HAL
archives-ouvertes.fr

Molecular cloning and analysis of four cDNAs from the heads of Apis cerana cerana nurse honeybees coding for major royal jelly proteins

Songkun Su, Stefan Albert, Shenglu Chen, Boxiong Zhong

▶ To cite this version:

Songkun Su, Stefan Albert, Shenglu Chen, Boxiong Zhong. Molecular cloning and analysis of four cDNAs from the heads of Apis cerana cerana nurse honeybees coding for major royal jelly proteins. Apidologie, Springer Verlag, 2005, 36 (3), pp.389-401. hal-00892149

HAL Id: hal-00892149
https://hal.archives-ouvertes.fr/hal-00892149

Submitted on 1 Jan 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Apidologie 36 (2005) 389–401
© INRA/DIB-AGIB/ EDP Sciences, 2005
DOI: 10.1051/apido:2005026

Original article

Molecular cloning and analysis of four cDNAs from the heads of *Apis cerana cerana* nurse honeybees coding for major royal jelly proteins[1]

Songkun SU[a], Stefan ALBERT[b], Shenglu CHEN[a]*, Boxiong ZHONG[a]

[a] Laboratory of Apicultural Research, Animal Science College, Zhejiang University Hangzhou 310029, China
[b] Institute of Medical Radiation and Cell Research, University of Würzburg, Versbacherstrasse 5, 97078 Würzburg, Germany

Received 8 September 2004 – revised 10 January 2005 – accepted 13 January 2005

Published online 7 July 2005

Abstract – A cDNA library was constructed from 8-day-old worker heads of *Apis cerana cerana*. A DIG-labeled probe derived from part of an *Apis cerana mrjp3* genomic segment was used to screen the library. One hundred and twenty positive clones were identified and characterized. Thirty one clones were homologous with major royal jelly proteins (MRJPs) of *Apis mellifera*. The most abundant MRJP homologue was MRJP1 (11 clones), followed by MRJP3 (10 clones), MRJP2 (7 clones) and MRJP5 (3 clones). Clones containing *A. cerana cerana* MRJP1, MRJP2, MRJP3 and MRJP5 cDNAs were identified, completely sequenced, and analyzed with bioinformatics software. Several lines of evidence suggested that the identified cDNAs code for major royal jelly proteins of *A. cerana*. In addition to polymorphic regions of MRJP3 and MRJP5, another polymorphic repetitive region was found in AccMRJP2. The polymorphism of AcMRJP2 and AcMRJP5 repeat regions were tested by PCR with genomic DNAs of individual honeybees. Different properties of the repetitive regions of MRJP2 genes in two closely related *Apis* species were discussed.

Apis cerana cerana* / royal jelly protein / polymorphism / VNTR / midisatellite / *Apis mellifera

1. INTRODUCTION

Royal jelly (RJ), a secretion of both the hypopharyngeal and mandibular glands of nurse workers, is believed to play a central role in honeybee queen development (Knecht and Kaatz, 1990; Lensky and Rakover, 1983; Moritz and Southwick, 1992). Proteins are an important component of RJ, forming about 50% of the dry mass (Rembold, 1987). Five major proteins of *Apis mellifera* RJ, called MRJP1-5, form a significant part (~90%) of RJ total protein (Schmitzova et al., 1998). The cDNA sequences of MRJP1-MRJP5 were isolated from a *A. mellifera* cDNA library of nurse honeybee heads (Albert et al., 1999a; Klaudiny et al., 1994; Schmitzova et al., 1998). Corresponding MRJP1-5 proteins were identified in RJ separated by 2-D electrophoresis (Sano et al., 2004). Recently, three cDNAs coding for new but less abundant members of MRJP protein family were identified from a *A. mellifera* brain EST library (Albert and Klaudiny, 2004).

Biological activities of *A. mellifera* major royal jelly proteins have been studied in various systems. The protein fractions of royal jelly were confirmed to possess a high antioxidative activity and scavenging ability against reactive oxygen species (Nagai and Inoue, 2004). MRJP1 enhanced cell proliferation of rat hepatocytes (Kamakura et al., 2001b), stimulated the growth of human lymphocytes in a serum-free medium

* Corresponding author: susongkun@zju.edu.cn
[1] Manuscript editor: Klaus Hartfelder

Article published by EDP Sciences and available at http://www.edpsciences.org/apido or http://dx.doi.org/10.1051/apido:2005026

(Watanabe et al., 1996) and showed an antifatigue effect in mice (Kamakura et al., 2001a). MRJP3 exhibited potent immunoregulatory effects in vitro and in vivo (Okamoto et al., 2003).

In contrast to *A. mellifera,* information on RJ of other honeybee species (including *Apis cerana*) is scarce. Takenaka and Takenaka (1996) reported that chemical composition, i.e., proteins, 10-hydroxydecenoic acid, and glucose/fructose ratio, differed between *A. mellifera* and *A. cerana* royal jelly. Recently an EST library was prepared from the hypopharyngeal glands of *A. cerana indica* (Srisuparbh et al., 2003). From this library, the MRJP1 homologue and apisimin cDNAs of *A. cerana indica* were isolated and sequenced. The MRJP1 of *A. cerana indica* showed 93% and 90% homology with MRJP1 of *A. mellifera* at the nucleotide and amino acid levels, respectively.

There are about two million *A. cerana cerana* colonies in China (Chen et al., 2002). *A. cerana cerana* is widely used for commercial beekeeping in mountain areas of South China, primarily due to its resistance against diseases, wasps and bee mites.

In this report, we constructed a cDNA library from 8-day-old nurse honeybee heads of *A. cerana cerana,* screened it by hybridization and identified cDNAs encoding MRJP1, MRJP2, MRJP3, and MRJP5 homologues. We provided full cDNA sequences of these genes and partial genomic sequences of *A. cerana cerana* MRJP1, 5, and 7. The sequences were compared with the MRJPs of *A. mellifera* and several lines of evidence showed that isolated cDNAs encode functional homologues of MRJPs in *A. cerana*. Moreover, we identified and characterized an unexpected polymorphism of MRJP2 in *A. cerana.*

2. MATERIALS AND METHODS

2.1. Biological samples

Nurse honey bees (*Apis cerana cerana*) were obtained from the Laboratory of Apicultural Research (Huajiachi campus, Zhejiang University, Hangzhou, China) as follows: newly emerged workers (less than 1-day-old) were marked with paint mark pen and put back into the colonies, to enable trophallactic contacts with other bees and normal development. Eight-day-old nurse bees were collected and anesthetized on ice, the heads were removed, frozen in liquid nitrogen and stored at –80 °C.

European honeybees of two putative races were obtained from colonies separated by several hundred kilometers, *Apis mellifera mellifera* from Göttingen, Germany, *Apis mellifera carnica* from Bratislava, Slovakia and Würzburg, Germany.

2.2. Construction of a cDNA library from 8-day-old nurse bee heads

Total RNA from the heads of nurse honeybees was prepared using the TRIZOL Kit (Promega) according to manufacturer's instructions. mRNA was extracted using PolyATtract^{TM} (Promega) mRNA method. 5 µg of purified mRNA were reverse-transcribed into cDNA and Lambda-ZAP phage library was constructed using ZAP-cDNA Synthesis Kit and ZAP-cDNA Gigapack III Gold Cloning Kit (Stratagene) following the respective manuals (www.stratagene.com).

2.3. PCR amplification of MRJP genomic fragments and cDNA library screening

Two primers termed P212 and P218 designed to prime within regions conserved among five MRJPs of *A. mellifera* [P212: AAA GT(G/A) T(T/G)G GAA G(T/A)C AAT CGA TG; P218: TGC CT(T/C) GG(C/T) ATA G(C/T)T TGT C] were used for PCR amplification of genomic DNAs of *A. cerana.* Reaction mixtures included commercial PCR buffer, containing 1 µM primers, 1 µg *A. cerana* genomic DNA and polymerase mixture (Taq:DeepVent = 5:1; AmpliTaq, Perkin Elmer and NEB, respectively).

After initial heating at 94 °C and polymerases addition at 80 °C, 32 cycles of 30 s at 94 °C, 60 s at 52 °C and 120 s at 72 °C were run. Amplified products were cloned into PCR2.1 vector (Invitrogen).

Studies of the polymorphic MRJP2 and MRJP5 alleles were done as above with the following primers: P28 (TTA ATG AGA AAT ACT CAT TGC G) and P29 (AAC GAC GAA CTT GAT TAT CAT TC) for MRJP2. The primers P126 (AGA CTC TTC AAA CGG TCG TTG) and P127 (CTG TAA TTT CAT ACT TAA AGC CAT C) were designed to amplify the DRM repetitive region of MRJP5.

