Orthologues of genetically identified *Drosophila melanogaster* chitin producing and organising genes in *Apis mellifera*

Richard Odemer[1], Peter Rosenkranz[1] & Bernard Moussian[2]

1 Apicultural State Institute, University of Hohenheim, August-von-Hartmannstr. 13, 70599 Stuttgart, Germany

2 Animal Genetics, Interfaculty Institute for Cell Biology, Eberhard-Karls University of Tübingen, Auf der Morgenstelle 8, 72076 Tübingen, Germany

Corresponding author: bernard.moussian@uni-tuebingen.de

Abstract

Insects face their environment through their cuticle that is produced during embryogenesis and before stage transitions when a new cuticle is needed to accommodate growth. Some factors that drive cuticle differentiation are being genetically characterised notably in the model insect *Drosophila melanogaster*. To date, most if not all of these factors, among others Chitin synthase-1, the chitin deacetylases Vermiform and Serpentine and the transcription factor Grainyhead are involved in synthesis and organisation of chitin, an essential component of the cuticle.

In the present work, we identified orthologues of these factors in the honeybee *Apis mellifera* and monitored their expression at different developmental stages. Organisation of the analysed genes differs considerably in *A. mellifera* and *D. melanogaster*. For instance, we detected species-specific splicing variants of *chitin synthase-1* transcripts, and dramatic differences in the 5-prime coding regions of respective *grainyhead* genes. Interestingly, *vermiform* and *serpentine* are expressed at different stages in *A. mellifera*, while in *D. melanogaster* their expression profiles are largely identical. Overall, differences in gene organisation and expression pattern between *D. melanogaster* and *A. mellifera* may compile different cuticle compositions that reflect their life style and ecology. Our data may serve to elucidate the honeybee-specific mechanisms of cuticle formation.

Introduction

Insects are covered by a cuticle that protects them against dehydration and pathogen entry and allows locomotion (Moussian, 2010). In addition, the cuticle stabilises

internal organs such as the foregut, the hindgut and the tracheae. It is a stratified extracellular matrix that is produced by the underlying epithelial cells at their apical site. Genetic approaches using the fruit fly *Drosophila melanogaster* has advanced our understating on the mechanisms of cuticle differentiation. Several factors identified and characterised in *D. melanogaster*, have been studied in the red flour beetle *Tribolium castaneum* by RNA interference (RNAi), largely confirming results obtained with *D. melanogaster*, occasionally adding new notions to their function. For instance, Knickkopf, a membrane-bound chitin organising factor in *D. melanogaster* (Moussian et al, 2006), is additionally needed to protect chitin from chitinase-driven degradation in *T. castaneum* (Chaudhari et al, 2011). Hence, despite the equivalence of chitin synthesising and organising factors in *D. melanogaster* and *T. castaneum*, functional differences exist. The significance of these differences remains to be investigated.

In order to deepen our understanding in molecular cuticle variability in insects, we sought to identify those cuticle factors in the honeybee *Apis mellifera* that had been previously characterised genetically in *D. melanogaster*. These factors are Krotzkopf verkehrt/Chitin synthase-1 (Kkv/CS-1), the chitin monomer (GlcNAc) producing enzyme Mummy (Mmy, UDP-GlcNAc pyrophosphorylase), the membrane-bound Knickkopf (Knk) and Retroactive (Rtv) and the extracellular proteins Obstructor-A, Serpentine and Vermiform, two chitin deacetylases (Luschnig et al, 2006; Moussian et al, 2005a; Moussian et al, 2005b; Moussian et al, 2006; Petkau et al, 2012; Tonning et al, 2006). They are mainly involved in chitin synthesis and organisation. Mutations in the respective genes are lethal and cause chitin deficiency (*kkv/CS-1*, *mmy*) or chitin disorganisation (*knk*, *rtv*, *verm*, *serp*). In addition, we identified the *A. mellifera* Grainyhead (Grh) transcription factor that in *D. melanogaster* has been shown to control the transcription of several chitin producing and organising factors such as *knk* (Bray & Kafatos, 1991; Gangishetti et al, 2012; Pare et al, 2012). In order to provide first indications of molecular dynamics during cuticle differentiation in *A. mellifera*, we also monitored the expression of these genes at different developmental stages.

Results and Discussion

In order to identify the orthologues of *D. melanogaster* genetically characterised chitin producing and organising factors in *A. mellifera*, we first searched the bee

genome for respective loci with the coding sequence of *D. melanogaster* factors using the tBLASTx software at BeeBase (http://hymenopteragenome.org/beebase/) (Munoz-Torres et al, 2011). Next, the exons of the annotated sequence were assembled to a coding sequence that was translated to a protein sequence. To reconfirm the highest hit, the protein sequence was subsequently used to search the *D. melanogaster* protein sequence database at FlyBase for homologous sequences by BLASTp. Concomitantly, our deduced honey bee protein sequences were aligned with respective annotated protein sequences at the NCBI gene bank by BLASTp. Ambiguous exons were tested by quantitative real time PCR (qPCR) to be part of respective transcripts. Occasionally, gene organisation in the genome of the honey bee-related species *Bombus impatiens* was used as an additional reference. Co-linearity of *D. melanogaster* and *A. mellifera* protein sequences (i.e. lack of alignment gaps larger than few amino acids) is considered as an indication that the conceptual *A. mellifera* protein is the correct one. Expression of genes was recorded at the first instar larval and yellow-eyed pupal stages. Correctness of this selection of stages is supported by the observation that generally in arthropods like *D. melanogaster* and *T. castaneum*, expression of cuticle genes is continuous.

The chitin synthase CS-1/Kkv

Chitin is synthesised by the membrane-inserted chitin synthase. Mutations in the *D. melanogaster* chitin synthase coding gene *krotzkopf verkehrt* (*kkv*) are lethal and cause a chitin deficient collapsed cuticle (Moussian et al, 2005a). The *kkv* locus codes for two enzyme isoforms with 1615 amino acids derived from alternative splicing of exon 7a and 7b (Irion, 2012). The *kkv*-similar *A. mellifera GB49845* (*Amkkv*) locus is annotated to contain 23 exons coding for a conceptual protein with 1783 amino acids (Figure 1). Sequence comparison suggests that duplicated stretches of sequences encoded by alternative exons account for the non-linearity and over-length of the *A. mellifera* sequence. For instance, exons 14 & 15 (102 aa) both contain the WGTRE motif that has been proposed to be essential for chitin extrusion (Merzendorfer, 2006). Other insect chitin synthases contain only one WGTRE motif arguing that exons 14 and 15 are mutually exclusive. Underlining the importance of this motif, a missense mutation changing the glycine to an aspartate completely abrogates enzyme activity in *D. melanogaster* (Moussian et al, 2005a). The etoxazole sensitive isoleucine^{1056} (Van Leeuwen et al, 2012) preceding the

WGTRE motif is conserved in both sequences. By qPCR, we confirmed that these exons are indeed alternatively spliced in larvae and pupae (Figure 2). Another pair of exons that like their respective exons in other insects (Arakane et al, 2004; Arakane et al, 2005; Hogenkamp et al, 2005) may be alternatively spliced is exons 18 (180 bps) and 19 (174 bps). Both exons are predicted to code for transmembrane domains (TMPred), which are highly similar (66% identical bases coding for 69% similar and 80% identical amino acids, respectively). However, our qPCR data using two independent pairs of primers suggest that in *A. mellifera* the exons 18 and 19 may be present in the same transcript in larvae and pupae (Figure 2). If both exons are present in the same transcript, the C-terminus of the resulting protein would be cytoplasmic instead of extracellular as compared to the situation in *D. melanogaster*.

In summary, the conceptual *A. mellifera* chitin synthase with only one WGTRE motif has 1681 residues and, apart from the duplicated transmembrane domain encoded by exon 18 and 19, is largely collinear with the respective *D. melanogaster* protein. Over 1599 residues, they are 66,4% (1061 amino acids) similar and 80,2% (1282 amino acids) identical to each other.

The major non-aligning region is present at the N-terminus, which is variable in chitin synthases in all insects (Hogenkamp et al, 2005; Merzendorfer, 2006). The predicted NCBI *A. mellifera* chitin synthase protein has 1632 amino acids that are, except from ten unique N-terminal amino acids, identical to the conceptual BeeBase sequence (Suppl. Figure 1). In qPCR experiments, we could not detect the 5' region of the *A. mellifera* chitin synthase transcript deposited at the NCBI site, whereas the BeeBase 5' sequence of the chitin synthase transcript could be amplified (Figure 2).

GlcNAc pyrophosphorylase

The monomer of chitin, N-acetylglucosamine (GlcNAc), which is also a major component of N-glycans, is produced by the Leloir pathway (Moussian, 2008). The last enzyme of this pathway is the cytoplasmic GlcNAc pyrophosphorylase, named Mummy (Mmy) in *D. melanogaster* (Araujo et al, 2005; Schimmelpfeng et al, 2006; Tonning et al, 2006). *D. melanogaster* larvae homozygous for loss-of-function *mmy* alleles have, among others, a strongly reduced cuticle and die before hatching. The *mmy* gene codes for two isoforms with 520 (A) and 483 (B) amino acids, isoform A having a longer N-terminus (Figure 3). The *A. mellifera* locus *GB44897* codes for a protein with 469 amino acids encompassing the GlcNAc pyrophosphorylase

signature from residue 84 to residue 406 (PF01704). It is collinear with the *D. melanogaster* isoform A and shows 53,8% identity and 73,8% similarity to it over the entire length of the protein, and 58,3% identity and 76,9% similarity over the PF01704 domain. Since *GB44897* (*Ammmy*) is a housekeeping gene, we did not monitor its expression at different developmental stages.

The chitin deacetylases Serpentine and Vermiform

Chitin modification by the two secreted chitin deacetylases Serpentine (Serp) and Vermiform (Verm) have been reported to be essential for chitin organisation (Gangishetti et al, 2012; Luschnig et al, 2006). The *A. mellifera GB45151* locus codes for a protein with 532 amino acids, which is 89,9% identical and 95% similar to the *D. melanogaster* Serp protein (isoform B: 541 amino acids) over the entire protein sequence (516 residues), excluding the N-terminal signal peptide (Figure 4). Over 511 residues the GB45151 protein is 61,4% similar and 77,5% identical to Verm (549 aa). The neighbouring locus *GB45152* encodes another chitin deacetylase with 549 amino acids that over 519 residues are 84% similar and 91,3% identical to Verm, and over 545 residues 56,1% similar and 73,4% identical to Serp.

The putative *A. mellifera* Serp (GB45151) and Verm (GB45152) orthologues are differentially expressed (Figure 2). While the *GB45151* transcript is detected at pupal stages, the *GB45152* transcript is larval specific. By contrast, in *D. melanogaster verm* and *serp* are expressed concomitantly at embryonic stages and act redundantly (Luschnig et al, 2006). The tandem organisation of *GB45151* (*Amserp*) and *GB45152* (*Amverm*) in *A. mellifera* and *serp* and *verm* in *D. melanogaster* suggests an early duplication of a chitin deacetylase-coding gene in the insect linage. These genes have probably a common enhancer in *D. melanogaster*, whereas in *A. mellifera* different enhancers control their expression at different stages. This observation suggests a level of complexity in chitin organisation that is missing in *D. melanogaster*.

The chitin organising factor Knickkopf

The *D. melanogaster* Knickkopf protein is a membrane-anchored protein that is needed for correct chitin organisation during larval cuticle formation (Moussian et al, 2006). In the flour beetle *T. castaneum*, Knk additionally protects chitin against degradation by chitinases before moulting (Chaudhari et al, 2011). The *A. mellifera GB50061* locus consisting of 13 exons codes for a protein with 686 amino acids that

displays 57,6% identity and 72,4% similarity over its entire length to the *D. melanogaster* Knk protein (689 amino acids, Figure 5). The *D. melanogaster* Knk protein has an N-terminal signal peptide, a tandem of DM13 domains, a central DOMON domain, and a C-terminal signature allowing the addition of a GPI anchor. GB50061 (AmKnk) is predicted to have an N-terminal signal peptide that is cleaved at Gly30. A tandem of DM13 domains (residues 52-157 and 167-273) and the DOMON domain (residues 288-443) are present in GB50061, as well. In addition, the C-terminus possesses the signature for GPI anchor attachment (omega-site 663). Taken together, the overall architectures of GB50061 and Knk are identical. The transcript of *GB50061* is detected at larval and pupal stages (Figure 2).

The chitin organising factor Retroactive

The *D. melanogaster* Retroactive (Rtv) protein is a membrane-bound Ly6-type protein that assist chitin organisation during embryogenesis (Moussian et al, 2005b; Moussian et al, 2006). In *T. castaneum*, Rtv assists trafficking of Knk to the apical plasma membrane and the extracellular space (Chaudhari et al, 2013). The *A. mellifera GB49764* locus that contains three exons codes for a protein with 146 amino acids displaying high similarity to the *D. melanogaster* Rtv protein (Figure 6). Over the length of 136 amino acids, this protein is 52,2% identical and 66,9% similar to the *D. melanogaster* Rtv protein (151 amino acids). GB49764 (AmRtv) has an N-terminal signal peptide and a C-terminal signature for GPI modification (omega-site N122). Structure-based sequence homology searches using the HHpred software (Soding et al, 2005) confirm that the sequence in-between the termini adopts a Ly6-like folding (data not shown), which in some cases has been shown to mediate partner recognition (Moussian et al, 2005b). Taken together, like *D. melanogaster* Rtv the *A. mellifera* orthologue is predicted to be directed to the plasma membrane where it presumably interacts with a partner to organise chitin microfibrils. Consistently, the *GB49764* transcript is detected at larval and pupal stages (Figure 2).

