *Apis mellifera scutellata* (*Insecta: Hymenoptera*) assessed using 16S rRNA sequences. *J. Invertebr. Pathol.* 84, 96–103. doi: 10.1016/j.jip.2003.08.007
- Kandemir, Y., Kence, M., and Kence, A. (2000). Genetic and morphometric variation in honeybee (*Apis mellifera*) populations of Turkey. *Apidologie* 31, 343–356. doi: 10.1051/apido:2000126
- Kevan, P. G., Guzman, E., Skinner, A., and van Englesdorp, D. (2007). Colony Collapse Disorder (CCD) in Canada: do we have a problem? *Hivelights* 20, 15–18. Available online at: <http://www.honeycouncil.ca/hivelights.php?hlID=21>
- Klee, J., Besana, A. M., Genersch, E., Gisder, S., Nanetti, A., Tam, D. Q., et al. (2007). Widespread dispersal of the microsporidian *Nosema ceranae*, an emergent pathogen of the western honey bee, *Apis mellifera*. *J. Invertebr. Pathol.* 96, 1–10. doi: 10.1016/j.jip.2007.02.014
- Kukiella, D., Esperon, F., Higes, M., and Sanchez-Vizcaino, J. M. (2008). A sensitive one-step real-time RT-PCR method for detection of deformed wing virus and black queen cell virus in honeybee *Apis mellifera*. *J. Virol. Methods* 147, 275–281. doi: 10.1016/j.jviromet.2007.09.008
- Lange, C. E., and Lord, J. (2012). "Protistan entomopathogens," in *Insect Pathology, 2nd Edn.*, eds B. Vega and H. Kaya (Amsterdam: Elsevier), 367–394. doi: 10.1016/B978-0-12-384984-7.00010-5
- Langridge, D. F., and McGhee, R. B. (1967). *Crithidia mellificae* n. sp. an acidophilic trypanosomatid of the honey bee *Apis mellifera*. *J. Protozool.* 14, 485–487. doi: 10.1111/j.1550-7408.1967.tb02033.x
- Ma, M., Huang, Y., Gong, Z., Zhuang, L., Li, C., Yang, H., et al. (2011). Discovery of DNA viruses in wild-caught mosquitoes using small RNA high throughput sequencing. *PLoS ONE* 6:e24758. doi: 10.1371/journal.pone.0024758
- Maheshwari, J. K. (2003). Endangered pollinators. *Environ. News Arch.* 9, 32–45.
- Maori, E., Lavi, S., Mozes-Koch, R., Gantman, Y., Peretz, Y., Edelbaum, O., et al. (2007). Isolation and characterization of Israeli acute paralysis virus, a dicistrovirus affecting honeybees in Israel: evidence for diversity due to intra- and interspecies recombination. *J. Gen. Virol.* 88, 3428–3438. doi: 10.1099/vir.0.83284-0
- Markham, P. G., and Townsend, R. (1981). Spiroplasmas. *Sci. Prog. Oxf.* 67, 43–68.
- Martinson, V. G., Magoc, T., Koch, H., Salzberg, S. L., and Moran, N. A. (2014). Genomic features of a bumble bee symbiont reflect its host environment. *Appl. Environ. Microbiol.* 80, 3793–3803. doi: 10.1128/AEM.00322-14
- Merzlyak, E., Yurchenko, V., Kolesnikov, A. A., Alexandrov, K., Podlipaev, S. A., and Maslov, D. A. (2001). Diversity and phylogeny of insect trypanosomatids based on small subunit rRNA genes: polyphyly of *Leptomonas* and *Blastocrithidia*. *J. Eukaryot. Microbiol.* 48, 161–169. doi: 10.1111/j.1550-7408.2001.tb00298.x
- Minoche, A. E., Dohm, J. C., and Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.* 12, R112. doi: 10.1186/gb-2011-12-11-r112
- Mohr, K. I., and Tebbe, C. C. (2006). Diversity and phylotype consistency of bacteria in the guts of three bee species (*Apoidea*) at an oilseed rape field. *Environ. Microbiol.* 8, 258–272. doi: 10.1111/j.1462-2920.2005.00893.x
- Moran, N. A., Hansen, A. K., Powell, J. E., and Sabree, Z. L. (2012). Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees. *PLoS ONE* 7:e36393. doi: 10.1371/journal.pone.0036393
- Morimoto, T., Kojima, Y., Yoshiyama, M., Kimura, K., Yang, B., Peng, G., et al. (2013). Molecular detection of protozoan parasites infecting *Apis mellifera* colonies in Japan. *Environ. Microbiol. Rep.* 5, 74–77. doi: 10.1111/j.1758-2229.2012.00385.x
- Mouches, C., Bove, J., Tully, J., Rose, D., McCoy, R., Carle-Junca, P., et al. (1983). Spiroplasma apis, a new species from the honey-bee (*Apis mellifera*). *Ann. Inst. Pasteur. Microbiol.* 134, 383–397.
- Muz, M. N., Girgin, A. O., Muz, D., and Aydin, L. (2010). Molecular detection of *Nosema ceranae* and *Nosema apis* infections in Turkish apiaries with collapsed colonies. *J. Apic. Res.* 49, 342. doi: 10.3896/IBRA.1.49.4.09
- Naug, D. (2009). Nutritional stress due to habitat loss may explain recent honey bee colony collapses. *Biol. Conserv.* 142, 2369–2372. doi: 10.1016/j.biocon.2009.04.007
- Nielsen, S. L., Nicolaisen, M., and Kryger, P. (2008). Incidence of acute bee paralysis virus, black queen cell virus, chronic bee paralysis virus, deformed wing virus, Kashmir bee virus and sacbrood virus in honey bees (*Apis mellifera*) in Denmark. *Apidologie* 39, 310–314. doi: 10.1051/apido:2008007
- Nordstrom, S., Fries, A., Aarhus, H., Hansen, H., and Korpela, S. (1999). Virus infections in Nordic honey bee colonies with no, low or severe *Varroa jacobsoni* infestation. *Apidologie* 30, 457–466. doi: 10.1051/apido:19990602
- Olofsson, T. C., and Vasquez, A. (2008). Detection and identification of a novel lactic acid bacterial flora within the honey stomach of the honeybee *Apis mellifera*. *Curr. Microbiol.* 57, 356–363. doi: 10.1007/s00284-008-9202-0
- Orantes-Bermejo, F. (1999). A scientific note on the prevalence of trypanosomatid parasites of honeybees (*Apis mellifera* L.) in southern Spain. *Apidologie* 30, 77–78. doi: 10.1051/apido:19990110
- Ozkirim, A., and Schiesser, A. (2013). Israel acute paralysis virus (IAPV) in Turkish bees. *J. Apic. Res.* 52, 56–57. doi: 10.3896/IBRA.1.52.2.09
- Palacios, G., Hui, J., Quan, P. L., Kalkstein, A., Honkavouri, K. S., Busetti, A. V., et al. (2008). Genetic analysis of Israel acute paralysis virus: distinct clusters

