Characterization of major royal jelly protein-like DNA sequences in *Apis dorsata*

Štefan Albert*[1]; Jürgen Schmitz[2]

[1]University of Würzburg, Institute of Medical Radiation and Cell Research (MSZ), Versbacherstr. 5, 97078 Würzburg, Germany

[2]Primate Genetics, German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany

(Received 28 May 2002, accepted subject to revision 4 September 2002, accepted for publication 25 September 2002)

SUMMARY

Major proteins of royal jelly (MRJP) are well characterized in the honey bee, *Apis mellifera*. They are secreted by hypopharyngeal glands and provide major proteinaceous components of the larval food. To date nothing is known about royal jelly proteins in other bee species. To fill this gap, we screened the genome of the Asian giant honey bee, *Apis dorsata*, for the two MRJP protein coding sequences MRJP3 and MRJP5. With PCR primers originally designed for *A. mellifera* we were able to amplify the presumably orthologous sequences in *A. dorsata*. Sequence analyses revealed high similarity in the two honey bees. As described in *A. mellifera* we found an orthologous extensive repetitive region in the MRJP3 sequence of *A. dorsata*. The number of reiterated repetitive motifs is highly variable in different *A. dorsata* individuals and is therefore well suited as a marker in population studies. The MRJP5 sequence region contains another repetitive region at the orthologous position as described for *A. mellifera*. These findings confirm the presence and indicate the functional equivalence of MRJP proteins in *A. dorsata* compared to *A. mellifera*. The origin of the MRJP protein family can therefore be dated back probably 22 million years ago to a common ancestor of *A. mellifera* and *A. dorsata*.

Keywords: *Apis dorsata*, *Apis mellifera*, honey bees, royal jelly, proteins, genes, alleles, tandem repeats, polymorphism, genotyping

*Corresponding author: Stefan.Albert@mail.uni-wuerzburg.de

INTRODUCTION

Royal jelly, a milky secretion of the hypopharyngeal glands of nurse honey bees, plays a central role in honey bee queen feeding and development (Rembold & Hanser, 1964). Proteins constitute about 50% of the dry weight of royal jelly. Major proteins of royal jelly (MRJP) were characterized in *A. mellifera* (Albert *et al.*, 1999b; Hanes & Simuth, 1992; Klaudiny *et al.*, 1994; Schmitzova *et al.*, 1998). They form a family of highly related proteins and show similarity to the *Drosophila* yellow protein and its homologues (Albert *et al.*, 1999a). Interestingly, while the yellow protein of *Drosophila* is involved in larval pigmentation (Geyer *et al.*, 1986; Walter *et al.*, 1991), MRJP proteins of honey bees form major components of larval food. It has been proposed that the MRJPs acquired a novel nutritive function in the course of evolution when becoming secreted by hypopharyngeal glands (Albert *et al.*, 1999a). This is supported by the fact that additional members of the MRJP/yellow protein family were found in salivary secretions of the sand flies *Lutzomyia longipalpis* (Charlab *et al.*, 1999) and *Phlebotomus papatasi* (Valenzuela *et al.*, 2001). Their function however remains unknown, and very little is known about the function and expression of 13 additional homologues of yellow genes found in the genome of *Drosophila melanogaster* (Drapeau, 2001). It cannot be ruled out that some of the yellow-like proteins are secreted by salivary glands.

Two of the five MRJPs characterized (MRJP3 and MRJP5) contain extensive but mutually unrelated repetitive regions, which were hypothesized to be connected with the nutritive function as storage of biologically useable nitrogen-rich amino acids (Albert *et al.*, 1999a, 1999b). The repeat region of the honey bee MRJP3 protein is highly polymorphic with as many as five alleles found in 10 individuals from the same colony (Albert *et al.*, 1999b). Its alleles differ in the number of basic repetitive units and can be distinguished easily by agarose gel electrophoresis of PCR products and/or by SDS-PAGE of proteins. On the other hand, the repetitive region of MRJP5 is located in a different part of the protein sequence and differs in length, composition, and repetitive unit size (Albert *et al.*, 1999a). The consensus sequence represents a tripeptide motif (aspartate-arginine-methionine; DRM in single-letter code).