The amplified cDNA library was screened by hybridization using standard protocols. We used an *mrjp3* gene fragment amplified from *A. cerana* genome as a hybridization probe employing DIG High Prime DNA Labeling and Detection Starter Kit I (Roche, Germany).

2.4. Primary characterization of positive clones

The phage extracts from the positive clones were transferred from plate into test tubes with 200 µL SM buffer. To determine the length of cDNA fragments inserted and check the purity of isolated clones, we amplified the cDNA inserts with T3 (5'-AAT TAA CCC TCA CTA AAG GG) and T7 (5'-TAA TAC GAC TCA CTA TAG GG) primers. Polymerization reactions were done as above in 50 µL volume with 5 µL positive clone extract (Lambda-ZAP, Stratagene).

2.5. DNA sequencing and bioinformatics analyzes

Plasmid DNAs were sequenced by the cycle sequencing method using the Prism Ready Reaction Dyedeoxy Terminator kit on an ABI PRISM 377 DNA Sequencer according to the manufacturer's instructions. Obtained sequences were compared with GenBank using the BLASTN and BLASTX programs (http://www.ncbi.nlm.nih.gov).

DNA sequences were assembled with the help of DNATOOLs and DNASTAR program packages.

To distinguish between proteins of different origin, the MRJPs of *Apis mellifera*, *Apis cerana indica*, *Apis cerana cerana* were termed AmMRJPs, AciMRJPs and AccMRJPs, respectively.

The cDNAs identified in this work were deposited in GenBank under following accession numbers: *A. cerana cerana MRJP1* (*AccMRJP1*): **AY279539**; *AccMRJP2*: **AY392758**; *AccMRJP3*: **AY394726**; *AccMRJP5*: **AY392757**. The genomic fragments encoding AcMRJPs were deposited in GenBank under following accession numbers: AcMRJP1: **AY862495**; AcMRJP7: **AY862496**; AcMRJP5: **AY862497**.

Phylogenetic analyses were done with CLUSTAL W and TREEPUZZLE.5.0.

3. RESULTS

3.1. Generation of MRJP genomic fragments of *Apis cerana*

To obtain initial information about *mrjp* genes of *A. cerana* and to prepare hybridization probes for screening a cDNA library, we looked for regions that were conserved among known *mrjp* cDNAs of *A. mellifera*. The rationale was that the regions conserved among *mrjps* of *A. mellifera* would also be conserved in other species. The cDNAs encoding *mrjp1–5* of *A. mellifera* were aligned to identify the regions of high conservancy. The primers P212 and P218 (see Materials and Methods), each with two degenerated positions were designed to anneal to these conserved regions. The PCR reaction was run at a relatively low annealing temperature (52 °C) using a mixture of proofreading polymerase and Taq polymerase. A broad band of 1350–1500 bp was amplified, which might be a mixture of several products of similar size.

We cloned the PCR products and sequenced five individual clones. Each of the clones contained an *mrjp*-like insert; two of them were identical. Cloned inserts showed homology to *mrjp1* (1×), *mrjp7* (1×), *mrjp3* (2×), and *mrjp5* (1×). We termed them *Acmrjp1*, *Acmrjp7*, *Acmrjp3* and *Acmrjp5*, respectively.

The significantly larger size of the amplified genomic fragments (1350–1500 bp versus expected size of cDNA fragments ~ 680 bp) indicated the presence of introns, which was confirmed by DNA sequencing. Three introns located at the same positions were found in all amplified genomic fragments (see Figs. 2, 3).

3.2. Screening of the cDNA library of *Apis cerana cerana*

A cDNA library containing about 90% recombinant clones was prepared from the heads of 8-day-old nurse bees. More than 200 positive clones were found from the cDNA library with a DIG-labeled Ac*mrjp3* genomic fragment as a probe. One hundred and twenty of these clones were extracted into SM buffer and subjected to PCR with T3/T7 primers. Fifty five of them showed single cDNA inserts that were 1200–2500 bp in size. The PCR products amplified from these clones were sequenced. Thirty one clones were homologous to *mrjps* of *A. mellifera*. The most abundant *mrjp* homologues were *mrjp1* (11 clones) followed by *mrjp3* (10 clones), *mrjp2* (7 clones) and *mrjp5* (3 clones). The complete cDNA sequences of *Accmrjp1*, *2*, *3*, and *5* were obtained by sequencing of the clones #103, #46, #50 and #91 respectively.

3.3. Characterization of sequences encoding MRJPs of *Apis cerana cerana*

AccMRJP1 cDNA contained an open reading frame (ORF) of 1445 nucleotides (poly[A]

tail not included) encoding a protein of 433 amino acids, which was highly similar (99.5% and 90.5% identity) to its homologues from *A. cerana indica* (Srisuparbh et al., 2003) and *A. mellifera* (Schmitzova et al., 1998) at the protein level. According to N-terminal sequencing of the *A. cerana* RJ protein (Srisuparbh et al., 2003), the cleavage site for signal peptidase was localized between Ser_{20} and Ser_{21}. Three potential N-linked glycosylation sites were found at amino acids 29, 145, and 178. The polyadenylation signal AATAAA was located 14 bp upstream of the poly(A) tail.

The sequence of **AccMRJP2** cDNA was 1590 bp long and included an open reading frame (ORF) of 1404 nucleotides encoding a protein of 468 amino acids (see Fig. 1). The sequence of the encoded protein contained all three internal peptide sequences (Fig. 1, underlined) determined by sequencing of the protein from RJ of *A. cerana indica* (Srisuparbh et al., 2003). The putative signal peptidase cleavage site was between Gly_{17} and Ala_{18}. Two potential N-linked glycosylation sites were found at asparagines 145 and 178. Interestingly, the repetitive region similar to that of MRJP3 (see below), consisting of 9 copies of NQKNN pentapeptide, was found at the C-terminal part of the deduced AccMRJP2 protein.

AccMRJP3 cDNA was 1977 bp long. The cDNA and inferred amino acid sequences were shown in Figure 2. The cDNA sequence contained an ORF (nucleotides 46–1824), which encoded a polypeptide of 593 amino acid residues. The deduced amino acid sequence of mature peptide began with AAVNHQRKS, which was identical to N-terminal sequence of the protein eluted in the peak A1 of separated RJ proteins of *A. cerana,* [(G/A)AVNHQRKSA] and nearly identical to N-terminal sequence of MRJP3 of *Apis mellifera* [AAVNHQ(R/K)KSANNLAHS] (Schmitzova et al., 1998). Similar to AmMRJP3, a repetitive region was also found at the C-terminal region of the AccMRJP3 ORF (Fig. 2, underlined). The basic segment was repeated 27 time in tandem. The sequence AATAAATAAAATAAA, containing two separated or three partially overlapping consensus polyadenylation signal sequences (AATAAA), was located 14 bp upstream from the poly(A) tail.

The cDNA of **AccMRJP5** comprised 1970 nucleotides. The nucleotide and the inferred amino acid sequences were depicted in Figure 3. The cDNA sequence contained an ORF encoding a protein of 598 amino acid residues with high identity with AmMRJP5 (83.6%). As described for the AmMRJP5, the AccMRJP5 protein also contained an extensive repeated region located between amino acid residues 367–540. Its basic repetitive unit was 9 bp and encoded a tripeptide with consensus asp-arg-met (DRM in single-letter code). The repetitive segments exhibited large differences between AccMRJP5 and AmMRJP5.

3.4. Studies of the polymorphic repeats of AcMRJP2 and AcMRJP5

The sequences of MRJP2 and MRJP5 cDNAs isolated from another *A. cerana* race, *A. cerana indica,* were deposited in GenBank by the group of S. Sittipraneed, Chulalongkorn University, Thailand (accession numbers **AF525777** and **AY532369** respectively). They were nearly identical with our cDNAs, except for their repetitive regions. For example, AccMRJP2 contained 9 copies of the reiterated pentapeptide unit, whereas Thailand isolation contained only 8 of them. This was surprising because no polymorphism has been reported for MRJP2 of *A. mellifera* (Schmitzova et al., 1998). Prompted by this observation, we set out to study the polymorphism of the MRJP2 repeat in detail. Two primers (P28 and P29; see Fig. 1, arrows) were designed to flank the repetitive region of MRJP2 and used in PCR with the genomic DNAs of *A. cerana* and *A. mellifera* individuals collected from geographically distant colonies. Figure 4A shows that there was a high variability of the repeat size in *A. cerana* even within individuals from the same colony, whereas the analogous region of *A. mellifera* showed the same size among all individuals originating from distant locations.

