

The MRJP/YELLOW protein family of *Apis mellifera*: Identification of new members in the EST library

Štefan Albert*, Jaroslav Klaudiny¹

Institute of Medical Radiation and Cell Research, Versbacherstr. 5, D-97078 Würzburg, Germany

Received 25 July 2003; received in revised form 26 September 2003; accepted 26 September 2003

Abstract

Major royal jelly proteins (named MRJP1-5) of honeybee (*Apis mellifera*), yellow proteins of *Drosophila*, together with putative proteins found in several bacteria, form a protein family termed the MRJP/yellow family. Members of the family exert diverse physiological functions and amongst eukaryotes appear to be restricted to the order Insecta. MRJPs constitute about 90% of total protein of royal jelly, which is secreted by nurse bees to feed the queen and growing larvae.

We looked for *mrjp* and *yellow* homologues in a honeybee brain expressed sequence tags (EST) library. In addition to the five *mrjp* cDNAs previously characterized, we found three additional cDNAs encoding novel MRJPs and importantly, two cDNAs coding for orthologues of *Drosophila* yellow proteins. One *yellow* cDNA and all three cDNAs coding for the novel MRJPs were assembled completely, the sequence of the other *yellow* homologue was partially assembled.

The data we present here supports the view that repeated duplications and functional divergence occurred during the evolution of MRJPs in honeybees, with even closely related MRJPs appearing to perform diverse physiological functions. Conversely, yellow protein orthologues appear to be conserved and thus candidates for maintaining the former function(s) of yellow proteins.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Royal jelly proteins; MRJP; Yellow; EST; Sequencing; Phylogeny; Nutrition; Brain

1. Introduction

The honeybee, *Apis mellifera*, a model organism for studying insect social behavior, is well-studied physiologically, but knowledge of its molecular characterization lags. However, two recent articles presenting an EST set expressed in honeybee brain (Kucharski and Maleszka, 2002; Whitfield et al., 2002) and the announcement of the honeybee genome sequencing project (Kaufman et al., 2002) have presented an unprecedented opportunity to improve genetic characterization of *A. mellifera*. The group of Gene Robinson (Whitfield et al., 2002) undertook the ambitious challenge to sequence the collection of random cDNAs from a normalized cDNA library of honeybee brain, whereas

the group of Ryszard Maleszka (Kucharski and Maleszka, 2002) employed a non-normalized cDNA library and microarrays for selection and sequencing of cDNAs differentially expressed during development. The data obtained were promptly made accessible to the scientific community. Robinson's team annotated the ESTs on their homology to the well-characterized *Drosophila* gene set. This approach led to the identification of many honeybee orthologues (genes performing the same function) to *Drosophila* counterparts, such as housekeeping genes and other conserved genes. However, many genes fell below the similarity threshold and remained unclassified. The latter set of genes is of particular importance, as it encompasses the genes performing unique functions and thus characterizing differences between the compared species.

One of the hallmarks of the advanced social behaviour of honeybees is that worker bees forfeit their own reproduction and take care of the brood laid by the queen instead. The task of nursing is performed by a specialized caste, called nurse bees, which produce lar-

* Corresponding author. Tel.: +49-931-201-45850; fax: +49-931-201-45887.

E-mail address: Stefan.Albert@mail.uni-wuerzburg.de (Š. Albert).

¹ Present address: Institute of Chemistry, Slovak Academy of Sciences, Dubravská cesta, SK-845 38 Bratislava, Slovakia.

val food called royal jelly or bee milk, and deliver it to the developing larvae. The queen is fed royal jelly throughout her life, which is believed to contribute to her longevity and enormous physiological performance.

We report here on members of the MRJP/yellow family (Protein Family Database code PFAM 03022), which also includes yellow protein, and its homologues. MRJPs were initially identified as major protein components of royal jelly, constituting 80–90% of the total protein content (Klaudiny et al., 1994a; Ohashi et al., 1997; Schmitzova et al., 1998). We identify the ESTs coding for all known MRJP proteins plus three new *mrjp* genes in the honeybee EST database. The cDNA sequences of new MRJP members are assembled based on information available from the EST database, our own experimental sequencing data and the recent release of shotgun sequencing of the *A. mellifera* genome. In addition, we identify two ESTs which are homologous with *yellow* genes of *D. melanogaster*, assemble their sequences and propose them to be honeybee *yellow* orthologues. Evolution of MRJP proteins is reconstituted from the obtained data.

2. Material and methods

2.1. Database searches and sequence assembly

BLAST searches (Altschul et al., 1990) of the honeybee EST database were done with cDNA sequences coding for MRJP proteins (GenBank accession numbers: *mrjp1*, AF000633; *mrjp2*, AF000632; *mrjp3*, Z26318, *mrjp4*, Z26319; *mrjp5*, AF004842, *yellow-f*, BT003613; *yellow*, NM_057444) on the honeybee EST Internet server http://titan.biotech.uiuc.edu/bee/honeybee_project.htm using standard parameters.

Sequence reads of honeybee genomic DNA were searched for similarity as described for EST. DNA sequences were assembled with the help of GCG (Devereux et al., 1984) and EMBOSS (Rice et al., 2000) program packages.

Newly identified cDNAs were deposited in GenBank under following accession numbers:

mrjp6: AY313893, *mrjp7*: BK001420, *mrjp8*: AY398690.

2.2. PCR analyses and DNA sequencing

The following primers were synthesized to amplify *mrjp6* and *mrjp8* cDNAs from the nurse honeybee head cDNA library (Klaudiny et al., 1994a).

MRJP6 forward: GTG TTC TCC AAA ATT ACT TGC, MRJP6 reverse: GCA AGT AAT TTT GGA GAA CAC, MRJP8 forward: TTG CGA AGT GAA TGG ATC, MRJP8 reverse: TTA TTT TTG GCA ACC ACT TCG, T3 primer: AAT TAA CCC TCA

CTA AAG GG, T7 primer: TAA TAC GAC TCA CTA TAG GG.

Polymerization reactions were done in 50 µl volume as follows: 5 µl of cDNA library stock (Lambda ZAP, Stratagene) after one or more freeze-thaw cycles were used as a template in a modified hot start procedure. After initial denaturation for 5 min at 94 °C, subsequent cooling to 85 °C and addition of 1.5 U of Taq polymerase (AmpliTaq, Perkin Elmer), PCR was performed in a commercial buffer containing 1.5 mM MgCl₂. We used 35 PCR cycles (50 s at 94 °C, 50 s at 50 °C and 100 s at 72 °C), followed by 10 min incubation at 72 °C.