Apis dorsata, the giant Asian honey bee, is a relative of *A. mellifera*. However, it is comparatively less well characterized than its European, cavity nesting, domesticated relatives due to its open nest building, seasonal migration of colonies and aggressive behaviour. Genetic studies done with these eusocial bees revealed a high degree of polyandry (Moritz *et al.*, 1995; Oldroyd *et al.*, 1996).

In this article we identify homologues of MRJP3 and MRJP5 genes in *A. dorsata* termed adMRJP3 and adMRJP5, respectively. We describe the polymorphism of the adMRJP3 repetitive region and provide unequivocal evidence for the existence of the MRJP gene family in *A. dorsata*.

MATERIALS AND METHODS

Biological samples

Two *A. dorsata* worker bees, originating from Borneo, were provided by Per Kryger (University of Pretoria, S Africa). Three further samples were obtained from Samui Island, Thailand, provided by Duangporn Sihanuntavong, Chulalongkorn University, Thailand.

Preparation of genomic DNA

Genomic DNA was isolated from individual thoraces by phenol/chloroform extraction according to Fondrk *et al.* (1993) or on QIAQUICK columns (QIAGEN) according to the manufacturer's instructions.

PCR amplifications

PCR reactions were performed with genomic DNA using the following primers. For the amplification of the repetitive region of adMRJP3 (Albert *et al.*, 1999b):

- primer #1: 5'-ATG TAA TTT TGA AGA ATG ATG AAC TTG-3'
- primer #2: 5'-TGT AGA TGA CTT AAT GAG AAA CAC-3'

For the amplification of the 280 bp region immediately upstream to the repeat of the adMRJP3:

- primer #3: 5'-GGC GTC CTT TTC TTG GGA C-3'
- primer #4: 5'-CAT TCG CAC CCA AAA TTC GG-3'

For the amplification of the DRM repeat region of adMRJP5:

- primer #5: 5'-AGA CTC TTC AAA CGG TCG TTG C-3'
- primer #6: 5'-CTG TAA TTT CAT ACT TAA AGC CAT C3'

Standard PCR analyses were performed in a 50 µl volume under the following conditions: an initial 2 min denaturation and subsequent addition of DNA polymerase (AmpliTaq, Perkin Elmer) was followed by 30 cycles of 30 sec at 94°C, 30 sec at 54°C and 1 min at 72°C. The reaction was completed with 10 min at 72°C. PCR products were size-separated by agarose gel electrophoresis (Small DNA or Metaphor Agarose, FMC Bioproducts). For fine resolution of MRJP3 alleles of *A. dorsata*, 20 µl PCR aliquots were extracted (QIAGEN) mixed with 5 units of *Bst*EII (NEB) in appropriate buffer, incubated at 60°C for 1 hour and electrophoresed as described above.

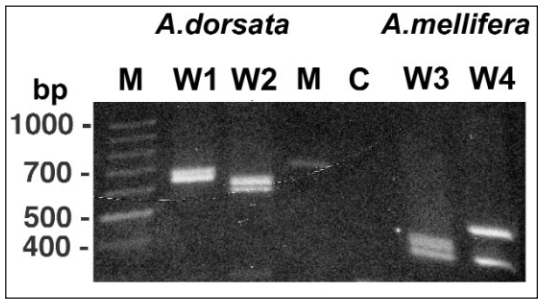

FIG. 1. Products of amplification of the MRJP3 repetitive region. Genomic DNAs of two *Apis dorsata* workers (lanes W1 and W2), a control reaction without DNA (lane C) and two *A. mellifera* workers (lanes W3 and W4) were used as templates for PCR to amplify the repetitive region of MRJP3. Reaction products were separated by electrophoresis on 2% high-resolution agarose (Metaphor) at high voltage (8V/cm) to improve resolution. Lanes marked M contain a 100 bp ladder and a lambda/*BstEII* molecular weight marker respectively.

Cloning and sequencing of MRJP sequences

PCR products of the two Borneo bees were cloned into the pCR2.1 vector (INVITROGEN) for further characterization by DNA sequencing of individual alleles. One PCR product originating from the Samui bees was also cloned and characterized. Plasmid inserts were sequenced by the cycle sequencing method (Prism Ready Reaction Dyedeoxy Terminator kit, Perkin Elmer) on an ABI 373A sequencing device with automatic data collection. DNA sequence analyses were done using the GCG program package (Devereux *et al.*, 1984) and the National Center for Biotechnology Information Internet Server (www.ncbi.nlm.nih.gov). DNA sequences presented in this work are deposited in the GenBank database under following accession numbers: AY082887 (repetitive region of adMRJP5), AY082888 (repetitive region of adMRJP3), AY082889 (region upstream of adMRJP3 repeat).