The chitin-binding factor Obstuctor-A

Obstractor-A (Obst-A) is an extracellular chitin-binding protein that interacts with Knk and Serp to protect chitin from degradation in *D. melanogaster* (Petkau et al, 2012). The Obst-A protein has 237 residues constituting an N-terminal signal peptide and three class 2 chitin-binding domains (CBD2, Figure 7). The *A.mellifera* potential

orthologous sequence with 233 amino acids that is coded by the locus *GB50636* (*AmobstA*) has the same domain composition. The proteins align over their entire length with 64,6% identical and 74,7 % similar residues. *GB50636* is expressed at larval and pupal stages. Thus, *GB50636* (AmObstA), *GB50061* (AmKnk), *GB49764* (AmRtv) are expressed at the same time points suggesting that the respective proteins may interact in *A. mellifera* as they do in *D. melanogaster* and *T. castaneum* to organise the chitin matrix (Figure 8). AmObstA and AmKnk may also interact with AmSerp in pupae. Since AmSerp is not expressed at the larval stage tested, another chitin deacetylase has to be postulated to be the partner of AmKnk and AmObstA in larvae.

The transcription factor Grainyhead

The *D. melanogaster grainyhead* (*grh*) locus is complex. According to FlyBase, the *grh* gene has 19 exons coding for eight isoforms of a CP2-type transcription factor ranging from 155 to 1333 residues (isoforms PH-PO) that are needed for terminal epidermal and tracheal differentiation and in a subset of neurons in the *D. melanogaster* embryo (Bray & Kafatos, 1991; Gangishetti et al, 2012; Uv et al, 1997). Experimentally, four isoforms (N, N', O and O') have been described coded by 17 exons (Uv et al, 1997). The N-isoforms (N 1063 and N' 1032 residues) are epidermal, while the O-isoforms (O 1333 and O' 1303 residues) are expressed in the nervous system. The *grh* homologous *A. mellifera* locus *GB46725* at BeeBase has 11 exons coding for a protein with 843 amino acids (Figure 9). The *GB46725* (*Amgrh*) transcript is detected at larval and pupal stages (Figure 2). Whole transcriptome shotgun sequencing (RNAseq) data at BeeBase indicate the additional existence of an alternative isoform with 636 amino acids. Alignment of the large *A. mellifera* protein with the *D. melanogaster* Grh epidermal isoform N reveals an identity of 61,4% and a similarity of 69,2% that is confined to the C-terminal half of the *D. melanogaster* protein (after around 600 amino acids, see Suppl. Fig. 3). The CP2 domains in the middle of the proteins are less divergent and share 85,8% similarity and 90,7% identity (Figure 9). The N-terminal half of the *D. melanogaster* N and O isoforms is characterised by the presence of several stretches of multiple glutamines (12%), which have been shown to enhance the activity of transcription factors (Atanesyan et al, 2012). The non-aligning N-terminal region of the *A. mellifera* Grh-homologous protein GB46725 is not glutamine- but serine-rich (15,5%), which in

Sp1-related transcription factors are needed for transcription activation (Suske, 1999). Consistently, the N-terminal half of the *B. impatiens* Grh-homologue Bimp10162 is serine-rich, as well (data not shown). Because of the comparably shorter N-terminal region, it is possible that the GB46725 protein might have additional unmatchable and low-complex N-terminal sequences that have escaped annotation. Scanning of the 5-prime region as well as the large first and second introns of *GB46725* for repetitive sequences failed to identify undetected exons coding for such sequences manually and by using the software Fgenesh+ (www.softberry.com) and Genscan (http://genes.mit.edu/GENSCAN.html) (data not shown).

Overall, diversity of *D. melanogaster* Grh isoforms relies mainly on variation of the sequence N-terminal to the CP2 domain. Consistently, organisation of exons coding for the N-terminal half of the *D. melanogaster* protein is complex with eight exons and three alternative translation-starting sites. In comparison, the respective region in the genome of *A. mellifera* is rather simple with four exons and two translation-starting sites. Considering the non-homologous N-terminal sequences of *D. melanogatser* and *A. mellifera* Grh proteins, this region of the protein seems to be frequently modified during evolution. This is obviously a common difference between CP2 transcription factors in *D. melanogaster* and *A. mellifera*. The closest homologue of Grh, the CP2 transcription factor Gemini (Gem) has a glutamine-rich N-terminal half in *D. melanogaster*, whereas in *A. mellifera* the respective region of the annotated Gem orthologue GB48238 is arginine-rich (data not shown).

Conclusion and Outlook

Despite the identical set of chitin producing and organising factors in the fruit fly and the honeybee, differences exist in their sequence composition (*kkv* and *grh*) and expression pattern (*serp* and *verm*). Recent publications on honeybee cuticle proteins underline considerable variances in cuticle composition in these two species. Whereas, for example, the *D. melanogaster* genome harbours 27 Tweedle proteins, only two have been shown to be present in the *A. mellifera* genome (Guan et al, 2006; Soares et al, 2011). Moreover, sequences homologous to the three *A. mellifera* low-complex cuticle proteins called apidermins are missing in the *D. melanogaster* genome (Kucharski et al, 2007). Hence, data from the fruit fly are only partially translatable to other insects. Dissection of the molecular mechanisms of

honeybee cuticle formation combined with a thorough histological study is crucial to contribute to the understanding of honeybee biology and ecology. In particular, we expect to learn more about the interaction between honeybee and the pathogen vector *Varroa destructor* that transmits viruses through the cuticle (Rosenkranz et al, 2010).

Experimental procedures

Animal husbandry

Honey bee (*Apis mellifera*) larvae and pupae were obtained from the Apicultural State Institute at the University of Hohenheim, Germany.

Bioinformatics

Transcripts of *D. melanogaster* genes were translated and aligned to conceptual *A. mellifera* protein sequences by tBlastx at the BeeBase (http://hymenopteragenome.org/beebase/?q=apis_blast). Retrieved genomic sequences were processed with the SerialCloner 2-6 software. Alignment of potential *A. mellifera* protein sequences with *D. melanogaster* protein sequences were performed at FlyBase using Blastp. *A. mellifera* protein sequences at GeneBank were identified by Blastp against *Apis mellifera* (taxon 7460) proteins.

Molecular biology

Total RNA was extracted from larvae (50 animals) and yellow-eyed pupae (2 animals) using the Qiagen RNEasy kit (Hilden, Germany). RNA extraction from pupae was carried out twice. Total RNA was applied to produce cDNA using the Roche Transcriptor First Strand cDNA Synthesis Kit (Mannheim, Germany). Quantitative real-time PCR (qPCR) was performed on a LightCycler Nano from Roche using the Roche FastStart SYBR Green Master kit (Mannheim, Germany). Data were analysed with the respective software and Microsoft Excel. The primers used for transcript amplification were designed with Primer3 software and are listed in table 1. For normalisation, primers to amplify the *tubulin* (*GB44134,* orthologue of *D. melanogaster tub56*) transcripts were used in each experiment. The *GB44134* amplification was performed for each experiment; supplementary figure 2C shows one example.

References

Arakane Y, Hogenkamp DG, Zhu YC, Kramer KJ, Specht CA, Beeman RW, Kanost MR, Muthukrishnan S (2004) Characterization of two chitin synthase genes of the red

296 flour beetle, Tribolium castaneum, and alternate exon usage in one of the genes
297 during development. *Insect Biochem Mol Biol* 34: 291-304
298 Arakane Y, Muthukrishnan S, Kramer KJ, Specht CA, Tomoyasu Y, Lorenzen MD,
299 Kanost M, Beeman RW (2005) The Tribolium chitin synthase genes TcCHS1 and
300 TcCHS2 are specialized for synthesis of epidermal cuticle and midgut peritrophic
301 matrix. *Insect Mol Biol* 14: 453-463
302 Araujo SJ, Aslam H, Tear G, Casanova J (2005) mummy/cystic encodes an enzyme
303 required for chitin and glycan synthesis, involved in trachea, embryonic cuticle and
304 CNS development--analysis of its role in Drosophila tracheal morphogenesis. *Dev*
305 *Biol* 288: 179-193
306 Atanesyan L, Gunther V, Dichtl B, Georgiev O, Schaffner W (2012) Polyglutamine
307 tracts as modulators of transcriptional activation from yeast to mammals. *Biol Chem*
308 393: 63-70
309 Bray SJ, Kafatos FC (1991) Developmental function of Elf-1: an essential
310 transcription factor during embryogenesis in Drosophila. *Genes Dev* 5: 1672-1683
311 Chaudhari SS, Arakane Y, Specht CA, Moussian B, Boyle DL, Park Y, Kramer KJ,
312 Beeman RW, Muthukrishnan S (2011) Knickkopf protein protects and organizes
313 chitin in the newly synthesized insect exoskeleton. *Proc Natl Acad Sci U S A* 108:
314 17028-17033
315 Chaudhari SS, Arakane Y, Specht CA, Moussian B, Kramer KJ, Muthukrishnan S,
316 Beeman RW (2013) Retroactive maintains cuticle integrity by promoting the
317 trafficking of Knickkopf into the procuticle of Tribolium castaneum. *PLoS Genet* 9:
318 e1003268
319 Gangishetti U, Veerkamp J, Bezdan D, Schwarz H, Lohmann I, Moussian B (2012)
320 The transcription factor Grainy head and the steroid hormone ecdysone cooperate
321 during differentiation of the skin of Drosophila melanogaster. *Insect Mol Biol* 21: 283-
322 295
323 Guan X, Middlebrooks BW, Alexander S, Wasserman SA (2006) Mutation of
324 TweedleD, a member of an unconventional cuticle protein family, alters body shape
325 in Drosophila. *Proc Natl Acad Sci U S A* 103: 16794-16799
326 Hogenkamp DG, Arakane Y, Zimoch L, Merzendorfer H, Kramer KJ, Beeman RW,
327 Kanost MR, Specht CA, Muthukrishnan S (2005) Chitin synthase genes in Manduca
328 sexta: characterization of a gut-specific transcript and differential tissue expression of
329 alternately spliced mRNAs during development. *Insect Biochem Mol Biol* 35: 529-540

330 Irion U (2012) Drosophila muscleblind codes for proteins with one and two tandem
331 zinc finger motifs. *PLoS One* 7: e34248
332 Kucharski R, Maleszka J, Maleszka R (2007) Novel cuticular proteins revealed by the
333 honey bee genome. *Insect Biochem Mol Biol* 37: 128-134
334 Luschnig S, Batz T, Armbruster K, Krasnow MA (2006) serpentine and vermiform
335 encode matrix proteins with chitin binding and deacetylation domains that limit
336 tracheal tube length in Drosophila. *Curr Biol* 16: 186-194
337 Merzendorfer H (2006) Insect chitin synthases: a review. *J Comp Physiol B* 176: 1-15
338 Moussian B (2008) The role of GlcNAc in formation and function of extracellular
339 matrices. *Comp Biochem Physiol B Biochem Mol Biol* 149: 215-226
340 Moussian B (2010) Recent advances in understanding mechanisms of insect cuticle
341 differentiation. *Insect Biochem Mol Biol* 40: 363-375
342 Moussian B, Schwarz H, Bartoszewski S, Nusslein-Volhard C (2005a) Involvement of
343 chitin in exoskeleton morphogenesis in Drosophila melanogaster. *J Morphol* 264:
344 117-130
345 Moussian B, Soding J, Schwarz H, Nusslein-Volhard C (2005b) Retroactive, a
346 membrane-anchored extracellular protein related to vertebrate snake neurotoxin-like
347 proteins, is required for cuticle organization in the larva of Drosophila melanogaster.
348 *Dev Dyn* 233: 1056-1063
349 Moussian B, Tang E, Tonning A, Helms S, Schwarz H, Nusslein-Volhard C, Uv AE
350 (2006) Drosophila Knickkopf and Retroactive are needed for epithelial tube growth
351 and cuticle differentiation through their specific requirement for chitin filament
352 organization. *Development* 133: 163-171
353 Munoz-Torres MC, Reese JT, Childers CP, Bennett AK, Sundaram JP, Childs KL,
354 Anzola JM, Milshina N, Elsik CG (2011) Hymenoptera Genome Database: integrated
355 community resources for insect species of the order Hymenoptera. *Nucleic Acids*
356 *Res* 39: D658-662
357 Pare A, Kim M, Juarez MT, Brody S, McGinnis W (2012) The functions of grainy
358 head-like proteins in animals and fungi and the evolution of apical extracellular
359 barriers. *PLoS One* 7: e36254
360 Petkau G, Wingen C, Jussen LC, Radtke T, Behr M (2012) Obstructor-A is required
361 for epithelial extracellular matrix dynamics, exoskeleton function, and tubulogenesis.
362 *J Biol Chem* 287: 21396-21405

