Exploring integument transcriptomes, cuticle ultrastructure, and cuticular hydrocarbons profiles in eusocial and solitary bee species displaying heterochronic adult cuticle maturation

Tiago Falcon[1,2*], Daniel G. Pinheiro[3], Maria Juliana Ferreira-Caliman[4], Izabel C. C. Turatti[5], Fabiano C. Pinto de Abreu[1], Juliana S. Galaschi-Teixeira[4], Juliana R. Martins[1], Moysés Elias-Neto[4], Michelle P. M. Soares[1], Marcela B. Laure[1], Vera L. C. Figueiredo[4], Norberto Peporine Lopes[5], Zilá L. P. Simões[4], Carlos A. Garófalo[4], Márcia M. G. Bitondi[4*]

[1]Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil;
[2]Núcleo de Bioinformática, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil;
[3]Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista "Júlio de Mesquita Filho", Jaboticabal, Brazil;
[4]Departamento de Biologia, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil;
[5]Departamento de Física e Química, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil.

*Corresponding authors:
E- mail: tiagofalconlopes@gmail.com (TF); mmgbit@usp.br (MMGB)

Abstract

Differences in the timing of exoskeleton melanization and sclerotization are evident when comparing eusocial and solitary bees. This cuticular maturation heterochrony may be associated with life style, considering that eusocial bees remain protected inside the nest for many days after emergence, while the solitary bees immediately start outside activities. To address this issue, we characterized gene expression using large-scale RNA sequencing (RNA-seq), and quantified cuticular hydrocarbon (CHC) through gas chromatography-mass spectrometry in comparative studies of the integument (cuticle plus its underlying epidermis) of two eusocial and a solitary bee species. In addition, we used transmission electron microscopy (TEM) for studying the developing cuticle of these and other three bee species also differing in life style. We found 13,200, 55,209 and 30,161 transcript types in the integument of the eusocial *Apis mellifera* and *Frieseomelitta varia*, and the solitary *Centris analis*, respectively. In general, structural cuticle proteins and chitin-related genes were upregulated in pharate-adults and newly-emerged bees whereas transcripts for odorant binding proteins, cytochrome P450 and antioxidant proteins were overrepresented in foragers. Consistent with our hypothesis, a distance correlation analysis based on the differentially expressed genes suggested delayed cuticle maturation in *A. mellifera* in comparison to the solitary bee. However, this was not confirmed in the comparison with *F. varia*. The expression profiles of 27 of 119 genes displaying functional attributes related to cuticle formation/differentiation were positively correlated between *A. mellifera* and *F. varia*, and negatively or non-correlated with *C. analis*, suggesting roles in cuticular maturation heterochrony. However, we also found transcript profiles positively correlated between each one of the eusocial species and *C. analis*. Gene co-expression networks greatly differed between the bee species, but we identified common gene interactions exclusively between the eusocial species. Except for *F. varia*, the TEM analysis is consistent with cuticle development timing adapted to the social or solitary life style. In support to our hypothesis, the absolute quantities of n-alkanes and unsaturated CHCs were significantly higher in foragers than in the earlier developmental phases of the eusocial bees, but did not discriminate newly-emerged from foragers in *C. analis*. By highlighting differences in integument gene expression, cuticle ultrastructure, and CHC profiles between eusocial and solitary bees, our data provided insights into the process of heterochronic cuticle maturation associated to the way of life.

Author Summary

From our previous observation that bees with distinct habits of life, eusocial and solitary, exhibit different degrees of cuticle melanization and sclerotization at the emergence, we decided to analyze the genetic signatures and ultrastructure of the integument, as well as the CHC profiles that could be involved in cuticle maturation. The expression patterns of certain genes involved in the melanization/sclerotization pathway, chitin metabolism, cuticle structure, and also regulators of cuticle renewal and tanning, in addition to other genes, might be grounded the slow process of cuticle maturation in the eusocial bees in comparison to the solitary ones. The electron micrographs revealed differences in the timing of cuticle deposition for the eusocial and solitary species. Among the identified CHCs, the proportions and quantities of n-alkanes in the developing cuticle are consistent with the faster cuticular maturation in the solitary bee, thus supporting our hypothesis.

Introduction

The exoskeleton (cuticle) enables arthropods to exploit a multitude of ecological habitats, and is central to the evolutionary success and worldwide expansion of insects. It is necessary for muscles attachment, for protection against predators, injuries, and pathogens [1]. In addition, its thickness is positively correlated with the resistance to some types of insecticides [2]. The exoskeleton is periodically shed and a new, larger one is formed, this characterizing the successive molting episodes that allow for insect growth and development. Its composition is defined by the secretion of products synthesized by the epidermis as well as by the uptake of molecules from other sources, for instances, hemolymph [3]. These products are used for cuticle renewal at each molting episode coordinated by changes in the titer of 20-hydroxyecdysone (20E), the active product of ecdysone hydroxylation. The Ashburner model postulated to explain 20E-induced chromosomal puffs in the larval salivary glands of *D. melanogaster* have ultimately led to the knowledge of molecular elements regulating molting and metamorphosis [4]. When 20E binds to the heterodimeric receptor consisting of EcR (Ecdysone receptor) and Usp (Ultraspiracle) proteins, its trigger a transcription factor regulatory cascade. Upstream elements of this cascade respond to the high 20E titer that also induces apolysis and initiates molting, whereas most downstream elements are only induced by the subsequent decrease in 20E titer. Binding sites for several of the transcription factors in this cascade were identified in many cuticular protein genes [5], suggesting that they, and other genes involved in cuticle remodeling [6, 7] are indirectly regulated by 20E.

The exoskeleton comprises an inner procuticle formed by layers of endocuticle and exocuticle, an outer epicuticle and the superficial envelope. The procuticle consists of a variety of proteins and chitin, a polymer of the glucose-derived N-acetylglucosamine. Chitin is a major compound in the insect exoskeleton [8]. Key enzymes in the chitin biosynthetic pathway starting from trehalose are the highly conserved chitin synthases that catalyze the transformation of UDP-*N*-acetylglucosamine to chitin. Chitin-modifying enzymes, specifically chitin deacetylases (Cdas), catalyze the conversion of chitin to chitosan, a polymer of β-1,4-linked d-glucosamine residues. Mutations in Cda genes are lethal to insect embryos, suggesting that these enzymes play critical roles during development, including the molting process [9]. Molting involves digestion of the actual cuticle, a process mediated by chitin-degrading enzymes, chitinases, which accumulate in the molting fluid [10]. The epicuticle does not contain chitin, but contains proteins and lipids and is rich in quinones, which are oxidized derivatives of aromatic compounds [11]. Together with chitin, the structural cuticular proteins constitute the bulk of insect cuti-

cle. Based on defining sequence domains, they have been classified into twelve families [12]. Proteins in the CPR family, with the largest number of members, contain the R&R Consensus [13, 14]. Some other structural cuticular proteins pertain to the Tweedle (Twdl) class [15], or were classified as Cuticular Proteins of Low Complexity – Proline-rich (CPLCP), Cuticular Proteins with Forty-four amino acid residues (CPF), Cuticular proteins analogous to peritrophins (Cpap), Glycine-Rich cuticular Proteins (GRP), and apidermins, among other classes. Some cuticular proteins, however, do not fill the features for inclusion in the pre-established classes. The main components of the envelope are the cuticular hydrocarbons (CHC) [16] that play roles in chemical communication (unsaturated CHC) [17, 18] and, together with other lipids, act as a barrier against insect desiccation by preventing water loss (mainly n-alkanes) [19, 17]. Key enzymes in CHC biosynthetic pathways occurring in the epidermis-associated oenocytes are the desaturases and elongases [20-22]. We previously determined gene expression profiles of six desaturases and ten elongases in the developing integument of *A. mellifera*, and correlated them with n-alkanes, methyl-alkanes, dimethyl-alkanes, alkenes and alkadienes quantification profiles [23]. Besides highlighting the CHC composition underlying envelope formation, these data provided clues to predict the function of these genes in CHC biosynthetic pathways.

In addition to chitin, cuticular proteins, CHCs, and other compounds, melanin pigments are crucial for the exoskeleton formation in insects. The chemical reactions in the core of the melanin biosynthetic pathway are evolutionary conserved. This pathway comprises the conversion of tyrosine into 3,4-dihydroxyphenylalanine (dopa) by the action of tyrosine hydroxylase (TH). Dopa is converted to dopamine, the primary precursor of insect melanin, via a decarboxylation reaction catalyzed by dopa decarboxylase (Ddc). Dopa or dopamine is further oxidized to dopaquinone or dopaminequinone, and finally these pigment precursors are converted into dopa-melanin or dopamine-melanin through reactions catalyzed by dopachrome conversion enzyme, a product of the *yellow* gene, and laccase2. Alternatively, dopamine is acetylated to N-acetyl-dopamine (NADA), and in conjugation with α-alanine originates N-β-alanyldopamine (NBAD). Both catechols are precursors for production of colorless and yellowish sclerotins [24, 25]. Thus, melanization occurs concomitantly to sclerotization through a shared biosynthetic pathway. Both processes are fundamental for the exoskeleton development [26], and are developmentally regulated by 20E [27, 28].

Among bees, we can distinguish the solitary and eusocial species. In the solitary species, every female constructs its own nest where it lay eggs, but does not provide care for the ecloded larvae. In contrast, the social organization is grounded on the division of labor between fertile queens and more

or less sterile, or completely sterile, workers that are engaged in nest construction and maintenance, besides caring for the queen's offspring [29, 30]. The search for genomic signatures of eusociality evolution in bees has grown since the publication of the *A. mellifera* genome [31] and gained force with the recent release of two *Bombus* species genomes [32] and the study of Kapheim *et al.* [33] comparing the genomes of ten bee species.

In this context, we draw our attention to the fact that bees greatly vary in the grade of cuticle melanization/sclerotization at the emergence time (adult ecdysis). In a previous study on the morphology of the developing adult cuticle [34], we observed that in eusocial bees, but not in the solitary ones, the process of cuticle melanization/sclerotization leading to cuticle maturation is extended to the adult stage. After emergence, workers from eusocial species (including the primitively eusocial bees from Bombini) spend some days performing inside nest activities, and during this period they stay protected in a safe and provisioned environment [35] where the hygienic behavior provides a certain level of immunity [36]. In contrast, the newly emerged solitary bees immediately leave the nest. Therefore, they need a fully mature cuticle to protect them in the external environment. This shift in the timing of cuticle maturation seems a case of heterochrony, which is defined as a change in the timing of development of a tissue or anatomical part relative to an ancestor, or between taxa [37]. If this assumption proves to be true, it can entail a link between the rate of cuticle maturation and the evolution of sociality in insects.

Here, we used the integument (cuticle and its subjacent epidermis) in an approach based on large-scale RNA sequencing (RNA-seq), transmission electron microscopy (TEM) and gas chromatography-mass spectrometry (GC/MS) to describe cuticle maturation in two eusocial bee species, *Apis mellifera* (Apini) and *Frieseomelitta varia* (Meliponini), and a solitary bee species, *Centris analis* (Centridini), the solitary lifestyle being considered the ancestral condition for bees [38]. TEM was also used for studying the ultrastructure of the cuticle of the primitively eusocial bee, *Bombus brasiliensis* (Bombini), the facultatively eusocial *Euglossa cordata* (Euglossini), and the solitary bee, *Tetrapedia diversipes*. This combined approach allowed us to compare the integuments at the morphological and molecular levels, besides highlighting differences that could be related to the heterochronic process of cuticle maturation. Among the genes expressed in the integument, we focused on those involved in the melanization/sclerotization pathway, chitin metabolism, genes encoding structural cuticular proteins, regulators of cuticle renewal and tanning, desaturase and elongase genes potentially involved in CHC biosynthesis, circadian clock genes that could determine the rhythm of

cuticle layers deposition [39, 40], and genes encoding pigments other than melanin.

The comparison of integument transcriptomes of three bee species at developmental points corresponding to adult cuticle formation, ecdysis, and at a mature age (foragers) gave us back the discovery of distinct genetic signatures of the integument, and highlighted differences in gene set expression profiles. The use of TEM and CHC analysis complemented these data by adding new information on cuticle ultrastructure and chemical profiles of its superficial layer, the envelope.

Results

Differential gene expression in the integument of *A. mellifera*, *F. varia* and *C. analis* during adult cuticle formation/maturation

We identified the expression of 13,200 genes in the developing integument of *A. mellifera*, and 55,209 and 30,161 contigs in the developing integument of *F. varia* and *C. analis*, respectively (S1 File). The data obtained from the three biological samples of each developmental phase, Pbm (pharate adult), Ne (newly-emerged) and Fg (forager) of each bee species, in a total of 27 transcriptomes, were used in Pearson correlation analysis in order to check reproducibility. A hierarchical clustering on pairwise correlation is shown in S1 Fig. In general, the samples of the same developmental phase (biological triplicates) joined together, indicating that they are more similar to each other than to samples of the other developmental phases. As expected, for the three bee species, the least correlated samples were those originated from the Pbm and Fg integuments. When filtering these data sets for the genes (DEGs) or contigs (DECs) differentially expressed between the developmental phases, we found 3,184 DEGs for *A. mellifera*, 5,959 DECs for *F. varia* and 2,543 DECs for *C. analis*, representing 24.1%, 10.8%, and 8.4% of the identified genes, respectively. Fig 1 shows the number of genes that were upregulated in the comparisons between the developmental phases of each of the three bee species. In *A. mellifera*, 14.8% and 17.8% of the DEGs were upregulated in the Pbm phase in comparison to the Ne and Fg phases, respectively; 20.9% and 7.8% were upregulated in Ne in comparison to Pbm and Fg; 24.6% and 10.4% were more expressed in Fg than Pbm and Ne. In *F. varia*, 21.1% and 39.3% of the DECs were upregulated in Pbm compared to the Ne and Fg phases, respectively; 27.9% and 21.1% DECs were more expressed in Ne than in Pbm and Fg; 39.6% and 16.0% showed higher expression in Fg than in Pbm and Ne. In *C. analis*, the Pbm phase showed higher

expression for 39.2% and 31.1% of the DECs in comparison to the Ne and Fg phases; 32.7% and 6.1% of the DECs were upregulated in Ne in comparison to Pbm and Fg; 31.9% and 3.6% were more expressed in Fg than in Pbm and Ne. These proportions of upregulated genes significantly vary in the comparisons of the developmental phases of each bee species and also between the bee species. In addition, the proportions of genes upregulated in the adult phases (Ne *versus* Fg) were significantly lower in the solitary *C. analis* than in the eusocial *A. mellifera* and *F. varia* bee species (z test, $p \leq 0.001$, except for one of the comparisons where $p = 0.014$).

To make more comprehensive the RNA-seq analysis of the integument, we searched the Gene Ontology (GO) functional terms for all *A. mellifera* DEGs and all *F. varia* and *C. analis* DECs. The GO annotations for Molecular Function, Cellular Component and Biological Process categories are described in S3 File. We then extracted from this analysis the functional terms more evidently related to cuticle development (Fig 2). Structural molecule activity, chitin-binding, and chitin metabolic process were categories overrepresented in the younger phases, i.e., the Pbm and Ne phases of the three bee species. Structural constituent of cuticle, structural constituent of chitin-based cuticle, and other cuticular components-related GO categories also included genes more expressed in the Pbm and Ne integuments of both, or one of the eusocial species. Functional categories related to the epidermis, which is the tissue responsible for secreting the cuticle, specifically epithelium development, epithelial cell differentiation/development, cell adhesion, cell junction organization/assembly, among other categories, were also more represented in the younger Pbm and Ne bees, but only of the eusocial species. For the three bee species, the DEGs and DECs more related to the functionality of the integument of newly-emerged (Ne) and forager bees (Fg) (here named older phases for simplification), were included in the following overrepresented GO terms: fatty acid biosynthetic process, lipid biosynthetic process, organic acid biosynthetic process, and carboxylic acid biosynthetic process. These terms and others overrepresented in the older Ne and Fg phases of *F. varia* and *C. analis*, i.e., very-long-chain fatty acid metabolic process, and fatty acid metabolic process, could be tentatively related to CHC biosynthetic pathways. For *F. varia* and/or *C. analis*, functional terms related to pigmentation pathways (pigmentation, pigment metabolic process, pigment biosynthetic process, pigmentation during development, and terms related to eye pigments), were also significantly more represented in the Ne and Fg phases. These GO results (Fig 2) evidenced the similarities and differences in terms of cuticle-related functional attributes between the developmental phases and bee species. Some functional categories were shared by the three bee species, and a larger number of categories were

shared by the two eusocial species than by one of them and *C. analis*.