PCR reactions were size-separated by gel electrophoresis (1.5% agarose), purified and ligated into PCR2.1 TA cloning vector (Invitrogen) or sequenced directly.

DNA sequences were determined by the cycle sequencing method (Prism Ready Reaction BigDye Terminator kit, Perkin Elmer) on an ABI 373A automatic sequencer. Obtained DNA sequences were assembled and analyzed using program packages mentioned above.

3. Results

3.1. Identification of known *mrjp* cDNAs in the EST library of *A. mellifera* and their genetic variations

Initially, we performed a BLAST search with the cDNA sequences of five known *mrjps* against a honeybee brain EST library. Sequences with >98% identity over more than 200 bases were considered to code for a given *mrjp* cDNA. The similarity search data are summarized in Table 1.

Table 1
Sequences coding for MRJP and YELLOW proteins found in the honeybee brain EST database

Contig/read number	Singlets in the contig	Encoded cDNA
654	14	<i>mrjp1</i>
185	44	<i>mrjp2</i>
908	5	<i>mrjp3</i>
1368	12	<i>mrjp3</i>
1203	4	<i>mrjp4</i>
1463	2	<i>mrjp4</i>
2106	11	<i>mrjp5</i>
1510	2	<i>mrjp6</i>
380	11	<i>mrjp7</i>
BB170012B2E06	1	<i>mrjp8</i>
662	4	<i>am-yellow-f</i>
BB160008B2E04	1	<i>am-yellow</i>

3.1.1. *mrjp1*

Contig 654, consisting of 14 single sequence reads (singlets), shows 100% identity with *mrjp1* cDNA sequences deposited in GenBank.

The sequence of *mrjp1* cDNA has previously been determined by three independent research teams located in different continents. All *mrjp1* sequences deposited in the databases were shown to be 100% identical with each other (Kucharski et al., 1998; Ohashi et al., 1997; Schmitzova et al., 1998). The sequence of contig 654, derived from a honeybee cDNA library produced in a fourth continent, the USA, adds further proof that the *mrjp1* sequence is highly conserved and without a single nucleotide change (Whitfield et al., 2002). The recently published sequence of the *mrjp1* genomic locus contains seven substitutions, which do not change the amino acid sequence of the encoded protein. It also shows that the genomic locus of *mrjp1* gene contains 6 introns (Malecova et al., 2003).

3.1.2. *mrjp2*

mrjp2 cDNA is identical with contig 185 consisting of 44 singlets, the highest number found in the entire brain EST library (Whitfield et al., 2002), which makes it the most abundant *mrjp* transcript identified in the honeybee brain EST database. Similar to *mrjp1*, no nucleotide substitutions were found between contig 185 and *mrjp2* cDNA (Schmitzova et al., 1998).

3.1.3. *mrjp3*

Two contigs were found to be overlapping with the *mrjp3* cDNA. Contig 908 (5 singlets) and contig 1368 (12 singlets) were not assembled due to differences in the repetitive region of this gene, which is highly polymorphic in *A. mellifera* (Albert et al., 1999b) and also in *A. dorsata* (Albert and Schmitz, 2002). Interestingly, in addition to the polymorphic repetitive region, many nucleotide substitutions (termed single nucleotide polymorphism, SNP), creating in several cases amino acid changes in the encoded protein, were identified in the deposited sequences. Observed SNPs are listed in Table 2. Some of these SNPs, observed between two cloned *mrjp3* cDNAs have been previously described (Albert et al., 1999b).

3.1.4. *mrjp4*

Two contigs align with the *mrjp4* sequence (Klaudiny et al., 1994a), with a gap of 98 bp between them. Contig 1203 (4 singlets) overlaps with bases 69–834 and contig 1463 (2 singlets) with bases 931–1593 of the *mrjp4*. Similar to *mrjp3*, several SNPs have been identified in *mrjp4* (see Table 2).

3.1.5. *mrjp5*

Contig 2106 (11 singlets) aligns with bases 720–1480 of the *mrjp5* cDNA (Albert et al., 1999a). As with

Table 2

Nucleotide substitutions found by comparison of EST cDNA sequences with cDNAs of *mrjps* deposited previously in GenBank

Position	Substitution	Inferred amino acid substitution
<i>mrjp3</i>		
398	G→A	Asp→Lys
448	T→C	Phe→Phe
589	C→T	Thr→Thr
610	T→A	Val→Val
804	A→G	Tyr→Cys*
853–5	TCT→CTC	Leu→Ser*
895	C→G	Ser→Arg*
<i>mrjp4</i>		
337	C→A	Asp→Glu
427	T→C	His→His
469	T→C	Asp→Asp
478	C→T	Leu→Leu
515	T→C	Phe→Leu
529	T→C	Leu→Leu
628	G→T	Ser→Ser
794	G→A	Val→Ile
1583	A-insertion	non-coding region
<i>mrjp5</i>		
722	A→T	(Glu→Asp)
784	A→C	(Asp→Ala)
1327	C→G	(His→Arg)

*Substitutions at these positions, plus changes at 5 additional positions were described previously (Albert et al., 1999b).

mrjp3, the sequence alignment breaks down in the repetitive region of the *mrjp5* sequence. Three SNPs were found in the overlapping region, each changing the sequence of the MRJP protein (Table 2).

3.2. New *mrjp* cDNAs in the honeybee brain EST database

Besides the previously characterized *mrjp* cDNAs described above, the BLAST search revealed three further sequences which are highly homologous, but not identical with the known *mrjp* cDNAs. They apparently code for new members of the MRJP protein family in *A. mellifera*. One (contig 380) harbours a complete open reading frame, which codes for an MRJP-like protein. The other two were incomplete cDNA fragments. We set out to complete the sequences of these cDNAs, which we present here.