363 Rosenkranz P, Aumeier P, Ziegelmann B (2010) Biology and control of Varroa
364 destructor. *J Invertebr Pathol* 103 Suppl 1: S96-119
365 Schimmelpfeng K, Strunk M, Stork T, Klambt C (2006) Mummy encodes an UDP-N-
366 acetylglucosamine-dipohosphorylase and is required during Drosophila dorsal
367 closure and nervous system development. *Mech Dev* 123: 487-499
368 Soares MP, Silva-Torres FA, Elias-Neto M, Nunes FM, Simoes ZL, Bitondi MM
369 (2011) Ecdysteroid-dependent expression of the tweedle and peroxidase genes
370 during adult cuticle formation in the honey bee, Apis mellifera. *PLoS One* 6: e20513
371 Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein
372 homology detection and structure prediction. *Nucleic Acids Res* 33: W244-248
373 Suske G (1999) The Sp-family of transcription factors. *Gene* 238: 291-300
374 Tonning A, Helms S, Schwarz H, Uv AE, Moussian B (2006) Hormonal regulation of
375 mummy is needed for apical extracellular matrix formation and epithelial
376 morphogenesis in Drosophila. *Development* 133: 331-341
377 Uv AE, Harrison EJ, Bray SJ (1997) Tissue-specific splicing and functions of the
378 Drosophila transcription factor Grainyhead. *Mol Cell Biol* 17: 6727-6735
379 Van Leeuwen T, Demaeght P, Osborne EJ, Dermauw W, Gohlke S, Nauen R, Grbic
380 M, Tirry L, Merzendorfer H, Clark RM (2012) Population bulk segregant mapping
381 uncovers resistance mutations and the mode of action of a chitin synthesis inhibitor
382 in arthropods. *Proc Natl Acad Sci U S A* 109: 4407-4412
383
384

Table

Table 1. Primers used for amplification by quantitative PCR

locus	forward primer	reverse primer
Amkkv *5' NCBI*	CCGAACAAATTAATCCGATTTC	TTGTGTAGAAACATTCAACCATTTT
5' BeeBase	TGATCGATGTTCAATTACTTACTTTG	TTCCTTCCACAGCCATTTTC
exon 8-9	TGTAAGAAGCAGCGATCACG	TGTTCACTTGCGACTCGTTC
exon 14	ATCGTAGGTACGGCTCTCCA	TCCCCATGAGACAACATTCA
exon 15	GGGACTGCGATTAACATCGT	CCCTTGTGCCCCATGTAATA
exons 14-15	GTTGTCTCATGGGGAACTCG	AACGATCACAGCGACCATTA
exon 18	CCTACGTGACCAGAGCGTTT	TCGTGAAAGTTATCGGCATT
exon 19	TCGAGCTACGGAACAAGAGC	CCCCGTCTTCTTTCATGGT
exons 18-19 a	AATGCCGATAACTTTCACGA	ATTAGGTCGCCGGCAATAC
exons 18-19 b	AATGCCGATAACTTTCACGA	ATTAGGTCGCCGGCAATAC
Amserp *exons 4-5*	CCTATGGCCCTACACCATGT	GTAATGGCGGTCGAAGTTGT
Amverm *exons 4-5*	GAGGGCAAATTATCGTGTGG	CGTCGGCTGAACAGTAACAG
Amknk *exons 5-8*	GCGTCAAATTTTGGCTCTGT	TCTTTAAATGCGGTGTGCAA
Amrtv *exons 1-2*	GCACAAGTTTTCTTAATATCCACGA	TCCATTTTAATCGCCGTTGT
Amobst-A *exons 4-5*	CCACCGGATTACATTTCGAC	AATTTCGGATGATCGACGAC
Amgrh *exons 5-6*	GTCTGAGCACCGACTTCTCC	TTCTCTTCGCTGCTCTCCTC
exons 7-8	CGAGAAGAGAAGAGCCCAGA	GTTCCAATAAACGGCAATCG
Amtubulin *(GB44134)*	TCACTCATTCGGTGGTGGTA	AGATTGCGTCGGCAAATATC

Figure legends

Figure 1. The *chitin synthase*/*kkv* locus.

A) The *A. mellifera chitin synthase* gene has 22 exons (orange boxes), whereas the *D. melanogaster kkv* gene has 12 exons. In *D. melanogaster*, exons one and two are separated by two genes, *CG14668* and *CR31593* (light grey). The chitin synthase signature (Glycosyltransferase like family 2: PF13641) is encoded by four exons (exons 9-12), whereas in *D. melanogaster* this domain is encoded by only one exon (exon 6). B) The respective amino acid sequences share 95.5% identical and 97,5% similar residues. C) In contrast to the situation in *D. melanogaster*, the sequence coding for the WGTRE motif that has been proposed to be involved in chitin extrusion through the plasma membrane is separated from the signature coding exons by an interjacent exon. Interestingly, the WGTRE coding exon is duplicated in *A. mellifera*. As shown by our expression analyses, the respective exons 14 and 15 are alternatively spliced. The exon 14 encoded protein sequence displays a higher identity (88,4%) and similarity (93,7%) to the respective *D. melanogaster* sequence than does the exon 15 encoded sequence (53,7% and 77,9%). The sizes of introns are indicated between the exon boxes.

Figure 2. Expression profiles of chitin producing and organising factors in *A. mellifera*.

406 The expression of genes coding for chitin producing and organising factors in *A.*
407 *mellifera* was recorded in first instar larvae and yellow-eyed pupae by quantitative
408 PCR (qPCR) using the primers listed in table 1. The values were normalised against
409 the amount of *tubulin* transcripts, i.e. a value of 1 means no difference to *tubulin*
410 expression. Taken together, since the abundance of the *tubulin* transcript can be
411 considered as similar at both stages tested, expression levels of genes coding for
412 chitin synthesising and organising factors are generally lower at the pupal stage than
413 at the larval stage.
414 The value of zero (0) indicates no expression. Expressions of the 5' end of *Amkkv*
415 (*kkv* 5' Beebase and *kkv* 5' NCBI) and of exons 5 and 6 of *Amgrh* were not
416 determined in pupae (n.d.). Examples of amplification curves are shown in
417 supplementary figures 2A-D.

418 Figure 3. The *mmy* locus.
419 The *D. melanogaster mmy* locus has two alternative start codons (ATG) and three
420 coding exons (orange boxes). The *A. mellifera* gene *GB44897* has only one
421 predicted start codon and three coding exons. RNA sequencing (RNAseq) data at
422 BeeBase do not contradict the comparably simple composition of the *GB44897*
423 locus. The sizes of introns are indicated between the exon boxes.

424 Figure 4. The *serp and verm* loci.
425 The *D. melanogaster* genes *serp* and *verm* are arranged in tandem, both coding for
426 chitin deacetylases composed of an N-terminal chitin-binding domain (CBD,
427 PF01607, lined box) followed by an LDLa motif (PF00059, dark grey box) that
428 precedes the enzyme signature (PF01522, light grey box). In *A. mellifera*, the genes
429 coding for the Serp and Verm orthologues are also arranged in tandem. The
430 *GB45151* locus coding for the Serp orthologue has five exons (orange boxes). The
431 *GB45152* locus that encodes the Verm orthologue has six exons. The exons 3, 5 and
432 6 of the *D. melanogaster verm* may be alternatively spliced. Exons 5 and 6
433 correspond to exons 2 and 3 in *A. mellifera*, arguing that they may be alternatively
434 spliced, as well. By contrast, *verm* exon 3 does not show any homology to any
435 sequence in the *A. mellifera* genome. The sizes of introns are indicated between the
436 exon boxes.

437 Figure 5. The *knk* locus.

A) *GB50061* has 13 exons (orange boxes) adding up to 2058 bps of coding sequence. The resulting protein has 686 amino acids. The *D. melanogaster knk* gene has 6 exons constituting a coding region of 2067 bps giving rise to a protein with 689 residues. B) These proteins align over 658 residues. Both proteins have an N-terminal signal peptide (SP), followed by two DM13 domains (PF10517), a central DOMON domain (PF03351) and a C-terminal signature for GPI modification (ω). The DM13 domains are 66 and 56% identical, and 80,2 and 74,8% similar to each other, respectively. The DOMON domains share 64,1% identical and 78,8% similar amino acids. The sizes of introns are indicated between the exon boxes.

Figure 6. The *rtv* locus.

GB49764 has three exons (orange boxes) that together constitute an open reading frame of 438 bps. The 453 bps open reading frame of the *D. melanogaster rtv* gene is split into two exons. The position of the first intron is conserved separating the respective sequences coding for the signal peptide from the Ly6 domain (PF00021). The sizes of introns are indicated between the exon boxes.

Figure 7. The *obst-A* locus.

A) The *A. mellifera GB50636* locus that codes for an Obst-A homologous protein has five exons whereas the *D. melanogaster* reference locus has four exons (orange boxes). B) The GB50636 protein and Obst-A are co-linear and have both three chitin-binding domains (I-III). The sequences C-terminal to the last chitin-binding domain are highly conserved (underlined) suggesting an important function in organising chitin. The sizes of introns are indicated between the exon boxes.

Figure 8. Scheme postulating interactions between chitin producing and organising factors in *A. mellifera*.

In the pupa, AmRtv is required to deliver AmKnk to the extracellular space. There, AmKnk interacts with AmObst-A and AmSerp to organise chitin and to protect it against chitinase before moulting.

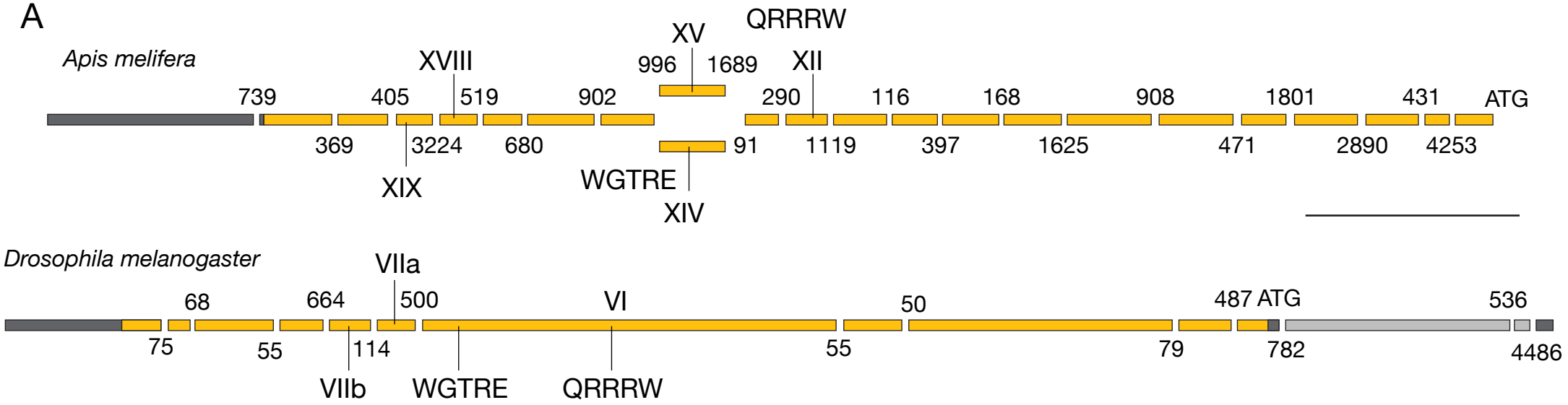
Figure 9. The *grh* locus.

A) The *A. mellifera* locus *GB46725* is predicted to have 11 exons (orange boxes) coding for a protein with 843 amino acids. According to the Flybase database, the *D. melanogaster grh* gene has 19 exons coding for eight isoforms ranging from 155 to 1333 residues (isoforms PH-PO). Alignment of *A. mellifera* protein with the *D.*

melanogaster Grh isoform PK that has a similar size (784 amino acids), reveals an identity of 61,4% and a similarity of 68,2%. Sequence comparison with the longer *D. melanogaster* isoforms shows that homology is confined to the C-terminal half of the protein.

B) The CP2 domains (PF004516) of *D. melanogaster* and *A. mellifera* are 85,8% identical and 90,7% similar to each other. The sizes of introns are indicated between the exon boxes.