S2 File specifies the genes upregulated between the developmental phases and bee species. Among the DEGs and DECs, it was clear that those encoding structural cuticular proteins, such as those in the CPR, Twdl, and Cpap families, and also chitin-related genes with roles in chitin metabolism, modification and degradation, were upregulated in the Pbm and/or Ne phases of the three bee species here studied. A series of sequences containing the chitin-binding peritrophin A domain were similarly overrepresented in the integument of the Pbm and/or Ne phases of *F. varia* and C. *analis*, thus being candidates to participate as structural proteins or enzymes in cuticle formation. In contrast, genes encoding odorant-binding proteins that bind to pheromones thus serving as insect chemoreceptors, as well as genes encoding a variety of CYPs (cytochrome P450), and antioxidant proteins like glutathione-S-transferase (GST), glutathione peroxidase (GTPx), thioredoxin peroxidase (TPX), and superoxide dismutase (SOD), were more expressed in the mature integument of foragers of the three bee species. Transcripts for genes related to the activity of juvenile hormone (JH), which is produced in a greater quantity in foragers [41], specifically *Krüppel homolog 1* (*Kr-h1*), and *JH-esterase* (*jhe*), were found in higher levels in the Fg integument of *F. varia* and *C. analis* than in the younger phases; transcripts for a JH-inducible (JHI-1) protein were overrepresented in the Fg and Ne integuments of *A. mellifera* in comparison to the Pbm integument. The Fg integument of *C. analis* showed a higher expression of the *ecdysone receptor* (*EcR*) and *seven-up*, an orphan nuclear receptor belonging to the steroid receptor gene superfamily [42]; *seven-up* is also overexpressed in the Fg integument of *F. varia*. Defense response genes (*defensin, apidaecin*) were also highly expressed in the Fg integument of *A. mellifera* (S2 File). Such developmental differences in gene expression in the integument reflect the dynamics of cuticle formation and acquisition of its functionality in adult bees

Distance correlation analysis based on the RNA-seq data is consistent with the earlier cuticle maturation in the solitary *C. analis* in comparison to the eusocial *A. mellifera*

We used the DEGs and DECs in a distance correlation analysis in order to measure the clustering potential of the studied developmental phases of each bee species (Fig 3). This strategy allowed us to know for each of the bee species how near, or distant from each other are the Pbm, Ne and Fg developmental phases in terms of gene expression levels/patterns in the integument. Assuming that the cuticle of solitary bee species is sufficiently mature at the emergence, the hypothesis approached here was that the integument samples of the Ne and Fg phases of *C. analis* would cluster

together, and separately from the Pbm samples. In contrast, in the eusocial species, the Pbm and Ne samples would group together, with the Fg samples forming a more distant group. Indeed, the results of the distance correlation analysis using all the *C. analis* DECs and *A. mellifera* DEGs were consistent with this hypothesis. In terms of differential gene expression in the integument, the Ne and Fg phases are nearest to each other in the solitary bee than they are in the eusocial *A. mellifera.* However, in *F. varia*, the distance correlation analysis grouped the Ne and Fg phases in a statistically supported cluster, in spite of the very distinct cuticle melanization patterns and hardness that they exhibit.

Gene expression profiles in the integument of the eusocial (*A. mellifera* and *F. varia*) and solitary (*C. analis*) bee species

Heatmaps representing the expression profiles of classes of cuticle-related genes through the Pbm, Ne and Fg developmental phases were constructed and clearly showed differences between the bee species (Fig 4).

We found in the RNA-seq libraries seven genes involved in the biosynthesis of melanin and sclerotizing compounds (see a representation of the melanin/sclerotin biosynthetic pathway in Shamim *et al.* [24]. The genes with roles in the melanization/sclerotization pathway, except for *Dat*,-were more expressed in the younger phases (Pbm and/or Ne) of *A. mellifera*. Similarly, these genes, including *Dat*, were more expressed in the younger phases of *F. varia*. In contrast, in *C. analis*, the majority of the genes in this class (*tan*, *Ddc*, *Lac2*, *yellow-y*) did not significantly change their expression levels, *ebony* was highly expressed in the Ne and Fg phases, and the expression profile of *Dat* also differed from both eusocial species. *TH* was the only gene in this class showing a significantly higher expression level in the very same developmental phase (Ne) of the three bee species.

Searching for genes related to pigmentation pathways other than the melanin biosynthetic pathway in the integument RNA-seq libraries, such as those genes involved in pterin, ommochromes, and heme formation, we found 17 genes in *A. mellifera*, and 18 genes in *F. varia* and also in *C. analis*, including *cardinal*, *scarlet, brown*, *vermillion*, *light*, *sepia*, and *henna* (this one involved in both biopterin formation, and tyrosine formation for the melanization process), thus indicating that their products are necessary in the adult cuticle. We also observed that a higher proportion (66.7%) of these genes displayed higher expression levels in the adults (Ne, Fg, or both phases) of *F. varia* in comparison to *A. mellifera* (29.4%) and *C. analis* (27.8%).

Concerning genes encoding chitin-related enzymes, we found 17, 16 and 33 of these genes in *A. mellifera*, *F. varia* and *C. analis*, respectively. The four *Cda* genes (*Cda4*, *Cda5*, *verm* and *serp*) found in the eusocial species, and five (*Cda4-like*, *Cda-like-1*, *Cda-like-2*, *verm*, *serp*) of the six *Cda* genes found in *C. analis* showed the higher expression in the Pbm, Ne, or both developmental phases, the other *C. analis Cda* gene (*Cda5-like*) showed no significant expression levels variation throughout the developmental phases. One of the two *ChS* genes found in *A. mellifera* (*kkv*) and *F. varia* (*ChS6*), and two of the four *ChS* genes found in *C. analis* (*ChS-kkv-like-1*, *ChS6-like1*) were also more expressed in the Pbm and/or Ne phases whereas the other *ChS* genes of the three bee species did not show significant expression level variation. Six (*Cht-like2*, *Cht-2 like*, *chitooligosaccharidolytic-domain-like*, *Cht5*, *chitotriosidase*, *Cht3*) of the eleven *Cht* genes of *A. mellifera*, and four (*Cht-like1*, *Cht-like*, *chitooligosaccharidolytic-domain-like*, *Cht2-like2*) of the ten *Cht* genes of *F. varia* were also more expressed in the Pbm and/or Ne phases, the remaining showing no significant variation in expression levels, except for the chitinase-encoding gene, *Idgf-4*, which is significantly more expressed in *A. mellifera* foragers. In contrast, only a small number (*Cht-like12*, *Cht-like4*, *Cht-like10*, *Cht-like1*) of the 22 *Cht* genes of *C. analis* were more expressed in these phases, the remaining showing no significant changes in expression levels, except for *Idgf-4*, which is more expressed in foragers.

The majority of the CPR genes (encoding cuticle proteins containing the RR1 or RR2 Consensus types) in the eusocial species showed significant variation in expression levels through the studied developmental phases, the proportions of RR1 and RR2 genes showing variable expression corresponding to 94.1% and 80.9% in *A. mellifera*, and 66.7% and 70.6% in *F. varia*, respectively. In contrast, lower proportions of RR1 and RR2 genes in *C. analis*, corresponding to 40% and 35% respectively, showed significant variation in transcript levels. For the three bee species, most of the genes showing changing transcript levels, in the range of 75 to 100%, were more expressed in the Pbm or both Pbm/Ne phases. Interestingly, a few CPR genes were significantly more expressed in the Ne phase (*AmCPR19*, *AmCPR27*, *AmSgAbd1-like*, *FvUnCPR-1*), or in both Ne and Fg phases (*CaSgAbd1-like* and *AmUnCPR-RR2-5*), and only a CPR gene, the RR1 motif *AmCPR13* gene, showed a higher expression exclusively in foragers. Similarly, a higher proportion of the non-RR cuticular protein genes showed significant transcript levels variation in *A. mellifera* and *F. varia*, 90.6% and 72.2% respectively, in comparison to *C. analis* (64.5%). These genes were also mostly more expressed in the Pbm or Pbm/Ne phases of the three bee species. However, like some CPR genes, there were non-RR genes displaying the highest expression in adults (Ne, Fg or both phases), specifically, *Apd* genes in *F.*

varia (*FvApd-1*) and *C. analis* (*CaApd-1* and *CaApd-2*), and *Cpap* genes in *C. analis* (*CaUnCpap-3*, *CaUnCpap-4*, *CaUnCpap-9, CaCpap3-e*).

For the three studied bee species, a higher proportion of genes encoding elongases (Elo-genes) and desaturases (Desat-genes) putatively involved in CHC biosynthesis were more expressed in adults (Ne, Fg, or both phases) than in the Pbm phase. However, in *C. analis*, a higher proportion (66.7%) of these genes increased significantly their expression levels from the Pbm to the Ne phase in comparison to *F. varia* (26.7%) and *A. mellifera* (39.1%).

A higher proportion of the regulatory genes was significantly more expressed in the Pbm phase of *A. mellifera* (50%) and *F. varia* (28.6%) than in *C. analis* (4.5%) in which the majority of the genes (72.7%) did not show significant difference in expression levels between the developmental phases. Some regulatory genes had a higher expression in adults (Ne, Fg or both phases) of *A. mellifera* [*Ammirr* (*mirror*), *AmUsp* (*Ultraspiracle*), *AmCCAP* (*Crustacean Cardioactive Peptide*), *AmKr-h1* and *Amhairy*], *F. varia* (*FvKr-h1*), and *C. analis* (*CaKr-h1*, *Camirr*, *CaE75*, *CaEcR* and *Cahairy*). Two of the regulatory genes in *A. mellifera*, *AmE75* and *AmMblk* (*Mushroom body large type Kenyon cell specific protein-1* or *E93-like*), which were highly expressed in the younger Pbm phase, were also highly expressed in the older Fg phase.

Four among the seven circadian rhythm genes of *A. mellifera* [*Clk* (*Clock*), *Cry* (*Cryptochrome*), *Per* (*Period*) and *Tim2* (*Timeless2*)] showed the highest expression in the Pbm phase. This is in contrast to the majority of the circadian rhythm genes in *F. varia* [*Clk*, *Per*, *Pdp1* (*Par domain protein I*), *Vri* (*vrille*), *Tim2*] and *C. analis* (*Per*, *Vri*, *Cry*, *Clk*, *Tim2*), which did not significantly change their expression levels. The genes *Vri*, *Cyc* (cycle), and *Pdp1* in *A. mellifera*, *Cyc* in *F. varia*, and *Cyc* and *Pdp1* in *C. analis* showed the highest expression in adults (Ne, Fg or both phases).

In summary, the main differences between the social and solitary bee species were highlighted in the heatmaps (Fig 4) displaying integument gene expression profiles: **(a)** A higher proportion of genes involved in the melanization/sclerotization pathway, cuticle formation (RR1, RR2, and non-RR genes), and regulation (regulatory genes) showed significant transcript levels variation through the studied developmental phases of *A. mellifera* and *F. varia* in comparison with *C. analis*. Most of these genes showing transcript levels variation were more expressed in the Pbm or Pbm/Ne phases. In *C. analis*, the higher proportion of genes displaying no differences in expression levels through the studied phases, were possibly highly expressed earlier, before the Pbm phase, for faster cuticle formation and

maturation, but this assumption requires further investigations; **(b)** The number of chitin-related genes, higher in *C. analis*, and not their expression patterns, distinguished this species from the eusocial *A. mellifera* and *F. varia*; **(c)** A higher proportion of desaturase and elongase genes putatively involved in CHC biosynthesis showed significantly increased expression levels at the emergence (Ne phase) of *C. analis* in comparison to the eusocial ones, which is consistent with an accelerated process of cuticle maturation in the solitary bee.

Importantly, all the gene classes here studied included representatives showing increased or high expression levels in the mature integument of foragers indicating that the mature cuticle is a dynamic structure requiring structural and regulatory elements for its maintenance.

Correlation among expression profiles of genes candidates to play roles in cuticle formation/ maturation in the eusocial (*A. mellifera* and *F. varia*) and solitary (*C. analis*) bees

We used Pearson's correlation in order to measure the strength of the linear association between the expression profiles of 119 genes related to cuticle development and maturation shown in Fig 4, which shared potential orthology relationships between the bee species. A fraction of these ortholog genes showed non-significantly correlated transcript levels fluctuation among the bee species, thus highlighting peculiarities in cuticle development for each species. However, 76 orthologs (S1 Table**;** Fig 5) displayed expression profiles significantly correlated at least between two of the three bee species. Importantly, the expression profiles of 21 among these 76 genes were positively correlated between the eusocial species, and negatively or non-correlated with the solitary bee ($r \geq 0.6$ and $p \leq 0.1$). In addition, other six genes, whose transcripts were not identified in *C. analis*, showed expression profiles positively correlated between the eusocial species. Therefore, these 27 genes are possibly contributing to differences in the processes of cuticle development and maturation in the eusocial bees versus the solitary bee. Thus, the expression profiles of genes related to the melanization/sclerotization pathway (*ebony, tan*) and chitin metabolism [*Idgf4-like, Cda5* (*Chitin deacetylase 5*), *chitooligosaccharidolytic-domain-like*], genes encoding cuticular structural proteins containing the RR1 or RR2 domains (*CPR14, CPR17, CPR18, CPR23, CPR25, CPR26*), or lacking these domains (*Apd-3, Apd-like*), and also genes in CHC pathways (*Desat-GB40659*, *Elo-GB54401*, *Elo-GB54302*, *Elo-GB45596, Elo-GB46038*), regulators of cuticle development [*Ethr* (*Ecdysis triggering hormone receptor*), *E74, Hr4* (*Hormone receptor 4*), *Hr38* (*Hormone receptor 38*), *FTZ-F1* (*Fushi tarazu-factor*

1), *rickets, Ptx-1* (bicoid-related *Paired-type homeobox gene D*), circadian rhythm genes (*Tim2*) and a gene in the non-melanin pigmentation pathways, *ALA*S (*δ-aminolevulinic acid synthase*)], suggest roles in the differential cuticle development in the solitary versus eusocial bees.

Among the above cited 76 orthologs, we also found genes whose expression profiles were positively correlated between the solitary and eusocial bees. Thus, the following 23 genes shared expression profiles positively correlated between *A. mellifera* and *C. analis*: *yellow-y* (melanization /sclerotization pathway), *Cda4* and *ChS-kkv-like1* (chitin metabolism), *SgAbd2-like* and *97Ea-like* (CPR-RR1 class), *CPR10* (RR2 class), *Twdl(Grp)*, *Cpap3-a*, *Cpap3-b* and *Cpap3-c* (non-RR class), *Desat-GB48195*, *Desat-GB45034*, *Desat-GB42217*, *Elo-GB51249* and *Elo-GB54404* (CHC pathways), *Usp*, *CCAPR*, *Mirr* and *hairy* (regulatory genes), *Cyc* (Circadian rhythm), *verm*, *sepia* and *pinta-like* (other pigmentation biosynthetic pathways than melanin). Similarly, the following 12 genes shared expression profiles positively correlated between *F. varia* and *C. analis*: *Cda4* and *ChS6* (chitin metabolism), *Cpap3-c* (non-RR class), *Elo-GB54404* (CHC pathways), *Bursβ*, *CCAP* and *E75* (regulatory genes), *Cyc* (Circadian rhythm), *verm-like*, *cardinal-like*, *light* and *scarlet* (other pigmentation biosynthetic pathways than melanin).

Co-expression networks reconstructed with genes related to cuticle development and maturation, and common interactions between the networks of the eusocial *A. mellifera* and *F. varia* bees

The genes related to cuticle formation and maturation in *A. mellifera*, *F. varia*, and *C. analis* were separately used for co-expression networks reconstruction (S2-S4 Figs). The gene co-expression networks for the eusocial species, *A. mellifera* and *F. varia*, showed common interactions among regulatory elements [*FTZ-F1*, *E74*, *Hr4*, *Hr46* (*Hormone receptor 46*)], genes encoding structural cuticular proteins (CPR14, CPR17, CPR23, CPR24, CPR25, Apd-3 and Apd-like), and encoding the elongase Elo-GB54302, Cdas [*verm* (*vermiform*), *serp* (*serpentine*), *Cda5*], and Lac2 (Fig 6). However, by intersecting the gene co-expression networks of the eusocial *A. mellifera* and the solitary *C. analis*, we found only one common interaction comprising the genes *yellow-y* and *Cpap3-a*. Similarly, only the interactions between *CPR16* and *Eh-like* (*Eclosion hormone*), and *tan*/Elo-GB45596 were highlighted as being common to the eusocial *F. varia* and the solitary *C. analis* after superimposing their respective gene co-expression networks.