RESULTS

The repetitive region of the MRJP3 gene is polymorphic in *Apis dorsata*

To test for the presence of *A. mellifera* MRJP3 homologues in *A. dorsata*, genomic DNAs of *A. dorsata* workers from Borneo and Samui Island have been isolated and used in PCR reactions employing *A. mellifera* primers (see Materials and Methods, primer #1 and #2). The PCR yielded fragments larger than 550 bp, so that these fragments are significantly larger than the respective fragments in *A. mellifera*, which are about 450 bp long (fig. 1). The PCR fragments of *A. dorsata* exhibited intraspecific length variation as do those of *A. mellifera* (fig. 1, fig. 2A).

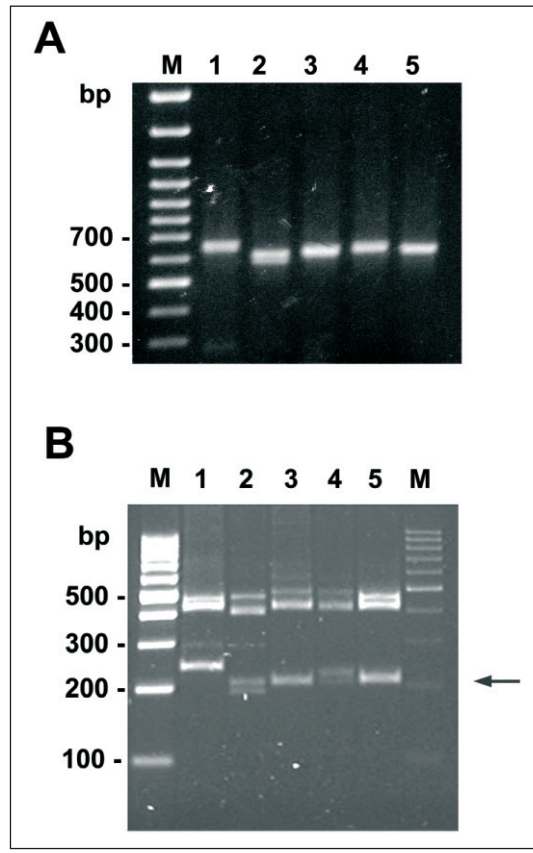

FIG. 2. Detection of adMRJP3 polymorphism by agarose gel electrophoresis. (A) Repetitive regions of MRJP3 of five *Apis dorsata* workers were amplified by PCR and electrophoresed on a 2% agarose gel. (B) aliquots of PCR reactions were digested with *BstEII* restriction endonuclease and electrophoresed on a 3% agarose gel. Polymorphic alleles around 200 bp long are indicated by an arrow.

The four alleles originating from the two Borneo *A. dorsata* individuals were cloned and sequenced. Figure 3 shows the DNA sequence and translation of one MRJP3 repeat region of *A. mellifera* (Albert *et al.*, 1999b) and *A. dorsata*. The translational reading frame was chosen congruent to *A. mellifera*. The region upstream of the repeat (not shown) shows clear homology with that of *A. mellifera* MRJP3 and the repeat itself shares all the properties described for *A. mellifera* MRJP3 repeats (Albert *et al.*, 1999b):

- A. The basic repeat unit is 15 nucleotides long and encodes a pentapeptide with the consensus sequence (K/R/N)QN(A/D/G)(D/N).
- B. The repetitive region can be divided into two contiguous segments differing slightly in their sequence