A



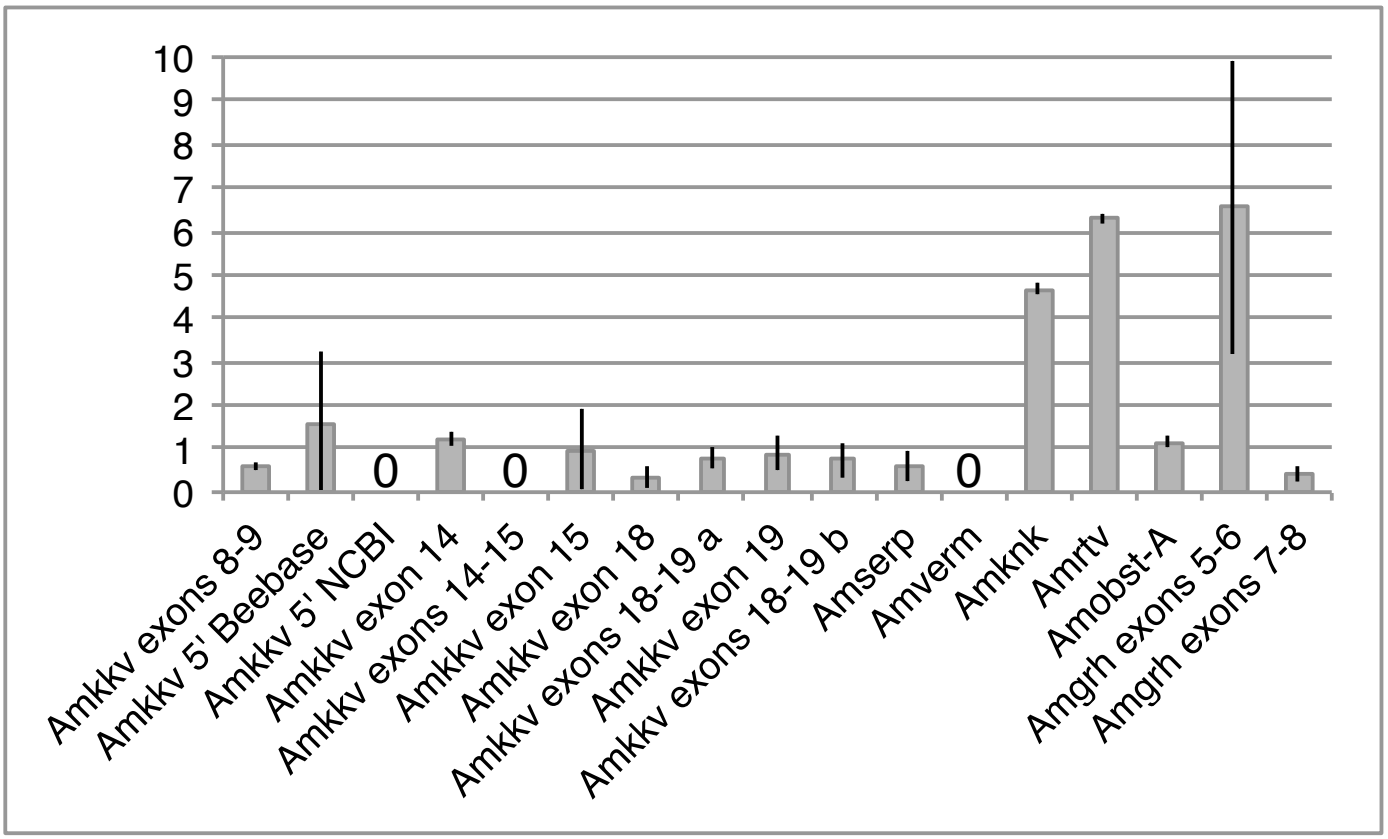
B

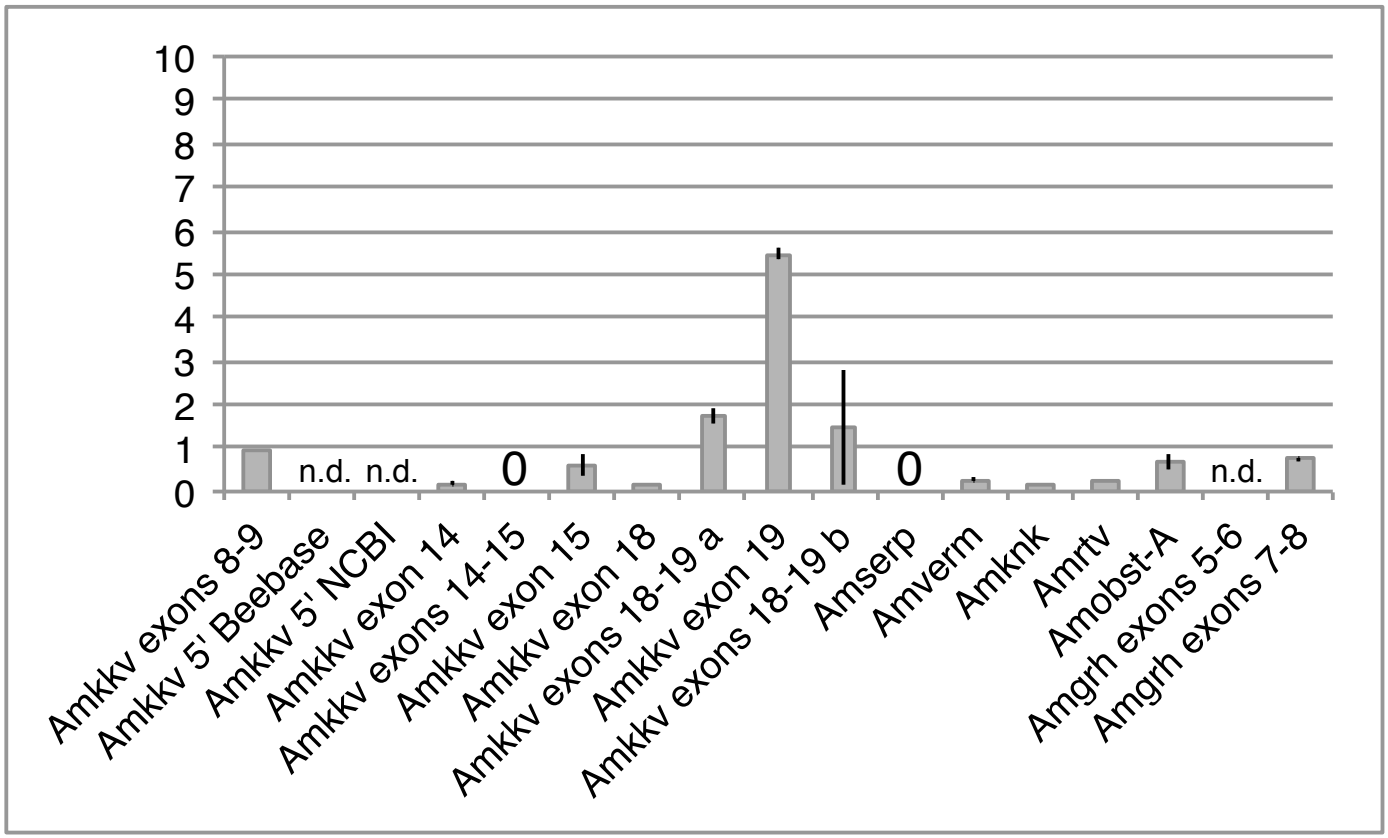
```
Dmel-AC DKDRIRHRKRWSQVMYMYYLLGHRLMELPISVDRKDAIAENTYLLTLDGDIDFKPNAVTL 60
Amel-AC DKSKIRHRKRWSQVMYMYYLLGHRLMELPISVDRKEVIAENTYLLTLDGDIDFQPAAVKL 60
        **.:*******************************:.***************:* **.*

Dmel-AC LVDLMKKNKNLGAACGRIHPVGSGPMVWYQLFEYAIGHWLQKATEHMIGCVLCSPGCFSL 120
Amel-AC LVDLMKKNKNLGAACGRIHPVGSGPMVWYQMFEYAIGHWLQKATEHMIGCVLCSPGCFSL 120
        ****************************:****************************

Dmel-AC FRGKALMDDNVMKKYTTRSDEARHYVQYDQGEDRWLCTLLLQRGYRVEYSAASDAYTHCP 180
Amel-AC FRGKALMDDNVMKKYTTRSDEARHYVQYDQGEDRWLCTLLLQRGYRVEYSAASDAYTHAP 180
        ************************************************************.*

Dmel-AC EGFNEFYNQRRRWVPSTIAN 200
Amel-AC EGFNEFYNQRRRWVPSTIAN 200
        ********************
```

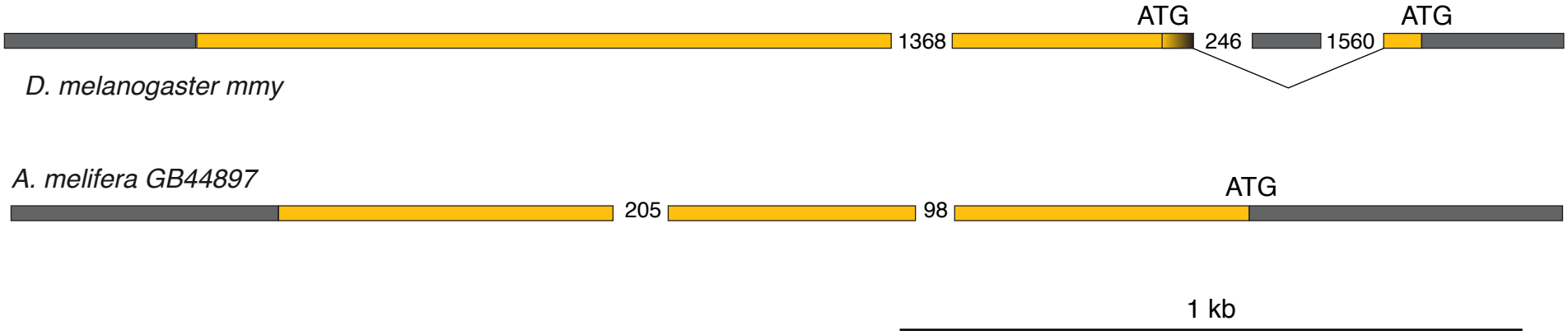
C

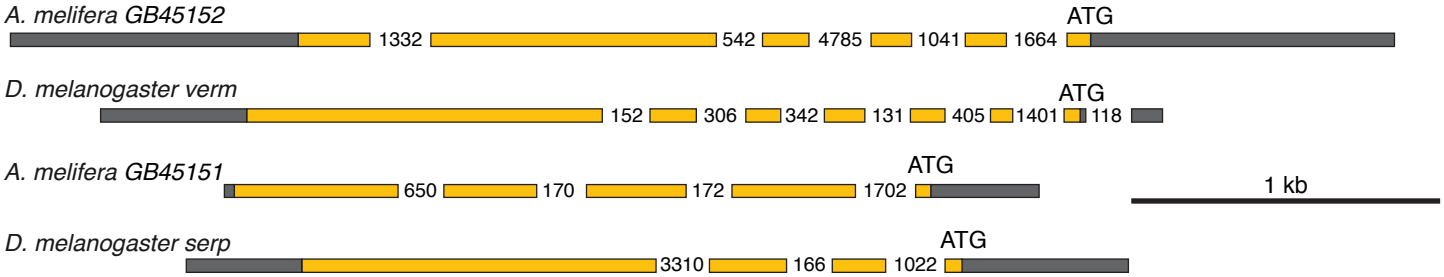
```
Dmel    LFVAQVLSTAYALIMMAVIVGTALQLGEDGIGSPSAIFLISMVGSFFIAACLHPQEFWCI 60
Amel14  LLCAQILSTGYAMIMMAVIVGTALQLGEDGIGSPSAIFLIALSGSFFIAACLHPQEFWCI 60
Amel15  VFVAAIISTLYGMIMVAVIVGTAINIVQDGILSPSAMFLMLLMSQIVITAILHTEEIFCI 60
        :: * ::** *.:**:*******::: :*** ****:**: : ..:.*:* **.:*::**

Dmel    TCGLIYLLSIPSMYLLLILYSIINLNVVSWGTREV------- 95
Amel14  VPGIIYLLSIPSMYLLLILYSIINLNVVSWGTREVQVKKTKK 102
Amel15  IYSVIYFITIPSMYVLLIIYSVCNLNDITWGTREIKTRKTAA 102
          .:**:::*****:***:**: *** ::******:
```