Ultrastructure and thickness of the developing adult cuticle shows conspicuous differences among the eusocial, primitively eusocial, facultatively eusocial, and solitary bee models

The morphology of the developing adult cuticle is shown for the eusocial *A. mellifera* and *F. varia* bees, for the primitively eusocial *Bombus brasiliensis*, for the facultatively social *Euglossa cordata*, and for two solitary bees, *C. analis* and *T. diversipes* (Fig 7). For *A. mellifera*, there were no noticeable modifications in cuticle ultrastructure from the pharate-adult phase (Pbm) to 48h after emergence. Up to this time, only the exocuticle was deposited. At 72h, endocuticle layers became apparent in the micrographs (Fig 7A). Cuticle ultrastructure was very similar in 96h-aged *A. mellifera* bees and foragers (Fig 7A). We then measured the thickness of the cuticle in seven time points of *A. mellifera* development (Fig 7A'). As the cuticle measurements in the groups of bees aging 0h to 96h post-emergence, and in the group of foragers, did not show a normal distribution (Shapiro-Wilk normality test, $p = 0.0074$) we used the Kruskal-Wallis test associated with the *post hoc* Conover-Iman test and Bonferroni correction to compare the sample collection data. Foragers have a significantly thicker cuticle in comparison to the earlier developmental phases, i.e., the Pbm phase, and bees at 0h, 24h and 48h after emergence (Fig 7A'). At 72h and 96h post-emergence, cuticle measurements values did not significantly differ from foragers. Differently, the cuticle of the eusocial *F. varia* showed very little variation in morphology (Fig 7B), and no significant variation in thickness (Fig 7B') from the Pbm phase to the forager time. For the solitary species, *C. analis*, we observed remarkable differences in cuticle ultrastructure (Fig 7C) and thickness (Fig 7C') between the Pbm and Ne phases, whereas the cuticles of the Ne and Fg phases were very similar. Pore canals are abundant in the Pbm cuticle of *C. analis*. At the Ne and Fg phases, the *C. analis* cuticle can be described as a succession of lamellae, the most superficial ones, i.e., those first deposited, became thicker and reached a higher degree of differentiation (Fig 7C). Like *C. analis*, the cuticle of *B. brasiliensis* (Fig 7D, 7D'), *E. cordata* (Fig 7E, 7E'), and *T. diversipes* (Fig 7F, 7F'), did not show noticeable ultrastructural changes, or statistically significant thickness differences, from the emergence (Ne phase) to the forager time (Fg phase).

Together, these data indicate that cuticle deposition in the solitary species, *C. analis* and *T. diversipes*, and the primitively and facultatively eusocial species, *B. brasiliensis* and *E. cordata*, respectively, is completed or almost completed at the time of adult emergence. In contrast, in *A. mellifera*, the endocuticle was deposited only after the emergence. Surprisingly, the cuticle of the eusocial *F. varia* species did not undergo significant variation in ultrastructure and thickness from the

Pbm to the Fg phases, although a great increase in pigmentation and sclerotization has been clearly noticed in *in vivo* observations.

Cuticular n-alkanes mark the earlier cuticle maturation in the solitary *C. analis* compared to the eusocial *A. mellifera* and *F. varia* bee species

The CHC composition of the superficial cuticle layer, the envelope, was determined for *A. mellifera*, *F. varia* and *C. analis* as another strategy potentially able to uncover differences that could be associated to the cuticle maturation heterochrony. The proportion of CHCs in the chromatograms, the significance level of each peak and the contribution of these peaks for discriminating the developmental phases of the eusocial and solitary species are shown in S4 File. The Euclidean distance clustering analysis applied to the total CHC quantification data clearly discriminated the Fg phase from the earlier Pbm and Ne phases in the eusocial bees, *A. mellifera* and *F. varia*, as well as in the solitary *C. analis* (S5 Fig). Total CHC quantification data grouped together the Pbm and Ne samples of *A. mellifera* (AU=100; BT=100), *F. varia* (AU= 100; BT=100), and *C. analis* (AU=96; BT=88). For *F. varia*, the group including Ne samples showed AU=95 and BT=87, which is a moderate to high BT value usually associated with Bayesian posterior probabilities ≥ 95% [43]. The same was verified for the *F. varia* Pbm samples (AU=94; BT=87). For the two eusocial species, the Fg samples grouped with maximal AU (100) and BT (100) values. For *C. analis*, however, these values were significantly lower (AU=78; BT=57) (S5 Fig).

When we analyzed separately the CHC classes, n-alkanes discriminated the *A. mellifera* foragers (Fg) (AU=94; BT=84) from the Pbm and Ne developmental phases, which were clustered together (AU=85; BT=77). As in *A. mellifera*, n-alkanes also discriminated the *F. varia* foragers (Fg) as a separate group (AU=97; BT=76), and the Ne and Pbm phases were clustered together (AU=98; BT=77). However, the n-alkanes did not significantly distinguish the developmental phases of *C. analis* (S5 Fig).

The unsaturated CHCs data from *A. mellifera* did not give us back a strong support for distinguishing the developmental phases. Although all the Ne samples and the majority of the Pbm samples have been grouped with a high AU value (99%), the BT=1 value was low. Three of the *A. mellifera* foragers (Fg) escaped from the main cluster formed by twelve foragers (AU=96; BT=23). In contrast, the unsaturated CHCs discriminated each of the developmental phases of *F. varia*. The groups

of Pbm samples (AU=99; BT=93) and Ne samples (AU=99; BT=94) were maintained together in a larger cluster (AU=99; BT=98), and separately from the group of Fg samples (AU=99; BT=97). This CHC class clustered together the Pbm and Ne samples of *C. analis* (AU=96; BT=80). The Fg samples of *C. analis* were separated into two main clusters, respectively supported by AU=94; BT=70 and AU=93, BT=83 (S5 Fig).

Branched CHCs from *A. mellifera* clearly clustered the Fg samples (AU=97; BT=94). The Ne and Pbm phases were joined together in a single well-supported group (AU=100; BT=100). In *F. varia*, separation of Fg from the earlier phases was not clear: three of the fifteen Fg samples joined to the group encompassing the Pbm and Ne samples, this group being supported by 98% AU, but showing a low BT value (BT=3). The *F. varia* forager samples were also clustered with low BT values. In the solitary *C. analis*, the branched CHCs clustered six of the seven Fg samples into a single group (AU= 97; BT=72), and all the Ne samples plus two of the four Pbm samples were clustered together in another group supported by AU=99, but presenting a low BT value (BT=39) (S5 Fig).

These data on the Euclidean distance based on the relative quantification of CHCs was contrasted with the results on the absolute quantification of CHCs (CHC µg per bee) (Table 1; S4 File). Table 1 shows that Fg bees of the eusocial species have significantly higher quantities of n-alkanes than the Ne and Pbm bees, which is not true for *C. analis*. In addition, absolute quantification of unsaturated CHCs also distinguished the foragers from the earlier developmental phases of *A. mellifera*, but not of *C. analis*. For *F. varia*, the mass of unsaturated compounds could not be quantified due to their very low quantities.

In summary, the Euclidean distance analysis based on the relative quantifications of n-alkanes, as well as the absolute quantifications of n-alkanes and unsaturated CHCs, were consistent with the hypothesis of interdependence between cuticle maturation timing and the eusocial/solitary ways of life. These analyses distinguished the foragers from the younger bees, but only in *A. mellifera* and *F. varia*, this being interpreted as the cuticle achieving its complete maturation tardily in the eusocial species, whereas the solitary bee emerges with an already mature cuticle.

Table 1. Absolute quantification of n-alkanes and unsaturated CHCs in the cuticle of eusocial and solitary bee species. Developmental phases are indicated: Pbm (pharate-adults), Ne (newly emeged bees), Fg (foragers). Means and standard deviations (STD) of 3 samples (N=3) per developmental

528 phase. Different lowercase letters in the Sig (statistical significance) column indicate difference
529 between the developmental phases of each species.

N-alkanes				
A. mellifera				
Phase	Mean	±	STD	Sig.
Pbm	9.35965938	±	2.89275421	a
Ne	7.36272669	±	1.44535498	a
Fg	18.2360314	±	4.35877417	b
F. varia				
Phase	Mean	±	STD	Sig.
Pbm	1.63324436	±	0.15390427	a
Ne	3.40142407	±	1.35387231	a
Fg	9.28337077	±	3.03839358	b
C. analis				
Phase	Mean	±	STD	Sig.
Pbm	6.56549413	±	1.62457012	a
Ne	14.8349947	±	0.32610609	b
Fg	13.3273848	±	5.07830924	ab
Unsaturated CHC				
A. mellifera				
Phase	Mean	±	STD	Sig.
Pbm	0.60042039	±	0.17210242	a
Ne	0.92421769	±	0.09047864	a
Fg	6.5543118	±	2.38207067	b
C. analis				
Phase	Mean	±	STD	Sig.
Pbm	9.24954719	±	2.48578756	a
Ne	19.242	±	2.40516304	ab
Fg	28.3380901	±	11.855003	b

Standard deviation (STD). Different red letters in significance (Sig.) column represent statistical significance (ANOVA associated to Tukey's HSD post hoc test) between developmental phases of a species.

531

532

533 **Discussion**

534

535 The RNA-seq analysis revealed the set of genes expressed in the integument of three bee

species, and also the changes in gene expression as the adult cuticle is deposited and differentiates in a mature and fully functional cuticle. For *A. mellifera*, for which we have the sequenced genome, the genes expressed in the integument represented 95.07% of the genes in the released genome assembly version 4.5. Similar proportions will likely be found for *F. varia* and *C. analis* in the near future, after the sequencing of their respective genomes. Selected genes with potential roles in cuticle formation and maturation were characterized in terms of differential expression profiles. Co-expression networks were reconstructed. In parallel, we examined the ultrastructure of the developing adult cuticle of bee species. Furthermore, the CHC composition of the envelope, the less known cuticle layer, was also characterized. Our data expanded the knowledge on the insect integument. It is our expectation that the obtained data provide a valuable resource for future studies on exoskeleton formation and maturation in arthropods.

Expression profiles of cuticle-related genes may significantly differ during adult cuticle formation/maturation, and among bee species.

Genes involved in adult cuticle formation in *A. mellifera* in general show higher expression soon after the ecdysteroid titer peak that signalizes pupal cuticle apolysis and the beginning of the pharate-adult stage [44, 45]. Consistently, the majority of the integument genes showing expression levels variation in the three bee species, and identified as playing roles in cuticle melanization/sclerotization, cuticle structure (RR1, RR2, and non-RR genes), and regulation of the molting events (regulatory genes), displayed a higher expression in pharate-adults (Pbm phase), sometimes extending their higher expression up to the emergence time (Ne phase). However, we found genes, including those related to melanization/sclerotization and other pigmentation pathways, and also genes related to chitin metabolism, and structural cuticle protein genes, which showed the highest expression later, at emergence (Ne phase), and even in foragers (Fg phase), suggesting that their products are incorporated into the mature cuticle. Moreover, all transcripts identified in higher quantities during cuticle formation in pharate-adults were also identified in the newly emerged and forager bees, although in lower quantities. Their products may be involved in adult cuticle maintenance. Our gene expression findings indicate that the structure of the mature cuticle entails a dynamism, which has been up to now mainly characterized in studies on CHC composition of its most superficial layer, the envelope [23, this work].

Among the genes identified in the RNA-seq analysis of the integument, we focused on classes

of genes playing roles in cuticle formation and maturation, such as those below discriminated.

Genes related to cuticle pigmentation and sclerotization

The expression patterns of the first gene in the pigmentation/sclerotization biosynthetic pathway, *TH*, were positively correlated between *A. mellifera*, *F. varia* and *C. analis*, and apparently, *TH* does not contribute to the differential timing of cuticle pigmentation among them. Lower levels of *TH* transcripts were verified for the forager bees of the three bee species, which is consistent with the reported reduction in *TH* transcripts levels in *T. castaneum* [46, 47] and *Diacamma* sp [48] following the emergence. However, the expression patterns of *ebony* and *tan*, whose protein products act in a reversible reaction between dopamine and NBAD sclerotin [49], were positively correlated exclusively between the eusocial species, thus differentiating these species from the solitary one. The expression profiles of the remaining genes in the melanization/sclerotization pathway, including the *Lac2* gene previously characterized in *A. mellifera* [50], did not show such correlation patterns. Interestingly, *Dat* showed significantly increased expression in the mature cuticle of *A. mellifera* foragers, which is an uncommon pattern for genes in the melanization/sclerotization pathway.

We also observed that in general, the genes involved in the biosynthesis of other pigments except melanin displayed a higher expression levels in adults (Ne, Fg, or both phases) of *F. varia*, which may be tentatively interpreted as these genes playing roles in the process of post-ecdysial cuticle pigmentation in this bee species. Two of these genes, *cardinal* and *scarlet,* are both necessary for ommochromes formation in *B. mori* [51], and are associated to the formation of red and brown pigments [52]. The expression profiles of *light*, which is required for pigment granules formation [53], were positively correlated in *F. varia* and *C. analis*, and might be related to the brownish and reddish color pattern typical of the cuticle of these two species. The expression profiles of the gene encoding *ALAS*, which catalyzes the first enzymatic step in heme biosynthesis, were positively correlated exclusively between the eusocial species, *F. varia* and *A. mellifera*. *ALAS* might be involved in detoxification, as suggested for *D. melanogaster* [54, 55], and in prevent dehydration [56]. Interestingly, in contrast to the eusocial bees, the expression of *ALAS* is higher in the Pbm phase of *C. analis*, which may suggests that mechanisms of protection against cuticle dehydration develop anticipatedly in the solitary species.

Genes involved in chitin synthesis, modification and degradation

598 In insects, *Cht*, *Cda*, and *ChS* genes have been described as highly expressed during cuticle
599 renewal at the pharate-adult development [8, 57-62]. This was also observed in the bee species here
600 studied, but with variations: the expression of a putative chitinase, *Idgf4-like*, increased in newly-
601 emerged *C. analis*, and like reported for *T. castaneum* [63], this may be important for the transition to
602 the adult stage. In *A. mellifera* and *F. varia*, the expression of *Idgf4-like* is high in foragers, supporting
603 roles in the mature adult cuticle. Therefore, the decay in the expression of chitinase genes in adult
604 insects seems not a standard pattern. Concerning the *Cda* genes, in *Drosophila*, they have a strict
605 relationship with the mechanical properties of the exoskeleton [64], and this might be true for the *Cda*
606 genes expressed in the integument of the bee species. The other class of chitin-related genes encodes
607 ChS enzymes, which catalyze the last step in the chitin biosynthetic pathway and have been implied in
608 the synthesis of epidermal cuticle in *T. castaneum* [65]. A *ChS* gene, *CS-1*, also called *krotzkopf*
609 *verkehrt* (*kkv*), is required for procuticle formation, stabilization of the epicuticle, and attachment of the
610 cuticle to the epidermis in *D. melanogaster* [66]. We found a *kkv* gene in *A. mellifera* (*Amkkv*) and three
611 potential orthologs in *C. analis* (*CaChS-kkv-like 1*, *CaChS-kkv-like 2*, *CaChS-kkv-like 3*); this gene was
612 not identified in the *F. varia* integument transcriptome.
613
614 ***Genes encoding structural cuticular proteins***
615 The large number of different cuticular protein genes found in insect genomes suggested that
616 their products display redundant and complementary functions [67]. A variable number of genes encode
617 the different classes of structural cuticular proteins in the three bee species and other hymenopterans
618 (S2 Table). Thirty-two CPR genes had been previously identified in *A. mellifera* [12]. We detected
619 other six CPR genes in our RNA-seq analysis of the *A. mellifera* integument, and also 32 and 35 CPR
620 genes in the integument of *F. varia* and *C. analis*, respectively. In addition to have roles as structural
621 proteins in the horizontally arrayed cuticular laminae, the function of some CPR proteins in *T.*
622 *castaneum* was associated to the formation and organization of the pore canals vertically extended
623 across the cuticle [68, 69]. This finding and the variety of CPR genes identified up to now suggest that
624 distinct and additional functions are yet to be discovered for members of the CPR protein class.
625 Like the class of CPR proteins, Twdl proteins are structural cuticular components that
626 effectively bind chitin, as demonstrated in *Bombyx mori* [70]. Two *Twdl* genes were previously
627 characterized in the thoracic integument of *A. mellifera* [44], and now in the abdominal integument,
628 thus indicating that Twdl proteins participate of both rigid (thoracic) and more flexible (abdominal)

cuticles. Like *A. mellifera*, *C. analis* has two *Twdl* genes, but we identified only one in *F. varia*.

Two CPLCP-encoding genes as reported in Willis [12] were herein confirmed in *A. mellifera*. Genes in this family were identified in insect genomes in general and are very enriched in mosquito genomes [71]. Based on sequence homology, we could not identify *CPLCP* transcripts in the *F. varia* and *C. analis* abdominal integument.

CPF proteins were associated to the outer cuticle layers of *A. gambiae* and, apparently, do not bind chitin [72]. Three *CPF* genes were previously reported for *A. mellifera* [12] and one of them, *AmCPF1*, was validated in the thoracic integument through microarray analysis [45]. Here we found *CPF1* and *CPF2* transcripts in the abdominal integument of *A. mellifera*, and in addition, transcripts for two other CPF proteins, *AmUnCPF1* and *AmUnCPF2*. We also identified one *CPF* gene in *F. varia* and one in *C. analis*.

Apd genes seem exclusive of hymenopterans and three of these genes were previously identified in *A. mellifera* [73]. Their transcript levels in the thoracic integument were higher in pharate-adults compared to earlier developmental phases [45]. Here, we detected one more *Apd* gene in *A. mellifera*, *AmApd-like*, and three *Apd* genes in *F. varia* as well as in *C. analis*.