- are circulating in the United States. *J. Virol.* 82, 6209–6217. doi: 10.1128/JVI.00251-08
- Paxton, R. J., Klee, J., Korpela, S., and Fries, I. (2007). *Nosema ceranae* has infected *Apis mellifera* in Europe since at least 1998 and may be more virulent than *Nosema apis*. *Apidologie* 38, 558–565. doi: 10.1051/apido:2007037
- Rath, W., Boecking, O., and Drescher, W. (1995). The phenomena of simultaneous infestation of *Apis mellifera* in Asia with the parasitic mites *Varroa jacobsoni* Oud. and *Tropilaelaps clareae* Delfinado-Baker. *Am. Bee J.* 135, 125–127
- Ravoet, J., Maharramov, J., Meeus, I., De Smet, L., Wenseleers, T., Smagghe, G., et al. (2013). Comprehensive bee pathogen screening in Belgium reveals *Crithidia melli?cae* as a new contributory factor to winter mortality. *PLoS ONE* 8:e72443. doi: 10.1371/journal.pone.0072443
- Round, J. L., and Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* 9, 313–323. doi: 10.1038/nri2515
- Runckel, C., Flenniken, M. L., Engel, J. C., Ruby, J. G., Ganem, D., Andino, R., et al. (2011). Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, *Nosema*, and *Crithidia*. *PLoS ONE* 6:e20656. doi: 10.1371/journal.pone.0020656
- Schmid-Hempel, R., and Tognazzo, M. (2010). Molecular divergence defines two distinct lineages of *Crithidia bombi* (Trypanosomatidae), parasites of bumblebees. *J. Eukaryot. Microbiol.* 57, 337–345. doi: 10.1111/j.1550-7408.2010.00480.x
- Schwarz, R. S., Bauman, G. R., Murphy, C. A., Ravoet, J., de Graaf, D. C., and Evans, J. D. (2015). Characterization of two species of trypanosomatidae from the honey bee *Apis mellifera*: *Crithidia mellificae* Langridge and McGhee, 1967 and *Lotmaria passim* n. gen., n. sp. *J. Euk. Microbiol.* doi: 10.1111/jeu.12209. [Epub ahead of print].
- Schwarz, R. S., and Evans, J. D. (2013). Single and mixed-species trypanosome and microsporidia infections elicit distinct, ephemeral cellular and humoral immune responses in honey bees. *Dev. Comp. Immunol.* 40, 300–310. doi: 10.1016/j.dci.2013.03.010
- Schwarz, R. S., Teixeira, E. W., Tauber, J. P., Birke, J. M., Martins, M., Fonseca, M., et al. (2014). Honey bee colonies act as reservoirs for two symbiotic *Spiroplasma* species and incur complex, multiyear infection dynamics. *Microbiol. Open* 3, 341–355. doi: 10.1002/mbo3.172
- Spivak, M., and Reuter, G. S. (2001). Resistance to American foulbrood disease by honey bee colonies, *Apis mellifera*, bred for hygienic behavior. *Apidologie* 32, 555–565. doi: 10.1051/apido:2001103
- Teixeira, E. W., Chen, Y., Message, D., Pettis, J., and Evans, J. D. (2008). Virus infections in Brazilian honey bees. *J. Invertebr. Pathol.* 99, 117–119. doi: 10.1016/j.jip.2008.03.014
- Tentcheva, D., Gauthier, L., Zappulla, N., Dainat, B., Cousserans, F., Colin, M. E., et al. (2004). Prevalence and seasonal variations of six bee viruses in *Apis mellifera* L. and *Varroa destructor* mite populations in France. *Appl. Environ. Microbiol.* 70, 7185–7191. doi: 10.1128/AEM.70.12.7185-7191.2004