For studying the polymorphism of the DRM repeat of MRJP5, PCR primers P126 and P127 (Materials and Methods) were designed to amplify this region. Similar to MRJP2, a clear size polymorphism with allele sizes ranging from ~560–650 bp was observed in *A. cerana* (Fig. 4B). MRJP5 of *A. mellifera* was also polymorphic (our unpublished data and Fig. 4B), but observed size differences among individual alleles were smaller. Additional bands of the intermediate size presumably representing

```
   1  GCACGAGGACATTCTGCAGTATCCTAAAAAAATGACAAAGTGGTTGTTTATGGTGGCATGCCTTGGCATAGCTTG   75
   1                                 M  T  K  W  L  F  M  V  A  C  L  G  I  A  C    15

  76  TCAAGGTGCCATTATTCGACAAAATTCTGCAAAAAACTTGGAAAATTCGTTGAACGTAATTCACGAATGGAAATA   150
  16   Q  G  A  I  I  R  Q  N  S  A  K  N  L  E  N  S  L  N  V  I  H  E  W  K  Y    40

 151  TATCGATTATGATTTCGGTAGCGAAGAAAGAAGACAAGCTGCGATTCAATCTGGCGAATACGATCATACGAAAAA   225
  41   I  D  Y  D  F  G  S  E  E  R  R  Q  A  A  I  Q  S  G  E  Y  D  H  T  K  N    65

 226  TTATCCCTTCGATGTCGATCAATGGCATGATAAGACTTTTGTCACCATACTAAAGTACGATGGTGTGCCTTCTAC   300
  66   Y  P  F  D  V  D  Q  W  H  D  K  T  F  V  T  I  L  K  Y  D  G  V  P  S  T    90

 301  TTTGAACATGATATCTAACAAAATCGGTAAGGGTGGACGCCTTCTACAACCATATCCTGATTGGTCGTGGGCAGA   375
  91   L  N  M  I  S  N  K  I  G  K  G  G  R  L  L  Q  P  Y  P  D  W  S  W  A  E    115

 376  GAATAAAGATTGCTCTGGAATCGTGAGCGCTTTCAAAATTGCGATTGACAAATTCGACAGATTGTGGGTTTTGGA   450
 116   N  K  D  C  S  G  I  V  S  A  F  K  I  A  I  D  K  F  D  R  L  W  V  L  D    140

 451  TTCAGGTCTTATCAATAGAACTGAACCTATATGTGCTCCAAAGTTGCATGTCTTTGATCTGAAAAACACAAAGCA   525
 141   S  G  L  I  N  R  T  E  P  I  C  A  P  K  L  H  V  F  D  L  K  N  T  K  H    165

 526  CCTTAAGCAAATCGAAATACCGCATGATATTGCCGTAAATGCCACCACAGGAAAGGGAGGGCTAGTCTCTCTAGT   600
 166   L  K  Q  I  E  I  P  H  D  I  A  V  N  A  T  T  G  K  G  G  L  V  S  L  V    190

 601  TGTTCAAGCCATGGATCCTATGAATACTTTAGTATACATAGCAGACCATAAGGGTGATGCTTTGATCGTCTATCA   675
 191   V  Q  A  M  D  P  M  N  T  L  V  Y  I  A  D  H  K  G  D  A  L  I  V  Y  Q    215

 676  AAATTCCGATGATTCCTTCCATCGAATGACTTCCAACACTTTCGATTACGATCCCAGATATGCCAAAATGACGAT   750
 216   N  S  D  D  S  F  H  R  M  T  S  N  T  F  D  Y  D  P  R  Y  A  K  M  T  I    240

 751  CAATGGAGAAAGTTTCACATTGAAAAATGGAATTTGTGGAATGGCTCTTAGTCCCGTGACGAACAATCTTTATTA   825
 241   N  G  E  S  F  T  L  K  N  G  I  C  G  M  A  L  S  P  V  T  N  N  L  Y  Y    265

 826  CAGTCCTCTCGCTTCTCACGGTTTGTATTATGTCAACACGGAACCATTTATGAAATCACAATTTGGAGACAATAA   900
 266   S  P  L  A  S  H  G  L  Y  Y  V  N  T  E  P  F  M  K  S  Q  F  G  D  N  N    290

 901  TAACGTGCAATATGAAGGATCCCAAGATACTTTGAACACGCAATCATTGGCTAAAGCAGTATCGAAAGATGGCGT   975
 291   N  V  Q  Y  E  G  S  Q  D  T  L  N  T  Q  S  L  A  K  A  V  S  K  D  G  V    315

 976  CCTCTTCGTCGGACTTGTGGGTAATTCAGCTCTTGGATGCTTGAACGAGCATCAACCACTTCAGAGAGAAAATTT   1050
 316   L  F  V  G  L  V  G  N  S  A  L  G  C  L  N  E  H  Q  P  L  Q  R  E  N  L    340

1051  AGAACTGGTCGCCCAAAATGAAAAAACACTTCAAATGATCGCAGGTATGAAAATTAAGGAAGAGCTTCCACATTT   1125
 341   E  L  V  A  Q  N  E  K  T  L  Q  M  I  A  G  M  K  I  K  E  E  L  P  H  F    365

1126  CGTAGGAAGTAACAAACCTGTAAAGGACGAATATATGTTAGTTTTAAGTAACAAAATGCAGAAAATAGTAAATAA   1200
 366   V  G  S  N  K  P  V  K  D  E  Y  M  L  V  L  S  N  K  M  Q  K  I  V  N  N    390

1201  TGATTTTAATTTCAACGACGTAAACTTCCGAATTTTGGGTGCGAATGTAAAGGAATTAATGAGAAATACTCATTG   1275
 391   D  F  N  F  N  D  V  N  F  R  I  L  G  A  N  V  K  E  L  M  R  N  T  H  C    415

1276  CGCAAATTTTAACAATAAAAATAATCAGAAGAATAACAATCAGAAGAATAACAATCAGAACAATAACAATCAGAA   1350
 416   A  N  F  N  N  K  N  N  Q  K  N  N  N  Q  K  N  N  N  Q  N  N  N  N  Q  K    440

1351  GAATAACAATCAGAAAAATAACAATCAGAAGAATAACAATCAGAAGAATAACAATCAGAAGAATAACAATCAGAA   1425
 441   N  N  N  Q  K  N  N  N  Q  K  N  N  N  Q  K  N  N  N  Q  K  N  N  N  Q  N    465

1426  TACTAACAATTAGAATGATAATCAAGTTCGTCGTTCTTCAAAATCGCATTAAAATCAATTAGGATGTAAACCAAA   1500
 466   T  N  N  *                                                                   468

1501  TTATTTTTAAAATATTTTTTCGATGTAAACAAAATTTTTTAAAATATTTCATTATATTATAAATAAATAAAATA   1575

1576  ATATCGTTTTCGCAT  1590
```