Figure 2

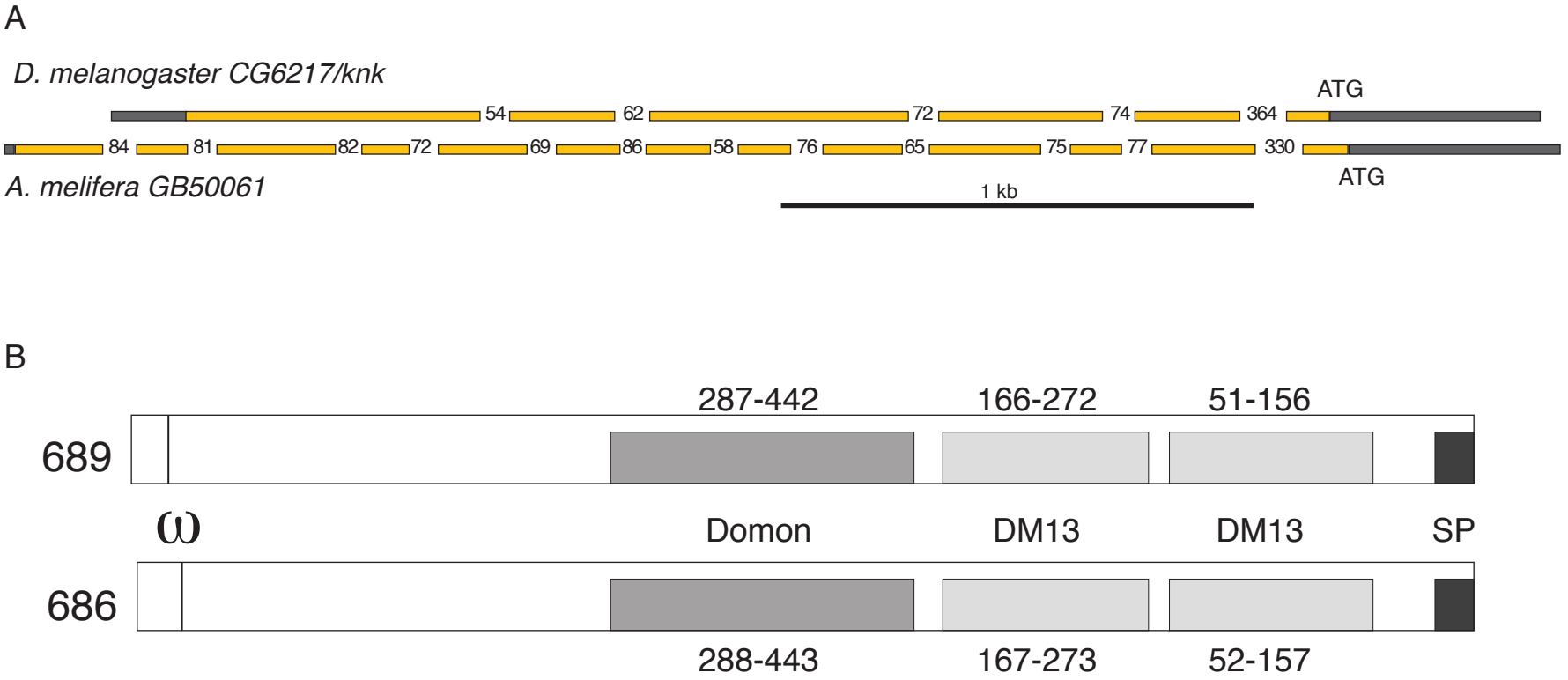

10
9
8
7
6
5
4
3
2
1
0
0
0
0
Amkkv exons 8-9
Amkkv 5' Beebase
Amkkv 5' NCBI
Amkkv exon 14
Amkkv exons 14-15
Amkkv exon 15
Amkkv exon 18
Amkkv exons 18-19 a
Amkkv exon 19
Amkkv exons 18-19 b
Amserp
Amverm
Amknk
Amrtv
Amobst-A
Amgrh exons 5-6
Amgrh exons 7-8


10
9
8
7
6
5
4
3
2
1
0
n.d.
n.d.
0
0
n.d.
Amkkv exons 8-9
Amkkv 5' Beebase
Amkkv 5' NCBI
Amkkv exon 14
Amkkv exons 14-15
Amkkv exon 15
Amkkv exon 18
Amkkv exons 18-19 a
Amkkv exon 19
Amkkv exons 18-19 b
Amserp
Amverm
Amknk
Amrtv
Amobst-A
Amgrh exons 5-6
Amgrh exons 7-8


ATG
ATG
1368
246
1560
D. melanogaster mmy
A. melifera GB44897
ATG
205
98
1 kb




```
DmelVerm        MARWWPILSVVCLCLAVDGLYVKNSFEPAIDQDPSRRYQRKPHGQDKLEGVDVEEVCADR 60
AmelVerm        MTATTTTATMSPRFLLLLGSLFLLAATQVRAQDE--------EGADGID-ANAEELCQDR 51
DmelSerp        MAKLFVVFAVLALAAFNEASASDPLLRYKRQATT---------EETKKEESFEKELCKDK 51
AmelSerp        -MRSSLVLLFLAAIVLAEGTS-----RVKRQ------------EDKKEESFESEICKDK 41
                             .       .                          :    .*:* *:
DmelVerm        PADEYFRLETDG--DCREVYRCTKS------GLKEIQCPSGLAFDVIKQTCDWKAKVTNC 112
AmelVerm        PGDEYFRLNVDG--DCRDVVRCDKASEIGVTRLATVRCPTGLAFDIERQTCDWKTNVKNC 109
DmelSerp        DAGEWFRLVAGEGDNCRDVIQCTSS------GLQAIRCPAGLYFDIEKQTCDWKESVKNC 105
AmelSerp        DAGEWFRLVAGEGDNCRDVIQCTSS------GLQAIRCPAGLYFDIDKQTCDWKDSVNNC 95
                  ..*:*** ..    :**:* :* .:       *  ::**:** **: :****** .*.**
DmelVerm        DEKEKPRKAKPILKTDEPICPEGKLSCGDGECLDKELFCNGKSDCKDESDENACSVDEDP 172
AmelVerm        DQLEKPRKVLPILRTDEPVCPEGKLSCGNGECVDKELFCNGKPDCKDESDENACTVETDP 169
DmelSerp        KSKNKERRVKPLLHTDEPLCQDGFLACGDGNCIERGLFCNGEKDCSDGSDENTCDIDNDP 165
AmelSerp        KLKNKERKAKPLLYTEEPLCQDGFLACGDGSCIERGLFCNGEKDCTDGSDENICDMDNDP 155
                .  :* *:. *:* *:**:* :* *:**:*.*::: *****: **.* **** * :: **
DmelVerm        NRAPECDPTQCALPDCFCSADGTRIPGGIEPQQVPQMITITFNGAVNVDNIDLYEDIFNG 232
AmelVerm        NRAPDCDPTQCVLPDCYCSADGTRIPGNIEPSQVPQMITITFNGAVNVDNIDLYEEIFNG 229
DmelSerp        NRAPPCDPAVCVLPDCFCSEDGTSIPGDLPAKDVPMMITITFDDAINNNNIELYKEIFK- 224
AmelSerp        NRAPPCDPSVCVLPDCFCSEDGTTIPGDLPPKDVPQMITITFDDAINNNNIGLYKEIFNG 215
                **** ***: *.****:** *** ***.: ..:** ******:.*:* :** **::**:
DmelVerm        QRQNPNGCSIKGTFFVSHKYTNYSAVQDLHRRGHEISVFSLTHKDDPNYWTGGSYDDWLA 292
AmelVerm        QRQNPNGCQIRGTFFVSHKYTNYSAVQDLHRRGHEIAVFSLTHKDDPQYWTQGSYDDWLA 289
DmelSerp        DRKNPNGCSIKATYFVSHKYTNYSAVQETARKGHEIAVHSITHNDEERFWSNATVDDWAK 284
AmelSerp        KRKNPNGCDIKATFFVSHKYTNYSAVQEMHRKGHEIAVHSISHNDDERFWSDATVDDWAK 275
                .*:*****.*:.*:************:  *:****:*.*::*:*: .:*: .: ***
DmelVerm        EMAGSRLIVERFANITDGSIIGMRAPYLRVGGNKQFEMMADQFFVYDASITASLGRVPIW 352
AmelVerm        EMAGARLIIERFANITDGSIIGMRAPYLRVGGNKQFEMMADQFFVYDASITASLGRVPIW 349
DmelSerp        EMAGMRIITEKFANITDNSVVGVRAPYLRVGGNNQFTMMEEQAFLYDSTITAPLSNPPLW 344
AmelSerp        EMAGMRIIAEKFANLTDNSVVGVRAPYLRVGGNNQFTMMEEQAFLYDSTITAALNNPPLW 335
                **** *:* *:***:**.*::*:********** ** :* *:**::***.*.. *:*
DmelVerm        PYTLYFRMPHKCNGNAHNCPSRSHPVWEMVMNELDRRDDPTFDESLPGCHMVDSCSNVAS 412
AmelVerm        PYTLYFRMPHKCNGNGGNCPSRSHPVWEMVMNELDRRDDPTFDESLPGCHMVDSCSNIQT 409
DmelSerp        PYTMYFRMPHRCHGNLQSCPTRSHAVWEMVMNELDRREDPVNDEYLPGCAMVDSCSNILT 404
AmelSerp        PYTMYFRMPHRCHGNLQHCPTRSHAVWEMVMNELDRREDPQNDEYLPGCAMVDSCSNILT 395
                ***:******:*:**   **:***.***********:**  ** **** *******: :
DmelVerm        GDQFARLLRHNFNRHYNSNRAPLGLHFHASWLKSKKEYRDELIKFIEEMLG-RNDVFFVT 471
AmelVerm        GEQFARLLRHNFNRHFNSNRAPLGLHFHASWLKSKKEYREELIKFIEEMLA-RSDVYFVT 468
DmelSerp        GDQFYNFLNHNFDRHYDQNRAPLGLYFHAAWLKNNPEFLDAFLYWIDEILANHNDVYFVT 464
AmelSerp        GDQFYNFLNHNFDRHYEQNRAPLGLYFHAAWLKNNPEFLDAFLYWIDEVLSNHNDVYFVT 455
                *:** .:*.***:**:: .*******:***:***.: *: : :: :*:*:*. :.**:***
DmelVerm        NLQVIQWMQNPTELNSLRDFQEWKEKCDVKGQPYCSLPNACPLTTRELPGETLRLFTCME 531
AmelVerm        MVQVIKWMQTPTELSALRDFQDWKETCDEKGQPYCSLPNACPLTTRELPGETLRLFTCME 528
DmelSerp        MTQVIQWMQNPRTISEVKNFEPWREKCVVEGKPACWVPNTCKLTSKEVPGETINLQTCVR 524
AmelSerp        MTQVIQWIQNPRTITESKSFEPWKEKCVVDGPPACWVPHTCKLTSKEVPGETINLQTCVR 515
                  ***:*:*.*  :.  :.*: *:*.*  .* * * :*::* **::*:****:.* **:.
DmelVerm        CPNNYPWILDPTGDGFSV--- 549
AmelVerm        CPNYYPWLLDPTGDGFTANKK 549
DmelSerp        CPNNYPWVSDPTGDGFF---- 541
AmelSerp        CPNNYPWVNDPTGDGFF---- 532
                *** ***: *******
```


A
D. melanogaster CG6217/knk
ATG
54
62
72
74
364
84
81
82
72
69
86
58
76
65
75
77
330
ATG
A. melifera GB50061
1 kb
B
287-442
166-272
51-156
689
ω
Domon
DM13
DM13
SP
686
288-443
167-273
52-157

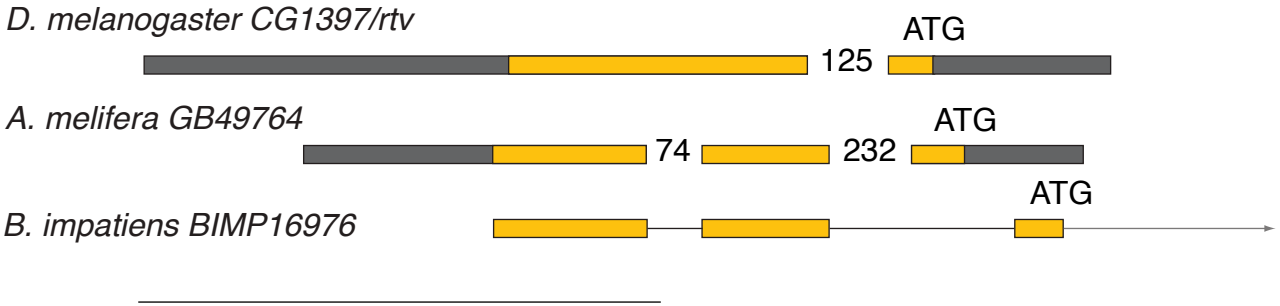
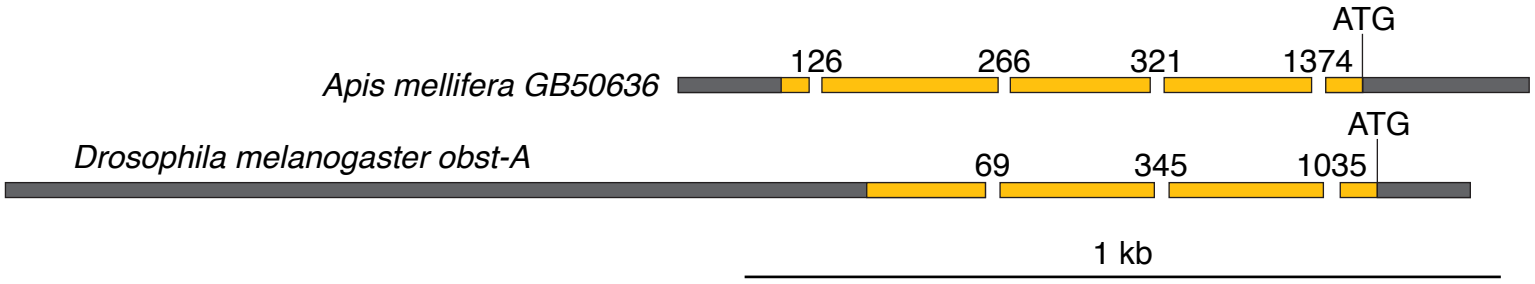