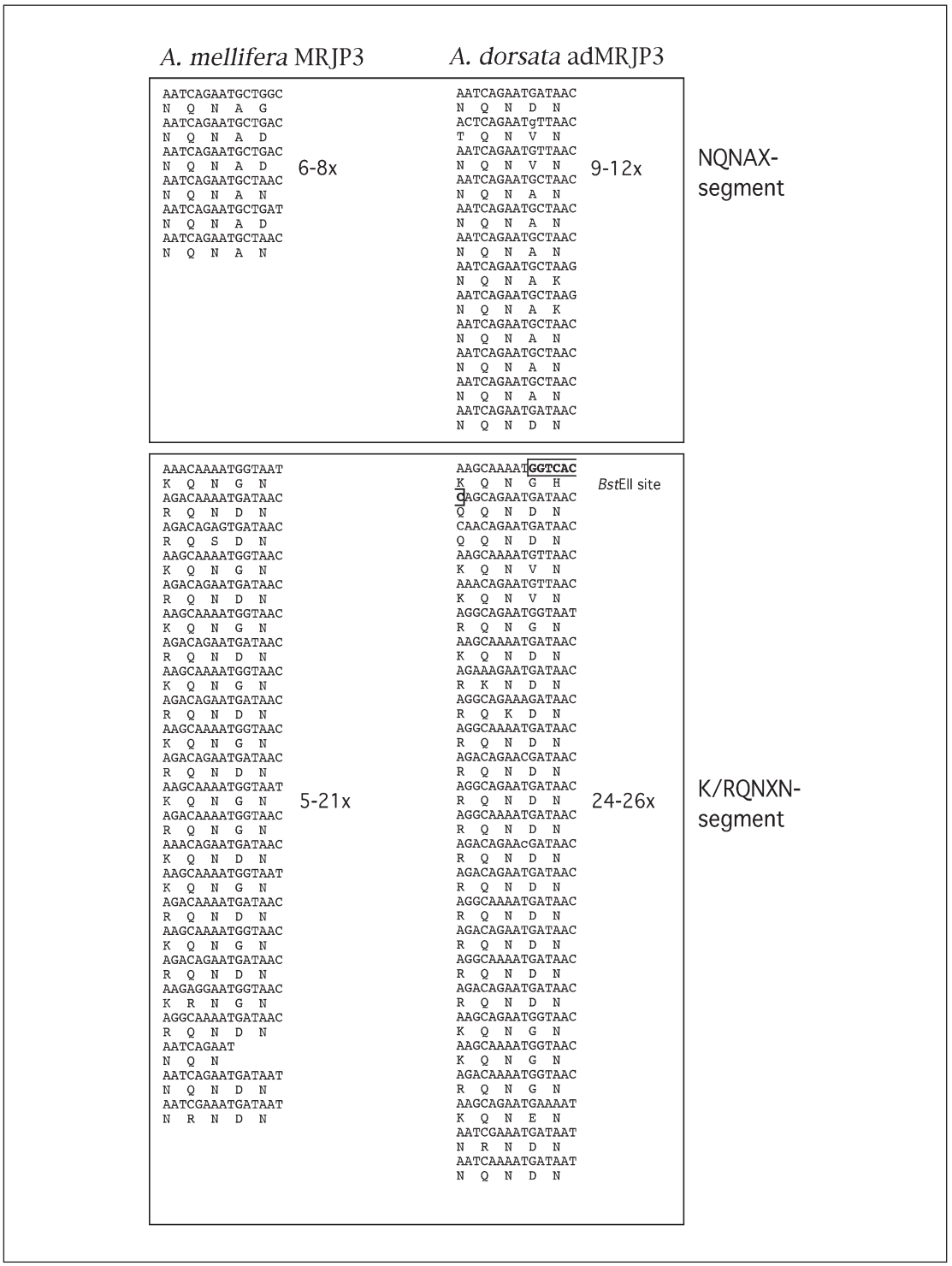

FIG. 3. Comparison of nucleotide and inferred amino acid sequences of an MRJP3 repetitive region of *Apis dorsata* and *A. mellifera*. A row length was chosen, which corresponds to the length of the basic repeat unit (15 nucleotides). The boxes separate two segments differing in their properties as described in the text. Bold characters indicate the recognition sequence of *Bst*EII restriction endonuclease, which occurs in *A. dorsata* but not in *A. mellifera*.

(see fig. 3). The N-terminal and more conserved segment encodes multiples of an NQNAX pentapeptide. Except for the initial three, each repeat unit of this segment contains a recognition sequence for *Bsm*I restriction endonuclease. The second segment has a more restricted consensus sequence (K/R)QN(D/G)N. Each unit begins with a positively charged amino acid.

- C. Polymorphic alleles differ in size because of the variable number of basic repeat units in their sequence (33–38 fold).