3.2.1. *mrjp6*

Contig 1510 consisted of two singlets. The closest relative appeared to be the *mrjp5* cDNA, which was nearly identical with the 3' terminus of contig 1510. *mrjp5*-specific primers that should prime both *mrjp5* and contig 1510, did not work with contig 1510. Only the 3' terminal fragment of *mrjp5* was amplified. This prompted us to inspect contig 1510 in detail. We carried out BLAST searches with each of the two singlets separately. We found that the singlet BB160020A20H03

really encodes a novel MRJP protein which we have called MRJP6, whereas singlet BB170018B10A09 represented a part of *mrjp5* cDNA (>99% identity over 260 bp). Primers were designed from the sequence of the BB160020A20H03 singlet. These in combination with T3 and T7 primers (priming in the multiple cloning site of Lambda ZAP phage near the cloned cDNA ends [Klaudiny et al., 1994a]) gave us PCR products. Sequencing of these PCR products enabled us to complete the sequence of *mrjp6* cDNA both upstream and downstream of singlet BB160020A20H03. Sequence data confirmed that the two singlets used for construction of contig 1510 in the EST database do not align. The assembled *mrjp6* cDNA is 1529 bp long and highly homologous to *mrjp5*. It encodes a protein of 437 aa (see Fig. 1). In contrast to the sequence of contig 1510, there is no repeat coding for a recurring aspartate-arginine-methionine tripeptide motif (DRM in single-letter code) in the sequence of *mrjp6* [Albert et al., 1999a]).

The sequence of the *mrjp6* cDNA was nearly completely covered by sequence readings from shotgun sequencing of the *A. mellifera* genome. This enabled us to identify the introns in the genomic locus of *mrjp6*. All introns in the protein coding region are located at the same positions as described for *mrjp1* (Malecova et al., 2003). Surprisingly, intron 0, which is 86 bp long and located just four bases upstream of the ATG initiation codon of *mrjp1* was not found in *mrjp6* although region homologous to its 3' part could be identified. The 22 nucleotides upstream of ATG are identical between the *mrjp6* cDNA and the *mrjp1* genomic locus, including the 3' part of intron with the conserved AG motif determining the 3' splice site of introns (see Fig. 1, bold and underlined characters). However, the 5' splice GT motif of the intron could not be identified. cDNA sequence of this region was perfectly co-linear with the corresponding genomic locus.

Taken together, the 5' non-coding region of *mrjp6* contains a "genetic fossil" of intron 0, which, probably due to deletion, lost its ability to be spliced out during mRNA processing and has become a stable part of *mrjp6* mRNA. In line with this, the 5' non-coding region of *mrjp6* cDNA (100 bp) is significantly longer than those of other *mrjps* (34–55 bp).

3.2.2. *mrjp7*

Contig 380 (11 singlets) is 1430 bp long. It contains a complete open reading frame, which encodes a protein of 442 amino acids named MRJP7, GenBank accession number BK001420, with high homology to MRJP2 protein (73.0% identity in 433 aa overlap with MRJP2, 66.7% with MRJP6). Two SNPs were found in assembled singlets (A190G, G210C). Neither of them affected the encoded protein sequence.

ATTAAATATTTTCAGCTTCCCTCCATAAGTGTTCATATATCTTAATTGTATTATTG	60
CAATCTTTTCATTATCTAACACGAAATATTTGTAGAAAATGACAAATGGTTACTGCT	120
M T N W L L L	7
GATAGTGTCTTAGCATAGCTTGTCAAGATGTCACAGTGCATTCATCAAGAAATC	180
I V C L S I A C Q D V T S A I H Q R K S	27
TTCAAAAAATTGGAACATTCGATGAACGTGATTCAGAAATGAAATATATCGATTATGA	240
S K N L E H S M N V I H E W K Y I D Y D	47
TTTGTGTAGTATGAAAAAGACAAGCTGCGATTCATTCGGCAATACGATTATACGAA	300
F G S D E K R Q A A I Q S G E Y D Y T K	67
AAATTATCCTTCGACGTGATCAATGGCATAATAAGACTTTCTCGCTGTAATAAGATA	360
N Y P F D V D Q W H N K T F L A V I R Y	87
CGATGGTGTACCTTCCTCTTTGAACGTGATATCGAGAAATGGCAACGGTGGATCGCT	420
D G V P S S L N V I S E K I I G N D S D G M L	107
TCTACAAACCGTATCTGATGGTGGTGGGCAACTATAAGATTGTTCTGGAATAGTGAG	480
L Q P Y P D W S W A N Y K D C S G I V S	127
CGCTTACAAATTCGATCGACAAATTCGACAGATTGTGGTCTTGACTCGGGTCTTAT	540
A Y K I A I D K F D R L W V N S D G M L	147
TAATAATATTCACCTTATGTGTCTCCAAATTAATTCGCTTTGATCTCAATACCTCAAA	600
N N I Q L M C S P K L L A F D L N T S K	167
GTTGCTTAAACAAATCGAGATACACATAATATGCCGTAAATGCCAGCACAGGAATGGG	660
L L K Q I E I P H N I A V N S D G M L	187
AGGACCCGTATCGCTAGTTGTTCAAGCTATGGATCTATGAATACTACGGTGTATATAGC	720
G P V S L V V Q A M D P M N T T V Y I A	207
AGACGACAGAGGTGACGCTTTAATCATCTATCAAAATCTGATGATTCTTTCCATCGATT	780
D D R G D A L I I Y Q N S D S F H R I	227
GACTTCCAAATCTTTGATAACGATCTCAGATATCTGAACGTGGCGTGGCGGAGAAAG	840
T S K T F D N D L R Y S E L A V A G E S	247
TTTCACAGTGCATGATGGAATTTTGGAAATGGCACTTAGTCTGTGACGAACAATCTTFA	900
F T V H D G I F G M A L S P V T N N L Y	267
TTACAGCCCTCTCACTTCTCACAGTTTGTATTATGTTAACATGGAACCATTCATGAATC	960
Y S P L T S H S L Y Y V N M E P F M K S	287
ACAATATGAAGAAATAATATAGAATATGAAGAAATCCAAAGATATTTTCAACACTCAATC	1020
Q Y E E N N I E Y E G I A Q N I F N T Q S	307
GTCGTCTAAAGTAATGTCGAAAAATGGCGTCTTTCTTCGGACTTGTGAATAATTCAGC	1080
S A K V M S K N G V L F F G L V N N S A	327
TATTGGTTGTGGAAACGAGCATCAACCACTTCAGAGACAAAATATGGATATGGTCGCTCA	1140
I G C W N E H Q P L Q R Q N M D M V A Q	347
GAATGAAAAGACACTTCAAATGATCATTAGCGTGAAATATACAAAATCTTGCAATATTC	1200
N E K T L Q M I I S V K I I Q N L A Y S	367
CGAAGGATGAATAGAATTCACAAGAAATGAATATATGTTGGCTTTAAGTAACAGAATGCA	1260
G R M N R I H K N E Y M L A L S N R M Q	387
GAAATAGTAAACAATGATTTTAATTCGACGAAGTAAACTTCGAATTTTGGGTGCGAA	1320
K I V N N D F N F D E V N F R I L G A N	407
TGTAAATAACTTAATAAAAAACACTCGTTGTGCAAGTCTAAACATCAGAATAACAATCA	1380
V N N L I K N T R C A K S N Q N N A Q	427
AAATAAATATAAGAACTCAAGCTCATTAGATTAGATCTTTTCAAGATCATATTAAATC	1440
N K Y K N Q A H L D *	438
TATAGATCAACTTTTTTCGTGGTAAATCAAATATTTTAAAAATTTATTTCATTATAA	1500
ATGAATAAAATAAATATCGTTTTCGCATG	1529