- Valera, F., Martín-Hernández, R., and Higes, M. (2011). Evaluation of large-scale dissemination of *Nosema ceranae* spores by European bee-eaters *Merops apiaster*. *Environ. Microbiol. Rep.* 3, 47–53. doi: 10.1111/j.1758-2229.2010.00186.x
- vanEngelsdorp, D., Evans, J. D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B. K., et al. (2009). Colony collapse disorder: a descriptive study. *PLoS ONE* 3:e6481. doi: 10.1371/journal.pone.0006481
- vanEngelsdorp, D., and Meixner, M. D. (2010). A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *J. Invertebr. Pathol.* 103, 80–95. doi: 10.1016/j.jip.2009.06.011
- Vayssier-Taussat, M., Moutailler, S., Michelet, L., Devillers, E., Bonnet, S., Cheval, J., et al. (2013). Next generation sequencing uncovers unexpected bacterial pathogens in ticks in western Europe. *PLoS ONE* 8:e81439. doi: 10.1371/journal.pone.0081439
- Ward, L., Waite, R., Boonham, N., Fisher, T., Pescod, K., Thompson, H., et al. (2007). First detection of Kashmir bee virus in the UK using real-time PCR. *Apidologie* 38, 181–190. doi: 10.1051/apido:2006072
- Welch, A., Drummond, F., Tewari, S., Averill, A., and Burand, J. P. (2009). Presence and prevalence of viruses in local and migratory honeybees (*Apis mellifera*) in Massachusetts. *Appl. Environ. Microbiol.* 75, 7862–7865. doi: 10.1128/AEM.01319-09
- Whitaker, J., Szalanski, A. L., and Kence, M. (2010). Molecular detection of *Nosema ceranae* and *N. apis* from Turkish honey bees. *Apidologie* 42, 174–180. doi: 10.1051/apido/2010045
- Williamson, D. L., Tully, J. G., and Whitcomb, R. F. (1989). “The genus *Spiroplasma*,” in *The Mycoplasma V: Spiroplasmas, Acholeplasmas, and Mycoplasmas of Plants and Arthropods*, eds R. F. Whitcomb and J. G. Tully (San Diego, CA: Academic Press), 71–111. doi: 10.1016/B978-0-12-078405-9.50011-2
- Yang, B., Peng, G. D., Li, T. B., and Kadowaki, T. (2013). Molecular and phylogenetic characterization of honey bee viruses, *Nosema* microsporidia, protozoan parasites, and parasitic mites in China. *Ecol. Evol.* 3, 298–311. doi: 10.1002/ece3.464
- Yoshiyama, M., and Kimura, K. (2011). Distribution of *Nosema ceranae* in the European honeybee, *Apis mellifera* in Japan. *J. Invertebr. Pathol.* 106, 263–267. doi: 10.1016/j.jip.2010.10.010
- Zander, E. (1909). *Tierische Parasiten als Krankheitserreger bei der Biene*. *Muenchener Bienenzeitung*. 31, 1961tung (Munich).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Tozkar, Kence, Kence, Huang and Evans. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read,
for greatest visibility



COLLABORATIVE PEER-REVIEW

Designed to be rigorous
– yet also collaborative,
fair and constructive



FAST PUBLICATION

Average 85 days from
submission to publication
(across all journals)



COPYRIGHT TO AUTHORS

No limit to article
distribution and re-use



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



SUPPORT

By our Swiss-based
editorial team



IMPACT METRICS

Advanced metrics
track your article's impact



GLOBAL SPREAD

5'100'000+ monthly
article views
and downloads



LOOP RESEARCH NETWORK

Our network
increases readership
for your article

Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on