Cpap proteins are essential for the correct formation of the cuticular exoskeleton and elytra in *T. castaneum* [74]. In our RNA-seq analysis, we identified transcripts of three *Cpap1* genes (encoding Cpap proteins containing one chitin-binding domain) in *A. mellifera* and two *Cpap1* genes in *F. varia*, and also verified that the *C. analis* integument is very enriched in *Cpap1* transcripts (n=12), and also in *Cpap* transcripts (n=11) that we could not classify as encoding Cpap1 or Cpap3 (containing three chitin-binding domains). The number of *Cpap3* genes (5 genes) in *A. mellifera* [12] is here confirmed, and two and seven *Cpap3* genes were found in the *F. varia* and *C. analis* integument transcriptomes, respectively. It is important to observe that the genes originally named as *Am-C* and *Am-D* by Soares *et al.* [45] were here renamed as *AmCpap3-c* and *AmCpap3-d*.

The genes, *dumpy* (*dp*), *knk* (*knickkopf*) and *Rtv* (*Retroactive*) have also been identified as encoding cuticular proteins. In *D. melanogaster*, *dp* play roles in cuticle formation [12]. We detected transcripts for *dp* in the abdominal integument of *A. mellifera*, but not in the integument of the other two bee species. The genes *knk* and *Rtv* are both involved in cuticle stabilization in *Drosophila* [75]. In *T. castaneum*, *Rtv* activity is essential for localization of the Knk protein, facilitating its transport to the cuticle [76, 77]. The co-expression of *Rtv* and *knk* in *A. mellifera*, as shown in the reconstructed co-expression network, supports interaction of their respective products, as verified in *T. castaneum*. We

also found *knk* transcripts in *C. analis* integument transcriptome, but not in *F. varia*. *Rtv* transcripts were not detected in the integument of these two bee species.

Genes encoding desaturases and elongases potentially involved in CHC biosynthesis

CHC biosynthesis occurs in the epidermis-associated oenocytes [20] through biosynthetic pathways where desaturase and elongase enzymes have essential roles. Previously, we characterized the gene expression profiles of six desaturases and ten elongases in the developing integument of *A. mellifera* [23]. Our RNA-seq data confirmed these findings, besides identifying three more desaturase genes and other four genes encoding elongases potentially involved in CHC biosynthesis for deposition in the cuticular envelope. For *A. mellifera*, *F. varia* and *C. analis*, a higher proportion of the differentially expressed desaturase and elongase genes showed increased expression in the adults (Ne and/or Fg phases), and only for the eusocial species there were genes more expressed in the pharate-adults (Pbm phase). Among the desaturase and elongase genes, we highlight the expression profiles of *Desat-GB40659*, *Elo-GB54401*, *Elo-GB54302*, *Elo-GB45596* and *Elo-GB46038* orthologs, all showing positive correlation exclusively between the eusocial species.

Genes of the ecdysone signaling cascade regulating cuticle formation and ecdysis in the integument

We detected in the integument the expression of genes that are part of the signaling cascade underlying insect molting and ecdysis, such as *EcR*, *Usp*, *E74*, *E75*, *FTZ-F1*, *CCAP*, *CCAPR* (*Crustacean Cardioactive Peptide Receptor*), *Eth* (*Ecdysis triggering hormone*), *Ethr*, and *Eh* [78]. Importantly, transcripts for these regulators were also detected in greater or lesser levels after ecdysis, in the integument of adult bees. *Usp*, which together with *EcR* forms the nuclear receptor complex that binds 20E and regulates the expression of a cascade of ecdysone-responsive genes, showed the higher expression in *A. mellifera* foragers. This is here tentatively related to the elevated JH titer at this phase of *A. mellifera* worker life [41] once *Usp* also has been proposed as a mediator of JH action [79].

CCAP, *hairy*, *mirr*, and *Kr-h1* in *A. mellifera*, *CCAP*, *Kr-h1*, *and Met* (*Methoprene-tolerant*) in *F. varia*, and *Kr-h1*, *E75*, *EcR, hairy,* and *mirr* in *C. analis* showed increased expression levels at the Ne and/or Fg phases. The roles of theses genes in adult bees, evidently dissociated from the molting events and metamorphosis, are yet to be determined. *Kr-h1* is a direct JH-response gene. *Met*, the JH receptor, has roles in the crosstalk of JH and 20E signaling pathways, which are critical in the

691 regulation of insect metamorphosis [80]. Since *Met*, and also *hairy*, mediate the action of JH on gene
692 regulation [81], they certainly are needed in adult bees where JH has important physiological roles. The
693 *mirr* gene encodes a homeodomain transcription factor with roles in *Drosophila* oogenesis [82]. To our
694 knowledge, its role in the integument has not yet been studied.

695 Some of the identified regulatory genes have been described as playing roles in cuticular
696 melanization, as an example, the *Abdominal B* (*Abd-B*) Hox gene, which regulates *yellow* in the
697 pigmentation/sclerotization pathway in *Drosophila* [83]. *Hairy*, which is a pair-rule gene in *Drosophila*
698 embryos [84], may be involved in the polarity of abdominal segment melanization. The heterodimeric
699 neuropeptide bursicon, composed by the gene products Bursα and Bursβ, is responsible for the
700 regulation of the laccase2-encoding gene, and is crucial for the melanization/sclerotization of the newly
701 formed cuticle [85, 86]. Bursicon interacts with the target tissue through its receptor, the product of the
702 *rickets* gene, whose transcripts were also identified in our RNA-seq analysis of the integument of the
703 three bee species.
704
705 **Searching for clues linking cuticle maturation heterochrony to eusocial or solitary life**
706 **styles in the RNA-seq analysis**
707
708 Our RNA-seq analyses were used to discover active genes in the integument of three bee
709 species and, in addition, we looked for differences in gene expression profiles that could be linked to
710 the heterochronic cuticle maturation dependent on the social/solitary ways of life. The following main
711 findings highlighted differences in integument gene expression distinguishing the eusocial *A. mellifera*
712 and *F. varia* from the solitary *C. analis*: **(a)** In contrast to the eusocial species, a smaller proportion of
713 the genes differentially expressed in the integument was upregulated in *C. analis* foragers in
714 comparison to the newly emerged bees, which is consistent with the cuticle of the solitary bee reaching
715 maturity already at the emergence time; **(b)** The GO analysis including all the integument genes
716 displaying orthology relationship with *Drosophila* genes highlighted functional categories that were
717 mainly shared by both eusocial species in detriment of the solitary *C. analis*; **(c)** The Euclidean distance
718 analysis based on the set of differentially expressed genes clearly separated the *A. mellifera* foragers
719 from the newly-emerged, whereas in *C. analis* these bee groups were clustered, thus suggesting a
720 greater similarity between the integument of newly-emerged and foragers of *C. analis*; **(d)** In contrast
721 to the eusocial species, most of the genes for melanization/sclerotization, genes encoding RR1, RR2, or

non-RR structural proteins, and also regulatory genes, did not show significant expression level variations in *C. analis*. Such differential fluctuation in transcript levels during development may have possibly contributed to the molecular heterochrony of cuticle maturation associated with bee life style. In addition, consistent with the comparatively earlier cuticle maturation process in the solitary bee, we found a higher proportion of CHC biosynthesis-related genes (desaturase and elongase genes) with significantly increased expression levels at the emergence (Ne phase) of *C. analis* in comparison to the eusocial bees. Furthermore, correlation analysis showed that a fraction of cuticle-related genes displayed congruent expression profiles between the eusocial species, but not with the solitary one; **(e)** By superposing the integument gene co-expression networks constructed for the three bee species, we found common interactions for the eusocial species, which were not seem when we compared these species with the solitary one.

Taken together, the comparative approach of the RNA-seq data highlighted suitable gene expression signatures related to adult cuticle formation and maturation in the bee species, in addition of revealing differences in gene expression that may possibly be involved in cuticle maturation heterochrony. Yet, this process may have entailed changes in the expression profiles of regulators of molting and metamorphosis.

Abdominal adult cuticle deposition timing and its ultrastructure exhibit marked differences between the bee species

Cuticle ultrastructure and thickness did not significantly vary between the pharate adults (Pbm phase), newly-emerged (Ne) and foragers (Fg) of *F. varia*, as evidenced by TEM analysis. This was an unexpected result, considering that at the emergence time, *F. varia* workers visibly show an immature cuticle, i.e., incompletely pigmented and sclerotized. Therefore, the evident intensification of cuticular pigmentation and sclerotization in *F. varia* in the subsequent days after the emergence, which is necessary for flight and task performances outside the nest, do not imply in changes in abdominal cuticle thickness. It is possible, however, that thickness measurements taken from cuticular regions other than the abdominal, could evidence a different result, considering that regions of the insect body may diverge in the number of cuticle layers [87] and, consequently, in cuticle thickness.

In contrast, in *A. mellifera*, cuticle deposition is extended through the initial adult stage. Only in the honeybee we could identify post-ecdysially-deposited cuticle layers. Both, the pre- and post-

753 ecdysially-deposited cuticle layers, or laminae, form the procuticle, which corresponds to the largest
754 portion of the cuticle in insects in general. The term exocuticle has been used synonymously with pre-
755 ecdysial cuticle, whereas those layers deposited post-ecdysially form the endocuticle. However, there is
756 some divergence concerning these concepts [88]. In beetles, for example, up to three endocuticle layers
757 are already present in specific areas of the body surface at the time of the adult ecdysis [87]. In
758 *Sarcophaga bullata* flies, deposition of endocuticle occurs before the adult ecdysis [89].

759 As expected, the solitary bees, *C. analis* and *T. diversipes*, and even the primitively eusocial *B.*
760 *brasiliensis* and the facultatively eusocial *E. cordata*, showed a fully deposited cuticle as soon as they
761 emerge, and newly-emerged and forager bees in each of these species displayed similar cuticle
762 ultrastructure, pigmentation and sclerotization. The rapid cuticle maturation in *E. cordata* is consistent
763 with its nesting biology and social structure. *E. cordata* nests are founded by a single female that build
764 up until ten brood cells. The offspring will leave the nest immediately after the emergence for founding
765 new nests. However, daughters may return to the maternal nest, thus resulting in a facultatively social
766 organization with a dominant female (the mother) and its subordinate daughters. There are also nests
767 formed by sisters' females or even by unrelated females, the oldest one showing dominance over the
768 youngest. The dominant female produces all the offspring and rarely leaves the nest, whereas the
769 subordinates assume the tasks of nest provisioning and maintenance, and they also lay trophic eggs that
770 are eaten by the dominant [90-92]. Such female associations may have preceded the highly eusocial
771 way of life [93]. Therefore, in *E. cordata*, as well as in the truly solitary bees, *C. analis* and *T.*
772 *diversipes*, rapid cuticle maturation is the condition for the immediate exit from the nest after
773 emergence.

774 This situation is somewhat diverse for the primitively eusocial *Bombus*. In *B. brasiliensis*, as
775 demonstrated here, the final adult cuticle ultrastructure and thickness are achieved at the emergence.
776 This would allow the workers start foraging soon, as reported for *B. atratus* workers that may leave the
777 nest as soon as at the emergence day (0 day). However, workers of this species may start foraging later,
778 at the age of 10-20 days [94], thus similar to the eusocial bees. Moreover, younger workers in the genus
779 *Bombus* have, in general, incompletely pigmented cuticle and hairs, denoting that cuticle maturity was
780 not yet completely achieved. Such characteristics that seem intermediary to the eusocial and solitary
781 condition may be inherent to the primitively eusocial species, but this requires further investigation.
782 Studies correlating the grade of cuticle pigmentation with the age of starting foraging among
783 primitively eusocial bee species should clarify this issue.

Our TEM analysis and thickness measurements showed that in the same abdominal segment, cuticle ultrastructure greatly differs between the bee species, not only in the number of the adjacently arranged chitin/protein sheets (laminae), but also in the morphology of the most superficial layers. Except for *F. varia*, these results are consistent with a cuticle development timing adapted to the life style, as observed for the highly eusocial *A. mellifera*, the facultatively eusocial *E. cordata*, the primitively eusocial *B. brasiliensis*, and the solitary *C. analis* and *T. diversipes* bees.

Considering that the timing of cuticle deposition is peculiar to bee species, and that cuticle deposition rhythm in *Drosophila* is regulated by a peripheral circadian oscillator in the epidermal cells, which requires the expression of the clock genes *Per*, *Tim2*, *Cyc*, and *Clk* [39], and also that a *Cry* clock gene regulates the rhythm of cuticle deposition in the bean bug *Riptortus pedestris* [40], we compared the expression of seven circadian rhythm genes (*Per*, *Tim2, Cyc, Clk*, *Cry, Vri and Pdp1*) in the developing integument of *A. mellifera*, *F. varia* and *C. analis*. Consistent with the differences in the timing of cuticle deposition, the expression profiles of *Clk* in *A. mellifera*, *F. varia*, and *C. analis* were negatively or non-correlated. Similarly, the expression profiles of *Cry* in *A. mellifera* and *C. analis* were negatively correlated (*Cry* was not identified in *F. varia*), as well as the expression of *Per*. It is likely that *AmPer* has roles in adult cuticle organization. Interaction of *AmPer* and other genes involved in cuticle formation was specifically observed in *A. mellifera*, whose sequenced genome gives more support for gene co-expression network reconstruction. In *A. mellifera*, *Per* was co-expressed with the *knk* gene, which in *T. castaneum* was associated with stabilization of the cuticular laminae [95]. Both genes were co-expressed with structural cuticular protein genes such as *AmCpap3-a*, *AmTwdl*(*Grp*-Glycine-rich protein), *AmUnCPR-RR2-2*, *AmCPR26*, *Am49Ah-like* and *AmSgAbd2-like*, and also with *Amyellow-y*, a gene in the yellow family, involved in cuticle pigmentation [96]. The expression profiles of another clock gene, *Tim2*, were positively correlated between the eusocial species, with a marked decrease in expression levels at the emergence, suggesting roles in the final step of adult cuticle formation in these bees. The *Pdp1* gene encodes a basic leucine zipper transcription factor and is expressed at high levels in the epidermis and other tissues of *Drosophila* embryos. *Pdp1* is an essential clock gene linked to the circadian rhythm. It is a regulator of *Clk* and other clock genes, such as *Tim*, *Per*, and *Pdf* (Pigment dispersing factor), a neuropeptide controlling circadian behavioral rhythms [97, 98]. *Pdp1* seems an important gene in the *C. analis* integument since it is connected with nine structural cuticle protein genes, three chitin-related genes, and two desaturase encoding genes in the co-expression network. However, it is significantly more expressed after the emergence, when the cuticle of the solitary

bee is already formed, thus virtually excluding a role in cuticle laminae deposition rhythm. Differently from *C. analis*, *Pdp1* was not co-expressed in the networks reconstructed with the *A. mellifera* and *F. varia* genes involved in cuticle formation and maturation. Some of the cuticular genes were also co-expressed with *Cyc* in the integument of *A. mellifera* and *C. analis*.

Cuticular n-alkanes as markers of cuticle maturity in bees

N-alkanes are structural lipids in the insect cuticle [99, 17], where they compose the envelope [100]. The absolute quantities of n-alkanes and unsaturated CHCs were significantly higher in the foragers than in the earlier developmental phases of the eusocial *A. mellifera* and *F. varia* species. The n-alkanes detected in higher proportions in *A. mellifera* foragers than in the newly-emerged were C_{23}, C_{24}, C_{25}, C_{26}, C_{27}, C_{29}, C_{31}, and C_{33}, the C_{25} and C_{27} n-alkanes presenting the highest proportions. The analysis of the individual CHC peaks obtained from *F. varia* also showed higher proportions of C_{27} and C_{29}, besides a higher proportion of C_{22}, in the foragers. All these n-alkanes, except C_{27} and C_{33}, were also proportionally increased in foragers than in newly-emerged bees of the eusocial *Melipona marginata* [101]. These data are consistent with previous reports on higher levels of n-alkanes in *A. mellifera* foragers [102] and in foragers of an ant species, *Pogonomyrmex barbatus* [103]. In contrast, the proportions and absolute quantities of n-alkanes did not differentiate foragers from the newly-emerged in *C. analis*. Together, these findings may be interpreted as the solitary bee displaying an accelerated process of cuticle maturation in comparison to the eusocial ones. N-alkanes may be markers of cuticle structure maturation. Long-chain alkanes are thought to increase cuticle waterproofing [104, 103], suggesting that this essential ability for the performance of extra-nidal activities was acquired earlier in the development of *C. analis*. At the adult emergence, the solitary bee already has the chemical profile needed for a prompt interaction with the environment outside the nest. Consistently, the levels of n-alkanes also did not significantly differ between young and old females of the solitary leafcutter bee species, *Megachile rotundata* [105].