Fig. 1. Nucleotide sequence of *mrjp6* cDNA and inferred sequence of the encoded protein. Polyadenylation signal is shown in bold characters. Borders of singlet BB160020A20H03 are indicated by vertical arrows. Bold letters upstream of ATG show the region of homology between *mrjp1* genomic locus and *mrjp6* cDNA; bold and underlined nucleotides indicate the 3'-terminus of intron 0, which is not spliced out of *mrjp6* mRNA.

3.2.3. *mrjp8*

Another MRJP-like singlet was deposited in the EST library. BB170012B20E06 is apparently a new member of the MRJP family named MRJP8. Using genomic sequence database and total cDNA preparation of nurse heads, we amplified, cloned and sequenced the *mrjp8* cDNA (GenBank accession number AY398690). The 5' non-coding part of the cDNA could not be entirely determined due to the presence of intron (see above).

3.3. Yellow proteins of *A. mellifera*

3.3.1. *am-yellow-f*

When the first *mrjp* cDNAs were characterized, the only distant homolog to be found in databases was the *yellow* gene of *D. melanogaster* (Albert et al., 1996). Based on different expression patterns of MRJP and yellow proteins as well as the phenotype of *yellow* mutants, it was proposed that the physiological functions of *yellow* and *mrjp* genes had diverged early in honeybee development. We hypothesized that functional homologues of *yellow* could exist in *A. mellifera*. Low stringency Southern analyses did not previously succeed in identifying these homologues (Albert et al., 1999a).

Therefore we performed a similarity search of the translated *A. mellifera* EST database of *A. mellifera* for homologues of *Drosophila*'s yellow proteins. It led to identification of contig 662 (4 singlets), which shares a significantly higher similarity with *yellow* than with any MRJP (60% vs 50%, using FASTA). The contig appeared to be truncated at its 3' terminus. We succeeded in extending it using *A. mellifera* genomic sequences (Fig. 2). Due to its homology with the *yellow-f* gene product of *Drosophila*, the cDNA has been designated *am-yellow-f*. The assembled cDNA of *am-yellow-f* is about 1447 nucleotides long with the 3' end placed 12 bases after the putative polyadenylation signal. The encoded protein is 411 aa long. It is possible that the 3' end of cDNA differs from that derived from genomic sequence due to interruption by introns. However, we believe for the following reasons that the 3' end of the cDNA is as depicted in Fig. 2. First, the double-stop codon (TAA TGA) occurs in the predicted exonic sequence. Second, there is no GT motif (indicating the 5' splice site of the intron) in the 45 bases upstream of the stop codon TAA. Third, the AATAAA polyadenylation signal (Fig. 2, underlined) is located 38 bp downstream of STOP codon. Fourth, we could not find any region homologous to the *yellow* gene in the 1100 bp of genomic sequence downstream. Fifth, the closest *Drosophila* homolog, *yellow-f* is of a very similar length (429 aa). However, experimental proof of the cDNA length and the length of predicted protein is awaited.

3.3.2. *am-yellow*

The translated singlet BB160008B20E04 showed the highest homology to the original *Drosophila* YELLOW protein. Using genomic sequence reads we could extend it in its 5' direction so that the N-terminal part of the encoded protein could be inferred. Unfortunately, no extension of this cDNA downstream of 130th aa of the encoded protein could be carried out due to the lack of flanking sequences in the databases. Therefore this *am-YELLOW* protein was not included in protein sequence alignment (Fig. 3) and phylo-