In spite of the similarities described above, significant differences have been detected between repeats of *A. mellifera* and *A. dorsata*. The repeat regions of *A. dorsata* (fig. 3) are longer (34–38 units; 510–570 bp) than those of *A. mellifera* (11–29 units; 165–435 bp). Their sequence is less conserved: units #1, #2, #3 and #12 (NQNDN, TQNVN, NQNVN, NQNDN) differ from the consensus sequence (NQNAX) in *A. dorsata* but not in *A. mellifera*. Similarly, units #2 and #3 in the second segment begin with glutamine instead of arginine/lysine. There is one exception from the XQNXX consensus found in all five *A. dorsata* alleles analysed, where glutamine is replaced by arginine. No interrupted or incomplete repeats as described in *A. mellifera* (Albert *et al.*, 1999b) were found in *A. dorsata* (see also fig. 3).

The genome of *A. dorsata* contains a close orthologue to the *A. mellifera* MRJP3 gene

To characterize the region 5' upstream of the adMRJP3 repeat region, we used primers #3 and #4 (see Materials and Methods) derived from *A. mellifera* sequences. We amplified a specific fragment of about 900 bp whose orthologous fragment in *A. mellifera* cDNA is about 280 bp. The DNA sequence of the 900 bp fragment, as well as the translation product, were subjected to a BLAST search (Altschul *et al.*, 1997) in the GenBank non-redundant database. High similarity with *A. mellifera* MRJP3 could be confirmed (fig. 4). The homologous region was interrupted by a 595 bp sequence different from all known MRJP cDNAs. A detailed analysis of the flanking regions revealed potential splicing sites (GT-AG rule; Mount, 1982) suggesting that the inserted sequence is intronic. The splicing product is a continuous sequence coding for a protein with high homology to MRJP3 (fig. 4). Finally, the intron in the MRJP3 gene of *A. dorsata* is located at exactly the same position as in *A. mellifera* genomic DNA (fig. 4, our unpublished data).

In summary, our data strongly suggest that the genome of *A. dorsata* contains a closely related orthologue to the *A. mellifera* MRJP3 gene, which also contains a polymorphic repetitive region.

The adMRJP3 repeat is a highly informative locus for population studies of *A. dorsata*

The adMRJP3 repeat is polymorphic in *A. dorsata* (see figs 1 and 2). Similar to that of *A. mellifera*, individual alleles differ in the number of basic pentadecanucleotide units. However, due to the larger size of the PCR products it is more difficult to separate individual alleles that differ by multiples of 15 bases (compare *A. dorsata* and *A. mellifera* lanes in fig. 1). In order to make the size differences in alleles more apparent, we used the conserved singular *Bst*EII restriction site (see fig. 3, framed characters) that cuts inside the repeat region of adMRJP3, resulting in two smaller, easily distinguishable fragments. PCR reactions were carried out using genomic DNAs of a total of five *A. dorsata* individuals. The PCR products were separated on a 2% agarose gel

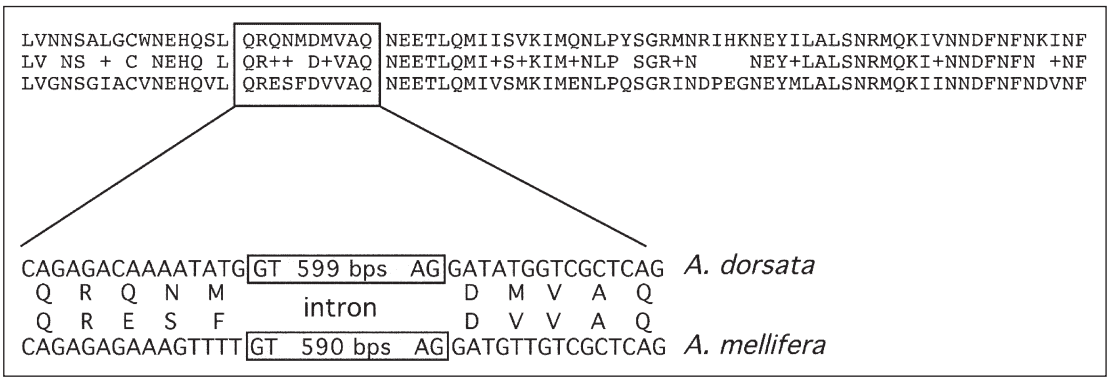

FIG. 4. Similarity between MRJP3 genes and proteins of *Apis mellifera* and *A. dorsata*. Regions upstream of the repetitive motifs of MRJP3 were amplified by PCR, cloned and subsequently sequenced. The upper part shows the BLASTA comparison of the *A. dorsata* gene product with that of *A. mellifera*. In the lower part, regions of nucleotide sequences around the presumptive introns are shown. Apart from high homology of nucleotide sequences and inferred proteins (71% identity, 84% similarity, BLASTA), the introns are positioned at the same site and have the same phase.