D. melanogaster CG1397/rtv
ATG
125
A. melifera GB49764
ATG
74
232
B. impatiens BIMP16976
ATG

Figure 7

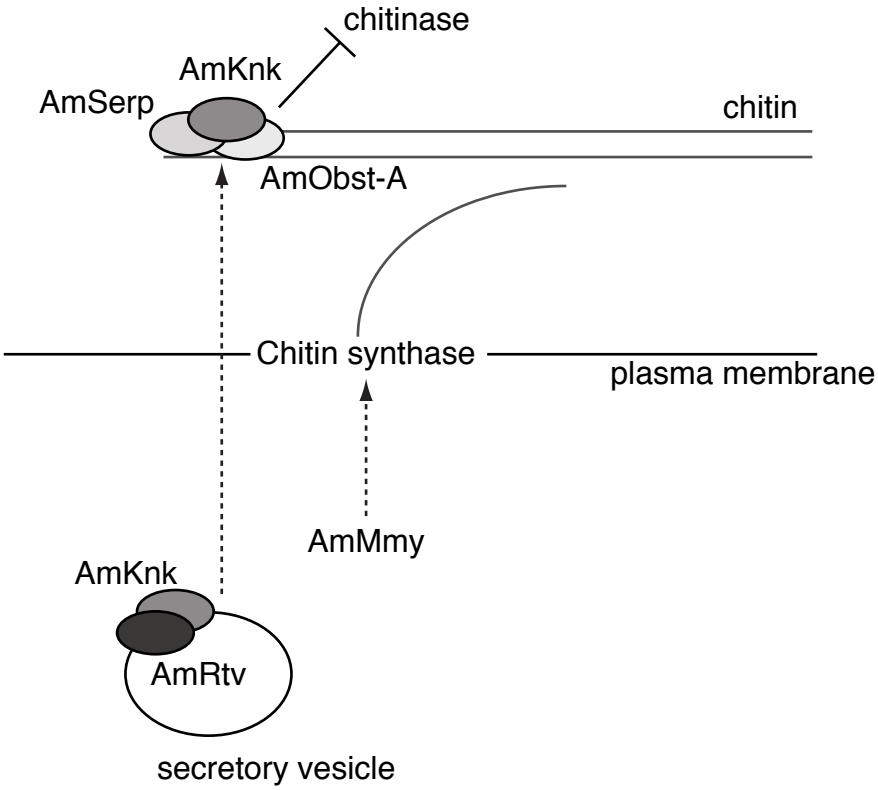
A



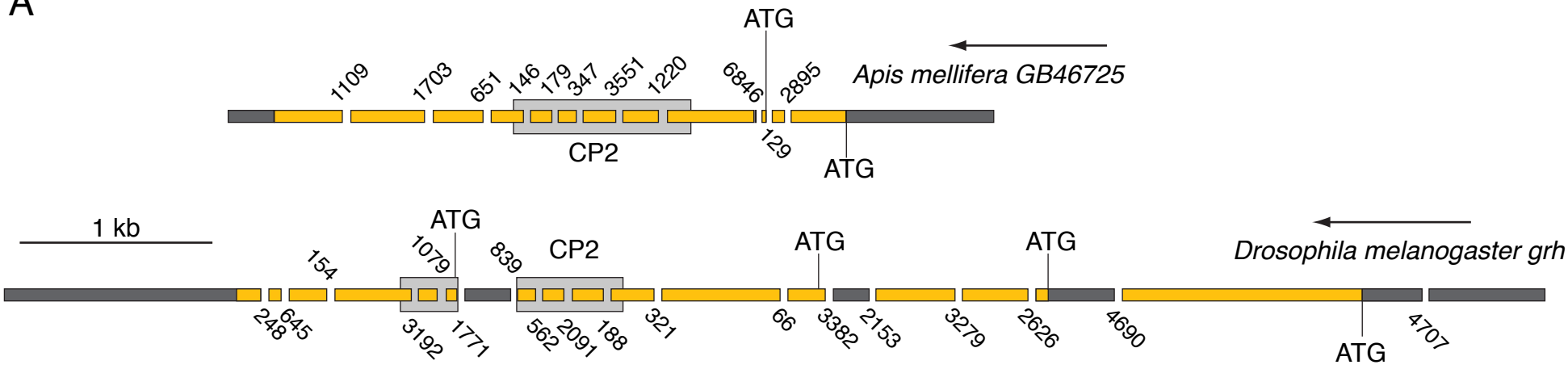
B

```
                                                                I
Amel:      7 VTILAVVAVTPDGA-FNCPSKDGQYEDPKQCDKYYECIDGIATEKLCPDGLVFDPLNRKV 65
               + +   VA T   A F CP  +GQ+ D  QCDK+Y C DG+A  KLCPDGLVFDPLNRK
Dmel:      8 IAVTLCVATTVSAANFECPKPNGQFADEVQCDKFYVCDDGVAKAKLCPDGLVFDPLNRKF 67
                                                                    II
Amel:     66 NKCDHVFNVDCGDRLELQPPQPTKKCPRRNGFFAHPDASVCNIFYNCIDGEAIEITCTTG 125
             NKCD  FNVDC DR ELQ P+ +K CPR+NGFFAHPD +VCNIFYNCI+G+A+E  CT G
Dmel:     68 NKCDQPFNVDCEDRTELQEPKSSKYCPRKNGFFAHPDPAVCNIFYNCIEGDALETKCTVG 127

Amel:    126 LHFDEYSGTCVWPDSAGREGCGVVDKKLKDGFECPRES-QVDTRGMVVDHPKFAHPDDCQ 184
             LHFDEYSGTCVWPD+A REGC    +  + GF CP++  + D RG VV HPK+ HP DCQ
Dmel:    128 LHFDEYSGTCVWPDTAKREGCNPEQRTSETGFVCPKDQPKTDDRGQVVTHPKYPHPTDCQ 187
                                III
Amel:    185 KFYVCLNGVTPREQGCSDGTVYNEEQQRCDAPENVPGCEDWYKD-DDKK 232
             KFYVCLNG  PR+ GC  G VYN+  + CDAPENVPGCEDWYKD DDKK
Dmel:    188 KFYVCLNGEDPRDLGCQLGEVYNDATEMCDAPENVPGCEDWYKDVDDKK 236
```

Figure 8


chitinase
AmKnk
AmSerp
chitin
AmObst-A
Chitin synthase
plasma membrane
AmMmy
AmKnk
AmRtv
secretory vesicle

A



B

```
Dmel    TRSRPWHDFGRQNDADKIQIPKIFTNVGFRYHLESPISSSQRREDDRITYINKGQFYGIT 60
Amel    PRSRPWHDFGRQNDADKIQIPKIFSAYGFKYHLESPISTSQRREDDRITYINKGQFYGIT 60
        .**********************:  **:********:**********************
Dmel    LEYVHDAEKP-IKN-TTVKSVIMLMFREEKSPEDEIKAWQFWHSRQHSVKQRILDADTKN 118
Amel    LDYVPDPDKPSLKAGQTVKSVVMLMFREEKSPEDEIKAWQFWHGRQHSVKQRILDADTKN 120
        *:** *.:** :*   *****:**********************.****************
Dmel    SVGLVGCIEEVSHNAIAVYWNPLESSAKINIAVQCLSTDFSSQKGVKGLPLHVQIDTFED 178
Amel    SVGLVGCIEEVAHNAIAVYWNPLESSAKINVAVQCLSTDFSSQKGVKGLPLHIQVDTYED 180
        ***********:******************:**********************:*:**:**
Dmel    P-----RDTAVFHRGYCQIKVFCDKGAERKTRDEERRAAKRKMTAT 219
Amel    PPPHTHTYTPPSHRGYCQIKVFCDKGAERKTRDEERRAAKRKMTAT 226
        *       *.  ***********************************
```