TGTTACCACTTCTGATGTGGACCGTTTACGAGAAGGTTCTTGGGTACAAATAACGGAT	60
AATTCACCTTCTGAGCGTTCAACATTACATTTTAAAGAGTTACTATAACCTGTGGCCCGT	120
CTTAAGATACAAAAATAAAGATATGTGGCACTTTTATGGATCGTATTTCTTGCAT	180
M W H F L W I V F L A L	12
TGGCAATGGCGAAGAAATTAACAATTTATTCATGGAACGTGATCGAATATAATTTTC	240
A N G E E I K T I Y S W N V I E Y N F P	32
CAATGATAATATTCGAAATACACTGATATCTAATGGTGATTATCGAAGAGAATAATA	300
N D N I R N T L I S N G D Y I E E N N M	52
TGCCAAATGGTATGCAAAATTGGAATGATAAAGTATTTATCACTATACCAGATGGAAAA	360
P N G M Q I W N D K V F I T I I P R W K N	72
ATGGTGTCCCTCTAATTTAAATTTCTTTTAAAAATGACGAATCAGAATCTCCAAAT	420
G V P S N L N F F L K N D E S E S P K L	92
TAAACCTTATCCCAATGGGAAATGAACGATATTAATAAATGATAGTATAAATA	480
N P Y P N W E M N D I N K N T I S I N I	112
TAATCCGTGTAGAGTTGATGCGCTGTGATCGATTATGGGGTGTGATACGGGTGTGATG	540
I R V R V D A C D R L W G V D T G V D D	132
ATATTTGGCAATAACACTGTATACATCAACCGAGAATTATCATAATCGATTTAAAAA	600
I L G N T V I H Q I I I D L Y K I	152
CAGATAAGATATTGAGAATTTATCCTCTAAAAAGTTTCAGATCAACATCCGATTCGTTTT	660
D K I L R I Y P L K S S D Q T S D S F F	172
TCGTGTATTAGTAATCGATGTAGATCCAAATAATTCGCGCAATACGTATGCTTACATCT	720
V D L V I D V D P N N C D N T I Y D L I S	192
CTGATCTCAGTGGATATGCTTTAGTAGTTTACAGTTGGGCCAAGAATGATTTCTGGCGTA	780
D L S G Y A L V V Y S W A K N D S W R I	212
TCACGCAATATTTTATTTTTCGATCTCGCTACGGTAATTAACAATCAATGGATTCA	840
T H N F F Y F D P R Y G N Y N I N G F N	232
ACTTTCAATGGAAGATGGTTTATTTGGATTATCTTTGTCGCGATTCGAAACCGATGGCT	900
F Q W K D G L F G L S L S A L Q T D G Y	252
ATAAAATTTTACTTCCATGCGATGCTAGTATCGCTGAATCTCAGTATCTACGGAAG	960
K I L Y F H A M S S I A E F S V S T E V	272
TTTACAGGATCACAGTTAGAAAAATCTAACGATTATATGCTTTTCACTTCGAGGGTG	1020
L Q D H T L E K S N D Y Y A F H F E G E	292
AAAAAGTCCAAATAGTCAGGGCCCGAGTTCTGTGATCGACGAATACCGGTGTGAT	1080
K G P N S Q G P S S V I D T N T G V D Y	312
ACTTCACCTCAATTAACAGAAATGGCATCGCTTGTGTTGGATACCAATCTGAATGAATC	1140
F T Q I N R N G I A C W D T N T E L N P	332
CAATACATTCATTTTATAGTCGCGAAAAATAACAACGATGGTCTTTTTCACAGATTTAT	1200
N T F I L V A E N N T T M V F C N D L S	352
CCATCGATCGGTCTACTAACACGATGTACGTGCTGTGATAATTTCCAGCACTTTTGT	1260
I D R S T N T M Y V L S D N F Q Q L L F	372
TCTCTAAATATGATGCGAAGAACGTAATTTTTCATCACAATGTTGATCTCGATTTCT	1320
S K Y D A K K R N P F I T M F S V S T E V	392
TAACGAATGCTGTGAAGAAAAAGACGATAAACCTAAAGACGGCTTCTCATATTTCTTT	1380
T N A C K K K D D K P K R R L P H I L *	412
AATGAGTTCATGCCAATGTATAAATATATTTATATATCGAAATATTAATAAATACTT	1440
*	
TGCTCAG	

Fig. 2. Assembled sequence of *am-yellow-f* gene and its translation. The cDNA was assembled from the sequences of contig 662 and genomic DNA reads. The end of the cDNA was set arbitrarily 12 nucleotides downstream of presumed polyadenylation signal (bold).

genetic analyses. Nevertheless, the cDNA of the singlet BB160008B20E04 seems to be a strong candidate for being a *yellow* ortholog in *A. mellifera*.

3.4. Phylogeny of MRJP/YELLOW proteins of *A. mellifera*

A phylogenetic tree was calculated from the aligned protein sequences (Fig. 3) using exhaustive maximum parsimony search (Fig. 4). Branch-and-bound bootstrap analysis confirmed the monophyly (100% bootstrap value) of MRJP proteins and both *A. mellifera* and *D. melanogaster* yellow proteins. Within the MRJP branch, an early divergence of MRJP8 was confirmed in all tests, whereas the branching order of other