Figure 1. The sequence of AccMRJP2 cDNA and inferred protein. The vertical arrow points to deduced signal peptidase cleavage site. Consensus N-glycosylation sites are enclosed in boxes. The underlined are the peptide sequences determined by sequencing A2 protein of *A. cerana* royal jelly (Srisuparbh et al., 2003). Shorter regularly spaced lines highlight the repetitive units of the pentapeptide repeat region. Horizontal arrows show the positions of the PCR primers used for amplification of the polymorphic pentapeptide repeat region (see Fig. 4A). Triple dots indicate the continuation of the underlined feature in the following row.

```
   1 ACGAGGGTTTCGTCAATCAGAAAATATCTATAATATCCTAGAAAAATGACAAAGTGGTTGTTGCTGGTGGTGTGT   75
   1                                                 M  T  K  W  L  L  L  V  V  C     10
  76 CTTGGTATAGCTTGTCAAGATGTGACAAGCGCAGCTGTGAACCATCAAAGAAAATCTTCAAAAAATTTGGCACAT  150
  11 L  G  I  A  C  Q  D  V  T  S  A  A  V  N  H  Q  R  K  S  S  K  N  L  A  H     35
 151 TCGATGAAGGTGATCTACGAATGGAAACATATTGATTATGATTTCGGTAGCGTTGAAAGAAGAGATGCTGCGATT  225
  36 S  M  K  V  I  Y  E  W  K  H  I  D  Y  D  F  G  S  V  E  R  R  D  A  A  I     60
 226 AAATCTGGCGAATTTGATCACACAAAAAATTACCCTTTCGATGTGGATAGATGGCGTGATAAGACATTTGTCACC  300
  61 K  S  G  E  F  D  H  T  K  N  Y  P  F  D  V  D  R  W  R  D  K  T  F  V  T     85
 301 GTAGAAAGGTTCGATGGTGTACCTTCTTCTTTGAACGTGGTAACTAATAAAAAAGGCAAAGGTGGTCCTCTTCTA  375
  86 V  E  R  F  D  G  V  P  S  S  L  N  V  V  T  N  K  K  G  K  G  G  P  L  L    110
 376 CATCCATATCCTGATTGGTCGTGGGCGAACTATAAAGATTGCTCTGGAATTGTGAGCGCTTTCAAAATTGCGGTC  450
 111 H  P  Y  P  D  W  S  W  A  N  Y  K  D  C  S  G  I  V  S  A  F  K  I  A  V    135
 451 GACAAATTCGACAGATTATGGGTTCTGGACTCAGGTCTTGTCAATAATAATCAACCCATGTGCTCTCCAAAATTG  525
 136 D  K  F  D  R  L  W  V  L  D  S  G  L  V  N  N  N  Q  P  M  C  S  P  K  L    160
 526 GTAACCTTCGATTTGAATACCTCAAAATTGCTTAAGCAAGTCGAGATACCACATAATATTGCCGTAAATGCCACC  600
 161 V  T  F  D  L [N  T  S] K  L  L  K  Q  V  E  I  P  H  N  I  A  V [N  A  T]   185
 601 ACAGGAATGGGAGAATTAGTATCACTAGCTGTTCAAGCTATAGATCCTACGAATACTATGGTGTACATAGCAGAC  675
 186 T  G  M  G  E  L  V  S  L  A  V  Q  A  I  D  P  T  N  T  M  V  Y  I  A  D    210
 676 GAAAAAGGTGAAGCTTTAATCATCTATCAAAATTCCGACGATTCCTTCCATCGATTGACTTCCAATACTTTCGAT  750
 211 E  K  G  E  A  L  I  I  Y  Q  N  S  D  D  S  F  H  R  L  T  S  N  T  F  D    235
 751 TACGATCCCAGATATACCAAACTGACAGTCGCTGGAGAAAGTTTCACAGTGAAAAATGGAATTTGTGGAATTGCA  825
 236 Y  D  P  R  Y  T  K  L  T  V  A  G  E  S  F  T  V  K  N  G  I  C  G  I  A    260
 826 CTTAGTCCCGTGACGAACAATCTTTATTACAGTCCTCTCGCTTCTCACAGTTTGTATTATGTTAACACAGAACAA  900
 261 L  S  P  V  T  N  N  L  Y  Y  S  P  L  A  S  H  S  L  Y  Y  V  N  T  E  Q    285
 901 TTCAGGAATCCACAATATGAAGAAAGTAACGTCCAATATGAAGGATCCCAAGATATTTTGAACACTCAATCATTC  975
 286 F  R  N  P  Q  Y  E  E  S  N  V  Q  Y  E  G  S  Q  D  I  L  N  T  Q  S  F    310
 976 GCTAAAGCAGTATCGAAAAATGGCGTCGTTTTCTTGGGACTCGTGAGTAATTCAGCTGTTGGCTGCGTGAATGAA 1050
 311 A  K  A  V  S  K  N  G  V  V  F  L  G  L  V  S  N  S  A  V  G  C  V  N  E    335
1051 CATCAAGTACTTCAGAAAGAAAATTTTGATGTTGTCGCTCAGAATGAAGAGACACTTCAAATGATCGTTAGTATG 1125
 336 H  Q  V  L  Q  K  E  N  F  D  V  V  A  Q  N  E  E  T  L  Q  M  I  V  S  M    360
1126 AAAATCATGCAAGATCTTCCACAATCCGGCAGAATTAATGATCCAGGAAATGAATATATGTTGGCTTTAAGTAAC 1200
 361 K  I  M  Q  D  L  P  Q  S  G  R  I  N  D  P  G  N  E  Y  M  L  A  L  S  N    385
1201 AAAATGCAAAAAATAATAAACAATGATTTTAATTTCAACGATGTAAATTTCCGAATTTTGGGTGCGAATGTAAAT 1275
 386 K  M  Q  K  I  I  N  N  D  F  N  F  N  D  V  N  F  R  I  L  G  A  N  V  N    410
1276 GATTTAACAAGAAACACTCGTTGCGCAAAATCTAATAATCAGAATGCTAACAATCAGAATGCTAATAATCAAAAT 1350
 411 D  L  T  R  N  T  R  C  A  K  S  N  N  Q  N  A  N  N  Q  N  A  N  N  Q  N    435

1351 GCTAACAATCAGAATGATAACAACCAGAATGATAATGGTAACAACAGGAGAAATGGTAACAACCAAAATGGTAAC 1425
 436 A  N  N  Q  N  D  N  N  Q  N  D  N  G  N  N  R  R  N  G  N  N  Q  N  G  N    460

1426 AGACAAAATGATAATAAACAGAATGATAACAAGCAGAATGCTAACAAGCAGAATGCTAACAAGCAAAATGATAAC 1500
 461 R  Q  N  D  N  K  Q  N  D  N  K  Q  N  A  N  K  Q  N  A  N  K  Q  N  D  N    485

1501 AAGCAAAATGGTAACAGACAAAATGATAATAGGCAGAATGATAACAAGCAAAATGATAATAGGCAGAATAATAAC 1575
 486 K  Q  N  G  N  R  Q  N  D  N  R  Q  N  D  N  K  Q  N  D  N  R  Q  N  N  N    510

1576 AAGCAAAATGGTAACAGACAAAATGATAATAGACAGAATGATAACCAGCGGAATGGTAACAGGCAAAATGATAAT 1650
 511 K  Q  N  G  N  R  Q  N  D  N  R  Q  N  D  N  Q  R  N  G  N  R  Q  N  D  N    535

1651 AGACAGAATGATAACAAGCGGAATGGTAACAGGCAAAATGATAATAGACAGAATGATAACAAGCGGAATGGTAAC 1725
 536 R  Q  N  D  N  K  R  N  G  N  R  Q  N  D  N  R  Q  N  D  N  K  R  N  G  N    560

1726 AGGCAAAATGATAACAAGCAAAATGATAACAGGCAGAATGATAACAATCAGAATGATAATCAGAATGATAATAAT 1800
 561 R  Q  N  D  N  K  Q  N  D  N  R  Q  N  D  N  N  Q  N  D  N  Q  N  D  N  N    585

1801 CGAAATAATCAAGCTCATCATTCTTAAAAATCACATTAAATCAATTAATTATCAATTAAAATCAATTAATTAGGA 1875
 586 R  N  N  Q  A  H  H  S  *                                                       594
1876 TGTAAACCAAATTATTTTTTAAAATATTTTTTCGATGTAAACAAAATTTTTTAAAATCTTTCATTATATTATAAA 1950
1951 TAAATAAAATAAATATCGTTTTCGCAT  1977
```

Figure 2. The sequence of AccMRJP3 cDNA and inferred protein. The vertical arrow points to signal peptidase cleavage site. N-glycosylation sites are boxed. The underlined is the N-terminal peptide sequence of A1 protein of *A. cerana* royal jelly (Srisuparbh et al., 2003). Shorter regularly spaced lines highlight the repetitive units of the pentapeptide repeat region. Vertical triangles show the positions of introns found in the genomic sequence.