Supplementary figure 1

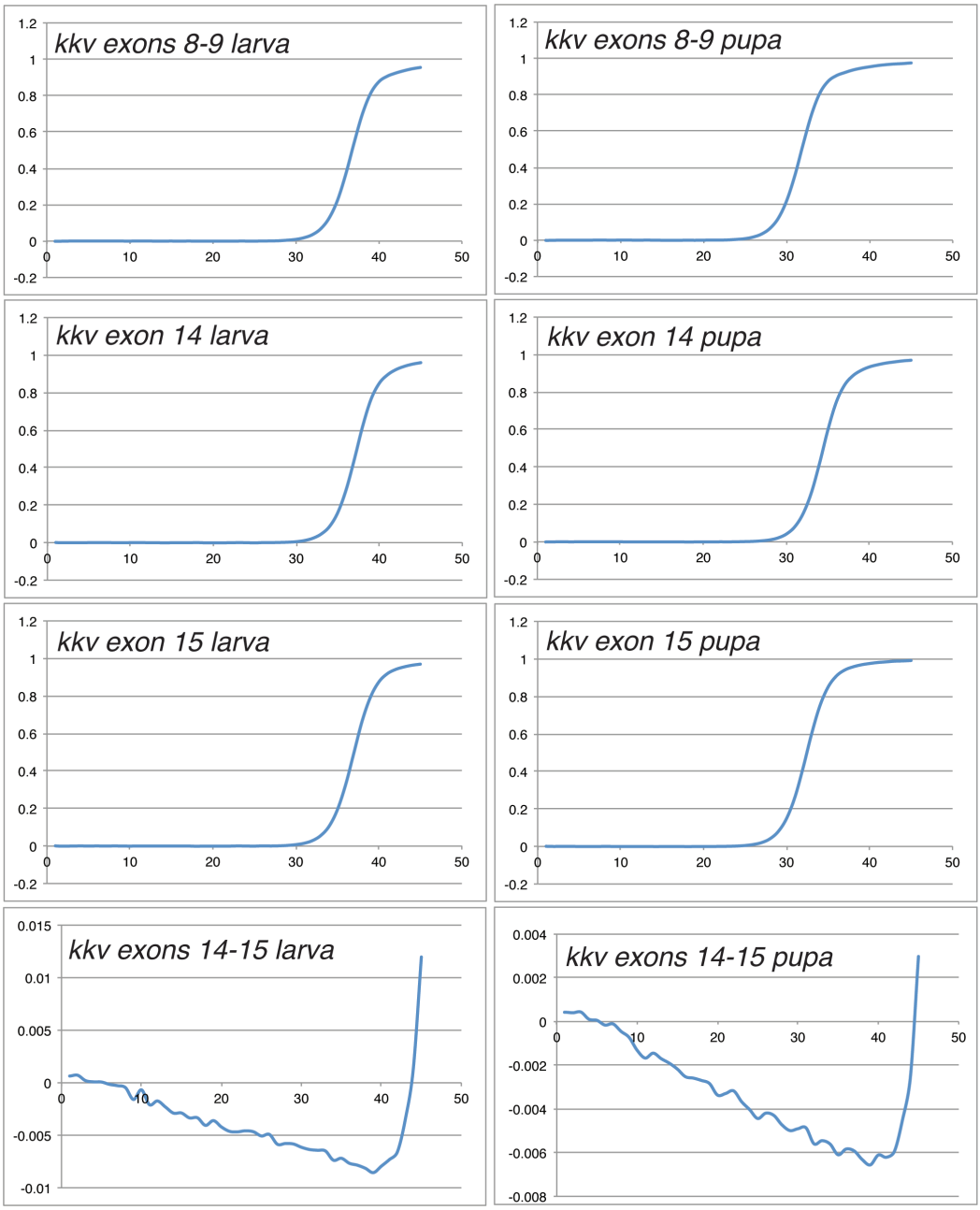
A) Alignment of the *A. mellifera* chitin synthase sequences from BeeBase and from NCBI

```
Beebase         MIDVQLLTLKMKRSKNQKSFERVRGLFQKKIYHLLWLNPKSKEFSIKKMAVEGMYRKGVM 60
NCBI            ------------------------------MILYKTNDSR----------------M 11
                                             :*: . .*:                 *
Beebase         SKIQQQNGMMPGNGTMPDDDDFSDGESTPLTQDYGDSQRTVVETKAWDVFRNPPPKIDSG 120
NCBI            SKIQQQNGMMPGNGTMPDDDDFSDGESTPLTQDYGDSQRTVVETKAWDVFRNPPPKIDSG 71
                ************************************************************
Beebase         SMANQRCLEVTVQITKVIVYLLVFVIVLGSGVVAKGTILFMTSQLRANRTIVHCNRDIGR 180
NCBI            SMANQRCLEVTVQITKVIVYLLVFVIVLGSGVVAKGTILFMTSQLRANRTIVHCNRDIGR 131
                ************************************************************
Beebase         DKYFEVTLPEQERIAWIWCIIIAFAVPEFGTLIRSIRICIFKSWKKPPASHFLVVFVMET 240
NCBI            DKYFEVTLPEQERIAWIWCIIIAFAVPEFGTLIRSIRICIFKSWKKPPASHFLVVFVMET 191
                ************************************************************
Beebase         FHVVGLALMFMAVLPDLDVVKGAMLTNCVCFVPGVLGLLSRNKKKDESRFVLVLIDIAAL 300
NCBI            FHVVGLALMFMAVLPDLDVVKGAMLTNCVCFVPGVLGLLSRNKKKDESRFVLVLIDIAAL 251
                ************************************************************
Beebase         VAQGTSFVLWPLLDSSRFSLWLIPPSLFLVSCGWWENYVSTQSPIGFVRSLGKIKQEMQL 360
NCBI            VAQGTSFVLWPLLDSSRFSLWLIPPSLFLVSCGWWENYVSTQSPIGFVRSLGKIKQEMQL 311
                ************************************************************
Beebase         TRYFTYMFMSVWKIIVFFTSTILILYIKGETVGHLFSMFGDAFGNHTIVVRSMYDVTGRT 420
NCBI            TRYFTYMFMSVWKIIVFFTSTILILYIKGETVGHLFSMFGDAFGNHTIVVRSMYDVTGRT 371
                ************************************************************
Beebase         TDIADIVDIDDNKIAIPANVKSPIYVLLLQIFSAYFMYIFGKFACKILIQGFSYAFPVNL 480
NCBI            TDIADIVDIDDNKIAIPANVKSPIYVLLLQIFSAYFMYIFGKFACKILIQGFSYAFPVNL 431
                ************************************************************
Beebase         TIPVSISLLIAACGLRHTDPCIFHNTIPDYLFYESPPLHFLNDFVSKQYAWVWLLWLLSQ 540
NCBI            TIPVSISLLIAACGLRHTDPCIFHNTIPDYLFYESPPLHFLNDFVSKQYAWVWLLWLLSQ 491
                ************************************************************
Beebase         TWITLHVWTPKCERLASTEKLFVVPMYNSLLIDQSMGLNRKRDDQPEVKVEDLAEIEKEK 600
NCBI            TWITLHVWTPKCERLASTEKLFVVPMYNSLLIDQSMGLNRKRDDQPEVKVEDLAEIEKEK 551
                ************************************************************
Beebase         GDGDYETIYEQTDGTTTPPSTVRSSDHVTRIYACATMWHENKEEMMEFLKSILRLDEDQC 660
NCBI            GDGDYETIYEQTDGTTTPPSTVRSSDHVTRIYACATMWHENKEEMMEFLKSILRLDEDQC 611
                ************************************************************
Beebase         ARRVAQKYLKVVDPDYYEFETHIFFDDAFELSDHDENESQVNRFVKLLVGTLDEAASDVH 720
NCBI            ARRVAQKYLKVVDPDYYEFETHIFFDDAFELSDHDENESQVNRFVKLLVGTLDEAASDVH 671
                ************************************************************
Beebase         QTRMHVRAPKKYPTPYGGRLVWTLPGKTKMIAHLKDKSKIRHRKRWSQVMYMYYLLGHRL 780
NCBI            QTRMHVRAPKKYPTPYGGRLVWTLPGKTKMIAHLKDKSKIRHRKRWSQVMYMYYLLGHRL 731
                ************************************************************
Beebase         MELPISVDRKEVIAENTYLLTLDGDIDFQPAAVKLLVDLMKKNKNLGAACGRIHPVGSGP 840
NCBI            MELPISVDRKEVIAENTYLLTLDGDIDFQPAAVKLLVDLMKKNKNLGAACGRIHPVGSGP 791
                ************************************************************
Beebase         MVWYQMFEYAIGHWLQKATEHMIGCVLCSPGCFSLFRGKALMDDNVMKKYTTRSDEARHY 900
NCBI            MVWYQMFEYAIGHWLQKATEHMIGCVLCSPGCFSLFRGKALMDDNVMKKYTTRSDEARHY 851
                ************************************************************
Beebase         VQYDQGEDRWLCTLLLQRGYRVEYSAASDAYTHAPEGFNEFYNQRRRWVPSTIANIMDLL 960
NCBI            VQYDQGEDRWLCTLLLQRGYRVEYSAASDAYTHAPEGFNEFYNQRRRWVPSTIANIMDLL 911
                ************************************************************
Beebase         MDAKRTIKINDNISLPYISYQILLMGGTILGPGTIFLMLVGAFVAAFKIDNWTSFYYNII 1020
NCBI            MDAKRTIKINDNISLPYISYQILLMGGTILGPGTIFLMLVGAFVAAFKIDNWTSFYYNII 971
                ************************************************************
Beebase         PILLFMLVCFTCKANIQLLCAQILSTGYAMIMMAVIVGTALQLGEDGIGSPSAIFLIALS 1080
NCBI            PILLFMLVCFTCKANIQLLCAQILSTGYAMIMMAVIVGTALQLGEDGIGSPSAIFLIALS 1031
                ************************************************************
Beebase         GSFFIAACLHPQEFWCIVPGIIYLLSIPSMYLLLILYSIINLNVVSWGTREVQVKKTKKE 1140
NCBI            GSFFIAACLHPQEFWCIVPGIIYLLSIPSMYLLLILYSIINLNVVSWGTREVQVKKTKKE 1091
                ************************************************************
Beebase         LEQEKKEAEEAKRKAKQKSLLGFLQNGVGSNDDEEGSIEISLAGLFKCMFCTHGQTSNEK 1200
NCBI            LEQEKKEAEEAKRKAKQKSLLGFLQNGVGSNDDEEGSIEISLAGLFKCMFCTHGQTSNEK 1151
                ************************************************************
```

```
Beebase         QQLVAIAQSMENVNKRLEIIERAVDPHGVTSRRRASSVGSRGDHLGAIGEDPAEGQDGHS 1260
NCBI            QQLVAIAQSMENVNKRLEIIERAVDPHGVTSRRRASSVGSRGDHLGAIGEDPAEGQDGHS 1211
                ************************************************************

Beebase         EPETVTSQNTEGNREGSNFLSRPYWLSDEGLKKGEIDVLSMQEEQFWKDLLEKYLYPIDE 1320
NCBI            EPETVTSQNTEGNREGSNFLSRPYWLSDEGLKKGEIDVLSMQEEQFWKDLLEKYLYPIDE 1271
                ************************************************************

Beebase         DKAEKA------------------------------------------------------ 1326
NCBI            DKAEKARIAKDLKDLRDQSVFAFFMMNALFVLIVFLLQLNKDLLHVKWPFGIKTNISFNA 1331
                ******

Beebase         ------RIAGDLIELRNKSVYAFFMFNTLFVLIVFLLQLNKDQLHVVWPLGVKENITMKE 1380
NCBI            DNFHEARIAGDLIELRNKSVYAFFMFNTLFVLIVFLLQLNKDQLHVVWPLGVKENITMKE 1391
                      **************************************************

Beebase         DGEVYVTKEYLQLEPIGLVFVFFFALILVIQFTAMLFHRFGTFAHILASTSLDWYCCKKT 1440
NCBI            DGEVYVTKEYLQLEPIGLVFVFFFALILVIQFTAMLFHRFGTFAHILASTSLDWYCCKKT 1451
                ************************************************************

Beebase         KDLSEEALLSKHAVEIVRDLQRLDGMEGDYEEDSGTGPGRRKTITNIEKSRKKTQAINTL 1500
NCBI            KDLSEEALLSKHAVEIVRDLQRLDGMEGDYEEDSGTGPGRRKTITNIEKSRKKTQAINTL 1511
                ************************************************************

Beebase         DVAFRQRFFSMSEEGNGLPRNMSTRRSAKAFKAFEGRRNSIMAMKRKSQMQTLGANNIYG 1560
NCBI            DVAFRQRFFSMSEEGNGLPRNMSTRRSAKAFKAFEGRRNSIMAMKRKSQMQTLGANNIYG 1571
                ************************************************************

Beebase         VAGNPLGIQGRPSRSSQISVKDVFEGHSGHTNPSYEPDENTGSSLRLHSLSQNAWREQNN 1620
NCBI            VAGNPLGIQGRPSRSSQISVKDVFEGHSGHTNPSYEPDENTGSSLRLHSLSQNAWREQNN 1631
                ************************************************************

Beebase         I 1621
NCBI            I 1632
                *
```

B) Alignment of the N-termini of the *D. melanogaster* and the two *A. mellifera* chitin synthase sequences

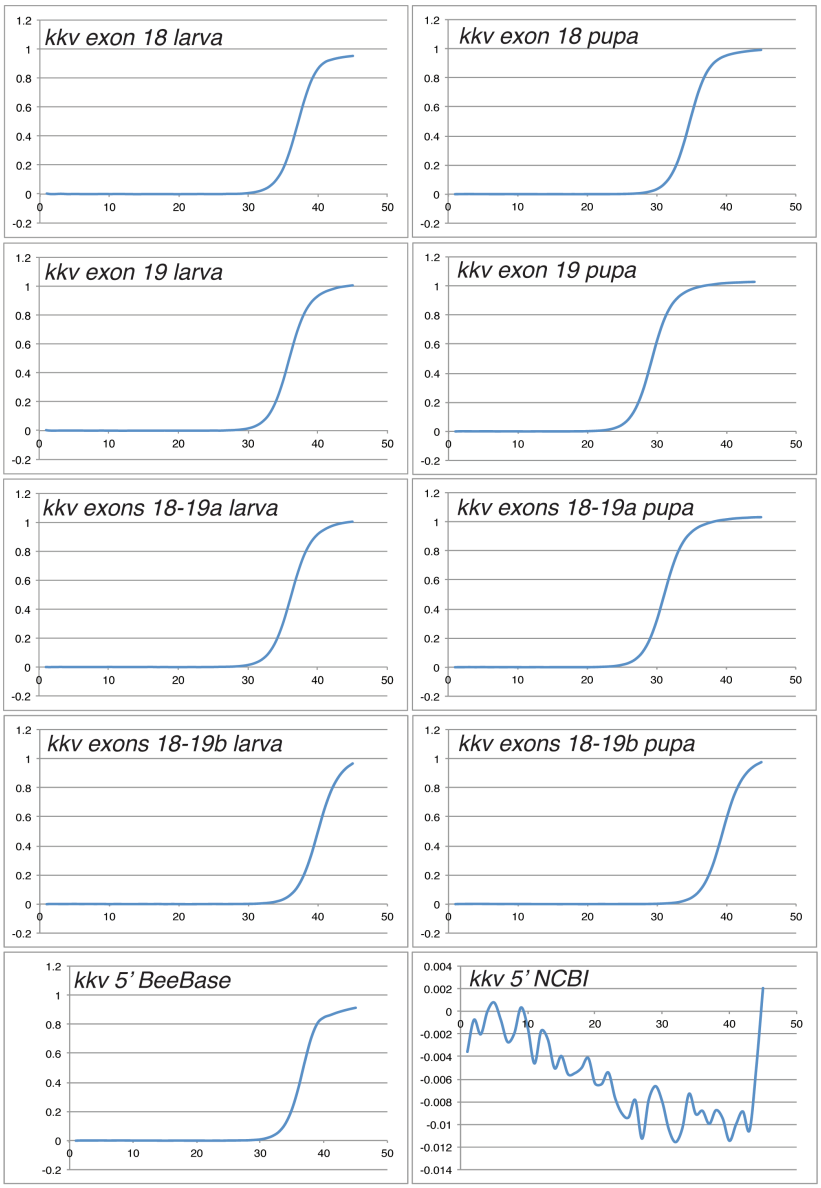
```
NCBI            ------------------------------MILYKTNDSR-----------------M 11
BeeBase         MIDVQLLTLKMKRSKNQKSFERVRGLFQKKIYHLLWLNPKSKEFSIKKMAVEGMYRKGVM 60
Dmel            ------------------------------MSAMRHRPMAP------------------ 11
                                                    :     :

NCBI            SKIQQQNGMMPGNGTMPDDDDFSDGESTPLTQD-YGDSQRTVVETKAWDVFRNPPPKIDS 70
BeeBase         SKIQQQNGMMPGNGTMPDDDDFSDGESTPLTQD-YGDSQRTVVETKAWDVFRNPPPKIDS 119
Dmel            -PGQGPGAGTAGEHVDSDDNNFTDDESSPLTHDIYGGSQRTIQETKGWDVFRDPPIKIET 70
                  *  ..  .*: . .**::*:*.**:***:* **.****: ***.*****:** **::
```



Supplementary figure 2A

As shown by examples of qPCR amplification curves, transcripts of the *Apis mellifera* chitin synthase are expressed in larvae and pupae. These data indicate that exons 14 and 15 that encode similar sequences (Fig. 1) are not present on the same transcripts.