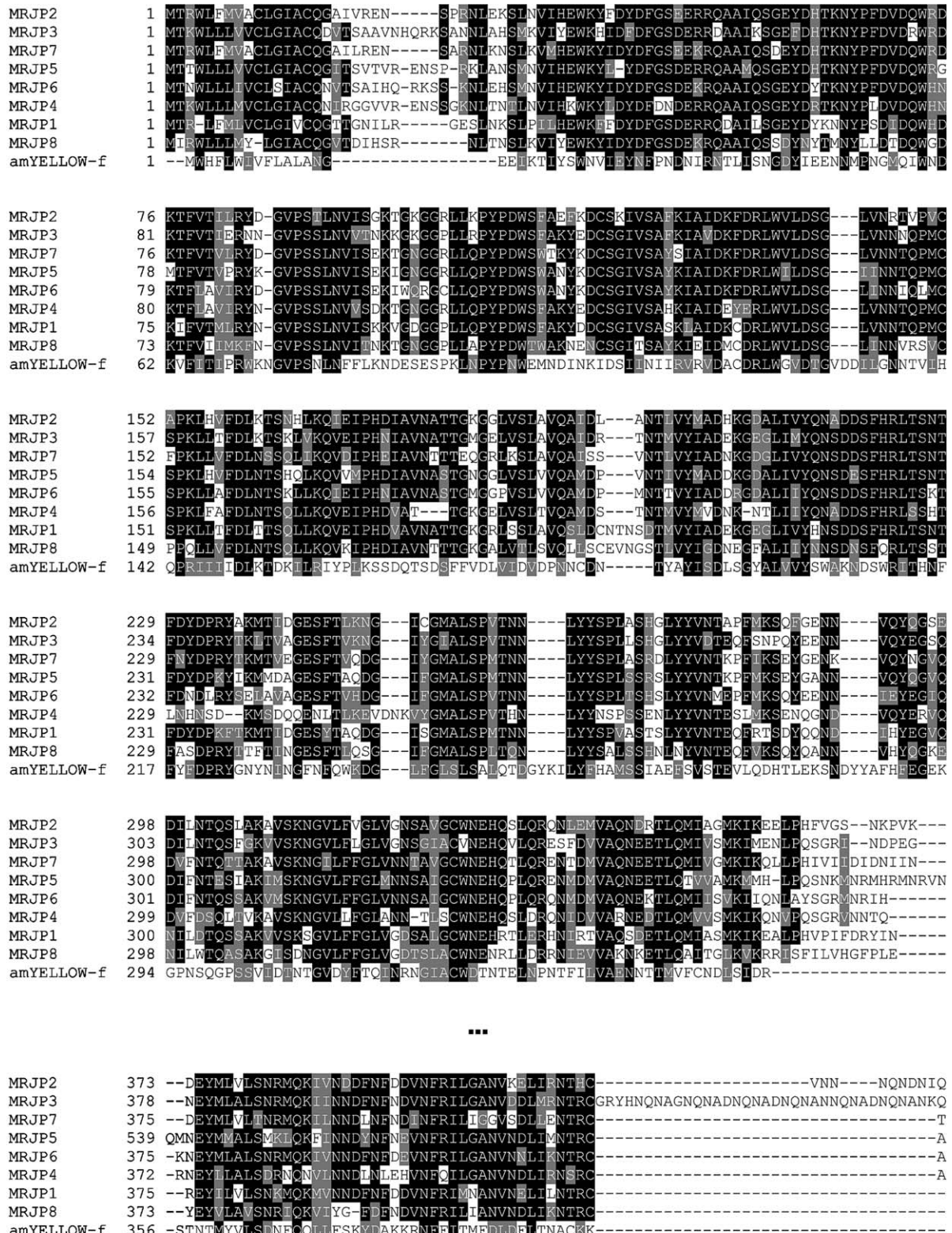


Fig. 3. Alignment MRJP/YELLOW proteins found in *A. mellifera*. The protein sequences were aligned using CLUSTAL W and visualized with BOXSHADE software. Amino acids that are identical in more than 50% of aligned sequences are shaded black, conserved substitutions are grey. For simplicity, the repetitive regions of MRJP5 and MRJP3, which are not present in other MRJPs were removed. Removed blocks of repetitive sequences and missing ends of some protein sequences are indicated by three dots (...).

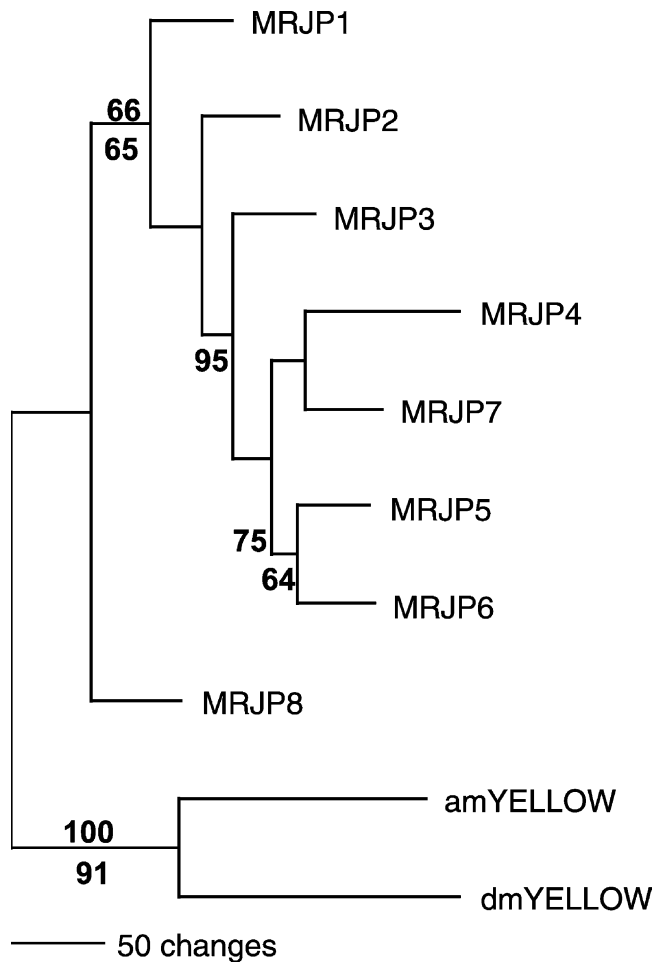


Fig. 4. Phylogeny of MRJP/YELLOW proteins of *A. mellifera*. This is an unweighted maximum parsimony tree (938 steps) found with an exhaustive search. The results of bootstrap analyses using the branch-and-bound method (200 replications) are shown above the branches. The numbers below the branches are maximum likelihood quartet-puzzling values found using TREEPUZZLE V5.0 (WAG + Γ evolutionary model, 5000 puzzling steps). Only bootstrap or puzzle values >60% are shown.

MRJPs could not be clearly resolved. Interestingly, MRJP5 and MRJP6 group together, which underlines their relatively recent divergence.

The analysis shows that both honeybee and *Drosophila* YELLOW-f proteins clearly form a monophyletic group distant from MRJPs, with MRJP8 appearing to be the earliest divergence.

4. Discussion

4.1. MRJPs in honeybees: still poorly understood proteins

It is surprising that cDNAs encoding all five major proteins of royal jelly characterized so far, plus three new MRJPs were found in honeybee brain cDNA

library. Only 40 cDNAs were reported to consist of more than 10 singlets in this normalized library (Whitfield et al., 2002). Five of them code for MRJPs, including the most abundant cDNA, *mrjp2*, with 44 singlets (see Table 1 for singlets number). Although the *mrjp1* transcript was found in honeybee brain (see below), it is improbable that all MRJPs are expressed in honeybee brain. It is possible that hypopharyngeal glands of nurses, which express high level of MRJPs, contaminated the honeybee brain preparations used in construction of the EST library (Whitfield et al., 2002).

MRJP1 is the most abundant and arguably best characterized glycoprotein of royal jelly with a molecular mass of 56–57 kDa (Hanes and Simuth, 1992; Ohashi et al., 1997; Schmitzova et al., 1998; Simuth, 2001). In the royal jelly, MRJP1 associates with a small peptide termed apisimin (Bilikova et al., 2002) and possibly with other compounds in a large complex of 420 kDa (Kimura et al., 1996; Simuth, 2001). MRJP1 mRNA was found to be differentially expressed in heads of early emerged honeybees (Kucharski et al., 1998), nurses and also forager honeybees (Klaudiny et al., 1994b). Its expression was localized to hypopharyngeal glands (Ohashi et al., 1997), and also to a subset of Kenyon cells (intrinsic neurons) of mushroom bodies—presumed centers of learning and memory in the honeybee brain (Kucharski et al., 1998). Therefore it would seem that MRJP1 not only functions as a component of larval food but it also plays a role in the honeybee's brain. This may increase the selective pressure acting on the gene. MRJP1 shows diverse biological activities in several heterologous systems. It enhances the cell proliferation of rat hepatocytes (Kamakura et al., 2001b), stimulates the growth of human lymphocytes in a serum-free medium (Watanabe et al., 1996), and shows an antifatigue effect in mice (Kamakura et al., 2001a).