Conclusions

Using RNA-seq analysis of the integument of two eusocial bee species, *A. mellifera* and *F. varia*, and a solitary bee, *C. analis*, we identified genes involved in cuticle (exoskeleton) formation and

maturation. The expression profiles of these genes were determined at three developmental time points corresponding to adult cuticle deposition/differentiation at the pharate-adult stage, newly-ecdysed cuticle, and fully developed cuticle of forager bees. TEM analysis of the cuticle at these time points, including other bee species, and CHC profiles determination were performed in addition to the transcriptome analysis. Together, these experimental approaches provided novel data on integument developmet. We also searched for clues in integument gene expression, structure, and CHC profiles that could be consistent with the premise that eusociality might have entailed heterochronic changes in cuticle development, resulting in faster cuticle maturation in the solitary bee, thus allowing flight and forager activities immediately after emergence, and in slow cuticle maturation in the eusocial bees, which benefit from the protected nest environment for a period of time after the emergence. This study expands our understanding on the molecular biology and structure of the developing integument, besides highlighting differences in the process of cuticle maturation related to the eusocial/solitary behaviors.

Materials and Methods

Sample collection

We collected workers of *A. mellifera* (Africanized) and *F. varia* from colonies maintained in the Experimental Apiary of the Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil. Trap-nests to collect samples of the solitary species *C. analis* and *T. diversipes* were made [106] and placed in the Experimental Apiary area. Additional bee species (*B. brasiliensis*, *E. cordata* and *T. diversipes*) were obtained from donations (see acknowledgments section).

We used females from three developmental phases: pharate adults in process of cuticle pigmentation (Pbm phase), newly emerged (Ne) adults and foragers. Carrying pollen bees from the solitary and social species, and building-nest females from the solitary species, were identified as foragers (Fg). The *B. brasiliensis*, *E. cordata* and *T. diversipes* species were exclusively used for cuticle morphology studies through TEM. In this case, we used the Ne and Fg phases.

RNA extraction and sequencing

For each developmental phase (Pbm, Ne and Fg) of *A. mellifera* and *F. varia*, we prepared three independent samples, each made with five abdominal integuments. For the corresponding

876 developmental phases of *C. analis,* we prepared three independent samples, each containing three
877 abdominal integuments. The RNA extractions were made using TRIzol® reagent (Invitrogen)
878 following manufacturer's instruction. The extracted RNAs (2 µg/per sample) were sent to a facility
879 (Laboratório Central de Tecnologias de Alto Desempenho em Ciências da Vida, Universidade Estadual
880 de Campinas, Campinas, Brazil) to access sample quality through a 2100 Bioanalyzer and for library
881 preparation (TruSeq™ RNA - Illumina®) and RNA sequencing in an Illumina HiSeq 2500 equipment
882 (paired-end reads, 2 x 100 bp read length). We obtained an average of 30 million reads per sample, with
883 90% of the bases showing quality scores > Q30. The RNA-seq data is deposited at the National Center
884 for Biotechnology Information (NCBI) database under the BioProject ID PRJNA490324.

885 **Adapters trimming and quality check**

886 We used the software Scythe v. 0.991 (https://github.com/vsbuffalo/scythe) for trimming 3'
887 standard Illumina adapter sequence. We followed a Cutadapt v. 1.4.1 [107] trimming at 5' ends of the
888 reads, and we filtered reads with Phred quality > 20. The trimmed sequences were filtered using the
889 software PRINSEQ-lite v. 0.20.3 [108] and sequence quality was evaluated through the software
890 FastQC v. 0.11.2 [109].

891 **Transcriptome assembly and gene expression**

892 We aligned the high quality reads from *A. mellifera* against its genome v. 4.5 using the software
893 TopHat v. 2.0.9 [110]. The *A. mellifera* aligned sequences were quantified and the developmental
894 phases compared using the software Cufflinks v. 2.1.1 [111]. The extension Cuffmerge integrated the
895 reads to the mapping results and the tool Cuffdiff checked the expression levels for each sample and the
896 significance of comparisons. CuffmeRbund R package v. 2.8.2 allowed us to access all this information
897 [112].

898 For *F. varia* and *C. analis*, we used the software Trinity (trinityrnaseq_r2014717) [113, 114].
899 The N50 contig length (smallest contig length for which the sum of fewest contigs corresponds to 50%
900 or more of the assembly [115]) of all transcripts of *F. varia* was 2372 and of *C. analis* was 2440.
901 Orthology search was performed through the software InParanoid 8 [116]. We only accepted those
902 transcripts with higher similarity with *A. mellifera* than to *Drosophila melanogaster*. Statistical
903 evaluation of these data was done with the R software v. 3.1.2, using the packages R DESeq2 v. 1.6.3
904 [117] and edgeR v. 3.8.6 [118]. We considered as differentially expressed between developmental
905 phases, those contigs with significant results for both R packages.

All heat maps were designed using the function heatmap.2 from gplots R package [119]. For all groups of genes, we measured the clustering potential of the samples for each phase and species. For this approach, we used the R package pvclust v. 1.3.2 [120] based on correlation distances, with a complete linkage method, and 10,000 bootstrap replication. We used unbiased p values (AU) and bootstrap values as measurements of clusters' significance. Clusters showing AU > 95% were considered statistically significant [120].

Molecular and functional characterization of differentially expressed genes

For the analysis of gene expression in *A. mellifera*, we filtered the differentially expressed genes using the following thresholds: q-value < 0.05; Log 2 Fold Change ≤ -1 or ≥ 1 and Fragments Per Kilobase of transcript per Million mapped reads (FPKM) ≥ 5. In the case of the other two bee species, *F. varia* and *C. analis*, we used the parameters cited in the previous section. For the Gene Ontology (GO) enrichment analysis we used the *A. mellifera* gene IDs to look for *D. melanogaster* ortologues in Fly Base, through the support of the online softwares g:Profiler (http://biit.cs.ut.ee/gprofiler/gorth.cgi), and g:Orth function [121, 122]. The same was done for *F. varia* and *C. analis* but using *A. mellifera* ortologues. We filtered the *Drosophila* IDs to avoid ID repetition and used them to generate an input list for the software DAVID v. 6.7 (http://david.abcc.ncifcrf.gov) [123, 124], used to perform the Gene Ontology analysis. The annotated functions belonged to Biological Process (BioP), Cellular Components (CC) and Molecular Function (MF) categories. Structural cuticular protein encoding genes were classified in accordance with the software CutProtFam-Pred (http://aias.biol.uoa.gr/CutProtFam-Pred/home.php) [125]. Venn diagrams were plotted with the online version of the software jvenn [126] (<http://bioinfo.genotoul.fr/jvenn/example.html>).

Transcription factor binding sites search with TRANSFAC®

Transcription factors whose binding sites could be enriched in specific groups of *A. mellifera* gene models were searched using TRANSFAC® [127] against insects database. For this enrichment analysis, we searched the 5' UTR regions covering -3,000 bases relative to the transcription start sites of the genes involved in sclerotization/melanization processes, chitin metabolism, CHC biosynthetic pathways, regulation of cuticle formation and maturation, circadian rhythm, and non-melanization pigmentation pathways (see gene IDs S1 File). We excluded the genes with 5' UTR < 500 bases. We used the *A. mellifera* genes in this analysis once it is the only among the species here studied with an available reference genome. For the transcription factor FTZ-F1 binding sites, we used the TRANSFAC database (DROME$FTZF1_01, DROME$FTZF1_02, and DROME$FTZF1_03 from *D.*

melanogaster, and BOMMO$CPR92_01, and BOMMO$CPR92_02 from *B. mori*) to generate a positional matrix. Here, we only highlighted those transcription factor binding sites (TFBS), which could be relevant for insect cuticle formation and maturation (see **Discussion section)**.

Gene co-expression networks

The networks were plotted based on the correlation of gene expression in the integument of the analyzed bee species. We used the software Cytoscape v. 3.3.0 [128] – for Linux, and its plugin ExpressionCorrelation App v. 1.1.0. We only accepted correlations $\geq +0.95$ or ≤ -0.95 and $p \leq 0.05$.

Transmission electron microscopy (TEM)

We dissected the integument from the right anterior region of the third abdominal tergite of the studied bee species. The ultrastructure of the integument was compared between species and between the developmental phases (Pbm, Ne and Fg) using 11 *A. mellifera* integument pieces (4 from Pbm, 3 from Ne, and 4 from Fg phases), 9 integuments from *F. varia* (3 Pbm, 3 Ne, 3 Fg), 11 integuments from *C. analis* (4 Pbm, 4 Ne, 3 Fg), 4 integuments from *B. brasiliensis* (2 Ne, 2 Fg), 6 integuments from *E. cordata* (3 Ne, 3 Fg), and 6 integuments from *T. diversipes* (3 Ne, and 3 Fg). The Ne and Fg phases of *B. brasiliensis* were recognized based on the grade of body pigmentation criterion (less intense in the Ne bees) and for the Fg phase we also examined the wings in the search for erosion signals that could indicate intense foraging activity. The integument samples were fixed in 5% glutaraldehyde in cacodylate buffer 0.1 M, pH 7.2, during 2 h under shaking, washed 3X in the cacodylate buffer, fixed in osmium tetroxide 1% diluted in phosphate buffer 0.1M, pH 7.2, dehydrated in acetone and propylene oxide and embedded in resin. We used uranyl acetate for enhancing image contrast. The ultrathin sections were examined in a Jeol-Jem-100cx-II Electron Microscope and the software ImageJ v. 10.2 [129] was used to measure integument thickness. Measurements were compared among the developmental phases and species using Analysis of Variance (ANOVA) associated with the Tukey's Honestly Significance Difference (Tukey's HSD) post hoc test in R software v. 3.1.2, except for the *A. mellifera* and *B. brasiliensis* data. As for *A. mellifera* the data did not present a normal distribution (Shapiro-Wilk normality test) we used the Kruskal-Wallis test associated with the *post hoc* Conover-Iman test and Bonferroni correction, and for *B. brasiliensis* we used the Student's t-test [130].

Cuticular hydrocarbon profiles

The quantification of CHCs was based on their peak area in each chromatogram. For the analysis of relative peak area, we collected 15 bees per developmental phase for each of the highly eusocial species, while for *C. analis* we obtained four Pbm-staged bees, seven Ne and seven Fg bees.

968 The *F. varia* hind legs were removed to avoid resin contamination. Except for this species, we bathed
969 each sample in 1.5 ml of n-hexane 95% (Mallinckrodt Chemicals) for 1 min and 30 s to extract the
970 CHCs. Due to the small size of *F. varia* species, we used 500 µl of n-hexane to extract the CHCs [131].
971 The extracts were dried under N_2 flow and resuspended in 160 µl of n-hexane (100 µl for *F. varia*
972 extracts) before running the analysis. CHC identification was made in a Gas Chromatograph / Mass
973 Spectrometer (GC-MS) system (Shimadzu GCMS model QP2010), equipped with a 30 m DB-5MS
974 column using helium as the carrier gas (1 ml/min), through electronic ionization (EI) mode. CHC
975 relative quantification and normalization of the peak areas were performed following Falcón *et al.* [23]
976 description. We also compared each developmental phase considering the classes of CHC (n-alkanes,
977 unsaturated, and branched alkanes), repeating the normalization process for each case.
978 In order to verify differences between the developmental phases and between the bee species,
979 we performed a clustering approach as described in the section **Transcriptome assembly and gene**
980 **expression**, but using the Euclidean distance instead of the correlation distance. We also verified the
981 compounds that better explained the detected differences. With this purpose, we performed a Principal
982 Component Analysis (PCA) using the R software. The variation of each CHC peak area between the
983 developmental phases was accessed through a Tukey's HSD test in R software.
984 Additionally, we calculated the absolute quantities of n-alkanes and unsaturated CHC per bee
985 for the species *A. mellifera*, *F. varia*, and *C. analis*. An analytical curve [132] was built to establish the
986 correlation between the quantities of the used standards and the CHCs. It is described by the equation:
987 y=ax +b, where y is the known amount of the standard, x is the quantity of the unknown CHC, a is the
988 peak area of the standard, and b is the area of the intercepted background. We prepared the curve based
989 on the n-alkanes standards C_{23}, C_{25}, and C_{32} (Alltech Corporation) and using 1.25, 2.5, 5, 10, 15 and 20
990 µg/ml of each alkane. To the standard solutions and to each sample, we added 100 µl of the internal
991 standard α-colestane (6.25 µg/µl) (Sigma-Aldrich) under the previously cited chromatography
992 conditions. The values of the correlation curve (R) between CHCs and standards were ≥ 0.99. After
993 curve preparation, CHCs were quantified in three independent samples, each prepared with individual
994 bees, for each developmental phase and species. Due to the reduced body size of *F. varia*, we used
995 pools of three bees for each one of the three independent samples per developmental phase and the
996 obtained CHC values were corrected accordingly. To calculate the concentrations of compounds up to
997 C_{24}, we used the curve of C_{23}; for compounds from C_{25} up to C_{29}, we used the C_{25} curve; and for
998 compounds larger than C_{29} up to C_{35} we used the curve of C_{32}. We followed an ANOVA associated with

the *post hoc* Tukey's HSD test to compare the absolute quantity of each CHC (n-alkanes and unsaturated) between developmental phases. The absolute quantities of the *F. varia* unsaturated compounds were not calculated once their quantities are very low.

Acknowledgments

We thank to Sidnei Matheus, Solange Augusto, Gabriela Freiria and Anete Lourenço by the samples provided. This work was supported by São Paulo Research Foundation (FAPESP): Grants to MMGB (2014/13136-0), ZLPS (2011/03171-5), and NPL (2014/50265-3). Fellowships to: TF (2012/24284-5) – during PhD, MEN (2012/09108-6). And later, fellowship from Programa Nacional de Pós-Doutorado (PNPD) CAPES/HCPA to TF (Process number: 88887.160608/2017-00). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

1. Hopkins TL, Kramer KJ. Insect cuticle sclerotization. Annul Rev Entomol. 1992;37: 273-302.

2. Wood OR, Hanrahan S, Coetzee M, Koekemoer LL, Brooke BD. Cuticle thikening associated with pyrethroid resistance in the major malaria vector *Anopheles funestus*. Parasit Vectors. 2010;3: 67.

3. Csikós G, Molnár K, Borhegyi NH, Talián GC, Sass M. Insect cuticle, an in vivo model of protein trafficking. J Cell Sci. 1999;112: 2113-2124.

4. Hill RJ, Billas IML, Bonneton F, Graham LD, Lawrence MC. Ecdysone receptors: from the Ashburner model to structural biology. Annu Rev Entomol. 2013;58: 251–271.

5. Ali MS, Iwanaga M., Kawasaki H. Ecdysone-responsive transcriptional regulation determines the temporal expression of cuticular protein genes in wing discs of *Bombyx mori*. Gene 2013;512: 337-347.

6. Hiruma K, Riddiford LM. The coordination of the sequential appearance of MHR4 and dopa decarboxylase during the decline of the ecdysteroid titer at the end of the molt. Mol Cell Endocrinol. 2007;276: 71-79.

7. Gu J, Huang L-X, Gong Y-J, Zheng S-C, Liu L, Huang L-H, et al. De novo characterization of transcriptome and gene expression dynamics in epidermis during the larval-pupal metamorphosis of common cutworm. Insect Biochem Mol Biol. 2013;43: 794-808.

8. Kramer JK, Dziadik-Turner C, Koga D. Chitin metabolism in insects. In: Kerkut GA, Gilbert LI, editors. Comprehensive insect physiology, biochemistry and pharmacology. Oxford: Pergamon

Press; 1985. pp. 75-115.

9. Dixit R, Arakane Y, Specht CA, Richard C, Kramer KJ, Beeman RW, et al. Domain organization and phylogenetic analysis of proteins from the chitin deacetylase gene family of *Tribolium castaneum* and three other species of insects. Insect Biochem Mol Biol. 2008;38: 440–451.

10. Merzendorfer H, Zimoch L. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. J Exp Biol. 2003;206: 4393-4412.

11. Locke M, Krishnan N. The distribution of phenoloxidases and polyphenols during cuticle formation. Tissue Cell. 1971;3: 103-126.

12. Willis JH. Structural cuticular proteins from arthropods: annotation, nomenclature and sequence characteristics in the genomic era. Insect Biochem Mol Biol. 2010;40: 189-204.

13. Rebers JE, Riddiford LM. Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. J Mol Biol. 1988;203: 411-423.

14. Andersen SO. Amino acid sequence studies on endocuticular proteins from the desert locust, *Schistocerca gregaria*. Insect Biochem Mol Biol. 1998;28: 421-434.

15. Guan X, Middlebrooks BW, Alexander S, Wasserman SA. Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in *Drosophila*. Proc Natl Acad Sci. 2006;103: 16794–16799.

16. Hepburn HR. Structure of integument. In: Kerkut GA, Gilbert LI, editors. Comprehensive insect physiology, biochemistry and pharmacology. Oxford: Pergamon Press; 1985. pp. 1-58.