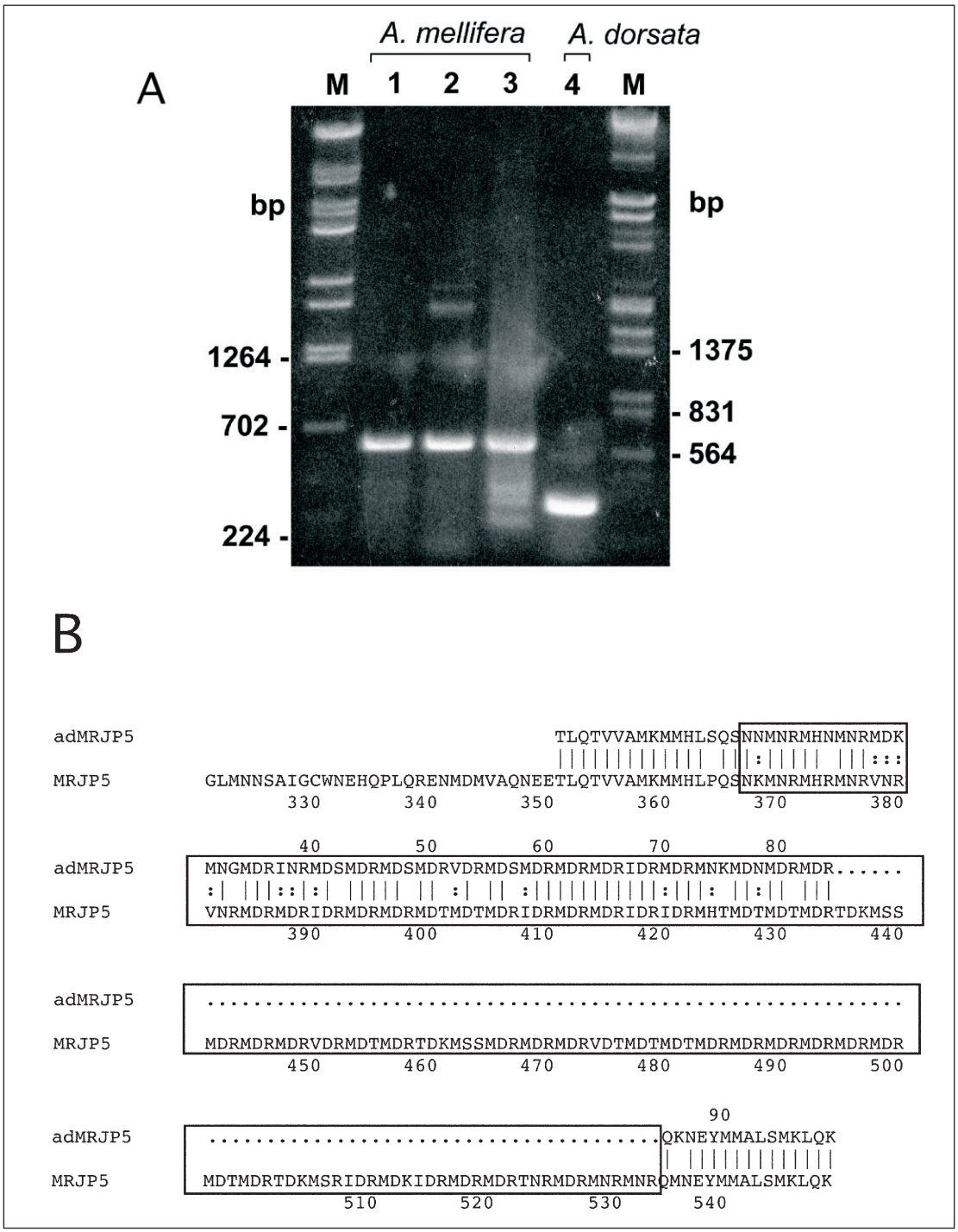

FIG. 5. Molecular characterization of the regions around the DRM repeat of MRJP5. (A) Plasmid bearing an MRJP5 cDNA (lane 1), genomic DNAs of two *Apis mellifera* (lanes 2 and 3), and genomic DNA of *A. dorsata* (lane 4) were PCR amplified with primers #5 and #6 and electrophoresed. The *A. dorsata* fragment was cloned and sequenced. Lanes marked M contain lambda/*Bst*EII and lambda/*Eco*RI+*Hind*III molecular weight markers respectively. (B) Comparison of MRJP5 amino acid sequences of *A. mellifera* and *A. dorsata*. Note the length polymorphism between the two MRJP5 repetitive regions (boxes).

(fig. 2A). An aliquot of the PCR reactions was digested with *Bst*EII, electrophoresed in 3% agarose gel (fig. 2B). Polymorphic DNA fragments appear in the ranges of 200 and 400 bp respectively. The resolution of the cleaved PCR products is much higher than that of non-cleaved ones (compare figs 2A and 2B). As an example, the nearly identical banding pattern of bees 1 and 4 could be distinguished clearly after restriction with *Bst*EII. Overall, four different banding patterns of small fragments were found after *Bst*EII digestion in the sample of five randomly collected individual *A. dorsata* bees. This indicates that the repeat of adMRJP3 is highly polymorphic and therefore highly suitable for population studies.

A homologue of another MRJP exists in *A. dorsata*

A phylogenetic comparison of MRJP proteins suggested that the MRJP proteins were created by nearly simultaneous multiplication of an ancestral gene (Albert *et al.*, 1999a). Therefore, also genes homologous to other *A. mellifera* MRJP genes are expected to be present in *A. dorsata*. To test this hypothesis, primers #5 and #6 (Materials and Methods) designed originally to amplify a region of the MRJP5 gene in *A. mellifera* containing the DRM repeat (Albert *et al.