Comparisons of available *mrjp* sequences with their counterparts in the EST database revealed an unequal distribution of nucleotide sequence diversity. Besides *mrjp1*, no SNPs were detected in *mrjp2* cDNAs, whereas *mrjp4*, *mrjp5* and particularly *mrjp3* showed a number of nucleotide variations in the honeybee population, which often altered the encoded amino acid residues (see Table 2). No conclusions about sequence diversity can be drawn for the three new *mrjps* (*mrjp6*–*8*) presented here since no comparison data are available.

It can only be speculated as to what causes these differences among closely related cDNAs. The location of *mrjp1* and *mrjp2* loci close to centromeres or chromosomal tips (regions with lower recombination rates) might explain their low nucleotide diversity (Begun and Aquadro, 1992). Alternatively, unequal selection pressures might have been acting on different *mrjp* genes. This is supported by the identification of MRJP1 in

royal jelly and brain described above. Differences between *mrjps* in sequence variability may reflect the function of their proteins. It is apparent that at least four of them (identified in royal jelly at high amounts; MRJP1-3 and MRJP5) (Hanes and Simuth, 1992; Kubo et al., 1996; Ohashi et al., 1997; Schmitzova et al., 1998) exert a nutritional function. Their mRNAs (*mrjp5* not analyzed) are highly expressed in the heads (hypopharyngeal glands) of nurse honey bees (Klaudiny et al., 1994a; Klaudiny et al., 1994b; Kucharski et al., 1998; Ohashi et al., 1997; Schmitzova et al., 1998) and down-regulated later upon onset of foraging behaviour (Ohashi et al., 2000). The nutritional function of MRJPs coincides with the relatively high percentage of essential amino acids (Schmitzova et al., 1998) and with the appearance of extensive nitrogen-rich repetitive regions in the primary structure of several MRJPs (e.g. MRJP3 and MRJP5 (Albert et al., 1999a; Albert et al., 1999b)). There is insufficient information to suggest what function the new MRJPs presented here play.

4.2. *mrjp/yellow* gene family multiplication and functional diversity of its members

As we show here, there are at least eight MRJPs, which are expressed in *A. mellifera* heads. All MRJPs are highly homologous to each other (Figs. 3 and 4). As proposed previously, it seems that MRJPs diverged from yellow-like proteins to gain a novel nutritional function in honeybees. Based on expression studies it also appears that some MRJPs have acquired an additional function in honeybee brain. To date, no YELLOW protein homolog has been identified in the honeybee. In this work, two novel honeybee yellow-like cDNAs called *am-yellow* and *am-yellow-f*, were partially characterized.

Although 14 yellow-like genes have been found in the *Drosophila* genome (Drapeau, 2001), the function of most of them is unknown. YELLOW-f and YELLOW-f2 proteins were recently shown to possess a dopachrome conversion activity (Han et al., 2002). YELLOW protein itself, identified originally for its crucial role in cuticle melanisation (Nash, 1976) was also found to be temporarily expressed in *Drosophila* brain where it somehow controls male courtship behavior (Drapeau et al., 2003). It is conceivable that *am-YELLOW* proteins described here play similar roles in *A. mellifera*.

How MRJPs evolved from YELLOW precursors is unknown. However, proteins belonging to the MRJP/YELLOW family have been identified in the saliva of several blood-sucking insect parasites such as *Lutzomyia longipalpis* (Charlab et al., 1999) and *Phlebotomus papatasi* (Valenzuela et al., 2001). With one exception, their physiological functions have not been resolved,

the exception being the MRJP-like saliva protein of sand fly *Phlebotomus duboscqui*, to which a lectin activity was attributed (Volf et al., 2002). The possibility that MRJP-like saliva proteins are the intermediates in the evolution from yellow-like function towards royal jelly components is to us an attractive hypothesis. It would therefore be interesting to investigate the tissue-specific expression of *Drosophila yellow* genes or analyze the secreted components of *Drosophila*'s saliva for the presence of yellow homologues.

5. Conclusions

This article gives a present-day view on *mrjp/yellow* genes in honeybee. It includes only genes, which are unequivocally expressed in the head. Completion of the genomic sequence of *A. mellifera* could well reveal additional *mrjp/yellow* genes expressed in other tissues, and/or exerting different physiological function(s). It will also show whether MRJP genes have a tandem organization as described for several yellow genes of *Drosophila* (Drapeau, 2001). It may also explain whether the different degrees of genetic variability observed among MRJP genes are due to their chromosomal localization (Begun and Aquadro, 1992).

Knowledge of the developmental stage- and tissue-specific expression pattern of *mrjp* genes would help us better understand the biological function of individual MRJPs, which may differ from the nutritive function found for some members. Of particular interest is the question, what triggered the evolutionarily "sudden" genomic amplification and following modification of *mrjp* genes? We believe that this process could have accompanied the evolution of siblings-feeding behaviour, one of the crucial characteristics of highly advanced social insects.

Further work will be required to learn more about the function of this family of proteins, which, in spite of mutual homology, are apparently involved in the regulation of many diverse physiological processes.

Acknowledgements

The authors wish to acknowledge Margita Plšková and Viera Albertová for their technical assistance, Jürgen Schmitz and Debashish Bhattacharya for their help with phylogenetic analyses, Barbora Malecová and Kirsty MacPherson for critical reading of the manuscript.

References

- Albert, S., Bhattacharya, D., Klaudiny, J., Schmitzova, J., Simuth, J., 1999a. The family of major royal jelly proteins and its evolution. *Journal of Molecular Evolution* 49, 290–297.