17. Gibbs AG. Lipid melting and cuticular permeability: new insights into an old problem. J Insect Physiol. 2002;48: 391-400.

18. Howard RW, Blomquist GJ. Ecological, behavioral and biochemical aspects of insect hydrocarbons. Annu Rev Entomol. 2005;50: 371-393.

19. Edney EB. Water balance in desert arthropods. Science. 1967;156: 1059-1065.

20. Piek T. Synthesis of wax in the honeybee (*Apis mellifera* L.). J Insect Physiol.1964;10: 563-572.

21. Blomquist GJ, Dillwith JW. Cuticular lipids. In: Kerkut GA, Gilbert LI, editors. Comprehensive Insect Physiology, Biochemistry and Pharmacology. Oxford: Pergamon Press; 1985. vol. 3, pp. 117-154.

22. Blomquist GJ, Jurenka R, Schal C, Tittiger C. Pheromone production: biochemistry and molecular biology. In: Gilbert LI, editor. Insect Endocrinology. New York: Elsevier; 2012. pp. 523-567.

23. Falcón T, Ferreira-Caliman MJ, Nunes FMF, Tanaka ED, Nascimento FS, Bitondi MMG. Exoskeleton formation in *Apis mellifera*: cuticular hydrocarbons profiles and expression of

desaturase and elongase genes during pupal and adult development. Insect Biochem Mol Biol. 2014;50: 68-81.

24. Shamim G, Ranjan SK, Pandey DM, Ramani R. Biochemistry and biosynthesis of insect pigments. Eur J Entomol. 2014;111: 149-164.

25. Solano F. Melanins: skin pigments and much more – types, structural models, biological functions, and formation routes. New J Sci. 2014: 1-28. Article ID 498276. http://dx.doi.org/10.1155/2014/498276.

26. Andersen SO. Insect cuticular sclerotization: a review. Insect Biochem Mol Biol. 2010;40: 166-178.

27. Hiruma K, Riddiford LM, Hopkins TL, Morgan TD. Roles of dopa decarboxylase and phenoloxidase in the melanization of the tobacco hornworm and their control by 20-hydroxyecdysone. J Comp Physiol B. 1985;155: 659-669.

28. Zufelato MS, Bitondi MMG, Simões ZLP, Hartfelder K. The juvenile hormone analog pyriproxyfen affects ecdysteroid-dependent cuticle melanization and shifts the pupal ecdysteroid peak in the honey bee (*Apis mellifera*). Arthropod Struct Dev. 2000;19: 111-119.

29. Michener CD. The social behavior of the bees. Massachusetts: Harvard University Press; 1974.

30. Wilson EO. Sociobiology. The New Synthesis. Cambridge: Harvard University Press; 1975.

31. Honeybee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature. 2006;443: 931-949.

32. Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, Elsik CG, et al. The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol. 2015;16: 76. DOI: 10.1186/s13059-015-0623-3.

33. Kapheim KM, Pan H, Li C, Salzberg SL, Puiu D, Magoc T, et al. Genomic signatures of evolutionary transitions from solitary to group living. Science. 2015;348: 1139-1143. DOI: 10.1126/science.aaa4788.

34. Elias-Neto M, Nascimento ALO, Bonetti AM, Nascimento FS, Mateus S, Garófalo CA, Bitondi MMG. Heterochrony of cuticular differentiation in eusocial corbiculate bees. Apidologie. 2013;45: 397-408.

35. Hansell MH. The ecological impact of animal nests and burrows. Funct Ecol. 1993;7: 5-12.

36. Evans JD, Spivak M. Socialized medicine: individual and communal disease barriers in honey bees. J Invertebr Pathol. 2010;103: S62–S72.

37. West-Eberhard MJ. Developmental plasticity and the origin of species differences. Proc Natl Acad Sci. 2005;102: 6543-6549.

1133 **38.** Wilson EO, Hölldobler B. Eusociality: origin and consequences. Proc Natl Acad Sci. 2005;205:
1134 13367-13371.
1135
1136 **39.** Ito C, Goto SG, Shiga S, Tomioka K, Numata H. Peripheral circadian clock for the cuticle
1137 deposition rhythm in *Drosophila melanogaster*. Proc Natl Acad Sci. 2008;105: 8446-8451.
1138
1139 **40.** Ikeno T, Katagiri C, Numata H, Goto SG. Causal involvement of mammalian-type cryptochrome in
1140 the circadian cuticle deposition rhythm in the bean bug *Riptortus pedestris*. Insect Mol Biol.
1141 2011;20: 409-415.
1142
1143 **41.** Huang ZY, Robinson GE, Tobe SS, Yagi KJ, Strambi C, Strambi A, et al. Hormonal regulation of
1144 behavioural development in the honey bee is based on changes in the rate of juvenile hormone
1145 biosynthesis. J Insect Physiol. 1991;37: 733-741.
1146
1147 **42.** Mlodzik M, Hiromi Y, Weber U, Goodman CS, Rubin GM. The *Drosophila seven-up* gene, a mem-
1148 ber of the steroid receptor gene superfamily, controls photoreceptor cell fates. Cell. 1990;60: 211-
1149 224.
1150
1151 **43.** Zander RH. Minimal values for reliability of bootstrap and jackknife proportions, decay index, and
1152 Bayesian posterior probability. PhyloInformatics. 2004;2: 1-13.
1153
1154 **44.** Soares MPM, Silva-Torres FA, Elias-Neto M, Nunes FMF, Simões ZLP, Bitondi MMG.
1155 Ecdysteroid-dependent expression of the *Tweedle* and *Peroxidase* genes during adult cuticle
1156 formation in the honey bee, *Apis mellifera*. PloS One. 2011;6: e20513.
1157
1158 **45.** Soares MPM, Barchuk AR, Simões ACQ, Cristino AS, Freitas FCP, Canhos LL, et al. Genes
1159 involved in thoracic exoskeleton formation during the pupal-to-adult molt in a social insect model,
1160 *Apis mellifera*. BMC Genomics. 2013;14: 576.
1161
1162 **46.** Arakane Y, Lomakin J, Beeman RW, Muthukrishnan S, Gehrke SH, Kanost MR. Molecular and
1163 functional analyses of amino acids decarboxylases involved in cuticle tanning in *Tribolium*
1164 *castaneum*. J Biol Chem 2009a*;* 284: 16584-16594.
1165
1166 **47.** Gorman MJ, Arakane Y. Tyrosine hydroxylase is required for cuticle sclerotization and
1167 pigmentation in *Tribolium castaneum*. Insect Biochem Mol Biol. 2010;40: 267-273.
1168
1169 **48.** Miyazaki S, Okada Y, Miyakawa H, Tokuda G, Cornette R, Koshikawa S, et al. Sexually dimorphic
1170 body color is regulated by sex-specific expression of *Yellow* gene in ponerine ant, *Diacamma* sp.
1171 PLoS One. 2014;9: e92875.
1172
1173 **49.** Wright TR F. The genetics of biogenic amine metabolism, sclerotization, and melanization in
1174 *Drosophila melanogaster*. Adv Genet. 1987; 24: 127-222.
1175
1176 **50.** Elias-Neto M, Soares MPM, Simões ZLP, Hartfelder K, Bitondi MMG. Developmental
1177 characterization, function and regulation of a Laccase2 encoding gene in the honeybee, *Apis*
1178 *mellifera* (Hymenoptera, Apinae). Insect Biochem Mol Biol. 2010;40: 241-251.
1179

51. Osanai-Futahashi M, Tatematsu KI, Futahashi R, Narukawa J, Takasu Y, Kayukawa T, et al. Positional cloning of a *Bombyx* pink-eyed white egg locus reveals the major role of *cardinal* in ommochrome synthesis. Heredity. 2016;116: 135-145.

52. Sugumaran M. Complexities of cuticular pigmentation in insects. Pigment Cell Melanoma Res. 2009;22: 523-525.

53. Lloyd V, Ramaswami M, Krämer H. Not just pretty eyes: *Drosophila* eye-colour mutations and lysosomal delivery. Trends Cell Biol. 1998;8: 257-259.

54. Hamza I, Dailey HA. One ring to rule them all: trafficking of heme and heme synthesis intermediates in the metazoans. Biochim Biophys Acta. 2012;1823: 1617-1632.

55. Stubenhaus BM, Dustin JP, Neverett ER, Beaudry MS, Nadeau LE, Burk-McCoy E, et al. Light-induced depigmentation in planarians models the pathophysiology of acute porphyrias. eLife. 2016;5: e14175.

56. Shaik KS, Meyer F, Vázquez AV, Flötenmeyer M, Cerdán ME, Moussian B. δ-Aminolevulinate synthase is required for apical transcellular barrier formation in the skin of the *Drosophila* larva. Eur J Cell Biol. 2012;91: 204-215.

57. Xi Y, Pan PL, Ye YX, Yu B, Zhang CX. Chitin deacetylase family genes in the brown planthopper, *Nilaparvata lugens* (Hemiptera: Delphacidae). Insect Mol Biol. 2014;23: 695-705.

58. Arakane Y, Hogenkamp DG, Zhu YC, Kramer KJ, Specht CA, Beeman RW, et al. Characterization of two chitin synthase genes of the red flour beetle, *Tribolium castaneum*, and alternate exon usage in one of the genes during development. Insect Biochem Mol Biol. 2004;34: 291-304.

59. Arakane Y, Dixit R, Begum K, Park Y, Specht CA, Merzendorfer H, et al. Analysis of functions of the chitin deacetylase gene family in *Tribolium castaneum*. Insect Biochem Mol Biol. 2009b;39: 355-365.

60. Tellam RL, Vuocolo T, Johnson SE, Jarmey J, Pearson RD. Insect chitin synthase: cDNA sequence, gene organization and expression. Eur J Biochem. 2000;267: 6025-6042.

61. Gagou ME, Kapsetaki M, Turberg A, Kafetzopoulos D. Stage-specific expression of the chitin synthase *DmeChSA* and *DmeChSB* genes during the onset of *Drosophila* metamorphosis. Insect Biochem Mol Biol. 2002;32: 141-146.

62. Ampasala DR, Zheng S, Zhang D, Ladd T, Doucet D, Krell PJ, et al. An epidermis-specific chitin synthase cDNA in *Choristoneura fumiferana*: cloning, characterization, developmental and hormonal-regulated expression. Arch Insect Biochem Physiol. 2011;76: 83-96.

63. Zhu Q, Arakane Y, Beeman RW, Kramer KJ, Muthukrishnan S. Functional specialization among insect chitinase family genes revealed by RNA interference. Proc Natl Acad Sci. 2008;105: 6650-6655.

64. Luschnig S, Bätz T, Armbruster K, Krasnow MA. *serpentine* and *vermiform* encode matrix proteins with chitin binding and deacetylation domains that limit tracheal tube length in *Drosophila*. Curr Biol. 2006;16: 186-194.

65. Arakane Y, Muthukrishnan S, Kramer KJ, Specht CA, Tomoyasu Y, Lorezen MD, et al. The *Tribolium* chitin synthase genes *TcCHS1* and *TcCHS2* are specialized for synthesis of epidermal cuticle and midgut peritrophic matrix. Insect Mol Biol. 2005;14: 53-463.

66. Moussian B, Schwarz H, Bartoszewski S, Nüsslein-Volhard C. Involvement of chitin in exoskeleton morphogenesis in *Drosophila melanogaster*. J Morphol. 2005;264: 117-130.

67. Pan PL, Ye YX,Lou YH, Lu JB, Cheng C, Shen Y, et al. A comprehensive omics analysis and functional survey of cuticular proteins in the brown planthopper. Proc Natl Acad Sci. 2018;115: 5175-5180.

68. Noh MY, Kramer KJ, Muthukrishnam S, Kanost MR, Beeman RW, Arakane A. Two major cuticular proteins are required for assembly of horizontal laminae and vertical pore canals in rigid cuticle of *Tribolium castaneum*. Insect Biochem Mol Biol. 2014;53: 22-29.

69. Noh MY, Muthukrishnam S, Kramer KJ, Arakane Y. *Tribolium castaneum* RR-1 cuticular protein TcCPR4 is required for formation of pore canals in rigid cuticle. PLoS Genet. 2015;11: e1004963.

70. Tang L, Liang J, Zhan Z, Xiang Z, He N. Identification of the chitin-binding proteins from the larval proteins of silkworm*, Bombyx mori*. Insect Biochem Mol Biol. 2010;40: 228-234.

71. Cornman RS, Willis JH. Annotation and analysis of low-complexity protein families of *Anopheles gambiae* that are associated with cuticle. Insect Mol Biol. 2009;18: 607-622.

72. Togawa T, Dunn WA, Emmons AC, Willis J. CPF and CPFL, two related gene families encoding cuticular proteins of *Anopheles gambiae* and other insects. *Insect Biochem Mol Biol.* 2007;37: 675-688.

73. Kucharski R, Maleszka J, Maleszka R. Novel cuticular proteins revealed by the honey bee genome. Insect Biochem Mol Biol. 2007;37: 128-134.

74. Jasrapuria S, Specht CA, Kramer KJ, Beeman RW, Muthukrishnan S. Gene families of cuticular proteins analogous to peritrophins (CPAPs) in *Tribolium castaneum* have diverse functions. PLoS One 2012;7: e49844

75. Moussian B, Tång E, Tonning A, Helms S, Schwarz H, Nüsslein-Volhard C, et al. *Drosophila* Knickkopf and Retroactive are needed for epithelial tube growth and cuticle differentiation through their specific requirement for chitin filament organization. Development. 2006;133: 163-171.

76. Chaudhari SS, Arakane Y, Specht CA, Moussian B, Boyle DL, Park Y, et al. Knickkopf protein protects and organizes chitin in the newly synthesized insect exoskeleton. Proc Natl Acad Sci. 2011;108: 17028-17033.

1274 **77.** Chaudhari SS, Arakane Y, Specht CA, Moussian B, Kramer KJ, Muthukrishnan S. et al. Retroactive
1275 maintains cuticle integrity by promoting trafficking of Knickkopf into the procuticle of *Tribolium*
1276 *castaneum*. PLoS Genet. 2013;9: e1003268.
1277
1278 **78.** Zitnam D, Adams ME. Neuroendocrine regulation of ecdysis. In LI Gilbert, editor. San Diego:
1279 Elsevier Academic Press; 2012. pp. 253-309.
1280
1281 **79.** Jones G, Sharp PA. Ultraspiracle: an invertebrate nuclear receptor for juvenile hormones. Proc Natl
1282 Acad Sci. 1997;94: 13499-13503.
1283
1284 **80.** Jindra M, Uhlirova M, Charles J-P, Hill RJ. Genetic evidence for function of the bHLH-PAS protein
1285 Gce/Met as a juvenile hormone receptor. PLoS Genetics. 2015;11: e1005394.
1286
1287 **81.** Saha TT, Shin SW, Dou W, Roy S, Zhao B, Hou Y, et al. Hairy and Groucho mediate the action of
1288 juvenile hormone receptor Methoprene-tolerant in gene repression. Proc Natl Acad Sci. 2016;7:
1289 E735-E743.
1290
1291 **82.** Zhao D, Woolner S, Bownes M. The Mirror transcription factor links signalling pathways in
1292 *Drosophila* oogenesis. Dev Genes Evol. 2000;210: 449-457
1293
1294 **83.** Jeong S, Rokas A, Carroll SB. Regulation of body pigmentation by the Abdominal-B Hox protein
1295 and its gain and loss in *Drosophila* evolution. Cell. 2006;125: 1387-1399.
1296
1297 **84.** Carroll SB, Laughon A, Thalley BS. Expression, function, and regulation of the hairy segmentation
1298 protein in the *Drosophila* embryo. Genes Dev. 1988;2, 883-890.
1299
1300 **85.** Song Q. Bursicon, a neuropeptide hormone that controls cuticle tanning and wing expansion. In:
1301 Gilbert LI, editor. Insect Endocrinology. New York: Academic Press; 2012. pp. 93-105.
1302
1303 **86.** Costa CP, Elias-Neto M, Falcon T, Dallacqua RP, Martins JR, Bitondi MMG. RNAi-mediated
1304 functional analysis of bursicon genes related to adult cuticle formation and tanning in the honeybee,
1305 *Apis mellifera*. PLoS One . 2016;11: e0167421.
1306
1307 **87.** Zelazny B, Neville AC. Endocuticle layer formation controlled by non-cyrcadian clocks in beetles.
1308 Insect Physiol. 1972;18: 1967-1979.
1309
1310 **88.** Andersen SO, Hojrup P, Roepstorff P. Insect cuticular proteins. Insect Biochem Mol Biol. 1995;25:
1311 153-176.
1312
1313 **89.** Whitten J. Coordinated development in the fly foot: sequential cuticle secretion. J Morphol.
1314 1969;127: 73-104.
1315
1316 **90.** Garófalo CA. Social structure of *Euglossa cordata* nests (Hymenoptera: Apidae: Euglossini).
1317 Entomol Gen. 1985;11: 77-83.
1318
1319 **91.** Augusto SC, Garófalo CA. Comportamento das fêmeas nas associações formadas em ninhos de
1320 *Euglossa cordata* (Hymenoptera; Apidae; Euglossini). In: Encontro sobre Abelhas, Ribeirão Preto,

SP, Brazil; 1994. pp. 171-181.