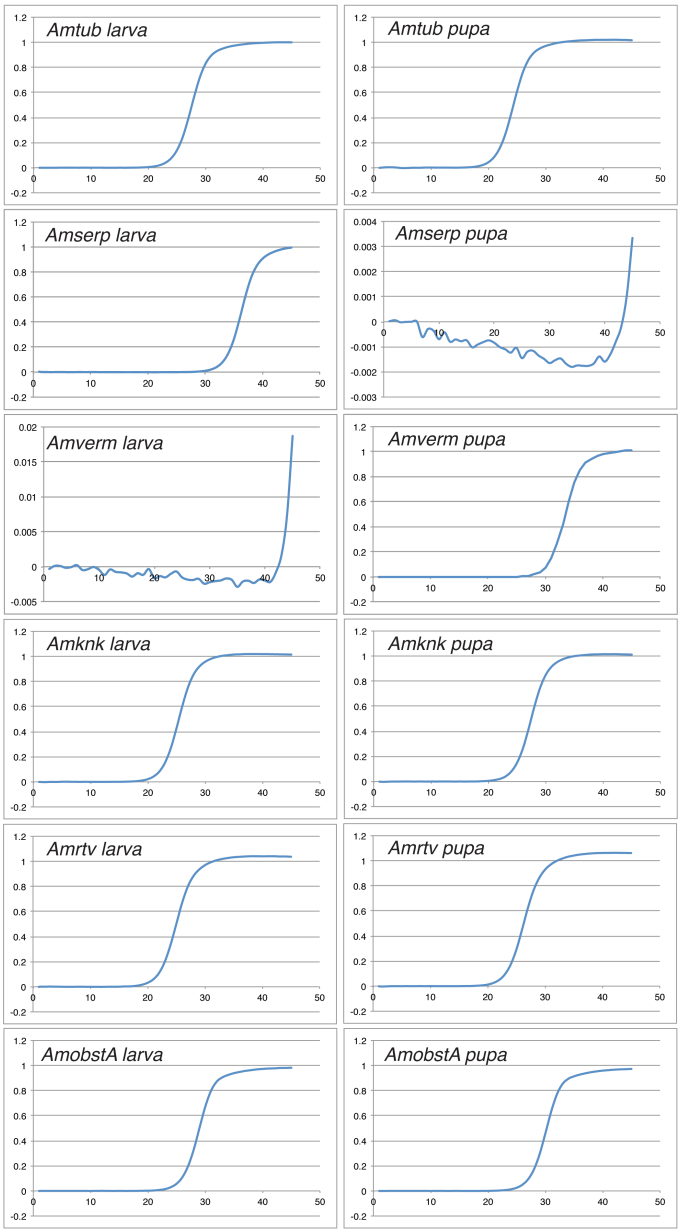
The relative expression levels normalised against the respective *tubulin* expression (not shown, see suppl. Fig. 2C) are shown in figure 2. Amplification cycles (x-axis) are plotted against arbitrary fluorescence values (y-axis) the amplification plateau being at 1 for the *tubulin* transcript in larvae (see suppl. Fig. 2C).



Supplementary figure 2B

Examples of qPCR amplification curves demonstrate that transcripts of the *Apis mellifera* chitin synthase are present in larvae and pupae. Transcripts amplified by two pairs of specific primers exist, which contain both exons 18 and 19 that code for highly similar sequences (Fig. 1). The respective exons in other insects are alternatively spliced (see text).

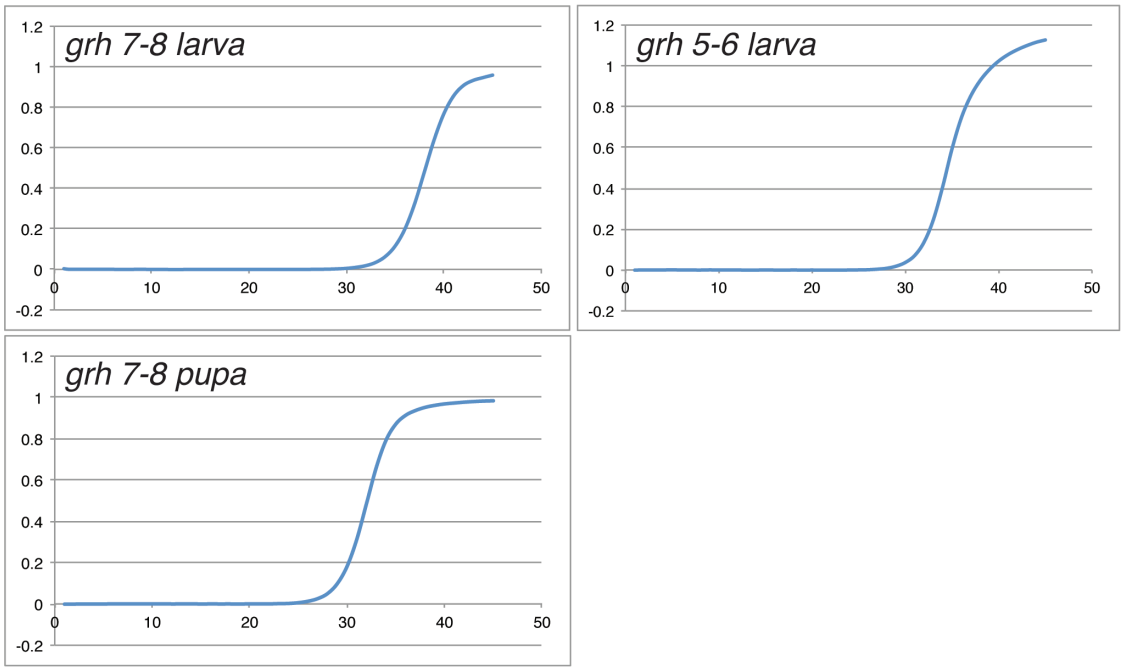
The relative expression levels normalised against the respective *tubulin* expression (not shown, see suppl. Fig. 2C) are shown in figure 2. Amplification cycles (x-axis) are plotted against arbitrary fluorescence values (y-axis) the amplification plateau being at 1 for the *tubulin* transcript in larvae (see suppl. Fig. 2C).



Supplementary figure 2C

Examples of qPCR amplification curves of transcripts of the *Apis mellifera* chitin organising factors *Amserp*, *Amverm*, *Amknk*, *Amrtv* and *AmobstA* show their expression in larvae and pupae. As a control, expression of *tubulin* (*GB44134*) was recorded.

The relative expression levels normalised against the respective *tubulin* expression are shown in figure 2. Amplification cycles (x-axis) are plotted against arbitrary fluorescence values (y-axis) the amplification plateau being at 1 for the *tubulin* transcript in larvae.



Supplementary figure 2D

Examples of qPCR amplification curves of *Apis mellifera grh* transcripts show their presence in larvae and pupae.

The relative expression levels normalised against the respective *tubulin* expression (not shown, see suppl. Fig. 2C) are shown in figure 2. Amplification cycles (x-axis) are plotted against arbitrary fluorescence values (y-axis) the amplification plateau being at 1 for the *tubulin* transcript in larvae (see suppl. Fig. 2C).

Supplementary figure 3

The *A. mellifera GB46725* locus codes for a CP2-type transcription factor with highest homology to the *D. melanogaster* Grh sequence. Homology starts after residue 605 of the *D. melanogaster* N isoform 244and residue 311 of the *A. mellifera* sequences. The N-terminal halves are, by contrast, highly divergent. Whereas the *D. melanogaster* Grh N-terminal sequence contains 12% glutamines, the respective *A. mellifera* sequence has only 9,3% glutamines. The predominant amino acid in the *A. mellifera* sequence is serine (16%).

```
Amel            ----------------------------------------MGQKLL-------VQEAP 11
Dmel-N          MSTSTATTSVITSNELSLSGHAHGHGHAHQLHQHTHSRLGVGVGVGILSDASLSPIQQGS 60
                                                       :*  :*       :*:..
Amel            GSGSGSANR-------------------------------------------------- 20
Dmel-N          GGHSGGGNTNSSPLAPNGVPLLTTMHRSPDSPQPELATMTNVNVLDLHTDNSKLYDKEAV 120
                *. **..*
Amel            -----------MEGGEGAG---------------GGELHHFLHQYNQQSHSP-------- 46
Dmel-N          FIYETPKVVMPADGGGGNNSDEGHAIDARIAAQMGNQAQQQQQQQQQTEHQPLAKIEFDE 180
                           :** * .               *.: ::  :* :* .*.*
Amel            ------------------------------------------------------------
Dmel-N          NQIIRVVGPNGEQQQIISREIINGEHHILSRNEAGEHILTRIVSDPSKLMPNDNAVATAM 240

Amel            -----------GQSLQGGLPLPLSPQSLRHDGSSSTGIAGVKR------EPEDLSSSRGA 89
Dmel-N          YNQAQKMNNDHGQAVYQTSPLPLDASVLHYSGGNDSNVIKTEADIYEDHKKHAAAAAAAA 300
                           **::    ****... *::.*...:.:  .:       : .  ::: .*
Amel            QQSSKRHKQAQPD-------------SPTPPGMYHHHQHQLQVQQYGSPYDPYSSCSPR- 135
Dmel-N          GGGSIIYTTSDPNGVNVKQLPHLTVPQKLDPDLYQADKHIDLIYNDGSKTVIYSTTDQKS 360
                   *  :. ::*:             .   *.:*: .:*   : : **    **: . :
Amel            --------------------LQSTTYTSTSATG------AANTGNPSGLHQEATAVYVTG 169
Dmel-N          LEIYSGGDIGSLVSDGQVVVQAGLPYATTTGAGGQPVYIVADGALPAGVEEHLQSGKLNG 420
                                  . .*::*:.:*      .*: . *:*:.:.  :  :.*
Amel            DALPPLASSSSS---------SLSTTTASYTRYEVVPSSYATTHAIRSSSSSSKVLTVDL 220
Dmel-N          QTTPIDVSGLSQNEIQGFLLGSHPSSSATVSTTGVVSTTTISHHQQQQQQQQQQQQQQQQ 480
                :: *  .*. *.         * .:::*: :   **.::  : *  :......:    :
Amel            PSPD-----------SGIG-----ADAVTPRQDHHPPTALHQSSFDYTELCPGGTAAGAA 264
Dmel-N          QHQQQQQHPGDIVSAAGVGSTGSIVSSAAQQQQQQQLISIKREPEDLRKDPKNGNIAGAA 540
                   :          :*:*     ..:.: :*:::   :::::. *  :   .*. ****
Amel            VVLESGAVIHHQPLQL-------QLQQSHAQAQVQRGALVSSAATNSNPNP--------- 308
Dmel-N          TANGPGSVITQKSFDYTELCQPGTLIDANGSIPVSVNSIQQRTAVHGSQNSPTTSLVDTS 600
                ..  .*:** ::.::         * ::::.  *. .:: . :*.:.. *.
Amel            --NPPRSRPWHDFGRQNDADKIQIPKIFSAYGFKYHLESPISTSQRREDDRITYINKGQF 366
Dmel-N          TNGSTRSRPWHDFGRQNDADKIQIPKIFTNVGFRYHLESPISSSQRREDDRITYINKGQF 660
                  ...**********************:  **:********:*****************
Amel            YGITLDYVPDPDKPSLKAGQTVKSVVMLMFREEKSPEDEIKAWQFWHGRQHSVKQRILDA 426
Dmel-N          YGITLEYVHDAEKP-IKN-TTVKSVIMLMFREEKSPEDEIKAWQFWHSRQHSVKQRILDA 718
                *****:** *.:** :*   *****:*********************.************
Amel            DTKNSVGLVGCIEEVAHNAIAVYWNPLESSAKINVAVQCLSTDFSSQKGVKGLPLHIQVD 486
Dmel-N          DTKNSVGLVGCIEEVSHNAIAVYWNPLESSAKINIAVQCLSTDFSSQKGVKGLPLHVQID 778
                **************:*****************:**********************:*:*

Amel            TYEDPPPHTHTYTPPSHRGYCQIKVFCDKGAERKTRDEERRAAKRKMTATGRKKLDELYH 546
Dmel-N          TFEDP-----RDTAVFHRGYCQIKVFCDKGAERKTRDEERRAAKRKMTATGRKKLDELYH 833
                *:***       *.  ****************************************
```

```
Amel            SVTERSEFYSMVDLHKPPVLFSPPAEHT------IDKFSTMELSGFYGGGGGGGGDTDTS 600
Dmel-N          PVTDRSEFYGMQDFAKPPVLFSPAEDMEKVGQLGIGAATGMTFNPLSNGNSNSNSHSSLQ 893
                .**:*****.* *: ********. :        *.  : * :. : .*.......:. .
Amel            SLSNGEGGGLKAGGSSPYPACG-RPSLPALKFHNHFPPDNLTSLHDKKDSLLMQVQELQG 659
Dmel-N          SFYGHETDSPDLKGASPFLLHGQKVATPTLKFHNHFPPDMQT---DKKDHILDQN----- 945
                *: . * .. .  *:**:   * : : *:**********  *   **** :* *
Amel            SVLQGVQQGLAGTVVSGVGNRLHLQQQPPHLRPTDAEERVMLYVRQESEDVYTPLHVTPP 719
Dmel-N          --------MLTSTPLTDFGPPMKRGRMTP-----PTSERVMLYVRQENEEVYTPLHVVPP 992
                         *:.* ::..*  ::  : .*      :.**********.*:*******.**
Amel            TVQGLLNAIESKYKIASSSINNLYRKNTKGITAKIDDDMIRYYVDEDLFLLEVTHSRVAN 779
Dmel-N              TTIGLLNAIENKYKISTTSINNIYRTNKKGITAKIDDDMISFYCNEDIFLLEVQ------
1046
                *. *******.****:::****:**.*.*********** :* :**:*****
Amel            PDPSNERNPANPGSPGDPGDPSDQPATAGYDVTLIELPSASAHIAHSPLNSAHGHAHVHE 839
Dmel-N              ----------------------QIEDDLYDVTLTELPNQ--------------------
1063
                                       *     ***** ***.
Amel            SGNT 843
Dmel-N          ----
```