heteroduplex DNAs (Kaiser et al., 2002) were observed in both PCR reactions. This phenomenon was more pronounced in MRJP5 products, possibly due to the different nature of the repetitive region or for other reasons that have not been investigated.

```
   1  GCACGAGGGTTCTCTTGAAACTGTCGTTTGCAAAATATTTGCAGCATCCAAGAACAATGACAAGTTGGTTGTTGC   75
   1                                                                M T S W L L L     7
  76  TGGTGGTGTGCCTTGGCATAGCTTGTCAAGGTATCACAGGCGCCACTGTTCGAGAAAATTCTTCGAGAAATTTGG  150
   8   V V C L G I A C Q G I T G A T V R E N S S R N L A   32
 151  CAAATTCGATGAACGTGATTCACGAATGGAAGTATCTTGATTATGACTTCGGTAGCGACGAAAAAAGACAAGCTG  225
  33   N S M N V I H E W K Y L D Y D F G S D E K R Q A A   57
 226  CGATTCAATCTGGCGAATATGATCATACGAAAAATTATCCCTTCGATGTCGATCGATGGCATGATATGACTTTTG  300
  58   I Q S G E Y D H T K N Y P F D V D R W H D M T F V   82
 301  TCACCGTACTAAGATACAAAGGTGTACCTTCCTCTTTAAACGTGATATCTAAGAAAATTGGCAACGGTGGACCTC  375
  83   T V L R Y K G V P S S L N V I S K K I G N G G P L  107
 376  TTCTGCAGCCATATCCTGATTGGTCGTGGGCGAACTATAAAGATTGCTCTGGAATCGTGAGCGCTTACAAAATTG  450
 108   L Q P Y P D W S W A N Y K D C S G I V S A Y K I A  132
 451  CGATCGACAAGTTCGACAGATTGTGGGTTCTGGACTCAGGTATTATCAATAATACTCAACCCATGTGTTCACCAA  525
 133   I D K F D R L W V L D S G I I N N T Q P M C S P K  157
 526  AATTGCATGTCTTTGATCTCAATACCTCACAGCAGATTAAGCAAGTTATGATGCCGCATGATATTGCCATAAATG  600
 158   L H V F D L N T S Q Q I K Q V M M P H D I A I N A  182
 601  CCACTACAGGAAAAGGAGGATTAGAAAATCTAGTTGTTCAAGCTATGGATCCTATGAATACTCTGGTGTATATGG  675
 183   T T G K G G L E N L V V Q A M D P M N T L V Y M A  207
 676  CAGATAACAAAGGTGATGCTTTAATTGTTTATCAAAATTCCGATGATTCCTTCCATCGATTGACTTCCAACACTT  750
 208   D N K G D A L I V Y Q N S D D S F H R L T S N T F  232
 751  TCGATTACGATCCCAAATATATCAAAATGATGGCCGCAGGAGAAAGTTTCACATTGCAAGATGGAATTTTTGGAA  825
 233   D Y D P K Y I K M M A A G E S F T L Q D G I F G M  257
 826  TGGCACTCAGTCCCATGACAAACAATCTTTATTACAGTCCTCTCGCTTCTCGCAGTTTGTATTATATTAATACGA  900
 258   A L S P M T N N L Y Y S P L A S R S L Y Y I N T K  282
 901  AACCCTTCATGAAATCACAATATGGAACAAATAACGTACAACATGAAGGTGTTCAAGATATTTTCAATACTCAAT  975
 283   P F M K S Q Y G T N N V Q H E G V Q D I F N T Q S  307
 976  CAATTGCTAAAATAATGTCGAAAAATGGCGTTCTCTTTTTCGGTCTCATGAATAATTCAGCTATTGGTTGTTGGA 1050
 308   I A K I M S K N G V L F F G L M N N S A I G C W N  332
1051  ATGAGCACCAACCACTTCAGAGACAAAATATGGATATGGTCGCTCAGAATGAAGAGACACTTCAAACGGTCGTTG 1125
 333   E H Q P L Q R Q N M D M V A Q N E E T L Q T V V A  357
1126  CTATGAAAATGATGCATCTCCCACAATCCAACAGGATGAATAGGATGCATAAGATGAATAGAGTGAATAGTATGA 1200
 358   M K M M H L P Q S N R M N R M H K M N R V N S M N  382
1201  ATAGAATGGATAGAATGGATAGGATGGATAAGATGGATAGGATGGATAGGATGGATAGGATGGATAGAATAGATG 1275
 383   R M D R M D R M D K M D R M D R M D R M D R I D G  407
1276  GGATGGATAGGATGGATAGGATGGATAGGATGGATAGAATGCATACAATGGATACAATGTATAGAATGGATAGAA 1350
 408   M D R M D R M D R M D R M H T M D T M Y R M D R I  432
1351  TAGATAGGATGGATAGGATGGATATAATGGATAGGACGAATAAAATGGATAGGATGGATAGGATGGATATAATGG 1425
 433   D R M D R M D I M D R T N K M D R M D R M D I M D  457
1426  ATAAGATGAATAAAATGGATAGGATGGATAGTATGATTAGAATAGATAAAATGGATAGAATGGATAGAATGGATA 1500
 458   K M N K M D R M D S M I R I D K M D R M D R M D R  482
1501  GAATAGATATAATGAATAGAATGGATAGAATGGATAGAATGGACACAATGGATAGAATAGATACAATGGATAGAA 1575
 483   I D I M N R M D R M D R M D T M D R I D T M D R M  507
1576  TGGACAGAATGGATAGAATGGATAAAATGGATAAGATAAATAAGATGCATAGGATGGGTAGGATGGATAGGATGG 1650
 508   D R M D R M D K M D K I N K M H R M G R M D R M D  532
1651  ATAGAATGAATAGAATGAATAGACAAATGAATGAATATATGATGGCTTTAAGTATGAAATTACAGAAATTTATAA 1725
 533   R M N R M N R Q M N E Y M M A L S M K L Q K F I N  557
1726  ACAATGATTATAATTTCAACGAAGTAAATTTCCGAATTTTGGCTGCAAATGTAAACGATTTAATAATGAACACTC 1800
 558   N D Y N F N E V N F R I L A A N V N D L I M N T R  582
1801  GTTGTGCAAATTCTAACAATCAGAATGATAATCAAAATAAGCATAATAATTAAGGTAATCGTTCTTTATATTAAA 1875
 583   C A N S N N Q N D N Q N K H N N *                  598
1876  ATCTGTTAATTAGTCTTTTCTCGACTATAAACCAAATATTGTTTCAAATTTCTTTATATTATAAATGAATAAAAT 1950
1951  AAATATCGTTTTTGCATGAT  1970
```

Figure 3. The sequence of AccMRJP5 cDNA and inferred protein. The vertical arrow points to putative signal peptidase cleavage site. Consensus N-glycosylation sites are enclosed in squares. Characteristic feature of MRJP5, the tripeptide repetitive region is underlined. Horizontal arrows indicate the positions of the PCR primers used for amplification of the polymorphic tripeptide repeat (Fig. 4B). Vertical triangles show the positions of introns in the genomic sequence.

3.5. Differences among MRJPs of *Apis cerana* and *Apis mellifera*

At the beginning of this project, no proteins of RJ or genes encoding them of honeybees other than *A. mellifera* were known. Therefore we sought cDNAs encoding MRJPs in another economically important honeybee species, *Apis cerana*.

Only 6 nucleotides and 2 amino acid residues differed between MRJP1 cDNAs of *A. cerana cerana* and *A. cerana indica* published in the meantime (Srisuparbh et al., 2003). AccMRJP1 showed high homology to AmMRJP1 at both nucleotide (93.8%) and protein (90.5%) level. Taken together, MRJP1, the most abundant protein of RJ, was highly conserved among honeybee species and subspecies.