*, 1999a) were used in *A. dorsata*. A specific product of about 300 bp was obtained, which is about half the size of that of *A. mellifera* (fig. 5A). A sequence analysis confirmed high homology to *A. mellifera* MRJP5 cDNA. Moreover, a repetitive region of the same properties is located at the same position as that in *A. mellifera* (see fig. 5B). The fragment size difference between *A. mellifera* and *A. dorsata* mentioned above is caused by reiterated basic nonanucleotide repeat motifs, occurring 58 and 23 times, respectively.

DISCUSSION

MRJP genes of *Apis dorsata*

This study shows that orthologues of *A. mellifera* MRJP3 and MRJP5 genes, both including repetitive sequence motives, exist in the distinct honey bee species *A. dorsata*. The oldest fossil *dorsata*-type bees are dated back to the early Miocene (22 million years ago, see Willis *et al.*, 1992). Consequently the common origin of the MRJP family identified in *A. mellifera* and *A. dorsata* is at least 22 million years old.

Repetitive regions of adMRJP3 and adMRJP5

The repeat regions of five adMRJP3 alleles in *A. dorsata* show striking similarity to repeats of the *A. mellifera* MRJP3 gene (for an example see fig. 3). Overall build-up and high sequence homology indicate that the repetitive region was present before *A. mellifera* and *A. dorsata* diverged from their common ancestor. The same holds true for the DRM repeat of the MRJP5 gene. An interesting finding is that like in *A. mellifera*, the number of reiterated motifs is highly polymorphic in adMRJP3. In the limited sample of bees investigated, the size of repeat region in alleles of the *A. dorsata* adMRJP3 was generally larger than that in *A. mellifera*. On the one hand selection pressure appears to exist which has maintained an exonic repetitive structure for probably more than 22 million years and across species boundaries. On the other hand the intraspecific, highly variable number of reiterated repetitive units seems to evolve more neutrally. We suggest gene conversion and replication slippage as possible mechanisms in generating new alleles.