92. Freiria GA, Garófalo CA, Del Lama MA. The primitively social behavior of *Euglossa cordata* (Hymenoptera, Apidae, Euglossini): a view from the perspective of kin selection theory and models of reproductive skew. Apidologie. 2017;48: 523-532. Doi: 10.1007/s13592-017-0496-4

93. Cardinal S, Danforth BN. The antiquity and evolutionary history of social behavior in bees. PloS One. 2011;6: e21086.

94. Silva-Matos EV, Garófalo CA. Worker life tables, survivorship, and longevity in colonies of *Bombus* (*Fervidobombus*) *atratus* (Hymenoptera: Apidae). Rev Biol Tropical. 2000;48: 657-664.

95. Chaudhari SS, Moussian B, Specht CA, Arakane Y, Kramer KJ, Beeman RW, et al. Functional specialization among members of Knickkopf family of proteins in insect cuticle organization. PLoS Genet. 2014;10: e1004537.

96. Hinaux H, Bachem K, Battistara M, Rossi M, Xin Y, Jaenichen R, et al. Revisiting the developmental and cellular role of the pigmentation gene *yellow* in *Drosophila* using a tagged allele. Dev Biol. 2018;438: 111-123

97. Reddy KL, Rovani MK, Wohlwill A, Katzen A, Storti RV. The *Drosophila* Par domain protein I gene, Pdp1, is a regulator of larval growth, mitosis and endoreplication. Dev Biol. 2006;289: 100-114.

98. Cyran SA, Buchsbaum AM, Reddy KL, Lin MC, Glossop NR, Hardin PE, et al. vrille, Pdp1, and dClock form a second feedback loop in the *Drosophila* circadian clock. Cell. 2003;112: 329-341.

99. Gibbs A, Pomonis JG. Physical properties of insect cuticular hydrocarbons: the effects of chain length, methyl-branchingand unsaturation. Comp Biochem Physiol. 1995;112B: 243-249.

100. Lockey KH. Lipids of the insect cuticle: origin, composition and function. Comp Biochem Physiol Part B: Comp Biochem. 1988;89: 595-645.

101. Ferreira-Caliman MJ, Nascimento FS, Turatti IC, Lopes NP, Zucchi R. The cuticular hydrocarbons profiles in the stingless bee *Melipona marginata* reflect task-related differences. J Insect Physiol. 2010;56: 800-804.

102. Kather R, Drijfhout FP, Martin SJ. Task group differences in cuticular lipids in the honey bee *Apis mellifera*. J Chem Ecol. 2011;37: 205-212.

103. Wagner D, Brown MJF, Broun P, Cuevas W, Moses LE, Chao DL, et al. Task-related differences in the cuticular hydrocarbon composition of the harvester ants, *Pogonomyrmex barbatus*. J Chem Ecol. 1998;24: 2021-2037.

104. Gibbs AG. Water-proofing properties of cuticular lipids. Amer Zool.1998;38: 471-482.

105. Paulmier I, Bagnères AG, Afonso CMM, Dusticier G, Rivière G, Clément JL. Alkenes as sexual

pheromone in the alfalfa leaf-cutter bee *Megachile rotundata*. J Chem Ecol. 1999;25: 471-490.

106. Jesus BMV, Garófalo CA. Nesting behaviour of *Centris* (*Heterocentris*) *analis* (Fabricius) in southeastern Brazil (Hymenoptera, Apidae, Centridini). Apidologie. 2000;31: 503-515.

107. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17: 10-12.

108. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27: 863-864.

109. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

110. Trapnell C, Patcher L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25: 1105-1111.

111. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. Nature Biotechnol. 2010;28: 511-515.

112. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protoc. 2012;7: 562-578.

113. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nature Biotechnol. 2011;15: 644-652.

114. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protoc. 2013;8: 1494-1512.

115. Miller JR, Koren S, Sutton G. Assembly algorithms for the next-generation sequencing data. Genomics. 2010;95: 325-327.

116. Sonnhammer ELL, Östlund G. InParanoid 8: ortology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. 2014;43: D234-D239.

117. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550.

118. Robinson MD, Smyth GK. Small sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics. 2008;9: 321-332.

119. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, et al. gplots: various R programming tools for plotting data. R package version 2.16.0. 2015. Available from:

http://CRAN.R-project.org/package=gplots

120. Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006;22: 1540-1542.

121. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler - a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 2007;35: W193-W200.

122. Reimand J, Arak T, Vilo J. g:Profiler - a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res. 2011;39: W307-W315.

123. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protocols. 2009a;4: 44-57.

124. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009b;37: 1-13.

125. Ioannidou ZS, Theodoropoulou MC, Papandreou NC, Willis JH, Hamodrakas SJ. CutProtFam-Pred: detection and classification of putative structural cuticular proteins from sequence alone, based on profile hidden Markov models. Insect Biochem Mol Biol. 2014;52: 51-59.

126. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. jvenn: an interactive Venn diagram viewer. BMC Bioinformatics. 2014;15. DOI: 10.1186/1471-2105-15-293.

127. Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, et al. TRANSFAC: and integrated system for gene expression regulation. Nucleic Acids Res. 2000;28: 316-319.

128. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13: 2498-2504.

129. Abramoff MD, Magalhães PJ, Ram SJ. Image processing with ImageJ. Biophotonics Internat. 2004;11: 36-42.

130. Fay DS, Gerow K. A biologist's guide to statistical thinking and analysis. In: Hobert O, editor. Wormbook. 2013. DOI: doi/10.1895/wormbook.1.159.1 .

131. Nunes TM, Nascimento FS, Turatti IC, Lopes NP, Zucchi R. Nestmate recognition in a stingless bee: does the similarity of chemical cues determine guard acceptance? Anim Behav. 2008;75: 1165-1171.

132. Analytical Methods Committee. Internal quality control of analytical data. Analyst 1995;120: 29-34.

Figures

Fig 1. Venn diagrams constructed with the genes and contigs differentially expressed in the integument of the developmental phases of (A) *A. mellifera*, (B) *F. varia*, and (C) *C. analis*. The number of genes upregulated in each pairwise comparison is indicated. Pbm: pharate adults; Ne: newly-emerged bees; Fg: foragers.

Fig 2. Gene Ontology (GO) functional terms attributed to integument genes during adult cuticle development and maturation. The functional terms more represented in the Pbm and Ne phases than in the Fg phase are indicated as Younger>Fg, and those more represented in the Ne and Fg phases than in the Pbm phase are reported as Older>Pbm. The green box includes GO terms related to the cuticle-producing tissue, the epidermis. Purple box: GO terms associated to structural components of the cuticle. Black box: GO terms potentially associated to CHC biosynthetic pathways. Yellow box: GO terms related to pigments and pigmentation.

Fig 3. Distance correlation analysis between developmental phases based on the expression of DEGs and DECs. (A) *A. mellifera*; (B) *F. varia*; (C) *C. analis*. Red values (BP): bootstrap support. Green values (AU): cluster support. Arrows point to significant clusters (AU > 95%). Branch edges are shown in gray. Pbm = pharate adults. Ne = newly emerged bees. Fg = foragers.

Fig 4. Representative heatmaps of gene expression profiles through the Pbm (pharate-adult), Ne (newly-emerged), and Fg (forager) developmental phases of (A) *A. mellifera*, (B) *F. varia*, and (C) *C. analis*. Genes were grouped according to their potential function in adult cuticle formation and maturation. Different lowercase letters on the heatmaps means statistically significant difference (see Materials and Methods) in the expression levels between the developmental phases of each bee species.

Fig 5. Ortholog genes showing significantly correlated expression profiles at least between two of the three bee species, *A. mellifera*, *F. varia* and *C. analis*. Expression profiles of *ebony*, *tan*, *Idgf4-like*, *Cda5*, *chitooligosaccharidolytic-domain-like*, *CPR14*, *CPR17*, *CPR25*, *CPR26*, *Apd-like*, *Elo-GB54302*, *Elo-GB54401*, *Elo-GB45596*, *Ethr*, *E74*, *Hr4*, *Hr38*, *FTZ-F1*, *rickets*, *Tim2*, and *ALAS* were positively correlated between the eusocial bee species, and negatively or non-correlated with the solitary bee. Expression profiles of *CPR18*, *CPR23*, *Apd-3*, *Desat-GB40659*, *Elo-GB46038*, and *Ptx-1* were positively correlated between the eusocial species, the basal line in the graphic representations indicating undetected orthologs in *C. analis*. Pbm: pharate adults, Ne: newly emerged, and Fg: foragers. Color key at the bottom of figure.

Fig 6. Overlapping interactions in the gene co-expression networks reconstructed with *A.*

***mellifera* (S2 Fig) and *F. varia* (S3 Fig) genes related to cuticle formation and maturation.**

Fig 7. Ultrastructure and thickness of the developing and mature adult cuticle of bees differing in ways of life. (A) *A. mellifera* (eusocial), (B) *F. varia* (eusocial), (C) *C. analis* (solitary), (D) *B. brasiliensis* (primitively eusocial), (E) *E. cordata* (facultatively eusocial), and (F) *T. diversipes* (solitary). Developmental phases are indicated: Pbm (pharate-adult); Ne (newly-emerged); 0h, 24h, 48h, 72h, and 96h after adult emergence; Fg (forager). The cuticle/epidermis junction was used to align the cuticle images. Means and standard deviations of cuticle thickness measurements are represented in red. The number of cuticle samples measured (N) is indicated for the Pbm, Ne and Fg phases of each bee species. (A'- F') Cuticle thickness measurements (μm) for the corresponding bee species. Different lowercase letters indicate significant statistical difference between the developmental phases of each species.

Supporting information

S1 Fig. Correlation heatmaps based on the RNA-seq data obtained from the integument of the pharate adults (Pbm), newly-emerged (Ne) and foragers (Fg) of the three bee species: (A) *A. mellifera*, (B) *F. varia* and (C) *C. analis*. The numbers 1, 2 and 3 following the Pbm, Ne, and Fg abbreviations indicate the independent samples of each developmental phase.

S2 Fig. Gene co-expression networks in the integument of *A. mellifera* for adult cuticle formation and maturation. The genes are indicated in the nodes, and the edges represent significant correlation among genes.

S3 Fig. Gene co-expression networks in the integument of *F. varia* for adult cuticle formation and maturation. The genes are indicated in the nodes, and the edges represent significant correlation among genes.

S4 Fig. Gene co-expression networks in the integument of *C. analis* for adult cuticle formation and maturation. The genes are indicated in the nodes, and the edges represent significant correlation among genes.

S5 Fig. Distances between the developmental phases of *A. mellifera*, *F. varia* and *C. analis* based on Euclidean distance analysis of total CHCs, n-alkanes, unsaturated CHCs, and branched CHCs relative quantifications. Red boxes indicate significant clusters with 95% of confidence. Arrows indicate significantly supported clusters. AU clusters' support (red values); BP bootstrap support (green values); Branches' edges (gray values). Pharate-adults (Pbm), newly emerged (Ne), and forager (Fg)

bees.

S1 Table. Ortholog genes displaying significantly correlated expression profiles. Comparisons of gene expression levels through the pharate-adult (Pbm), newly-emerged (Ne) and forager (Fg) developmental phases of *A. mellifera*, *F. varia*, and *C. analis*. Blue: significant positive correlation. Red: significant negative correlation. (-): undetectable gene expression.

S2 Table – Number of genes encoding the different classes of structural cuticular proteins in hymenopterans.

S1 File. Genes identified in the RNA-seq analysis of the integument of *A. mellifera*, *F. varia* and *C. analis*.

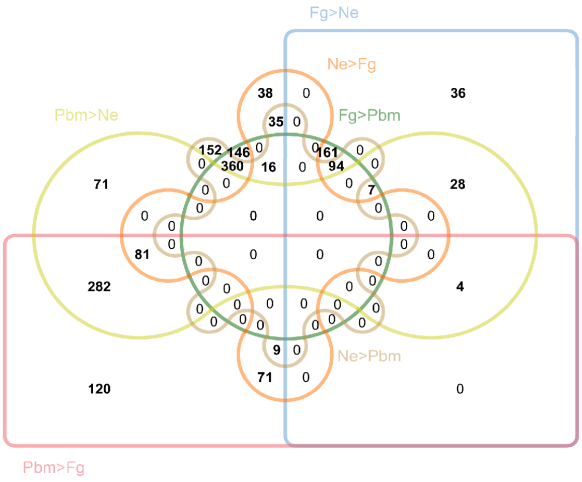
S2 File. Genes upregulated in the comparisons of the developmental phases of each bee species, *A. mellifera*, *F. varia* and *C. analis*. Pharate-adult (Pbm), newly-emerged (Ne) and forager (Fg) developmental phases. The bee species and developmental phases compared are specified in the inferior margin of each table in this File. For example: Amel_Pbm>Ne means the list of *A. mellifera* genes upregulated in Pbm in comparison to Ne.

S3 File. Gene Ontology (GO) functional analysis of the differentially expressed genes. The bee species and developmental phases compared are specified in the inferior margin of each table in this File. For example: Amel_Pbm_Ne>Fg_ID means the bees' gene IDs and Fly Base gene IDs from higher expressed genes at younger developmental phases; and Amel_Ne_Fg>Pbm_GO means the Gene Ontology of the higher expressed genes at older developmental phase based on their fly orthologues IDs.

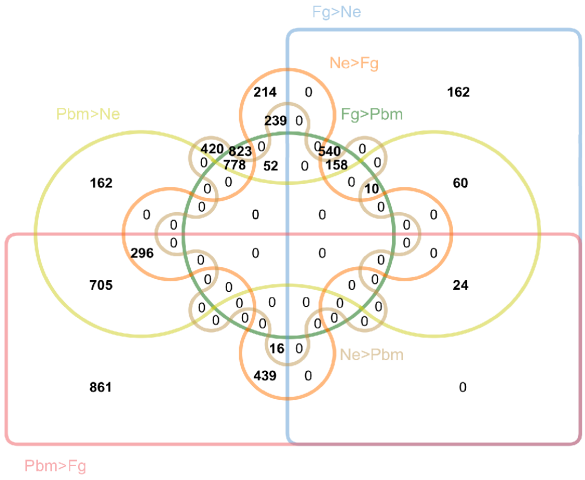
S4 File. Cuticular hydrocarbon (CHC) profiles determined for the developmental phases of *A. mellifera*, *F. varia*, and *C. analis*, variable contribution (total and per comparison) and mass quantification. Pharate-adult (Pbm), newly-emerged (Ne) and forager (Fg) developmental phases.

1554 Figure 1

A.



B.



C.