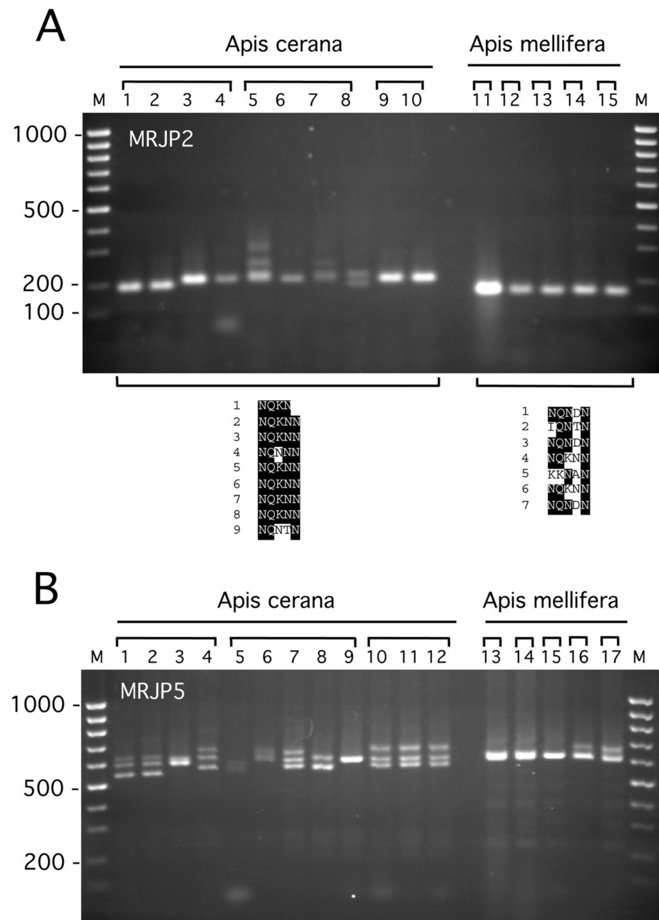

Figure 4. Studies of MRJP2 and MRJP5 polymorphism. (A) Polymorphism of the repetitive region of MRJP2. Primers flanking the repetitive region of MRJP2 (see Fig. 1) were used in PCR with genomic DNAs of individual honeybees from three colonies (horizontal square brackets) and separated in 2% agarose gel. With *A. cerana* genomic DNAs, alleles of different lengths were amplified among bees from the same colony. No polymorphism of the MRJP2 gene was observed among *A. mellifera* individuals originating from different colonies of two races (*A. mellifera carnica* (lanes 11–13) and *A. mellifera mellifera* (lanes 14, 15). The lower part depicts aligned consecutive units of one representative MRJP2 repetitive region of *A. cerana* (left) and *A. mellifera* (right). Preserved amino acids are shaded black. A correlation can be observed between repetitive unit conservancy and degree of polymorphism. (B) Polymorphism of the DRM tripeptide repeat region of MRJP5 genes. Genomic DNAs (the same as above) were used in PCR to amplify the DRM repeat of MRJP5. PCR products were resolved in 1.5% agarose. The size polymorphism seemed to be more pronounced in *A. cerana* than in *A. mellifera.* Bands of intermediate size might represent heteroduplex DNA products, are more common in the MRJP5 repeat, probably due to intrinsic properties of the repetitive region.

MRJP2 cDNAs were also nearly identical between *A. cerana cerana* and *A. cerana indica* (GenBank entry **AF525777**) except for the repetitive region (see above). The AccMRJP2 was also highly homologous with AmMRJP2 at both the nucleotide (89.2%) and protein (84.5%) levels.

The nearest homologue of AccMRJP3 was MRJP3 of *A. mellifera*. AmMRJP3 was shorter than AccMRJP3 due to the different length of

the repetitive region. As described for AmMRJP3 (Albert et al., 1999b), the repeat of AcMRJP3 was also highly polymorphic among individuals of *A. cerana* and other honeybees. (Albertova et al., unpublished data).

AciMRJP4 (Acc. Number **AY532368**) showed homology to its *A. mellifera* counterpart (90.0% and 80.0%). The most obvious difference between AciMRJP4 and AmMRJP4 was the length of encoded proteins. AciMRJP4 comprised 485 residues while AmMRJP4 only 464 residues. Again, the different length of the pentapeptide repetitive region between species caused the size difference.

The sequence identity of nucleotides and amino acid residues of MRJP5 was 99.3% and 97.2% between *A. cerana cerana* (Acc. Number **AY392757**) and *A. cerana indica* (Acc. Number **AY532369**). AccMRJP5 was homologous to MRJP5 of *A. mellifera* at nucleotide (90.7%) and protein level (83.6%), respectively. The DRM repetitive region (amino acids: 420–530) was highly variable between honeybee species and alleles of different size were also found among individuals (see Fig. 4B).

A phylogenetic tree was calculated from the aligned MRJP protein sequences of *A. cerana* (Fig. 5) and *A. mellifera* by neighbor-joining method using exhaustive maximum parsimony search with distantly related Yellow-f protein of *A. mellifera* (Albert and Klaudiny, 2004) as an outgroup (not shown). Each of the AcMRJPs formed a monophyletic group with its *A. mellifera* homologue. High bootstrap values supported the groups formed by *A. mellifera*-*A. cerana* pairs, further supporting the notion that the MRJPs of *A. cerana* isolated here were orthologs of the corresponding *A. mellifera* proteins.

4. DISCUSSION

4.1. The newly isolated cDNAs code for proteins of *A. cerana* royal jelly

Using PCR approach with primers priming to conserved regions of A. *mellifera* MRJPs, we have cloned and characterized fragments of four MRJP genes of *A. cerana*. Using one of the four genomic fragments (MRJP3) as a hybridization probe, we isolated complete cDNAs of AccMRJP1, 2, 3 and 5. On the other hand, screening of 120 positive clones was not sufficient for isolation of less abundant MRJP cDNAs, such as MRJP4, 6–8 (Albert and Klaudiny, 2004) from the cDNA library. Several lines of evidence, such as features of encoded proteins, homology with AmMRJPs, relative abundance of the isolated cDNAs in the cDNA library, and finally the presence of the peptide sequences proteins of *A. cerana* RJ (Srisuparbh et al., 2003) supported the view that the cDNAs presented here code for protein components of *A. cerana* RJ.

4.2. Are MRJP genes a heaven for polymorphic repetitive segments?

Phylogenetic analysis suggested that MRJP proteins evolved by subsequent but nearly simultaneous duplications (Albert et al., 1999a). Several MRJPs contained a pentapeptide repetitive region or its leftovers, which is extremely extended and highly polymorphic in MRJP3 (Albert et al., 1999b; Albert and Schmitz, 2002). Here we show that the same repetitive region is highly polymorphic in AcMRJP2. It was argued that in other MRJPs, the repeat units accumulated mutations which disabled their further rearrangements by a slipped strand mechanism. Three of five MRJP genes/proteins of *A. cerana* characterized so far (AcMRJP2, AcMRJP3, AcMRJP5) exhibit a repeat length polymorphism (Fig. 4 and unpublished data). The distribution of polymorphic alleles seems to be phenotypically neutral as there was no repeat length bias found in the population. Thus, the family of MRJP genes alone provides three midisatellite polymorphic loci of VNTR (variable number of tandem repeat) type, which are suitable for genotyping *A. cerana* individuals (Beye et al., 1998).