- Albert, S., Klaudiny, J., Simuth, J., 1996. Newly discovered features of the updated sequence of royal jelly protein RJP57-1; longer repetitive region on C-terminus and homology to *Drosophila melanogaster* yellow protein. *Journal of Apicultural Research* 35, 63–68.
- Albert, S., Klaudiny, J., Simuth, J., 1999b. Molecular characterization of MRJP3, highly polymorphic protein of honeybee (*Apis mellifera*) royal jelly. *Insect Biochemistry and Molecular Biology* 29, 427–434.
- Albert, S., Schmitz, J., 2002. Characterization of major royal jelly protein-like DNA sequences in *Apis dorsata*. *Journal of Apicultural Research* 40, 75–82.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410.
- Begun, D.J., Aquadro, C.F., 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356, 519–520.
- Bilikova, K., Hanes, J., Nordhoff, E., Saenger, W., Klaudiny, J., Simuth, J., 2002. Apisimin, a new serine-valine-rich peptide from honeybee (*Apis mellifera* L.) royal jelly: purification and molecular characterization. *FEBS Letters* 528, 125–129.
- Charlab, R., Valenzuela, J.G., Rowton, E.D., Ribeiro, J.M.C., 1999. Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis*. *Proceedings of the National Academy of Sciences of the United States of America* 96, 15155–15160.
- Devereux, J., Haeberli, P., Smithies, O., 1984. A comprehensive set of sequence analysis programs for VAX. *Nucleic Acids Research* 12, 387–395.
- Drapeau, M.D., 2001. The family of yellow-related *Drosophila melanogaster* proteins. *Biochemical and Biophysical Research Communications* 281, 611–613.
- Drapeau, M.D., Radovic, A., Wittkopp, P.J., Long, A.D., 2003. A gene necessary for normal male courtship, yellow, acts downstream of fruitless in the *Drosophila melanogaster* larval brain. *Journal of Neurobiology* 55, 53–72.
- Han, Q., Fang, J., Ding, H., Johnson, J.K., Christensen, B.M., Li, J., 2002. Identification of *Drosophila melanogaster* yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes. *Biochem J* 368, 333–340.
- Hanes, J., Simuth, J., 1992. Identification and partial characterization of the major royal jelly protein of the honey bee (*Apis mellifera* L.). *Journal of Apicultural Research* 31, 22–26.
- Kamakura, M., Mitani, N., Fukuda, T., Fukushima, M., 2001a. Antifatigue effect of fresh royal jelly in mice. *Journal Nutrition Sciences and Vitaminology (Tokyo)* 47, 394–401.
- Kamakura, M., Suenobu, N., Fukushima, M., 2001b. Fifty-seven-kDa protein in royal jelly enhances proliferation of primary cultured rat hepatocytes and increases albumin production in the absence of serum. *Biochemical and Biophysical Research Communications* 282, 865–874.
- Kaufman, T.C., Severson, D.W., Robinson, G.E., 2002. The *Anopheles* genome and comparative insect genomics. *Science* 298, 97–98.
- Kimura, Y., Kajiyama, S., Kanaeda, J., Izukawa, T., Yonekura, M., 1996. N-linked sugar chain of 55-kDa royal jelly glycoprotein. *Biosciences Biotechnology Biochemistry* 60, 2099–2102.
- Klaudiny, J., Hanes, J., Kulifajova, J., Albert, S., Simuth, J., 1994a. Molecular cloning of two cDNAs from the head of the nurse honey bee (*Apis mellifera* L.) for coding related proteins of royal jelly. *Journal of Apicultural Research* 33, 105–111.
- Klaudiny, J., Kulifajova, J., Crailsheim, K., Simuth, J., 1994b. New approach to the study of division of labour in the honeybee colony (*Apis mellifera* L.). *Apidologie* 25, 596–600.
- Kubo, T., Sasaki, M., Nakamura, J., Sasagawa, H., Ohashi, K., Takeuchi, H., Natori, S., 1996. Change in the Expression of Hypopharyngeal-Gland Proteins of the Worker Honeybees (*Apis mellifera* L.) with Age and/or Role. *Journal of Biochemistry (Tokyo)* 119, 291–295.
- Kucharski, R., Maleszka, R., 2002. Evaluation of differential gene expression during behavioral development in the honeybee using microarrays and northern blots. *Genome Biology* 3, (RESEARCH0007).
- Kucharski, R., Maleszka, R., Hayward, D.C., Ball, E.E., 1998. A royal jelly protein is expressed in a subset of Kenyon cells in the mushroom bodies of the honey bee brain. *Naturwissenschaften* 85, 343–346.
- Malecova, B., Ramser, J., O'Brien, J.K., Janitz, M., Judova, J., Leh-rach, H., Simuth, J., 2003. Honeybee (*Apis mellifera* L.) mrjp gene family: computational analysis of putative promoters and genomic structure of mrjp1, the gene coding for the most abundant protein of larval food. *Gene* 303, 165–175.
- Nash, W.G., 1976. Patterns of pigmentation color states regulated by the y locus in *Drosophila melanogaster*. *Developmental Biology* 48, 336–343.
- Ohashi, K., Natori, S., Kubo, T., 1997. Change in the mode of gene expression of the hypopharyngeal gland cells with an age-dependent role change of the worker honeybee *Apis mellifera* L. *European Journal of Biochemistry* 249, 797–802.
- Ohashi, K., Sasaki, M., Sasagawa, H., Nakamura, J., Natori, S., Kubo, T., 2000. Functional flexibility of the honey bee hypopharyngeal gland in a dequeened colony. *Zoological Science* 17, 1089–1094.
- Rice, P., Longden, I., Bleasby, A., 2000. EMBOS: the European Molecular Biology Open Software Suite. *Trends in Genetics* 16, 276–277.
- Schmitzova, J., Klaudiny, J., Albert, S., Schroeder, W., Schreckengost, W., Hanes, J., Judova, J., Simuth, J., 1998. A family of major royal jelly proteins of the honeybee *Apis mellifera* L. *Cellular and Molecular Life Sciences* 54, 1020–1030.
- Simuth, J., 2001. Some properties of the main protein of honeybee (*Apis mellifera*) royal jelly. *Apidologie* 32, 69–80.
- Valenzuela, J.G., Belkaid, Y., Garfield, M.K., Mendez, S., Kamhawi, S., Rowton, E.D., Sacks, D.L., Ribeiro, J.M.C., 2001. Toward a defined anti-Leishmania vaccine targeting vector antigens: Characterization of a protective salivary protein. *Journal of Experimental Medicine* 194, 331–342.
- Volf, P., Skarupova, S., Man, P., 2002. Characterization of the lectin from females of *Phlebotomus duboscqi* sand flies. *European Journal of Biochemistry* 269, 6294–6301.
- Watanabe, K., Shinmoto, H., Kobori, M., Tsushida, T., Shinokara, K., Kanaeda, J., Yonekura, M., 1996. Growth stimulation with honey royal jelly DIII protein of lymphocytic cell lines in a serum-free medium. *Biotechnology Techniques* 10, 959–962.
- Whitfield, C.W., Band, M.R., Bonaldo, M.F., Kumar, C.G., Liu, L., Pardinas, J.R., Robertson, H.M., Soares, M.B., Robinson, G.E., 2002. Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Research* 12, 555–566.