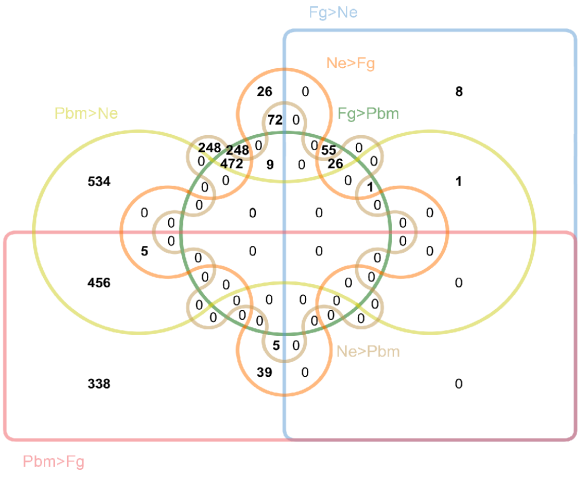
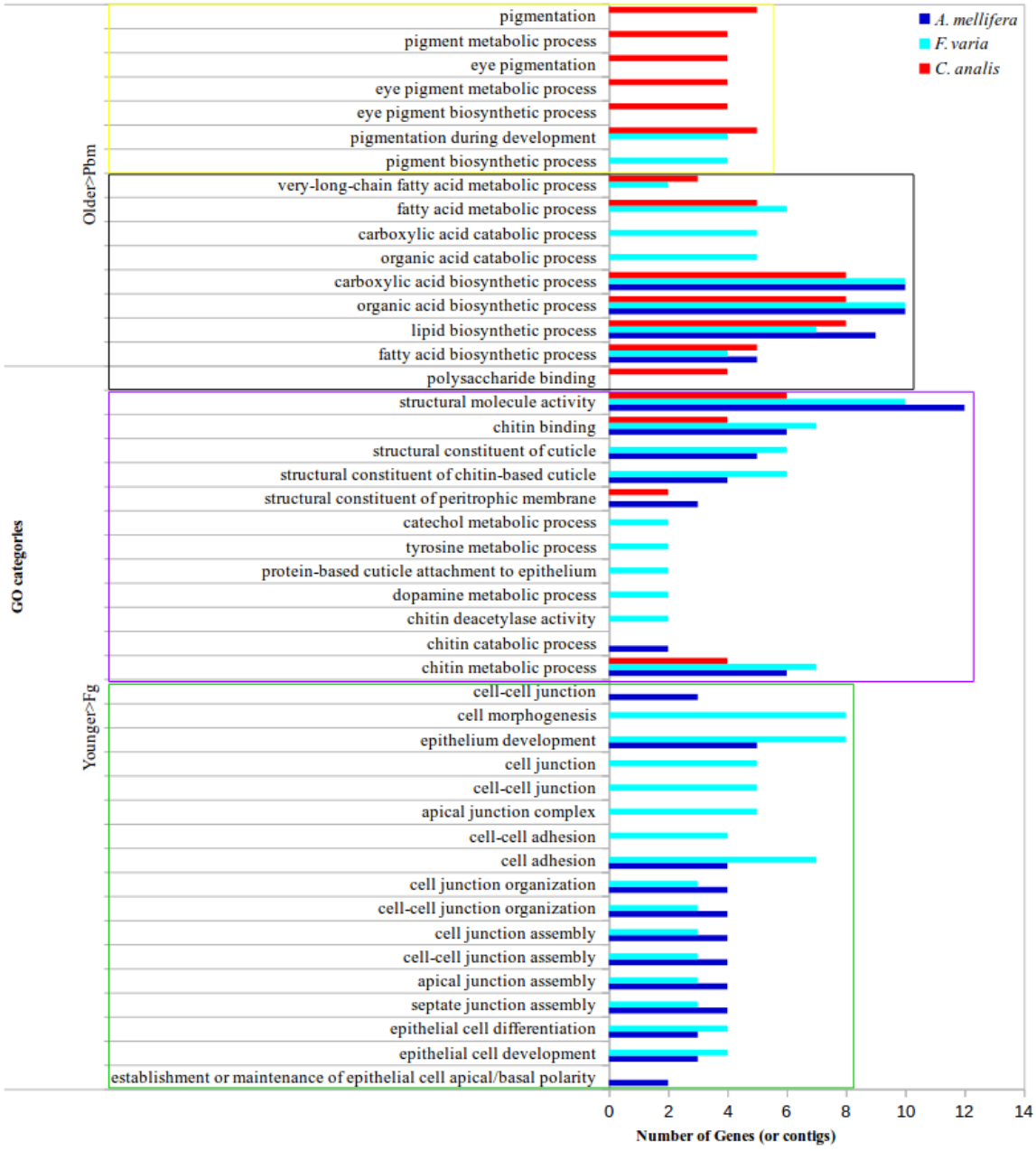
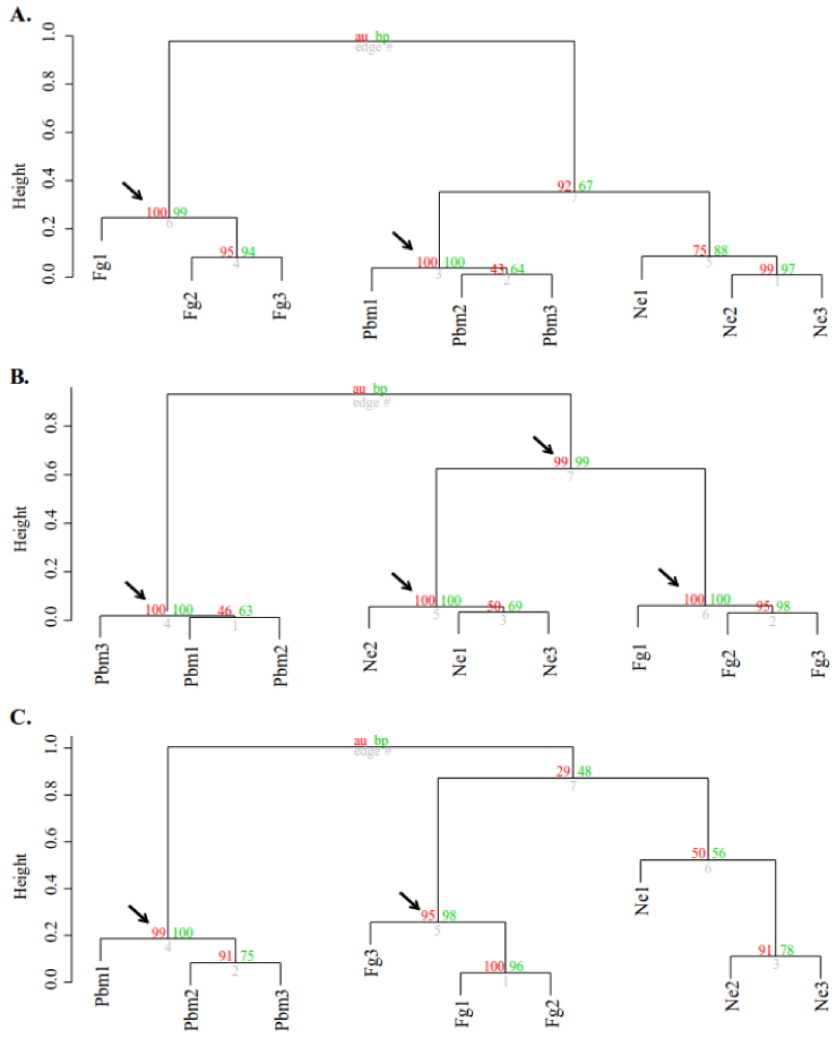


Figure 2



1563 Figure 3

1564



1566

1567

1568

1569

1570

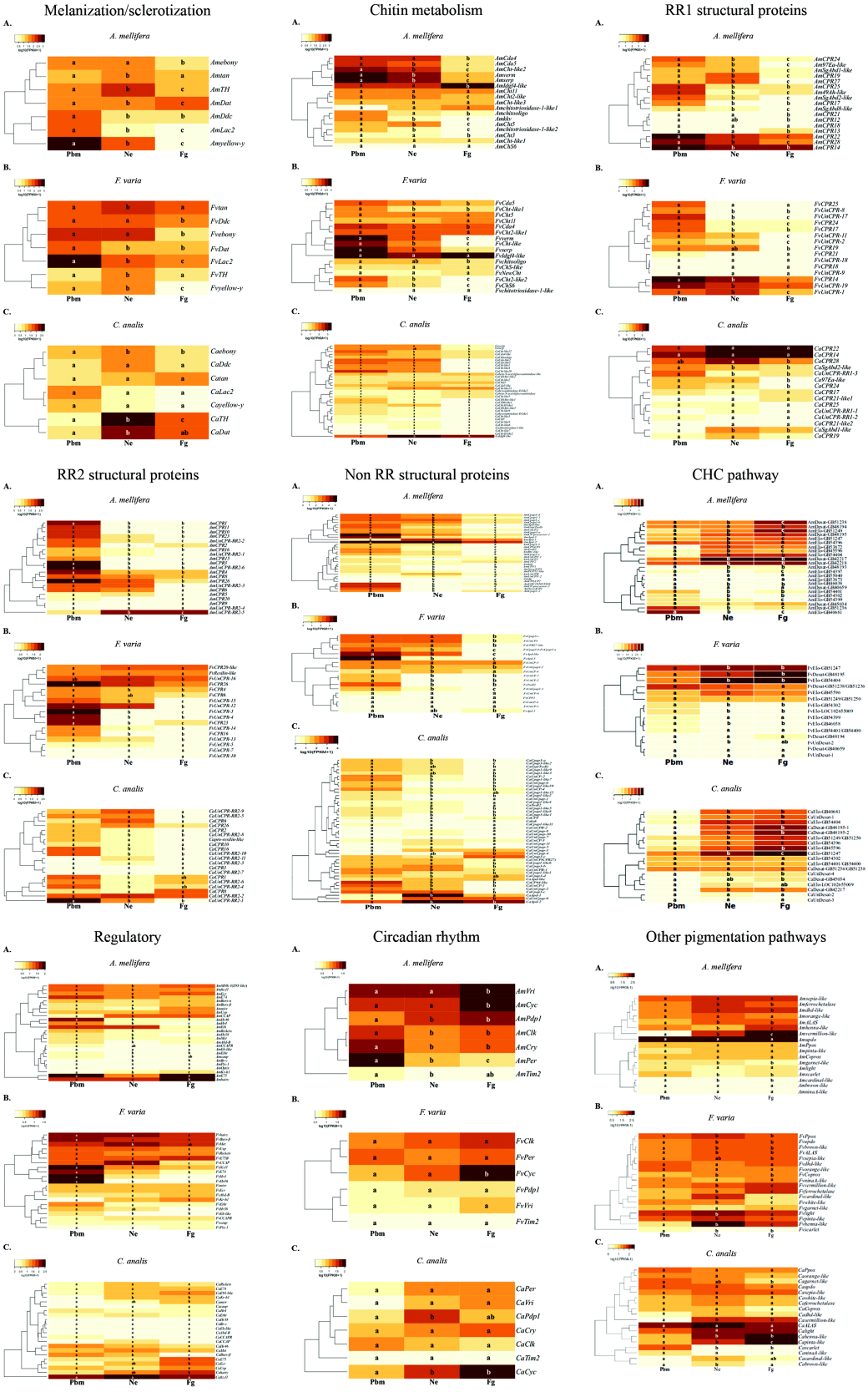
1571

1572

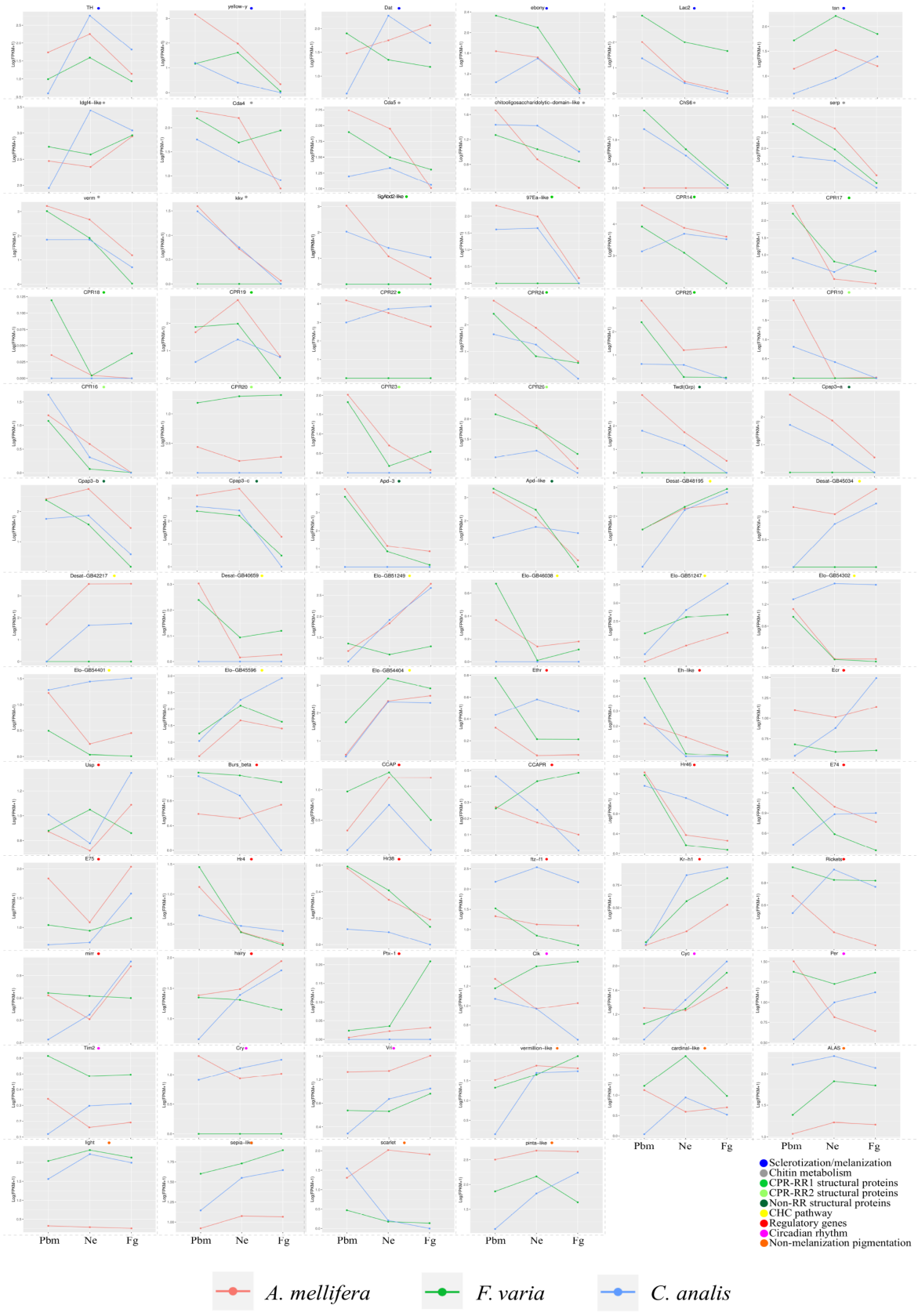
1573

1574

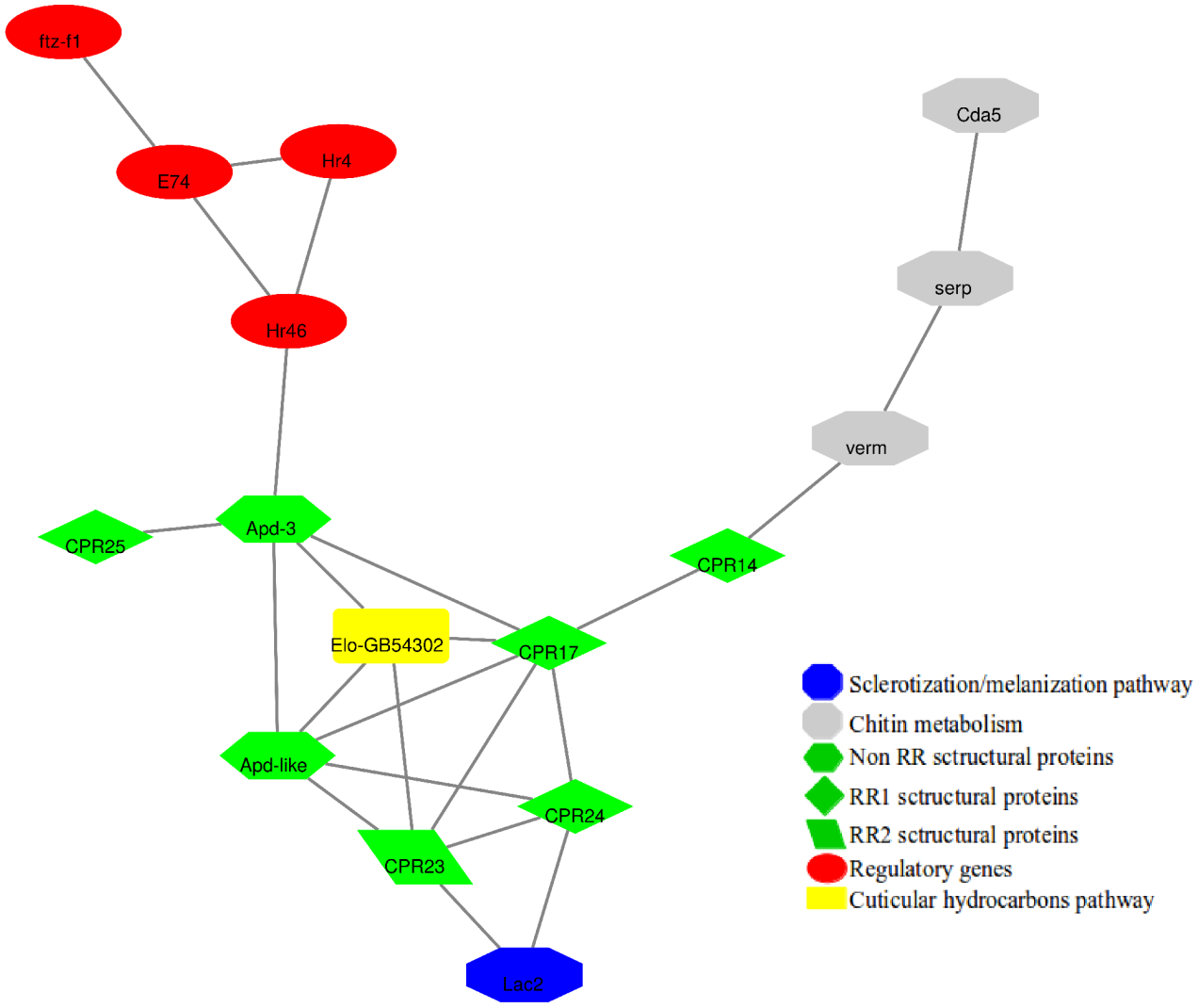
1575 Figure 4



1577 Figure 5



1579 Figure 6



1581

1582

1583

1584

1585

1586

1587

1588

1589

Figure 7