Of the remaining two known MRJPs of *A. cerana*, AcMRJP1 does not harbor any repetitive region. In line with this observation, the two AcMRJP1 cDNAs originating from two different *A. cerana* races (this work and Srisuparbh et al., 2003) are nearly identical and highly homologous with AmMRJP1. Another member of the family, AcMRJP4, although not tested yet, might also be polymorphic. At least in comparison to AmMRJP4, there is an apparent segment duplication seen in the repetitive region of AcMRJP4.

```
AccMRJP1     1 MTKWLFMVVCLGIVCQGTTS--SILRGES---LNKSISVIHEWKEFDYDFDSDERRQDAI
AccMRJP5     1 MTSWLLLVVCLGIACQGITG--ATVRENSSRNLANSMNVIHEWKYLDYDFGSDEKRQAAI
AccMRJP2     1 MTKWLFMVACLGIACQG----ATIRQ-NSAKNLENSINVIHEWKYIDYDFGSEERRQAAI
AccMRJP3     1 MTKWLLLVVCLGIACQDVTSAAVNHQRKSSKNLAHSMKVIYEWKHIDYDFGSVERRDAAI

AccMRJP1    56 LSGEYDYRKNYPSDVDQWHGKIFVTMLRYNGVPSSLNVISKKIGDGGPLLQPYPDWSEAK
AccMRJP5    59 QSGEYDHTKNYPFDVDRWHDMTFVTVLRYKGVPSSLNVISKKIGNGGPLLQPYPDWSWAN
AccMRJP2    56 QSGEYDHTKNYPFDVDQWHDKTFVTILKYDGVPSTLNMISNKIGKGGRLLQPYPDWSWAE
AccMRJP3    61 KSGEEDHTKNYPFDVDRWRDKTFVTVERFDGVPSSLNVVTNKKGKGGPLIHPYPDWSWAN

AccMRJP1   116 YDDCSGIVSATKLAIDKCDRLWVLDSGLVNNTQPMCSPKLLTFDLTTSQLLKQVEIPHDV
AccMRJP5   119 YKDCSGIVSAYKIAIDKFDRLWVLDSGIINNTQPMCSPKLHVFDLNTSQQIKQVMMPHDI
AccMRJP2   116 NKDCSGIVSAFKIAIDKFDRLWVLDSGLINRTEPICAPKLHVFDLKNTKHLKQIEIPHDI
AccMRJP3   121 YKDCSGIVSAFKIAVDKFDRLWVLDSGLVNNNQPMCSPKIVTFDLNTSKLLKQVEIPHNI

AccMRJP1   176 AVNATTGKGRLSSLAVQPLDCNINGDTMVYIADEKGEGLIVYHDSDYSFHRLTSKTFDYD
AccMRJP5   179 AINATTGKGGLENLVVQAMDP-MN--TLVYMADNKGDALIVYQNSDDSFHRLTSNTFDYD
AccMRJP2   176 AVNATTGKGGLVSLVVQAMDP-MN--TLVYIADHKGDALIVYQNSDDSFHRMTSNTFDYD
AccMRJP3   181 AVNATTGMGELVSLAVQAIDP-TN--TMVYIADEKGEALIIYQNSDDSFHRLTSNTFDYD

AccMRJP1   236 PKETKMTINGESFTTQNGISGMALSPMTNNLYYSPVASTSLYYVNTEQFRTSNYEQNA-V
AccMRJP5   236 PKYIKMMAAGESFTLQDGIFGMALSPMTNNLYYSPLASRSLYYINTKPFMKSQYGTNN-V
AccMRJP2   233 PRYAKMTINGESFTLKNGICGMALSPVTNNLYYSPLASHGLYYVNTEPFMKSQFGDNNNV
AccMRJP3   238 PRYTKLTVAGESFTVKNGICGIALSPVTNNLYYSPLASHSLYYVNTEQFRNPQYEESN-V

AccMRJP1   295 HYEGVQNILDTQSSAKVVSKSGVLFFGLVGDSALGCWNEHRSLERHNIRTVAQSDETLQM
AccMRJP5   295 QHEGVQDIFNTQSIAKIMSKNGVLFFGLMNNSAIGCWNEHQPLQRQNMDMVAQNEETLQT
AccMRJP2   293 QYEGSQDTLNTQSIAKAVSKDGVLFVGLVGNSALGCLNEHQPLQRENLEIVAQNEKTLQM
AccMRJP3   297 QYEGSQDILNTQSFAKAVSKNGVVFLGLVSNSAVGCVNEHQVLQKENFDVVAQNEETLQM

AccMRJP1   355 IVGMKIK-----------------------------------------------------
AccMRJP5   355 VVAMKMMHLPQSNRMNRMHKMNRVNSMNRMDRMDRMDKMDRMDRMDRMDRIDGMDRMDRM
AccMRJP2   353 IAGMKIK-----------------------------------------------------
AccMRJP3   357 IVSMKIM-----------------------------------------------------

AccMRJP1   362 ------------------EALPHVPIFDR-------------------------------
AccMRJP5   415 DRMDRMHTMDTMYRMDRIDRMDRMDIMDRTNKMDRMDRMDIMDKMNKMDRMDSMIRIDKM
AccMRJP2   360 ------------------EELPHFVGSNKPVK----------------------------
AccMRJP3   364 ------------------QELPQSGRINDPG-----------------------------

AccMRJP1   373 ------------------------------------------------------------
AccMRJP5   475 DRMDRMDRIDIMNRMDRMDRMDTMDRIDTMDRMDRMDRMDKMDKINKMHRMGRMDRMDRM
AccMRJP2   374 ------------------------------------------------------------
AccMRJP3   377 ------------------------------------------------------------

AccMRJP1   373 -YINR---EYIILVLSNRMQKMANNDYNFNDVNFRIMDANVNDLILNTRCENPNNDNTPFK
AccMRJP5   535 NRVNRQMNEYMMALSMKLQKFINNDYNFNEVNFRILAANVNDLIMNTRCANSNNQNDNQN
AccMRJP2   374 -------DEYMLVLSNKMQKIVNNDFNFNDVNFRILGANVKELMRNTHCANFNNK--NNQ
AccMRJP3   377 -------NEYMLALSNKMQKIINNDFNFNDVNFRILGANVNDITRNTRCAKSNNQNANNQ

AccMRJP1   429 ISIHL-------------------------------------------------------
AccMRJP5   595 KHNN--------------------------------------------------------
AccMRJP2   425 KNNNQKNNNQNNNNQKNN------------------------------------------
AccMRJP3   430 NANNQNANNQNDNNQNDNGNNRRNGNNQNGNRQNDNKQNDNKQNANKQNANKQNDNKQNG

AccMRJP1       ------------------------------------------------------------
AccMRJP5       ------------------------------------------------------------
AccMRJP2   443 ------------------------------------------------------------
AccMRJP3   490 NRQNDNRQNDNKQNDNRQNNNKQNGNRQNDNRQNDNQRNGNRQNDNRQNDNKRNGNRQND

AccMRJP1       --------------------------------------------
AccMRJP5       --------------------------------------------
AccMRJP2   443 ----------NQKNNNQKNNNQKNNNQKNNNQNTNN--------
AccMRJP3   550 NRQNDNKRNGNRQNDNKQNDNRQNDNNQNDNQNDNNRNNQAHHS
```

Figure 5. Alignment of AccMRJP proteins. The sequences were aligned using CLUSTAL W and visualized using BOXSHADE. Black-shaded residues are those identical between at least two proteins, grey-shaded indicate the conservative substitutions.

The advantage of using MRJP polymorphism for genotyping lies in easy detection of the polymorphic alleles by PCR and subsequent electrophoresis in standard agarose gels (Fig. 4). In addition, the polymorphism of MRJP genes affects the molecular mass of encoded proteins, although the resolution at the protein level is lower (Albert et al., 1999b). Therefore a precise analysis of *A. cerana* RJ proteins by 2D-electrophoresis, as done with *A. mellifera* RJ (Sano et al., 2004), would provide further identification of the polymorphism of its protein components.

4.3. Different fate of the MRJP2 repetitive region in *A. cerana*

An interesting phenomenon revealed by our studies of MRJP2 polymorphism is that the MRJP2 is highly polymorphic in *A. cerana* but not in *A. mellifera* (Fig. 4A). For further support of the AmMRJP2 monomorphism we looked at the MRJP2 cDNAs in the normalized honeybee brain EST library (prepared from 400 brains of another honeybee race, *A. mellifera ligustica*) from another continent (Whitfield et al., 2002). MRJP2 with 44 contigs is the most abundant cDNA in this library. Nine independent sequence reads covered the repetitive region of MRJP2, but no repeat length or sequence variants could be found among them. Finally, recent assembly of *A. mellifera* genome (Amel 1.2) contains a single MRJP2 locus (GroupUn 1795) with the same sequence of the repeat region; this is additional evidence for the absence of MRJP2 polymorphism in *A. mellifera.*

It has been reported that the length polymorphism of repetitive regions depends on the equilibrium between repeat rearrangements rate caused by slipped-strand mispairing during replication and/or crossing-over (promoting the uniformity of the recurring units) and local mutation rate (disrupting their uniformity). It is improbable that the units with altered sequences form the slipped-strand duplexes, which seem to be essential for repeat expansions/contractions (Levinson and Gutman, 1987).

The different fate of the MRJP2 repeat is an interesting example of a dual outcome of the same process in two closely related species. Apparently, the equilibrium in repeat evolution inclined towards repeat rearrangements in *A. cerana* and towards length fixation and accumulation of mutations in *A. mellifera*. In line with the above statement, the basic units of the MRJP2 pentapeptide repeat region differ from each other in *A. mellifera* but are identical in *A. cerana* (Fig. 4A, alignments beneath the gel picture, black-shaded regions).

ACKNOWLEDGEMENTS

We thank Prof. Dr. Gao Qikang, Institute of Biotechnology, Zhejiang University for his help on technology. We also thank Professor Miao Yungen for critical reading of the manuscript. This project was supported by National Natural Science Foundation of China (No. 30200206) and Zhejiang Provincial Natural Science Foundation of China (No. 302113).

Résumé – Clonage moléculaire et analyse de quatre ADNc issus de têtes de nourrices d'*Apis cerana cerana* codant pour les principales protéines de la gelée royale. La gelée royale (GR) est un composant crucial de la nutrition de l'abeille (*Apis* sp.). Elle est synthétisée par les glandes hypopharyngiennes des nourrices et cette sécrétion sert à nourrir la reine et les larves. Les principales protéines de la GR sont très proches les unes des autres et éloignées des protéines trouvées chez les autres insectes qui sont impliquées dans la pigmentation de la cuticule et dans d'autres processus physiologiques. Bien que les principales protéines de la gelée royale (MRJP) de l'abeille domestique (*Apis mellifera*) soient bien caractérisées sur le plan génétique et de leurs séquences protéiniques, on sait peu de choses concernant les MRJP des autres espèces du genre *Apis*. Nous avons caractérisé les ADN complémentaires (ADNc) et des portions des séquences génomiques qui codent pour les protéines de la GR d'*Apis cerana cerana* par le clonage, le séquençage et la PCR. Une bibliothèque d'ADNc a été construite à partir de nourrices d'*A. c. cerana* âgées de 8 j. Elle a été testée par hybridation à l'aide de protocoles standard, ce qui a permis d'identifier quatre ADNc codant pour les protéines d'*A. c. cerana.* Ces protéines ont été nommées AccMRJP1, AccMRJP2 (Fig. 1), AccMRJP3 (Fig. 2) et AccMRJP5 (Fig. 3), en fonction de leur ressemblance avec les protéines de la GR d'*A. mellifera*. Les protéines de la GR d'*A. cerana* caractérisées ici présentent une forte ressemblance avec les protéines respectives d'*A. mellifera*, y compris la présence de séquences répétitives chez certaines d'entre elles (Fig. 5). La longueur des séquences répétitives diffère selon l'espèce, mais aussi entre individus d'*A. cerana* en raison du nombre d'unités de répétition (Fig. 4). Nous avons pu montrer par la PCR et le séquençage de l'ADN que la région répétitive de MRJP2 avait évolué différemment après la scission des espèces *mellifera* et

cerana. Ce fait se reflète dans le degré de conservation des motifs répétitifs individuels (Fig. 4). En résumé, les protéines de la GR de l'abeille asiatique, *A. cerana*, présentent une forte homologie avec celle de l'abeille européenne, *A. mellifera*, mais également des différences frappantes.