Possible reasons for the high degree of polymorphism of the repetitive region of MRJP3 have been discussed previously (Albert *et al.* 1999a). However, easy detection and relatively easily discernible alleles — especially after digestion with *Bst*EII restriction endonuclease — render the adMRJP3 repetitive region a very attractive tool for easy and cost-effective genotyping of *A. dorsata* individuals, requiring only a PCR thermocycler and an agarose gel electrophoretic apparatus. In comparison, microsatellite methods employ denaturing polyacrylamide sequencing gels to detect polymorphic alleles (Moritz *et al.*, 1995; Neumann *et al.*, 2000). However, we are aware that for precise population genetic studies, the information on several polymorphic alleles need to be combined.

Polymorphic MRJP3 genes give rise to polymorphic MRJP3 proteins in *A. mellifera*. Polymorphic MRJP3 proteins can be detected by SDS polyacrylamide gel electrophoresis of royal jelly, the hypopharyngeal glands, or head extracts (Albert *et al.*, 1999b; Kubo *et al.*, 1996). Whether MRJP genes are also components of *A. dorsata* royal jelly remains to be established. The close relationship between MRJP proteins in *A. mellifera* and *A. dorsata* reported here calls for further analysis of royal jelly proteins within the genus *Apis*.

Acknowledgements

We are grateful to Per Kryger and Duangporn Sihanuntavong for kindly providing *A. dorsata* bees. Our thanks go to Dieter Gallwitz for enabling us to perform the experiments in his laboratory. We are also grateful to Jana Schmitzova and Michael Haberl for critically reading the manuscript, and Kira Gee for revising the English text.

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(2): 290–297.

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(5): 427–434.

ALTSCHUL, S F; MADDEN, T L; SCHÄFFER, A A; J., Z; ZHANG, Z; MILLER, W; LIPMAN, D J (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389–3402.

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(26): 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(3): 611–613.

FONDRK, M K; PAGE, R E; HUNT, G J (1993) Paternity analysis of worker honey bee using random amplified polymophic DNA. *Naturwissenschaften* 80: 226–231.

GEYER, P K; SPANA, C; CORCES, V G (1986) On the molecular mechanism of gypsy-induced mutations at the yellow locus of *Drosophila melanogaster*. *EMBO Journal* 5(10): 2657–2662.

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(1): 22–26.

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.) coding for related proteins of royal jelly. *Journal of Apicultural Research* 33(2): 105–111.

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* 119(2): 291–295.

MORITZ, R F A; KRYGER, P; KOENIGER, G; KOENIGER, N; ESTOUP, A; TINGEK, S (1995) High degree of polyandry in *Apis dorsata* queens detected by DNA microsatellite variability. *Behavioral Ecology & Sociobiology* 37(5): 357–363.

MOUNT, S M (1982) Catalogue of splice junction sequences. *Nucleic Acid Research* 10: 459–472.

NEUMANN, P; KOENIGER, N; KOENIGER, G; TINGEK, S; KRYGER, P; MORITZ, R F A (2000) Home-site fidelity in migratory honeybees. *Nature* 406(6795): 474–475.

OLDROYD, B P; SMOLENSKI, A J; CORNUET, J M; WONGSIRI, S; ESTOUP, A; RINDERER, T E; CROZIER, R H (1996) Levels of polyandry and intracolonial genetic relationships in *Apis dorsata* (Hymenoptera, Apidae). *Annals of the Entomological Society of America* 89(2): 276–283.

REMBOLD, H; HANSER, G (1964) On the royal jelly of honey bees. 8. Demonstration of the determining principle in the jelly of queen bee larvae. *Hoppe-Seylers Zeitschrift der Physiologische Chemie* 339(1): 251–254.

SCHMITZOVA, J; KLAUDINY, J; ALBERT, S; SCHRODER, 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(9): 1020–1030.

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(3): 331–342.

WALTER, M F; BLACK, B C; AFSHAR, G; KERMABON, A-Y; WRIGHT, T R F; BIESSMANN, H (1991) Temporal and spatial expression of the yellow gene in correlation with cuticle formation and DOPA decarboxylase activity in *Drosophila* development. *Developmental Biology* 147: 32–45.

WILLIS, L G; WINSTON, M L; HONDA, B M (1992) Phylogenetic relationships in the honeybee (genus *Apis*) as determined by the sequence of the cytochrome oxidase II region of mitochondrial DNA. *Molecular Phylogenetic Evolution* 1(3): 169–178.