***Apis cerana cerana* / *Apis mellifera* / gelée royale / protéine / polymorphisme / midisatellite / VNTR**

Zusammenfassung – Klonierung und molekulare Analyse von vier für Gelée Royale-Proteine kodierende cDNAs aus Köpfen von *Apis cerana cerana* Ammenbienen. Das auch als Bienenmilch bezeichnete Gelée Royale ist eine kritische Komponente in der Ernährung von Honigbienen. Es wird in den Hypopharynxdrüsen von Ammenbienen produziert und als Sekret an die Königin und an Larven verfüttert. Proteine sind die Hauptbestandteile von Gelée Royale, und seine Hauptproteinkomponenten zeigen untereinander große molekulare Ähnlichkeit. Etwas weiter verwandt sind diese mit Proteinen, die bei anderen Insekten in der Pigmentierung der Cuticula und in anderen physiologischen Prozessen involviert sind. Die Hauptproteinkomponenten von Gelee royale (MRJPs) der europäischen Honigbiene, *Apis mellifera*, sind sowohl genetisch als auch in ihren Proteinsequenzen gut charakterisiert. Die Gelée Royale-Proteine anderer Honigbienenarten sind hingegen vergleichsweise wenig untersucht. In der vorliegenden Arbeit charakterisieren wir mittels Klonierung, Sequenzierung und Polymerasekettenreaktion die komplementären DNA-Sequenzen (cDNAs) und Stücke der genomischen Sequenzen, die für die Gelée Royale-Proteine von *Apis cerana cerana* kodieren. Eine cDNA-Bibliothek wurde aus RNA-Extrakten 8-Tage-alter Ammenbienen von *A. cerana cerana* erstellt und mittels klonarer Hybridisierung getestet. Dies führte zur Identifizierung von vier cDNAs, die für Futtersaftproteine von *A. cerana* kodieren. Basierend auf ihrer jeweiligen Sequenzähnlichhkeit mit Futtersaftproteinen von *A. mellifera* erhielten diese die Bezeichnungen AccMRJP1, AccMRJP2 (Abb. 1) AccMRJP3 (Abb. 2) und AccMRJP5 (Abb. 3). Die insgesamt hohe Ähnlichkeit der *A. cerana* Proteine mit den entsprechenden Gelée Royale-Proteinen von *A. mellifera* erstreckte sich auch auf die repetitiven Sequenzen in einigen dieser Proteine (Abb. 5). Die Länge der repetitiven Sequenzen wies nicht nur artspezifische Unterschiede auf, sondern variierte auch zwischen Individuen von *A. cerana* aufgrund unterschiedlicher Kopiezahlen der repetitiven Einheiten (Abb. 4). Mittels Polymerasekettenreaktion und DNA-Sequenzierung konnten wir zeigen, dass die repetitive Region von MRJP2 sich nach der Artaufspaltung von *A.mellifera* und *A. cerana* evolutiv unterschiedlich entwickelte. Dies spiegelt sich im jeweiligen Grad der Konservierung der einzelnen repetitiven Motive wider (Abb. 4). Zusammengefasst lässt sich sagen, dass sich die Proteine des Königinnenfuttersafts der asiatischen Honigbiene *A. cerana* trotz des hohen Grads an Sequenzhomologie doch deutlich von den entsprechenden Proteinen der europäischen Honigbiene unterscheiden.

Apis cerana cerana* / Gelée Royale Proteine / Polymorphismus / VNTR / Midisatellit / *Apis mellifera

REFERENCES

Albert S., Schmitz J. (2002) Characterization of major royal jelly protein-like DNA sequences in *Apis dorsata*, J. Apic. Res. 41, 75–82.

Albert S., Klaudiny J. (2004) The MRJP/YELLOW protein family of *Apis mellifera*: identification of new members in the EST library, J. Insect Physiol. 50, 51–59.

Albert S., Klaudiny J., Simuth J. (1999b) Molecular characterization of MRJP3, highly polymorphic protein of honeybee (*Apis mellifera*) royal jelly, Insect Biochem. Mol. Biol. 29, 427–434.

Albert S., Bhattacharya D., Klaudiny J., Schmitzova J., Simuth J. (1999a) The family of major royal jelly proteins and its evolution, J. Mol. Evol. 49, 290–297.

Beye M., Neumann P., Schmitzova J., Klaudiny J., Albert S., Simuth J., Felder M., Moritz R.F.A. (1998) A simple non-radioactive DNA fingerprinting method for identification of patrilines in honeybee colonies, Apidologie 29, 255–263.

Chen S.L., Su S.K., Lin X.Z. (2002) An introduction to high-yielding royal jelly production methods in China, Bee World 83, 69–77.

Kaiser R., Tremblay P.B., Roots I., Brockmoller J. (2002) Validity of PCR with emphasis on variable number of tandem repeat analysis, Clin. Biochem. 35, 49–56.

Kamakura M., Mitani N., Fukuda T., Fukushima M. (2001a) Antifatigue effect of fresh royal jelly in mice, J. Nutr. Sci. Vitaminol. (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, Biochem. Biophys. Res. Commun. 282, 865–874.

Klaudiny J., Hanes J., Kulifajova J., Albert S., Simuth J. (1994) Molecular cloning of two cDNAs from the head of the nurse honey bee (*Apis mellifera* L.) for coding related proteins of royal jelly, J. Apic. Res. 33, 105–111.

Knecht D., Kaatz H.H. (1990) Patterns of larval food production by hypopharyngeal glands in adult worker honey bee, Apidologie 21, 457–468.

Lensky Y., Rakover Y. (1983) Separate body compartments of the worker honey bee (*Apis mellifera* L.), Comp. Biochem. Physiol. B 75, 607–615.

Levinson G., Gutman G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution, Mol. Biol. Evol. 4, 203–221.

Moritz R.F.A., Southwick E.E. (1992) Bees as superorganism, An evolutionary reality, Springer Verlag, Berlin, Heidelberg.

Nagai T., Inoue R. (2004) Preparation and functional properties of water extract and alkaline extract of royal jelly, Food Chem. 84, 181–186.

Okamoto I., Taniguchi Y., Kunikata T., Kohno K., Iwaki K., Ikeda M., Kurimoto M. (2003) Major royal jelly protein 3 modulates immune responses in vitro and in vivo, Life Sci. 73, 2029–2045.

Rembold H. (1987) Die Kastenbildung bei der Honigbiene, *Apis mellifica* L., aus biochemischer Sicht, Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart.

Sano O., Kunikata T., Kohno K., Iwaki K., Ikeda M., Kurimoto M. (2004) Characterization of Royal Jelly Proteins in both Africanized and European Honeybees (*Apis mellifera*) by Two-Dimensional Gel Electrophoresis, J. Agric. Food Chem. 52, 15–20.

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., Cell. Mol. Life Sci. 54, 1020–1030.

Srisuparbh D., Klinbunga S., Wongsiri S., Sittipraneed S. (2003) Isolation and characterization of major royal jelly cDNAs and proteins of the honey bee (*Apis cerana*), J. Biochem. Mol. Biol. 36, 572–579.

Takenaka T., Takenaka Y. (1996) Royal jelly from *Apis cerana japonica* and *Apis mellifera*, Biosci. Biotechnol. Biochem. 60, 518–520.

Watanabe K., Shinmoto H., Kobori M., Tsushida T., Shinokara K., Kanaeda J., Yonekura M. (1996) Growth stimulation with honeybee royal jelly DIII protein of lymphocytic cell lines in a serum-free medium, Biotechnol. Tech. 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 Res. 12, 555–566.

To access this journal online:
www.edpsciences.org